

# 150 Years Journal of Economics and Statistics

Edited by

Wolfgang Franz and Peter Winker

With Contributions by

Aus dem Moore, Nils, RWI (Berlin)

Burret, Heiko T., Walter Eucken Institut  
und Universität Freiburg

Dinsenhacher, Natalie, Statistisches  
Bundesamt Wiesbaden

Egeler, Roderich, Statistisches Bundesamt  
Wiesbaden

Feld, Lars P., Walter Eucken Institut  
und Universität Freiburg

Feuerstein, Switgard, University  
of Heidelberg and GEP, University  
of Nottingham

Issing, Otmar, Goethe University  
of Frankfurt

Kendzia, Michael J., Institute for the  
Study of Labor GmbH (IZA), Bonn

Kleber, Birgit, Statistisches Bundesamt  
Wiesbaden

Köhler, Ekkehard A., Walter Eucken  
Institut und Universität Freiburg

Schmidt, Christoph M., RWI (Essen),  
Ruhr-Universität Bochum, IZA (Bonn),  
and CEPR (London)

Von der Lippe, Peter, Universität  
Duisburg-Essen

Wieland, Volker, Goethe University  
of Frankfurt

Zimmermann, Klaus F., Institute for the  
Study of Labor GmbH (IZA), Bonn

## Guest Editors

Prof. Dr. Dr. h.c. mult. Wolfgang Franz  
Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW) Mannheim  
L 7, 1  
68161 Mannheim  
franz@zew.de

Prof. Dr. Peter Winker  
FB 02 – Statistik und Ökonometrie  
Justus-Liebig-Universität Gießen  
Licher Straße 64  
35394 Gießen  
peter.winker@wirtschaft.uni-giessen.de

## Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar

ISBN 978-3-8282-0583-3

© Lucius & Lucius Verlagsgesellschaft mbH · Stuttgart · 2013  
Gerokstraße 51, D-70184 Stuttgart

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen und Mikroverfilmungen sowie die Einspeicherung und Verarbeitung in elektronischen Systemen.

Satz: inmedialo, Digital- und Printmedien UG, Plankstadt  
Druck und Bindung: Neumann Druck, Heidelberg  
Printed in Germany

## Inhalt / Contents

Guest Editorial . . . . .	260-265
---------------------------	---------

### Abhandlungen / Original Papers

<i>Aus dem Moore, Nils, Christoph M. Schmidt</i> , On the Shoulders of Giants: Tracing Back the Intellectual Sources of the Current Debate on “GDP and Beyond” to the 19th Century . . . . .	266-290
<i>Burret, Heiko T., Lars P. Feld, Ekkehard A. Köhler</i> , Sustainability of Public Debt in Germany – Historical Considerations and Time Series Evidence . . . . .	291-335
<i>Von der Lippe, Peter</i> , Recurrent Price Index Problems and Some Early German Papers on Index Numbers . . . . .	336-366
<i>Feuerstein, Switgard</i> , From the Zollverein to the Economics of Regionalism . . . . .	367-388
<i>Egeler, Roderich, Natalie Dinsenhacher, Birgit Kleber</i> , The Relevance of Census Results for a Modern Society . . . . .	389-405
<i>Kendzia, Michael J., Klaus F. Zimmermann</i> , Celebrating 150 Years of Analyzing Fertility Trends in Germany . . . . .	406-422
<i>Issing, Otmar, Volker Wieland</i> , Monetary Theory and Monetary Policy: Reflections on the Development over the last 150 Years . . . . .	423-445

### Buchbesprechung / Book Review

<i>Holub, Hans-Werner</i> , Eine Einführung in die Geschichte des Ökonomischen Denkens . . . . .	446
---	-----

## Guest Editorial

With this special issue, the “Jahrbücher für Nationalökonomie und Statistik” celebrate 150 years of publishing. To the best of our knowledge, the “Jahrbücher” are, hence, the oldest journal in Germany devoted specifically and exclusively to economics and statistics.<sup>1</sup> The first editor was Bruno Hildebrand who died in 1878. Nearly at the same time the “Gustav Fischer Verlag” became publisher of the “Jahrbücher”. In 1996 that publisher passed his publishing activities in economics over to the “Lucius & Lucius Verlagsgesellschaft” newly founded by a great-grandson of Gustav Fischer. In order to stress that the “Jahrbücher” are relevant for an international audience the German title, although retained, was complemented by “Journal of Economics and Statistics” in 1999 and, nowadays, most papers are written in English. This is, however, not a *conditio sine qua non*, i.e., still also contributions in German are refereed and published solely depending on their quality.

The characteristics of an academic journal should mirror the progress made in the field of science under consideration, which might not always be a simple monotonically increasing function. In 2011, the guest editors published a call for papers for a special issue to celebrate the occasion of the 150th anniversary of the “Jahrbücher”. The aim was attracting papers establishing links from articles published during the early years of the journal to the current scientific discussion in economics and related fields. To this end we also approached several colleagues directly, and to our pleasant surprise, all people contacted by us agreed more or less spontaneously to submit such a paper. We consider this as an appraisal of this journal and would like to express our deep gratitude to them. Obviously, it was not always an easy task to link this historical discussion to actual developments in economics. Fortunately, a large number of reviewers provided their assistance to cut a crystal out of the submitted papers eventually accepted for publication in this special issue. A short overview of the papers collected in this volume after having gone through the standard refereeing procedure of the journal, will be given at the end of this short editorial.

Before turning to these articles, some more general remarks on the history of the “Jahrbücher” may be in order. The first reason is that such a historical perspective teaches modesty. Many ideas, now standard in economics or recently put forward as the latest fashion, have already been stated and published by our predecessors some considerable number of decades ago. It goes without saying, that most papers published in the first decades of the “Jahrbücher” did not contain much algebra, if some at all, let alone econometrics. But in those papers much effort is undertaken in collecting data, preparing adequate tables, and interpreting them carefully, with much common sense and, occasionally, with some speculative arguments. Clearly, not all topics discussed at that time are still on the top agenda of current research papers, but many are still discussed such as issues in agriculture or railway systems and others are just coming back on the agenda as the contributions in this special issue demonstrate. Reading those early papers is, therefore, highly informative but, on the other hand, also frustrating: almost everything has been said already long ago.

---

<sup>1</sup> The „Zeitschrift für die gesamte Staatswissenschaft“ (now „Journal of Institutional and Theoretical Economics“) was founded in 1844. But as has been pointed out by Lütge (1963), the „Zeitschrift“ was merely devoted to issues such as constitutional law, international law, politics, history of state, and much less to economics and statistics.

A second reason for taking a historical perspective is to look at the development of the journal itself. What was, in 1863, the motive to establish a journal, how did the former editors organize the journal, and what has been changed looking at the “Jahrbücher” of the year 2013?

To begin with the first question, in the preface of the first issue editor Bruno Hildebrand described the purpose of the “Jahrbücher” as follows. They should accompany and stimulate progress in scientific knowledge in the fields of economics and statistics. This is not so different what a contemporary editor of a similar journal would write nowadays. Moreover, Hildebrand goes on to claim, the “Jahrbücher” should also examine whether all great economic thoughts and reforms, which take place in the life of European people, are well placed with respect to their historic context and scientific justification. This also sounds rather familiar, especially when we translate that into “devoted to actual economic policy” reminding us of the battles between Keynesians and monetarists in the 1970s. However, at the end of his preface Hildebrand clearly states that he is an opponent of “French centralism and governmental paternalism” but strongly prefers the British economic model of “full liberty, self-administration and self-responsibility of people”. He promises to stay with this opinion in the “Jahrbücher” deliberately. But he adds that he will accept papers expressing opinions different from his own, provided that these are justified seriously and meet methodological standards. Despite this concession it remains unclear how many supporters of the mentioned French model refrained from submitting a paper to the “Jahrbücher” at that time. Presumably no editor today would himself come out like this because he is running the risk of a type 2 error, i.e., not to publish a highly qualified and influential paper.

Be that as it may, Hildebrand organized the “Jahrbücher” along four sections, namely scholarly papers, new developments in legislation, review articles on notable books and articles in European journals, and short notes on economic and statistical facts including short papers and communications. Perhaps with the exception of legislation, this arrangement is not far away from other journals edited by national economic associations (not existing at that time in Germany), possibly distributed among several journals like the AER, JEL and JEP. However, research papers have gained an overwhelming predominance today including contemporary issues of the “Jahrbücher”.

While being a success story after all, the historic development of the “Jahrbücher” is not free of turbulences. The success can be documented with the names of very influential authors such as von Böhm-Bawerk, Eucken, Gumbel, Hayek, Laspeyres, Menger, von Nell-Breuning, Schmoller, von Stackelberg, and Wicksell, to mention only a few. It might be remarkable that already back in the 19<sup>th</sup> and early 20<sup>th</sup> century the Journal attracted some international contributions, including, e.g., a paper by Knut Wicksell (1897), which became published in German language at that time. A first crisis occurred in the 1870s. According to Lütge (1963), these years experienced an alarming decline of manuscript submissions so that the editor saw himself forced to publish all offered material, publishable or not. Consequently, the number of subscribers went down and it was due to the efforts of the new editor of the “Jahrbücher”, Johannes Conrad, who took sole responsibility in 1878 that the turnaround succeeded. Another crisis occurred after 1933 due to serious conflicts with the Nazi regime and, after World War II, with the Soviet occupying power. Both crises resulted in interruptions of publishing.

During the first decades, the term “statistics” in the title of the journal related mainly to issues of how to define appropriate quantitative indicators in economics, e.g., for

changes in the overall price level, and to the distribution of actual data from different spheres of the German economies in absence of a national state, not to speak from a national office of statistics. With the establishment of public infrastructure for the collection and publication of economic data, these issues became less relevant for the journal. Thus, after 100 years of existence, it was even discussed to delete “statistics” from the title “Journal of Economics and Statistics”. The editors refused to follow this advice bowing to tradition. They were right. In fact, statistics did not suffer from a loss of importance as it is evidenced by the ongoing discussion about an appropriate measurement of price movement, starting from Laspeyres price indices to hedonic prices and perceived inflation, or the discussion about measurement of economic well being beyond GDP. Furthermore, statistics in the form of econometrics became one of the major tools of economic research and is a relevant part, if not the main content of many papers in the journal nowadays.

Taken together, we might state that the “Jahrbücher” do have their merits, and that the founding father, Bruno Hildebrand, has rendered outstanding services to economics and statistics at least in Germany. He and his successors offered economists and statisticians an academic forum to publish their work. Moreover, from the beginning the “Jahrbücher” have emphasized the importance of economic history and the relevance of cultural science. The implications of an ethic responsibility of economic science have been a topic during 150 years. Finally, the “Jahrbücher” have highlighted the relevance of empirical research and the importance of adequate data from their first issues.

In 2063, will there be still editors of a journal named “Jahrbücher” to publish an editorial for the 200<sup>th</sup> anniversary of the journal? Hopefully, yes. Maybe, all papers will be written in English and the name “Jahrbücher” is fully replaced by “Journal of Economics and Statistics”. Possibly, there will not be a printed version anymore, but just an online-version with open access. If so, what will the editors of the 200 year volume write in their editorial? How will they judge upon the upcoming, then past 50 years? It will depend on whether the editors stay with the basic principles for the “Jahrbücher”: accept innovative papers only, put emphasis on papers which combine theoretical and empirical issues, encourage submissions of papers which deal with statistics and economic history, and, last but not least, take care of a speedy referee process. Together with the efforts of the publisher, the editors, the scientific board, and the referees we estimate a fairly high probability that in 2063 someone will write a 200 years editorial and, on this occasion, will read that of 50 and 100 years ago as well as ours. Therefore, great thanks to all who helped the “Jahrbücher” and to those who will do so in the future!

Let us now turn to the content of this special issue devoted to 150 years of existence of the “Jahrbücher”. All seven contributions finally accepted for publication share a common property. By presenting a specific research topic, they all refer to research published in the journal during the first decades of its existence and provide a sketch of the further development of economic thinking and empirical assessment up to current debates. Thus, it is not obvious on how to order these contributions addressing quite different topics. Eventually, we decided to present them in an order starting with those providing the earliest reference to a paper published in the “Jahrbücher”.

The first paper by Nils aus dem Moore and Christoph M. Schmidt (2013) stresses in its title the close link between research done in the early years of the journal’s activities and current debates: “On the Shoulders of Giants: Tracing Back the Intellectual Sources of the Current Debate on “GDP and Beyond” to the 19th Century”. In fact, a substantial amount of publications in the Journal starting in the 1860s concentrated on how to

measure economic activity and on providing actual numbers. The authors refer to these early sources and point out that besides the search for methods to obtain reasonable numbers (keep in mind that there was no system of national accounts at that time, not even the concept of GDP was established as such), the researchers have also discussed the limitations of different concepts of measuring economic activities. In particular, topics such as “sustainability” have been formulated in the *Journal* as early as in Eggert (1883) in the context of forestry. However, the contribution does not only provide traces to the origins of some ideas, but also presents an expert view on the current debate on extending measurement systems for taking into account aspects beyond GDP, i.e., it opens the horizon for future developments in this field.

While measuring economic activity and well being is a topic of permanent interest, the issue of public debt crises seems to appear on the scene of economic debates just in the aftermath of the most recent financial market crisis. However, as the contribution on “Sustainability of Public Debt in Germany – Historical Considerations and Times Series Evidence” by Heiko T. Burret, Lars P. Feld, and Ekkehard A. Köhler (2013) demonstrates, the scientific debate about the sustainability of public debt is at least as old as the *Journal*. In fact, contributions to this topic can be found in many issues over the full 150 years period from the discussion about the financing of the American War of Independence in Hildebrand (1863) to the discussion about the TARGET II balances in Spahn (2012). The contribution not only sketches this debate in the “*Jahrbücher*”, but also provides a historical overview on the actual development of public debt from the 1850s to 2010 – following the argument of the first editors that economic science also requires a proper analysis of the historical context. Obviously, theories and empirical methods regarding public debt have developed over time: The cointegration analysis presented in the paper has not been on the agenda yet 150 ago.

The first two contributions contain references to papers in the “*Jahrbücher*” dating back to 1863. In contrast, the discussion on measuring price levels and inflation rates seems to start in the *Journal* only in 1864 by a first contribution of Laspeyres (1864), i.e., seven years before his famous paper on the price index formula later becoming known as “Laspeyres price index” (Laspeyres 1871). Starting with these early contributions, Peter von der Lippe (2013) in his paper on “Recurrent Price Index Problems and Some Early German Papers on Index Numbers Notes on Laspeyres, Paasche, Drobisch and Lehr” traces back the history of price index formulae in the “*Jahrbücher*”. Although the price index formulae according to Laspeyres, Paasche and Drobisch are still in common use and part of standard statistical procedures, not all practitioners might be aware of the fact that these concepts have been introduced first in papers published in the *Journal of Economics and Statistics* in the 1870s, followed by some controversy about the pros and cons of the methods for specific settings. Peter von der Lippe’s contribution sketches these developments over more than a century. In fact, the *Journal* continues to publish papers devoted to price index theory and inflation measurement until today, e.g. in a recent special issue (Diewert/von der Lippe 2010).

Also dating back to references in 1864, Switgard Feuerstein’s (2013) contribution “From the Zollverein to the Economics of Regionalism” describes the discussion in the mid 19<sup>th</sup> century about tariff unions preceding the founding of the German State in 1871. In a series of papers starting with Fischer (1864), the author shapes the perception of the tariff union “Zollverein”. Also the idea of preferential trade agreements are traced over time making analogies between the discussions about the not yet unified German states prior to 1871 and the European Union obvious. Feuerstein’s analysis demonstrates which of

the arguments put forward in the 1860s are still valid today and might even be used for an assessment of the state of the European Union.

Population censuses appear to be almost as old as written history. However, they became a major source to official statistics, in particular with regard to the national economy much later. The references to publications in the “Jahrbücher” provided in the contribution “The relevance of census results for a modern society” by Roderich Egeler, Natalie Dinsenhacher, and Birgit Kleber (2013) date back to two articles published by Fabricius (1866) and Meyer (1866). Their contribution describes the relevance of population censuses for political decision making. Furthermore, they describe how the technique of population censuses have changed from the traditional approaches in the 19<sup>th</sup> and 20<sup>th</sup> century to the current register based approach. It becomes evident that not only the census techniques have changed, but also concepts and definitions, e.g., regarding the concept of “population” did not remain the same over time.

The current discussion on the demographic transition in Germany and most European countries might also be considered in the context of a long scientific tradition dealing with demographic trends, their causes and effects, and forecasts of future developments. In their contribution “Celebrating 150 years of analyzing fertility trends in Germany” Michael J. Kendzia and Klaus F. Zimmermann (2013) provide a summary of this discussion in the “Jahrbücher”. They come up with more than twenty papers putting a specific focus on these issues. The authors argue that all relevant tendencies in demographic research have also been taken up in papers published in the “Jahrbücher”. As pointed out by the authors, this includes also some contributions reflecting the *Zeitgeist* during the time of National Socialism.

The last contribution to this special issue by Ottmar Issing and Volker Wieland (2013) on “Monetary Theory and Monetary Policy - Reflections on the development over the last 150 years” also refers to Laspeyres (1871) for the obvious reason that a useful measure of inflation is of central importance for any empirically founded monetary policy. Clearly, a German monetary policy could not start earlier given the founding date of the Deutsches Reich in the same year and the establishment of a German National Bank (“Deutsche Reichsbank”) as successor of the Prussian Bank only in 1876. The authors provide evidence that publications in the “Jahrbücher”, including the already mentioned paper by Wicksell (1897), contributed to the academic discussion about monetary theory and implication for monetary policy. Given the eventful history of central banking and monetary policy in Germany, a full coverage of 150 years of history – even when concentrating on publications in the “Jahrbücher” – is beyond the scope of a single article. Therefore, the authors concentrate on some aspects of the interaction between monetary theory and policy with economic policy making and the financial system following their discussion in the “Jahrbücher” and beyond over the last 150 years.

To conclude, the contributions in this special issue have put some spotlight on a few relevant issues in economics and statistics which have been on the agenda for a large part of the last 150 years in all industrialized countries. In fact, we might postulate that the “Jahrbücher” over the 150 years of their existence have been more than just a “German” journal – the contributions covered almost all those areas of economics and statistics which have been perceived as of general importance not only for the German economy making use of the models, methods and data available at the time. The contributions in this special issue also demonstrate that it might be worth to have a look back to our academic predecessors since topics such as public debt, demo-



graphic transition or the value of money have already been analyzed by them in the context of their time.

## References

- Aus dem Moore, N., C. M. Schmidt (2013), On the Shoulders of Giants: Tracing Back the Intellectual Sources of the Current Debate on “GDP and Beyond” to the 19th Century. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 266-290.
- Burret, H. T., L. P. Feld, E. A. Köhler (2013), Sustainability of Public Debt in Germany – Historical Considerations and Times Series Evidence. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 291-335.
- Diewert, W. D., P. von der Lippe (2010), Introduction to the Special Issue on Index Number Theory and Price Statistics. *Jahrbücher für Nationalökonomie und Statistik* 230(6): 660-672.
- Egeler, R., N. Dinsenhacher, B. Kleber (2013), The Relevance of Census Results for a Modern Society. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 389-405.
- Eggert, U. (1883), Die ersten Beziehungen zwischen Forstwirtschaft und Volkswirtschaft. *Jahrbücher für Nationalökonomie und Statistik* 40: 305-343.
- Fabricius, K. C. A. (1866), Ueber Volkszählungen. *Jahrbücher für Nationalökonomie und Statistik* 6: 305-323.
- Feuerstein, S. (2013), From the Zollverein to the Economics of Regionalism. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 367-388.
- Fischer, G. (1864), Ueber das Wesen und die Bedingungen eines Zollvereins: Die Idee eines deutschen Zollvereins und ihre Ausführung geschichtlich entwickelt. *Jahrbücher für Nationalökonomie und Statistik* 2: 317-385, 397-432.
- Hildebrand, B. (1863), Das Papiergeld der Vereinigten Staaten Nordamerikas während ihres Unabhängigkeitskrieges von 1775-1781. *Jahrbücher für Nationalökonomie und Statistik* 1: 392-396.
- Issing, O., V. Wieland (2013), Monetary Theory and Monetary Policy – Reflections on the Development over the last 150 Years. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 423-445.
- Kendzia, M. J., K. F. Zimmermann (2013), Celebrating 150 Years of Analyzing Fertility Trends in Germany. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 406-422.
- Laspeyres, E. (1864), Hamburger Waarenpreise 1851-1863 und die californisch-australischen Goldentdeckungen seit 1848 – Ein Beitrag zur Lehre von der Geldentwerthung. *Jahrbücher für Nationalökonomie und Statistik* 3: 81-118, 209-236.
- Laspeyres, E. (1871), Die Berechnung einer mittleren Warenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik* 16: 296-314.
- Lütge, F. (1963), Zum 100. Geburtstag der “Jahrbücher”. *Jahrbücher für Nationalökonomie und Statistik* 175: 1-11.
- Meyer, G. (1866), Das Princip der rechtlichen und der factischen Bevölkerung. *Jahrbücher für Nationalökonomie und Statistik* 6: 97-110, 425-432.
- Spahn, P. (2012), David Hume und die Target Salden. *Jahrbücher für Nationalökonomie und Statistik* 232: 460-481.
- Von der Lippe, P. (2013), Recurrent Price Index Problems and Some Early German Papers on Index Numbers – Notes on Laspeyres, Paasche, Drobisch and Lehr. *Jahrbücher für Nationalökonomie und Statistik* 233(3): 336-366.
- Wicksell, K. (1897), Der Bankzins als Regulator der Warenpreise. *Jahrbücher für Nationalökonomie und Statistik* 13: 228-243.

*Wolfgang Franz  
Peter Winker*

# On the Shoulders of Giants: Tracing Back the Intellectual Sources of the Current Debate on “GDP and Beyond” to the 19th Century

**Nils aus dem Moore\***

RWI (Berlin)

**Christoph M. Schmidt**

RWI (Essen), Ruhr-Universität Bochum, IZA (Bonn), and CEPR (London)

JEL A13; B16; E01

GDP; well-being; quality of life; sustainability; indicator systems.

Received: 21.06.2012

Revision received: 08.10.2012

Accepted: 13.10.2012

---

## Summary

An eternal motive of human existence is the search for guidance. While values and beliefs retain their high relevance, today's enlightened societies also tend to rest their aspirations and decisions on the actual facts and on a sober assessment of possible courses of events emerging from different choices. Given the complexity of modern life, it is by now well understood that this strategy requires objective, comprehensive and accessible statistical reporting. Today, the desire to provide such a valuable basis for individual decisions and policy-making finds one of its most important expressions in the international debate on “GDP and beyond”. In contrast to similar efforts displayed in previous decades, the current projects emphasize sustainability issues and focus on the accessibility of the information, using modern tools of measurement and presentation. Yet, there is ample evidence that even by the mid-19<sup>th</sup> century economists aspired to use the objectifying power of statistical analysis as an instrument to improve policy-making and to achieve societal progress. Many of the approaches entertained today have thus to be viewed as an extension of attempts started at that time.

## 1 Introduction

In these early years of the 21<sup>st</sup> century, the world is searching for new and better guidance. The second global recession that resulted from a financial and banking crisis in the United States, and the growing stock of evidence that the global economy puts substantially more pressure on the ecological systems of our planet than those are able to carry have contributed to a widespread conviction that “business as usual” is no longer a sensible option. The definite direction that the course of events should take is, naturally,

---

\* We are grateful to Andrea Cassel, Claudia Schmiedchen, and Hartmut Westram for their support, and to two anonymous referees and to the editor for their constructive comments and suggestions.

highly controversial. But one lesson from the global economic crisis and the multitude of ecological damages that are looming or already occurred seems to be widely accepted: A compass is needed that provides policy makers, economy and society with a reliable bearing in matters of economic performance, quality of life and sustainability.

Across the globe, governments, researchers and social activists are currently discussing better individual indicators for specific aspects of our present condition in material as well as in non-material dimensions. They are also working on their coherent integration into comprehensive but still accessible measurement systems that should form the basis for evidence-based public deliberation and policy-making. This movement benefits from recent advancements in statistical techniques and indicator research and is based on the widespread utilization of powerful and yet still improving information and communication technologies. The ultimate aim of these initiatives is the establishment of user-friendly statistical monitoring systems that are at the same time comprehensive and comprehensible.

But in a year in which the *Journal of Economics and Statistics* (*JNS – Jahrbücher für Nationalökonomie und Statistik*) completes its first 150 years, it seems worthwhile to look back and search for traces of this debate in earlier times. This article documents that this search turns out to be much more successful, though, than just detecting some vestiges of the current debate. Instead, many of the ideas and concepts discussed today were already present in the late 19<sup>th</sup> century. We document these intellectual roots meticulously, both with respect to economic reasoning and with respect to the potential of statistical methods to support this endeavor.

And in addition to the intriguing parallels, we also discuss the most important differences that distinguish the current and the past contributions to the issue. Most importantly, we find sustainability issues to have conquered much of the current attention, aspects which were arguably of a less urgent character in the late 19<sup>th</sup> century. Moreover, for the adept user, the potential to access a wealth of information is bigger than ever quite in contrast to the late 19<sup>th</sup> century statistician who could only have dreamed of these possibilities. This potential is creating a new challenge, though, since now the task is to utilize the technological capabilities to make the information accessible to a wider public.

The outline of the paper is as follows. The second section presents an overview of the current status and the evolution of the international debate on “GDP and Beyond”, giving an outline of the guiding principles of this work and concrete examples of measurement systems that have been recently proposed. The third section traces the intellectual sources of the current debate back to the second half of the 19<sup>th</sup> century, focusing for obvious reasons on publications in the *JNS*. Becoming more concretely, section 4 documents for the three spheres of the current debate, i.e. for (i) material well-being and economic performance, for (ii) non-material well-being and quality of life, and for (iii) various aspects of sustainability, that many of the indicators currently being discussed, and the respective reasoning behind them, can be traced back to 19<sup>th</sup> century precursors debated in the early volumes of the *JNS*. Finally, the concluding fifth section draws some conclusions from this voyage into the history of economic thought.

## 2 “GDP and beyond”: Evolution and current status of the debate

Despite the great variety of new indicators and dashboards that have been devised to support policy-making and public deliberation, participants in the current debate on improving the measurement of societal progress and human welfare share a common point of departure. Their overwhelming consensus is that in recent decades policy makers, business leaders and economists, as well as the media and society at large, placed too much emphasis on short-run economic performance. Concentrating on measuring the growth rate of Gross Domestic Product (GDP), they neglected to adequately address the long-term. Although economists have pointed out since its inception during WWII that GDP is a measure of economic activity and not of economic well-being, apparently they did not do this fervently enough. Consequently, in the public perception it rose to a quasi-hegemonic status and was often interpreted as a gauge of the overall welfare of economies.

There are good reasons for the assessment that GDP alone is not sufficient to convey a reliable impression of a society's material wealth and non-material well-being. Inter alia this reservation reflects that (i) in the *economic domain* all non-market activities, such as housework and parenting or unpaid voluntary activities, and a substantial part of publicly provided services in education, health and security remain unaccounted for by GDP, (ii) in the *social domain* the distributions of income and wealth are not documented and inequalities of opportunity are not captured, (iii) in the *ecological domain* damages and losses as well as the consumption of non-renewable resources are not factored in adequately, and (iv) aspects of the *political and institutional domain* that are clearly relevant for overall welfare, like the extent of democratic participation, accountability and the rule of law are not considered whatsoever.

Notwithstanding the dominance of GDP in the public realm, the work on a broader framework for measuring human progress and well-being continued in the background. Scientific expert communities and international institutions have addressed these questions at least since the publication of “The Limits to Growth” by the Club of Rome in 1972 (Meadows et al. 1972). Starting with Nordhaus and Tobin (1972) and their calculation of a Measure of Economic Welfare, a wide array of alternative measurement frameworks and concrete indicators have been proposed, as is documented today in a vast and well-surveyed literature (Fleurbaey 2009). Perhaps the most prominent framework that has been designed to measure well-being, quality of life, human development and sustainable development in the last three decades is the Human Development Index which ranks countries by their level of “human development” through a composite indicator that takes GDP but also health and education into account (Hall et al. 2010: 7).

An important step towards a broader understanding of welfare and progress and the development of adequate indicators has been the establishment of sustainability as a guiding policy principle, at least in rhetorical terms. The famous Brundtland report (WCED 1987) defined “sustainable development” as the kind of development “that meets the needs of the present without compromising the ability of future generations to meet their own needs”. While this definition highlights the intertemporal aspect of the sustainability concept and focuses on the ethical requirement of intergenerational equity, most of the numerous implementations followed a more pragmatic interpretation, aiming at the reconciliation of environmental, social equity and economic demands which came to be known as the “three pillars” of sustainability. Recent implementations often

add governance as a fourth dimension that takes into account aspects of democratic participation, accountability and the rule of law.

By now, many advanced countries have established indicator systems along these lines. As a prominent example, the biennial Monitoring Report of Eurostat covers 111 indicators which are linked to the key challenges and objectives of the EU Sustainable development strategy (European Commission 2011). A consolidation of these numerous indicators into 11 so-called headline indicators and the visualization of their current trends allow a quick overview of the progress made towards the specific goals. Following the same approach, the German government has established a National Sustainability Strategy that is monitored biennially by a progress report released by the Federal Statistical Office. It contains 35 indicators that have been politically agreed upon to cover 21 subject areas (Statistisches Bundesamt 2012).

There is a second reason, in addition to the emergence of the sustainability paradigm, for the recent multiplication of efforts in many countries, by governments on all levels as well as in the civil society, to go “beyond GDP”. This is the growing body of evidence that the tight co-evolution of economic performance with other aspects of material and non-material welfare might have weakened considerably in recent decades, at least in some societies (GCEE/CAE 2010: 6). As long as economic performance and general welfare evidently marched in lock-step, the focus on GDP growth as a proxy for broader welfare was completely warranted. Yet today, ever mounting work pressure, reduced security of employment, the breakdown of family structures, rising poverty rates and the spread of mental-health problems and diseases of civilization have been identified as potential welfare reducing phenomena that are not captured in traditional measures, let alone by GDP (Wahl et al. 2010: 9).

The consequence is obvious: GDP alone is not sufficient to measure adequately the sustainable wealth of nations. This insight is confirmed by popular assessment. Two thirds of the respondents in a Eurobarometer poll of the European Commission said in 2008 that indicators for social, economic and ecologic dimensions should receive equal value in measuring progress. Governments and international organizations tied in with the popular dissatisfaction of traditional welfare measurement and started several initiatives towards the establishment of broader concepts. The Organisation for Economic Cooperation and Development (OECD) has without doubt given a decisive impetus to these endeavors through a series of high-level conferences in Palermo (Italy 2004), Istanbul (Turkey 2007) and Busan (South Korea 2009) and accompanying publications. Most importantly, the “Istanbul Declaration” acknowledged “an emerging consensus on the need to undertake the measurement of societal progress in every country, going beyond conventional economic measures such as GDP per capita” with the aim of producing “high-quality, facts-based information that can be used by all of society to form a shared view of societal well-being and its evolution over time” (OECD 2007).

Within Europe, the European Commission acted as an extra pacemaker. At the conference “Beyond GDP”, Commission President Barroso declared that, in measuring welfare and well-being, the EU must aim at “the sort of breakthrough that we saw in the 1930s, a breakthrough that adapts GDP, or complements it with indicators that are better suited to our needs today, and the challenges we face today” (cited in Kroll 2011: 16). The ensuing report “GDP and Beyond: Measuring Progress in a Changing World” (EU Commission 2009) contained a roadmap with the following five key steps towards a comprehensive and comprehensible reporting on progress: (i) complementing GDP with environmental and social indicators, (ii) providing near real-time information

for decision making; (iii) generating more accurate reporting on distribution and inequalities, (iv) developing a European Sustainable Development Scoreboard; and (v) extending National Accounts to environmental and social issues (Kroll 2011: 16).

The final breakthrough into the public view came with the publication of the so called Stiglitz-Sen-Fitoussi-Report (SSFC 2009). It was conducted by a commission compiled around the two Nobel prize winners Joseph Stiglitz and Amartya Sen, and the French economist Jean-Paul Fitoussi. The commission concluded that “those attempting to guide the economy and our societies are like pilots trying to steering a course without a reliable compass. [...] For many purposes, we need better metrics. Fortunately, research in recent years has enabled us to improve our metrics, and it is time to incorporate in our measurement systems some of these advances” (SSFC 2009: 9).

As an immediate follow-up to the SSFC report, the Franco-German Council of Ministers asked the German Council of Economic Experts (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, SVR) and its French counterpart, the Conseil d'Analyse économique (CAE) to develop a concrete proposal how the recommendations of the SSFC report could be implemented. The resulting report was published in December 2010 and proposed a comprehensive set of indicators to measure economic performance, quality of life and sustainability (GCEE/CAE 2010).

In both countries, the quest for a better measurement “beyond GDP” has continued since then with various activities: In France, the national statistical institute INSEE (Institut national de la statistique et des études économiques) has begun to enrich its surveys with questions to include new dimensions that were recommended in the SSFC report (Kroll 2011: 9). In Germany, the Bundestag established the Study Commission on “Growth, Wellbeing and Quality of Life – Paths to Sustainable Economic Activity and Social Progress in the Social Market Economy” that took up its deliberations in January 2011 (German Bundestag 2010). In its interim report the respective working group argued against the construction of a holistic composite index and proposed instead to work out a concrete proposal for an indicator suite that would be as large as necessary and as compact as possible, thereby aiming to strike the ideal balance between comprehensiveness and comprehensibility (Deutscher Bundestag 2012).

Similar activities are under way in many other countries, and a growing number of them has already completed the development phase of their indicator suites or composite indices and presented them to the public (Kroll (2011) provides a comprehensive overview). We restrict ourselves here to examples from four countries (Australia, Canada, UK, US) to document the breadth of different approaches that are currently pursued.

Australia is without doubt among the pioneers for a broader welfare measurement. As early as 1996, there was a Senate Inquiry into “National Wellbeing: A system of national citizenship indicators and benchmarks”, followed by a national conference on measuring progress in 1997 and the release of the publication “Measuring Progress: Is life getting better?” in 1998 (Eckersley 1998). Building on these initiatives, the Australian Bureau of Statistics (ABS) has published the indicator system “Measures of Australia's Progress” on a regular basis since 2002. The official brochure contains 17 headline indicators that are grouped into the three pillars Society, Economy and Environment (ABS 2010).

A second national initiative that has attracted international attention is the “Canadian Index of Well-Being” (CIW) that was started by scientists and social activists at the University of Waterloo in 2001. Ten years later, the first index brochure “How are Canadians *Really* doing?” was released (CIW 2011). The CIW is a composite indicator

calculated by two steps of aggregation: Firstly, 64 baseline indicators that depict the evolution of different aspects of quality of life between 1994 and 2008 are normalized to percentage values and summarized in indicators regarding eight distinct dimensions. Secondly, the overall CIW is calculated as the average of these dimension indicators. Although the authors are well aware of the weighting problem in the construction of their composite indicator (Michalos et al. 2011: 29-31), they nevertheless interpret the difference between the GDP growth of 31 per cent and a substantially smaller CIW rise of only 11 per cent as evidence that GDP overdraws the real gain in quality of life: "(...), our economic performance outpaces our quality of life." (CIW 2011: 12).

In the UK, more emphasis than anywhere else is placed on the integration of comprehensive measures of the country's subjective well-being into the official statistical apparatus. The heavy weight that economists around Sir Richard Layard have given to so-called happiness research at the London School of Economists and the apt popularization of its findings and potential policy implications (Layard 2005) laid the ground for this orientation. In November 2010, Prime Minister David Cameron launched a large-scale initiative for the introduction of a measure of "general well-being" (GWB) that should complement GDP, based on surveys of well-being. Starting in April 2011, 200 000 Britons have been asked the following four questions each year in the Integrated Household Survey, answering on a scale of 0 to 10: (i) "Overall, how satisfied are you with your life nowadays?", (ii) "Overall, how happy did you feel yesterday?", (iii) "Overall, how anxious did you feel yesterday?", and (iv) "Overall, to what extent do you feel that the things you do in your life are worthwhile?" (Matheson 2011).

In the United States, President Barack Obama signed the Key National Indicators Act into law in March 2010, which should lead to the creation of a Key National Indicator System (KNIS). Congress authorized 70 Million US-Dollars of public funding for the project that is carried out by the interdisciplinary National Academy of Sciences in collaboration with the newly founded non-profit institute "State of the USA". The ambition of KNIS is to provide US citizens with statistical information in their country and region on a broad variety of issues in an accessible way on a user-friendly website. The comprehensive system will eventually include around 300 individual indicators to cover the 14 topic areas, even on the disaggregated levels of states, regions and social subgroups (Kroll 2011: 10f.).

### 3 Intellectual sources of the current debate in the 19th century

Across the globe we currently find ambitious projects that are aiming for a comprehensive and internationally comparable account of the state of economic performance, well-being and quality of life. These initiatives have arguably led to remarkable progress in the quest of statistical analysis to generate deeper insights and to pave the ground for better policy. Yet, they are certainly not the starting point of this endeavor, but rather its most recent expression. Indeed a recurring theme in the current debate regarding the design of indicator systems and the organization of the statistical groundwork behind their publication is the insight that many of the aspects that capture our attention today have been discussed before, sometimes even for several decades. Examples are the social indicator systems devised in the 1960s and 1970s and the indicator systems for environmental sustainability that have been around for more than a decade.

But, what might perhaps come as a very surprising discovery for the protagonists of the current debate, its intellectual sources can be traced backed much further, at least



to the second half of the 19<sup>th</sup> century. To provide convincing evidence for this claim, our focus in this article will selectively lie on contributions made to the *JNS* during its first decades of existence, along the lines of the following thought experiment: If some of the early contributors to the *JNS* could have used a time machine to arrive in the year 2012, the year preceding the 150<sup>th</sup> anniversary of this journal, what would they have said to the current state of the debate? More concretely, what would be their assessment of the motives behind these initiatives, of the conceptual ambition of this endeavor, and of the statistical issues pertaining to the construction of the various systems of indicators?

### 3.1 The motives behind “GDP and beyond”

With respect to their reaction to the motives behind the current discussion, one might fear that we will only be able to speculate. After all, when the *JNS* was founded, German per capita GDP – even though the concept as such was yet to be developed – was arguably much lower than it is today in any OECD country. While contemporaneous observers were rightfully in awe of the technical achievements already reached in their era of “steam and lightning” (Hildebrand 1864: 136, our translation) and the fast growth in economic prosperity observed after the industrial revolution had started, they would probably not have been prepared whatsoever to the exponential growth in prosperity which we have seen since, in particular during the post-WWII era. Maddison (2001), for instance, demonstrates more than convincingly that (not only) in the developed world real per capita income today is multiple times the income experienced one and a half centuries ago.

Thus, as speculative as this has to be, it seems quite likely that our time travelers’ first reactions to the current debate on human welfare would be expressions of astonishment: How could societies which are that rich by any historical standard display such concern about their precarious condition? And, correspondingly, can it actually be true that outside of the current “GDP and beyond” debate the typical focus displayed by economic reports published by international organizations, such as the *Organisation for Economic Co-Operation and Development* (OECD) and the *International Monetary Fund* (IMF), by advisory groups such as Germany’s *Council of Economic Experts* (GCEE), and by basically all economic research institutes lies on deviations of aggregate economic activity from its trend value? That is, the major attention is not given to the level of prosperity, but its growth relative to an upward trend? How rich must human societies have become in the developed world, and how regular and stable must past growth have been?

But after this initial shock subsided, the intellectual basis of the current debate would undoubtedly receive endorsement. Hildebrand (1863a: 14, 17), for instance, acknowledges the dangers to the coherence of any society caused by fast economic development that is carried by strong market forces, rapid investments into the capital stock and turbulent scientific progress. From this vantage point, the major questions of the current debate, (i) “Is the accumulation of material welfare really generating more happiness?”, (ii) “Are there insurmountable, albeit distant limits to growth?”, and (iii) “How could well-being be ascertained without economic growth?” are merely the most recent expression of age-old questions posed by uncountable analysts of human prosperity and of the functioning of society. In addition, 19<sup>th</sup> century social scientists would certainly be well acquainted with the phenomenon that crises always tend to bring these fundamental questions on human prosperity to the surface (Hildebrand 1863a: 17).



What perhaps most distinguishes the current discussion most from earlier debates on the connexion between economic prosperity and well-being is the strong emphasis given today to the theme of economic, social, and environmental sustainability. But since they were confronted with quite similar problems at their time, earlier economists would probably quite easily be convinced that the accumulated stocks of private and public debt currently provide a serious challenge to the objective of ascertaining sustainable public finances, requiring intense analysis and discussion (Müller 1912: 372, 375). Somewhat different, although probably not less supportive, might be the time travelers' reaction to the recent concerns about environmental quality and sustainability, as environmental degradation and, in particular, global challenges such as climate change, have not been an issue during their times. But even though the emphasis on these questions is predominantly a modern development, they also find their reflection in earlier writings (Hildebrand 1863a: 19).

The rest of this section will argue that we find important roots in 19<sup>th</sup> century thinking for both dimensions of the current debate as well, the rather abstract discussion about what should be the ambition of and the concepts behind such an indicator system on the one hand, and the predominantly applied discussion about the statistical principles and approaches to fulfilling this task on the other.

### 3.2 The conceptual ambition: “Nationalökonomie”

Much of what we can infer about the position the earlier contributors to the *JNS* would have taken, derives from the articles written by Bruno Hildebrand, the founding editor of the *JNS*. His contributions are acknowledged impressively in an obituary by Johannes Conrad (1878: IV, XI, XIII), his comparably important successor as *JNS* editor, documenting the extraordinary respect Hildebrand enjoyed among his peers. Reading Hildebrand's articles reveals that, irrespective of his personal position on these questions, he would adamantly have maintained that only careful and encompassing statistical analysis could bring about the desired enlightenment in these matters. Already the programmatic “mission statement” (our term) of the *JNS* (Hildebrand 1863d) outlines his vision of the economy as one of the fundamental elements of the life of any society, equal in importance to its language, its literature, its legal system and its art (Hildebrand 1863d: 3), making it not only worthwhile but critical to explore its functioning on the basis of solid statistical analysis.

Hildebrand (1863d: 3) reckons that the most demanding intellectual challenge for economic analysis is the fact that, although the functioning of an economy follows some important regularities, it nevertheless does not obey any fundamental laws. This, he realizes, distinguishes economics from the – by then already comparatively well-established – natural sciences. For him, the reason behind this contrast is clear: It derives from the fact that aggregate economic activity is the condensation of a multitude of individual actions and decisions. Consequently, economics as a science would need to utilize the experiences made in the past in order to identify the regularities characterizing human behavior and the causes and consequences of human actions, with the aim of supporting the principles of individual freedom and individual responsibility. Hildebrand leaves no doubt in his contributions that this objective would necessitate a plenitude of statistical work.

This summary assessment is highly reminiscent of the ambitions behind “GDP and beyond”, and, thus, deserves a closer look. In several highly programmatic articles in the *JNS*, Hildebrand (1863a, 1863c, 1866) carefully outlines the role of economics

and statistics in their historical context in a way that in many respects could serve as a guide to the current debate. As a starting point for his observations, Hildebrand characterizes the – by then – relatively new discipline of economics (“Nationalökonomie” to contemporaneous observers) as the result of the enlightenment movement of the late 18<sup>th</sup> and early 19<sup>th</sup> century (1863a: 5-6, 1866: 2). In its early expression, this new discipline had the ambition to equal the natural sciences in its quest for discovering the laws governing human behavior, a perspective that went hand in hand with the idea of an “atomistic society” in which the pursuit of individual utility was the exclusive driving force of all activities (Hildebrand 1863a: 6-7, 1873: 2-3).

Hildebrand quite openly reveals his frustration with the obstinate resistance of many contemporary economists to a more nuanced view on the importance of moral sentiments (Hildebrand 1863a: 8-9), but also offers very little sympathy for the contrasting position taken by contemporaneous socialists that the pursuit of individual utility rather be the root of all evil (Hildebrand 1863a: 9-10). For him, the evidence seems overwhelming that free choice of occupation and division of labor, the uninhibited allocation of production factors according to their marginal returns, steady investments into the stock of productive capital, and incessant scientific progress had become powerful driving forces of economic progress. These developments, which were intimately associated with the intellectual success of contemporaneous economic reasoning, were lifting human productivity to unprecedented heights – and were even viewed to enhance the ability for education and enjoyment in the population (Hildebrand 1863a: 10-14).

But despite his highly critical assessment of the socialist arguments against contemporaneous economic thinking, Hildebrand also explicitly rejects the notion emphasized by Adam Smith and other economists that pursuing an untamed *laissez faire* – which would render any statistical analysis to become a useless husk – could be a sensible foundation of a prosperous society (1863a: 10). He rather views the tendencies of any *laissez faire*-economy to an ever-increasing concentration of economic power and wealth with a serious dose of skepticism, especially since the proponents of *laissez faire* could allegedly refer to unconditioned scientific arguments (Hildebrand 1863a: 14-17). His view is echoed in Knapp (1871: 238-239). Similarly, proponents of a balanced view on economic growth today need to defend their position both against blind believers as well as against fanatic opponents of growth.

As a first conclusion of his considerations, Hildebrand challenges his peers to engage into a critical review of the contemporaneous science of economics, culminating in the clarification of the questions whether human behavior followed strict natural laws and, consequently, economics should be accepted as a natural science, after all (1863a: 19-20). If that were to be the case, neither moral sentiments nor political interventions would have to play any role. Personally, Hildebrand rejects this notion vehemently (1863a: 21-25, 137-140). Rather, in his view individual economic freedom, as worthy of protection as it is, cannot, taken by itself, ascertain societal prosperity and human progress. In addition to the individual pursuit of utility, ethical considerations and a sense of responsibility lie at the heart of any successful society (Hildebrand 1863a: 140-143). Knapp (1871: 241-243, 247) essentially takes the same position.

Today, economics is well established as a social science, not as a natural science, and economists would definitely shy away from arguing that the regularities which their work is uncovering carry the encompassing explanatory power of a natural law. Nevertheless, both in the 19<sup>th</sup> century and today, the recipients of the empirical results derived by economists and statisticians might misunderstand the limited conclusions facilitated

by the evidence. Knapp (1871: 240), for instance, argues that the idea – prominent among contemporaneous observers – that statistical regularities were proving the notion that individuals acted according to laws of nature, unable to alter their life course by their own will, decisions, or ethical considerations, reflected the incompetence of commentators to see statistical analyses for what they are: While statisticians themselves would rightfully and soberly regard statistics merely as a set of procedures to uncover regularities and causal links, its findings might lead outside observers to inflated conclusions. This assessment is highly reminiscent of the misunderstanding of the concept of GDP as a welfare measure, a misunderstanding that has not been suggested by economists or statisticians, but has nevertheless become the epitome of the motives behind the “GDP and beyond” movement. In the 19<sup>th</sup> century as well as today, analytical concepts which are taken out of context might lead to completely erroneous conclusions. In particular, many 19<sup>th</sup> century economists apparently tended to argue that, since economic freedom was widely realized, circumstances could hardly get any better, and there was not need whatsoever for economic or statistical analysis of social inequality. By contrast, Schmoller (1873: 6-7) not only argues that there was actually a deep societal divide, but also blamed the overemphasis on economic performance in the short run to lie behind it. In his view, the exclusive aspiration on fostering economic performance, without giving any consideration to the potentially negative consequences of increasing material wealth for other aspects of human existence was putting social coherence at risk (1873: 10). Similarly, Knapp (1871: 247) observes that the regularities in income inequality and poverty provide evidence enough for concern about the future prosperity of society. Schmoller (1873: 9, 11-12) advocates a strong state as a factor correcting unbearable social inequality, without interfering in the general functioning of the market economy. It is not difficult to detect the current competition between various forms of capitalism, most prominently a US-style free market economy and a continental European-style social market economy, as the present-day expression of this intellectual conflict between the concept of *laissez faire* and its critics. And the balance advocated by Hildebrand and Schmoller in their writing is clearly reflected in the setup of the modern social market economy as well.

Hildebrand's second, closely related, conclusion regards the agenda for statistical research, and finds its direct reflection in today's initiatives aiming at the construction of indicator systems measuring the state of human welfare and progress: From the perspective of the current discussion on measuring the state of human welfare, the highly practical consequence of such a balanced position would be that collecting evidence on the genuine state of affairs and their development over time has to be viewed as both worthwhile and necessary. Most importantly, rejecting the notion that human behavior follows natural laws and exploiting the fact that human behavior displays a high degree of regularity are no contradiction whatsoever. Rather, these systematic patterns reflect both regularities in human experiences and similarities in the environment in which individuals make their decisions, and, thus, they offer the opportunity to statistical researchers to discover patterns that can form the basis for good economic policy (Hildebrand 1863a: 142-143, 1863c: 482, 1872: 9-10).

This position on the proper role of the discipline of statistics (“Moralstatistik” to contemporaneous observers) contrasts the scattered statistical attempts that characterized earlier times (Hildebrand 1866: 2) and is emphatically shared by Knapp (1871) and von Neumann-Spallart (1885). In these contributions, statistics is ultimately an instrument that allows abstracting from anecdotes and identifying the general pattern behind

observable phenomena (Hildebrand 1866: 3; Knapp 1871: 248; von Neumann-Spallart 1885: 225). Thus, it objectifies experiences and facilitates the formulation of hypotheses and the test of their implications against the evidence, far beyond the potential offered by mere plausibility considerations (Hildebrand 1863c: 482, 1866: 3, 5). Taken together, these insights on the appropriate role of statistical analysis and the arguments regarding economics as a social science build up a fervent plea for pursuing – in today's words – evidence-based economic research and policy advice (Schmidt 2007).

In essence, this 19<sup>th</sup> century literature already appreciates that statistics can be viewed as the reporting system of the economy as a whole, whereby it is learning about the true state of its economic circumstances, based on meticulous accounting and routinely conducted closure of accounts (Hildebrand 1863c: 482). And given that this work is continued on a regular basis, without major gaps, the resulting time series will offer even more opportunity to learn about patterns of stability and instability and perhaps even causal mechanisms underlying the results (Hildebrand 1866: 9). This understanding forms the basis for the systems of national income accounting which were subsequently developed in the 20<sup>th</sup> century and which are the staples of statistical reporting on human welfare today.

### **3.3 The construction of indicator systems: "Statistik"**

Today it is well understood that any researcher approaching the task of constructing an encompassing system of indicators of human welfare should first consider a range of statistical issues. The most important items on the agenda concern (i) the practical setup of the analysis and the governance of the institutions involved in this work, (ii) the distinction of correlation and causality, and (iii) a clear grasp on sampling error, hypothesis testing, and measurement error. This basis for the statistical work has necessarily to be secured before the applied work is starting, or otherwise either the statistician or the recipients of the results will be headed for disappointment. Again with a focus on the contributions to the *JNS* during its early years in the late 19<sup>th</sup> century, it will be documented here that many of these issues have already been discussed in the articles published by our statistical forefathers.

Turning, first, to the practical setup of the analysis and the governance of the institutions involved in the construction of indicators and indicator systems, the simple, yet easily forgotten starting point is necessarily a careful conceptual delineation of the task. As a general rule of thumb, the more ambitious the indicator system is conceptually, the more likely it will run into serious problems of measurability and comparability. In particular, one needs to ascertain whether the desired statements to be derived merely capture the current status in a snapshot perspective or should be extrapolated to outline future conditions in terms of a projection. Classical statistical work concerns the former (Hildebrand 1864: 137), while the modern-day emphasis on sustainability issues provides additional challenges.

Most importantly, one needs to clarify at the outset, whether the desired comparability is inter-temporal or even international in nature. The fact that international comparability is not easy to ascertain was a matter of intense discussion among statisticians of the 19<sup>th</sup> century. Von Neumann-Spallart (1885: 223), for instance advocates the formation of an international statistical institute to reach more uniformity in statistical concepts and procedures. After all, he goes on to argue, statistics can only fulfill its tasks when it is understood as an international science (1885: 223), especially when the

objective is explaining the consequences of regulations or institutions which vary across, but not within nations (1885: 225). This insight had already been behind the formation of national statistical offices in the new nation states of the 19<sup>th</sup> century and of the formulation of the rules governing their operation (N.N. 1870: 111). Regarding the governance of such institutions, it seems obvious that these statistical agencies need to be independent of their respective governments (von Neumann-Spallart 1885: 229).

Another trade off concerning the organization of the task regards the recurring nature of the desired report. In particular, the current “GDP and beyond” initiatives are aiming at the construction of indicator systems which are published in regular, not all too distant intervals, which are published timely and not with all too long lags, and whose messages are robust and not altered too severely, if definitions, timing or other practical issues involved in the derivation of their indicators are varied slightly. Timeliness was a concern already 150 years ago. Hildebrand (1864: 136, 137), for instance, judged it to be unacceptable that in his fast-living times, one had to wait for several years until important statistical results were published. The pressure for short lags between the realization of the economic phenomenon and its publication has not become smaller since that time.

A further issue which most of the “GDP and beyond” literature has been silent about (an exception is Kassenböhmer/Schmidt 2011), concerns the trade off between the conceptual desirability of candidate indicators and their measurability. As a general rule of thumb, those items which, in principle, would offer important additional information on human prosperity on top of GDP, are typically difficult to measure precisely. Examples would be individual freedom or social contacts. By contrast, items which are typically measured quite precisely, such as life expectancy, are often highly correlated with GDP and, thus, unable to provide valuable additional information. In fact, the close resemblance between economic prosperity and life expectancy already occupied the literature 150 years ago (Hildebrand 1863b).

Finally, all statistical work involves time and monetary resources. This almost trivial, yet easily forgotten insight has already been a matter of concern for 19<sup>th</sup> century statisticians (Hildebrand 1864: 136; von Neumann-Spallart 1885: 225). Thus, as one guiding principle of all applied statistical work, one should make use of the generous reservoir of previous conceptual work and already established data collection wherever possible. Only if applied statisticians follow this principle, will they be able to avoid redundant efforts and to trade-off the marginal benefits and costs of collecting additional data or constructing another indicator properly.

Regarding the organization of statistical work, Hildebrand (1863c: 487, 1866: 5, 7-8) argues that the work of statistical offices will always be indispensable, most importantly, since the statistical analysis should cover the nation as a whole, not merely a specific section of it, and frequently needs to rest on official registers. But he also acknowledges the resistance which respondents might develop when being interviewed by government agencies for statistical purposes, for reasons of privacy and fear of oppression, and the limits of the willingness of civil servants in fulfilling statistical reporting duties. These ideas are certainly reflected in the modern consensus that statistical offices, research institutes and universities, and private data providers might work fruitfully together to derive an encompassing portrait of the actual state of affairs.

The second statistical issue to be discussed is the distinction between correlation and causality. While it is most obvious to experienced practitioners and academic economists and statisticians, mere correlations are all too often mistaken as causal relations in the political debate. But one has to be absolutely adamant in maintaining that the indicator

systems which are constructed in any of the various expressions of the current “GDP and beyond” movement are exclusively serving descriptive purposes. That is, their concentration is on positive questions such as “What is the true state of human welfare?” or “What are the connections between material and non-material aspects of quality of life?”. A sensible normative debate “What should be done about the state of affairs?” can only be conducted after these questions have been answered – and it cannot rest its answer on these indicator systems alone.

This was already clear to 19<sup>th</sup> century economists and statisticians (Hildebrand 1866: 9), and yet it is often forgotten in the discussions in the political and public sphere even today. Knapp (1871: 242-243) explains brilliantly how the conclusions derived from statistical results – here the detection of regular patterns of behavior – might differ substantially, depending on the direction of causality implicitly presumed by their commentators. While the contemporaneous proponents of economics as a natural science identified – in modern words – the similarities in behavior as the outcome of some external driving forces, their critics emphasized that inherent similarities between individuals and contextual effects generated homogeneous behavior instead. From the perspective of modern-day literature on social interactions (Manski 1995), Knapp’s position was highly sensible.

While the potential and the limits of observational analyses to facilitate causal statements are the matter of intense academic discussions today, with randomized controlled trials serving as a hypothetical benchmark, already Hildebrand (1866: 3) indicates that, by contrast to the natural sciences, the social sciences suffer from the impediment that they will typically not be able to resort to experimental evidence. The modern-day evaluation literature cautions that, even if causality might have been established in the analysis, one might be reluctant to extrapolate the result from the situation under scrutiny to a more general setting. This is due to a potential conflict between the internal and the external validity of evaluation studies. Typically, social experiments are characterized by a high internal validity and low external validity. Being unaware of the experimentalist movement of today, reservations about external validity are already present in the discussion of causal issues by Hildebrand (1866: 10).

As a third precondition that has to be fulfilled, both, statistical researchers and the recipients of their work need to find a good grasp on the essential statistical issues of sampling error, hypothesis testing, and measurement error. Statistical analysis is always attempting to reduce complexity by abstracting from the multitude of individual influences on the phenomenon, which are irrelevant for the question at hand. This task necessarily involves the construction of average figures, representing relative frequencies of the condition in question within a sample that represents the population. In that sense, all statistical work is abstracting from anecdotes, that is, individual cases, to derive a summary picture of the state of affairs (Hildebrand 1866: 3-4).

To receive a more detailed portrait of the state of affairs in a population, one might resort to stratification, and construct relative frequencies in the corresponding sub-samples. But, as Hildebrand (1866: 7) stated in his words already some 150 years ago, all statistical work needs to rest on some – in modern words – identification assumptions. That is, in order to characterize the contrasts between the relative frequencies across sub-samples, one has to decide, in the first place, how to distinguish these sub-samples. Statistical analysis can only detect relative frequencies when analysts are deliberately looking for them, that is, it is pursuing to discover the known unknown, not the unknown unknown.



The reported figures in any statistical analysis can never be anything else, but an approximation to the underlying population concept. While it was a comparatively recent development at the time, 19<sup>th</sup> century statisticians started to understand that probability theory is providing a sound analytical basis for addressing the question whether numerical deviations of the derived average figures have to be assessed as relevant or not (Westergaard 1885: 1). What made the idea somewhat unpalatable to many contemporaneous observers, was obviously the abstraction involved in its underlying thought experiment – which is well-established as the intellectual basis of all statistical work today – of imagining the current calculations (in the “sample”) as being embedded in a multitude of hypothetical calculations of the same kind (Westergaard 1885: 2).

Following this abstraction, it becomes possible to judge whether some movement of average figures over time or some deviation across sub-samples are indeed indicating anything substantial. It all depends on the precision of the estimates involved, and “large” or “negligible” have to be understood in relation to the unavoidable residual uncertainty, which is inherent in any such approximation (Westergaard 1885: 23). Most importantly, given the same phenomenon, large samples tend to produce more precise approximations than small samples (Westergaard 1885: 2). The concept of statistical testing is intimately related to this notion of statistical precision. Its principal idea is to devise a decision rule according to which a deviation is large enough to be assessed as “significant”. Again, *ceteris paribus* precision tends to increase with sample size (Westergaard 1885: 4). Correspondingly, any causal effect itself can only be approximated with more or less precision, not with certainty (Westergaard 1885: 22). And any statement about causality necessitates tailoring the analysis to the situation, instead of following a one-size-fits-all approach (Westergaard 1885: 12).

One of the central insights emerging from the current-day literature on the evaluation of treatment effects is that identification and sampling error are intimately related. On the one hand, it is clear that proper causal analysis is resting on the principle of comparing the comparable, and that homogeneity within a sub-sample can more easily be ascertained, if the sub-samples are defined according to more demanding stratification rules. On the other hand, given that the overall sample size is limited, an extensive stratification geared towards ascertaining within sub-sample homogeneity, corresponding to less-demanding identification assumptions, will lead to smaller and smaller sub-samples, and ultimately to highly imprecise average figures. Thus, there is no escaping from finding the right balance between identification assumptions and potential sampling error. The 19<sup>th</sup> century statistical literature was well aware of this intricate problem (Westergaard 1885: 22).

As it is a problem inherent in any statistical work, measurement error has always been regarded as a serious obstacle to empirical analysis. By contrast to the multitude of irrelevant influences on the phenomenon under scrutiny, which are fruitfully subsumed under the notion of sampling error, measurement errors tend to bias the results systematically (Westergaard 1885: 23). But as if that was not challenging enough, while condensation always requires averaging across individual observations, the current discussion regarding “GDP and beyond” also involves the idea of summary measures of welfare, as an amalgamation of the indicators representing the multitude of facets of human existence. If one was indeed pursuing this idea, this would require researchers to average even across different types of (quantitative and qualitative) indicators. Potentially, this would lead to an exponentiation of the many measurement problems which tend to plague the analysis even in the most benign circumstances.

Thus, even if statisticians are completely aware of the intricate measurement problems involved in their work, the attempt to construct a single, encompassing welfare indicator might face conceptual obstacles, if indicators are difficult to compare across persons in terms of content (“apples and oranges”). Even more dangerously, it might run into problems of manipulability, if the practical application of the concept is open to interpretation and suggestive interviewing, as might be the case, for instance, with issues of self-assessed well-being. In the face of such problems, the current consensus reached in the “GDP and Beyond” debate seems to be that an enlightened society has to be able to deal with a (limited) wealth of information, if the complex reality is to be captured adequately. Thus, instead of a single, encompassing welfare index, one should construct indicator systems instead. It very much seems that time-travelling 19<sup>th</sup> century economists and statisticians would have understood.

#### **4 Themes and indicators today – and their precursors**

One way to delineate the task of measuring human welfare comprehensively that is chosen by both SSFC (2009) and GCEE/CAE (2010) is to take standard measures of economic performance as a starting point, and to improve upon these standard measures in three directions. Regarding (i) measures of economic performance and material welfare, one should aim at advancing the traditional well-matured measures, such as GDP and the unemployment rate, while retaining their well-known strengths. To address (ii) non-material aspects of human welfare, one should enhance the indicator set regarding various facets of quality of life, respecting the conceptual limits to measurability of emotions and preferences. And to acknowledge that there are important (iii) forward-looking aspects of sustainability, one should construct projections of possible courses for the future state of affairs.

It is the latter set of sustainability indicators that distinguishes the current discussion most visibly from earlier attempts at the construction of encompassing indicator systems. Gauging sustainability typically entails new conceptual challenges as well: Quite importantly, these projections always necessitate an underlying assumption of behavioral stability. They are projections of what could happen under a specific set of circumstances, business as usual, for instance, not forecasts of what is likely to happen. By contrast, forecasts also attempt to consider how economic agents or policy-makers might react to a given tendency, thereby potentially altering the course of events altogether. In addition, and not at all less relevant, many of the issues of concern in these considerations have an important international dimension, precluding a sensible reporting that would be confined to the national level.

Figure 1 documents the dashboard that was developed on request of the Franco-German Ministerial Council by the German Council of Economic Experts and the French Conseil d'Analyse économique to monitor economic performance, quality of life and sustainability (GCEE/CAE 2010). Its three pillars “Economic Performance”, “Quality of Life” and “Sustainability” that together contain a total of 25 indicators closely follow the reasoning discussed above. The following sub-sections demonstrate that at least some of these indicators had precursors already in the statistical work of the late 19<sup>th</sup> and early 20<sup>th</sup> century. Documenting this early statistical work leads to impressive evidence that these ideas and concepts left their imprint in the volumes of the *JNS* which were published during this era.



Economic Performance (A)	Quality of Life (B)	Sustainability (C)
GDP per capita	Health: Potential years of life lost	Private sector net fixed capital formation (% of GDP)
GDP per hours worked	Education: Students (ISCED 1-6) aged between 15 and 24 years	R&D investment (% of GDP)
Employment rate (15 - 64 age group)	Personal activities: Employees working on shift work	Cyclically adjusted fiscal balance (% of GDP)
Net national income per capita	Political voice and governance: Voice and Accountability	Fiscal sustainability gap S2
Final consumption expenditure per capita (including government consumption)	Social connections and relationships: Frequency of time spent with people at sport, culture, communal organization	Total private credit to GDP gap
Distribution measure of net income per consumption unit (income quintile share ratio (S80/S20); internationally harmonized)	Environmental conditions: Urban population exposure to air pollution by particulate matter	Real equity price gap
	Personal and economic insecurity: Not-at-risk-of-poverty rate	Real property price gap
		Level of greenhouse gas emissions
		Greenhouse gas emissions per capita
		Resource productivity (GDP relative to non-renewable Domestic Material Input, DMI)
		Resource consumption (non-renewable Domestic Material Consumption - DMC per capita)
		Biodiversity (preliminary indicator: bird index)

Source: GCEE/CAE 2010: 27.

**Figure 1** Dashboard for Monitoring Economic Performance, Quality of Life, and Sustainability

#### 4.1 Economic performance

What can today be regarded as “classic” measures of economic performance, most prominently, the growth rate of (per capita) GDP, are serving a range of practical purposes. They are, first of all, a gauge of the current state of the economy, they indicate whether monetary and fiscal policies have to be set into action, and they serve as the basis of reliable fiscal planning. Of course, as stated in section 2, measuring economic performance is at best an approximation to capturing the state of human welfare. Well-known weaknesses of these measures are twofold. They might suffer from problems of observability, as it will be the case, for instance, for activities in the shadow economy, and from problems of valuation and measurement, a problem which tends to plague measurement in the field of public services. And without doubt, distributional questions have not been addressed sufficiently in national accounting up to now.

Against this backdrop, the SSFC report recommended to (i) measure income or consumption in per capita terms, (ii) emphasize the household over the individual perspective, (iii) analyze also wealth and not only current income, (iv) analyze the distributions

of income, consumption and wealth, and (v) give more emphasis to capturing non-market activities (SSFC 2009: 12-14). After careful evaluations weighing the desirable versus the feasible, the GCEE/CAE report proposed the six indicators that are depicted in Figure 1 to monitor material well-being.

And yet, in the face of the critical undertone of the current discussion, one might easily forget the important achievements which have led to the comfortable situation modern analysts have been handed down by their forefathers. In fact, the contributors to the *JNS* during its early decades obviously would have dreamed of a measure with the qualities of GDP, surely not only to assess the economic performance but also as a tool for approximating the welfare of the population. The development of coherent national accounts and the calculation of an all-encompassing indicator for the performance of the market economy along the lines of the later concept of GDP were indeed already present as a latent ideal in their writings.

Yet for several decades, they still had to content themselves with indirect approaches to estimate the national product and to draw conclusions about the situation and evolution of material welfare from these estimates. The most important basis for such work were the tax statistics of the time as is documented by several contributions from Soetbeer (1879, 1882) in the *JNS* that are concerning the national income of Prussia. Soetbeer does not confine himself to report estimates of the overall national income in Prussia. Instead, he provides further insights to his contemporaries through international comparisons and quite detailed analyses of the income distribution. His comparisons of results for Prussia rested on methodologically comparable, i. e. income tax-based estimates of national income for the Kingdom of Saxony and the United Kingdom, respectively. These comparisons served two distinct purposes: He used, first, the tax statistics from nearby Saxony mainly with the scientific purpose to check the plausibility of his results for Prussia, since it was believed that the compilation of the Saxon income tax statistic was the most accurate and the Saxon revenue authorities were the most reliable of his time (Soetbeer 1882: 235). Second, his comparisons of results for Prussia and the United Kingdom clearly follow the ambition, to provide empirical evidence for political decisions. Therefore, he does not only compare the mean values of per capita income, but also proceeds to comparisons of the respective income distributions (Soetbeer 1882: 238-239). In doing so, he was clearly aware that his data at hand allowed only approximate estimates. But he was likewise firmly convinced that even these limited pieces of statistical evidence were highly important to inform (economic) policy.

At his time, the social question was hotly debated in Prussia. The unification of the Reich in 1871 had led to the abolishment of the remains from the medieval guild system. As Schmoller observed, the guarantees of unprecedented economic freedom led to a substantial improvement in economic performance, but it also created social problems and mounting tensions in the society. His perception, that “our society is threatened to resemble more and more a ladder which grows rapidly at the bottom and at the top, but in which the steps in the middle are more and more breaking-out, allowing a sure foothold only at the very bottom or at the very top” (Schmoller 1873: 11, our translation) was widespread and closely resembles the concerns about growing inequality in many countries that are voiced in contemporary debates about the need to go “beyond GDP”.

For Soetbeer, the concerns of growing inequality first and foremost pointed towards an empirical question that he tried to resolve for the period 1872-1879 in Prussia through analyses of changes in the income distribution (Soetbeer 1879). At the outset, he con-

cedes that based on personal experiences, non-specialist observers could get the impression of a declining income in most parts of the society but ever rising incomes in the richest classes (Soetbeer 1879: 113). He then proceeds with calculations of the income shares of six income classes for the year 1879 and tabulates the yearly evolution in the three highest income classes for the period 1872-1879 (Soetbeer 1879: 114-115). He finds that the income shares of all classes remain close to constant and that the public perception of a growing inequality does not hold up in the light of empirical scrutiny (Soetbeer 1879: 115-118).

Three years later, he confirms this result for the period 1872-1881 and, as a methodological innovation that resembles the income quintile share ratio S80/S20 included in the dashboard proposed in GCEE/CAE (2010), calculates the ratio of the income in the two highest income classes relative to the whole national income as a summary indicator to track the evolution of income inequality over time (Soetbeer 1882: 233-234). Comparing the evolution of the two classes with the highest incomes in Prussia and the United Kingdom, Soetbeer is puzzled by the fact that the introduction of a progressive income tax is hotly debated in Prussia although the data show no significant widening of the social gap, whereas in the UK with a much higher and ever growing income share of the two highest classes, this tax seems to be a non-issue (Soetbeer 1882: 239-240). Soetbeer comes to the conclusion that “there is no worse illusion than the opinion that an artificial complication of big capital accumulations could raise the general welfare and especially better the lot of the working class.” (Soetbeer 1882: 240; our translation).

With respect to one prominent recommendation of the SSFC report regarding economic performance, namely to “consider income and consumption jointly with wealth” (SSFC 2009: 13), the contrast between the desirable and the feasible has barely changed over the centuries. Already Hildebrand (1863c: 479-480) deplores that so little was known about the true value of real estate and the corresponding proceeds, but also about the debt associated with real estate. A comparable frustration is provoked in contemporary endeavors to go “beyond GDP” when politicians with a high preference for social equity have to accept the sobering fact, that a wealth inequality indicator comparable to the income quintile share ratio S80/S20 included in the GCEE/CAE dashboard is not (yet) feasible, due to the lack of reliable wealth information especially for high and highest fortunes.

## 4.2 Quality of life

Undoubtedly, a comprehensive perspective on quality of life comprises both material and non-material aspects. Facets of material quality of life are therefore indispensable pieces of information when aiming at an encompassing indicator, and one should not exclude material aspects altogether. Quite the contrary, we need supplementary information that augments the “classic” indicators of economic performance and material well-being. But addressing the non-material facets of human existence poses its own conceptual challenges. Most specifically, the aggregation of the various non-material aspects of quality of life confronts a twofold problem of aggregation: Not only do we have to condense the information by aggregating across people, as in any statistical approach, but we also have to find aggregate representations regarding the various facets of quality of life for any individual.

Even more importantly, one might assess the typical impact of deviations in a specific aspect of life, living with and without a physical impairment, say, more or less convincingly, depending on the information content of the data utilized and the econometric

approach chosen to extract this assessment from the data. For this purpose, one would construct contrasts between self-assessed levels of well-being, for instance, for individuals which are observed at two different points in time, some of which change their status (from “without an impairment” to “with an impairment” in the example). But as fruitful as such an approach of using the resulting difference in an outcome measure could be in understanding the typical reaction to misfortune, measuring the *level* of well-being is substantially more complicated: There is simply no normal state of affairs which could allow statisticians to calibrate different answers of self-assessed well-being.

In particular, utilizing surveys on happiness to engage into international comparisons is as tempting as it will necessarily be misleading. Too substantial are the many problems of measurement and observability, since long-term happiness and short-term affects are different, cognitive problems and strategic behavior might distort the answers to the interview questions, and there is a danger of political manipulability. Instead of such a “top-down approach” based on overall measures of self-assessed well-being, GCEE/CAE decided in favor of a “bottom-up approach” that rather collects information on seven different spheres of non-material quality of life. The seven dimensions included in the second pillar of the GCEE/CAE dashboard (Figure 1) follow the respective recommendation in the SSFC report quite closely.

The “capability approach” (Sen 1999) provided the conceptual framework for the choice of these dimensions. The core focus of the capability approach is on what individuals in a given society are able to do, i. e. which so-called “functionings” they can freely choose to promote and achieve. If our time-traveling contributor to the *JNS* from the 19<sup>th</sup> century could have a look at this second pillar of the dashboard, he would surely not be too surprised, neither by the selected dimensions nor by the chosen indicators. That dimensions like health, education, working conditions and security were already seen as playing a crucial role for the quality of life in a society is well documented for instance in the contributions from Neumann (1872), Schmoller (1873) and Sartorius von Waltershausen (1882a, 1882b).

The rapid evolution of applied statistics as a discipline during the first 50 years of the *JNS* and its expansion to ever more subject areas is by itself a proof that the statisticians of the late 19<sup>th</sup> and early 20<sup>th</sup> century already had the clear ambition to go “beyond GDP”, even though the tool of GDP was yet to be introduced. The progress made towards this aim becomes visible through a comparison of Neumann’s (1872) account of the social situation in Prussia that was still plagued with the non-availability of many fundamental statistics (e. g. on wages, p. 284) and had to make do with more or less suitable “proxy variables” (like counts of the numbers of dress- and shoemakers or of physicians in a given region as a proxy for the wealth of the population, p. 283, p. 294). A comprehensive overview of the statistical infrastructure and its latest results is provided by Kollmann (1912).

### 4.3 Sustainability

The standard conceptualization of sustainability encompasses the three essential dimensions of social, economic and environmental sustainability. Social sustainability addresses issues like fairness in distribution and opportunity, adequate provision of social services, gender equity, and political accountability and participation. Economic sustainability demands that the economic system is able to produce goods and services on a continuing basis under the side conditions that a manageable level of government and

external debt is maintained and extreme imbalances between the different sectors are avoided. Environmental sustainability requires (i) that a stable resource base is maintained, (ii) that the over-exploitation of renewable resource systems or environmental sink functions is avoided, (iii) that non-renewable resources are depleted only to the extent that investment is made in adequate substitutes, (iv) that biodiversity, atmospheric stability and other ecosystem functions are maintained which are ordinarily not classified as economic resources. Only if the social, economic and environmental requirements are satisfied simultaneously, is a sustainable state achieved (GCEE/CAE 2010: 102).

From a methodological perspective, sustainability is a concept regarding long-term developments. Essentially, it requires that we answer the question “whether we can hope to see current levels of well-being at least maintained for future periods or future generations” (GCEE/CAE 2010: 101). Statements regarding future developments can never be deterministic – their construction requires identification assumptions on future paths of behavior. Here we are not talking at all about forecasts, though. The desired projections are “What would happen, if ...” statements instead. These questions are especially difficult to answer in the realm of social sustainability. Consequently, the indicators in the third pillar of the dashboard concentrate on aspects of economic and environmental sustainability (Figure 1).

To answer the question whether economic growth can be assumed to continue in an uninhibited fashion, the first two measures address the maintenance and improvement of the productive capital stock. The following two indicators are both concerned with the budget restrictions of governments, but measure fiscal sustainability with different time horizons (GCEE/CAE 2010: 111-112). Another set of three indicators constitutes a small set of early-warning indicators that could alert policy makers and the general public to the build-up of excessive developments in different spheres of the financial sector. Their construction follows the principle of “cumulative gaps”, i.e. the deviations from the respective trends during a moving time window are summed up. If for any indicator the cumulative gap exceeds a pre-determined threshold, this is interpreted as a signal that a crisis in the financial sector could be looming and counteractions might be warranted (GCEE/CAE 2010: 115-125).

The remaining five indicators of the third pillar in the GCEE/CAE dashboard (Figure 1) cover the environmental domain. The inclusion of ecological indicators in a dashboard that uses nations as the unit of observation generates a serious conceptual problem: Many environmental challenges cannot be treated from a purely national perspective in any meaningful way. Keeping this important caveat and, thus, the necessity of tracking the corresponding international developments closely, GCEE/CAE decided to focus on three dimensions: greenhouse gas emissions, resource productivity and the maintenance of biodiversity. To respect the global budget of 705 gigatons of CO<sub>2</sub> that could presumably be emitted until 2050 without jeopardizing the aim of keeping global warming below 2 °C compared with the pre-industrial level, the CO<sub>2</sub> budget per capita was calculated to be 2.7 tons annually between 2010 and 2050. The fact that Germany emits currently around 10 tons of CO<sub>2</sub> per capita documents the size of the challenge and the urgency to resolve it (GCEE/CAE 2010: 128-132).

To monitor the use of non-renewable resources in the national economy, GCEE/CAE recommend the inclusion of indicators for resource productivity and for resource consumption. Due to a number of remaining methodological problems that limit their precision and international comparability, these two indicators can merely serve as warning signals concerning the qualitative trends of resource use (GCEE/CAE 2010:

132-139). Last but not least, the preservation of biodiversity is also addressed in the GCEE/CAE dashboard. Biodiversity, conceived as the totality of genes, species and ecosystems of a region and all their interactions, can be seen as a form of capital that is essentially needed for food and nutrition security, medical progress, the chemical industry, industrial raw materials, as well as ecosystem services like the absorption of carbon dioxide by oceans and forests (Baumgärtner 2006). However, economic research on biodiversity is still in its infancy and the few biodiversity measures that are also based on economic reasoning are still far from being operational.

The two areas of sustainability that are prominently addressed in the third pillar of the GCEE/CAE dashboard, namely economic and environmental sustainability, have been of concern already to economists and statisticians in the 19<sup>th</sup> century, as a respectable number of contributions in the early decades of the *JNS* reveal. In the economic realm, the analysis of sovereign debts and debates concerning their longer-term sustainability was already a prominent topic. In the environmental realm, questions of the sustainable consumption of renewable and non-renewable resources were also discussed.

That states face something like an intertemporal budget constraint was a well-known fact of life to the political economists of the 19<sup>th</sup> century. At their time, naturally, the fiscal sustainability was not put at risk by the implicit liabilities of a welfare state, but instead by the very explicit liabilities that were caused by the high costs of financing a standing army or even by waging war. In Prussia, public expenditures for civilian purposes exceeded the military budget for the first time in 1841 (Gerstfeldt 1883: 47). Although the use of public financial means changed considerably, and for the better, since the founding years of the *JNS*, the problem of excessive debt financing and the resulting burden of a high debt service were a topic already back then as it is now (Gerstfeldt 1883: 48-49; von Kaufmann 1887: 97-99).

The practice of comparing the structure of national budgets, so common today due to international organizations like the OECD and the European Commission, was still in its infancy time during the early decades of the *JNS*, but first steps in this direction were already made. Gerstfeldt, after he finished a meticulous compilation of summary financial statistics for Prussia whose administration did not yet publish these figures on a coherent and regular basis, compared the most relevant ratios to the respective values for France and the United Kingdom (Gerstfeldt 1883: 46). Likewise, von Kaufmann (1887) compares financial indicators for the six European Great Powers of his time (Germany, United Kingdom, France, Austria-Hungary, Italy, and Russia). The fiscal situation in the United States, which received considerable attention by 19<sup>th</sup> century economists and statisticians due to its rapid progress in many areas, were regularly a topic in the early volumes of the *JNS* (e.g. N.N. 1864b, N.N. 1879).

The most obvious difference to contemporary statistics concerning fiscal sustainability, apart from the patchy data sources in the late 19<sup>th</sup> century, is the fact that GDP was still lacking as the natural reference figure on which ratios of debt levels and debt burdens could be based. Hence, debt levels and burdens were usually calculated with reference to the total sum of revenues and made comparable across nations via per capita terms. The inadequacy of this method, given that the debt bearing capacity does not depend on the population size of a country but on its productive capacity, is only but one proof for the significant progress that the establishment of coherent national accounts and the concept of GDP provided to the disciplines of economics and economic statistics (see Bos 2011 for a comprehensive account of three centuries of macro-economic statistics).

In the ecologic realm, modern assessments of the importance of ecological sustainability find an early precursor in the statements by Hildebrand that man is merely a part of nature, ultimately acting at the mercy of nature and its unyielding laws (1863a: 19, 143). However, today's ecological concerns about the transgression of planetary boundaries and the destruction of ecosystems, obviously did not yet play a role during the early decades of the *JNS*. Nowadays, they are still on their way from the niche of ecological economics into the centre of the discipline (Leipprand/aus dem Moore 2012). At the time, environmental questions were addressed exclusively from a perspective that saw nature for and foremost as a resource that man can use.

Among the examples for this resource view on nature in the early volumes of the *JNS* are the short analysis on the range of hard coal reserves in England (N. N. 1864: 300-301), the discussion of a map that depicted production, consumption and transport of brown and hard coal in Prussia (Laspeyres 1863: 230-231) as well as treatments on the principles of forestry (Mayr 1864) and their nexus with the then emerging scientific discipline of economics (Eggert 1883). Indeed, the concept of sustainable economic activity was developed in the context of forestry and spelled out explicitly under the denomination of sustainability in the *JNS* by Eggert: "It is the concern over a sufficient use in the future that leads to a well-regulated economic activity in which a sustainable mode of production aims at using only so much in a year or period, as can be compensated for by natural growth." (Eggert 1883: 306; our translation).

## 5 Conclusions for the current debate

In a year in which the *JNS* completes its first 150 years, it seems worthwhile to look back and search for the intellectual roots of the debate on "GDP and beyond" in these far earlier times. This article documents meticulously that many of the ideas and concepts discussed today were indeed already present in the late 19<sup>th</sup> century, with the *JNS* being an important marketplace of ideas. This assessment regards both the tough contemporaneous debates on how to advance economic reasoning and the argumentation regarding the potential of statistical methods to support this endeavor. In fact, what most economists and statisticians consider to be the natural perspective on their fields of study today, was ground out by their predecessors many decades ago.

In addition to these remarkable parallels, in this paper we discuss the most important differences that distinguish the current and the past contributions to the issue. Most importantly, the three essential aspects of sustainability, economic, social and environmental sustainability enjoy a lot of emphasis today. As the world community seems to be seriously testing its planetary boundaries at the present time, it is no surprise that these issues were not as important in the late 19<sup>th</sup> century. In addition, where these early-day economists and statisticians were dreaming of a better statistical information base, current users of statistical reports and indicator systems rather face the danger of drowning in a sea of information. Consequently, now the task is making this information accessible to a large, potentially uneducated public.

We might be able to master this assignment for the very same reasons which have created the nearly overwhelming wealth of information, namely on the basis of advances in information technologies. We can indeed be hopeful that better statistics and better reporting techniques might combine to promote the state of information in the population, prepare better individual decisions and policy-making, and ultimately provide a foundation for democracy. These are the motives behind the many present-day endeavors



for the formulation of combined indicators and comprehensive and at the same time well-structured and compact indicator systems and their website presentations.

This spirit was already present some 150 years ago. Engel, for instance, states “Statistical analysis that should be useful for the present does not only need to present its results as quickly as possible after the facts it describes, but also needs to find the utmost dissemination, since the public is the inspiring and corrective element for statistical analysis.” (cited in Hildebrand 1864: 137, our translation). In conclusion, the recent attempts at constructing encompassing indicator systems reflecting the state of human welfare and societal progress, such as the prominent attempts by SSFC (2009) and GCEE/CAE (2010), are indeed “standing on the proverbial shoulders of giants” (GCEE/CAE 2010: 3). Since its early years, the *JNS* has been an indispensable place for the publications of these great minds of the late 19<sup>th</sup> century, and now, some 150 years later, economics and statistics owe them a tremendous amount of respect.

## References

- ABS – Australian Bureau of Statistics (2010), Measures of Australia’s Progress – Is Life in Australia getting better? Internet: [www.abs.gov.au/about/progress](http://www.abs.gov.au/about/progress).
- Baumgärtner, S. (2006), Measuring the Diversity of What? And for What Purpose? A Conceptual Comparison of Ecological and Economic Biodiversity Indices. Internet: <http://ssrn.com/abstract=894782> (last accessed: 14.11.2012)
- Bos, F. (2011), Three centuries of macro-economic statistics. MPRA Paper No. 35391. Internet: <http://mpira.ub.uni-muenchen.de/35391/> (last accessed: 14.11.2012)
- Bötzow, C. (1883), Die in Europa und den Vereinigten Staaten von Amerika erteilten Patente 1842-1882. *Jahrbücher für Nationalökonomie und Statistik* 41: 571.
- CIW (2011), How are Canadians really doing? Highlights: Canadian Index of Wellbeing 1.0. Canadian Index of Wellbeing and University of Waterloo, Waterloo, Ontario.
- Conrad, J. (1878), Bruno Hildebrand †. *Jahrbücher für Nationalökonomie und Statistik* 30: I-XVI.
- Deutscher Bundestag (2012), Arbeitsbericht der Enquete Projektgruppe 2 “Entwicklung eines ganzheitlichen Wohlstands- bzw. Fortschrittsindicators”. Enquete-Kommission Wachstum, Wohlstand, Lebensqualität, Kommissionsdrucksache 17(26)72 neu, Berlin.
- Drobisch, M. W. (1871), Ueber die Berechnung der Veränderungen der Warenpreise und des Geldwerths. *Jahrbücher für Nationalökonomie und Statistik* 16: 143-155.
- Eckersley, R. (ed.) (1998), Measuring Progress: Is life getting better? CSIRO Publishing, Collingwood, Victoria, pp.3-34.
- Eggert, U. (1883), Die ersten Beziehungen zwischen Forstwirtschaft und Volkswirtschaft. *Jahrbücher für Nationalökonomie und Statistik* 40: 305-343.
- European Commission (2009), GDP and beyond – Measuring progress in a changing world. Communication from the Commission to the Council and the European Parliament, COM 2009-433, Brussels.
- European Commission (2011), Sustainable development in the European Union – 2011 monitoring report of the EU sustainable development strategy. Luxembourg.
- GCEE/CAE – German Council of Economic Experts and Conseil d’Analyse économique (2010), Monitoring Economic Performance, Quality of Life and Sustainability: Joint Report as Requested by the Franco-German Ministerial Council. Wiesbaden: Statistisches Bundesamt.
- German Bundestag (2010), Setting up of a Study Commission on “Growth, Wellbeing and Quality of Life – Paths to Sustainable Economic Activity and Social Progress in the Social Market Economy”. Printed paper 17/3853, Berlin.
- Gerstfeldt, P. (1883), Beiträge zur Statistik der Finanzen in Preußen. *Jahrbücher für Nationalökonomie und Statistik* 41: 1-61.
- Hall, J., E. Giovannini, A. Morrone, G. Ranuzzi (2010), A Framework to Measure the Progress of Societies. OECD Statistics Working Paper 2010-05.



- Hildebrand, B. (1863a), Die gegenwärtige Aufgabe der Wissenschaft der Nationalökonomie. *Jahrbücher für Nationalökonomie und Statistik* 1: 5-25, 137-146.
- Hildebrand, B. (1863b), Die neuesten Untersuchungen über die mittlere Lebensdauer. *Jahrbücher für Nationalökonomie und Statistik* 1: 605-631.
- Hildebrand, B. (1863c), Die statistische Aufgabe der landwirthschaftlichen Vereine. *Jahrbücher für Nationalökonomie und Statistik* 1: 478-488.
- Hildebrand, B. (1863d), Vorwort. *Jahrbücher für Nationalökonomie und Statistik* 1: 1-4.
- Hildebrand, B. (1864), Die amtliche Statistik Preussens und ihre neuesten Publicationen. *Jahrbücher für Nationalökonomie und Statistik* 2: 136-154.
- Hildebrand, B. (1866), Die wissenschaftliche Aufgabe der Statistik. *Jahrbücher für Nationalökonomie und Statistik* 6: 1-11.
- Hildebrand, B. (1872), Die Verdienste der Universität Jena um die Fortbildung und das Studium der Staatswissenschaften. *Jahrbücher für Nationalökonomie und Statistik* 18: 1-11.
- Hildebrand, B. (1873), Die Eisenacher Versammlung zur Besprechung der socialen Frage und Schmoller's Eröffnungsrede. *Jahrbücher für Nationalökonomie und Statistik* 20: 1-6.
- Kassenböhrmer, S. C., C. M. Schmidt (2011), Beyond GDP and Back: What is the Value-added by Additional Components of Welfare Measurement? *Ruhr Economic Papers* No. 239.
- Knapp, G. F. (1871), Die neuern Ansichten über Moralstatistik. *Jahrbücher für Nationalökonomie und Statistik* 16: 237-249.
- Kollmann, P. (1912), Die Statistik in Deutschland nach ihrem heutigen Stand. *Jahrbücher für Nationalökonomie und Statistik* 99: 414-423.
- Kroll, C. (2011), Measuring Progress and Well-Being – Achievements and Challenges of a New Global Movement. Friedrich-Ebert-Stiftung, Berlin.
- Laspeyres, E. (1863), Karte über die Production, Consumption und Circulation der mineralischen Brennstoffe in Preussen während des Jahres 1862. *Jahrbücher für Nationalökonomie und Statistik* 1: 230-232.
- Layard, R. (2005), *Happiness – Lessons from a New Science*. New York: Penguin.
- Leipprand, A., N. aus dem Moore (2012), Die natürlichen Grenzen der Ökonomie: Plädoyer für eine ökologische Schuldenbremse. *Leviathan – Berliner Zeitschrift für Sozialwissenschaft* 40(2): 173-201.
- Maddison, A. (2001), *The World Economy: A Millennial Perspective*. Paris: OECD.
- Manski, C. F. (1995), *Identification Problems in the Social Sciences*. Harvard University Press.
- Matheson, J. (2011), *National Statistician's Reflections on the National Debate on Measuring National Well-being*. London: Office for National Statistics.
- Mayr, G. (1864), *Ueber Forstverwaltungsgrundsätze*. *Jahrbücher für Nationalökonomie und Statistik* 2: 433-454.
- Meadows, D. H., D. L. Meadows, J. Randers, W. W. Behrens III (1972), *The Limits to Growth*. New York: Universe Books.
- Michalos, A. C., B. Smale, R. Labonté, N. Muharjarine, K. Scott, K. Moore, L. Swystun, B. Holden, H. Bernardin, B. Dunning, P. Graham, M. Guhn, A. M. Gadermann, B. D. Zumbo, A. Morgan, A.-S. Brooker, I. Hyman (2011), *The Canadian Index of Wellbeing*. Technical Report 1.0, Canadian Index of Wellbeing and University of Waterloo, Waterloo, Ontario.
- Müller, J. (1912), Versuch einer Statistik des Vermögens deutscher Städte. *Jahrbücher für Nationalökonomie und Statistik* 98: 344-375.
- N. N. (1864a), Nachhaltigkeit der englischen Steinkohlen-Ausbeute. *Jahrbücher für Nationalökonomie und Statistik* 2: 300-301.
- N. N. (1864b), Die Staatsschulden der Vereinigten Staaten v. Nord-Amerika. *Jahrbücher für Nationalökonomie und Statistik* 2: 301-302.
- N. N. (1870), Erlass, betreffend die Zusammensetzung, Stellung und Geschäftsführung der statistischen Centralkommission im Königreich Preussen. *Jahrbücher für Nationalökonomie und Statistik* 15: 111-113.
- N. N. (1879), Brutto- und Netto-Schuld der Ver. Staaten, ferner Gesamt-Bevölkerung, Schulden und Zinsen pro Kopf. *Jahrbücher für Nationalökonomie und Statistik* 33: 340-341.
- Neumann, Fr. J. (1872), Unsere Kenntniss von den socialen Zuständen um uns. *Jahrbücher für Nationalökonomie und Statistik* 18: 278-341.

- OECD (2007), Istanbul Declaration,  
Internet: <http://www.oecd.org/newsroom/38883774.pdf> (last accessed: 14.11.2012)
- Sartorius von Waltershausen, A. (1882a), Arbeitszeit und Normalarbeitstag in den Vereinigten Staaten von Amerika, I. Arbeitsbedingungen in der Vergangenheit und der Gegenwart, im Osten und Westen. *Jahrbücher für Nationalökonomie und Statistik* 38: 461-473.
- Sartorius von Waltershausen, A. (1882b), Arbeitszeit und Normalarbeitstag in den Vereinigten Staaten von Amerika, II. Agitation für die Abkürzung der Arbeitszeit und für den Normalarbeitstag. *Jahrbücher für Nationalökonomie und Statistik* 39: 107-146.
- Schmidt, C.M. (2007), Policy Evaluation and Economic Policy Advice. *AStA – Advances in Statistical Analysis* 91: 379-389.
- Schmoller, G. (1873), Die Eisenacher Versammlung zur Besprechung der socialen Frage und Schmoller's Eröffnungsrede. *Jahrbücher für Nationalökonomie und Statistik* 20: 6-12.
- Sen, A. (1999), *Development as Freedom*. University Press, Oxford.
- Siebert, H. (1975), *Ökonomie der Umwelt: Ein Überblick*. *Jahrbücher für Nationalökonomie und Statistik* 188: 119-151.
- Soetbeer, A. (1879), Umfang und Vertheilung des preussischen Volkseinkommens im Jahre 1879. *Jahrbücher für Nationalökonomie und Statistik* 34: 112-119.
- Soetbeer, A. (1882), Preussisches Volkseinkommen i. J. 1881. *Jahrbücher für Nationalökonomie und Statistik* 39: 230-240.
- SSFC – Stiglitz, J.E., A. Sen, J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris.
- Statistisches Bundesamt (2012), *Nachhaltige Entwicklung in Deutschland – Indikatorenbericht 2012*. Wiesbaden.
- von Kaufmann, R. (1887), Finanzstatistisches zu den Schulden der europäischen Großmächte. *Jahrbücher für Nationalökonomie und Statistik* 49: 97-150.
- von Neumann-Spallart, F.X. (1885), Ein internationales statistisches Institut. *Jahrbücher für Nationalökonomie und Statistik* 45: 220-237.
- Wahl, S., M. Schulte, E. Butzmann (2010), *Das Wohlstandsquartett – Zur Messung des Wohlstands in Deutschland und anderen früh industrialisierten Ländern*. Denkwerk Zukunft – Stiftung kulturelle Erneuerung, Bonn.
- WCED – United Nations World Commission on Environment and Development (1987). *Our Common Future*, Oxford.
- Westergaard, H. (1885), Zur Theorie der Statistik. *Jahrbücher für Nationalökonomie und Statistik* 44: 1-23.

Dipl.-Vw. Nils aus dem Moore, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Büro Berlin, Hessische Straße 10, 10115 Berlin, Germany.  
[nils.ausdemmoore@rwi-essen.de](mailto:nils.ausdemmoore@rwi-essen.de)

Prof. Dr. Christoph M. Schmidt, Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, 45128 Essen, Germany, and Ruhr-Universität Bochum, IZA (Bonn), and CEPR (London).  
[praesident@rwi-essen.de](mailto:praesident@rwi-essen.de)

# Sustainability of Public Debt in Germany – Historical Considerations and Time Series Evidence

Heiko T. Burret, Lars P. Feld and Ekkehard A. Köhler\*

Walter Eucken Institut and Universität Freiburg

JEL H62; H63

Cointegration; impulse response functions; public debt; structural breaks; sustainability; unit roots; VAR; VECM.

Received: 07.03.2012

Revision received: 02.12.2012

Accepted: 29.01.2013

---

## Summary

We analyse German public finances against a theoretical background using a unique database, retrieved from multiple sources covering the period between 1850 and 2010. Multiple currency crises and force majeure offer anecdotal evidence contradicting the historical perception of Germany being the poster child of European public finance. Given these corresponding breaks in time series, the empirical analysis is conducted for the sub-periods 1872-1913 and 1950-2010. In addition to anecdotal historical analysis, we conduct formal tests on fiscal sustainability, including tests on stationarity and cointegration and the estimation of Vector Autoregression (VAR) and Vector Error Correction Models (VECM). While we cannot reject the hypothesis that fiscal policy was sustainable in the period before the First World War, the tests allow for a rejection of the hypothesis of fiscal sustainability for the period from 1950 to 2010. This evidence leads to the conclusion that Germany's public debt is in dire need of consolidation. Albeit constituting a much needed reform to this development, the incompleteness of the German debt brake and fiscal federalism will have to be addressed in the coming years, in order to ensure that fiscal consolidation actually takes place – for the sake of public debt sustainability.

*Experience, however, shows that neither a State nor a Bank ever have had the unrestricted power of issuing paper money without abusing that power.*

David Ricardo (1817/1932: ch. XXVII)

## 1 Introduction: sovereign debt as a recurring issue in economics – the JNS at 150 years

Sovereign debt sustainability has attracted attention in the wake of the Great Recession and its European sibling, the Euro crisis. Both events have painfully reconfirmed the negative effects of the growing costs of servicing an ever increasing amount of public

---

\* The authors are grateful to two anonymous referees and the managing editor, Peter Winker, for valuable comments and suggestions that have led to a considerable improvement of the paper.

debt on economic growth, monetary stability and public finance. Despite having a seemingly sound economy when compared with its crisis-ridden southern European periphery Germany, too, may account for unsustainable government finances. The German debt-to-GDP (Gross Domestic Product) ratio has never been higher during peacetime; the six year average increase in nominal debt has only been outperformed once – during the German hyperinflation. A unique database, retrieved from multiple sources covering the period from 1850 to 2010, provides us with anecdotal and empirical evidence that rejects the hypothesis of German fiscal sustainability – at least for the period after the Second World War. Therefore, we support a fiscal consolidation strategy and refute the perception that German fiscal policy is on a sustainable path.

The Journal of Economics and Statistics (JNS) has reported on the phenomenon of sovereign debt in Germany since its inception. With its publications spanning a period of 150 years, the JNS proves to be an outstanding resource for a re-evaluation of the scholarly discourse on public debt over generations of economists, mirroring paradigm shifts and large scale exogenous events. In addition and most relevant to empirical economists, the JNS offers a primary source for historical data, especially before official records of the *Statistisches Jahrbuch für das Deutsche Reich* (Statistical Yearbook of the German Reich) were published.<sup>1</sup> Both characteristics turn the JNS into a unique resource for our attempt to re-discuss the development of public debt in Germany from 1850 to 2010 in a conjoint anecdotal and empirical analysis.

As the complete record of JNS volumes offers a unique literature set regarding the historicity of the discussion on public debt in the discipline of economics we first outline some major contributions to prove the JNS' impact on this recurring issue. The first volume of the JNS already features an article dealing with a debt related problem in the realm of public finance: Hildebrand (1863) explores the monetary effects of the debase-ment of money in the United States that was undertaken to finance the American War of Independence. In fact, the majority of the following 232 volumes include analyses of issues related to sovereign debt. The geographic focus does not rest exclusively on Germany and Prussia (Warschauer 1883; Gerstfeldt 1883; Frick 1890): JNS' authors cover the public finances of many of the young European nation states, e.g., Ireland (Raffalovich 1885), Denmark (Peterson-Studnitz 1888), Hungary (Mandello 1897), and other European states (Paasche 1878; Kaufmann 1887; Raffalovich 1888; Eheberg 1892; von Heckel 1900). The American continents were subject to analyses concerning public debt as early as the second half of the 19<sup>th</sup> century, e.g., Argentina (Hübner 1891) or the United States (Hildebrand 1863).

The early 20<sup>th</sup> century volumes are especially devoted to issues of war financing, e.g., Heinemann (1914), Graner (1915), Köppe (1916a, 1916b, 1916c, 1916d, 1916e, 1918a, 1918b, 1919), Inhülsen (1919), Regensburger (1921), and war related economic consequences, particularly with respect to the relation between debt and inflation, e.g., Menger (1892), Seidler (1894), Helander (1915), Heyn (1919), Terhalle (1923), Rittershausen (1943). The JNS thus offers evidence for the emergence of a remarkably strong bias among German economists against inflation and non-sustainable fiscal policy, even before WWII. However, by 1933 a paradigm shift had swayed German economic thought in the direction of nationalist economics. The resulting tensions between the editorial

---

<sup>1</sup> The widely used historical debt dataset of Hoffmann (1965) refers, among others, to the contributions by Warschauer (1883) and Kaufmann (1887) that were first published in the JNS. Without these unique contributions, our empirical analysis would probably have started in 1871.

board of the JNS and the National Socialists led to a surge in contributions that focused more on foreign countries and theoretical economics rather than on the national socialist economics (Krawehl 1986). Aside from minor concessions, the commitment of the editors Otto von Zwiedineck and Gerhard Albrecht was successful in curbing the NS influence on the JNS until its cessation in 1944 (Ott/Strecker 1986).<sup>2</sup> The Journal was relaunched five years later.

The initial focus of the post-war period lay on emerging Keynesian ideas and the methodological re-orientation of economics towards the neoclassical school of thought and towards macroeconomics, marking another paradigm shift in German economic thought. Accordingly, public debt was hardly an issue until the 1970s (Lampert et al. 1986). Subsequent volumes include both theoretical and empirical analyses of issues related to public debt, with the frequency of the latter increasing over time. The interdependency of public credit and price stability is the subject of numerous analyses in the JNS (e. g., Wittmann 1966; Lang/Welzel 1992). In addition, problems regarding the measurement of different fiscal variables (Scherf 1989; Larch 1993; Brümmerhoff/Reich 1999), limits to public borrowing (Mückl 1985), effects of budget deficits (Jaeger 1982, Lachmann 1984; Aschinger 1985), and the impact of anti-cyclical fiscal policy (Oberhauser 1985; von Zameck 1988) are addressed.

More recent contributions focus mainly on the effect of various institutional arrangements regarding public debt (Lang/Welzel 1992; Wenzel/Wrede 2000; Blume et al. 2008; Schulz/Wolff 2009). Institutional changes in the wake of the European debt crisis are covered by Meyer (2011) regarding the costs of the European Financial Stability Facility and the independence of the European Central Bank. In one of the latest volumes of the JNS, Spahn (2012) addresses the issue of Target 2 balances in the European Monetary Union. If there is a common denominator in the publication record of the JNS, it can be summed up in the two lessons economists have learned from the debt crisis: first, that institutions matter, and second, that sound fiscal policy works best in combination with a rule-oriented approach.

On the basis the JNS has laid down, we are conducting an analysis of the sustainability of public finances in Germany. The remainder of this paper is organized as follows: Section 2 draws on some theoretical considerations regarding fiscal sustainability and reviews empirical studies. In Section 3, anecdotal evidence on the development of German public debt and the relation between economic growth and interest rates is provided. Section 4 presents the empirical strategy and the results. Section 5, finally, concludes the paper.

## 2 Theoretical framework and empirical literature review

Most studies on fiscal sustainability are based on the theoretical foundation by Domar (1944). According to his widely accepted theoretical approach, public finances can be classified as sustainable, if the growth rate of GDP exceeds, or at least equals the growth rate of public debt. The Present Value Budget Constraint (PVBC) that is closely related to the approach of Domar can be described as:

$$d_0 = - \sum_{t=1}^{\infty} \left( \frac{1+y}{1+r} \right)^t * p_t + \lim_{T \rightarrow \infty} \left( \frac{1+y}{1+r} \right)^T * d_T, \quad (1)$$

<sup>2</sup> For a further analysis of the JNS during the Third Reich see Jansen (2000: 181 ff.).

where  $y$  is real GDP growth,  $r$  is the real interest rate,  $d$  denotes the debt-to-GDP ratio,  $p$  is the primary budget balance to GDP ratio and  $t$  indicates the period.

Equation (1) implies two complementary definitions of fiscal sustainability, which may be tested in empirical research. One line of empirical studies tests if the transversality condition holds, i.e. whether the second term on the right hand side equals zero, by conducting stationarity tests for public debt and deficits, respectively. If the transversality condition is satisfied, public finances are expected to be stationary, i.e., sustainable. Hamilton and Flavin (1986) conducted one of the first empirical studies on the sustainability of public finances by applying this approach to the U.S. fiscal policy.<sup>3</sup> The second line of empirical studies tests, whether current public debt equals the first term on the right hand side of eq. (1), i.e. if current public debt equals the discounted future primary budget surpluses (Kirchgässner/Prohl 2008). This approach commonly applies cointegration models for testing. Early studies in this line of research include Elliot and Kearney (1988), Hakkio and Rush (1991), Smith and Zin (1991), MacDonald (1992), Tanner and Liu (1994), Liu and Tanner (1995), Ahmed and Rogers (1995), Haug (1995) and Quintos (1995).

The first analysis of German public finances was presumably conducted by Grilli (1988). Using data for the period from the early 1950s to 1986, he concludes that the hypothesis of non-stationarity (unsustainable) fiscal deficits can be rejected. While most subsequent studies basically apply similar approaches, the findings are not unambiguous (see Afonso 2005 for an overview). Since the late 1990s, the German debate has been enhanced by the results of cointegration tests, predominantly providing evidence in support of the hypothesis of fiscal sustainability (e.g., Payne 1997; Bravo/Silvestre 2002; Afonso 2005). Due to their strong assumptions, the tests for stationarity and cointegration have been criticized, however, by Bohn (1995, 1998). He proposes to test whether the reaction of primary surplus to GDP ratio is sufficient to offset changes in the public debt-to-GDP ratio. If this condition holds, i.e., if primary surplus increases at least linearly with debt, then public finances are sustainable. By applying this method, Greiner et al. (2006) and Greiner and Kauermann (2007) conclude that public finances in Germany are basically sustainable. Similar conclusions are offered by Greiner and Kauermann (2008). They perform semi-parametric tests using penalized spline smoothing. Polito and Wickens (2011) find contrary results by analysing the fiscal adjustments required to reach a targeted debt ratio and by testing for stationarity. Unit root tests are also used by Kitterer (2007) in order to analyse the finances of the German states. He concludes that fiscal policy does not comply with the PVBC in 13 out of 16 states.

In contrast to the above-mentioned studies, this paper focuses on a much longer time span (1872-2010) and applies a more in-depth analysis of the sustainability of German public finances. In fact, we apply a wide range of approaches and tests, including various stationarity tests (partly allowing for structural breaks), cointegration tests, VAR tests as suggested by Bohn (1995, 1998) and the estimation of a VECM. While previous studies focus mainly on one period, we additionally cover different sub-periods, thus allowing for period-specific peculiarities in German fiscal policy. Furthermore, by performing statistical tests for all four indicators of fiscal policy, i.e., public revenue, expenditure,

<sup>3</sup> Notable subsequent stationarity tests for the U.S. debt and deficits have been conducted by Kremers (1988), Trehan and Walsh (1988, 1991) and Wilcox (1989). Other empirical tests on the sustainability of the U.S. finances have been performed by Haug (1995), Bohn (1998, 2008), Greiner and Kauermann (2007) and Polito and Wickens (2011).

deficit and debt, our approach differs from previous studies.<sup>4</sup> In consequence, this paper contributes to existing empirical literature by providing a thorough and extensive analysis of German fiscal policy since 1872. In addition to empirical evidence, anecdotal evidence is drawn upon to arrive at a more complete understanding of the development of German public debt.

### 3 History of public debt in Germany 1850-2010: anecdotal evidence

Before formally testing the sustainability of German public finances, it is useful to consider Germany's fiscal history. Figure 1 illustrates the development of German public debt from 1850 to 2010. This time span can be divided into four separate phases: The first phase spans from 1850 to 1870 and it is characterized by relatively steady levels of sovereign debt on the one hand and changing governance structures on the other. The second phase describes the peaceful era of the German Empire leading up to World War I, the so called Kaiserreich (1871-1913). According to Obstfeld and Taylor (2003), the first two phases coincide with the first era of globalization (1850-1913). The third phase includes the two World Wars and their aftermaths, each culminating in a severe debt crisis and currency reform (1923 and 1948; see also Figure 2). The fourth phase starts with the founding of the Federal Republic of Germany and comprises large scale events as the collapse of the Bretton Woods system, the German Reunification and the Great Recession. The fourth phase coincides with the second era of globalization.

#### 3.1 Phase I: debt and fiscal policy in the (North-) German Confederation (1850-1870)

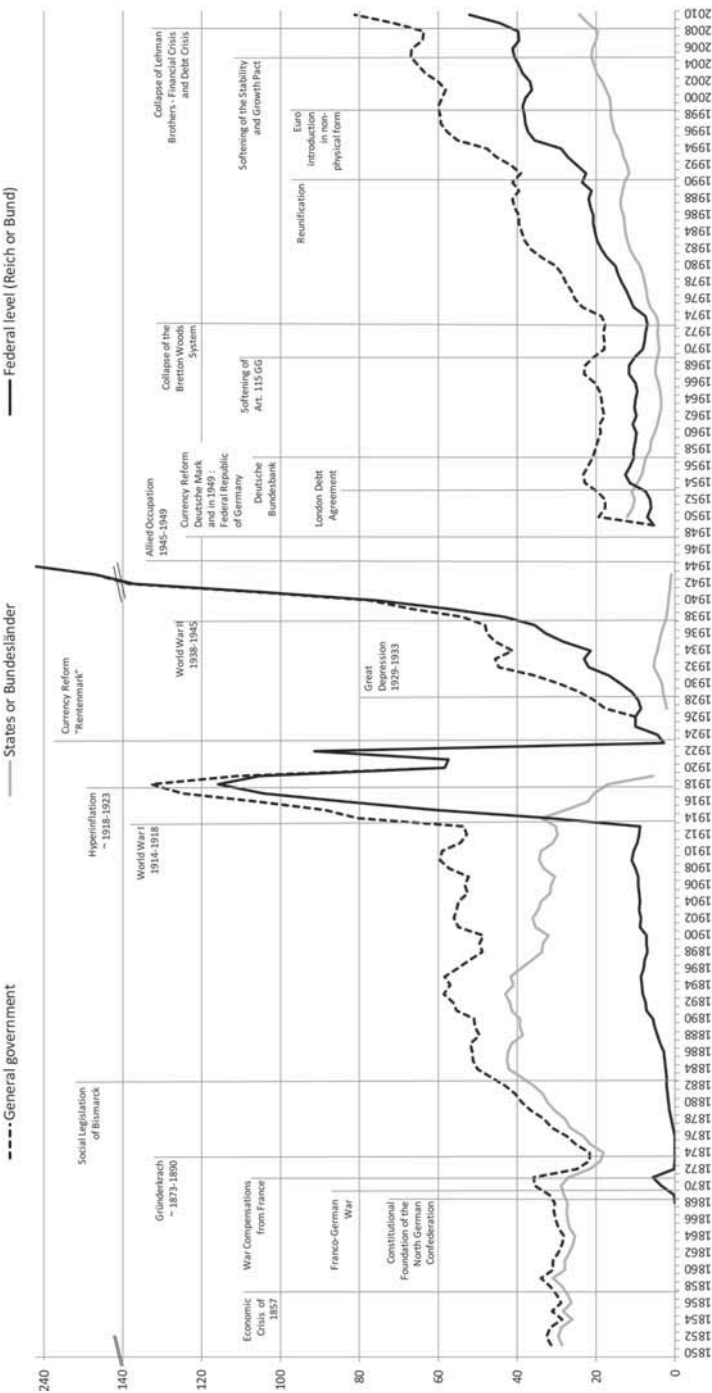
Following the unsuccessful attempt to establish a German nation state in the wake of the so-called March-Revolution of 1848, federal government debt did not accrue due to the mere lack of a central government. Public debt was, however, not non-existent: The sovereign states of the German Confederation (*Deutsche Bund*) had run fiscal deficits even before the starting point of our analysis (Figure 1). While debt statistics for most states are only available sporadically, coherent data series on public debt can be retrieved for the larger states such as Prussia, Bavaria, Saxony and Baden.<sup>5</sup> In 1851, the aggregated debt-to-GDP<sup>6</sup> ratio of these states was 30 percent. This debt can be explained by the fiscal burden of the Napoleonic Wars. However, in comparison to France (200 %) or the United Kingdom (149 %), German public debt was still relatively low (Abbas et al. 2010). This large difference might have been caused by the debt defaults of several German states, e.g., Prussia in 1807 and 1813, Hesse in 1814, Schleswig-Holstein in 1850 or Westphalia in 1812 (Reinhart/Rogoff 2009). The two decades following the March-Revolution of 1848 are characterised by ostensibly sound finances, the only exception being the economic crisis of 1857, which resulted in the Bank of Hamburg receiving a 15 million Marks bail-out from Vienna (Bordo/Schwartz 1999). Besides the relatively stable debt-to-GDP ratio, the amount of nominal debt had nearly doubled

<sup>4</sup> Indeed, Afonso and Jalles (2011) analyse a similar period (1880-2009), yet they apply almost exclusively stationarity tests regarding public debt levels, only. In addition, they do not focus on different sub-periods.

<sup>5</sup> The German Empire was later composed of 26 states (including three Hanseatic cities and Alsace-Lorraine). Missing data are derived by interpolation (Hoffmann 1965). Hoffmann (1965) retrieves some of the data from the early volumes of the JNS.

<sup>6</sup> Before 1913 no consistent data on nominal GDP are available. Instead, GNP was used (see Appendix A1).





Missing data: 1945-1948; for 1920-1926 and 1944-1949 no data for states; for 1915-1926 and 1940-1945 no data for municipalities. Before 1949 fiscal year ends 31<sup>st</sup> March/1<sup>st</sup> April. The reference date for 1923 is 15<sup>th</sup> November, 1945 debt refers to the level at the end of World War II. General government data includes federal, state and local tiers; since 1954, it includes Federal Special Funds; since 2006, it includes most, and since 2010, all public funds, institutions and companies. Federal level data includes the Mefo bills for 1932-1944. For further details refer to data sources listed in Appendix A1.

**Figure 1** Public debt as percentage of GDP in Germany, 1850-2010

by the time of the foundation of the North German Confederation in July 1867 after the victory over the Habsburg Empire in the Austro-Prussian war in 1866 (Figure 2).<sup>7</sup> The main tasks of the North German Confederation were to enlarge its navy, to build a coastal defence system and, at least partly, to finance the Franco-German War of 1870/71. Aside from matters of military, bank regulation and the standardization of weights and measures hardly any responsibilities were assigned to the federal level. Nevertheless, beginning in 1868, federal government debt had risen to 692 million Marks by March 1871.

### 3.2 Phase II: struggling decentralization of fiscal policy in the Kaiserreich (1871-1913)

The Kaiserreich, constituted in April 1871, inherited not only the debt burden of its predecessor states, but also a high degree of political autonomy, a system of bottom-up fiscal federalism and a dominant influence of Prussia. While most of the states were relatively small, Prussia comprised 60 percent of the total population and economic power, and delegated both the Chancellor (Reichskanzler) and the German Emperor (Kaiser). Despite Prussian dominance, the federal states were granted extensive legislative competence, including the power to tax. While the lower chamber of parliament, i.e., the Reichstag, remained rather weak, the Federal Council, i.e., the Bundesrat was the dominant legislative body, having been granted a veto right pertaining to all federal legislation, including taxation and the right to dissolve the Reichstag. Since small states were represented disproportionately strongly in the Bundesrat, they were – at least formally – able to overrule Prussia. In general, the competencies between the federal and the subnational level were clearly assigned, remaining mostly at the state level (Stegarescu 2005).

The high degree of fiscal autonomy allowed for (fiscal) competition between the states. Hence, fiscal responsibility of the jurisdictions was fostered and political incentives to incur debt had been curbed. Due to their political (veto) power the states were relatively successful at obstructing the expansion of central taxation powers. The federal government had to rely on revenues from tariffs (27 %), indirect taxes and matricular contributions (29 %).<sup>8</sup> The states were obliged to pay for the latter in order to balance the ordinary federal budget. Therefore, the federal government's means of accruing debt were limited. Since matricular contributions were collected in accordance with the state's population, the federal government was highly dependent on the Prussian state (Gerloff 1913). In addition to these contributions, no further fiscal equalization scheme was established.

After receiving war indemnities from France, totalling 4.2 billion Marks, the federal debt was completely paid for in 1874 and regional debt decreased significantly (Gerloff 1913).<sup>9</sup> This increase in liquidity fuelled economic growth and led to a period known as the *Gründerzeit*. The Kaiserreich experienced an extraordinary increase in production

<sup>7</sup> In contrast to the North German Confederation, the German Confederation and German Customs Union (Deutsche Zollverein) were relatively noncommittal unions of sovereign states without a common constitution.

<sup>8</sup> The numbers in parentheses indicate the respective percentages of total federal revenue, excluding debt obligations in 1871.

<sup>9</sup> The outstanding debt could be repaid at any time since the *Zwangstilgung* (coerced coverage) was abolished in 1870 and a *freie Tilgung* (free principal) established (Neumark 1976). Aside from debt repayment, the indemnities were used in order to obtain gold and subsequently introduce a gold currency.

**Table 1** Adjusted matricular contributions to the Reich (in million Marks)\*

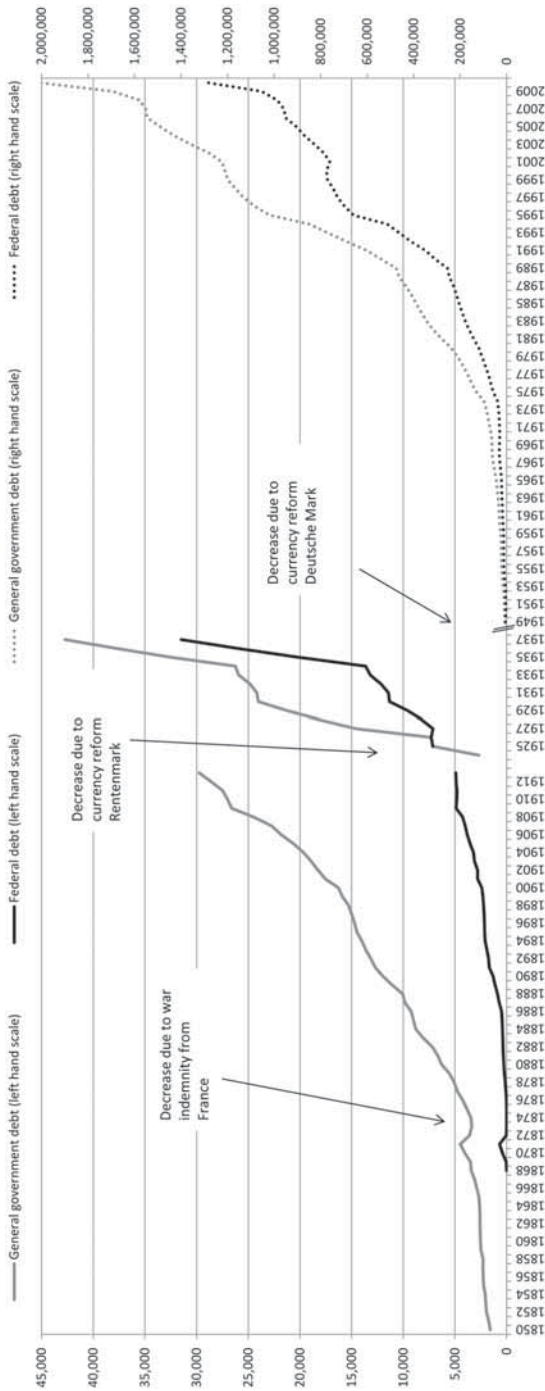
1879/80	1881/82	1883/84	1885/86	1887/88	1889/90	1891/92	1893/94	1895/96	1897/98
64.1	17.2	-11.5	-13.0	-5.4	-139.8	-66.9	30.3	-17.4	-13.2

\* The numbers are derived by subtracting the Frankenstein Payments to the states from the matricular contributions to the Reich. Source: Gerloff (1913: 522).

capacities and in the number of companies and banks. Most stock prices in Berlin doubled between 1870 and 1872. It is often argued that the Kaiserreich transitioned from an agricultural to an industrialized country during this time. Although the spoils of war and the subsequent boom of the *Gründerzeit* initially had a stabilising effect on public finances, this stability retarded the progress of fiscal reforms. The existing tax system was not designed to cope with the new structural pattern of the economy, resulting in decreased tax revenues. In 1873, economic growth slowed down due to global financial turmoil that was described as the *Gründerkrach* in Germany. This resulted in the most extensive economic crisis of the 19<sup>th</sup> century in Germany, lasting almost three decades (Plumpe 2010). The subsequent rise in public debt on the regional level (Fig. 1 and 2) is particularly remarkable, as the federal government was, at least in part, able to finance its growing expenditures by collecting higher matricular contributions. To do so, these contributions had to rise by an average of 40 percent during the period from 1874 to 1879 (Gerloff 1913: 522). Though the debt level continued to increase in the years following the *Gründerkrach*, an international comparison reveals that German public finances were in better shape than those of other countries (Appendix A3). Nevertheless, a (weak) debt constraint was imposed on the Kaiserreich. While article 73 of the federal constitution of 1871 restricted debt obligations to extraordinary circumstances, i. e., war finance, the law was obviously interpreted quite loosely.

As part of its protectionist trade policy, and in order to improve its finances, the federal government increased tariff rates in 1879. Although central government revenue had doubled by 1887, the effort was hardly rewarded. In order to prevent fiscal independence by the federal government, the states enacted the so-called Frankenstein Clause in 1879. According to this law, all federal revenues from tariffs and the tobacco tax exceeding 130 million Marks per year (the amount was increased later) had to be transferred to the states (*Gesetz betreffend den Zolltarif des Deutschen Zollgebiets und den Ertrag der Zölle und der Tabacksteuer*, § 8). The excess revenue was deducted from the states' matricular obligations, regardless of potential fiscal deficits at the federal level. The resulting reverse matricular payments (Table 1) may explain the decrease of the debt-to-GDP ratio at the state level in the following decades and, in combination with the armament of the German fleet, the rise of the federal debt-to-GDP ratio (Figure 1).

The additional increase in federal debt predating WWI was fuelled by rising colonial and educational expenses from 1885 onwards, the introduction of the social insurance systems by Bismarck (e. g., health insurance in 1883, pensions insurance in 1891) and the costs of war preparations. While the military expenses accounted for roughly 60 percent of total federal expenditures in the pre-war years, this share increased to an average of 86 percent during the war (Table 2). The percentage of social spending increased from an average of 18 percent in 1876-1880 to 31 percent in 1911-1913 (Hefeker 2001). In addition to new functions being assigned to the federal level, the loss of the states' autonomy is also reflected by the introduction of the inheritance tax in 1906, a state tax shared with the Reich, and the *Wehrbeitrag* (war contribution) in 1913, the first direct tax on the



\*Public debt in Mark, Reichsmark, and Euro, respectively. For further definition and sources, refer to Figure 1. For 1914-1923 and 1938-45 debt is not mapped due to distortive outliers. For 1945-48 no data are available. 1850-1937 refer to the left hand scale, 1949-2010 refer to the right hand scale.

**Figure 2** Nominal public debt in Germany, 1850-2010 (in millions) \*

**Table 2** Ordinary and extraordinary expenses of the Reich (in billion Reichsmark)

	Total	of which:			
		Military and War		Debt Service	
<b>1911</b>	2.12 (100 %)	1.28	(60.34 %)	0.27	(12.63 %)
<b>1913</b>	2.64 (100 %)	1.66	(63.08 %)	0.25	(9.29 %)
<b>1914</b>	8.78 (100 %)	7.76	(88.36 %)	0.47	(5.36 %)
<b>1915</b>	25.80 (100 %)	24.06	(93.22 %)	1.35	(5.22 %)
<b>1916</b>	27.84 (100 %)	24.81	(89.13 %)	2.62	(9.40 %)
<b>1917</b>	52.20 (100 %)	42.23	(80.91 %)	6.52	(12.49 %)
<b>1918</b>	44.43 (100 %)	33.95	(76.42 %)	6.77	(15.24 %)
<b>1919</b>	53.16 (100 %)	*30.64	(57.64 %)	8.40	(15.80 %)

\* Including fiscal burdens of war consequences, demobilization and reparations. *Source:* Roesler (1967).

federal level (Neugebauer 2000: 133; Stegarescu 2005). This centralization tendency is accompanied by an increase in public debt, particularly at the federal level. Figure 1 shows that the total public debt-to-GDP ratio increased by 148 percent between 1873 and 1913. As indicated by Figure 2, the amount of total nominal debt rose by 783 percent during this period. Aware of this increase in public debt, the Reich followed the example of the states and enacted a law specifying that, beginning in 1908, 3-5 percent of federal debt is to be repaid each year. However, the law never really came into effect (Neumark 1976). The state of German public finances provoked a political debate on tax reforms, particularly on the introduction of direct federal taxes. Dissent on this matter finally resulted in the resignation of chancellor Bülow in 1909.

### 3.3 Phase III: two World Wars, sovereign defaults and currency reforms (1914-1949)

The amount of debt accumulated during the war years indicates how the war was financed. With the outbreak of WWI, numerous laws with the purpose of enhancing the ability to generate public revenue came into effect. The Reichsbank, not independent in 1914,<sup>10</sup> was finally released from its duty to exchange Marks for gold, implying flexible exchange rates. As a result, the gold standard, which Schumpeter (1952) called “Die goldene Bremse an der Kreditmaschine” (the golden brake on the credit machine) was abolished. Additionally there was a de facto cessation of almost all quantitative restrictions on the money supply (Pfleiderer 1976). These amendments led the government to finance its expenditures directly through the central bank. The rise in short term treasury bills held by the Reichsbank is evidence for the extensive use of this instrument, resulting in an increasing inflation rate (Table 3). Due to the already high inflation rate, neither the introduction of new taxes nor the increase of existing ones could offset the fiscal deficits anymore. In the period between 1912 and the end of WWI, the increase in public debt exceeded the inflation-driven increase in nominal GDP, resulting in an upward shift of the debt-to-GDP ratio. In 1915 the debt-to-GDP ratio and the nominal debt of the central government exceeded the debt level of the states for the first time. Apart from a short time period after WWII, this has not changed up to today (Figure 1).

<sup>10</sup> “Die dem Reiche zustehende Leitung der Bank wird vom Reichskanzler...ausgeübt...” (Deutsches Reichsgesetzblatt Vol. 1875, Nr. 15, 177-198, as amended on: 14<sup>th</sup> March 1875, § 12). This could be translated to: The administration of the Bank in the powers of the Reich is conducted by the Chancellor (Translation by the authors).

When the Weimar Republic was established in 1919, Art. 87 of the new constitution restricted the use of public credit to extraordinary needs and to projects leading to a return. The crisis and the confusion after the First World War enabled the new Republic to reduce the fiscal and political autonomy of the states in such a way that the Reichstag became the dominant authority. In order to improve public finances, the Erzberger reform of 1919/20 altered the structure of fiscal federalism further towards a more centralized top-down approach. The federal level of government was institutionally supplemented with a financial administration and the authority to levy most taxes, leading to a unification and equalization of regional tax laws. The states were provided with shares of certain tax revenues in order to compensate them for the loss of fiscal autonomy. Furthermore, a vertical equalization scheme was established (Stegarescu 2005). The states were now fiscally dependent on the federal government. The growing federal sector may also explain the evident increase in the share of federal debt (Figure 1).

In 1919/20 the Treaty of Versailles obliged the Weimar Republic, contrary to its own intentions,<sup>11</sup> to pay reparations to the Entente. Indebtedness and inflation prevented the Weimar Republic from issuing government bonds on the capital markets.<sup>12</sup> The methods of generating public revenue were limited to taxes and the further (ab)use of the Reichsbank. Although the Erzberger reform led to an increase in various taxes, no substantial revenues could be generated in real terms (Haller 1976). Thus the government had to rely on further increases in the amount of floating debt held by the Reichsbank and in money supply (Table 3). The hyperinflation induced a decrease of the total debt-to-GDP ratio from its high level of 131 percent of GDP in 1918, with an exception in 1922, when the increase in debt exceeded the inflation driven rise in nominal GDP. In 1923, almost 90 percent of state revenue was generated by issuing debt obligations. Due to inflation, the possibility of paying contributions in government bonds and then repurchasing them, the debt-to-GDP ratio approximated zero in 1923.<sup>13</sup> The debt-to-GDP ratio of the states had already been zero in 1919, as most of their debt had been transferred to the federal level as compensation for the acquisition of the railway (Gesetz betreffend den Staatsvertrag über den Übergang der Staatseisenbahnen auf das Reich, § 4).

The hyperinflation ended in November 1923, with the currency reform and the Reichsbank ceasing to discount treasury bills. The newly issued currency, the Rentenmark (backed by real estate, land and later by gold), was exchanged for 1 trillion Marks. While smaller amounts of debt were simply deleted, the majority of outstanding public obligations were transferred according to the principle “Mark für Mark” (Mark for Mark). Thus, public debt was notably reduced. Outstanding liabilities still existed in the form of reparations, which Germany had previously suspended. Negotiations on restructuring

<sup>11</sup> In 1916 the State Secretary of the Reich Treasury announced in the Reichstag: “Das Bleigewicht der Milliarden haben die Anstifter dieses Krieges verdient; sie mögen es durch die Jahrzehnte schleppen, nicht wir.” (The burden of the billions is the merit of the instigators of this war; they may carry it through the decades, not us. Translation by the authors.) (Helfferich 1916: 224).

<sup>12</sup> Exceptions are some “value stable” bonds expressed in Goldmark, rye or other goods (Pfleiderer 1976).

<sup>13</sup> Since no data on nominal GDP are available for the time period between 1914 and 1924, we had to derive them from real GDP estimates deflated by the respective CPI. The amplitude of our data in 1922 may be due to the fact that CPI is measured at the end of the year, while debt is measured at the end of the fiscal year (i. e., the end of March of the following year). However, our estimated debt-to-GDP ratios are rather similar to those of Holtfrerich (1996) and to data kindly made available by Schularick. For further details see Appendix A1 and A2.

**Table 3** Macroeconomic indicators, 1915-1923

	Floating debt (in million Marks)			Money in circulation <sup>1,2</sup>	Exchange rate to U. S. Dollar <sup>2</sup>	Consumer Price Index
	total	Held by the Reichsbank		1913=1	1914 = parity	1913=100
<b>1915</b>	7.2	6.0	83.33 %	1.5	1.16	135
<b>1916</b>	9.3	7.3	78.49 %	1.7	1.26	180
<b>1917</b>	18.5	13.1	70.81 %	2.3	1.69	225
<b>1918</b>	33.0	15.7	47.58 %	5.5	1.97	310
<b>1919</b>	63.7	29.9	46.94 %	8.3	11.1	490
<b>1920</b>	91.5	42.7	46.67 %	13.4	17.4	1'044
<b>1921</b>	166.3	64.5	38.79 %	20.3	45.7	1'337
<b>1922</b>	271.9	146.5	53.88 %	213.0	1'808.0	15'036
<b>1923<sup>3</sup></b>	191.6tri	189.8tri.	99.06 %	81.8bn.	1'000bn.	15'897bn.

<sup>1</sup>Stückgeldumlauf (Reichsbanknoten, Privatbanknoten, Reichskassenscheine, Darlehenskassenscheine and coins). <sup>2</sup>1915-1917 at mid year, 1918-1922 end of year. <sup>3</sup>Until 23<sup>th</sup> November. Sources: Deutsche Bundesbank (1976) and Haller (1976).

these external liabilities resulted in the Dawes Plan of 1924. According to this agreement, Germany had to pay a rate of 2.5 billion Marks per annum, starting in 1928/29. An expiry date was not agreed upon. Discontent of the involved parties and the onset of the Great Depression led to a new agreement in 1929. The so-called Young Plan reduced annual payments to 2 billion Marks for the next 59 years. In order to service the payment schemes, Germany issued the Dawes and Young bond, respectively (Glasemann 1993).

Soon after reparations had been agreed upon, public finances were hit by the Great Depression, placing a high burden on public finances, especially on the local level. Two developments played a major role in this: First, the main sources of local revenue (corporate and personal income tax) decreased by over one third and, second, higher costs of social services caused local expenditures to rise (Petzina 1986: 246-251). Due to the fiscal situation during the Great Depression, the Weimar Republic introduced capital controls in 1931 and stopped servicing most of its debt in compliance with the contracts. The capital controls led to an erosion of the gold standard, which gave the Reichsbank leeway for discretionary monetary policy. After the takeover of the National Socialists in 1933, most debt payments were discontinued, the Reichsrat was dissolved and the federal structure of Germany was factually abolished. Subsequently, the share of federally determined tax revenues increased from 50 percent in 1913/14 to 98 percent in 1938 (Terhalle 1952: 317).

While there was a legal ceiling on the amount of treasury bills discountable at the Reichsbank before 1933, the National Socialists found various ways to circumvent fiscal and other disciplining rules of the Weimar Republic constitution. The rule of the National Socialists can be divided into four stages, each characterised by different methods of public finance. The practice predominantly used during the first stage of the regime (1933-1936) was the issue of bills of exchange by dummy firms redeemable at par in Reichsmark and expansive fiscal policy. The latter was already common before the National Socialists came to power: Job-creating measures were initially introduced in 1930, and their intensive use in 1932/1933 was re-financed by tax credits (Schiller 1936: 54).<sup>14</sup>

<sup>14</sup> This included the installation of institutions that financed work programmes, such as the Deutsche Gesellschaft für öffentliche Arbeiten (Oeffa) (e.g., Papen-Programm, Sofort-Programm and the Reinhardt-Programm).



In addition to these waivers, several shadow and off-budget activities became the backbone of the National Socialist work programme.<sup>15</sup> The second stage (1936-1939) of National Socialist fiscal policy was directed to install a system of “silent” war financing operations. Wages and prices were frozen in June and November 1936, while money supply was increased. In 1939, the *Kriegswirtschaftsverordnung* (§ 1, RGBl. I, 1939, 1609-1613) imposed penalties on the hoarding of money, extending to bank deposits at commercial banks. Consumer demand was capped by the introduction of consumption stamps in September 1939, while the investment operations of banks were limited and restricted to locally issued products. This boosted savings even further. Since commercial banks were prohibited to invest savings abroad (capital controls had been previously established) and were regulated to buy national securities, savings indirectly, but inevitably, ended up at the Reichsbank – or in government bonds. This “silent” war financing procedure turned the German citizens into creditors of the National Socialist war finance strategy. In 1939, by means of the “Gesetz über die deutsche Reichsbank” the bill financing arrangement became obsolete, since public expenditures could now be financed, to a nearly unlimited extent, directly through the Reichsbank (Hansmeyer/Caesar 1976).<sup>16</sup>

The beginning of the third stage (1940-1943) was marked by the occupation of neighbouring countries. Additional revenue was obtained by integrating them into the above mentioned scheme, while selling off their central bank’s securities (including their gold) on the still existent international capital market (Vogler et al. 2012). The implementation of subsidiaries of the Reichsbank in Krakow (December 1939), Brussels (June 1940) and Bohemia (August 1940) illustrates the extent of this strategy. The leverage of the war financing operations increased, as more and more citizens were forced into this scheme. The fourth and final stage (1943-1945) saw the capital inflow from the occupied territories slow down. Additionally, the limits of the war financing operations between the Reichsbank and private savings had been reached. The Reichsbank started financing public spending by directly buying bonds and printing money. The cumulated debt overhang and the loss of credibility of the institutions of the Reich finally led to a rejection of the Reichsmark in May 1945.

The pursued strategies of war finance, combined with decreasing economic growth, resulted in a rise of the debt-to-GDP ratio, exceeding 240 percent in 1944 – not including shadow budgets and other liabilities by foreigners (Figure 1). This increase in the debt-to-GDP ratio surpassed the experience of WWI and exceeded the debt-to-GDP ratios of most other nations (Appendix A3). In contrast to WWI, the freezing of prices and wages curbed inflation. A currency reform was enacted in June 1948, three years after the German capitulation (Goldschmidt/Köhler 2008).

<sup>15</sup> An often cited example is the Metallurgische Forschungsanstalt (Mefo) which was installed to emit bonds to further increase the fiscal power of the Reich. The bonds served as bills of exchange to finance rearmament, while at the same time camouflaging the war preparation from foreign observers (Cohn 1997: 271). In an interrogation, Hjalmar Schacht (1945), president of the Reichsbank from 1933-1939, said about this technique that it “enabled the Reichsbank to lend by a subterfuge to the government what it normally or legally could not do”. From a quantitative point of view though, the Mefo bills were not very important (Table 4).

<sup>16</sup> According to the “Gesetz über die deutsche Reichsbank”, as amended on 15<sup>th</sup> June 1939, the Reichsbank was essentially assigned directly to Hitler.

**Table 4** Debt and money circulation 1938-1945 (in billion Marks)

	1938/39	1939/40	1940/41	1941/42	1942/43	1943/44	1944/45 <sup>2</sup>
<b>General Government debt</b>	53.2	73.5	102.6	153.3	208.0	284.9	387.9
of which:							
Mefo Bills	11.9	11.4	10.8	10.1	9.5	8.8	8.1
New Domestic short-term debt	6.5	18.0	38.2	66.9	103.5	154.2	241.0
in %	12.2	24.6	37.2	43.6	49.8	54.1	62.2
of which:							
Reichswchsel	0.4	6.5	14.9	26.0	37.3	61.2	116.0
<b>Money circulation<sup>1</sup></b>	11.0	15.2	18.3	23.8	29.8	38.7	73.0

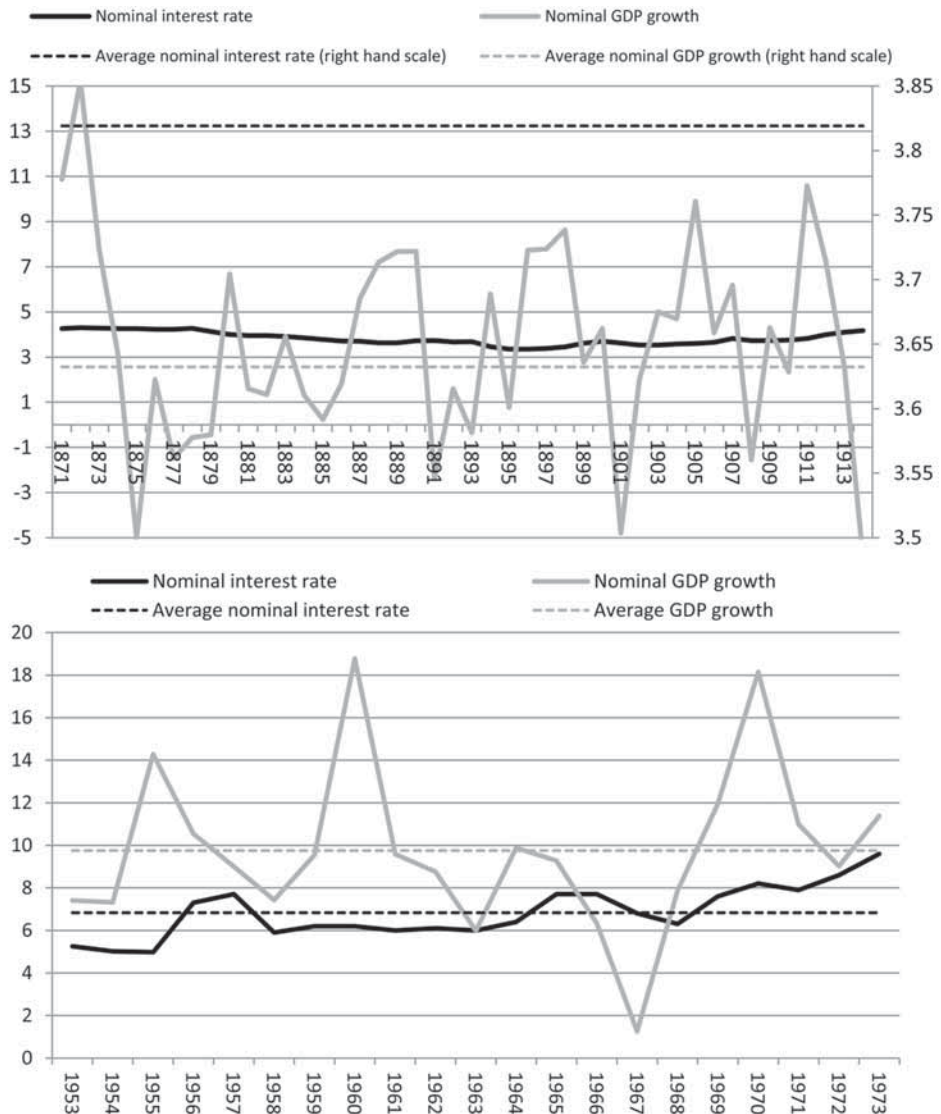
<sup>1</sup>Including Reichsbanknoten, Rentenbankscheine and Scheidemünzen; 31<sup>st</sup> June. <sup>2</sup>Partly estimates; data are valid until the end of WWII. Source: Hansmeyer and Caesar (1976).

### 3.4 Phase IV: fiscal policy in the Federal Republic of Germany (1950-2010)

The currency reform of 1948 went hand in hand with a default on liabilities of the National Socialist regime. The claims of outstanding German debt were not decided and settled before the London Debt Agreement in 1953. By that time, the Bank deutscher Länder (BdL) had already gained autonomy from the Allies and had successfully organized an asset re-allocation (i. e., Ausgleichsforderungen) to maintain the supply of capital and credit from the start of the Deutsche Mark. The Ausgleichsforderungen accounted for nearly 50 percent of the BdL balance on December 31<sup>st</sup> 1950. Larger liabilities were haircut against newly issued securities, eligible to be discounted at the Federal Reserve Banks (i. e. Landeszentralbanken) of the newly established Central Bank System headed by the BdL. In short, the debt-to-GDP ratio had decreased, since most public debt was converted into Deutsche Mark with an exchange ratio of 1:10. In 1952, Germany joined the Bretton Woods System and pegged its exchange rate to the U.S.-Dollar. This can be seen as an attempt to return to a rule-based monetary policy. In the two decades following the formation of the Federal Republic of Germany, the general debt-to-GDP ratio fluctuated around 22 percent (Figure 1). However this situation owes far more to the extraordinary economic growth during the “economic miracle” than to fiscal prudence. In fact, the average nominal GDP growth between 1953 and 1973 reached almost 10 percent (Figure 3), while the nominal debt increased from 14.78 to 86.42 billion Marks (Figure 2).

In the late 1960s, a realignment of macroeconomic policy in Germany took place. A new law authorized the government to stabilize economic cycles, using countercyclical fiscal and economic policy (StabG 1967). The implementation of this law is often interpreted as a move towards Keynesian demand side economics. In addition, the constitutional rule restricting public deficits to the volume of investment (Art. 115 GG) was amended in 1969.<sup>17</sup> In the following years both the debt-to-GDP ratio and the amount of nominal debt increased (Figure 1 and 2). While the initial rise in public debt might have resulted

<sup>17</sup> Until then article 115 GG, adopted from the Weimar constitution, allowed public credit in exceptional situations and for projects leading to a return. Following the financial reform of 1967/69, Art. 115 GG allowed the accrual of debt for public investments and to fight macroeconomic disequilibrium. Further exceptions were allowed for off-budget special funds. The article was amended with the introduction of the debt brake in 2009. For an analysis of this debt brake see Feld and Baskaran (2010).



**Figure 3** Nominal interest and growth rates in percentage points, different periods

from the first serious recession of the post-war era in the 1970s, the German Council of Economic Experts (2005: 478) and the Academic Advisory Board at the Federal Ministry of Economics and Technology (2008), among others, report a bias towards public debt since the constitutional amendment. The debate on the evident increase in public debt marked the starting point of the break-up of the social liberal coalition in 1982. The following government under Chancellor Kohl succeeded in consolidating the structural deficit until the eve of German Unification.

However, the costs of the German Unification, combined with the recession of the 1990s resulted in a new increase in public debt (Figure 1 and 2). After the establishment of the European Monetary Union and the implementation of the Stability and Growth Pact (SGP), which limits the amount of national debt and deficits as percentages of GDP to 60 percent and 3 percent, respectively, the German debt-to-GDP ratio has continuously exceeded the Maastricht benchmark of 60 percent. Due to an intervention by Germany and France, the SGP was weakened in 2005. Finally, the Great Recession has led to an increase of the general debt level to over 80 percent or above 2 trillion Euro in 2010 (Figure 1 and 2). An international comparison, however, shows that the fiscal situation is even worse in several other states (see Appendix A3).

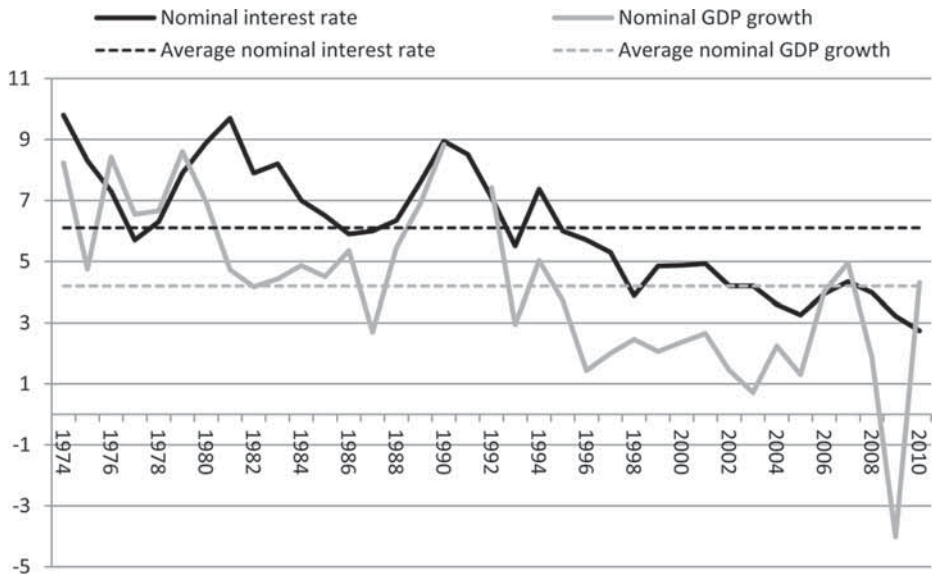
Although Germany has experienced steep increases in its public debt before, both the debt-to-GDP ratio and the nominal amount of debt have never been higher in peacetime than today. The introduction of a constitutional debt brake replacing the former Art. 115 GG is a reaction to this development. Today, the debt rule is in a transition period, giving the state and federal governments time to consolidate their budgets. Starting in 2016, the federal budget has to be close-to-balance after adjusting for cyclical fluctuations. Four years later, the states will be no longer permitted to run a structural deficit. Exceptions are narrowly defined and tied to repayment rules, but automatic triggers in the case of contravention are missing. At this stage, several states are struggling to balance their budget, thereby risking an unbalanced budget in 2020 (Deutsche Bundesbank 2011; German Council of Economic Experts 2011). Since legal restrictions to the accumulation of public deficits have been evaded in Germany since the Kaiserreich, and nowadays even on the EU level, the sustainability of public finances is still questionable.

### 3.5 Summary and further evidence

The anecdotal evidence on the development of public debt in Germany has revealed some key characteristics. First, public debt has particularly increased during economic crises, wars or extraordinary circumstances (i.e., unification). Apart from potentially “good” reasons for fiscal deficits in those times, it is evident that debt has hardly decreased afterwards. Second, constitutional rules restricting the amount of public debt were in place since the formation of the first German nation state. They have obviously been interpreted very loosely. These two characteristics can be explained by Buchanan and Tullock’s (1962) theory of the fiscal commons (Schaltegger/Feld 2009). Third, since 1949 the share of the federal debt level has increased. Fourth, the largest rises in the public debt ratios were accompanied by a loss of independence of the central bank. Fifth, although the Bundesbank gained independence and committed itself to price stability, German fiscal policy was not constrained effectively, as supposed by Sargent and Wallace (1981).<sup>18</sup>

In the current debate on the European debt crisis the call for a growth-oriented policy instead of austerity measures gains popularity. Such a policy is supposed to enable economies to grow out of their debt without the need of (primary) surpluses. As discussed in Section 2, this is only feasible under the assumption that the nominal interest rate remains below the nominal GDP growth rate in the long run. This is not the case. After

<sup>18</sup> In contrast, the fiscal theory of price level (Leeper 1991, Sims 1994, Woodford 1995) indicates that fiscal policy determines the price level, while monetary policy plays at best an indirect role. Permanent fiscal deficits may induce inflation by their effect on aggregate demand irrespective of monetary policy. In monetary unions – but also in federal states – excessive spending of one state over a long horizon may therefore lead to heterogeneous inflation patterns (Neck and Sturm 2008).



Sources and description see Appendix A1.

**Figure 4** Nominal interest and growth rates in percentage points, different periods

1974, economic growth in Germany was lower than the interest rate (Figure 4). Average economic growth is notably below the average interest rate for this period. In the period before WWI, regarding average values for this period, the interest rate also remained above the rate of economic growth (Figure 3, top). Only in the years of the “economic miracle” the average growth rate of nominal GDP comes close to 10 percent. It therefore exceeds the average interest rate (Figure 3, bottom). Given these facts, it is hardly feasible – at least for Germany – to grow out of its debt. The years of the “economic miracle” are the exception rather than the rule.

## 4 Empirical analysis

In order to test whether the development of public debt described in Section 3 is sustainable, we conduct a time series analysis. Any of the assumptions made and set in the estimations will be discussed subsequently. The four staged approach of this paper is framed by methodological remarks in the introductory (4.1) and in a concluding section (4.6).

### 4.1 Data and empirical test strategy

The empirical analysis is based on annual data covering the period from 1872 to 2010. The dataset is composed of multiple sets of institutions and sources – including the JNS (see Appendix A1). The two most commonly stated problems of long time series are inconsistency and timeliness of the data (Reinhart/Rogoff 2009). The available data is indeed not fully consistent, due to varying reporting standards and statistical procedures, causing statistical discrepancies within public spending and income statistics. The

**Table 5** Descriptive statistics

Variable	Unit	Frequency		Obs**	Mean	Std. Dev.	Min	Max
<b>Public Debt</b>	Share in GDP*	Year	1872-2010	135	0.472	0.319	0.029	2.427
			1872-1913	42	0.476	0.110	0.215	0.596
			1950-2010	61	0.365	0.183	0.176	0.812
<b>Public Expenditures</b>	Share in GDP*	Year	1872-2010	135	0.296	0.192	0.000 <sup>1</sup>	1.095
			1872-1913	42	0.147	0.058	0.362	0.356
			1950-2010	61	0.403	0.897	0.213	0.514
<b>Public Revenues</b>	Share in GDP*	Year	1872-2010	134	0.271	0.155	0.000 <sup>2</sup>	0.609
			1872-1913	42	0.143	0.071	0.013	0.356
			1950-2010	61	0.383	0.081	0.216	0.481
<b>Budget Surplus</b>	Share in GDP*	Year	1872-2010	133	-0.032	0.068	-0.521	0.143
			1872-1913	42	-0.004	0.020	-0.085	0.012
			1950-2010	61	-0.020	0.015	-0.062	0.014
<b>Primary Budget Surplus</b>	Share in GDP*	Year	1950-2010	61	-0.001	0.016	-0.048	0.042
<b>CPI Inflation</b>	Percentage change in CPI	Year	1950-2010	61	0.025	0.022	-0.062	0.075
<b>Yield</b>	Percentage points	Year	1950-2010	61	6.305	1.720	2.743	9.8

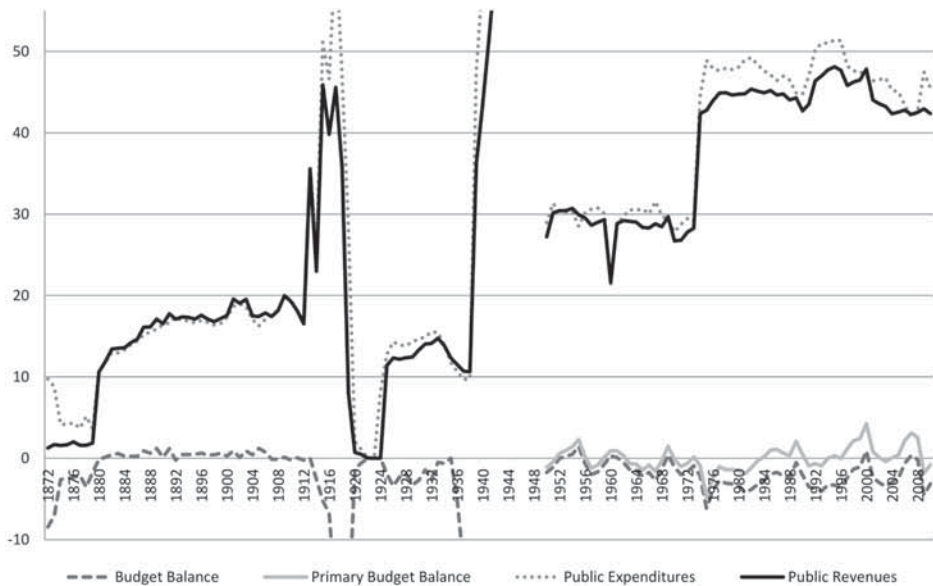
\*See footnotes 6 and 13. \*\*Observations for budget deficit in 1924 and 1949, and for revenues in 1924 are not included. Between 1945 and 1949 data are unavailable. For further details see Appendix A1. <sup>1</sup> i.e.  $7.85e^{-13}$ ; <sup>2</sup> i.e.  $8.75e^{-14}$ .

consistency of the data is improved, however, when we consider the two sub-periods 1872-1913 and 1950-2010 (see Appendix A1). In addition, some data are unavailable before 1950: Budget surplus is therefore calculated from the annual differences between revenues and expenditures, until the officially approved data series for annual surpluses started in 1950. Data on primary budget surpluses could not be obtained for this early period.

The data are measured in relation to GDP. Drawing on Bohn (2008), such a procedure guarantees a similarly scaled series that offers more credible information, in contrast to raw data or their respective logs. According to Kirchgässner and Prohl (2008) the ratios, as indicated by eq. (1), provide a more natural definition of sustainability. This holds especially with regard to the widely observed phenomenon in time series analysis on debt sustainability of a common stochastic trend in GDP on the one hand and public debt and deficits on the other hand, indicating that “stationarity of the latter two is not necessary for a sustainable fiscal policy” (ibid.: 68). Table 5 provides a summary of the descriptive statistics of all variables used in the empirical analysis. The development of expenditure, revenue and budget surpluses (in % of GDP) is shown in Figure 4.

In order to test for fiscal sustainability in a more formal manner, we basically follow the approach summarized in section 2. In a first step, we test whether the transversality condition is met (i.e. whether the second term of the right hand side of eq. (1) converges to zero over time) by conducting various stationarity tests on public debt, deficits, and, if available, primary surpluses. These tests are also performed with respect to expenditures and revenues.<sup>19</sup> In order to examine the stationarity properties of the time series, we

<sup>19</sup> For a discussion of the limits of stationarity as an indicator for sustainability, see Bohn (2008).



For reasons of clarity, values below -10 % of GDP and above 55 % of GDP are not depicted. Own calculations.  
Source: Refer to Appendix A1.

**Figure 5** Public expenditures, revenues and budget surplus (in % GDP), 1872-2010

apply different unit root tests: first, the Augmented Dickey Fuller (ADF) test, determining the number of lags using the Hannan-Quinn criterion; second, the Philipps-Perron (PP) test, selecting the bandwidth automatically in accordance with the Newey-West procedure using Bartlett kernel; and third the Kwiatkowski-Test (KPSS) with equivalent bandwidth selection procedures (Hamilton 1994; Kirchgässner/Wolters 2006; Kwiatkowski et al. 1992). The tests differ with respect to their null hypotheses: The null hypothesis of the ADF and the PP tests is the existence of a unit root in the time series, whereas the null hypothesis of the KPSS test is trend stationarity of the time series.

Since unit roots in fiscal data imply that economic shocks have a sustaining effect on the data over time, the identification of a unit root denotes a non-stationary (unsustainable) time series. All tests are applied in levels allowing for a constant and a constant with trend, respectively, and in first differences. In the presence of structural breaks, the power of standard unit root tests is decreased, e.g., the ADF test is biased towards a non-rejection of the null hypothesis. In order to take a possible distortion of structural breaks into account, we follow a twofold approach: First, we conduct the unit root and stationarity tests on the entire sample (1872-2010) and on the two sub-samples (1872-1913 and 1950-2010). Second, we additionally apply a fourth test suggested by Zivot and Andrews (1992). It tests the null hypothesis of a unit root against the break-stationarity alternative using the Akaike Information Criterion (AIC) in order to determine the optimal number of lags. The break date is chosen where the t-statistics from the ADF test is most negative, i.e., the evidence is “least favourable for the unit root null” (Glynn et al. 2007: 68). The Zivot and Andrews test (ZA) is applied in levels allowing for a structural break in the intercept and in the intercept and trend, respectively.



In a second step we perform Johansen cointegration tests to examine whether current debt equals discounted future primary budget surpluses over time (see first term of the right hand side of eq. (1) in section 2). If variables are integrated of order one  $I(1)$ , there may be a linear combination of a lower order of integration, such that a cointegration relation exists. This procedure allows for analysing whether the time series share a common stochastic drift, and to detect the rank of a cointegration matrix ( $r$ ) for a VECM analysis. Hence, a long-term relation, a cointegration vector, and a short term adjustment may be detected once we identify a cointegrated system. Cointegration tests allow for a further analysis of the data beyond standard regressions: If, for example, debt and surpluses are cointegrated, the necessary condition for the sustainability of the intertemporal budget constraint holds (Afonso 2005). Two statistical tests are able to determine the number of cointegration vectors: the Trace test and the Maximum Eigenvalue test. The former tests the null hypothesis that there are at most  $r$  cointegration vectors against an unrestricted alternative. The latter tests the null hypothesis that there are  $r$  cointegration vectors against the alternative of  $r+1$  cointegration vectors.

In a third step, we consider standard VAR tests for those variables that are not  $I(1)$ . In a VAR the endogenous variables are explained by their own lagged values and the lags of the other variables in the model. In addition, the impact of shocks on the adjustment path is modelled with impulse response functions. This procedure allows for estimating the adjustment processes on increasing public debt ratios, as suggested by Bohn (1995, 1998). Furthermore, a VAR can disclose the assumed positive reaction patterns between primary surpluses and debt (Greiner/Kauermann 2007).

In a fourth step, we test for a simultaneous equilibrium in the dataset and estimate a VECM to further explore the relation between debt, yields and inflation. Impulse response and tests for Granger causality further explore this relation.

## 4.2 Results of the unit root tests

### 4.2.1 1872-2010 unit roots tests results

The stationarity tests on the period from 1872-2010 are only indicative, as the whole period is characterized by large scale events and structural breaks. The positively skewed distribution of total debt supports this introductory remark.

Regarding expenditures and revenues, the tests indicate that both are stationary in differences while the results for their levels are ambiguous: Revenues tend to be stationary in levels with trend, as the ADF test statistic allows for rejecting the hypothesis of a unit root on the one percent significance level. In contrast, the KPSS test rejects the null hypothesis of stationarity if we allow for a constant. If we also include a trend in the estimation, we fail to reject the hypothesis of trend stationarity. Hence, revenues seem to be trend stationary in levels if we consider the ADF and KPSS tests. The results of the PP test contradict these findings. To further explore this puzzle, we supplement a ZA unit root test that is sensitive to structural breaks in the intercept and trend. If we allow for a structural break in the intercept, we cannot reject the hypothesis that revenues have a unit root. If we also allow for a structural break in the trend, the hypothesis is retained. Both tests indicate a break point in 1918 and support non-stationarity of revenues.

Expenditures in levels appear to be non-stationary over the whole period and stationary in first differences – which suggests  $I(1)$  series – if we consider ADF and PP test results. The hypothesis of stationarity can be rejected by the KPSS tests allowing for a constant. Since KPSS neither rejects stationarity in expenditure levels with trend nor in first differ-

Table 6 Tests for unit roots and stationarity, 1872-2010

Variables	ADF test			PP test			KPSS test			ZA test	
	Level	Level with trend	First differences	Level	Level with trend	First differences	Level	Level with trend	First differences	Intercept	Intercept and trend
Public Debt (in % GDP)	-0.126	-0.100	-2.011	0.174	0.212	-8.892***	0.096	0.079	0.075	-0.828 (1945)	-0.469 (1956)
Gross Surplus/Deficit (in % GDP)	3.673	3.712	0.515	3.243	3.184	-10.455***	0.135	0.107	0.150	2.784 (1945)	2.283 (1955)
Expenditures (in % GDP)	-1.652	-2.347	-5.045***	-1.247	-1.839	-10.360***	1.062***	0.027	0.113	-2.714 (1939)	-2.395 (1918)
Revenues (in % GDP)	-2.206	-4.369***	-6.095***	-1.524	-2.856	-11.561***	1.170***	0.048	0.060	-4.304 (1918)	4.423 (1918)

Note: The values are the estimated t-statistics. '\*\*\*', '\*\*' or '\*' indicates that the corresponding null hypothesis of a unit root can be rejected (KPSS: null hypothesis: no unit root) at the 1, 5, or 10 percent level, respectively. ADF lag length is automatically selected at a maximum lag length of 12 based on the Hannan-Quinn criterion. PP and KPSS bandwidth is automatically selected using Bartlett kernel. ZA uses AIC max lag length of 8. Lag length criteria tests are available in the workfile. 'yyyyy' indicates the year for the ZA break point. ZA graphs are reported in the Online Appendix.

ences, the results are, again, puzzling: With a ZA test allowing for a structural break in the intercept, we cannot reject the hypothesis that expenditures have a unit root. Also, the null hypothesis of a unit root with a structural break in both the intercept and trend cannot be rejected. Break points are 1939 with respect to the intercept and 1918 regarding a break in intercept and trend. Comparing expenditure and revenue break points, the ZA results suggest that fiscal policy of the 19<sup>th</sup> century is significantly different from that of the 20<sup>th</sup> century. The significant breakpoint in 1939 is due to the sharp increase of expenditure growth to finance WWII.

With respect to surpluses, the ADF and PP tests suggest a unit root in levels with and without a trend. They differ with regard to in first differences: The PP test rejects the hypothesis of a unit root in first differences, while ADF test does not. Trend stationarity is rejected neither in levels nor in differences by the KPSS test. The ZA test statistic does not allow for a rejection of the hypothesis of a unit root in deficits while choosing 1945 as a break point in the intercept. If we allow for a break in the trend as well, the ZA retains the hypothesis of a unit root with 1955 as a break point – the year also well known for its peak growth performance of the German “economic miracle” (Richter 1998: 32). The latter surplus coincides with its maximum value of the second period (1950-2010), as reported in Table 5 at 1.4 percent. Surpluses appear to be non-stationary in line with the ADF and the PP tests in levels.

The ADF and PP tests suggest a unit root in levels with and without a trend for public debt. They differ with regard to their test results in first differences: While ADF does not reject the hypothesis of a unit root in first differences, the PP test does so. Trend stationarity is neither rejected in levels nor in differences by the KPSS test. The ZA test statistic does not allow for a rejection of the hypothesis of a unit root in debt while choosing 1945 as a break point in the intercept. If we allow for a break in the trend and intercept, the ZA retains the hypothesis of a unit root with 1956 as a break point. Public debt appears to be non-stationary in support of the ADF and the PP tests in levels.

Given these ambiguous results, we conclude that continuing the analysis with the whole time series is not meaningful. Moreover, the power of standard unit root tests decreases substantially if there are significant structural breaks in the time series. The ZA tests indicate a multiplicity of structural breaks, which reaffirms the results from the anecdotal evidence. Therefore, we divide the sample into two sub-periods as discussed above.

#### **4.2.2 1872-1913 unit roots tests results**

With regard to public debt during the first sub-period of the dataset (1872-1913), the PP unit root tests and the KPSS tests indicate a non-stationary debt time series in levels. The ADF, however, rejects the hypothesis of a unit root in levels, without a trend on a ten percent significance level. If we allow for a constant and a trend in the estimation, the ADF retains the null hypothesis. In addition, the ZA rejects neither the hypothesis for a unit root with a structural break in the intercept (1882), nor the hypothesis of a unit root with a structural break in the intercept and trend (1883). Both dates coincide with the start of the depression in Germany (i. e., the *Gründerkrach*) and the United States. ADF and PP tests on first differences reject unit roots on the one percent level. Also rejecting stationarity in first differences on the ten percent significance level, the KPSS questions whether the debt series is truly  $I(1)$ .

The annual budget surplus is stationary during this period with regard to the ADF and PP tests. For levels – either with a constant or with a constant and a trend – both unit root

Table 7 Tests for unit roots and stationarity, 1872-1913

Variables	ADF test			PP test			KPSS test			ZA test		
	Level	Level with trend	First differences	Level	Level with trend	First differences	Level	Level with trend	First differences	Intercept	Intercept and trend	Intercept and trend
Public Debt (in % GDP)	-2.740*	-1.782	-4.186***	-2.183	-1.019	-4.760***	0.603**	0.191**	0.352*	-2.780 (1882)	-3.571 (1883)	-3.571 (1883)
Gross Surplus/Deficit (in % GDP)	-6.665***	-5.861***	-6.484***	-7.477***	-5.603***	-6.514***	0.529**	0.214**	0.510**	-11.820*** (1880)	NA (1880)	NA (1880)
Expenditures (in % GDP)	-0.256	-2.359	-3.107**	0.292	-2.362	-2.978**	0.697**	0.121*	0.231	-3.801 (1880)	-3.964 (1880)	-3.964 (1880)
Revenues (in % GDP)	-0.955	-2.118	-3.088**	-0.219	-2.101	-3.244**	0.682**	0.172**	0.150	-4.556 (1880)	-4.527 (1880)	-4.527 (1880)

Note: The values are the estimated t-statistics. '\*\*\*', '\*\*', or '\*' indicate that the corresponding null hypothesis of a unit root can be rejected (KPSS: null hypothesis: no unit root) at the 1, 5, or 10 percent level, respectively. 'yyyyy' indicates the year for the ZA break point. For further information please see the Note to Table 6 and the EViews workfile, respectively.

tests reject the existence of a unit root at the 1 percent level. This is also reflected by the ZA test that rejects a unit root allowing for a structural break in 1880 for the intercept, but fails to estimate the results allowing for a break in the intercept and the trend as regressors may be perfectly collinear. However, the KPSS test rejects the null hypotheses of trend stationarity in first differences and levels on the five percent level. Hence, the results for budget surplus ratios are not unambiguous.

ADF and PP tests do not reject the hypothesis of a unit root in the data series of expenditures and revenues in levels. The ZA supports these results, reaffirming the structural break in 1880. The hypothesis of trend stationarity (KPSS) is rejected in levels without trend. Therefore, expenditures and revenues are  $I(1)$ . The hypothesis of a unit root in differences can be rejected at the 5 percent level according to the ADF and PP test. Similar results are obtained for revenues.

#### 4.2.3 1950-2010 unit roots test results

For the second sub-period (1950-2010), the officially reported annual budget (gross) surplus and primary surplus are available.<sup>20</sup> Public debt is non-stationary in levels as ADF and PP indicate. This is validated by the rejection of the hypothesis of trend stationarity in debt levels by the KPSS test. Unit roots can be rejected in first differences on the five percent significance level. In contrast to this finding, stationarity of the debt ratio in first differences can be rejected on the 5 percent level. Allowing for a trend in the difference estimation solves this ambiguity by not rejecting stationarity ( $0.036t_i < 0.119t_{crit,10\%}$ ). A linear trend in differences corresponds with a deterministic trend in levels. An economic explanation for this empirical result could be found in the ever increasing sovereign debt ratios since 1950. Structural breaks are found in 1968, allowing for a break in the intercept and 1972, allowing for an additional break in the trend. Both dates coincide with the paradigm shift in post-war German fiscal policy (see section 2). As unit roots cannot be rejected in levels, the ZA test supports the empirical evidence that public debt is integrated of order one.

Gross and primary deficits are, in contrast, not  $I(1)$ . ADF and PP tests conjointly reject a unit root in levels and first differences. The KPSS tests confirm these results for stationarity of both time series in first differences. The hypothesis of stationary primary deficits in levels without trend cannot be rejected, in contrast to gross surplus. Considering a trend in both tests allows for the rejection of the hypothesis of trend stationary deficits in both cases. If we allow for a structural break in the regression, however, we can reject the hypothesis of a unit root in gross and primary deficits at least on the five percent level. The break is significant in 1997 for gross deficits and in 1985 for primary deficits, and for 1975 and 1984, respectively, if we also allow for a trend in the estimation. Supporting the results of the ADF and PP test in levels, the ZA test provides further evidence that we cannot treat the series as  $I(1)$ .

The ADF and PP tests in levels reflect the results from the first period, with regard to government expenditures and revenues. The KPSS null hypothesis of trend stationarity can be rejected in levels, allowing for a constant with or without a trend at least at the 10

<sup>20</sup> We thank Mrs. Ilsemarie Lenz at the Federal Statistical Office of Germany (Statistisches Bundesamt), who has provided us all relevant missing data for interest payments during the years 1950-1962. We could not have conducted this analysis without her help and her instant replies when data related problems occurred.

Table 8 Tests for unit roots and stationarity 1950–2010

Variables	ADF test			PP test			KPSS test			ZA test	
	Level	Level with trend	First differences	Level	Level with trend	First differences	Level	Level with trend	First differences	Intercept	Intercept and trend
Public Debt (in % GDP)	1.529	-1.168	-3.403**	1.678	-1.133	-3.573**	0.911***	0.225***	0.717**	-2.315 (1968)	-4.270 (1972)
Gross Surplus/Deficit (in % GDP)	-3.558***	-3.943**	-7.161***	-3.426**	-3.892**	-13.434***	0.485**	0.191**	0.247	-5.595*** (1997)	-6.449*** (1975)
Primary Deficit (in % GDP)	-3.766***	-4.211***	-6.767***	-3.615***	-3.501**	-9.494***	0.213	0.135*	0.233	-5.207** (1985)	-5.408** (1984)
Expenditures (in % GDP)	-1.481	-1.792	-7.427***	-1.500	-1.863	-7.370***	0.729**	0.166**	0.096	-8.927*** (1974)	-8.850*** (1974)
Revenue (in % GDP)	-1.545	-1.745	-8.688***	-1.512	-1.806	-8.622***	0.745***	0.143*	0.124	-9.380*** (1974)	-9.309*** (1974)

Note: The values are the estimated t-statistics. '\*\*\*', '\*\*' or '\*' indicate that the corresponding null hypothesis of a unit root can be rejected (KPSS: null hypothesis: no unit root) at the 1, 5, or 10 percent level, respectively. '(yyyy)' indicates the year for the ZA break point. For further information please consult the Note of Table 6 and the EViews workfile, respectively.

percent level, but must be retained in first differences. The two variables therefore appear to be  $I(1)$ . Following the ZA test results, revenues and expenditures do not have unit roots if we allow for a break in the intercept, or both intercept and trend, while choosing 1974 as a break point. The frequent appearance of structural breaks in the 1970s underpins the fiscal implications of the reforms during this period, as mentioned in the anecdotal analysis.

Given the results of the unit root and trend stationarity tests of the period from 1950-2010, we conclude that the necessary condition for the sustainability of the intertemporal budget constraint is not met, as the transversality condition does not hold due to the non-stationarity of public debt. The determinants of public debt of this period will be further discussed in Section 4.5. In addition, we find evidence for a stationary deficit variable (see Section 4.4) and – as far as unit root and stationarity tests are concerned – evidence that revenues and expenditures are  $I(1)$ , which will be further explored in Section 4.3 comparing both periods.

### 4.3 Sustainability test of expenditures and revenues

In order to determine the number of cointegrating vectors in the system, we perform Johansen cointegration tests for the time periods of 1872-1913 and 1950-2010. The lag lengths are selected from two VAR models. The first model (1872-1913) retrieves the following lag lengths: 2 Akaike Information Criterion (AIC); 1 Schwarz Criterion (SC); 2 Hannan Quinn Criterion (HQ). The second model (1950-2010) retrieves the same lag structure [2 AIC; 1 SCO; 2 HQ]. Subtracting one lag length for the application of the Johansen cointegration test, which is tested in first differences, we obtain the test results as reported in Table 9a if we assume a stochastic trend in the two periods, and the test results as reported in Table 9b if we assume a trend in the series as well as in the cointegration relation.

With regard to Table 9a, we can reject the hypothesis of no cointegration at the 5 percent significance level for revenues and expenditures in both periods: Trace and Maximum Eigenvalue tests indicate one cointegration vector at the same significance level for the rank of one. Table 9b shows the values for the Johansen test assuming a trend in the data series and allowing for an intercept and trend in the cointegration relation. We reject the hypothesis of no cointegration at the 5 percent significance level for revenues and expenditures in the first period. Trace and Maximum Eigenvalue tests differ with regard to their results for the second period: The Trace test statistic does not allow for rejecting the hypothesis of no cointegration in contrast to the Maximum Eigenvalue test. Cheung and Lai (1995) show that the Trace test is more robust than the Maximum Eigenvalue test with regard to skewness and excess kurtosis of residuals of the estimation. Based on this, we reject a cointegration in the second period (1950-2010) if we assume a trend in the cointegration relation. Nevertheless, we will report a test on the above mentioned sustainability vector of  $[1, -1]$  to double-check Cheung and Lai (1995).

In line with the second step of our procedure as outlined above, we now estimate VECM models for the two sub-periods and both trend assumptions, respectively. Our main objective is to check whether the hypothesis that a one percentage point increase in revenues leads to a one percentage point increase in expenditures (and vice versa) can be rejected within the four systems: Econometrically, we analyse whether the cointegrating vector of rank 1 is  $[1, -1]$ . This vector is associated with fiscal sustainability of expenditures and revenues (Afonso 2005). Afonso (2005) assumes that sustainability



**Table 9a** Johansen test on expenditures and revenues for both sub-periods with intercept in cointegration

	Null Hypothesis	Eigenvalue	Trace Statistic	5 % critical value
<b>1872-1913</b>	None	0.469	26.370**	15.495
	At most 1	0.010	0.413	3.843
			<b>Max Eigenvalue</b>	
	0	0.469	25.956**	14.265
	1	0.010	0.414	3.843
<b>1950-2010</b>	None	0.273	21.343**	15.495
	At most 1	0.036	2.222	3.843
			<b>Max Eigenvalue</b>	
	0	0.273	19.121**	14.265
	1	0.036	2.222	3.842

Note: '\*\*\*' indicates that the null hypothesis of no cointegration can be rejected on the 5 percent level. The null hypothesis is the hypothesized number of cointegration relations, i.e. the rank of the matrix ( $r$ ). The number of cointegration relations (No. of CIR) is smaller than 1, i.e., „None“, following Trace test's null hypothesis. If the statistic is higher than the critical value, the null hypothesis is rejected. Hence, the null of „None“ is rejected as shown above. The hypothesis of „At most 1“ cannot be rejected by the Trace test. The Eigenvalue test has a slightly different null hypothesis: Null is „No. of CIR =  $r$ “ which is zero „0“ which is rejected at the 5 percent significance level. The critical values for both tests are derived from the Trace and Maximum Eigenvalue of the stochastic matrix.

**Table 9b** Johansen test on expenditures and revenues for both sub-periods with intercept and trend in cointegration

	Null Hypothesis	Eigenvalue	Trace Statistic	5 % critical value
<b>1872-1913</b>	None	0.500	34.183**	25.872
	At most 1	0.130	5.7512	12.518
			<b>Max Eigenvalue</b>	
	0	0.500	28.432**	19.387
	1	0.130	5.751	12.518
<b>1950-2010</b>	None	0.278	22.985	25.872
	At most 1	0.056	3.345	12.518
			<b>Max Eigenvalue</b>	
	0	0.278	19.538**	19.387
	1	0.056	3.446	12.518

Note: '\*\*\*' indicates that the null hypothesis of no cointegration can be rejected at the 5 percent level. The null hypothesis is the hypothesized number of cointegration equations, i.e., the rank of the matrix. (See also: Note Table 9a)

in fiscal policy is given, if the time series of expenditures and revenues are cointegrated and if the hypothesis of a “normality vector” of  $[1, -1]$  holds. Vice versa, fiscal unsustainability is the case if a cointegration relation  $[1, -1]$  can be rejected. Due to this, the vector, and its corresponding Chi-Square test, is used in recent contributions to detect whether revenues and expenditures have followed a sustainable path (e.g., Kirchgässner/Prohl 2008).

### 1872-1913 Results

The 1872-1913 VAR suggests a lag length of zero for the VECM of the cointegrated time series. Assuming a deterministic trend in the dataset and allowing for a constant in the cointegration relation provides the following system:

**Table 10** Test for sustainability vector [1, -1] without trend in cointegration relation between expenditures and revenues, 1872-1913

<b>Chi-Square(1)</b>	2.168
<b>Probability</b>	0.141
<b>Revenues(-1)</b>	1.000
<b>Expenditures(-1)</b>	-1.000
<b>Constant</b>	0.004

*Note:* The null hypothesis of the Chi-Square [1, -1] test is that a one percentage point increase in revenues leads to a one percentage point increase in expenditures.

The null hypothesis that the cointegrating vector is [1, -1] cannot be rejected by the Chi-Square on the underlying system (Table 10): The p-value of the Chi-Square statistic is higher than the five percent critical value (0.141). This is evidence that revenues and expenditures are cointegrated with a sustainability vector. Given the test result, fiscal policy has been sustainable over the period from 1872-1913, assuming a constant GDP growth rate of and allowing for a Wagnerian trend in the dataset of fiscal policy (see also Feld/Schaltegger 2010).

**Table 11** Test for sustainability vector [1, -1] with trend in cointegration relation between expenditures and revenues, 1872-1913

<b>Chi-Square(1)</b>	4.546**
<b>Probability</b>	0.033
<b>Revenues(-1)</b>	1.000
<b>Expenditures(-1)</b>	-1.000
<b>@TREND(71)</b>	-0.0001 (0.0003)
	[-0.390]
<b>Constant</b>	0.007

*Note:* '\*\*' indicates that the hypothesis of the Chi-Square test [1, -1] can be rejected on the 5 percent significance level.

We repeat the Chi-Square test, allowing for a trend in the cointegration relation (Table 11). The Johansen procedure allows for an identical variation, due to the Trace and Eigenvalue test results that conjointly indicate one cointegration equation, assuming a trend in the dataset as well as in the cointegration equation (Table 9b). As shown in Table 11, the hypothesis of a cointegrating vector of [1, -1] is rejected at least on the five percent significance level with a Chi-Square value of 4.546. Sustainability of fiscal policy could thus be doubted if we allow for a trend in the cointegration relation.

### 1950-2010 Results

Applying the same restrictions on the VECM for the second period reveals a different result compared to the fiscal policy in the German Empire before WWI.

**Table 12** Test for sustainability vector [1, -1] without trend in cointegration relation between expenditures and revenues, 1950-2010

<b>Chi-Square(1)</b>	7.128***
<b>Probability</b>	0.008
<b>Revenues(-1)</b>	1.000
<b>Expenditures(-1)</b>	-1.000
<b>Constant</b>	0.019

*Note:* '\*\*\*' indicates that the hypothesis of the Chi-Square test [1, -1] can be rejected on the 1 percent significance level.

The null hypothesis of a cointegrating vector of [1, -1] is rejected on the 1 percent significance level (Table 12). In contrast to the first period, the fiscal sustainability vector does not hold for the period between 1950 and 2010. A variation of the test with regard to a trend in the cointegration equation is supplemented (Table 13): Note that the trend is significant on the 10 percent level. However, the negative coefficient is very small, at 4 basis points.

We conclude that revenues and expenditures have not followed a sustainable path since 1950. If we allow for trends in the cointegration, neither period is associated with fiscal sustainability in revenues and expenditures. It therefore appears that unsustainability is robust even considering a GDP trend over the long run. If the series are  $I(1)$ , cointegrated and if a corresponding VECM does not reject the vector [1,-1], a trend in the cointegration equation should be minded – particularly for long time series. If the hypothesis of a normality vector is still rejected, one can assume that revenues and expenditures are unsustainable.

**Table 13** Test for sustainability vector [1, -1] with trend in cointegration relation between expenditures and revenues, 1950-2010

<b>Chi-Square(1)</b>	4.566**
<b>Probability</b>	0.033
<b>Revenues(-1)</b>	1.000
<b>Expenditures(-1)</b>	-1.000
<b>@TREND(71)</b>	-0.0004 (0.0002)
	[1.868]
<b>Constant</b>	0.008

*Note:* '\*\*' indicates that the hypothesis of the Chi-Square test can [1, -1] be rejected on the 5 percent significance level.

### 4.4 Multivariate analysis of budget deficits and debt ratios

The third step of our procedure is to estimate a VAR to further explore the determinants of gross and primary budget deficits that have been stationary in levels over each sub-period. This allows for a re-discussion of the existing empirical results on German fiscal

policy and general reaction patterns in fiscal policy with regard to the primary surplus. Bohn (1998, 2008) finds evidence for a positive conditional response of primary surpluses to increases in the debt ratio. A positive and significant reaction coefficient is commonly viewed as a condition for fiscal sustainability, following Bohn's Model-Based Sustainability (MBS) test. Even if the response is non-linear or time varying, a positive fiscal reaction to increasing public debt ratios by increasing the surplus is considered to be a necessary condition.

Mendoza and Ostry (2008) find such a reaction pattern for Germany and other countries during 1990-2005. Greiner and Kauermann (2008) find evidence for a positive reaction coefficient, however, "with a declining tendency" in a dataset from 1960 to 2003 (Greiner/Kauermann 2008: 1152). Fincke and Greiner (2009) repeat the analysis with a dataset from 1971 to 2006 that is later (2011) expanded to a period from 1950 to 2007. Their results reaffirm the positive reaction function on the one hand and the conclusion of public debt sustainability on the other hand (Fincke/Greiner 2011: 211). However they point out, that a rising debt ratio is not "compatible with sustainability in the long-run", admitting that German debt may "exceed a certain critical value beyond which sustainability is excluded" (*ibid.*). Due to these results, we ask whether there is such a positive and significant coefficient for the debt-to-GDP ratio in our data set. We are furthermore interested in the reaction of fiscal policy to changes in interest rates and inflation. Thus, we conduct this empirical analysis with two VARs, one for the primary surpluses, the other for gross surpluses.

At first, we retrieve the lag lengths for the two VARs. For the primary surpluses VAR we detect the following lags: AIC: 2 SC: 1 HQ: 1. These lengths coincide with the results for gross surpluses. Both VARs should thus be tested with one and two lags, which is also meaningful from the perspective of political economy: Fiscal path dependency on the one hand and the observation that democracies plan each fiscal year in advance on the other hand. Our objective is to analyse whether the lagged values have a significant effect on the two budget ratios. Annual yields and CPI inflation are included as controls. The results of the VARs with two lags are shown in Table 14 (and in Table A2.1 (one lag) in the Online Appendix).

With regard to Table 14, lagged debt ratios and the cost of debt (yields) are not significant for primary surpluses, while both, own lagged values and the lagged CPI, are significant at least at the five percent level. These results are supported by the gross surplus VAR. A constant is, however, only significant in the gross surplus VAR at the five percent level. Contrary to the findings mentioned above, both VARs do neither indicate a positive nor a significant coefficient for total debt. The gross surplus VAR has normally distributed residuals: We cannot reject the hypothesis of normality with a Jarque-Bera test value of 8.34. The residuals of the primary surplus VAR are however not normally distributed: With a Jarque-Bera of 65.10 we can reject the hypothesis that the residuals are multivariate normal. Therefore, the following analysis focuses on the gross surplus VAR only.

The next step is to continue the analysis with a VAR Granger causality and block exogeneity Wald Tests, which investigate and examine the causal relation of the variables in the model. This is important because the test results inform about the direction of causality among the variables, which can be unidirectional, bi-directional or neutral. The results are reported in Table 15. The Chi-Square statistic of 13.866 for debt with reference to gross surplus tests the hypothesis that lagged coefficients of debt in the regression equation of gross surplus are equal to zero. In accordance, the lagged coefficients of yields and CPI as well as the block of all coefficients are equal to zero. This is empirical

**Table 14** Results from VAR estimates of budget surpluses, 1950-2010

Variables	Primary Surplus in % of GDP	Variables	Gross Surplus in % of GDP
Primary Surplus Ratio $t - 1$	0.478 (0.155) [3.11]	Surplus Ratio $t - 1$	0.384 (0.161) [2.39]
Primary Surplus Ratio $t - 2$	-0.506 (0.140) [-3.60]	Surplus Ratio $t - 2$	-0.280 (0.118) [-2.38]
Yield $t - 1$	-0.001 (0.199) [-0.003]	Yield $t - 1$	-0.242 (0.169) [-1.43]
Yield $t - 2$	-0.042 (0.210) [-0.20]	Yield $t - 2$	-0.146 (0.190) [-0.77]
Debt Ratio $t - 1$	-0.031 (0.106) [-0.29]	Debt Ratio $t - 1$	-0.073 (0.093) [-0.78]
Debt Ratio $t - 2$	0.049 (0.111) [0.44]	Debt Ratio $t - 2$	0.030 (0.092) [0.322]
CPI $t - 1$	-0.331 (0.122) [-2.70]	CPI $t - 1$	-0.205 (0.101) [-2.04]
CPI $t - 2$	-0.174 (0.094) [-1.85]	CPI $t - 2$	-0.116 (0.078) [-1.45]
Constant	0.005 (0.011) [0.382]	Constant	0.032 (0.01) [2.83]
R-squared	0.548	R-squared	0.631
Adj. R-squared	0.476	Adj. R-squared	0.572
Sum sq. resids	0.007	Sum sq. resids	0.005
S. E. equation	0.012	S. E. equation	0.010
F-statistic	7.586	F-statistic	10.697
Log likelihood	182.852	Log likelihood	192.199
AIC	-5.894	AIC	-6.210
SC	-5.577	SC	-5.893
Mean dependent	-0.0004	Mean dependent	-0.019
S. D. dependent	0.016	S. D. dependent	0.015

Note: The numbers in parentheses indicate the standard error. The numbers in square brackets are absolute values of the estimated t-statistics.

evidence that gross surplus is influenced by yields, debt and CPI and by all three variables together. Debt, yields and CPI are Granger causal at least at the 5 percent significance levels. They are conjointly Granger causal for surplus at the 1 percent significance level. block exogeneity is, however, accepted when yields and debt are taken as dependent variables. Debt and yield are therefore not jointly influenced by the other variables. The hypothesis of block exogeneity is rejected for CPI. Surplus and yields are not Granger causal. Yields and surplus are unidirectional. Yields are Granger causal for surplus, but

**Table 15** VAR Granger causality and block exogeneity Wald Test results for budget surpluses and yields, 1950-2010

Dependent variable: Gross Surplus				Dependent variable: Yield			
Excluded	Chi-sq	df	Prob.	Excluded	Chi-sq	df	Prob.
Yield	5.993799	2	0.0499	Gross Surplus	2.029176	2	0.3626
Debt Ratio	13.86591	2	0.0010	Debt Ratio	4.356628	2	0.1132
CPI	6.442257	2	0.0399	CPI	0.130706	2	0.9367
All	28.11551	6	0.0001	All	9.051684	6	0.1707
Dependent variable: Debt Ratio				Dependent variable: CPI			
Excluded	Chi-sq	df	Prob.	Excluded	Chi-sq	df	Prob.
Gross Surplus	3.856132	2	0.1454	Gross Surplus	0.356698	2	0.8367
Yield	1.150911	2	0.5624	Yield	15.61314	2	0.0004
CPI	6.018748	2	0.0493	Debt Ratio	0.338673	2	0.8442
All	10.96006	6	0.0896	All	18.73896	6	0.0046

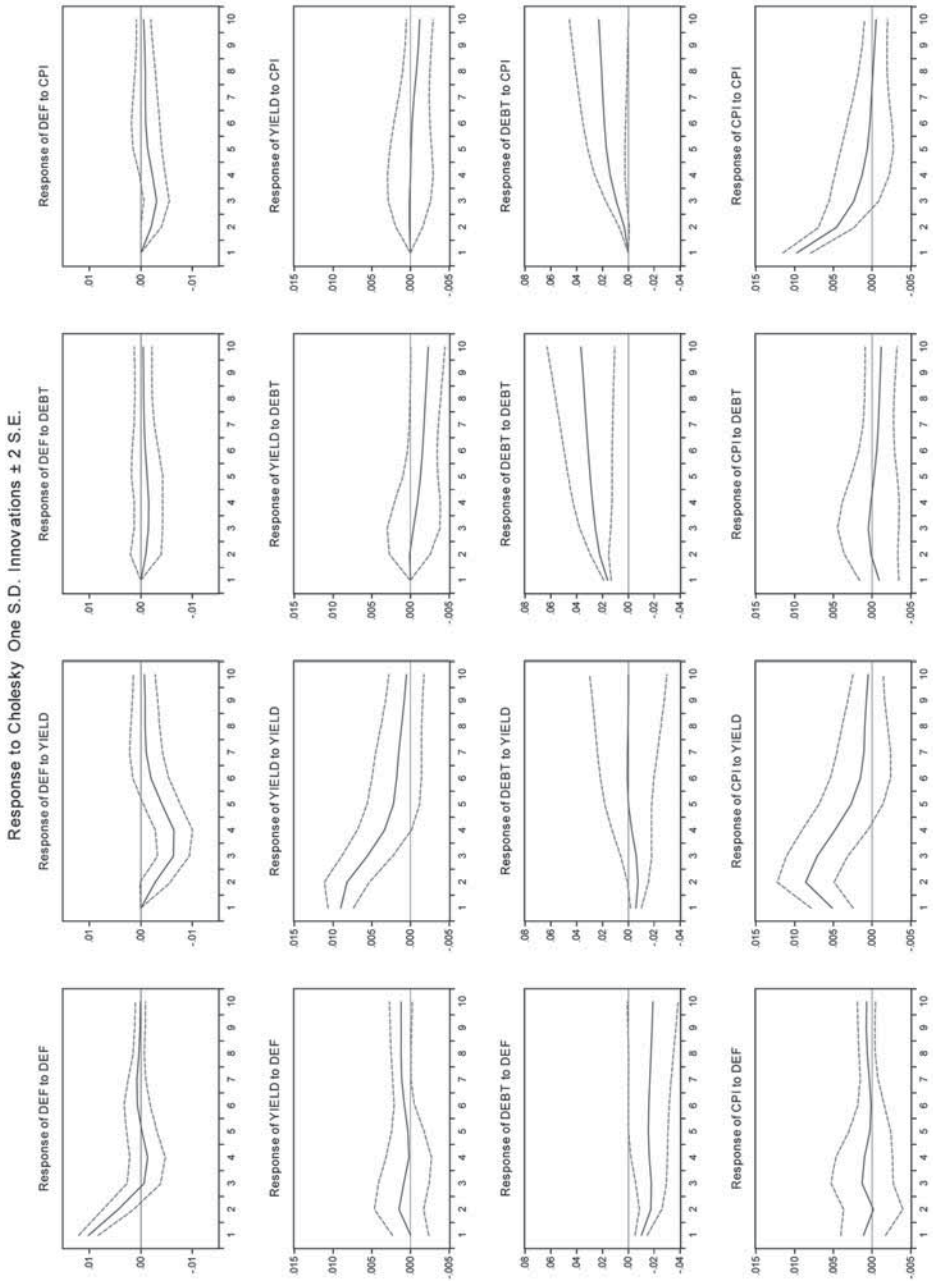
not vice versa. This is exactly the same with debt and CPI. Debt and CPI are Granger causal for surplus, but not vice versa.

The most common way to analyse VAR models is the use of impulse response functions. An impulse response analysis examines the effect of current and past shocks on the time series. We consider shocks to each residual in the system and the effects on all variables of the VAR. The impulse responses are shown in Figure 5 as solid lines. The dotted lines define confidence intervals of two standard deviations. If these confidence intervals include zero, it can be concluded that the impulse response is not significantly different from zero.

The first column of the graph shows the effect of a one standard deviation shock to the gross surplus residuals. The first of these four graphs illustrates the effect of the shock of gross surplus on itself. It rapidly disappears with a sustaining positive effect at the first lag and a small negative impact on its second lagged value as shown above in the VAR. The second, third and fourth graph of this first column show the effects of the shock on yields, debt ratios and CPI, respectively. The only evidence is that an unforeseen standard deviation increase in primary surpluses decreases CPI. With regard to the confidence intervals, CPI (yields) include zero, i. e., the impulse response is not significantly different from zero. Taking into account the complexity of determinants of CPI (and yields) this is not surprising. The confidence intervals of gross surplus and public debt do not include zero until the second and fourth year respectively. Though Granger causing surplus, debt is neither positive nor significant which is why we cast doubt on the empirical findings of the above mentioned papers. Again, primary surplus did not even pass the normality test, which is why we discarded a further analysis.

#### 4.5 Vector error correction model for government debt ratios

The fourth step is to identify a simultaneous equilibrium in the dataset with respect to government debt. Since we have not discussed debt ratios so far, we estimate a VECM to further explore the relation between yields and debt. The unit root and stationarity tests



**Figure 6** Impulse response tests for the gross surplus VAR, 1950-2010



**Table 16** Johansen test on yields and debt

	Null Hypothesis	Eigenvalue	Trace Statistic	5 % critical value
<b>1950-2010</b>	None	0.362	39.710**	29.797
	At most 1	0.196	13.235	15.495
			<b>Max Eigenvalue</b>	
	None	0.362	26.476**	21.132
	1	0.196	12.893	14.265

Note: '\*\*' indicates that the null hypothesis of no cointegration can be rejected at the 5 percent level. See also: Notes of Table 9a.

have shown that debt is I(1). A VAR is tested to retrieve the lag length for the procedure on CPI, yields and the debt ratio and throughout the procedure. The cointegration test indicates one cointegration equation at the five percent level as reported in Table 16. Allowing for a deterministic linear trend in the dataset, and a constant in the cointegration equation, the VECM results are shown in the eq. (2) to (4) below. The resulting residuals are tested for normality: The null hypothesis of normal distribution of the residuals cannot be rejected (the Jarque-Bera test statistic is 9.067). According to the KPSS test, the cointegration vector is stationary (0.112 in levels with a constant and 0.100 with trend). The following system provides robust evidence on the long and short run relation between the debt-to-GDP ratio, government bond yields and the CPI. While the long run relation is significant for yields and the debt-to-GDP ratio, it is not for the CPI. Looking at the debt equation (3), all three lagged coefficients are significant in the short run. Block exogeneity can be rejected for the debt equation at a significance level of 1 percent. CPI Granger causes debt but not vice versa. Similarly, yields Granger cause debt at a 1 percent significance level but not vice versa. This is evidence for a unidirectional relation. Block exogeneity cannot be rejected for equations (2) and (4). Yields and CPI are not jointly influenced by the other variables.

$$\begin{aligned}
 D(\text{YIELD}) = & -0.445* (\text{YIELD}(-1) + 0.004*\text{DEBTRATIO}(-1) - 0.001*\text{CPI2005}(-1) - 0.080) & (2) \\
 & [-3.040] & [5.479] & [-4.172] \\
 + 0.260*D(\text{YIELD}(-1)) + & 9.966e-05*D(\text{DEBTRATIO}(-1)) - 0.001*D(\text{CPI2005}(-1)) + 0.002 \\
 & [1.706] & [0.161] & [-1.118] & [0.727]
 \end{aligned}$$

$$\begin{aligned}
 D(\text{DEBTRATIO}) = & -83.946* (\text{YIELD}(-1) + 0.004*\text{DEBTRATIO}(-1) - 0.001*\text{CPI2005}(-1) - 0.080) & (3) \\
 & [-2.582] & [5.479] & [-4.172] \\
 + 56.140*D(\text{YIELD}(-1)) + & 0.641*D(\text{DEBTRATIO}(-1)) + 0.905*D(\text{CPI2005}(-1)) - 0.763 \\
 & [4.663] & [4.662] & [3.237] & [-1.583]
 \end{aligned}$$

$$\begin{aligned}
 D(\text{CPI2005}) = & 7.452 (\text{YIELD}(-1) + 0.004*\text{DEBTRATIO}(-1) - 0.001*\text{CPI2005}(-1) - 0.080) & (4) \\
 & [0.754] & [5.479] & [-4.172] \\
 + 23.790*D(\text{YIELD}(-1)) + & 0.034*D(\text{DEBTRATIO}(-1)) + 0.745*D(\text{CPI2005}(-1)) + 0.335 \\
 & [2.323] & [0.811] & [8.758] & [2.283]
 \end{aligned}$$

Due to the insignificance of the cointegration equation for CPI at the five percent level for the error correction model, we test whether it has a weak exogenous influence on the system by restricting the coefficient of the cointegration equation of CPI to be zero. The Chi-Square test does not allow for rejecting the hypothesis that the cointegration coefficient is zero at a p-value of 0.513. The residuals of this model are, too, normally distributed with a Jarque-Bera value of 9.686. The cointegration vector is, again, sta-

tionary (0.059 in levels with a constant and 0.058 with a constant and trend). We therefore have empirical evidence to conclude that yields have a negative influence, whereas CPI has a positive influence on debt.

A concluding remark of the test series of the unique data set in this section is straightforward: We have found econometric evidence, that Germany's fiscal policy has not been sustainable since 1950 (section 4.2). Sustainability of fiscal policy can only be assumed for the German Empire before 1913, but only if we do not allow for trends in the cointegration relation. In addition, we could not find any empirical evidence for a positive reaction of debt to primary surpluses (section 4.3). Instead, we report evidence that contradicts Fincke and Greiner (2011). The VECM (section 4.4) showed a simultaneous equilibrium. Granger causality and Block exogeneity tests showed that debt is influenced by yields and CPI, but not vice versa.

## 5 Conclusion

Public debt in Germany has continuously attracted attention. After the experience of the two World Wars, the German population is quickly alarmed when debt levels appear to be rising to unsustainable levels. This holds particularly for recent years, as Germany's debt-to-GDP ratio has never been higher in peacetime than today.

In this paper, we analyse sustainability of German public finances from 1872 to 2010. Given the breaks in the data series, in particular those induced by the two World Wars, the main analysis is conducted for the sub-periods 1872-1913 and 1950-2010. In addition to anecdotal historical evidence on the basis of studies published in the JNS, we conduct more formal tests on fiscal sustainability. While we cannot reject the hypothesis that fiscal policy was sustainable in the period before the First World War, this only holds if we do not allow for trends in the cointegration relation. The hypothesis of fiscal sustainability for the years 1950 to 2010, on the other hand, must be rejected. After the Second World War, German public finances have become unsustainable.

This evidence leads to the conclusion that public finances in Germany are in dire need of consolidation. In fact, the introduction of the debt brake in the year 2009 is a much needed reaction to this development. Although such fiscal rules always have their loopholes and are necessarily incomplete, they usually have some success in restricting public deficits and debt (Feld/Kirchgässner 2008; Feld/Baskaran 2010). The incompleteness of the German debt brake will have to be addressed in the coming years in order to ensure that fiscal consolidation actually takes place. One shortcoming of the new debt rule requires a wider ranging reform, however: The Länder (including their local jurisdictions) not only have huge consolidation requirements, they also do not have the tax autonomy to balance the spending demands on their budgets. The next major reform of the German fiscal constitution should thus allow for more tax autonomy at the sub-federal level.

## Appendix

### A1 Data

Variable	Name	Description	Source
<b>CPI</b>	cpi1913	consumer price index (1913/14=100).	1877-1944: Deutsche Bundesbank (1976)
	cpi2005	consumer price index (2005=100).	1949-2010: Deutsche Bundesbank (2012)
<b>General government debt</b>	debtgen	total public debt in million Mark/Euro. Including Reich/Bundesrepublik, states/Bundesländer, municipalities (where available), most public funds.	1850-1949: Own calculations (federal debt + regional debt + municipal debt) 1950-2010: Statistisches Bundesamt (2011a)
<b>Federal debt</b>	debtfc	debt of the Reich/Bundesrepublik in million Mark/Euro. Fiscal year until 1945: 31 <sup>st</sup> March, afterwards 31 <sup>st</sup> December. 1914-1944 including Mefo bills. 1949 including equalisation claims (Ausgleichsforderungen).	1850-1913: Hoffmann (1965) 1914-1944: Deutsche Bundesbank (1976) 1949: Bank deutscher Länder (1950) 1950-2010: Statistisches Bundesamt (2011a)
<b>Regional debt</b>	debtfl	Debt of the states/Bundesländer in million Mark/Euro. Fiscal year until 1945: 31 <sup>st</sup> March, afterwards 31 <sup>st</sup> December.	1850-1913: Hoffmann (1965) 1914-1944: Deutsche Bundesbank (1976) 1950-2010: Statistisches Bundesamt (2011a)
<b>Municipal debt</b>	debtm	Municipal debt in million Mark/Euro. Fiscal year until 1945: 31 <sup>st</sup> March, afterwards 31 <sup>st</sup> December. 1850-1913 including debt with Sparkassen and other local authority obligations.	1850-1913: Hoffmann (1965) 1914: Schremmer (1994) 1928-1940: Deutsche Bundesbank (1976) 1950-2010: Statistisches Bundesamt (2011a)
<b>Debt service</b>	dservice	Interest service on total public debt in million Mark/Euro. 1880-1913 for Reich and Prussia, Württemberg, Baden, Bavaria, Saxony and Hamburg.	1880-1913: Accominotti et al. (2011) 1950-2010: inquired at Statistisches Bundesamt
<b>Fiscal deficit</b>	deficit	Public deficit in million Mark/Euro	1871-1913: own calculation: debtgen(t)-debtgen(t-1) 1950-2010: inquired at Statistisches Bundesamt
<b>Public expenditures</b>	exp	General ordinary and extraordinary expenditures (in million Mark/Euro). 1871-1879, 1919-1925, 1933-1944 without states. 1920-1923 only current expenditures of the Reich.	1871-1879, 1924: Mitchell (2007) 1880-1913: Accominotti et al. (2011) 1914-1919: Roesler (1967) 1920-1923: Haller (1976) 1925-1932: Ritschl (2002) 1933-1944: Länderrat des Amerikanischen Besatzungsgebiets (1949) 1950: Statistisches Bundesamt (2011b) 1951-2010: Statistisches Bundesamt (2011c)

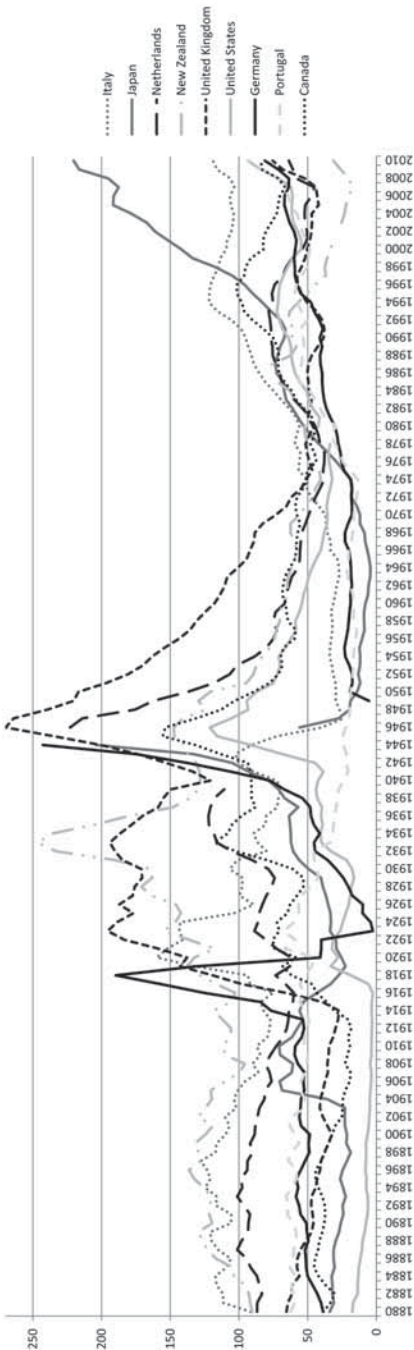
Variable	Name	Description	Source
<b>Nominal GDP</b>	nomgdp	in local currency, current prices.  Nominal GDP in local currency, current prices. 1914-1924: Real GDP (in 1913 prices) estimated by Ritschl and Spoerer (1997), GDP deflator as measured by CPI1913.	1851-1913: Burhop and Wolff (2005), kindly made available by Schularick. 1914-1924: Own calculations 1924-1939: Schularick and Taylor (2012) 1940-1944, 1946-1950: Ritschl and Spoerer (1997) 1950-2010: Statistisches Bundesamt (2012)
<b>Public revenues</b>	rev	General ordinary and extraordinary revenue (in million Mark/Euro). 1871-1879, 1919-1925, 1933-1949 without states; 1920-1923 only current expenditures.	1871-1879, 1946-1949: Mitchell (2007) 1880-1913: Accominotti et al. (2011) 1914-1919: Roesler (1967) 1920-1923: Haller (1976) 1925-1932: Ritschl (2002) 1933-1944: Länderrat des Amerikanischen Besatzungsgebiets (1949)  1950: Statistisches Bundesamt (2011b) 1951-2010: Statistisches Bundesamt (2011c)
<b>Primary deficit</b>	pdeficit	Primary deficit, i.e. budget deficit or surplus after deducting interest payments (in million Mark/Euro). constant prices of 1990, Int. GK\$	1950-2010: deficit – dservice
<b>Real GDP</b>	rgdp		1850-2008: Maddison (2010) 2009-2010: Schularick and Taylor (2012)
<b>Bond yields</b>	yield	Long term government bond yields, 5 year or 10 year maturity. 1850-1869: computed as average of two Prussian and one Bavarian bond yields	1850-1869: Homer (1963/77) 1870-2008: Schularick and Taylor (2012) 2009-2010: IMF International Financial Statistics

## A2 Comparison of estimated debt ratios (% of GDP), 1914-1924

	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924
<b>Our estimates*</b>	80.50	88.70	105.68	124.81	132.78	110.84	58.30	57.39	91.35	2.86	4.39
<b>Holtfrerich (1996)</b>	–	–	–	–	180	127	–	–	–	–	–
<b>Schularick**</b>	47.92	61.82	98.11	128.94	160.4	133.13	43.41	42.44	10.31	–	–

\*Our estimates include the Bundesländer until 1919. \*\*Data kindly made available by Schularick, referring to Niall Ferguson.

A3 Public debt as percentage of GDP in nine OECD countries, 1880-2010



Sources: For Germany refer to Figure 1, for the rest refer to Abbas et al. (2010).

## References

- Abbas, S.M., N. Belhocine, A.A. El-Ganainy, M.A. Horton (2010), A Historical Public Debt Database. IMF Working Paper WP/10/245.
- Academic Advisory Board at the Federal Ministry of Economics and Technology (2008), Zur Begrenzung der Staatsverschuldung nach Art. 115 GG und zur Aufgabe des Stabilitäts- und Wachstumsgesetzes. Gutachten 01/08. Berlin.
- Accominotti, O., M. Flandreau, R. Rezzik (2011), The Spread of Empire: Clio and the Measurement of Colonial Borrowing Costs. *The Economic History Review* 64: 385-407.
- Afonso, A. (2005), Fiscal Sustainability: The Unpleasant European Case. *FinanzArchiv* 61: 19-44.
- Afonso, A., J. T. Jalles (2011), A Longer-Run Perspective on Fiscal Stability. School of Economics and Management, Technical University Lisbon, Working Paper 17.
- Ahmed, S., J. Rogers (1995), Government Budget Deficits and Trade Deficits. Are Present Value Constraints Satisfied in Long-term Data? *Journal of Monetary Economics* 36: 351-374.
- Aschinger, G. (1985), Probleme der Staatsverschuldung. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 200: 582-601.
- Bank Deutscher Länder (1950), Geschäftsbericht für die Jahre 1948 und 1949. Frankfurt a. M.
- Blume, L., T. Döring, S. Voigt (2008), Fiskalische Effekte der Kommunalverfassungsreformen der 1990er Jahre in Deutschland. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 228: 317-344.
- Bohn, H. (1995), The Sustainability of Budget Deficits in a Stochastic Economy. *Journal of Money, Credit and Banking* 27: 257-271.
- Bohn, H. (1998), The Behaviour of U. S. Public Debt and Deficits. *Quarterly Journal of Economics* 113: 949-963.
- Bohn, H. (2008), The Sustainability of Fiscal Policy in the United States. Pp. 15-49 in: R. Neck, J. Sturm (eds.), *Sustainability of Public Debt*. Cambridge/London.
- Bordo, M.D., A. J. Schwartz (1999), Under What Circumstances, Past and Present, Have International Rescues of Countries in Financial Distress Been Successful? *Journal of International Money and Finance* 18: 683-708.
- Bravo, A., A. Silvestre (2002), Intertemporal Sustainability of Fiscal Policies: Some Tests for European Countries. *European Journal of Political Economy* 18: 517-528.
- Brümmerhoff, D., U.P. Reich (1999), Zur Problematik der Fixierung auf einzelne statistische Größen in der Wirtschafts- und Finanzpolitik – Dargestellt an der staatlichen Defizitquote. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 219: 575-590.
- Buchanan, J.M., G. Tullock (1962), *The Calculus of Consent*. Ann Arbor.
- Burhop, C., G. Wolff (2005), A Compromise of German Net National Product, 1851-1913, and Its Implications for Growth and Business Cycles. *Journal of Economic History* 65: 613-657.
- Cheung, Y.-W., K.S. Lai (1995), Lag Order and Critical Values of the Augmented Dickey-Fuller Test. *Journal of Business and Economic Statistics* 13: 277-280.
- Cohn, L.R. (1997), Fiscal Policy in Germany During the Great Depression. Pp. 261-283 in: J. Komlos, S.M. Eddie (eds.), *Selected Cliometric Studies on German Economic History*. Stuttgart.
- Deutsche Bundesbank (1976), *Deutsches Geld- und Bankwesen in Zahlen 1876-1975*. Frankfurt a.M.
- Deutsche Bundesbank (2011), *The Debt Brake in Germany – Key Aspects and Implementation*. Monthly Report, October.
- Deutsche Bundesbank (2012), *Zeitreihe UJFB99. Verbraucherpreisindex*. Available at: [http://www.bundesbank.de/cae/servlet/CSVDownload?tsId=BBK01.UJFB99&its\\_csvFormat=de&-mode=its](http://www.bundesbank.de/cae/servlet/CSVDownload?tsId=BBK01.UJFB99&its_csvFormat=de&-mode=its) (last access on 20 June 2012).
- Domar, E.D. (1944), The Burden of the Debt and International Income. *American Economic Review* 34: 789-827.

- Eheberg, K. T. (1892), Finanzverhältnisse europäischer Großstädte. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 58: 906-918.
- Elliot, G., C. Kearney (1988), The Intertemporal Government Budget Constraint and Tests for Bubbles. Research Discussion Paper 8809, Reserve Bank of Australia.
- Feld, L. P., T. Baskaran (2010), Federalism, Budget Deficits and Public Debt: On the Reform of Germany's Fiscal Constitution. *Review of Law and Economics* 6(3): 365-393.
- Feld, L. P., G. Kirchgässner (2008), On the Effectiveness of Debt Brakes: The Swiss Experience. Pp. 223-255 in: R. Neck, J.-E. Sturm (eds.), *Sustainability of Public Debt*. Cambridge.
- Feld, L. P., C. A. Schaltegger (2010), Political Stability and Fiscal Policy: Time Series Evidence for the Swiss Federal Level since 1849. *Public Choice* 144: 505-534.
- Fincke, B., A. Greiner (2009), *Public Debt and Economic Growth*. Heidelberg.
- Fincke, B., A. Greiner (2011), Do Large Industrialized Economies Pursue Sustainable Debt Policies? A Comparative Study for Japan, Germany and the United States. *Japan and the World Economy* 23: 203-213.
- Frick, O. (1890), Das Preußische Staatsschuldbuch. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 55: 337-391.
- Gerloff, W. (1913), *Die Finanz- und Zollpolitik des Deutschen Reiches*. Jena.
- German Council of Economic Experts (2005), *Die Chance nutzen – Reformen mutig voranbringen. Jahresgutachten 2005/2006*. Wiesbaden.
- German Council of Economic Experts (2011), *Verantwortung für Europa wahrnehmen. Jahresgutachten 2011/2012*. Wiesbaden.
- Gerstfeldt, P. (1883), Beiträge zur Statistik der Finanzen in Preußen. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 41: 150-161.
- Glasemann, H. G. (1993), *Deutschlands Auslandsanleihen 1924-1945*. Wiesbaden.
- Glynn, J., N. Perera, R. Verma (2007), Unit Root Tests and Structural Breaks: A Survey with Applications. *Revista de Métodos Cuantitativos para la Economía y la Empresa (Journal of Quantitative Methods for Economics and Business Administration)* 3: 63-79.
- Goldschmidt, N., E. Köhler (2008), Die Ursprünge der Sozialen Marktwirtschaft und die Währungsreform von 1948. Pp. 82-98 in: *Hessisches Wirtschaftsarchiv (ed.)*, *Beiträge zur hessischen Wirtschaftsgeschichte* 4. Darmstadt.
- Graner, H. (1915), Die französischen Finanzen in Kriegszeit. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 104: 337-343.
- Greiner, A., G. Kauermann (2007), Sustainability of US Public Debt: Estimating Smoothing Spline Regressions. *Economic Modelling* 24: 350-364.
- Greiner, A., G. Kauermann (2008), Evidence for Germany and Italy Using Penalized Spline Smoothing. *Economic Modelling* 25: 1144-1154.
- Greiner A., U. Köller, W. Semmler (2006), Testing the Sustainability of German Fiscal Policy: Evidence for the Period 1960-2003. *Empirica* 33: 127-140.
- Grilli, V. (1988), *Seigniorage in Europe*. National Bureau of Economic Research (NBER), Working Paper No. 2778.
- Hakkio, G., M. Rush (1991), Is the Budget Deficit Too Large? *Economic Inquiry* 29: 429-445.
- Haller, H. (1976), Die Rolle der Staatsfinanzen für den Inflationsprozeß. Pp. 115-155 in: *Deutsche Bundesbank (ed.)*, *Währung und Wirtschaft in Deutschland 1876-1975*. Frankfurt a. M.
- Hamilton, J. D. (1994), *Times Series Analysis*. Princeton.
- Hamilton, J. D., M. Flavin (1986), On the Limitations of Government Borrowing: A Framework for Empirical Testing. *American Economic Review* 76: 808-819.
- Hansmeyer, K.-H., R. Caesar (1976), *Kriegswirtschaft und Inflation (1936-1948)*. Pp. 367-429 in: *Deutsche Bundesbank (ed.)*, *Währung und Wirtschaft in Deutschland 1876-1975*. Frankfurt a. M.
- Haug, A. (1995), Has Federal Budget Deficit Policy Changed in Recent Years? *Economic Inquiry* 33: 104-118.
- Heckel, M. von (1900), Beiträge zur vergleichenden Finanzstatistik europäischer Großstaaten im Jahre 1898. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 74: 34-61.



- Hefeker, C. (2001), The Agony of Central Power: Fiscal Federalism in the German Reich. *European Review of Economic History* 5: 119-142.
- Heinemann, E. (1914), Der Kursstand der deutschen Staatsanleihen. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 102: 219-225.
- Helander, S. (1915), Das Inflationsproblem im Kriege. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 105: 246-252.
- Helfferrich, K. T. (1916), Stenographische Berichte der Verhandlungen des Reichstags Vol. 306. Available at: [http://www.reichstagsprotokolle.de/Blatt\\_k13\\_bsb00003402\\_00000.html](http://www.reichstagsprotokolle.de/Blatt_k13_bsb00003402_00000.html) (last access on 20 June 2012).
- Heyn, O. (1919), Zum Problem der Geldentwertung. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 113: 336-343.
- Hildebrand, B. (1863), Das Papiergeld der Vereinigten Staaten Nordamerikas während ihres Unabhängigkeitskrieges von 1775-1781. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 1: 392-396.
- Hoffmann, W. G. (1965), Das Wachstum der deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts. Heidelberg.
- Holtfrerich, C.-L. (1996), Bewältigung der deutschen Staatsbankrotte 1918 und 1945. Pp. 25-57 in: E. Kantzenbach (ed.), Staatsüberschuldung: Referate gehalten auf dem Symposium der Joachim-Jungius-Gesellschaft der Wissenschaften Hamburg am 9. und 10. Februar 1996. Göttingen.
- Homer, S. (1963/77), A History of Interest Rates (2<sup>nd</sup> ed.). New Jersey.
- Hübner, O. (1891), Der finanzielle Zusammenbruch Argentiniens. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 57: 239-254.
- Inhülsen, C. H. P. (1919), Englische Urteile über die deutsche Finanzlage. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 112: 570-578.
- Jaeger, F. (1982), Die Unterstellung der Staatsverschuldung unter ökonomisch fundierte Regelbindungen. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 197: 413-435.
- Jansen, H. (2000), Nationalökonomie und Nationalismus (Beiträge zur Geschichte der deutschsprachigen Ökonomie Bd. 10). Marburg.
- Kaufmann, R. (1887), Finanzstatistisches zu den Schulden der europäischen Großmächte. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 49: 97-150.
- Kirchgässner, G., S. Prohl (2008), Sustainability of Swiss Fiscal Policy. *Swiss Journal of Economics and Statistics* 144: 57-83.
- Kirchgässner, G., J. Wolters (2006), Einführung in die moderne Zeitreihenanalyse. München.
- Kitterer, W. (2007), Nachhaltige Finanz- und Investitionspolitik der Bundesländer. *Perspektiven der Wirtschaftspolitik* 8: 53-83.
- Köppe, H. (1916a), Die deutschen Kriegsanleihen. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 106: 321-348.
- Köppe, H. (1916b), Die Kriegsanleihen Englands. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 107: 289-338.
- Köppe, H. (1916c), Die Kriegsanleihen Frankreichs und die englisch-französische Anleihe in den Vereinigten Staaten. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 106: 721-775.
- Köppe, H. (1916d), Die Kriegsanleihen Österreich-Ungarns. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 106: 449-491.
- Köppe, H. (1916e), Die Kriegsanleihen Russlands und Italiens. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 107: 577-618.
- Köppe, H. (1918a), Die Kriegsanleihen der Ententemächte. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 111: 1-46.
- Köppe, H. (1918b), Die Kriegsanleihen Deutschlands und Österreich-Ungarn. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 110: 167-212.
- Köppe, H. (1919), Die Kriegsanleihen der europäischen Großmächte. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 113: 385-426.

- Krawehl, O.E. (1986), Zweihundert Bände "Jahrbücher für Nationalökonomie und Statistik". Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 202: 1-23.
- Kremers, J. (1988), The Long-Run Limits of U.S. Federal Debt. Economics Letters 28: 259-262.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, Y. Shin (1992), Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root. Journal of Econometrics 54: 159-178.
- Lachmann, W. (1984), Kann die staatliche Kreditaufnahme im Ausland ein "Crowding-out" verhindern. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 199: 522-536.
- Lang, G., P. Welzel (1992), Budgetdefizite, Wahlzyklen und Geldpolitik: Empirische Ergebnisse für die Bundesrepublik Deutschland 1962-1989. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 210: 72-85.
- Lampert, H., A. Oberhauser, A. Wagner (1986), Die "Jahrbücher für Nationalökonomie und Statistik" in der Zeit von 1949 bis 1985. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 202: 38-40.
- Länderrat des Amerikanischen Besatzungsgebiets (1949), Statistisches Handbuch von Deutschland 1928-1944. München.
- Larch, M. (1993), Measuring the Effectiveness of Budget Consolidation Efforts. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 211: 503-517.
- Leeper, E. (1991), Equilibria under 'Active' and 'Passive' Monetary Policies. Journal of Monetary Economics 27: 129-147.
- Liu, P., E. Tanner (1995), Intertemporal Solvency and Breaks in the U.S. Deficit Process: A Maximum-likelihood Cointegration Approach. Applied Economics Letters 2: 231-235.
- MacDonald, R. (1992), Some Tests of the Government's Intertemporal Budget Constraint Using U.S. Data. Applied Economics 24: 1287-1292.
- Maddison, A. (2010), Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD. Available at: [http://www.ggdc.net/maddison/Historical\\_Statistics/horizontal-file\\_02-2010.xls](http://www.ggdc.net/maddison/Historical_Statistics/horizontal-file_02-2010.xls) (last access on 29 November 2012)
- Mandello, K. (1897), Ungarns Finanzlage seit dem 1867er Ausgleiche. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 68: 93-102.
- Mendoza, E., J.D. Ostry (2008), International Evidence on Fiscal Solvency: Is Fiscal Policy "Responsible"? Journal of Monetary Economics 55: 1081-1093.
- Menger, C. (1892), Valutaregulierung in Österreich-Ungarn. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 58: 496-515.
- Meyer, D. (2011), Kosten des Europäischen Finanzstabilisierungssmechanismus (EFSM) aus deutscher Sicht. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 231: 288-303.
- Mitchell, B.R. (2007), International Historical Statistics: Europe 1750-2005 (6<sup>th</sup> ed.). New York.
- Mückl, W.J. (1985), Langfristige Grenzen der öffentlichen Kreditaufnahme. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 200: 565-581
- Neck, R., J.E. Sturm (2008), Sustainability of Public Debt: Introduction and Overview. Pp. 1-13 in: R. Neck, J.E. Sturm (eds.), Sustainability of Public Debt. Cambridge.
- Neugebauer, W. (2000), Handbuch der preussischen Geschichte, Band 3 Vom Kaiserreich zum 20. Jahrhundert und große Themen der Geschichte Preußens. Berlin.
- Neumark, F. (1976), Die Finanzpolitik in der Zeit vor dem I. Weltkrieg. Pp. 57-111 in Deutsche Bundesbank (ed.), Währung und Wirtschaft in Deutschland 1876-1975. Frankfurt a.M.
- Oberhauser, A. (1985), Das Schuldenparadox. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 200: 333-348.
- Obstfeld, M., A.M. Taylor (2003), Sovereign Risk, Credibility and the Gold Standard: 1879-1913 versus 1925-1931. Economic Journal 113: 241-275.
- Ott, A.E., H. Strecker (1986), Die "Jahrbücher für Nationalökonomie und Statistik" in der Zeit von 1929 bis 1944. Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 202: 33-37.

- Paasche, H. (1878), Die neueste Entwicklung der Banknoten- und Papiergeld-Zirkulation in den hauptsächlichsten Kulturländern der Gegenwart. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 30: 332-364.
- Payne, J. (1997), International Evidence on the Sustainability of Budget Deficits. *Applied Economics Letters* 12: 775-779.
- Peterson-Studnitz, A. (1888), Das dänische Staatsbudget. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 50: 456-476.
- Petzina, D. (1986), Kommunale Finanzen und Handlungsspielräume in der Weltwirtschaftskrise: Das Beispiel der Stadt Bochum. Pp. 231-254 in: H.-J. Teuteberg (ed.), *Stadtwachstum, Industrialisierung, sozialer Wandel*. Berlin.
- Pfleiderer, O. (1976), Die Reichsbank in der Zeit der großen Inflation, die Stabilisierung der Mark und die Aufwertung von Kapitalforderungen. Pp. 157-202 in: Deutsche Bundesbank (ed.), *Währung und Wirtschaft in Deutschland 1876-1975*. Frankfurt a.M.
- Plumpe, W. (2010), *Wirtschaftskrise: Geschichte und Gegenwart*. München.
- Polito, V., M. Wickens (2011), Assessing the Fiscal Stance in the European Union and the United States, 1970-2011. *Economic Policy* 26: 599-647.
- Quintos, C. (1995), Sustainability of the Deficit Process with Structural Shifts. *Journal of Business & Economic Statistics* 13: 409-417.
- Raffalovich, A. (1885), Die Finanzen Irlands. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 45: 476-477.
- Raffalovich, A. (1888), Englands Finanzen. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 51: 320-329.
- Regensburger, E.H. (1921), Frankreichs Finanzen seit 1914. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 117: 151-154.
- Reinhart, C.M., K.S. Rogoff (2009), *This Time Is Different: Eight Centuries of Financial Folly*. Princeton.
- Ricardo, D. (1817/1932), *Principles of Political Economy and Taxation*. London.
- Richter, R. (1998), *Deutsche Geldpolitik 1948-1998*. Tübingen.
- Ritschl, A. (2002), Deutschlands Krise und Konjunktur. Binnenkonjunktur, Auslandsverschuldung und Reparationsproblem zwischen Dawes-Plan und Transfersperre 1924-1934, Berlin.
- Ritschl, A., M. Spoerer (1997), Das Bruttosozialprodukt in Deutschland nach den amtlichen Volkseinkommens- und Sozialstatistiken 1901-1995. Pp. 28-54 in Lothar Baar et al. (ed.), *Jahrbuch für Wirtschaftsgeschichte 1997-2*. Berlin.
- Rittershausen, H. (1943), Die Zukunft der öffentlichen Schuld. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 158: 22-38.
- Roesler, K. (1967), *Die Finanzpolitik des Deutschen Reichs im Ersten Weltkrieg*. Duncker und Humblot, Berlin.
- Sargent, T.J., N. Wallace (1981), Some Unpleasant Monetarist Arithmetic. *Federal Reserve Bank of Minneapolis Quarterly Review* 5: 1-17.
- Schacht, H. (1945), Interrogation of Schacht 16 October 1945. Donovan Nuremberg Trials Collection XVI. Available at: [http://library2.lawschool.cornell.edu/donovan/pdf/Batch\\_7/Vol\\_XVI\\_48\\_04\\_01.pdf](http://library2.lawschool.cornell.edu/donovan/pdf/Batch_7/Vol_XVI_48_04_01.pdf) (last access on 20 June 2012).
- Schaltegger, C.A., L.P. Feld (2009), Do Large Cabinets Favor Large Governments?, Evidence on the Fiscal Commons Problem for Swiss Cantons. *Journal of Public Economics* 93: 35-47.
- Scherf, W. (1989), Zur Abgrenzung und finanzpolitischen Bedeutung verschiedener Arten der Staatsverschuldung. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 206: 136-151.
- Schiller, K. (1936), *Arbeitsbeschaffung und Finanzordnung in Deutschland. Zum Wirtschaftlichen Schicksal Europas: Arbeiten zur deutschen Problematik*. Berlin.
- Schremmer, E. (1994), *Steuern und Staatsfinanzen während der Industrialisierung Europas*. Berlin.
- Schularick, M., A.M. Taylor (2012), Credit Booms Gone Bust: Monetary Policy, Leverage Cycles and Financial Crises, 1870–2008. *American Economic Review* 102: 1029-1061.

- Schulz, A., G. B. Wolff (2009), The German Sub-national Government Bond Market: Structure Determinants of Yield Spreads and Berlin's Forgone Bail-out. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 229: 61-83.
- Schumpeter, J. A. (1952), Aufsätze zur Ökonomischen Theorie. Tübingen.
- Seidler, E. (1894), Die Schwankungen des Geldwertes (der Kaufkraft des Geldes) und die juristische Lehre von dem Inhalte der Geldschulden. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 62: 685-706.
- Sims, C. A. (1994), A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy. *Economic Theory* 4: 381-399.
- Smith, G., S. Zin (1991), Persistent Deficits and the Market Value of Government Debt. *Journal of Applied Econometrics* 6: 31-44.
- Spahn, P. (2012), David Hume und die Target Salden. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 232: 460-481.
- Statistisches Bundesamt (2011a), Schulden der öffentlichen Haushalte: Finanzen und Steuern, 2010 Fachserie 14 Reihe 5. Wiesbaden.
- Statistisches Bundesamt (2011b), Finanzen und Steuern – Rechnungsergebnisse des öffentlichen Gesamthaushalts 2008, Fachserie 14 Reihe 3.1. Wiesbaden.
- Statistisches Bundesamt (2011c), Entwicklung der öffentlichen Finanzen in Deutschland. Available at: <https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/SteuernFinanzen/Irfin01.html> (last access on 20 June 2012).
- Statistisches Bundesamt (2012), Volkswirtschaftliche Gesamtrechnungen – Bruttoinlandsprodukt, Bruttonationaleinkommen, Volkseinkommen – Lange Reihe ab 1950, 2011. Available at: [https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VGR/Inlandsprodukt/Tabellen/Volkseinkommen1950.xls?\\_\\_blob=publicationFile](https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VGR/Inlandsprodukt/Tabellen/Volkseinkommen1950.xls?__blob=publicationFile) (last access on 20 June 2012).
- Stegarescu, D. (2005), Centralization Tendencies in the Public Sector in Germany. Zentrum für Europäische Wirtschaftsforschung (ZEW) Discussion Paper 46.
- Tanner, E., P. Liu (1994), Is the Budget Deficit Too Large? Some Further Evidence. *Economic Inquiry* 32: 511-518.
- Terhalle, F. (1923), Die fiskalische Inflationsbilanz. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 121: 97-122.
- Terhalle, F. (1952), Geschichte der deutschen öffentlichen Finanzwirtschaft vom Beginn des 19. Jahrhunderts bis zum Schlusse des zweiten Weltkrieges. Pp. 273-326 in: W. Gerloff, F. Neumark (eds.), *Handbuch der Finanzwissenschaft* Vol. 1 (2<sup>nd</sup> ed.). Tübingen.
- Trehan, B., C. E. Walsh (1988), Common Trends, Intertemporal Budget Balance, and Revenue Smoothing. *Journal of Economic Dynamics and Control* 12: 425-444.
- Trehan, B., C. E. Walsh (1991), Testing Intertemporal Budget Constraints: Theory and Applications to the US Federal Budget and Current Account Deficit. *Journal of Money, Credit and Banking* 23: 206-223.
- Vogler, R. U., J. Baumberger, H. R. Reginbogin, J. Spiller (eds.) (2012), *Financial Markets of Neutral Countries in World War II*. Association for Financial History Switzerland und Principality of Liechtenstein. Zürich.
- Warschauer, O. (1883), Die fundierte Staatsschuld Preußens. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 41: 150-161.
- Wenzel, H. D., M. Wrede (2000), Public Debt in a Federation – Growth and Stability. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 220: 95-107.
- Wilcox, D. (1989), The Sustainability of Government Deficits: Implications of the Present-Value Borrowing Constraint. *Journal of Money, Credit, and Banking* 21: 291-306.
- Wittmann, W. (1966), Versuch einer volkswirtschaftlichen Theorie des öffentlichen Kredits. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 179: 273-285.
- Woodford, M. (1995), Price Level Determinacy Without Control of a Monetary Aggregate. *Carnegie-Rochester Conference Series on Public Policy* 43: 1-46.

Zamek, W. von (1988), Konjunkturbedingte Defizite und materieller Haushaltsausgleich. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics)* 204: 197-207.

Zivot, E., K. Andrews (1992), Further Evidence On The Great Crash, The Oil Price Shock, and The Unit Root Hypothesis. *Journal of Business and Economic Statistics* 10: 251-270.

Dipl.-Vw. Heiko T. Burret, Walter Eucken Institut e.V., Goethestrasse 10, 79100 Freiburg, Germany.  
burret@eucken.de

Prof. Dr. Lars P. Feld, Walter Eucken Institut e.V. and Albert-Ludwigs-University Freiburg, Goethestrasse 10, 79100 Freiburg, Germany.  
feld@eucken.de

Dipl.-Vw. Ekkehard A. Köhler, Walter Eucken Institut e.V., Goethestrasse 10, 79100 Freiburg, Germany.  
koehler@eucken.de

# Recurrent Price Index Problems and Some Early German Papers on Index Numbers

Notes on Laspeyres, Paasche, Drobisch, and Lehr

Peter von der Lippe\*

Universität Duisburg-Essen

JEL B16; C43; E31; E52

Index numbers; Laspeyres; Paasche; Drobisch; chain indices; unit values; inflation; geometric mean; time reversal test; index theory; pure price comparison.

Received: 14.02.1012

Revision received: 03.07.2012

Accepted: 18.07.2012

---

## Summary

With Laspeyres, Paasche and other authors such as Drobisch and Lehr, Germany made quite a promising start in index theory in the last decades of the 19<sup>th</sup> century. However, it soon lost ground after this period, which is described in this paper. The focus is not on biographies but on controversies where these persons acted as opponents and developed the views for which they are well known. The issues selected are primarily those which are still interesting and controversial today, e.g. the merits and demerits of certain index formulas, the definition and updating of weights, “pure” price comparison vs. chain indices etc. However, in order to aid a better understanding of how Laspeyres etc. arrived at their index formulas and views about the purposes index numbers should serve, some attention is also given to the typical disputes and prejudices of that time (e.g. regarding inflation under the regime of a gold currency).

## 1 Introduction

It appears attractive to take the occasion of the anniversary of this journal to review some papers that appeared in this journal during the last three or four decades of the 19<sup>th</sup> century, and which paved the way to modern price index numbers. It is particularly appealing as such a plan leads to names that were later to become famous around the world<sup>1</sup>,

---

\* I would like to thank two anonymous referees for their valuable comments and Mathew Harrison who checked and improved the language.

<sup>1</sup> It was anything but certain that names like Laspeyres and Paasche were to become so famous. Given the German literature on index numbers in the first half of the 20<sup>th</sup> century, it does not seem unlikely that these names would have fallen into oblivion had there not been some English speaking authors, in particular the Americans Walsh and Fisher, who constantly referred to their work (and names). Interestingly even L.v. Bortkiewicz in 1932 did not introduce the names Laspeyres and Paasche (unlike his 1927 article) as authors of his formulas 1 and 2, while he consistently mentioned the names in connection with the other eight formulas he discussed in his paper. It is perhaps also not by

for example Etienne Laspeyres (1834 – 1913) and Herrmann Paasche (1851 – 1925), as well as two probably far less well-known authors, namely Moritz Wilhelm Drobisch (1802 – 1896) and Julius Lehr (1845 – 1894).<sup>2</sup> The life and academic work of these celebrities in index theory, each of them taken in isolation, has already found a number of competent and detailed descriptions and recognitions. So we have for example biographies and appreciations of the academic achievements of Laspeyres (Rinne 1981; Diewert 1987; Roberts 2000), Drobisch (v. Auer 2010), and also much is known about Paasche, who for example once was a vice-president of the German *Reichstag* (parliament).

Our intention is therefore already at the outset quite different, namely to relate these persons to some controversial issues in index numbers in which they acted as opponents. We selected such issues which already occupied a lot of people at that time and which still continue to do so today. After a review of many articles (in this and other journals) and also books on indices of that time (roughly 1860 to 1920), it was clear that there should be enough material to carry out this plan, and that we can indeed see a number of such perennial controversies fought out on the long road to modern index number theory. Such controversies and the persons involved in them are

- the use of the geometric or the arithmetic mean<sup>3</sup> in a problem now known as “low-level aggregation” (or compilation of “elementary” indices); historically this is the “Jevons vs. Laspeyres” case and will be dealt with in Sec. 2;
- the introduction of weights (to account for the relative importance of goods), for example physical weights multiplied with prices to form so-called “unit values” and based on them the “unit value index”, which is the case “Drobisch vs. Laspeyres” (see Sec. 3), a case which also gave rise to a claim of authorship (on the part of Drobisch; see Sec. 4);
- the choice between a single (using quantity weights  $q_0$  or  $q_t$ )<sup>4</sup> and a double weighting system (using both the quantities  $q_0$  and  $q_t$  as weights); see Sec. 5;
- disputes as to whether an index should be compiled as an average of price relatives (price ratios) or rather represent a ratio of average prices (Sec. 6); and finally
- as a sort of logical continuation of the problem that a constant updating of the  $q_t$ ’s appears desirable, we find the idea of a “chain index”, in which the periods 0 (base period) and  $t$  (current period) are not compared directly with an index  $P_{0t}$  affected by prices and quantities of 0 and  $t$  only, but via “chaining” (multiplying)  $P_{01}P_{12}...P_{t-1,t}$ ; see Sec. 7.<sup>5</sup>

---

coincidence that I had problems with getting some German articles of Laspeyres and Paasche here in Germany. So I owe for example copies of Laspeyres 1875 and Paasche 1878 to Othmar Winkler (Georgetown University, Washington D.C.), and Laspeyres 1883 to Hellen Roberts (University of Illinois, Chicago). I also should express my gratitude to Erwin Diewert. I learned a lot from the historical remarks he made in his e-mails.

<sup>2</sup> The fourth memorable author (Lehr), to which Sec. 7 below will be devoted, is also repeatedly represented in this journal (in particular with papers about the then revolutionary [“Austrian school”] concepts of “marginal utility” etc.) He presented his considerations about index numbers, however, in a monograph (Lehr 1885). In addition to the fact that renowned authors published their papers on index numbers in this journal, it is noteworthy that J. Conrad, a former editor of the journal, did a lot to promote index number research (see footnotes 60, 62).

<sup>3</sup> This dispute is also summarized in Walsh (1901: 220).

<sup>4</sup> As is well known, the formulas of Laspeyres and Paasche differ in this respect.

<sup>5</sup> Nothing indicates a difference or even controversy between Laspeyres and Paasche; however, it is justified to speak of “Lehr vs. Paasche”, because Lehr, as an early proponent of chain indices, quite



Hence our focus is on recurrent and more or less still relevant index problems rather than biographies. We therefore exclude problems that used to trigger some passionate discussions and had a considerable impact on the study of index numbers at the time under consideration here, but have since lost much or all of their relevance. Many of the early debates can only be understood in the light of the (gold) currency problems of that time. For example, gold currency versus bimetallism proved to be a catalyst for Jevons' interest in index numbers. The relationship between inflation, money and prices was not yet well understood. Laspeyres ran into great difficulties with a distinction between rising prices of commodities ("*Waarenbewerthung*", revaluation of goods) on the one hand and devaluation of money ("*Geldentwerthung*") on the other, because both phenomena were observationally equivalent.<sup>6</sup> It is also noticeable that it was not yet generally accepted that inflation would call for a study of prices (as was Laspeyres' view) rather than for statistics of the "*Zunahme des Metallvorraths*" (increased availability of precious metals).<sup>7</sup> The problem people found at least as intriguing as index numbers was for example whether an increase in prices was primarily caused by gold discoveries, or by higher aspiration levels and consumption standards of the urban working class.<sup>8</sup>

Moreover, not only *what* was discussed in those days may appear strange, but also *how* it was discussed. In order to do justice to authors of that time it should be borne in mind that many now well-established methods to assess index formulas had not yet been developed, or at least were not yet familiar. To assess formulas in term of "axioms" (or "tests") was still uncommon, and *a fortiori* to interpret formulas in terms of utility maximizing by making substitutions in response to changes of relative prices. This was at best alluded to in rather vague verbal statements, but definitely not yet worked out mathematically. Of course mathematics for economists was in general only in its infancy.<sup>9</sup> It was not uncommon to content oneself with numerical examples, elaborated in detail over many pages. Also, lengthy deliberations about the correct definitions of certain concepts and logical conditions required for certain conclusions were common practice. It should also be noticed that academic communication across borders and detailed knowledge of foreign publications were but exceptions. It is well known and reported that e.g. Jevons and Laspeyres were in close contact. Finally, it should be added that in

---

vehemently criticized Paasche for his vague and slightly inconclusive position concerning the weights. In the view of Lehr, Paasche saw (in Paasche 1874) a need for continuously varying weights  $q_t$  but inconsequentially he did not go so far as to account also for all the intermediate periods  $q_{t-1}$ ,  $q_{t-2}$ , ... as vehemently required by Lehr (1885: 44).

<sup>6</sup> As Laspeyres (1864: 82) noted, they are "wearing the same outer garment" (*tragen "dasselbe äussere Gewand"*). Also in the title of Drobisch (1871a), we see the distinction between change of prices (*Veränderungen der Waarenpreise*) and change of the exchange value of money (*Veränderungen des Geldwerths*). As to the "exchange value", Walsh 1901 seems to have succeeded in maximizing confusion with a host of hair-splitting terminological distinctions.

<sup>7</sup> The title of Paasche 1878 might be mistaken as a book providing statistics of prices or empirical research with his formula, presented just four years ago, but in fact it is dealing almost exclusively with statistics about stocks and flows of gold and silver. There is no mention given to a "price level" let alone a price index.

<sup>8</sup> In this point opinions of Laspeyres and Paasche seemed to differ slightly; cp. Laspeyres (1883: 798).

<sup>9</sup> A comparison of some writings of German authors to for example Edgeworth's papers on index theory at that time clearly showed that the Germans began already soon after the time of Laspeyres and Paasche to lag behind and they did so even more pronouncedly some decades later. We could find a great number of articles on index numbers in English and American journals in the first couple of decades of the 20<sup>th</sup> century, but not many in German journals. Instead there was much in the form of futile sophistic philosophy about money and prices.

what follows the focus is on the formula, as is common in index theory, while many non-formula problems such as the selection of commodities for the price index, the organization of regular price quotations and family budget surveys, or how to make sure that the quality of goods at different points in time is comparable are not discussed here. Laspeyres in particular quite often referred to such aspects of indices.

Finally, it is noteworthy and should be borne in mind that in the time period dealt with in this paper there was hardly any systematic statistical data gathering, and official statistical agencies (on a regional or national level) were not yet established, or at least were quite rare. What is now provided by official statistics was simply unavailable or had to be compiled laboriously on a private initiative. Economists like Laspeyres etc. spent an enormous amount of time and effort on the compilation of statistical figures at the expense of mathematical or conceptual work in statistics.

## 2 Laspeyres vs. Jevons: arithmetic vs. geometric mean of price relatives

Laspeyres dealt mainly with three problems in two famous contributions to this journal. In his paper of 1864 he discussed:

1. the relationship between the quantity and value of gold on the one hand and the level of prices (of commodities) on the other;
2. whether to use the geometric mean of price relatives as suggested by Jevons, or rather to keep to the arithmetic mean as preferred by Laspeyres and most of the economists of his day; while
3. the choice of suitable quantity weights for prices, intended to indicate their relative importance, may be viewed as a third problem, and is dealt with only in Laspeyres 1871, where he also presented his well-known price index formula. The formula grew out of a controversy with Drobisch to which we will return in Section 3.2.

While the first problem is no longer relevant, the second is still an issue now, and is referred to as “low-level aggregation” of price quotations.<sup>10</sup>

In Laspeyres’ day, the formula generally in use was the arithmetic mean of price relatives (price ratios), now known as the index formula of Carli:<sup>11</sup>

$$P_{0t}^C = \frac{1}{n} \sum_{i=1}^n \frac{P_{it}}{P_{i0}}. \quad (1)$$

<sup>10</sup> Such (unweighted) “elementary indices” serve as building blocks for a second aggregation (this time inclusive of weights) when an index is compiled in two stages, which is common practice in official statistics.

<sup>11</sup> This index is also known as the “Sauerbeck index” (see also Balk (2008: 9). Laspeyres and some other contemporary authors made extensive use of this formula (and also of Sauerbeck’s price statistics for the British foreign trade; while Sauerbeck provided data for England, Soetbeer did the same for Germany). It was only in the 20<sup>th</sup> century (more precisely: owing to Walsh 1901) that it became generally known that the formula originated from Gian Rinaldo Carli (1720 – 1795). Walsh also discovered that Dutot was the author of the index  $P^D$  (see eq. 3). References to the books of Carli, Dutot and many other authors of the early history of index numbers can be found for example in Diewert 1993.

Jevons by contrast suggested the geometric mean (at that time unusual and unfamiliar):

$$P_{0t}^J = \left( \prod_{i=1}^n \frac{p_{it}}{p_{i0}} \right)^{1/n}. \quad (2)$$

As mentioned above, the problem of which mean to use when an unweighted (using prices only) or “elementary” price index is to be compiled is still relevant. However, we are nowadays in a better position in that we are used to discussing such problems with reference to “tests” (or “axioms”), like for example the time reversal test,<sup>12</sup> or other axioms which were basically unknown at the time of Jevons and Laspeyres. The way in which arguments were developed and advanced, e.g. by numerical examples, was quite different in those days.

For his contemporaries it was widely accepted (and criticized) that Jevons did not give many reasons for his choice of the geometric mean, and when he faced adverse opinions he did not make the issue any clearer by adding yet another candidate, namely the harmonic mean.<sup>13</sup>

Literature on the issue of geometric vs. arithmetic means abounds, and already did so at the time under consideration here, and space restrictions require us to limit ourselves to only those arguments that were expressly advanced by Laspeyres and Jevons in their controversy.

Laspeyres frankly admitted that he was impressed by Jevons’s example<sup>14</sup> according to which a rise in the cocoa price of 100 % ( $p_{1t}/p_{10} = 2$ ) will be neutralized by a drop of the price for cloves by 50 % ( $p_{2t}/p_{20} = 0.5$ ), so that  $P^J = 1$  should be correct by contrast to  $P^C = 1.25$ . In his rebuttal, Laspeyres (arguing again in terms of a numerical example) made the assumption that if we initially have one “Centner”, i.e. one hundredweight (1 cwt.) cocoa for  $p_{10} = 100$  Thaler (Tlr), and also 1 cwt. clove for  $p_{20} = 100$  Tlr, the change in prices means that we later (with new prices  $p_{1t} = 200$  and  $p_{2t} = 50$ ) will get only  $q_{1t} = 0.75$  cwt. (instead of 1 cwt.) cocoa and  $q_{2t} = q_{20} = 1$  cwt. cloves. The  $1/4$  cwt. cocoa less is worth 50 Tlr (or 25 % of the expenditure  $\Sigma p_0 q_0 = \Sigma p_t q_t = 200$ ). So prices in actual fact rose by 25 % instead of the 0 % according to Jevons, and  $P^C = 1.25$  is correct.<sup>15</sup>

<sup>12</sup> This test requires that interchanging 0 and t in an index should result in  $P_{t0} = (P_{0t})^{-1}$ . It is in no small measure due to this test that  $P^C$  came out on the losing end of the rivalry with  $P^J$ . But time reversibility was not yet an issue in the controversy Laspeyres vs. Jevons. The concept of this test is due to Pierson (1896: 128). More formal statements of the test were made by Walsh (1901: 324) and Fisher (1922: 64).

<sup>13</sup> Jevons (1865: 295). For the above criticism concerning Jevons see Padan (1900: 173, 181), and Cooley (1893: 287). Edgeworth even spoke in a footnote of Jevons’ “obscure dicta as to the grounds for preferring the geometric mean”, cf. Edgeworth (1918: 189). It is clear that, given our present state of index theory, we are now able to say more in favour of Jevons’ position.

<sup>14</sup> Es “*hat etwas Bestechendes und wollte auch mich anfangs verführen, allein eine genauere Betrachtung hat mir gezeigt, dass gerade das arithmetische Mittel das richtige ist*” Laspeyres (1864: 96), (“The example appealed to me and at first also almost seduced me as well; only a closer inspection revealed to me that only the arithmetic mean is the correct one”).

<sup>15</sup> Laspeyres also modified the example to quantities  $q_{1t} = 1$  cwt. cocoa and  $q_{2t} = 0$  cwt. cloves (the general equation is of course  $4 = 4q_{1t} + q_{2t}$  or simply  $200 = \Sigma p_t q_t = p_{1t} q_{1t} + p_{2t} q_{2t}$  where  $p_{1t} = 200$  and  $p_{2t} = 50$ ).

As to the reasoning of Jevons, it is notoriously disregarded that + 100 % (in good 1) is canceled by -50 % only when a *subsequent* decline refers to  $p_{1t} = 200$ , that is to the good of which the price has risen, rather than to the initial price of a second good  $p_{20} = 100$ . The correct average *over time* (i.e. over a number of adjacent intervals) of a single good  $i$ , that is, over  $p_{i1}/p_{i0}, p_{i2}/p_{i1}, p_{i3}/p_{i2}, \dots$  (with a constantly changing base of the relatives), is a geometric mean. However, this has to be kept distinct from an average *over different goods* referring to one interval only (that is, over  $p_{11}/p_{10}, p_{21}/p_{20}, \dots, p_{n1}/p_{n0}$ ), in which case it is far from clear that the geometric mean is appropriate.<sup>16</sup>

However, Laspeyres' consideration is also liable to at least two criticisms:

1. What Laspeyres actually defended was not  $P^C$  (the index formula he and most of his contemporaries used) but rather the price index of Dutot

$$p_{0t}^D = \frac{\sum p_{it}}{\sum p_{i0}} = \frac{\sum p_{it}/n}{\sum p_{i0}/n} = \frac{\bar{p}_t}{\bar{p}_0} \quad (3)$$

which coincides with Laspeyres' index

$$p_{0t}^L = \frac{\sum p_{it}q_{i0}}{\sum p_{i0}q_{i0}} \quad (4)$$

when all quantities in 0 are equal, as in Laspeyres' example ( $q_{10} = q_{20} = 1$  cwt.). That is why he could argue against Jevons in terms of *expenditures*, which always was (and continues to be) a rather popular paradigm<sup>17</sup> of conceiving inflation: a price "level" is rising to the extent that we get less for the same amount of money.

2. Already Pierson 1896 noticed that Laspeyres' example would not have worked so well had he started with unequal prices, for example  $p_{10} = 50$  and  $p_{20} = 100$ , because then  $p_{0t}^D = p_{0t}^L = 1$ . With  $p_{10} = 50$  and  $p_{20} = 200$ , he even got  $p_{0t}^D = 0,8 < p_{0t}^L = 1$ , in which case his argument against Jevons had broken down completely. However, Laspeyres apparently had not noticed this because he failed to see that with equal relative price weights  $P^C$  boils down to  $P^D$

$$p_{0t}^C = \sum \frac{p_{it}}{p_{i0}} \frac{1}{n} = \sum \frac{p_{it}}{p_{i0}} \frac{p_{i0}}{\sum p_{i0}} = \frac{\sum p_{it}}{\sum p_{i0}} = p_{0t}^D. \quad (5)$$

In his controversy with Jevons, Laspeyres unfortunately only considered *unweighted* arithmetic means. Had he introduced weights  $a$  and  $1-a$  in his example he should have seen why his result differs from Jevons'. It is easy to find  $a$  for which the weighted arithmetic mean of two price relatives  $r_1$  and  $r_2$ , i.e.  $P^A = ar_1 + (1-a)r_2$ , equals the index  $P^J$  of Jevons  $(r_1r_2)^{1/2}$ . Assuming  $r_1 > r_2$  (and thus  $r_2 < \sqrt{r_1r_2} < r_1$ ), we get  $a = (\sqrt{r_1r_2} - r_2)/(r_1 - r_2)$ . With  $a = r_1 = 2$  and  $r_2 = a^{-1} = 1/2$  (Jevons' example), we have  $a = 1/3$ . The greater  $a$  is, the more  $P^C = (a^2+1)/2a = (a+1)^2/2a - 1$  moves away from  $P^J = 1$  and the smaller the weight  $a = (1+a)^{-1}$  in a weighted arithmetic mean is for which  $P^A = P^J = 1$  holds.

<sup>16</sup> Cooley (1893: 287) had already drawn attention to this point.

<sup>17</sup> It will become apparent that all indices constructed as ratios of averages (ROA), also known as "generalized unit value" indices, are capable of being interpreted in this way (i.e. in terms of expenditures).

In his controversy with Jevons, Laspeyres referred to the amount of money spent for a certain quantity of commodities instead of solely looking at price relatives.<sup>18</sup> This is clearly another paradigm. It implies that both the type of (selected) goods in question as well as their respective quantities should be the same in the two periods compared (a point Laspeyres repeatedly stressed). This also brings us to the two (not necessarily closely related) problems of defining a price level and assigning weights to prices.

### 3 Laspeyres and Drobisch: quantities as weights and unit values

Not only Laspeyres but also Drobisch started with the problem arithmetic vs. geometric mean. We will see that Drobisch did not have much to say about this point but instead came up with some new ideas about index numbers. His innovations were twofold:

1. to account for quantities as “weights” assigned to the prices,<sup>19</sup> and
2. to conceive an index as being a ratio of average prices (ROA) rather than an average of price ratios (AOR), which clearly contradicted Jevons and Laspeyres (the latter applied the then prevailing “arithmetic mean”  $P^C$  as unweighted and an AOR).

The second point triggered a long-lived controversy (although there are obvious formal<sup>20</sup> relationships between AOR and ROA) that has many ramifications and thus deserves discussion in a separate section (see Sec. 6 below).

#### 3.1 Quantities to account for the relative “importance” of goods

As to the first point, at that time it was by no means clear that weighted means of prices or price relatives are preferable over unweighted means,<sup>21</sup> and secondly if weights were used at all – for example to account for the “relative importance” of goods – it was far from generally accepted that such aspects are best represented by the quantities consumed. Problems of this kind gave rise to a sort of fundamentalism regarding index numbers which is unimaginable for us today. The “importance method” was attacked on the grounds that it would require dubious speculations about satisfaction and perceived well being which have nothing to do with the value of gold<sup>22</sup> (then generally seen as pivotal for prices). Even if quantities were agreed upon (as somehow proportional to “importance”), it was still found worthwhile to discuss:

1. whether explicit quantities  $q_{i0}$  or  $q_{it}$  should enter the formula or implicit quantities in the form of reciprocal prices would be appropriate, and
2. whether a selection of goods would do, or figures comprising all goods are required.

<sup>18</sup> Not only Laspeyres referred to this notion of an increase of the price level (“inflation” was not yet widely in use) time and again; Oker 1896 and some other authors also expressed it very distinctively.

<sup>19</sup> The British Association (1902: 29) also acknowledges that Drobisch’s index had possibly been the first weighted index. The type of weights were called “fluctuating” weights (as they included not only  $q_0$  but in particular also  $q_t$ ).

<sup>20</sup> Conspicuously, at that time economists (above all in Germany) were habitually not content with purely formal arguments.

<sup>21</sup> The most prominent advocate of unweighted means was clearly F. Y. Edgeworth, but Laspeyres, Jevons and Giffen also repeatedly expressed the conjecture that adding weights might eventually not make a difference. For more details reference can be made in particular to Laspeyres (1883: 797–798) and Walsh (1901: 87–88).

<sup>22</sup> For Pierson (1895: 332) these were “two problems bearing a wholly different character”. For him this and the existence of different yet equally reasonable formulas as well as ambiguities with respect to the choice of the base period (we now would say violation of time reversibility) gave reason enough to demand that the system of index numbers “is to be abandoned altogether” (Pierson (1896: 127)).

**Ad 1:** Implicit *quantities* should be kept distinct from indirect *weights*. The latter is given when one good is represented by a great number of varieties, whereas another good is only represented by one price or no price quotation at all. This was very quickly and fairly generally recognized as a problem of indirect or implicit weighting. Notably Jevons made many experiments with omitting goods or allowing for “importance” by adding more variants (and he concluded that weights would not make a substantial difference).

Implicit quantities (as reciprocal prices) on the other hand are intended to take into consideration that e.g. a pound of silk costs more (viz.  $p_{s0}$ ) than a pound of bread ( $p_{b0} < p_{s0}$ ).  $1/p_{s0}$  and  $1/p_{b0}$  are therefore the quantities of silk and bread respectively that are affordable for one currency unit, and relating prices to such “implicit” quantities will provide suitable weights and a common denominator allowing a summation of prices across the board. Many unweighted index formulas, which have now fallen into oblivion, follow this kind of reasoning and account for reciprocal prices of period 0 or t or both periods (see below in Sec. 4 with the index  $P^Y$ ). This way of looking at “weights” (related to the preciousness of goods rather than their physical weight) in the form of “implicit” quantities is an elegant device to give some unweighted price indices a meaningful interpretation in terms of money expenditure and “quantity”. It also relates a ROA approach like

$$P_{0t}^D = \frac{\sum p_{it}}{\sum p_{i0}} = \frac{\bar{p}_t}{\bar{p}_0} = \frac{e_t}{e_0} \text{ to an AOR approach } \frac{\sum p_{it} \frac{1}{p_{i0}}}{\sum p_{i0} \frac{1}{p_{i0}}} = \frac{\sum p_{it} \frac{1}{p_{i0}}}{n} \text{ resulting in } P_{0t}^C. \text{ So } P^D$$

(using weights of  $1/n$ ) becomes  $P^C$  by using weights  $1/p_{i0}$ , and vice versa; we can easily translate an AOR index like  $P_{0t}^C$  into a ROA formula  $P_{0t}^D$  (as done in eq. (5)). Implicit quantities  $q_i^*$  may also allow an interpretation of  $e_t = \sum p_{it} q_i^* = \sum p_{it} (1/p_{i0})$  and  $e_0 = \sum p_{i0} (1/p_{i0}) = n$  as a sort of “expenditure”.

Nowadays, some authors provide interpretations of unweighted price indices (composed of only prices without any explicit quantities) in terms of substitution behavior which is said to be implied in the formula under consideration. They do so asking which type of weighted index (e.g. Laspeyres, Fisher etc.) will be approximated by such an unweighted index given that prices are sampled with probabilities of selection proportional to quantity shares  $q_{it}/\sum q_{it}$  or expenditure shares  $p_{i0}q_{i0}/\sum p_{i0}q_{i0}$  and the like.<sup>23</sup>

**Ad 2:** A not uncommon view at the time under consideration here was that “quantities” should comprise all sorts of goods (not only actually transacted goods) because money has “power over all goods” (stocks and current production of goods, as well as financial assets),<sup>24</sup> and as it appeared practically unfeasible to provide such weights some people called index numbers “intrinsically impossible”<sup>25</sup> and discarded them as a futile search for the philosopher’s stone<sup>26</sup>. Such a wide definition of the “price level” covering all transacted goods (including financial assets) also became popular when some decades later more and more attempts were made to verify the (definitional!) equation of exchange. While such considerations are no longer interesting now, the following aspects of weights are continually relevant.

<sup>23</sup> See for example Balk (2005).

<sup>24</sup> This definitely applies to Lehr (1885: 37) and also to Laspeyres, quite distinctively in (1883: 796): “Wir haben keine genügende Statistik der durchschnittlichen Konsumtion irgendeines Landes” [We do not have a sufficient statistic of the average private consumption of any country]). See also British Association (1902).

<sup>25</sup> In German: “aus inneren Gründen unmöglich” Held (1871: 321).

<sup>26</sup> Held (1871: 326). Interestingly Held praised Drobisch for the simple reason that he opposed Laspeyres and that this work is good for fostering mistrust in the at that point new method of index numbers.

Once recourse to a *selection of explicit* quantities is agreed upon, a decision has to be made on a “single weighting system” (making use of either  $q_{i0}$  or  $q_{it}$  as suggested by Laspeyres and Paasche) or a “double weighting system”<sup>27</sup> (using both the quantities  $q_{i0}$  and the “fluctuating”  $q_{it}$  in an index). Drobisch introduced quantities in the concept of a “unit value” of all  $n$  goods at time  $t$ , defined as follows:

$$\tilde{p}_t = \frac{\sum p_{it} q_{it}}{\sum q_{it}} = \sum p_{it} \frac{q_{it}}{\sum q_{it}} \quad \text{by contrast to} \quad \bar{p}_t = \sum p_{it} \frac{1}{n} \quad (6)$$

and  $\tilde{p}_0$  defined correspondingly, which are meant to reflect the price level of a rather comprehensive set of goods. The use of  $\tilde{p}$  instead of  $\bar{p}$  (as in Dutot’s index) has the advantage of avoiding a commensurability problem with respect to prices. Clearly  $\bar{p}$  is affected by a move from prices ( $p_i$ ) quoted in kilograms to prices quoted in pounds ( $1/2p_i$ ), whereas an expenditure  $\sum p_i q_i$  in the numerator of  $\tilde{p}$  is invariant to such changes. The troublemaker is, however, the denominator because  $\sum q_i$  is in general not defined across all goods. It is difficult, if not impossible, to add over bushels of wheat, tons of iron, yards of cloth and hours of bus rides.<sup>28</sup>

Drobisch felt sure that he had solved this problem properly by requiring that all quantities should be expressed uniformly in hundredweights (cwt., “Centner” in Drobisch’s text).<sup>29</sup> This would rule out different results due to isolated changes in only some of the prices but it still does not render the index uniquely determined. As Walsh later pointed out, this is because a change from physical weight in cwt. to another dimension, say bulk [capacity, volume] measured in gallons or cubic meters, again applied to all goods, would yield a different  $\tilde{p}$ .<sup>30</sup>

Drobisch was not the only author who made use of unit values. Eduard Segnitz (1870) also introduced  $\tilde{p}$  as an alternative to the then very popular “midpoint” of prices defined as  $(p_{\min} + p_{\max})/2$  and he was also (possibly unlike Drobisch) aware of the fact that  $\tilde{p}_t$  is quite sensitive with regard to the length and position of the time interval  $t$  to which its quantities (as a flow variable) refer. It is known from scanner data, now increasingly in use, that it matters a lot whether the  $q$ ’s and therefore  $\tilde{p}$  refer to a week or a month and whether the time interval covers some extraordinary events (e.g. sale promotions and the like) or not.<sup>31</sup>

<sup>27</sup> These terms appear to be introduced by Walsh who was, like Drobisch and many others, especially in the Anglo-American index theory, vigorously in favour of a “double system”. He wrote (1901: 383) “...the method first discovered by Drobisch of comparing the averages of prices at each period on the mass-quantities of each period, and so employing what we have called *double weighting*”. Apart from the double system, which was much to Walsh’s liking, however, Walsh had rather a low opinion of Drobisch. Walsh considered the formulas of Drobisch ( $P^{DR}$ , eq. 7) and Lehr ( $P^{LE}$ , eq. 13) as representatives of double weighting.  $P^{LE}$  amounts to taking averages of weights  $q_{i0}$  and  $q_{it}$  (for each commodity  $i = 1, \dots, n$ ), just like Walsh’s preferred solution ( $q_{i0}q_{it}$ )<sup>1/2</sup> which he called “Scrope’s emended method”; Walsh (1901: 540-543).

<sup>28</sup> The nonexistence of such sums over dissimilar quantities is the central shortcoming of unit values.

<sup>29</sup> Note that he did not seek a way to account for the different preciousness of the goods.

<sup>30</sup> Hence unit values are acceptable only for a fairly homogeneous set of goods and thus only for “low level aggregations” and sub-aggregates, that is, as building blocks (taking the part of genuine prices) for greater aggregates. In this sense we have in some countries “unit value indices” (not to be confused with Drobisch’s index), especially for the price levels of exports and imports because they are readily available as a by-product of foreign trade statistics.

<sup>31</sup> Segnitz for example maintained that the interval should be neither too short, nor too long. As to experiences with (and the treatment of) scanner data, see Ivancic, Diewert and Fox (2011).



From the definition of  $\tilde{p}$  we quite naturally arrive at Drobisch' price index as a ratio of unit values<sup>32</sup>

$$p_{0t}^{DR} = \frac{\tilde{p}_t}{\tilde{p}_0} = \frac{\sum p_{it}q_{it}/\sum q_{it}}{\sum p_{i0}q_{i0}/\sum q_{i0}} = \frac{\sum p_{it}q_{it}}{\sum p_{i0}q_{i0}} \cdot \frac{\sum q_{i0}}{\sum q_{it}} = \frac{V_{0t}}{Q_{0t}^D}. \quad (7)$$

Hence as  $\tilde{p}$  is the quotient of an expenditure (or a more general value) and a quantity, so  $p^{DR}$  is a quotient of the respective indices,  $V_{0t}$  and  $Q_{0t}$ .

Eq. 7 provides a sort of indirect definition of inflation: less quantity for the same amount of money.  $p^{DR}$  may be viewed as an "indirect" (Diewert) or "factor antithetic" (I. Fisher) price index gained by dividing  $V_{0t}$  by Dutot's quantity index  $Q_{0t}^D = \sum q_t / \sum q_0 = \bar{q}_t / \bar{q}_0$ .<sup>33</sup>

### 3.2 Double and single weights

It is beyond the scope of this historically oriented paper to discuss the altogether disappointing axiomatic record of  $p^{DR}$ .<sup>34</sup> Since in Drobisch's days great store was generally set by the chain test (transitivity), it is remarkable that  $p^{DR}$  is able to comply with this rarely met requirement because  $p_{0t}^{DR} = p_{01}^{DR} p_{12}^{DR} \dots p_{t-1,t}^{DR}$ . Most noteworthy is, however, that Laspeyres realized that  $p^{DR}$  violates the identity axiom,<sup>35</sup> which requires that a price index should be unity if all prices in  $t$  are equal to those of 0. This assumption  $p_{it} = p_{i0}$  for all  $i$  yields

$$p_{0t}^{DR} = \frac{\sum p_{0t}q_t / \sum q_t}{\sum p_{00}q_0 / \sum q_0} = \frac{Q_{0t}^L}{Q_{0t}^D},$$

and there is no reason to assume that  $Q^D = Q^L$ , or (equivalently) that for all goods quantity shares  $q_{i0}/\sum q_{i0}$  and expenditure shares  $p_{i0}q_{i0}/\sum p_{i0}q_{i0}$  coincide.

Most importantly, violation of identity implies that  $p^{DR}$  does not comply with the idea of "pure price comparison" (a price index should only reflect a price movement), which indeed is the cornerstone of Laspeyres' thinking (see below).

Drobisch was well aware of the fact that his formula specializes to

$$p_{0t}^L = \frac{\tilde{p}_t^*}{\tilde{p}_0} = \frac{\sum p_{it}q_{i0}/\sum q_{i0}}{\sum p_{i0}q_{i0}/\sum q_{i0}} = \frac{\sum p_{it}q_{i0}}{\sum p_{i0}q_{i0}} \quad (\text{Laspeyres price index}), \quad \text{and} \quad (7a)$$

$$p_{0t}^P = \frac{\tilde{p}_t}{\tilde{p}_0^*} = \frac{\sum p_{it}q_{it}/\sum q_{it}}{\sum p_{i0}q_{it}/\sum q_{it}} = \frac{\sum p_{it}q_{it}}{\sum p_{i0}q_{it}} \quad (\text{Paasche price index}). \quad (7b)$$

<sup>32</sup> They are meant as absolute price levels, so  $p^{DR}$  is a typical ROA index.

<sup>33</sup> Such an interpretation in terms of the factor reversal test linking a price and a quantity index to the value ratio (see Fisher (1911: 418)) was not familiar to Drobisch, who died in 1896. Also the name Dutot and the concept of a "quantity index" in general was not yet widely known in Drobisch's days. Balk (2008: 7, 73).

<sup>34</sup> For details cf. Balk (2008: 72) and von der Lippe (2007: 18-20).

<sup>35</sup> It is perhaps for this reason that Laspeyres is widely recognized as the "inventor" of this axiom (it is most likely, however, that Laspeyres was not yet aware of the fact that identity is a special case of proportionality). This achievement of Laspeyres and his critique of  $p^{DR}$  is also reported in British Association (1902: 30). As pointed out there, the violation of identity (as a disadvantage) may, however, be set against the advantage that  $p^{DR}$  can (unlike  $p^L$ ) reflect substitutions households make in response to changes in relative prices.

What makes the difference between  $P^L$  and  $P^P$  on the one hand and  $P^{DR}$  on the other is that in  $P^L$  and  $P^P$  reference is made to the same quantities in the numerator and the denominator. This, however, is most important as it avoids the problem with commensurability in  $P^{DR}$ . Also, use is made in  $P^L$  and  $P^P$  of “hybrid” values like  $\Sigma p_t q_0$  and  $\Sigma p_0 q_t$ , which Drobisch considered illegitimate and thus consistently avoided in his formula  $P^{DR}$ .

By way of some numerical examples, Laspeyres studied how his index  $P^L$  is related to  $P^{DR}$ . However, he did not come to conclusions that could be generalized beyond his example. It will be shown here, in an appendix, how these indices are related to one another.

Lehr (1885: 41) also realized that a unit value may indicate a change ( $\tilde{p}_t \neq \tilde{p}_0$ ) although no price in the aggregate has changed, which means that the index  $P^{DR}$  violates identity. Lehr therefore rightly maintained that prices are comparable only when the quantities in two periods to be compared either do not differ or at least are proportional<sup>36</sup>

It is interesting to see that Drobisch claimed to be credited with authorship of  $P^L$  and  $P^P$ , just because both formulas emerge as special cases of  $P^{DR}$  (which gives rise to our digression in Sec. 4), although he argued against these formulas, mainly by repudiating somehow hybrid average prices of the type  $\tilde{p}_t^*$  and  $\tilde{p}_0^*$  as allegedly being logically illegitimate.<sup>37</sup>

He did so in his rebuttal of Laspeyres’ critique concerning identity,<sup>38</sup> in which he saw an attempt by the latter to place a “death-blow” (“*Todesstoß*” as he put it) to his formula. His reaction was not only peeved but also in no small measure helpless. He argued that Laspeyres might be right “in calculo”, but that neither  $P^C$  nor  $P^L$  is “an authority” for him and that prices in  $t$  (and  $0$  respectively) should not be averaged with quantities other than  $q_t$  (or  $q_0$  in the case of  $p_0$ ).  $\tilde{p}_t^*$  is therefore logically illegitimate. He obviously did not see that  $\tilde{p}_t^* = \tilde{p}_t$  once the assumption is made. This kind of replying to a “formal” argument by appealing to “logic”, “plausibility” and allegedly wrong comparisons is another perennial game in index theory. It was to become very popular, especially in Germany from the 1920s onwards.<sup>39</sup>

Such epistemological issues bring us back to Jevons’ choice of the geometric mean. As remarked above, Drobisch did not come to a definite conclusion on this issue, the resolution of which was what was initially called for. In Drobisch (1871b: 154) we find the conjecture that Jevons might have chosen the geometric mean simply because

<sup>36</sup> *Durchschnittswerte* (unit values) sind “nur unter der Voraussetzung miteinander vergleichbar, daß die Mengen zu verschiedenen Zeiten sich überhaupt nicht oder doch nur im gleichen Verhältnisse ändern” Lehr (1885: 42) (Unit values “are only comparable under the assumption that quantities at different points in time are unchanged or have only changed in the same proportion”), that is,  $q_{it} = \lambda q_{i0}$ . This might be understood as an argument in favour of Laspeyres’ formula. Lehr’s second objection against Drobisch was the commensurability problem with  $\Sigma q_0$  and  $\Sigma q_t$  total quantities, which are not even defined across all goods.

<sup>37</sup> He also preferred his formula due its being ostensibly more general, and he erroneously believed that  $P^L$  is unrealistic as it requires all quantities to remain constant over time. He apparently seems to have overlooked that the  $q_0$ ’s are kept constant only for analytical purposes, that is, only in a kind of thought experiment.

<sup>38</sup> Apart from this case, to our knowledge Drobisch never entered into discussions about the rightly criticized flaws of his formula.

<sup>39</sup> This refers in particular to the many publications of Paul Flaskämper and his project to develop statistics (and index formulas in particular) solely from reflections on logical conditions of comparability (and mostly without mathematics, or at best mathematics only of the simplest kind). It is beyond the scope of this paper to go into details here, but it is interesting to see why this project was bound to fail.

it yields a lower inflation rate, and also an explicit critique of Jevons which reads as follows: “*Hier vermisst man nun ganz und gar einen positiven und allgemeinen Grund, aus welchem dem geometrischen Mittel vor dem arithmetischen der Vorzug gebühren soll*” (emphasis by Drobisch. “Here we miss completely a *positive* and *general* reason why we should prefer the geometric mean over the arithmetic.”). Furthermore Drobisch correctly pointed out “...*dass aus der Unzulässigkeit des arithmetischen Mittels nicht die Nothwendigkeit des geometrischen folgt, da es ja ausser diesen beiden noch viele andere Arten von Mittelgrößen giebt.*” (p. 154F, “...that from the inadmissibility of the arithmetic mean, it does not follow that the geometric mean must be taken, as there are many more means in addition to these two”). He was also not short of unsolicited epistemological advice addressed to Laspeyres. On the other hand there is much in Drobisch’s own writings which may well be criticised. To give an example, Drobisch rejected other formulas as inappropriate or unacceptable for the simple reason that they only coincide under very restrictive and unrealistic conditions with his formula (he studied  $P^C$ ,  $P^J$  and also  $P^L$  solely with this intention). The less realistic the assumptions must be in order to approximate his formula, the less meaningful a formula was for Drobisch, as if his formula were the indisputable standard against which everything else should be measured.

#### 4 Digression on priority claims and the re-discovery of formulas (Young’s formula)

There are reasons why disputes about priority, such as those that took place between Drobisch and Laspeyres, are sometimes quite difficult to settle. In our view it is not sufficient to realize that  $P^L$  (and also  $P^P$ ) emerges as a special case of  $P^{DR}$ , or to indirectly accept both formulas ( $P^L$  and  $P^P$ ) as equally valid or invalid by suggesting a simple arithmetic mean  $\frac{1}{2}(P_{0t}^L + P_{0t}^P)$  of them.<sup>40</sup> In order to claim authorship of a formula it is also desirable if not plainly necessary to demonstrate the comparative advantages of the respective formula over other formulas, and this is precisely what Drobisch clearly failed to do. Furthermore, it was Laspeyres who figured out some shortcomings of Drobisch’s formula  $P^{DR}$  and thereby advanced arguments to prefer  $P^L$  over  $P^{DR}$  so that he should rightly be credited for  $P^L$ , not Drobisch.

As Kuhn said, “discovery” is a complex process which involves at least two steps: “*that* something is and *what* it is”<sup>41</sup>. He illustrated this fact with the example of the discovery

<sup>40</sup> As done in Drobisch (1871: 425). It noteworthy that in this paper Drobisch was prepared to accept *any* kind of weighted arithmetic mean  $aP^L + (1-a)P^P$ , not only  $a = 1/2$ . He was quite indifferent about which  $a$  to choose and he suggested this index only in an interrogative sentence: “*Man könnte nun zwar davon das arithmetische Mittel nehmen, welches giebt ..., aber muss den der richtige Werth gerade in der Mitte ... liegen?*” (“One could now possibly take the arithmetic mean, giving ...; but why should the true result lie exactly in the middle between the two?” Drobisch (1871c: 425)). Interestingly Drobisch not only saw no reason to prefer one formula over the other, he also made use of “crossing” of formulas which later became very fashionable (Irving Fisher in particular made extensive use of it in his index theory). In the Anglo-American literature the above mentioned index Drobisch suggested (for  $a = 1/2$ ) is also known as the Sidgwick–Bowley index (see e.g. Diewert 1993 for more details). Also v. Bortkiewicz (1932: 24) remarked that the index  $1/2(P^L + P^P)$  should not be credited to (“the philosopher”) Drobisch – just because of his quoted indifference regarding the choice of  $1/2$  for  $a$  – and he also said that the formula was proposed by Henry Sidgwick (an English utilitarian philosopher 1838 – 1900).

<sup>41</sup> Kuhn (1996: 55).

of oxygen. There were at least three claimants of authorship: Scheele, whose experiments led him to infer its existence but who was too hesitant to publish his finding in time; then a bit later Priestley, who was the first person able to isolate this gas but was unable to understand its real nature; and finally Lavoisier, who after having received hints from Priestley was the first who analyzed and understood (almost) correctly what kind of gas it was.

The lesson to be learned from this example is that claims of priority are more often than not questionable. In addition, a potential forerunner is sometimes less precise and delivers only a more-or-less vague verbal description rather than a formula. For example, Jastram 1951 observed that Willard Phillips might be called a predecessor of Paasche because he wrote in his *Manual of Political Economy* in 1828 that an absolute (constant) measurement rod of value should not be strived towards and could not be established in the form of labour (which was generally accepted by his contemporaries). Instead, the standard of value should be different for different times, and depend on prices of goods. Phillips also suggested that “quantities of the different articles assumed ought to be in proportion of the consumption or the amount possessed in the country or district for which the measure is framed.”<sup>42</sup> Phillips also noticed that substitutions will make  $q_0$ -weights inappropriate: “Without changing the amounts of articles to correspond to the differences of consumption, the table would not be a fair representation.” This may be understood in such a way that Phillips “table” should include quantities  $q_{it}$  rather than  $q_{i0}$ . However a “table” (not even worked out empirically, and not reduced to a ratio of expenditures) is not yet a formula which in turn is more than just a (suggested) list of  $q_{it}$  quantities.

We can also easily quote various remarks which can be interpreted “with hindsight” as a very early allusion to the “economic theory of index numbers”. Such words can even be found in the writings of authors who plainly rejected index numbers altogether.<sup>43</sup> Yet such more-or-less vague and only occasional remarks are far from anticipating the mathematically developed economic approach as presented for the first time only as recently as 1924 by Konüs.<sup>44</sup> To my knowledge it is due to von Bortkiewicz that his work became known beyond Russia.<sup>45</sup>

It not only happens that somebody has dubious claims of priority and an *alleged* authorship, it may also easily happen that authorship is erroneously claimed because something new (subjectively) is found without knowing of a *real* predecessor and therefore independently of him. Such “re-discovery” is not unusual, even today. The following provides an example of “rediscovery” in index theory. Allyn Young (not to be confused with Arthur Young 1812) proposed the following seemingly weird and unmotivated formula of Young (1923: 357) which reads as follows:

<sup>42</sup> Jastram (1951: 125).

<sup>43</sup> This applies for example in Germany to Held (1871: 331) for whom with inflation the question was whether “...noch die alten Bedürfnisse im alten Umfang oder nur in geringerem Umfang befriedigt werden können” (...we can continue to satisfy our needs to the old extent or only to a limited degree) and interestingly he concluded (just like Paasche 1878) that much of what seemed to be inflation was only a self-deceit due to the growing needs of consumers.

<sup>44</sup> Held, who was an engrained skeptic as regards mathematics in economics and particularly index numbers (as were many others in Germany at that time) was of course light years away from Konüs.

<sup>45</sup> See also von Bortkiewicz (1932: 18), where he quotes the original Russian text of Konüs.

$$P_{0t}^Y = \frac{\sum p_t \sqrt{\frac{1}{p_0 p_t}}}{\sum p_0 \sqrt{\frac{1}{p_0 p_t}}} = \frac{\sum \sqrt{\frac{p_t}{p_0}}}{\sum \sqrt{\frac{p_0}{p_t}}}, \quad (8)$$

and which Irving Fisher (1927: 530 f.) later called “an ingenious anomaly, scarcely classifiable” (in the scheme of Fisher’s book) and “a scientific curiosity”. Not surprisingly, it soon fell totally into oblivion, possibly also because the derivation of  $P^Y$  was not well understood, although it can easily be explained. Using implicit quantities (by way of inverse prices), Young found that “base year weighting” in  $P_{0t}^C = \sum p_t \frac{1}{p_0} / \sum p_0 \frac{1}{p_0}$  “overweights rising prices”, by contrast to  $P_{0t}^H = \sum p_t \frac{1}{p_t} / \sum p_0 \frac{1}{p_t}$ , the harmonic mean, which tends to underweight them. Thus he was quite naturally led to the geometric mean  $\sqrt{\frac{1}{p_0 p_t}}$  as a compromise.

The formula was then rediscovered by Bert Balk, who called it the Balk-Walsh index,<sup>46</sup> because with *explicit* quantities we obtain  $P_{0t}^W = \frac{\sum p_t \sqrt{q_0 q_t}}{\sum p_0 \sqrt{q_0 q_t}}$ , i. e. Walsh’s formula, as the weighted counterpart. The geometric mean of  $P^C$  and  $P^H$ , called the CSWD-index<sup>47</sup> is also known to approximate  $P^J$  fairly well.

Another rediscovery of  $P^Y$  took place when Jens Mehrhoff – in a short note he contributed to von der Lippe (2007: 45 f.) – looked for a linear index able to approximate  $P^{CSWD}$  and thereby also  $P^J$ . He called it “hybrid index”, and later the BMW-index (Balk Mehrhoff Walsh), not knowing that it coincides with  $P^Y$ .

Young also saw that his index meets the time reversal test but not the circular test, which means that  $P^Y$  is not transitive

$$P_{02}^Y = \frac{\sum \sqrt{\frac{p_2}{p_0}}}{\sum \sqrt{\frac{p_0}{p_2}}} \neq \frac{\sum \sqrt{\frac{p_1}{p_0}} \sum \sqrt{\frac{p_2}{p_1}}}{\sum \sqrt{\frac{p_0}{p_1}} \sum \sqrt{\frac{p_1}{p_2}}}, \quad (8a)$$

and finally he also noticed (interestingly in view of Mehrhoff’s intentions that led him to  $P^Y$ ) as regards  $P^Y$  that, “In general it will agree very closely with the geometric average” (357) i. e. with  $P^J$ .

## 5 Laspeyres and Paasche: single quantity weights ( $q_0$ or $q_t$ ?) and “pure” comparison

It has often been stated (approvingly for example by Walsh) that Drobisch’s formula (as well as Lehr’s below) may be viewed as a double weighting formula while  $P^L$  and  $P^P$  represent formulas with single weights only. Today, so called “symmetric” index formulas in the sense of price index functions  $P(p_0, q_0, p_t, q_t)$  that treat price and quantity vectors of both the period 0 and the period  $t$  in a symmetric manner (such as Fisher’s or Törnqvist’s index) are often viewed (e.g. by Diewert) as being superior to indices like  $P^L$  and  $P^P$  that only make use of either  $q_0$  or  $q_t$  respectively. Symmetric indices particularly stand out in relief against other indices because all “superlative” indices

<sup>46</sup> Balk (2005 and 2008: 187-191).

<sup>47</sup> Proposed by Carruthers et al. (1980) and Dalen (1992); see Balk (2008: 184).

(i.e. indices able to approximate the “true cost of living index”, or “constant utility index” in the sense of the economic index theory) are symmetric, such as for example the indices of Fisher, Walsh and Törnqvist.<sup>48</sup> The definition of “symmetry” applies to Drobisch’s index  $P^{DR}$  as well, but  $P^{DR}$  is far from being superlative.

Of course the notion of superlative indices was unknown in the 19<sup>th</sup> century and it was definitely not the intention of Prof. Laspeyres or Prof. Paasche to provide an upper or lower bound to the cost of living index, and so it is better to restrict ourselves in the discussion of pros and cons of formulas to those ideas that were already known in the late 19<sup>th</sup> century.

Even before the economic index theory became influential, the proponents of double weights seemed to have prevailed over the “single weighters” and it fits to their view that they consider the  $P^L$  and  $P^P$  indices to be equally well reasoned.<sup>49</sup>

This being the situation, “single weighters” have always had a hard job. Many theories have been advanced as to why Laspeyres insisted upon  $q_0$ -weights and Paasche on  $q_t$ -weights. Interestingly, both were conspicuously taciturn as regards this issue. In his controversy with Drobisch, Laspeyres confined himself to exploiting the comfortable position that he could quote Drobisch (although both men obviously disliked one another considerably) for this purpose.<sup>50</sup> He apparently thought that this would be disarm Drobisch and save him the trouble of substantiating his position.

It is often stated that Laspeyres only took quantities  $q_0$  for practical reasons,<sup>51</sup> and that he would have taken  $q_t$  (or  $q_t$  in addition to  $q_0$ ) if only he had better access to timely data on such quantities. Lack of suitable data were admittedly the reason for initially only using the unweighted  $P^C$ -index, however, it is far from clear that he would have preferred a constant and timely update of weights, or even a double weighting system, if only he had had access to appropriate data.

Laspeyres (1883: 796) is one of the rare occasions where he discussed – explicitly referring to Conrad and Paasche – the problem of whether  $q_0$ ,  $q_t$  or some average of both should be taken. He concluded “*Doch sind dies praktisch noch unzweckmäßige Fragen*” (“However, these questions are still unsuitable from a practical point of view”). He was referring to the state of statistics on *all* quantities produced and consumed. As already mentioned, he was obviously misled by the then common belief that such quantities ought to refer to the whole economy rather than a sample of consumers. And as he saw that he was unlikely to get such statistics,<sup>52</sup> he decided to pay more attention to

<sup>48</sup> More about the notions “symmetric” and “superlative” index functions cf. Diewert (1976).

<sup>49</sup> “Nothing can be offered in proof of the superiority of the one over the other” (Walsh as discussant in Fisher 1921: 538), a statement which may serve as backing of the widely held view that some kind of average of the two indices (like Fisher’s “ideal” index  $(P^L P^P)^{1/2}$ ) should be taken. There was a discussion in Germany in the late 19<sup>th</sup> century about whether or not to average index functions like  $P^L$  and  $P^P$  or to average weights ( $q_0$  and  $q_t$ ). We will come back to this at the end of this section. At the moment our focus is on Laspeyres’ position (as opposed to Drobisch).

<sup>50</sup> He quoted Drobisch (1871b: 145): “*Wir nehmen dabei, zur Vereinfachung an, dass seine Lebensbedürfnisse in qualitativer Hinsicht sich gleich geblieben sind, und auch quantitativ sich weder vermehrt noch vermindert haben.*” (To make things easier we assume that needs did not rise nor fall, neither quantitatively nor qualitatively.)

<sup>51</sup> Roberts (2000: 10).

<sup>52</sup> The exceptions he saw were import statistics and consumption patterns of working class households, possibly based on a sample, because the variability of such patterns will tend to be smaller than for better-off families.

other points, especially a justification of unweighted means like  $P^C$  which he continued to prefer over his own formula for many purposes (e.g. measuring the purchasing power of money).

In other contexts, however, we can see clear indications that he was not indifferent concerning  $q_0$  or  $q_t$ . There are good reasons to assume that he had deliberately kept weights constant even for a relatively long interval in time and even if availability of data could have enabled him to do otherwise. Constant quantities were essential to him as a device to imitate and simulate an experiment as the only way to prove causality. He was always immensely interested in both causal inference<sup>53</sup> and the prerequisites needed for making valid comparisons. Ideally, statistical figures should reflect a hypothetical and “pure” process as a surrogate of an experiment. Obviously for him constant quantities were an artificial “*ceteris paribus*” that permit isolating the factor “price” from other correlated variables and influences such as demand, income etc. Constancy is not meant as a counterfactual description of a real process but rather as a kind of model, intended to achieve in the social sciences something analogous to an experiment in the natural sciences.<sup>54</sup> The underlying idea is particularly clearly spelled out in his paper on “*Kathedersocialisten*” (Laspeyres 1875), where he wrote:

*Um “den Charakter der Bewegung kennen zu lernen muß man nicht vorwärts, sondern lieber rückwärts schauen und diejenigen Objecte aussuchen, welche ausnahmsweise eine lange Zeit in vergleichbarer Qualität producirt wurden”* (p. 18).<sup>55</sup> He demanded that these objects “*in die Vergangenheit recht weit zurückverfolgt werden können*” (it should be possible to trace them back fairly far into the past; emphasis original). And finally he said as a kind of credo: “*Die statistische Untersuchungsmethode kann einen Schritt weitergehen, sie nimmt nicht an, daß die anderen Umstände alle gleich seien, sondern sie macht alle anderen Umstände gleich, mit Ausnahme des einen, dessen Wirkung sie untersuchen will, den einen Umstand aber, dessen Wirkung sie untersuchen will, macht sie möglichst verschieden ...*” (p. 32).<sup>56</sup>

The reason for using quantities of the base period in a number of subsequent periods, which is the characteristic feature of  $P^L$ , can be seen in the sequence

$$P_{01}^L = \frac{\sum P_1 q_0}{\sum P_0 q_0}, P_{02}^L = \frac{\sum P_2 q_0}{\sum P_0 q_0}, P_{03}^L = \frac{\sum P_3 q_0}{\sum P_0 q_0}, \dots \quad (9)$$

<sup>53</sup> He introduced his “mammoth number-crunching” work (Roberts), which is Laspeyres 1901, and which kept him busy for many years with the question “*Kann man statistisch ein post hoc als ein propter hoc nachweisen?*” (Can you prove statistically a “because” with observations of the “after” type?).

<sup>54</sup> This idea is rejected with much vigour in Winkler (2009: 101–110), who in a way represents the very opposite of Laspeyres and recommends going back to Dutot’s index.

<sup>55</sup> This means: “In order to understand the character of a movement you should not look ahead but rather backwards and choose such objects that coincidentally are produced for a long time in comparable quality.” In the light of some other statements it seems to be fair to say that Laspeyres also would have emphasized “long”.

<sup>56</sup> “The statistical experiment method can go one step further (than theory), it does not *assume* that everything else remains constant, it rather *makes* all other circumstances constant with the exception of the one whose effects it wishes to test, which is made as different as possible.”



in which subsequent indices differ only with respect to prices. By contrast for  $P^P$

$$P_{01}^P = \frac{\sum P_1 q_1}{\sum P_0 q_1}, P_{02}^P = \frac{\sum P_2 q_2}{\sum P_0 q_2}, P_{03}^P = \frac{\sum P_3 q_3}{\sum P_0 q_3}, \dots \quad (9a)$$

we get continually changing weights  $q_1, q_2, q_3, \dots$ . Prices in this sequence (as opposed to  $P^L$ ) are therefore not comparable among themselves but only to  $p_0$ .

There were (and still are) not many people who clearly distinguish between a “year-on-year” (or bilateral) comparison only, and a “comparison-in-series”, as Young (1923: 364) put it. In the former situation  $P^P$  may be as good as  $P^L$  (i.e. what applies to 0 in  $P^L$  simply applies to  $t$  in  $P^P$ , and  $t$  is one period just like 0), whereas in the case of a series  $t$  denotes not one period (like 0) but a number of periods ( $t = 1, 2, \dots$ ) and there  $P^L$  may well be preferred over  $P^P$  from the point of view of “consistent series” (Young), or “pure price comparison”,<sup>57</sup> a concept which is more difficult to define in exact terms than might appear at first glance.<sup>58</sup> While  $q_0$  is kept constant (for some periods),  $q_t$  is “fluctuating”, constantly changing with the passage of time.

Of course there were soon critics of the  $P^L$  formula in abundance, and they quickly got into the habit of deriding  $P^L$  predominantly because of its constant weights. This has continued to be the standard argument ever since, and it goes as follows: Keeping the selection of goods and their weights constant is difficult in a dynamic economy and results in the index sooner or later hopelessly losing touch with reality.

As to Paasche, the situation is quite different. It is difficult to find pronounced statements as to why he preferred weights  $q_t$  over the weights  $q_0$ , and why he did not chose both  $q_0$  and  $q_t$ .<sup>59</sup> We can find statements concerning the first point, but few (if any) concerning the second.

There are remarks on the part of Paasche that were intended to justify the preference for a single weight system (which also would apply to  $q_0$  instead of  $q_t$ ), intended in the first place to avoid ambiguity of the index (reflecting possibly both price *and* quantity movement). Similar arguments can also be found in the writings of Johannes Conrad, a long time editor of this journal (1878 – 1915) and a promoter of Paasche and many other authors in index theory.<sup>60</sup>

<sup>57</sup> We could not find suitable quotations published in Laspeyres’ time but only some fifty years later. In addition to Young, Persons also had a similarly clear position (and explicitly advocated  $P^L$ ) as a discussant in Fisher 1921. He said there: “An index number is not computed merely to compare the index number of one given year with that for the base year, but to compare the indices for a series of years with each other ...” (p. 545). He definitely argued in favour of pure price comparison, because for him double, and therefore *variable* weighting, “has the defect that we do not know whether changes in the indices result from changes in the prices or production” (p. 545). Conspicuously Fisher made clear in his rejoinder that the “time reversal test” (on which he and many others laid a disproportionate stress) rules out all indices “which do not have symmetrical or ‘double’ weighting” (p. 549), and that Person’s position is inconsistent and amounts to “demanding the impossible”.

<sup>58</sup> Cf. von der Lippe (2005) (the whole paper is on the problem of defining “pure comparison”).

<sup>59</sup> Lehr (1885: 44) argued that Paasche wanted the same quantities in numerator and denominator in order to avoid problems (of  $P^{DR}$ ) with commensurability. However, it is not double weighting that ensues commensurability problems but rather the summation of quantities (be they  $q_{i0}$  or  $q_{it}$ ) over  $n > 1$  goods ( $i = 1, \dots, n$ ). Lehr’s own index (see below) makes use of  $q_{i0}$  and  $q_{i1}$  but the summation takes place over periods (0 and 1) for each good  $i$  separately (just like the averages  $(q_{i0}q_{i1})^{1/2}$  in Walsh’s index) so that no commensurability problems can arise.

<sup>60</sup> Among them for example the American Samuel McCune Lindsay, who (like Paasche) received his Ph. D from Conrad in Halle and whose book on prices (in German) was extensively commented by Edgeworth (1894).

However, Paasche also expressly stated that he could well imagine that a double system would make sense. He mentioned possible studies of how consumption changed and households escape inflation by substituting goods, but in a nebulous and not at all satisfactorily substantiated remark he concluded that taking both periods into accounting would not be advisable:

*“Aber für die einfache Constatierung und Berechnung der Preissteigerung würde das wenig helfen, weit mehr verwirren, denn das sind allerdings wesentliche Momente für die Bestimmung der Preise, aber für die Aufstellung des Verhältnisses der einmal gegebenen Werthe nicht weiter von Einfluß.”* Paasche (1874: 173).<sup>61</sup>

In summary his position was: quantities of the past  $q_{t-1}$ ,  $q_{t-2}$ , ... may be interesting as factors determining the present situation, but they should not be taken into account when the task is to establish a price index comparing 0 to  $t$ . This of course caused him some trouble with Lehr.

It is true that he gave detailed comments on the plausibility of his empirical index calculations as regards specific commodities, but he gave only sparse comments, if any, on why he preferred  $q_t$  over  $q_0$ . The motivation was possibly (as it always is in the standard critique of  $P^L$ ) only that more recent and constantly updated quantities are considered better.<sup>62</sup>

A final remark relating to  $P^L$  vs.  $P^P$  and single vs. double weights may be in order:<sup>63</sup> Drobisch felt irritated by the fact that both formulas ( $P^L$  as well as  $P^P$ ) are equally possible (or perhaps even equally plausible) and he requested an unequivocal solution which he believed to have found in his formula  $P^{DR}$ .<sup>64</sup> The problem of  $P^{DR}$ , however, is that it fails identity, as Laspeyres rightly noticed. Avoiding this seemed to require a single weighting system,<sup>65</sup> which on the other hand requires making a choice between  $q_0$  and  $q_t$ . Lexis proposed (as an “improvement” of Drobisch’s method) to make use of  $\frac{1}{2}(q_0 + q_t)$ , which results in a formula now known as the Marshall-Edgeworth index.<sup>66</sup> Lexis also viewed

<sup>61</sup> The quotation reads as follows: “However, this would not be helpful for the simple identification and quantification of a price increase, but rather cause confusion, because these aspects may be relevant as determinants of the present prices, but have little influence on relating the given prices (to the past by way of an index, he meant).” Not surprisingly Lehr (1885: 44) quoted (most disapprovingly) precisely this very sentence, especially as regards the (alleged) “confusion”. Lehr’s message – we will see – was essentially that what was called for was not a binary comparison (0 to  $t$ ), but rather a time series in which all intermediate periods are to be taken into account.

<sup>62</sup> Later van der Borgh (1882), who continued Paasche’s and Conrad’s regular compilations of price statistics in this journal, argued in a similar vein in favour of  $P^P$ : This index is more convenient when it is difficult to find prices in  $t$  which match with those in 0, as with  $P^P$  there is no need to look back in time. This kind of reasoning was, and still is, notably popular among all those who advocate chain indices. Richard van der Borgh later became president of the German Imperial Office of Statistics (1904 – 1912).

<sup>63</sup> The following chain of reasoning is nicely developed in Lexis (1886: 117-121).

<sup>64</sup> As mentioned above (Footnote 40), it is therefore not quite correct to credit Drobisch for the formula  $\frac{1}{2}(P^L + P^P)$  although he took it into consideration.

<sup>65</sup> According to Lexis (1886: 118), Paasche advocated the single weight system even more than Laspeyres did.

<sup>66</sup> The German text introducing the averaged weights  $\frac{1}{2}(q_0 + q_t)$ , reads as follows: “daß man für jede Ware den Durchschnitt aus der verkauften Menge des Anfangs- und des Endjahres in Rechnung brächte” Lexis (1886: 119). I only discovered this paper of Lexis thanks to a quotation of von Bortkiewicz (1932: 24) who also maintained that the formula was proposed in 1886 by Lexis and thereafter also recommended by Marshall and Edgeworth.

the formula of Lehr (eq. 13a) as an attempt to resolve the dilemma of choosing among two weights. For Lehr a (price) weighted average between quantities of adjacent periods appeared to be the solution<sup>67</sup> which quite naturally made Lehr advocate chain indices. So there are various possibilities that may lead us to plea for chain indices, e.g. a choice among weights (Lehr), ambiguities with non-time-reversible formulas which makes independence of the base period attractive (an argument used for example by Pierson (1896: 128), or Flux (1907: 618)), and perhaps above all practical problems with new and disappearing goods and the fixed weights  $q_0$  becoming progressively obsolete with the passage of time (a point made by many authors, among them Lehr). We will discuss Lehr's approach in more detail later (Sec. 7).

## 6 Average of price relatives (AOR) or ratio of average prices (ROA)

We have already introduced the distinction between the ROA approach, of which Drobisch's  $P^{DR}$  is an example, and the (before Drobisch prevailing) AOR approach (as in the form of  $P^C$  and  $P^J$ ). The dichotomy triggered a host of ultimately useless controversies as early as in the 19<sup>th</sup> century and has continued to do so ever since. The problem for proponents of the ROA approach<sup>68</sup> is that they are tempted to view price indices (comprising  $n > 1$  goods) by analogy to simple price relatives (each in turn referring to only one good) and thus to demand that indices fulfill all those axioms that relatives necessarily fulfill. This applies in particular to transitivity, but also to Fisher's reversal tests, which were obviously patterned against the model of simple relatives.

Proponents of the AOR view, e.g. Jevons, habitually hesitate to aggregate over prices referring to such different quantity units as hours, cwt, gallons etc., but they have no problems with the same figures when transformed into relatives (as they then become dimensionless pure numbers). Getting rid of such problems with dimensions was seen as a main advantage of AOR.<sup>69</sup>

<sup>67</sup> Lexis (1886: 118) criticized the choice of only two adjacent periods, that is  $q_0/q_1, q_1/q_2, \dots$  (he asked: why not average over more than just two periods?) and for von Bortkiewicz (1932: 31) the problem with Lehr's formula is that it violates proportionality (while identity is satisfied) and Lehr's erroneous interpretation of his formula in terms of "utility". He also did not endorse the chain approach of Lehr. In Bortkiewicz's view Walsh and Edgeworth were also too cautious and indulgent in criticizing Lehr. See also Walsh (1901: 386, 547), Edgeworth (1894: 160) and Edgeworth (1901: 404) for their views on Lehr.

<sup>68</sup> In Germany for example the previously-mentioned Paul Flaskämper was a crusader for the cause of ROA indices, which he considered the only "logically" tenable indices. For him the only difference between an index and a relative is that the former has an *average* of prices rather than a single price in its numerator and denominator and thus (the conclusion is far from convincing as it denies all aggregation problems) should share all properties with relatives. He even went so far as to deny the relevance and validity of so many simple equations which show that the two approaches are often quite closely related. This is perhaps again a consequence of the then prevalent propensity to philosophize in German statistics (a major exception was L. von Bortkiewicz). We cannot go into more details here concerning the problem with Paul Flaskämper (1928) and the decline of German index theory in that period. It will be the subject of another paper.

<sup>69</sup> With this motivation e.g. Irving Fisher was perhaps the most prominent advocate of the AOR approach. "My book is devoted entirely to averages of ratios" Fisher (1923: 743). Some proponents of AOR were quick in ridiculing their ROA opponents: "Actually a price index number is not properly thought of as a ratio of average prices. An average of the number of horses and the number of apples has little, if any, meaning. Neither has an average of the price of horses and the price of apples. An index number should be an average of ratios, not a ratio of averages." Cowden and Pfouts (1952: 92).

It may seem strange that the alternative of AOR vs. ROA stirred up so many controversies, since many index functions can be written in both ways. We already demonstrated this with the translation of  $P^D$  (ROA) into  $P^C$  (AOR) and vice versa simply by introducing weights. It seems to be very well known that the two formulas exist for  $P^L$

$$P_{0t}^L = \frac{\sum P_t q_0}{\sum P_0 q_0} = \sum \frac{P_t}{P_0} \frac{P_0 q_0}{\sum P_0 q_0} \quad (10)$$

As to Young's index  $P^Y$  in the digression: Mehrhoff also remarked (like v. Bortkiewicz 1927 beforehand) that  $P^Y$  not only has a ROA interpretation (indicated in (8) with weighted means of prices), but also an AOR interpretation, namely  $P_{0t}^Y = \sum \frac{P_t}{P_0} \sqrt{\frac{P_0}{P_t}} / \sum \sqrt{\frac{P_0}{P_t}}$ , and thus is indeed classifiable in Fisher's scheme, which Fisher apparently failed to see.<sup>71</sup> It may be noticed in passing that Young well appreciated this double interpretation of his index  $P^Y$ .<sup>72</sup> This is what L. v. Bortkiewicz (1927: 747) elevated to the rank of a quality indicator in the form of his "*Zweiförmigkeitskriterium*"<sup>73</sup>. It may be viewed as an axiom which (by way of exception, given the nature of the other axioms) directly focuses on "meaningfulness" and "understandability" of a formula.

There are of course indices which allow only one interpretation.  $P^{DR}$  is for example not a weighted mean of relatives for sub-aggregates. Assume  $K$  sub-aggregates and "partial" unit values  $\tilde{p}_{kt}$  and  $\tilde{p}_{k0}$  ( $k = 1, \dots, K$ ). The ratio of unit values  $\tilde{p}_{kt}/\tilde{p}_{k0}$  is not a mean of

price relatives because  $\frac{\tilde{p}_{k1}}{\tilde{p}_{k0}} = \frac{Q_{k0}}{Q_{kt}} \sum_j \frac{P_{kjt}}{P_{kj0}} \left( \frac{P_{kj0} q_{kjt}}{\sum_j P_{kj0} q_{kj0}} \right)$  and the aggregated  $P^{DR}$  is not

simply a weighted mean of the  $\tilde{p}_{kt}/\tilde{p}_{k0}$  ratios because  $P_{0t}^{DR} = \sum_k \frac{\tilde{p}_{kt}}{\tilde{p}_{k0}} \left( \frac{\tilde{p}_{k0} \sigma_{kt}}{\sum_k \tilde{p}_{k0} \sigma_{k0}} \right)$ , where

the  $\sigma$  are quantity shares  $\sigma_{kt} = \sum_j q_{jkt} / \sum_k \sum_j q_{jkt}$  and  $\sigma_{k0}$  defined correspondingly.

Hence there is no AOR interpretation of  $P^{DR}$ ,<sup>74</sup> whereas both of the indices  $P^L$  and  $P^P$ , which Drobisch regarded as special (and inferior) cases of his formula, can be interpreted in both ways, i.e. ROA and AOR. The problem with  $P^L$  is of course the fictitious character of an average price  $\tilde{p}_t^* = \sum P_t \frac{q_0}{\sum q_0}$  and an expenditure  $\sum P_t q_0$  requiring that

recourse has to be made in period  $t$  to quantities  $q_0$  in the past. This brings us to another perennially contentious issue: chain indices.

<sup>70</sup> That indices such as  $P^L$  (and also  $P^P$ ) can be written in both ways had already been seen by Walsh (1901: 428, 539) and Fisher (1911: 365). Many students seem to be unaware of the fact that two such formulas exist for other indices than  $P^L$  (for  $P^P$  for example) as well, and they tend to mystify the fact that prices are multiplied by (absolute) quantities, while price *relatives* are multiplied by expenditure shares. This was also a problem that appeared puzzling and vexing to Flåskämper and other German statisticians of his time.

<sup>71</sup> Fisher's weights were, however, expenditure shares, not square roots of reciprocal prices relatives.

<sup>72</sup> "In a way Professor Fisher is right in holding that all true index numbers are averages of ratios.' But I should prefer to say that all true index numbers are at once averages of ratios and ratios of aggregates." (Young 1923: 359).

<sup>73</sup> For him this "two-way" (or twofold) interpretation had the rank of an axiom or test, just like time reversibility or proportionality. Note that there are examples of indices which allow both interpretations ( $P^L$  and  $P^P$ ), only one (as  $P^{DR}$ ), or none of them (unless in a quite farfetched manner), such as e.g. Fisher's highly esteemed "ideal index"  $P^F$ . Hence index theorists will most definitely disagree on Bortkiewicz's "*Zweiförmigkeit*" (existence of two forms) lest  $P^F$  will be downgraded for its poor performance in this respect.

<sup>74</sup> This also applies to Lehr's index  $P^{LE}$  in the next section.

## 7 Chain indices, Lehr and the ideal of “pure price comparison”

The idea of chain indices arose not only from the conviction that weights need to be continuously updated, but perhaps (as mentioned above) to an even greater extent from the undue embarrassment with contradictory empirical results of time series of index numbers when they referred to different base periods.<sup>75</sup> Chain indices were welcomed as a device to solve the problem of choosing a base (by finding means to be independent of the base), to avoid ambiguities in this respect and to update weights (which  $P^L$  fails to do).<sup>76</sup>

A chain index is defined as a product of indices (“links”), each of which refers to two adjacent periods (as a short sub-interval). To arrive at an index for the total interval, the links are multiplied to form a chain. When the index is transitive, such as for example  $P^{DR}$ , it is clear that the direct index coincides with the chain  $P_{0t}^{DR} = \bar{P}_{0t}^{DR} = P_{01}^{DR} P_{12}^{DR} \dots P_{t-1,t}^{DR}$ . The same applies to Jevons’ unweighted index  $P^L$ .

However, chaining of an index also takes place when the respective index is not transitive, which applies for example to  $P^C$  or  $P^L$ , where therefore as a rule  $P_{0t}^L \neq P_{01}^L P_{12}^L \dots P_{t-1,t}^L$ .<sup>77</sup> Usually much of what is argued in favour of chain indices grows out of a critique of so called “fixed weight” or “fixed base” indices like  $P^L$ . This seems to apply to Lehr too.

It should be noticed that the terms “fixed weight” or “fixed base” are incorrect and should be abandoned. This can easily be seen in the case of a sequence of a “fixed base” *Paasche* price index (see eq. 9a), where the  $q$ -weights are not fixed but instead constantly vary in much the same way as in the case of factors creating a *chain Paasche* price index given by

$$\bar{P}_{0t}^P = \frac{P_1 Q_1}{P_0 Q_1} \cdot \frac{P_2 Q_2}{P_1 Q_2} \cdot \dots \cdot \frac{P_t Q_t}{P_{t-1} Q_t} \neq P_{0t}^P = \frac{P_t Q_t}{P_0 Q_t}. \quad (9b)$$

The correct characterization should be chain index by contrast to direct index, because the alternatives are either to compare 0 to  $t$  indirectly via 0-1, 1-2, ...,  $t-1, t$  or to compare 0 to  $t$  directly.

The property of a chain index that is particularly often found desirable is to always have “realistic” (up-to-date) weights by constantly switching to more recent quantity-weights. However, this is clearly in conflict with making “pure comparisons” by keeping

<sup>75</sup> As mentioned above, such puzzles led many economists (for example Pierson) to a general rejection and ostracism of all sorts of index numbers. In the time period under consideration it was not uncommon to take averages over a *number of years* (e.g. five or even ten) as the base period (“standard”) in order to mitigate potential “resounding” effects of an inappropriate single base year.

<sup>76</sup> Interestingly, until recently most of the arguments in favour of chain indices were advanced very early on and have remained by and large the same until today. As many others, Flux (1907) for example was obsessed with quantity weights being as up-to-date as possible, because fixed weights get “thoroughly out of touch with the facts” (p. 619). Furthermore, chain indices or the “method of year-to-year steps” as he called it, has the advantage of being “not dependent on the location of the starting point” and to facilitate “introducing new articles or dropping old ones” (p. 625). Much the same can be read in Fisher’s writings. Such arguments have continued to dominate all debates of chain indices vs. Laspeyres’ formula ever since. For more details see von der Lippe (2001).

<sup>77</sup> It makes sense to call the left hand side of this inequality the “direct index” because it compares  $t$  directly to 0 without taking into account the intermediate periods. However, it is unfortunately more common to speak of a “fixed base” index as opposed to a chain index.

$q_0$  constant. Chain indices can be viewed as an option in favour of a constant update at the expense of making “pure comparisons”. Reconciling both advantages appears to be insoluble.<sup>78</sup>

Diewert mentioned Marshall and Lehr in the context of chain indices, and von Bortkiewicz wrote that such indices were first suggested by Lehr and then by Alfred Marshall.<sup>79</sup> It may therefore be pertinent to briefly present the index theory of Julius Lehr.

As mentioned above, unlike Laspeyres, Paasche and Drobisch, Lehr did not contribute papers on index numbers to this journal. He developed his somewhat peculiar formula (denoted in the following by  $P^{LE}$ ) in a small pamphlet (Lehr 1885). In this book, however,  $P^{LE}$  covered only a couple of pages, and it can be seen that he multiplied the respective links  $P_{01}^{LE}, P_{12}^{LE}, \dots$  as if this were a matter of course, but he did not say much about the properties of a chain index.

Before going into details of how Lehr justified chaining, we should introduce his formula which is, like  $P^{DR}$ , a typical ROA approach. Central to this index is the fictitious quantity  $g_{i,01}$ , called “*Genußseinheit*” (or “pleasure unit” in the translation of Edgeworth<sup>80</sup>) by Lehr. In the tradition of implicit quantities it is conceived as a reciprocal price level  $1/\check{p}_{i,01}$ . However,  $\check{p}_{i,01}$  depends on explicit (effective) quantities  $q_i$  and combines the prices of two periods. In

$$\check{p}_{i,01} = \frac{1}{g_{i,01}} = \frac{p_{i0}q_{i0} + p_{i1}q_{i1}}{q_{i0} + q_{i1}} = p_{i0} \frac{q_{i0}}{q_{i0} + q_{i1}} + p_{i1} \frac{q_{i1}}{q_{i0} + q_{i1}} \quad (11)$$

we may see a sort of mid-interval price of good  $i$  because averaging takes place over two adjacent periods in time, and not over two goods. This leads to Lehr’s definition of an absolute price level<sup>81</sup>

$$\check{P}_1 = \frac{\sum_{i=1}^n p_{i1}q_{i1}}{\sum_{i=1}^n q_{i1}\check{p}_{i,01}} = \sum p_{i1} \frac{q_{i1}}{\sum q_{i1}\check{p}_{i,01}} \text{ in period 1, and } \check{P}_0 = \frac{\sum p_{i0}q_{i0}}{q_{i0}\check{p}_{i,01}} \quad (12)$$

correspondingly in period 0, so that his index as a ratio of price levels is given by

$$P_{01}^{LE} = \frac{\check{P}_1}{\check{P}_0} = \frac{\sum p_{i1}q_{i1}}{\sum p_{i0}q_{i0}} \cdot \frac{\sum q_{i0}\check{p}_{i,01}}{\sum q_{i1}\check{p}_{i,01}} = \frac{\sum p_{i1}q_{i1}}{\sum p_{i0}q_{i0}} \cdot \frac{S_0}{S_1} = \frac{V_{01}}{Q_{01}^{LE}}, \quad (13)$$

<sup>78</sup> This dilemma was made particularly clear in a paper of Sir George Knibbs. On the one hand we can find a plea for an “unequivocal” index (a “price index of an indefinite or variable basis cannot possibly yield an unequivocal result” and “the whole purpose of a price-index is to reflect the effect of change of price solely...” Knibbs (1924: 46)), and on the other hand a plea for representativeness (what he called “reality”). For Knibbs the conflict between pure price comparison (what he called “definiteness”) and reality is “the crux of the whole matter” (p. 60), a problem of the squaring the circle type. He called a price index which also reflects quantity movement a “confused” as opposed to an unequivocal index, and he was well aware of the drift problem inherent in the chain index method.

<sup>79</sup> Das “Kettensystem”, das “zuerst von Lehr und bald darauf von Marshall in Vorschlag gebracht worden ist” (The “chain system” that “was put forward for consideration first by Lehr and soon thereafter by Marshall”, v. Bortkiewicz (1927: 749)).

<sup>80</sup> In this somewhat eccentric concept he primarily saw an attempt to measure utility on the part of Lehr.

<sup>81</sup> Note that this term (unlike the corresponding term in  $P^{DR}$ ) is a dimensionless ratio of expenditures.

and as this formula is intended to serve as a link  $P_{t-1,t}$  in a chain  $P_{0t} = P_{01}P_{12}...P_{t-1,t}$  the general formula is given by

$$P_{t-1,t}^{LE} = \frac{\check{P}_t}{\check{P}_{t-1}} = \frac{\sum P_{it}Q_{it} / \sum Q_{it}\check{P}_{i,t-1,t}}{\sum P_{i,t-1}Q_{i,t-1} / \sum Q_{i,t-1}\check{P}_{i,t-1,t}} \quad \text{with}$$

$$\check{P}_{i,t-1,t} = \frac{P_{i,t-1}Q_{i,t-1} + P_{it}Q_{it}}{Q_{i,t-1} + Q_{it}}. \quad (13a)$$

The formula (13) looks a bit outlandish<sup>82</sup> and its rightmost variant shows

- that Lehr meant that terms like  $S_1 = \sum q_{i1}\check{p}_{i,01} = \sum q_{i1}(1/g_{i,01})$  and  $S_0 = \sum q_{i0}\check{p}_{i,01}$  represent a sort of total quantity,<sup>83</sup> where (physical) quantities are made commensurable upon dividing by “*Genußseinheiten*”, or (equivalently) upon weighting (multiplying) by mid-interval prices  $p_{t-1,t}$ ,<sup>84</sup> so that we obtain with  $P_{0t}^{LE}$  a measure of how the price of a pleasure unit has changed,<sup>85</sup> and
- how  $P_{0t}^{LE}$  can be viewed as a ratio of a value and a quantity index  $P_{0t}^{LE} = V_{0t}/Q_{0t}^{LE}$  and is thereby comparable to  $P_{0t}^L = V_{0t}/Q_{0t}^P$ ,  $P_{0t}^P = V_{0t}/Q_{0t}^L$ , or also  $P_{0t}^{DR} = V_{0t}/Q_{0t}^D$  so that a ratio of price indices is tantamount to a ratio of quantity indices. In the appendix we will make use of this relationship:  $P_{0t}^{LE}/P_{0t}^L = Q_{0t}^P/Q_{0t}^{LE}$  etc.

In the appendix it also turns out that  $P^{LE}$  may in a way be regarded as standing between  $P^P$  and  $P^L$ . As  $\check{p}_{i,01}$  may be seen as the price in the middle of the interval  $(0, 1)$ , it is intuitively also plausible that  $S_1 = \sum q_{i1}\check{p}_{i,01}$  is in a way in the middle of  $\sum q_{i1}p_0$  and  $\sum q_{i1}p_1$ .<sup>86</sup> It follows also that the results of  $Q_{01}^{LE} = S_1/S_0$  and  $Q_{01}^{ME} = \sum q_1(p_0 + p_1)/\sum q_0(p_0 + p_1)$ , the Marshall-Edgeworth quantity index, might be close together.

Introducing price- ( $\lambda$ ) and quantity ( $\omega$ ) relatives we get

$$\check{p}_{i,01} = \frac{p_{i0}(1 + \lambda_{i1}\omega_{i1})}{1 + \omega_{i1}}. \quad (14)$$

It now can easily be seen that if  $\lambda_{i1} = 1$  (for all  $i$ )  $\check{p}_{i,01} = p_{i0} = p_{i1}$  so that  $\sum q_{i1}\check{p}_{i,01} = \sum q_{i1}p_{i0}$ , which Lehr had already noticed,<sup>87</sup> and  $P_{01}^{LE} = P_{01}^L = P_{01}^P$ , in which case, however, all these three indices amount to unity. This does not apply, however, for example to  $P_{01}^{DR}$ . Hence Lehr's index meets identity (unlike Drobisch's index) but not

<sup>82</sup> Edgeworth called it “cumbrous”, which may explain why it was hardly used and never had renowned supporters. Lehr was mentioned by Edgeworth, Walsh and Fisher most rarely in Germany (an exception is Lexis 1886).

<sup>83</sup> For Lehr  $S_0$  and  $S_1$  is the sales value of a number of pleasure units (“*verkaufte Genußseinheiten*” Lehr 1885: 39). Thus the term is expressed in currency units and comparable across all kinds of good. Hence by contrast to  $\sum q_0$  and  $\sum q_t$  in  $P^{DR}$  there is no need to express all quantities uniformly in hundredweights (cwt) in  $P^{LE}$ . However,  $S_0$  and  $S_1$  may also be regarded as expenditures (values) and  $\sum q_0$  and  $\sum q_t$  are clearly much more understandable.

<sup>84</sup> They may be viewed as “comparable” or “standardized” quantities and therefore much better than Drobisch's simple sums of hundredweights  $\sum q_{it}$ . Also the  $Q^{LE}$  is rightly seen as a sort of quantity index.

<sup>85</sup> “*erhalten wir das Maß, in welchem sich der Preis der Genußseinheit geändert hat*” Lehr (1885: 39).

<sup>86</sup> The same should apply to  $S_0$  in relation to  $\sum q_0p_0$  and  $\sum q_0p_1$ . Note that in this interpretation  $S_1$  (and  $S_0$ ) is “acting” as an expenditure rather than a “quantity”.

<sup>87</sup> Lehr (1885: 40). He also saw that  $q_{1t} = q_{2t} = \dots = q_{nt}$  implies that his index  $P^{LE}$  reduces to Dutot's index  $P^D$ .



proportionality.<sup>88</sup>  $\check{p}_{i,01} = p_{i0}$  implies  $p_{i1} = p_{i0}$  and therefore  $\lambda_{i1} = 1$ . Note that this result regarding “identity” does not mean that Lehr’s index already sufficiently complies with the principle of pure price comparison (i. e. to reflect price changes between 0 and t only) which seems to have been in Laspeyres’ thinking the most important criterion a good price index should fulfill. Assuming identity of all prices in 0 and 3 (and also all quantities  $q_{i0} = q_{i3}$ ),<sup>89</sup> multiplying links  $P_{01}^{LE} P_{12}^{LE} P_{23}^{LE}$  will in general not result in unity, that is  $\bar{P}_{03}^{LE} = P_{01}^{LE} P_{12}^{LE} P_{23}^{LE} \neq 1$  although each link  $P_{t-1,t}$  as such satisfies identity. As is well-known, this not only applies to the chaining of  $P^{LE}$ , but also to  $P^L$  and all sorts of chain indices, which thus violate pure price comparison in the sense of reflecting only the difference between two price vectors  $p_t$  and  $p_0$ . Instead chain indices are also affected by prices and quantities of all intermediate periods.

When on the other hand quantities remain constant, that is  $\omega_{i1} = 1$  (for all i),  $\check{p}_{i,01} = \frac{1}{2}(p_{i0} + p_{it})$  and  $Q_{0t}^{LE}$  reduces to  $Q_{0t}^{ME}$  and is unity<sup>90</sup>, as is  $Q_{0t}^L$ ,  $Q_{0t}^P$ , and also  $Q_{0t}^D$ . Note that  $Q_{t-1,t}^{LE} = S_t/S_{t-1}$  compares quantities multiplied by the same prices  $\check{p}_{i,t-1,t}$  and thus can be interpreted as a weighted mean of quantity relatives with weights  $q_{i0}\check{p}_{i,01}/\sum q_{i0}\check{p}_{i,01}$ . Hence unlike  $P_{t-1,t}^{LE}$ , the quantity index  $Q_{t-1,t}^{LE}$  satisfies the mean value property.

According to (11),  $\check{p}_{i,01}$  is a weighted mean  $\check{p}_{i,01} = ap_{i0} + (1-a)p_{it}$  of  $p_{i0}$  and  $p_{i1}$ , assigning a greater weight  $\frac{1-a}{a} = \frac{q_{i1}}{q_{i0}} > 1$  to the new price  $p_{i1}$  then to the old one ( $p_{i0}$ ) of a commodity when its quantity went up  $q_{i1} > q_{i0}$  (that is,  $\omega_{i1} > 1$ ) such that  $\sum q_{i1}\check{p}_{i,01}$  tends to  $\sum q_{i1}p_{i1}$  (Paasche) when quantities rise, and to  $\sum q_{i1}p_{i0}$  (Laspeyres) when they basically go down.

As to the difference between weights  $\check{p}_{i,01}$  and  $\check{p}_{i,01}$  in the sums  $S_1$  and  $S_0$  in  $Q^{LE}$  we see that they are related as follows  $\frac{\check{p}_{j,01}}{\check{p}_{i,01}} = \frac{g_{j,01}}{g_{i,01}} = \frac{p_{j0}}{p_{i0}} \frac{1 + \omega_j \lambda_j}{1 + \omega_i \lambda_i} \frac{1 + \omega_i}{1 + \omega_j}$ .<sup>91</sup> Now  $\frac{q_{j1}\check{p}_{j,01}}{q_{i1}\check{p}_{i,01}}$  in  $Q^{LE}$  may be compared to  $q_{j1}/q_{i1}$  in  $Q^D$  or to  $p_{j0}q_{j1}/p_{i0}q_{i1}$  in  $Q^L$ .

We should refrain from going more into the details of the underlying rationale of  $P^{LE}$  and the properties of the index. More importantly, however, it should be noted that Lehr’s index – unlike  $P_{0t}^{DR}$  – cannot be chained (notwithstanding Lehr had no qualms with multiplying his index numbers) because

$$P_{02}^{LE} = V_{02} \frac{\sum q_{i0}\check{p}_{i,02}}{\sum q_{i2}\check{p}_{i,02}} \neq P_{01}^{LE} P_{12}^{LE} = V_{02} \frac{\sum q_{i0}\check{p}_{i,01}}{\sum q_{i1}\check{p}_{i,01}} \frac{\sum q_{i1}\check{p}_{i,12}}{\sum q_{i2}\check{p}_{i,12}} \quad (15)$$

<sup>88</sup> If proportionality ( $p_{it} = \lambda p_{i0}$ ) then also identity (the special case where  $\lambda = 1$ ), but the converse is not true. If identity is violated, so is proportionality. Hence  $P^{DR}$  also fails proportionality.

<sup>89</sup> With  $P^{LE}$  it is not sufficient to multiply only two links under the assumption of price and quantity vectors  $p_0 = p_2$  and  $q_0 = q_2$  because the *Genusseinheiten*  $g_i$  relate two adjacent periods to one another (in a chained Laspeyres index there are no  $g_i$  terms, so a chain of two links only suffices to demonstrate that identity may be violated).

<sup>90</sup> Put differently: the quantity index of Lehr meets identity (but not proportionality) in the quantities.

<sup>91</sup> From this follows: When prices remain constant  $\lambda_i = \lambda_j = 1$  good i represents more *Genusseinheiten* than good j when its price is lower ( $p_{i0} < p_{j0}$ ). This is in line with the then widely held opinion that the weight of a price should be inversely proportional to its base period price.

where  $p_{12}^{LE} = \frac{\sum p_{i2}q_{i2}}{\sum p_{i1}q_{i1}} \cdot \frac{\sum q_{i1}\check{p}_{i,12}}{\sum q_{i2}\check{p}_{i,12}}$ , and  $\check{p}_{i,12} = \frac{p_{i1}q_{i1} + p_{i2}q_{i2}}{q_{i1} + q_{i2}}$  (for  $p_{01}^{LE}$  (13) applies), and in general we get  $p_{0t}^{LE} \neq p_{01}^{LE}p_{12}^{LE}...p_{t-1,t}^{LE}$  as opposed to  $p_{0t}^{DR} = p_{01}^{DR}p_{12}^{DR}...p_{t-1,t}^{DR}$ .

A final remark to Lehr's ideas on chaining may be added. We could not see that he advanced any noteworthy arguments in order to advocate for chain indices. The only advantage of his approach that he pointed out was the frequently mentioned ease in dealing with the emergence of new goods and disappearance of old goods (or "entry and withdrawal").<sup>92</sup> He argued that abrupt transitions causing extreme discontinuities will be unlikely. And even if there were such abnormal events he considered his method superior to the then possibly widely used strategy to simply cancel outliers in time series and to take averages over longer intervals in time.<sup>93</sup>

In summary his method consisted of:

- taking all observations (in the intermediate periods) into account, not only the endpoints 0 and t of the interval,<sup>94</sup>
- multiplying  $P_{t-1,t}$  indices ("links"), to form a chain (that is the chain index method), and
- estimating trends in the time series.

He did not study properties of chain indices and he was not very specific concerning the pros and cons of this method.<sup>95</sup> Such things were not so much in his focus. Instead the greater part of his book is devoted to various least squares estimations of trends in time series of prices (and not to his index formula nor to the rationale of chaining).

## 8 Some concluding remarks

Given the length of the paper it seems advisable to only very briefly point out some results:

1. To begin with Lehr, it is slightly ironic, and certainly widely unknown, that Germany was one of the first countries, if not the first country, where the idea of chain indices emerged. It is well-known that this country was particularly unhappy with the general move to chain indices in official statistics all over the world in the late 1990s. Chain indices were widely disapproved of as being irreconcilable with pure price comparisons, and were viewed with suspicion.

<sup>92</sup> "Tritt nun ein neues Gut ein, ... so kann dasselbe einfach in der oben mitgetheilten Formel in Rechnung gestellt werden. Ebenso ist zu verfahren, wenn ein bislang begehrtes Gut fortan ... nicht mehr in den Handel gebracht wird" Lehr (1885: 46). This reads as follows: When a new good emerges ... it can simply be accounted for in the above mentioned formula. One may proceed likewise when a hitherto demanded good is henceforth ... no longer on the marketplace.

<sup>93</sup> He criticized this method in which he viewed an attempt to detect a sort of trend by excluding extraordinary observations in an otherwise smooth time series. For him the problem was: to exclude the abnormal (e.g. outliers) requires to know what is "normal", and finding out exactly this is the purpose of smoothing (p. 48).

<sup>94</sup> Lehr set great store by taking all price and quantity observations of a time interval into account. Here he vigorously disagreed with Paasche. Also Lehr paid a great deal of attention to the estimation of a linear or exponential trend with the method of least squares. Compared to the index formula, a much greater part of his book is devoted to precisely this task.

<sup>95</sup> He seems to have seen no more (or different) advantages of chaining than other authors of the time also did.

2. It is possibly not a coincidence that all four of the authors presented here only temporarily worked on indices, that they had no idea of the increasing importance this topic would gain in the future and that they underrated the relevance of their index formula. They were occupied with many other research interests and price indices were not central to them. None of them dealt with index numbers for many years or even decades,<sup>96</sup> unlike for example Edgeworth, where around 16 papers on index numbers authored by him are known of, spanning the time period from 1883 to 1925. Furthermore, index problems did not seem to attract many discussants. The situation was fundamentally different in monetary theory where many economists contributed papers and were involved in theoretical disputes. Perhaps misconceptions in this field and the inability to recognize that prices pose intellectually challenging measurement problems could also be responsible for the lack of interest in index numbers.

3. On the other hand, possibly as an after-effect of historicism in economics, much effort was spent on meticulously compiled statistics in laborious monographs covering phenomena of regionally and temporally only rather limited relevance. To give one example only, Paasche extensively studied prices of manorial estates of Prussian knights and other nobles. Paasche might have considered such works, and he made quite a few of this kind, as no less important than his formula. Laspeyres complained (in Laspeyres 1875) that he had to spend some four hours every day only on performing mechanical and dull computations. Much of the work was also devoted to the procurement of statistical data, so there was not much room left for applying one's own index formula, because of the time-consuming preparation of detailed tables. As already mentioned, official statistics of the time did not yet provide statistical data to the extent we are used to today.

4. An astonishing observation for me was that neither Paasche nor Laspeyres were very clear and resolute as far as the specific features of their formulas are concerned. Laspeyres' arguments in favour of  $q_0$  were not very well substantiated (the same applies to Paasche with his choice of  $q_t$ ). It is not quite clear whether he preferred  $q_0$  to  $q_t$  on theoretical grounds, or because  $q_t$  might be less readily available than  $q_0$ . However, there was an abundance of other statements on methodological issues that may allow us to infer what motivated him to his formula. Surprisingly, he also still adhered to the unweighted Carli index many years after having developed his own formula.

5. In a similar vein, Lehr was most non-committal concerning the justification and effects of the operation of chaining. This is all the more astonishing as he was quite mathematically oriented for his time. In this situation it should have been an interesting exercise for him to do more in the unveiling of properties of his slightly peculiar formula and of chain indices in general.

6. Laspeyres' emphasis on "pure" comparisons had a lasting effect. This was to become distinctive of typical German index theorizing, but also by degrees more of a burden. In a good way it prevented overly "formal" considerations as an end in themselves, but in a bad way it carried on into futile sophistry about the logic of comparability, which characterized German economic statistics in the 1920s and 1930s. One of those ultimately useless topics discussed above is for example the alternative AOR or ROA (also given that many index functions can be written in both ways).

<sup>96</sup> In the case of Lehr it should be taken into account that he died deplorably early (shortly before his 49<sup>th</sup> birthday).

## Appendix

### Relationships between price indices

#### a) Drobisch, Laspeyres, and Paasche

To show how  $P^{DR}$  is related to  $P^L$  and  $P^P$  the theorem of Ladislaus von Bortkiewicz on linear indices will be used.<sup>97</sup> We also make use of the equations relating the indices to the value index, viz.  $P^{DR}_{0t} = V_{0t}/Q^D_{0t}$ ,  $P^L_{0t} = V_{0t}/Q^P_{0t}$  and  $P^P_{0t} = V_{0t}/Q^L_{0t}$ . The theorem then yields the following bias formulas

$$\frac{P^{DR}_{0t}}{P^L_{0t}} - 1 = \frac{Q^P_{0t}}{Q^D_{0t}} - 1 = \frac{\sum \left( \frac{q_t}{q_0} - Q^D_{0t} \right) (p_t - \tilde{p}_t^*) w}{Q^D_{0t} \tilde{p}_t^*} \quad (16)$$

with weights  $w = \frac{q_0}{\sum q_0}$ ,  $Q^D_{0t} = \sum \frac{q_t}{q_0} w$ , and  $\tilde{p}_t^* = \sum p_t w$ , and

$$\frac{P^{DR}_{0t}}{P^P_{0t}} - 1 = \frac{Q^L_{0t}}{Q^D_{0t}} - 1 = \frac{\sum \left( \frac{q_t}{q_0} - Q^D_{0t} \right) (p_0 - \tilde{p}_0) w}{Q^D_{0t} \tilde{p}_0} \quad \text{where } \tilde{p}_0 = \sum p_0 w. \quad (17)$$

The equations are closely related to equations in Diewert and von der Lippe 2010,<sup>98</sup> and they seem to make sense: When changes in quantities correlate negatively with the price level in  $t$  we expect the Laspeyres index  $P^L$  to exceed Drobisch's index  $P^{DR}$ , which according to Laspeyres (1871: 307) seems to consistently be the case.<sup>99</sup> For the bias of  $P^{DR}$  relative to  $P^P$ , what matters is the price level of the base period.

#### b) Drobisch and Lehr

A similar equation with quantity weights  $w = q_0/\sum q_0$  can be found with Bortkiewicz's theorem for the relationship between  $X_1 = P^{DR}$  and  $X_0 = P^{LE}$

$$\frac{P^{DR}_{0t}}{P^{LE}_{0t}} - 1 = \frac{Q^{LE}_{0t}}{Q^D_{0t}} - 1 = \frac{\sum \left( \frac{q_t}{q_0} - Q^D_{0t} \right) (\tilde{p}_{01} - \bar{Y}) w}{Q^D_{0t} \bar{Y}} \quad (18)$$

where  $\bar{Y} = Y_0 = \sum p_{01} w$  compared to  $\tilde{p}_t^* = \sum p_t w$  and  $\tilde{p}_0 = \sum p_0 w$  in (16) and (17) respectively. Note that the prices  $\tilde{p}_{01}$  are quantity weighted averages between prices

<sup>97</sup> See von der Lippe (2007: 198) for this theorem. We follow also the notation with  $X_0$ ,  $Y_0$ , and  $X_1$  used there. I only later realized that v. Bortkiewicz already presented his theorem in a form which I (and many other authors) assumed to be a generalization.

<sup>98</sup> Eq. (17) is equivalent to eq (20) in Diewert and v. d. Lippe, and (16) is basically the same as (29) and (30), where, however, reciprocal quantity relatives, i. e.  $r = q_0/q_t$ , are studied and all the covariance equations were derived without reference to v. Bortkiewicz. This shows that there is in general more than only one way to describe the relationship between any two linear indices as a function of a covariance. This can be seen here for example with the two equations (19) and (19a), both derived with Bortkiewicz's theorem, or with the fact that we may express  $P^{LE}/P^L - 1$  on the one hand and  $P^L/P^{LE} - 1$  (in (16a)) on the other hand using different covariances (the same applies to  $P^{LE}/P^P$  relative to  $P^P/P^{LE}$ ).

<sup>99</sup> The result also resembles the well-known fact (found by von Bortkiewicz) that  $P^L > P^P$  when price relatives and quantity relatives are negatively correlated. However, weights are then expenditure shares  $p_0 q_0 / \sum p_0 q_0$  rather than quantity shares  $q_0 / \sum q_0$  as above.

$p_0$  and  $p_1$ , and so in a way are mid-interval prices. The structure of the three formulas (16) through (18) is thus quite similar.

### c) Lehr, Laspeyres, and Paasche

It appears desirable to find an expression analogous to (16) with  $P^{LE}$  instead of  $P^{DR}$ , that

is  $\frac{P_{0t}^{LE}}{P_{0t}^L} - 1 = \frac{Q_{0t}^P}{Q_{0t}^{LE}} - 1$ , however, one way to construct the formula for this situation ac-

cording to Bortkiewicz's theorem leads to expenditure weights depending on prices  $\check{p}_{01}$  (and in the case of  $P^{LE}/P^P$  also to a sort of reciprocal price relatives  $p_0/\check{p}_{01}$ ) with no plau-

sible interpretation,  $\frac{P_{0t}^{LE}}{P_{0t}^L} - 1 = \frac{\sum \left( \frac{q_t}{q_0} - Q_{0t}^{LE} \right) \left( \frac{p_1}{\check{p}_{01}} - \bar{Y}_1 \right) \check{w}}{Q_{0t}^{LE} \bar{Y}}$  with weights

$$\check{w} = \frac{\check{p}_{01} q_0}{\sum \check{p}_{01} q_0} \text{ and } \bar{Y}_1 = \frac{\sum p_1 q_0}{\sum \check{p}_{01} q_0}, \text{ and}$$

$$\frac{P_{01}^{LE}}{P_{01}^P} - 1 = \frac{Q_{01}^L}{Q_{01}^{LE}} - 1 = \frac{\sum \left( \frac{q_1}{q_0} - Q_{01}^{LE} \right) (\check{p}_{01} - \bar{Y}_2) \check{w}}{Q_{01}^{LE} \bar{Y}} \text{ where } \bar{Y}_2 = \frac{\sum p_0 q_0}{\sum \check{p}_{01} q_0}$$

is a kind of reciprocal price index. It appears more reasonable to study the relation  $P^L/P^{LE} - 1$  (and accordingly  $P^P/P^{LE} - 1$ ) instead of  $P^{LE}/P^L - 1$  (analogous to (16)). This will at least in the case of  $P^P/P^{LE} - 1$  yield more meaningful weights representing now empirical expenditure shares  $p_0 q_0 / \sum p_0 q_0$ , which in turn allows a comparison of the result with the well known formula for the bias  $P^P/P^L - 1$ .

$$\frac{P_{01}^L}{P_{01}^{LE}} - 1 = \frac{Q_{01}^{LE}}{Q_{01}^P} - 1 = \frac{1}{Q_{01}^P \bar{Y}} \sum \left( \frac{q_1}{q_0} - Q_{01}^P \right) \left( \frac{\check{p}_{01}}{p_1} - \bar{Y}_3 \right) \frac{p_1 q_0}{\sum p_1 q_0} \quad (16a)$$

where  $\bar{Y}_3 = \frac{\sum \check{p}_{01} q_0}{\sum p_1 q_0} = \frac{1}{\bar{Y}_1}$  and

$$\frac{P_{01}^P}{P_{01}^{LE}} - 1 = \frac{Q_{01}^{LE}}{Q_{01}^L} - 1 = \frac{1}{Q_{01}^L \bar{Y}_4} \sum \left( \frac{q_1}{q_0} - Q_{01}^L \right) \left( \frac{\check{p}_{01}}{p_0} - \bar{Y}_4 \right) \frac{p_0 q_0}{\sum p_0 q_0} \quad (17a)$$

where  $\bar{Y}_4 = \frac{\sum \check{p}_{01} q_0}{\sum p_0 q_0} = \frac{1}{\bar{Y}_2}$ , and this equation can be compared to the well-known equation

$$\frac{P_{01}^P}{P_{01}^L} - 1 = \frac{1}{Q_{01}^L P_{01}^L} \sum \left( \frac{q_1}{q_0} - Q_{01}^L \right) \left( \frac{p_1}{p_0} - P_{01}^L \right) \frac{p_0 q_0}{\sum p_0 q_0} \quad (19)$$

and a formula analogous to (16a) can also be established by the lesser-known equation

$$\frac{P_{01}^L}{P_{01}^P} - 1 = \frac{1}{Q_{01}^P P_{01}^P} \sum \left( \frac{q_1}{q_0} - Q_{01}^P \right) \left( \frac{p_0}{p_1} - \frac{1}{P_{01}^P} \right) \frac{p_0 q_1}{\sum p_0 q_1}, \quad (19a)$$

so that we may compare (17a) to (19), or  $\bar{Y}_4$  to  $P_{01}^L$ , or (a bit less impressive because of different weights)  $\bar{Y}_3 = \frac{\sum \check{p}_{01}q_0}{\sum p_1q_0} = \frac{1}{\bar{Y}_1}$  to  $\frac{\sum p_0q_0}{\sum p_1q_0} = \frac{1}{P_{01}^L}$  in (16a) and (19a).

Moreover it should be borne in mind that the terms  $\frac{\check{p}_{01}}{p_0} = \frac{q_0}{q_0 + q_1} + \frac{q_1}{q_0 + q_1} \cdot \frac{p_1}{p_0}$  in (17a) are simply linear transformations of the price relatives  $p_1/p_0$  and the structure of  $\bar{Y}_4$  is similar to that of  $P_{01}^L$ . Likewise  $\frac{\check{p}_{01}}{p_1} = \frac{q_1}{q_0 + q_1} + \frac{q_0}{q_0 + q_1} \cdot \left(\frac{p_1}{p_0}\right)^{-1}$  in (16a) may be regarded as linear transformations of *reciprocal* price relatives and the corresponding indices  $\bar{Y}_3$  in (16a) and  $1/P_{01}^L$  in (19a) as reciprocal price indices.

As is well-known, according to (19) we expect  $P^L$  to exceed  $P^P$  price relatives and quantity relatives are negatively correlated. In this case linear transformations  $\check{p}_{01}/p_0$  of price relatives will also correlate negatively with quantity relatives, so that we expect  $P^P < P^{LE}$  just like  $P^P < P^L$ . We may also conclude that the negative correlation between  $p_1/p_0$  or  $\check{p}_{01}/p_0$  and the  $q_1/q_0$  amounts to a positive correlation between the reciprocal (transformed) price relatives and the quantity relatives so that we will have  $P^L > P^{LE}$  just like  $P^L > P^P$ . Hence there are good reasons to assume that Lehr's price index ( $P^{LE}$ ) lies within the bounds of Paasche ( $P^P$ ) and Laspeyres ( $P^L$ ), such that  $P^P < P^{LE} < P^L$  (or, less likely,  $P^P > P^{LE} > P^L$ ).

## References<sup>100</sup>

- Balk, B. M. (2005), Price Indexes for Elementary Aggregates: The Sampling Approach. *Journal of Official Statistics* 21: 675-699.
- Balk, B. M. (2008), Price and Quantity Index Numbers. Cambridge, New York etc. (Cambridge University Press).
- British Association for the Advancement of Science (1902), Memorandum on the Construction of Index Numbers of Prices. Reprinted in *Journal of the American Statistical Association* 9 (Mar. 1904): 21-37.
- Carruthers, A. G., D. J. Sellwood, P. W. Ward (1980), Recent Developments in the Retail Prices Index. *The Statistician* 29: 1-32.
- Cooley, C. H. (1893), Observations on the Measure of Change. *Journal of the American Statistical Association* 3: 285-292.
- Cowden, D. J., R. W. Pfouts (1952), Index Numbers and Weighted Means. *Southern Economic Journal* 9: 90-95.
- Dalen, J. (1992), Computing Elementary Aggregates in the Swedish Consumer Price Index. *Journal of Official Statistics* 8: 129-147.
- Diewert, W. E. (1976), Exact and Superlative Index Numbers. *Journal of Econometrics* 4: 114-145.
- Diewert, W. E. (1987), Laspeyres, Ernst Louis Etienne. Pp. 133-134 in: J. Eatwell, M. Milgate, P. Newman (eds), *The New Palgrave: A Dictionary of Economics*. Vol. L. London.
- Diewert, W. E. (1993), The Early History of Price Index Research. Chapter 2 of *Essays in Index Number Theory* (W. E. Diewert and A. Nakamura, eds., Amsterdam, Vol. 1.)
- Diewert, W. E., P. von der Lippe (2010), Notes on Unit Value Bias. *Jahrbücher für Nationalökonomie und Statistik* 230: 690-708.

<sup>100</sup> We refer to the present journal (*Journal of Economics and Statistics*) as "*Jahrbücher für Nationalökonomie und Statistik*" because we quoted a number of papers which appeared in this (German) journal in the 19<sup>th</sup> and early 20<sup>th</sup> century, and the name of the journal was German for most of the time since its foundation in 1863.

- Drobisch, M. W. (1871a), Ueber die Berechnung der Veränderungen der Waarenpreise und des Geldwerths. *Jahrbücher für Nationalökonomie und Statistik* 16: 143-156.
- Drobisch, M. W. (1871b), Ueber einige Einwürfe gegen die in diesen Jahrbüchern veröffentlichte neue Methode, die Veränderungen der Waarenpreise und des Geldwerthes zu berechnen. *Jahrbücher für Nationalökonomie und Statistik* 16: 416-427.
- Edgeworth, F. Y. (1894), Recent Writings on Index Numbers. *Economic Journal* 4: 158-165.
- Edgeworth, F. Y. (1901), Mr. Walsh on the Measurement of General Exchange Value. *Economic Journal* 11: 404-416.
- Edgeworth, F. Y. (1918), The Doctrine of Index-Numbers According to Professor Wesley Mitchell. *Economic Journal* 28: 176-197.
- Fisher, I. (1911), *The Purchasing Power of Money*. New York
- Fisher, I. (1921), The Best Form of Index Number. *Journal of the American Statistical Association* 17: 533-537.
- Fisher, I. (1922/1927), *The Making of Index Numbers*. Boston, New York (1<sup>st</sup> edition 1922, 2<sup>nd</sup> 1927).
- Fisher, I. (1923), Professor Young on Index Numbers. *Quarterly Journal of Economics* 37: 742-755.
- Flaskämper, P. (1928), *Theorie der Indexzahlen*. Beitrag zur Logik des sozialen Vergleichs, Berlin.
- Flux, A. W. (1907), Modes of Constructing Index Numbers. *Quarterly Journal of Economics* 21: 613-631.
- Held, A. (1871), Noch einmal über den Preis des Geldes. Ein Beitrag zur Münzfrage. *Jahrbücher für Nationalökonomie und Statistik* 16: 315-340.
- Ivancic, L., W. E. Diewert, K. J. Fox (2011), Scanner Data, Time Aggregation and the Construction of Price Indexes. *Journal of Econometrics* 161: 24-35.
- Jastram, R. W. (1951), Willard Phillips, A Predecessor of Paasche in Index Number Formulation. *Journal of the American Statistical Association* 46: 124-126.
- Jevons, W. St. (1865), On the Variation of Prices and the Value of Currency since 1728. *Journal of the Royal Statistical Society* 28: 294-320.
- Knibbs, G. H. (1924), The nature of an unequivocal price-index and quantity-index. *Journal of the American Statistical Association* 19: 42-60, 196-205.
- Konüs, A. A. (1924), The Problem of the True Index of the Cost of Living. Translated in *Econometrica* 7: 10-29 (1939)
- Kuhn, T. S. (1996), *The Structure of Scientific Revolutions*. 3<sup>rd</sup> edition, Chicago and London.
- Laspeyres, E. (1864), *Hamburger Waarenpreise 1851-1863 und die californisch-australischen Goldentdeckungen seit 1848*. Ein Beitrag zur Lehre von der Geldentwerthung. *Jahrbücher für Nationalökonomie und Statistik* 3: 81-118, 209-236.
- Laspeyres, E. (1871), Die Berechnung einer mittleren Warenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik*. 16: 296-314.
- Laspeyres, E. (1875), *Die Kathedersocialisten und die statistischen Congresse*. Gedanken zur Begründung einer nationalökonomischen Statistik und einer statistischen Nationalökonomie. Berlin.
- Laspeyres, E. (1883), Die Bewegung der Warenpreise in der zweiten Hälfte des 19. Jahrhunderts. *Meyers Konversationslexikon* Bd. 20, 3. Aufl. (Jahres-Supplement 1882/83), Leipzig: 795-803.
- Laspeyres, E. (1901), Statistische Untersuchung zur Frage der Steuerüberwälzung, geführt an der Geschichte der preussischen Mahl- und Schlachtsteuer. *Finanz-Archiv* 18: 46-282.
- Lehr, J. (1885), Beiträge zur Statistik der Preise, insbesondere des Geldes und des Holzes. Frankfurt/M.
- Lexis, W. (1886), *Neuere Schriften über Edelmetalle, Geld und Preise*. *Jahrbücher für Nationalökonomie und Statistik* 47: 96-121.
- Marshall, A. (1887), Remedies for Fluctuations of General Prices. *Contemporary Review* 51: 355-375.
- Oker, C. W. (1896), The Fallacy of Index-Numbers. *Journal of Political Economy* 4: 515-522.



- Paasche, H. (1874), Ueber die Preisentwicklung der letzten Jahre nach den Hamburger Börsennotierungen. *Jahrbücher für Nationalökonomie und Statistik* 23: 168-179.
- Paasche, H. (1878), Studien über die Natur der Geldentwertung und ihre praktische Bedeutung in den letzten Jahrzehnten. Aufgrund Statistischen Datenmaterials, Jena.
- Padan, R. S. (1900), Prices and Index Numbers. *Journal of Political Economy* 8: 171-202.
- Pierson, N. G. (1895), Index numbers and appreciation of gold. *Economic Journal* 5: 329-335.
- Pierson, N. G. (1896), Further Considerations on Index-Numbers. *Economic Journal* 6: 127-131.
- Rinne, H. (1981), Ernst Louis Etienne Laspeyres 1834 – 1913. *Jahrbücher für Nationalökonomie und Statistik* 196: 194-216.
- Roberts, H. (2000), Laspeyres and his index. Unpublished paper prepared for the European Conference on the History of Economics in Rotterdam (available in the Internet).
- Segnitz, E. (1870), Über die Berechnung der sogenannten Mittel, sowie deren Anwendung in der Statistik und anderen Erfahrungswissenschaften. *Jahrbücher für Nationalökonomie und Statistik* 14: 183-195.
- van der Borgh, R. (1882), Die Preisentwicklung während der letzten Dezennien nach den Hamburger Börsennotierungen. *Jahrbücher für Nationalökonomie und Statistik* 39: 177-185.
- von Auer, L. (2010), Drobisch's Legacy to Price Statistics. *Jahrbücher für Nationalökonomie und Statistik* 230: 673-689.
- von Bortkiewicz, L. (1927), Der gegenwärtige Stand des Problems der Geldwertmessung. *Handwörterbuch der Staatswissenschaften*, 4. Aufl. Bd. IV: 743-752.
- von Bortkiewicz, L. (1932), Die Kaufkraft des Geldes und ihre Messung. *Nordic Statistical Journal* 4: 1-68.
- von der Lippe, Peter (2001), Chain Indices. A Study in Index Theory. *Spectrum of Federal Statistics Vol 16*, Statistisches Bundesamt (editor).
- von der Lippe, Peter (2005), Das Ideal des "reinen Preisvergleichs". *Jahrbücher für Nationalökonomie und Statistik* 225: 499-509.
- von der Lippe, Peter (2007), *Index Theory and Price Statistics*. Frankfurt.
- Walsh, C. M. (1901), *The measurement of general exchange-value*. New York.
- Walsh, C. M. (1921), Discussion Fisher, I. (1921). *Journal of the American Statistical Association* 17: 537-544.
- Winkler, O. (2009), *Interpreting Economic and Social Data. A Foundation of Descriptive Statistics* (Springer).
- Young, A. (1923), Fisher's "The Making of Index Numbers". *Quarterly Journal of Economics* 37: 342-364.

Prof. Dr. Peter Michael von der Lippe, Universität Duisburg-Essen Campus Duisburg, Mercator School of Management, Lotharstraße 65, 47057 Duisburg, Germany.  
peter@von-der-lippe.org

# From the Zollverein to the Economics of Regionalism

Switgard Feuerstein\*

University of Heidelberg and GEP, University of Nottingham

JEL F15; N93; B00

Zollverein; customs union; economics of regionalism; European integration; preferential trade; sequencing.

Received: 05.02.2012

Revision received: 12.11.2012

Accepted: 25.01.2013

---

## Summary

The paper is a contribution to the special issue “150 years Journal of Economics and Statistics”. The starting point is a series of articles on the nature and the history of the German Zollverein by Gustav Fischer in the early volumes of the Journal. Both the perception of the Zollverein and the economics of preferential trade are traced over time. The focus is on the analogies between the Zollverein and the European Community and on the expansion of the Zollverein in view of the modern theory of sequencing in forming a trade bloc.

## 1 Introduction

When Viner (1950) wrote his seminal study *The customs union issue*, the topic was widely discussed and there were various plans for economic integration, but no notable customs union existed at that time. The main historic example of a customs union of sovereign states was the German Zollverein, a union of most German states realising free trade among each other and having a common tariff system applied to imports from outside countries from 1834–1871<sup>1</sup>. The early volumes of the *Jahrbücher für Nationalökonomie und Statistik* contain a series of articles on the nature of a customs union based on the history of the German Zollverein authored by Gustav Fischer (1864, 1866, 1867).

In contrast to the situation in 1950, the number of customs unions and free trade areas has recently exploded, and nearly 300 of such preferential trade agreements (PTAs) are currently in force and many more are negotiated or in the stage of planning (WTO 2011). Accordingly, research on preferential trade has increased, and the WTO World Trade Report 2011 focuses on PTAs as a special theme. The increase of PTAs is often referred

---

\* I thank two anonymous referees for helpful comments which have improved the paper. Moreover, I am grateful to Marcus Muggenburg of the Archive of the Friedrich Schiller University Jena for providing me with information on Gustav Fischer.

<sup>1</sup> The Zollverein was reformed in 1866/67, and after German unification in 1871, the German Reich was in charge of trade policy, although the Zollverein treaties formally remained in force.

to as regionalism, a term that is common but not precise, as countries forming a PTA need not necessarily be geographically close to each other.

This contribution to the anniversary issue of the *Jahrbücher* aims at linking Fischer's (1864, 1866, 1867) arguments developed in view of the Zollverein to the modern discussion of preferential trade agreements. After introducing Gustav Fischer and his articles in section 2, both the perception of the Zollverein and the literature on the economics of preferential trade are traced over time. Since there exist excellent surveys on the literature of regionalism since the seminal book of Viner (1950), the survey focuses on early discussions and in particular on the recent theory of sequencing when forming a trade bloc, which will subsequently be applied to the Zollverein.

Corresponding to Fischer's contributions, two topics are selected to be discussed in more detail. The Zollverein, which preceded the political unification of Germany, has long been discussed as a model for the European Community. Section 4 takes up this point in view of Fischer's second article which deals with the nature of a customs union and the institutions of the Zollverein. Fischer's first article contains a history of the Zollverein which emphasises the sequence of accession and often points out how the joining of one state affected the situation of others and their decision and eventual negotiations to join the Zollverein. Thereby, Fischer's work is related to the modern theory of sequencing, that includes Baldwin's (1995) domino theory and the theory of negotiating free trade by Aghion et al. (2007). In section 5, the history of forming and expanding the Zollverein is recapitulated in view of the modern theory, making use of recent work of Ploeckl (2010). Moreover, possible domino effects in the Cobden-Chevalier network of bilateral commercial treaties in the nineteenth century are briefly discussed. Finally, section 6 concludes the paper.

## 2 Gustav Fischer and the Zollverein

### 2.1 Gustav Fischer and his articles

The starting point of the paper is a discourse written by Gustav Fischer in the first volumes of the *Jahrbücher für Nationalökonomie und Statistik* with the title "Ueber das Wesen und die Bedingungen eines Zollvereins" (On the nature and the conditions of a customs union), published between 1864 and 1867. Gustav Fischer<sup>2</sup> (1803–1868) studied law and political sciences (Rechts- und Staatswissenschaften) in Jena and Göttingen, and he became professor at the University of Jena<sup>3</sup> in 1834. He was one of the two founders of the Department of Political Sciences (Seminar für Staatswissenschaften)<sup>4</sup> at the University of Jena in 1850 and one of its two directors. From 1860 onwards, the second director was Bruno Hildebrand (1812–1878), the founder and editor of the *Jahrbücher*. In 1848/49, both Gustav Fischer and Bruno Hildebrand, who at that time was a professor at the university of Marburg, were members of the Frankfurt national assembly.

<sup>2</sup> The information on Gustav Fischer was provided by Marcus Müggenburg of the Archive of the Friedrich Schiller University Jena.

<sup>3</sup> Both Jena and Buttstädt, the town where Gustav Fischer was born, belonged to the Duchy of Saxe-Weimar-Eisenach, which was one of the Thuringian principalities.

<sup>4</sup> On the ground of this experience, Fischer (1857) advocates the formation of departments of political sciences. According to Fischer, the "Staatswissenschaftliche Seminar" in Jena was the only one at a German university at that time.

Professor Gustav Fischer should not be confused with the founder of the Gustav Fischer Verlag in Jena, who was more than 40 years younger (1845–1910) and born in Hamburg. After 1953, the *Jahrbücher* were published in the Gustav Fischer Verlag, which by that time had moved to Stuttgart.

The German Zollverein was formed in 1834 as a true customs union of German sovereign states. It had predecessors, and it gained in members in the course of time. Prussia was the dominant country, always having more than 50 % of the population of the Zollverein. Eventually, the German Reich was formed in 1871 and the central government became responsible for trade policy.

In his articles, Gustav Fischer aims at investigating the nature and the conditions of a customs union in general, as in spite of the extensive literature on the German Zollverein existing at that time, this question had hardly been discussed (Fischer 1864: 317 f.). His discourse consists of two articles, each of which appeared in two parts in different issues and the second one even in two different volumes of the *Jahrbücher*. Altogether, the articles have more than 250 pages, which amounts to the size of a book.

The first article has the subtitle “Die Idee eines deutschen Zollvereins und ihre Ausführung geschichtlich entwickelt” (The idea of a German customs union and its implementation considered historically, vol. 2, 1864, 317–385 and 397–432) and contains a thorough treatment of the genesis of the German Zollverein and its development, in particular its various enlargements. In his introduction, Fischer (1864) points out that customs unions had only emerged among German states in the nineteenth century, and hence such a historical part of the discourse is worth the effort to understand the nature of a customs union. Indeed, in a review on new publications about the Zollverein in the *Preussische Jahrbücher* (1865), Fischer’s (1864) article is described as a valuable and clear contribution, that is the first complete history of the Zollverein written with large conscientiousness and exhaustive use of all published sources.

The subtitle of the second article is “Das Wesen eines Zollvereins und der Unterschied desselben von anderen Arten der Zollgemeinschaft unter mehreren Staaten” (The nature of a customs union and the difference to other kinds of customs communities among several states, vol. 7, 1866, 225–304, and vol. 8, 1867, 252–350). This article characterises and discusses the concept of a customs union (Zollverein) and distinguishes it from other kinds of common customs areas, namely a common tariff system within a federal state and a customs accession (Zollanschluss), where one state joins a customs area without receiving a vote on the future development of the tariff system. In the discussion on the Zollverein, there had been confusion on these concepts, and Roscher (1870: 207) writes that it is the merit of Fischer (1866) to be the first to discuss the difference between a customs union and a customs accession.

Fischer intended to write a third article on the conditions for the formation and the viability of a customs union, as announced in the title of the series. Moreover, he wanted to write on the role of Austria and to explain the distinction of a customs union from a customs area in a federal state in more detail. He could not finish this third article due to his death in March 1968. A fragment of the intended article was published in the *Jahrbücher* posthumously in the same year. In Fischer (1868), he points out that a common tariff of the North American states was only introduced with the Federal Constitution in 1789.

The United States were one example for the observation that a customs community had never existed in a confederation of states, but had only been realised when the confed-

eration was transformed into a federal state (Fischer 1867: 281). This argument was well-known and is already contained in List (1841): “All examples which history can show are those in which the political union has led the way, and the commercial has followed” (cited from Viner 1950: 95, who lists a number of examples and provides a comprehensive discussion).

### **3 Literature on the Zollverein and on Preferential Trade**

#### **3.1 The perception of the Zollverein**

In particular the first of Fischer’s two articles had some immediate impact. Fischer’s article inspired Aegidi (1865) to write a detailed study on the history of the run-up to the Zollverein, which in turn was reviewed by Fischer (1865) in the *Jahrbücher* in an article of 40 pages length. Fischer assessed large parts of Aegidi’s book as good. But there was a dispute on the statement of Aegidi (and several Prussian authors) that already at the time of the Prussian customs reform in 1818, Prussia had the plan of founding a German customs union – an opinion that Fischer considers as being absurd (Fischer 1864: 342 footnote; 1865: 383 ff.), and he is supported in his view by Roscher (1870).

Weber (1869) refers to Fischer (1864), which he calls an excellent article, many times in his book on the Zollverein. The interest in the Zollverein in the late 1860s is also reflected in several reviews of the literature on the Zollverein and of its publications in the *Jahrbücher* (e.g. *Jahrbücher* 1868). Weber’s book is reviewed in 1870.

At that time, there had been debates on various aspects of the Zollverein since its founding with a variety of different views. Hahn (1982: 13–14) points out that the bulk of the present-day patterns of interpretation of the Zollverein already emerged in the contemporary discussion. However, most of this knowledge sank into oblivion rapidly after 1866 (Hahn 1982). After the political unification of Germany in 1871, the view that the Zollverein was a vehicle for achieving national unity dominated for decades, and Viner (1950: 98) still took it for granted “that Prussia engineered the customs union primarily for political reasons, in order to gain hegemony or at least influence over the lesser German states.” However, the simplifying picture that the Zollverein from its beginning had the aim of national unification is meanwhile attributed to Prussian historiography represented in particular by Treitschke (1897), which interprets the Zollverein *ex post* based on the result of German unification. Wolfram Fischer (1960) points out that it was economic and financial interests from which the Zollverein originated. He concludes his case study on the Zollverein stating that “To-day we are in a better position than the generations before us to see it (the Zollverein) in its proper perspective: not as the glorious beginning of a glorious history, but as a system of expedients set up to meet urgent needs.” Similar reasonings can be found in Henderson (1981), Hahn (1982) and Dumke (1984).

At the end of the 19th and the beginning of the 20th century, economists’ research interest in the Zollverein remained limited. Preferential trade was, however, discussed in the context of commercial treaties and of the most favoured nation clause (MFN clause). The Cobden-Chevalier Treaty between Britain and France in 1860 started a wave of bilateral negotiations among other European states who aimed at equal market access to the contracting parties. This process resulted in a network of bilateral MFN-treaties in Europe and in fairly free trade among European countries (see section 5.4). However, towards the late 19th century, when these treaties were renegotiated, the attitude towards more

open trade had changed and the MFN clause started to be granted only with limitations (WTO 2011: 49 f.). In the *Jahrbücher*, there are two contributions on the question of the MFN clause, namely Grunzel (1893) and Posse (1928). Both authors refer to recent events diluting the MFN clause and strongly advocate to include the unconditional MFN clause into commercial treaties.

There were also ongoing discussions on the idea of forming customs unions among European states, summarised in Pentmann (1917). After World War I, when trade policies had become more protectionist, there existed numerous proposals and attempts to reach preferential trade agreements or to found customs unions, but no agreements were reached. Overviews on those plans can be found in United Nations (1947), Viner (1950) and Machlup (1977). Henderson (1966) covers plans to found customs unions in Central Europe from 1840-1940. The *Jahrbücher* contain an article by Zwiedinek-Südenhorst (1926), who discusses and advocates a German-Austrian customs union.

The interest in the Zollverein itself increased again when the 100th anniversary of its founding approached. In 1927, the Friedrich-List-Gesellschaft took the initiative to prepare an edition of documents of the founding of the Zollverein covering the years 1815–1834, which appeared as Eisenhart Rothe et al. (1934). In the *Jahrbücher*, Zwiedinek-Südenhorst (1935) reviews this edition and in particular the introduction written by Oncken. Both Oncken and Zwiedneck-Südenhorst keep some distance from the Prussian historiography, but the discussion sticks to the question on who deserves the merit of being the creator of the Zollverein – Prussian statesmen or Friedrich List or who else. Fischer (1864: 341) considers the Badian statesman Nebenius as the intellectual father of the Zollverein (see section 4).

A differentiated and more sober interpretation of the Zollverein was presented by Henderson (1939)<sup>5</sup>. His well-founded study is based on a large amount of material, including the edition of Eisenhart Rothe et al. (1934), own research in archives in Vienna and London and also the above-mentioned book by Weber (1869), which he cites several times as a general reference. Henderson (1939: 94) also cites Fischer (1864), albeit only with a translation of the short description of the night of the January 1, 1834, when the Zollverein came into effect.

Viner's (1950) book on *The customs union issue* is mainly known for its contribution to the theory of customs unions (see section 3.2), but it also covers historical and political aspects. Viner often refers to the Zollverein since “the German Zollverein was the pioneer and by far the most important customs union, and generalisations about the origin, nature and consequences of unification of tariffs tend to be based mainly or wholly on the German experiment” (page 97). The book contains a bibliography on customs unions, which contain the articles of Gustav Fischer (1864, 1866, 1867) in the *Jahrbücher*, but they are not directly cited<sup>6</sup>.

Interest in the Zollverein revived in the context of the beginning European economic integration. An early contribution that explicitly refers to the Zollverein as a model for European integration is Schmölders (1954), followed by Fischer (1960), Henderson

<sup>5</sup> There are two more editions of Henderson's book, which are actually reprints. The second edition was published in 1959, and a third edition appeared in 1984, the year of the 150th anniversary of the German Zollverein. Together with Hahn (1984), Henderson's book is still the standard history of the Zollverein.

<sup>6</sup> The same applies to the books by Aegidi (1865) and Weber (1869) and to the article of Roscher (1870).

(1981) and Dumke (1984). Some analogies and differences of the design of the Zollverein and the European Economic Community will be considered in section 4. In the preface to the second edition of his book on the Zollverein in 1959, Henderson briefly discusses what can be learnt for the European Common Market. One aspect he emphasises is that the “German experience of the nineteenth century shows that the successful working of a customs union does not require member states to be of approximately equal size or to have reached the same stage of economic development”. The post-war discussion on European integration is, however, not reflected in the *Jahrbücher* and there are no articles on this question<sup>7</sup>. The same is true for the evolving theory of customs unions briefly discussed in the next subsection, and none of the books by Viner (1950), Balassa (1962), Meade (1955) or Machlup (1977) was reviewed in the *Jahrbücher*. Finally, Dumke (1981, 1994) introduced a new assessment of the economic effects of the Zollverein, estimating them to have been quite small, which implies that also its relevance for industrialisation was limited. His interpretation of the Zollverein is that the primary reason for its formation was tariff revenues, based on economies of scale in tariff administration in a larger customs area. Moreover, those revenues were received without consent of the parliaments, which was an advantage for the reigning monarchs.

The Zollverein remains a classical example for a customs union of several sovereign states, and recently Baldwin (2006) has referred to the Zollverein as an example for his theory of domino effects (see section 5).

### 3.2 The economics of regionalism

The theory on preferential trade agreements (PTAs) is generally considered to begin with Viner (1950), who introduces the concepts of trade creation and trade diversion and shows that the welfare effects for the participating countries are generally ambiguous. Viner's book is on *The customs union issue*, but the effects of trade creation and trade diversion do not only apply to customs unions but to preferential trade in general, whenever a country imposes lower tariffs on goods of the contracting parties than on goods of other countries.

According to the rules of the GATT, reciprocal PTAs must take the form of free trade areas or customs unions. However, historically PTAs did not introduce free trade among the contracting parties, but commercial treaties in general only determined to lower tariffs for certain goods produced in the partner countries (e.g. the treaties of the Cobden-Chevalier network briefly discussed in section 5.4). Nevertheless, part of the literature formulates the economics of preferential trade referring to customs unions, regardless of whether the results apply only to customs unions or hold more generally.

#### Early contributions

Pomfret (1988, chapter 6) discusses some contributions written earlier than Viner (1950), stating that some of them recognise the possibility of trade diversion already, but none of them spells out the consequences for the welfare effects of forming a customs union. However, O'Brien (1976) tracks the notion that due to trade diversion, preferential trade agreements may be harmful to the participating countries back to Adam Smith, who opposed the Methuen Treaty between England and Portugal signed in

<sup>7</sup> The register of the years 1929–1985 of the *Jahrbücher* contains only two entries for “European Community”, which refer to very specialised problems.



1703. O'Brien also reports discussions on whether the Cobden-Chevalier Treaty in 1860 was beneficial to England or not, in which the relative amount of trade creation and trade diversion involved in the treaty was an issue<sup>8</sup>. Trade diversion was sometimes also mentioned in the context of the Zollverein. Dumke (1981: 262) points out a discussion carried out in newspapers and reported in the *Zollvereinsblatt* (1846) in which it is argued that the southern German states suffer from trade diversion, as Prussia and Saxonia have displaced France and England as a supplier of manufactures. O'Brien (1976: 551 f.) recalls that England regarded the Zollverein at first with suspicion fearing exclusion from the markets, but after some years a number of British writers acknowledged that the external tariffs of the Zollverein were sufficiently low to make it mainly trade-creating.

However, this knowledge on preferential trade and on the possibly dominating welfare effects of trade diversion sunk into oblivion (similar to the variety of approaches of interpreting the Zollverein), and "the early twentieth-century literature is disappointing" (O'Brien 1976: 541 f.). Before Viner's book was published in 1950, it was generally believed that preferential tariff reductions are better than no tariff reductions and are always beneficial to both countries involved, being possibly harmful only for third countries. This is true in spite of the fact that Viner (1931) had already published on the topic. For example, Haberler (1933, chapter 20.5) denies Viner's conclusions (see also Viner 1950: 53, footnote 12).

### Traditional and current theory

Viner's (1950) formulation of the concept of trade creation and trade diversion inspired a comprehensive literature on the theory of preferential trade. As there are excellent surveys on the theory of customs unions and regionalism, for instance Krauss (1972), Panagariya (2000) and WTO (2011), its development is only sketched very briefly. In the 1950s, the discussion was particularly vivid, and a comprehensive bibliography can be found in Balassa (1962).

Traditionally, the static effects of a customs union, consisting of trade creation, trade diversion and terms of trade effects, and dynamic effects emerging from an increased market, are distinguished. The latter comprise economies of scale internal to the firm, external economies of scale, fiercer competition and reduction of risks and uncertainty from foreign transactions (Kreinin 1964). The theory has also been extended to analyse the effects on outside countries, and the growing incentive to impose tariffs on imports from third countries according to the optimal tariff reasoning as the customs area becomes larger.

Since the 1990s, the focus of the literature has shifted to another kind of dynamics, namely the effects of forming preferential trade agreements on the global trading system. A first question is, whether PTAs hamper or foster multilateral trade liberalisation, whether PTAs are stumbling blocs or building blocs to reaching global free trade, as Bhagwati (1991) has formulated it. Jagdish Bhagwati is the most prominent economist who see the growing regionalism as a threat to the multilateral process (see e.g. Bhagwati/Panagariya 1996 and Bhagwati 2008). The main representative of the opposite view that PTAs are building blocs on the way to world-wide free trade is Richard Baldwin (2006).

<sup>8</sup> O'Brien (1976) in particular refers to McCulloch, a Scottish economist who lived from 1803–1864.

A further question is how the recent wave of new preferential trade agreements can be explained, leading to theories of endogenous formation of PTAs. Brief surveys of this literature are contained in Egger and Larch (2008) and Baldwin and Jaimovich (2012). Part of the literature deals with the question of sequencing in forming and expanding trade blocs. As the theory of sequencing will be applied to the Zollverein in section 5, it is explained in more detail in the next section. Endogenous formation of PTAs in the historical context of the nineteenth century is considered in Pahre (2008).

### **3.3 Sequencing in forming a trade bloc**

#### **Domino theory**

When explaining the growing regionalism and the expansion of existing preferential trade agreements, Baldwin (1995) argues that there is kind of a domino effect. Both the enlargement and the deepening of integration of a trade bloc change the situation for the outsiders, who may be induced to apply for membership. Thus there is a feed-back effect, which can be explained by trade diversion. When a preferential trade area grows, third countries will be affected by additional trade diversion and joining the trade bloc becomes more attractive. If becoming a member of the trade bloc is not feasible – for political reasons or because the union countries do not admit the country – a free trade agreement with other countries could become more attractive (Egger/Larch 2008). Classical examples for the latter are the founding of the EFTA in 1960 in response to the EEC (Kreinin 1960) and the Tax Union in nineteenth century Germany (see section 5).

Baldwin formalises his idea of a domino effect by analysing the incentive for outsiders to seek entry. He combines a version of the model that Krugman (1991) uses in investigating economic geography issues with the political-economy (protection-for-sale) model of Grossman and Helpman (1994). Countries are symmetric economically, but potential entrants differ in political resistance to membership in the trade bloc. Countries with larger non-economic costs of entry require a larger economic advantage to seek accession. The initial equilibrium is characterised by the feature that the economic benefit of being in the trading bloc of the marginal member equals its political costs. Baldwin assumes an exogenous shock disturbing the equilibrium, and gives the European single market initiative as an example. The advantages of joining the trade bloc increase by the shock, and according to the political economy model firms in the outside country at the margin will lobby for entry. The enlargement increases the benefits of membership for the remaining outsiders which may now offset political costs for the next potential member. Thus a domino effect emerges, which may lead to a new equilibrium size of the trade bloc or to global free trade. Of course, the basic idea of the model can be applied to other asymmetries of the countries, as e.g. different economic situations or different size.

Panagariya (2000) points out two criticisms of the model. A first limit is that trade barriers are modelled as iceberg transport costs, thus the revenue aspect of tariffs is not considered. Secondly, Baldwin assumes that insiders admit every country that wants to join and does not consider whether insiders have an incentive to block entry. Baldwin only looks at the demand side of joining a trading bloc and neglects the supply side. Moreover, it can be criticised that an exogenous shock is necessary to begin the process of expansion.

The domino theory has been tested empirically by Sapir (2001) and more recently by Baldwin and Rieder (2007), Egger and Larch (2008) and Baldwin and Jaimovich (2012) for the current regionalism and by (Lampe 2011) for the nineteenth century. These studies use long series of country pair data and all find some support for the domino theory. In the interpretation of the results it has to be watched out, however, that joining a trade bloc and forming a new bilateral preferential trade agreement is not the same (Baldwin/Jaimovich 2012).

### **Negotiating free trade**

Aghion et al. (2007) also tackle the question of sequencing, but they do not only take the perspective of potentially acceding countries, but consider both the demand and the supply of joining a trade bloc. They assume that a leading country can negotiate on forming a free trade area with the other countries. The leading country has the choice whether to negotiate at all and with whom to negotiate, and it may bargain sequentially or multilaterally. In the negotiations, it makes “take it or leave it” offers. Aghion et al. model a transferable-utility game (meaning that some kind of side-payments or concessions are possible) with three countries.

Two concepts turn out to be crucial: Grand coalition sub-additivity and coalition externalities. Grand coalition sub-additivity means that regardless which coalition structure prevails, the sum of the payoffs is lower than in a world with global free trade. Coalition externalities occur, when forming a free trade area has effects on the payoff of the outside countries, which are usually thought of to be negative, but which may also be positive. As long as grand coalition sub-additivity holds, global free trade will be realised in equilibrium. In this sense, preferential trade agreements are neither stumbling blocs nor building blocs of multilateral liberalisation. With grand coalition superadditivity, the structure of coalition externalities only determines, whether the leading country prefers to negotiate sequentially or multilaterally. If there are coalition externalities in at least one of the follower countries, the leader prefers sequential bargaining. The reason is that the offer to a country that is negatively affected by a free trade agreement of the other two countries may be lower than in a multilateral negotiation.

The condition of grand coalition sub-additivity is satisfied in many trade models, when countries aim at maximising welfare. It may cease to hold, however, when there are political costs of joining a trade bloc or if the payoff function of the government is not welfare, but corresponds to a different political objective, which may be influenced by special interest groups or – as may be relevant for the Zollverein – may focus on tariff revenue. In this case, the process of negotiating need not lead to global free trade, and in the model of Aghion et al. (2007), PTAs may both be building blocs or stumbling blocs for global free trade. As it is well developed on which assumptions and conditions the results depend on, the brief remark by Baldwin and Jaimovich (2012) that Aghion et al. (2007) “show that almost anything can happen” does not describe the contribution adequately.

In section 5, the history of the Zollverein is discussed in view of the theory of sequencing. To apply the theory of negotiating of Aghion et al. (2007) to the German Zollverein, Ploeckl (2010) generalises some results to the case of more than three countries. Generally, when the leading country has decided to bargain sequentially, countries whose accession will have strong negative externalities on the other potential member countries will be likely to be early in the sequence. Such an order of bargaining makes negotiations a better deal for the leading country.

#### 4 On the nature of a customs union

In his second article, Gustav Fischer (1866, 1867) thoroughly characterises the nature of a customs union (Zollverein) and distinguishes it from other kinds of common customs areas. To this end, he refers to the terms of the Zollverein, of the Bavarian-Wuerttemberg customs union and the Tax Union (see section 5). He both considers issues which he assesses to be essential for a customs union and issues that seem desirable for the free movement of goods, but are not constituting for a customs union. Starting from some key points made by Fischer, this section discusses some analogies and differences in the institutions and the functioning of the Zollverein and the European Community.

Fischer distinguishes a *customs union* from a common tariff system within a federal state and from a *customs accession* (Zollanschluss)<sup>9</sup>, where one state joins a customs area and participates in the tariff revenues without receiving a vote on the future development of the tariff system<sup>10</sup>. Today's distinction between a customs union and a free trade area is not focussed on, but Fischer (1866: 244 ff., footnote 36) discusses plans to ensure the free movement of goods without a common customs system in a five-page footnote. His judgement is that such a system is not practical.

Fischer (1866: 303) defines a *customs community* (Zollgemeinschaft or Zollverband) as “an association of sovereign states contracted for a limited period to ensure free movement of goods by a common system of border tariffs and to share the tariff revenues commensurately”<sup>11</sup>. This definition is close to what nowadays is called a customs union (Viner 1950: 5). The only difference is that Fischer emphasises that the agreement should be for a limited time (see below). Viner (1950: 7 ff.) attributes this definition to Cavour in 1857 in an exchange of diplomatic notes with Austria<sup>12</sup>. However, Fischer (1864: 340) refers to a memorandum of Friedrich Nebenius written already in 1820, in which for the first time the nature of a customs union was developed and its feasibility was shown (later published as Nebenius 1833).

As already mentioned above, Fischer distinguishes between a *customs union* and a *customs accession*. Therefore, he regards two additional features as belonging to the definition of a customs union. First, all changes in tariffs have to be decided unanimously, giving each member a veto, and second, the members keep their independent customs administration. Indeed, the central administration of the Zollverein was a small office in Berlin. As Henderson (1981: 502) puts it “in no way does the European Economic Community differ more from the Zollverein than in the size of its civil service”.

Moreover, Fischer considers a limited time period for the agreements as essential for the nature of a customs union<sup>13</sup>, as a treaty for an unlimited time would be incompatible with the sovereignty of the member states. Otherwise some sovereignty would be surrendered to the union forever, which would only happen if the members formed a federal

<sup>9</sup> There have been many German terms for customs unions and variants thereof, and the exact use of them is ambiguous. The translation “customs accession” follows Viner (1950: 83).

<sup>10</sup> Customs accessions in this sense took place, when the small enclaves joined the Prussian customs system between 1819 and 1829. A present-day example is Turkey, which in the customs union with the European Union does not take part in the determination of the EU trade policy.

<sup>11</sup> The quotations from Fischer's articles were translated by the author.

<sup>12</sup> Austria had a commercial treaty with Sardinia ensuring reciprocal most-favoured-nation treatment with the exemption of a “complete customs union”. Cavour argued on the side of Sardinia, that the customs union between Austria and Modena was not complete.

<sup>13</sup> At the beginning in 1834, the contract period of the Zollverein was eight years, and it was twelve years afterwards.

state with a common customs system, which would make the customs union redundant (Fischer 1867: 346 f.). Nevertheless, Fischer assumes that an exit from a customs union would hardly actually happen. The requirement that exiting the union has to be possible seems noteworthy in view of the fact that in the European Union the right to withdraw was only introduced in the Treaty of Lisbon in 2009. It is also interesting in view of the current discussion on Greece in the European Monetary Union, for which there are no rules for a possible exit.

Fischer (1867: 256 ff.) understands the difficulties emerging from giving a veto to each member of the union, and discusses the veto (and proposals to change that rule) in length. However, the right to veto is indispensable for him. One reason he gives – besides the sovereignty of the members – is that the administration of a state would not conscientiously conduct a law, which was imposed on it by majority vote and to which it had not agreed voluntarily. Henderson (1981) uses the same reasoning referring to the European Economic Community and points out that for a long time votes were made only unanimously although the Council of Ministers formally could decide by qualified majority. “But there was no means by which a member of the Community could be forced to accept a decision by a ‘qualified majority’”. In practice it was necessary to secure a unanimous vote ...” (Henderson 1981: 496).

In the Zollverein, however, the right to veto a decision and the limited duration of the agreements were linked, and the possibility to exit mitigated the problems emerging from vetos. If a decision was vetoed by one or more states, Prussia could wait until the renewal of the treaties, and the other states had the choice of accepting the change or leaving the Zollverein. Prussia acted like this several times, and in practice, a threat to use the veto could only delay decisions for some years (Henderson 1981: 495; Fischer 1867: 338 ff.). Two examples will be given in section 5.3.

Fischer (1866) also discusses issues that are not essentials of the customs union, but that in modern terms would refer to measures ensuring the functioning of the internal market. For instance, within the Zollverein, a harmonisation of weights and measures took place, and there was also a reform of coinage. Moreover, similar commercial laws are desirable within a customs union, and the treaties of the Zollverein expressed the intention to move in that direction, which however did not happen (page 294 ff.). Moreover, Fischer’s discussion on the desirable harmonisation of indirect taxes resembles very much the arguments in the lead-up to the internal market of the European Union before 1992 (page 277 ff.). The only excise tax that was unified within the Zollverein and jointly collected was the tax on beet root sugar. It was introduced in 1841, since the tariff revenue from imported cane sugar had fallen, and the tax rate was increased several times until 1858 to ensure that the revenue replaces the loss in tariffs (page 283 f. including footnote 76).

The taxing of sugar illustrates that in the Zollverein, tariffs were to a large extent fiscal tariffs<sup>14</sup>. Henderson (1981: 503) highlights the fiscal aspect of tariffs by pointing out that “In Prussia for example, just before the establishment of the Zollverein, the revenue from

<sup>14</sup> Note, however, that most European countries introduced fiscal policies for developing their own sugar production during the nineteenth century, combining tariffs, taxes, and direct and indirect subsidies for transportation, production or export. In 1864, France, Belgium, the Netherlands (the greatest exporters) and Great Britain (the largest importer) agreed upon a Sugar Convention limiting export subsidies. The Brussels Convention of 1902 was signed by a larger group of countries including Germany. The Sugar Convention is an early example of multilateralism. For details see Moura Filho (2006). The German sugar tax was abolished only when the European Single Market was introduced in 1992.

customs duties was a little over 40 per cent of total revenue.” For Wolfram Fischer (1960: 67), the intended purpose in levying tariffs is one of the major differences between the Zollverein and the European Community. Whereas in modern times, tariffs are introduced to protect industries, in the nineteenth century, the revenue aspect was dominant. In the Zollverein, tariffs were mainly levied on mass consumption goods, such e.g. coffee, tea and colonial food stuff. After a comprehensive reasoning based on data on the Zollverein, Dumke (1994: 22) concludes “that the tariff ... of the Zollverein ... was primarily a tariff for revenue”. These results are in line with Fischer (1866: 251 ff.), who also emphasises the relevance of tariff revenue after discussing both protective and fiscal tariffs.

## 5 Sequencing in the Zollverein

Among the examples Baldwin (2006, footnote 13) gives for domino effects in forming a trade bloc, he also refers to the Zollverein and cites Viner (1950, chapter V.3). Explaining the foundation and expansion of the Zollverein is indeed a relevant application of the modern theory of sequencing in forming customs unions. However, the cited chapter in Viner’s book only contains a brief summary of the history of the Zollverein based on Henderson (1939) as a principle source. These authors mostly explain why and in what order the German states would request to join the Zollverein, thereby looking at the demand side of membership. Ploeckl (2010) adds the aspect of supply of membership applying a generalised model of Aghion et al. (2007) to the Zollverein (see section 3.3.). He analyses how the structure of negative externalities explain which sequence Prussia as the leading country chose for the negotiations. Moreover, he points out that the negative externalities explain why Prussia did not opt for multilateral negotiations and in general only negotiated with the other states one by one. Negative externalities can for instance emerge from trade diversion and subsequent fall in trade volume and a loss in tariff revenue. Revenues may also be affected by the diversion of trade routes and thereby falling transit duties.

Many German states resisted to join the Zollverein in spite of financial and economic advantages of membership, because they feared to lose their sovereignty. This resistance can be interpreted as reflecting non-economic costs of membership as in the model of Baldwin (1995), and such costs would enter the pay-offs in the approach of Aghion et al. (2007).

The negative externalities, i.e. the worsening of the economic situation of the outsiders due to an accession of a state to the Zollverein, are already discussed by Fischer (1864), and his formulations describing the history of the Zollverein contain many elements of domino effects. Hence there is another example where the basic notion of a modern pattern of interpretation of the Zollverein already existed in the 1860s. In the following, the history of the Zollverein focussing on the sequencing of the accession of the German states is summarised, accentuating both Fischer’s work and the contribution of Ploeckl (2010)<sup>15</sup>.

<sup>15</sup> A map of the Deutsche Bund (German Confederation) showing the German states in 1818, which may be helpful in following the arguments, can be found on [http://commons.wikimedia.org/wiki/File:Deutscher\\_Bund.png](http://commons.wikimedia.org/wiki/File:Deutscher_Bund.png) or at the server for digital historical maps of the Leibniz Institute for European History Mainz (IEG), <http://www.iegmaps.de/mapsp/mapd820.htm>. Note that Prussia consisted of a western and an eastern part that had no connection. Moreover, Series 4 of IEG-MAPS contains maps on the economic unification of Germany 1828 to 1901, including the area of the Zollverein at different dates, <http://www.ieg-maps.uni-mainz.de/map4.htm>.

In addition, Austria's relation to the Zollverein and its vain attempts to join it are discussed in section 5.3, and section 5.4 considers possible domino effects in the Cobden-Chevalier network.

### 5.1 The lead-up to the Zollverein

Often, the first important date mentioned in the history of the Zollverein (not yet referring to the formation of a customs union) is 1818, when Prussia introduced a new border tariff system. Not only Germany was splintered into 38 sovereign states, but also Prussia was no uniform customs area. Before 1818, tariffs were not collected at the border, but customs duties were instead levied at places such as town gates, bridges, roads, crossings and the like. According to Fischer (1864: 329), there were 60 different systems of tariffs and excise taxes in Prussia, excise taxes were levied on 2775 different products, and there were import and export prohibitions. Trade within Prussia was impeded, revenues were low, administration costs high and smuggling was widespread. The new tariff law introduced free trade within Prussia and a tariff to be collected at the external borders. Internal duties could only be levied when a service was rendered, as e.g. a bridge toll (Henderson 1939: 40). The southern states Bavaria (1807), Wuerttemberg (1808) and Baden (1812) had conducted similar reforms some years earlier already, but essentially the Prussian tariff (with some modifications) was to become the tariff of the Zollverein in 1834 (Hahn 1984: 12).

The first expansion of the Prussian customs area was the adoption of the tariff system by the enclaves within Prussia – which did however not result in a customs union but were customs accessions. To prevent smuggling from these enclaves and to simplify the administration, Prussia treated the enclaves as if they belonged to the Prussian customs area and levied transit duties at its borders on goods destined for the enclaves. At the same time, Prussia offered them a share of the tariff revenue according to the share in population when they joined the Prussian customs system. Although this offer was financially attractive for the enclaves and in spite of the pressure imposed by the transit duties, many enclaves resisted accession. They obviously had non-economic costs of accession, as they were concerned about their sovereignty. The treaties with the nine enclaves were signed between 1819 and 1830 (Fischer 1864: 354; Ploeckl 2010: 15). Attempts in the 1820s to form a customs union among several Southern German states failed. In 1827, the two largest of them, Bavaria and Wuerttemberg, negotiated a customs union, which came into force on July 1, 1828. Baden did not become a member, as it wanted to keep its low tariffs. Hesse-Darmstadt was also invited to join, but it was to become the first state which completely voluntarily entered a customs union with Prussia.

An important aim of Prussia was to connect its eastern and the western territories within a common customs area. Hesse-Darmstadt was negatively affected by the Prussian tariff reform, as it lacked free access to the large Prussian market. Moreover, its financial situation was bad. Hesse-Darmstadt had approached Prussia already in 1825 to form a commercial treaty, but at that time Prussia only wanted to negotiate jointly with Hesse-Darmstadt and Hesse-Cassel. The latter would have linked the two parts of Prussia but refused to negotiate for political reasons (Henderson 1939; Hahn 1984: 51). When Hesse-Darmstadt again asked for negotiations in January 1828, agreement on a customs union was reached quickly. Fischer (1864: 357) points out that the change in Prussia's attitude was probably due to the formation of the Bavarian-Wuerttemberg customs union. Hesse-Darmstadt adopted the Prussian tariff system, but retained full sovereignty and changes in tariffs had to be agreed upon by the two member states.



Hesse-Darmstadt benefited a lot from the customs union, as it received a share of the tariff revenue according to its population, which amounted to substantial financial transfers from Prussia. Moreover, the two Prussian territories were not connected by Hesse-Darmstadt and the common border was short. Hence for Prussia, the agreement was less profitable (Ploeckl 2010: 17).

Prussia thus entered the union with Hesse-Darmstadt in view of a future expansion of the customs union. "Only the confident hope that further states would join this customs union could make Prussia enter an agreement that on its own was financially more disadvantageous than advantageous for her" (Fischer 1864: 358). Most other German states reacted with displeasure on the agreement. Ploeckl (2010: 17 f.) emphasises the negative externalities on other states. The agreement between Prussia and Hesse-Darmstadt impeded an enlargement of the customs union between Bavaria and Wuerttemberg, and it put pressure on the other small German states around Hesse-Darmstadt.

As a reaction, several central and northern German states, among them Hesse-Cassel, founded the *Middle German Commercial Union* (Mitteldeutscher Handelsverein) in 1828. Geographically, it covered the market places Frankfurt and Leipzig, the roads to the North sea and all roads between the two parts of Prussia. As Fischer (1864) and Ploeckl (2010) both point out, the agreement included only some minor reciprocal facilitation of trade and some obligations referring to the maintaining and building of roads. It was no customs union and did not even ensure free trade among its members. The main term of the contract was that member states committed themselves not to join any customs union without the consent of the whole Union until 1834, when the agreement between Prussia and Hesse-Darmstadt expired. The objective was thus to prevent the customs union between Prussia and Hesse-Darmstadt (and also the Bavarian-Wuerttemberg one) from expanding. Ploeckl (2010: 18) interprets the Middle German Commercial Union as an attempt to mitigate negative coalition externalities, that might arise from further states joining the customs union with Prussia. Moreover, the Middle German Commercial Union tried to force Prussia to negotiate multilaterally. However, Prussia refused collective negotiations, which is in line with the theory of sequencing by Aghion et al. (2007) and Ploeckl (2010).

Fischer (1864) and Ploeckl (2010) somewhat disagree in what the next step was. Ploeckl considers Hesse-Cassel as the next element of the entry sequence focussing on negative externalities of the coalition between Prussia and Hesse-Darmstadt. In contrast, Fischer describes the joining of Hesse-Cassel as an effect of the rapprochement between Prussia and the Bavarian-Wuerttemberg customs union. Of course, these two reasonings do not exclude each other but can be understood as complements. The Bavarian-Wuerttemberg customs union was still small and had long customs borders, hence the efficiency gains of the customs union were small. Fischer (1864: 364) indicates that the administration costs of tariff collection amounted to 44 percent of the gross revenue to show the imperfection of that union. According to Fischer, approaching Prussia was substantially facilitated by the forming of the customs union between Prussia and Hesse-Darmstadt. Henderson (1939: 89) points out that Prussia, Bavaria and Wuerttemberg had a common distrust of the Middle German Commercial Union which also fostered the cooperation. In 1829, the two customs unions signed a comprehensive commercial treaty facilitating trade among the four states significantly. Moreover, the customs systems of the two unions were to be harmonised.

As a reaction, in October 1829 it was agreed within the Middle German Commercial Union that it was to be continued an additional six years until 1841 (Fischer 1864:

362). However, in this change of the contract Saxony insisted on a clause that it could leave the Middle German Commercial Union if the Bavarian-Wuerttemberg and the Prussian-Hessian customs unions merged.

The geographical location of Hesse-Cassel was crucial in the situation at hand. It controlled the traditional north-south trade routes and an important east-west road. Moreover, it separated the eastern and the western provinces of Prussia. Conversely, Hesse-Cassel was negatively affected by the union between Prussia and Hesse-Darmstadt. It faced higher tariffs at the border to Hesse-Darmstadt, and Prussia tried to increase the negative externalities by its road building programme, which made the north-south connection through Hesse-Cassel less important (Ploeckl 2010: 19; Hahn 1984: 60). Moreover, Hesse-Cassel was in a very bad economic condition. In 1831, Hesse-Cassel decided to defect with the treaty of the Middle German Commercial Union and to accept Prussia's offer to join the customs union.

The membership of Hesse-Cassel in the Prussian-Hessian customs union had very strong externalities in particular on the southern states, including Bavaria and Wuerttemberg. The area of the customs union was now connected, and as it spread from the very east to the very west of Germany, it split up the area of the Middle German Commercial Union and controlled all north-south roads. "All states located south of this customs area ... suddenly came into a desperate commercial situation." (Fischer 1864: 367). In contrast, an expansion of the customs union by other states would not have had similar negative externalities on Hesse-Cassel. Hence, as Ploeckl (2010: 19) convincingly emphasises, Prussia's negotiations with Hesse-Cassel complies well with his theory of an optimal sequence.

Next, the Bavarian-Wuerttemberg and the Prussian-Hessian customs union negotiated and agreed on merging and forming the Zollverein in March 1833. The revenue sharing according to population amounted to a strong financial transfer to Bavaria and Wuerttemberg, and they also received some political concessions (Ploeckl 2010: 20). This agreement in turn had externalities both on Baden and on Saxony and the Thuringian principalities. The latter two were surrounded by customs union area and at the Austrian border, trade was impeded by prohibitive tariffs. Hence Saxony and the Thuringian principalities had strong motives to join the Zollverein urgently (Fischer 1864: 372). According to Ploeckl (2010: 20f.), an expansion of the Prussian-Hessian customs union by one of these areas would not have had comparable effects on Bavaria and Wuerttemberg. Moreover, the merger of the two customs unions allowed Saxony to use its exit clause from the Middle German Commercial Union and join the Zollverein. Saxony joined since it feared isolation, although the revenue sharing scheme put it at a disadvantage, as its share in imports was much higher than its share in population. Ploeckl (2010) further argues, that the externality of a membership of Saxony on the Thuringian states was larger than it would have been vice versa, explaining the order of negotiations. The Zollverein started on January 1, 1834 with the states of the two customs unions and Saxony and the Thuringian principalities.

As a reaction to the founding of the Zollverein, the northern states Brunswick and Hanover formed a customs union in 1834, called the *Steuerverein* (Tax Union), which Oldenburg joined in 1836 (Ploeckl 2010: 23).

## 5.2 Enlargement of the Zollverein

Soon after the formation of the Zollverein, Baden and the Hessian states Nassau and Frankfurt also entered the union. Baden had long borders with France and Switzerland, hence it was not as dependent on the north-south routes as the other southern states. Moreover, it had a free trade orientation and hesitated to introduce the higher tariffs of the Zollverein. However, as mentioned above, it was affected by Bavaria and Württemberg being part of the Zollverein, and the large markets of the Zollverein made joining worthwhile, although it had refused to become a member of the Bavarian-Württemberg customs union in 1828. Negotiations were completed in May 1835 (Fischer 1864: 378; Ploeckl 2010: 22).

Nassau had hesitated to join a customs union, one reason for which was that it had revenue from tolls on the Rhine and therefore a high reservation price for joining the Zollverein. But after the accession of the southern states, it feared isolation. After reaching compromises referring to shipping on the Rhine, the agreement of joining dates from December 1835. Nassau's accession in turn affected the town of Frankfurt, which had lost its unimpeded access to the Rhine through the area of Nassau. Moreover, its trade fairs had suffered from the fairs in the neighbouring town of Offenbach, which was located within the Zollverein. Negotiations were completed in the course of the year 1836. Frankfurt was conceded to receive a fixed sum from the tariff revenues, which was higher than it would have been according to the population based distribution scheme (Fischer 1864: 380 ff.; Henderson 1939: 110 ff.; Ploeckl 2010: 22). This higher payment was justified by the fact that in a trading center, per capita trade was higher, but it can also be interpreted as a side payment making membership attractive.

When there was a dispute about the building of roads and railways during the negotiations on the renewal of the Tax Union, Brunswick approached the Zollverein to avoid isolation, and it joined in 1842. In that year, also Luxembourg, which was surrounded by the German, French and Belgium customs borders, decided to accede the Zollverein.

## 5.3 Austria, the accession of Hanover and the crises of the Zollverein

In the course of the renewals of the Zollverein treaties, which expired in 1853 and again in 1865, the Zollverein experienced two crises. In both cases, the dualism between Prussia and Austria played a crucial role. Prussia succeeded in preventing Austria from joining the Zollverein, although some southern members of the Zollverein were in favour of an accession of Austria. Hence the enlargement of the Zollverein reached a limit due to political reasons.

By 1842, of the states of the German Confederation, only Austria and a number of northern states – among them Hanover and the seaports Hamburg and Bremen – had not joined the Zollverein. Austria had a prohibitive customs system that it did not want to give up, hence it had not taken part in any negotiations on a customs union. In the 1840s, there were some discussions on a cooperation with Austria, but “Austria's failure to reform her tariffs in the early forties made it useless to negotiate with Prussia for a commercial treaty, let alone for an Austro-Zollverein union” (Henderson 1939: 177). The trade of the northern states was oriented overseas and less to the other German states, and they were free trade oriented and did not want to accept the higher tariffs of the Zollverein, hence in earlier negotiations, they had demanded far-reaching concessions (Fischer 1864: 401 f.). Formulated in terms of the theory of sequencing, “their access to major ports .... had the implication that the related absence of strong coalition

externalities meant that Prussia found the required reservation price too high and did not secure their accession” (Ploeckl 2010: 23).

Moreover, there was a conflict among protectionists and free-traders within the Zollverein (Hahn 1984: 113 ff.). Among others, Bavaria and Württemberg were protectionist, whereas Prussia wanted to lower some tariffs, which however was only possible by mutual consent of all members. After the revolution of 1848/49, trade policy in Austria changed, and the prohibitive system was replaced by a protective system in 1852. Austria aimed at an Austro-German customs union, hoping that the Zollverein would raise its tariffs (Henderson 1939: 203). Prussia at that time did not want a customs union with Austria for political reasons, but some southern states had threatened to only renew the Zollverein treaties, which expired at the end of 1853, if the Zollverein would include Austria.

In this situation, Prussia and Hanover secretly agreed on a customs union in September 1851, that was to begin in 1854. Prussia was interested in an agreement, as Hanover’s territory was an alternative of connecting its eastern and western provinces, which weakened the bargaining power of the other members. The conditions for Hanover were very favourable. In particular, it was to receive 75 % more of the tariff revenues than its share according to population. Moreover, several tariffs and taxes in the Zollverein were to be lowered (Henderson 1939: 213 ff.; Fischer 1864: 413 ff.). “This action effectively created a wall against Austrian accession, because Austria was unlikely to reduce its duties to these lower levels” (Pahre 2008: 315). As Prussia could terminate the Zollverein treaties in 1853, the other states could only accept the treaty with Hanover or leave the Zollverein, which none of them did. As explained in section 4, Prussia could use the threat of terminating the agreement at the next possible date to circumvent the difficulties arising from the possibility of veto by each member state<sup>16</sup>. Oldenburg, that together with Hanover had been in the Tax Union, followed Hanover into the Zollverein in 1854.

Before, Prussia and Austria had agreed on a commercial treaty at the beginning of 1853, which also was to become part of the new Zollverein treaties. Austria abolished most import prohibitions and lowered tariffs, and it received preferential access to the Zollverein. Moreover, negotiations on an Austrian-German customs union were to begin in 1860.

In the lead-up to the next renewal of the Zollverein treaties, the conflicting interests and the action of Prussia followed similar patterns. Prussia wanted to lower tariffs and to keep Austria out of the Zollverein. A change in tariffs was, however, only possible by unanimous vote, and southern member states of the Zollverein supported Austria in its wish to form a Central European customs union. Again, Prussia strengthened its bargaining situation by signing a treaty with a third party. Prussia agreed on a commercial treaty with France in 1862, which in 1860 had signed the Cobden-Chevalier Treaty with England (see section 5.4). Tariffs were to be reduced significantly, and the Franco-Prussian Commercial Treaty contained a most-favoured-nation (MFN) clause, meaning that Austria would lose its preferential access to the markets of the Zollverein. Austria opposed the treaty, and some members of the Zollverein rejected

<sup>16</sup> Fischer (1864: 408 ff.) describes the crisis of the Zollverein at the beginning of the 1850s in detail, and defends the course of action of Prussia, as the accession of Hanover was in the interest of all members of the Zollverein (page 416 f.). Fischer does not cover the second crisis of the Zollverein 1860–1865 any more.

it, but Prussia again conditioned the renewal of the Zollverein on the acceptance of the lowering of tariffs and the treaty with France. Eventually a new treaty between Austria and the Zollverein was agreed upon in April 1865 and in May 1865 a treaty was signed to renew the Zollverein for another twelve years (Henderson 1939; Hahn 1984).

However, after the Seven-Weeks War in 1866 and the defeat of Austria, many northern German states (among them Schleswig-Holstein, that until 1866 was not in the Zollverein) became part of Prussia, the Northern German Confederation emerged, the Zollverein was reformed (in particular introducing a Customs Parliament deciding by majority vote) with new treaties between Prussia, Bavaria, Baden, Wuerttemberg and Hesse-Darmstadt. Hamburg and Bremen remained outside the Zollverein even after 1871 and joined the common customs area only in the 1880s.

In his recent book on trade cooperation in the nineteenth century, Pahre (2008: 311 ff.) discusses Austrian trade policy beyond the common interpretation that Austria first was not willing to enter a customs union due to its prohibitive system and later was not admitted to join the Zollverein by Prussia due to the rivalry between the two states. Pahre points out that Austria responded to the Zollverein very close to what the reasoning of political economy would predict: First, it lowered its own tariffs, second it sought cooperation with the Zollverein and third, it cooperated with outsiders, signing treaties on commerce and tariffs for instance with Britain, Belgium, Russia, the Ottoman Empire and Sardinia. Moreover, the customs border between Austria and Hungary was abolished in 1850.

#### 5.4 Domino effects in the Cobden-Chevalier network

Domino effects are also discussed in the context of the network of bilateral commercial treaties among European states that emerged 1860–1875. The Cobden-Chevalier Treaty between Britain and France signed in 1860, which reciprocally lowered tariffs, is regarded to have sparked a wave of bilateral trade agreements (WTO 2011: 49). Lampe (2011: 645) shows a map of the Cobden-Chevalier network in 1875, calling it “the mother of all spaghetti bowls”<sup>17</sup>.

The bilateral treaties contained the unconditional most-favoured-nation (MFN) clause, thereby together creating kind of a “plurilateral preferential trade agreement” ensuring that a country could not discriminate among those trading partners with whom it was linked by a commercial treaty (WTO 2011: 49)<sup>18</sup>.

Nevertheless, when applying the concept of domino effects to the development of the Cobden-Chevalier network, it has to be noted that a country could not join an existing (multilateral) trade bloc. Rather, a bilateral treaty should be interpreted as a “closed preferential agreement” with two members, inducing to form new PTAs as a reaction. (Lampe 2011: 651). In this respect, the enlargement of the Zollverein and the contagion of the bilateral commercial treaties have to be distinguished. In his empirical analysis,

<sup>17</sup> The term “spaghetti bowl” was coined by Bhagwati to describe today’s numerous and interwoven PTAs, and Baldwin (2006) took up this term for the title of his paper.

<sup>18</sup> However, the MFN clause may hamper the spread or the deepening of trade liberalisation, as there is an incentive to try to free-ride on tariff reductions agreed upon by others. Pahre (2008, chapter 11) finds some empirical support for this hypothesis on the discouraging effect of MFN cooperation based on the treaties of the nineteenth century. Such effects may contribute to explain, why the process did not continue with further tariff reductions “in a second round” after 1875 (Lampe 2011: 662 ff.; Accominotti/Flandreau 2008: 181).

Lampe (2011: 651) finds that the Cobden-Chevalier network can be explained by combining the approach of Baier and Bergstrand (2004), who focus on the economic fundamentals, political economy considerations as in Pahre (2008) and Baldwin's (1995) domino theory based on fears of being affected by trade diversion.

The Cobden-Chevalier treaty of 1860 has long been regarded as the starting point of an era of fairly free trade among European countries. Recent literature has challenged this conventional view by pointing out that the process of substantial trade liberalisation had started already in the 1840s (Tena-Junguito et al. 2012; Accominotti/Flandreau 2008). Estimating a gravity model with data from 1850–1870, Accominotti and Flandreau (2008) find that the treaties did not have any marked effects on overall trade flows. Lampe (2009) draws a somewhat more differentiated picture, looking at different goods. Although total trade was not raised, he shows that the treaties did have an effect on trade, in particular for trade in final goods.

In his discussion on the reaction of outside countries to the Zollverein in the 1840s and 1850s, Pahre (2008: 316 ff.) points out, that both unilateral trade liberalisation and cooperation between outside countries may have been induced by the forming of the Zollverein, and that “in this way, the Zollverein may have been a catalyst for the entire nineteenth-century trade regime.”

## 6 Final remarks

The German Zollverein of the 19th century is a particular suitable application of the recent theory of sequencing in forming a trade bloc. The states indeed negotiated one by one with Prussia, and domino effects could be observed. States that had resisted becoming a member would propose to join the Zollverein some years later after it had been enlarged. Prussia never initiated the negotiations but waited until the respective governments made a proposal (Fischer 1864: 383). Thus the demand of membership which is at the focus of Baldwin's (1995) theory played a role. But Ploeckl's (2010) considerations on the supply of an agreement based on Aghion et al. (2007) are also relevant, as Prussia for instance chose to only negotiate bilaterally and not multilaterally, and the sequence can be explained by externalities of the membership of one country on others.

Domino effects of enlarging or spreading preferential trade agreements were observed also in other instances, e.g. within Europe and in Asia (Baldwin 1995, footnote 13). However, most of these cases are more diffuse. For instance, most enlargements of the European Union followed fundamental political changes within the joining countries, which should be considered as exogenous for the accession process and cannot be attributed to domino effects.

Hence the Zollverein remains a classical historical example for the economics of regionalism. The situation can be compared to the discussion on European integration and the emergence of the theory of customs unions in the 1950s: The Zollverein at that time was not the only historical example of a customs union, but by far the most relevant one. Likewise, the forming of the Zollverein is not the only example for domino effects in forming a trade bloc, but probably the purest one.

## References

- Accominotti, O., M. Flandreau (2008), Bilateral treaties and the most-favored nation clause: The myth of trade liberalization in the nineteenth century. *World Politics* 60(2): 147–188.
- Aegidi, L. (1865), *Aus der Vorzeit des Zollvereins: Beitrag zur deutschen Geschichte*. Boyes & Geisler, Hamburg.
- Aghion, P., P. Antras, E. Helpman (2007), Negotiating free trade. *Journal of International Economics* 73(1): 1–30.
- Baier, S. L., J. H. Bergstrand (2004), Economic determinants of free trade agreements. *Journal of International Economics* 64(1): 29–63.
- Balassa, B. (1962), *The theory of economic integration*. London: Allen & Unwin.
- Baldwin, R. (1995), A domino theory of regionalism. Pp. 25–48 in: R. Baldwin, P. Haaparanta, J. Kiander (eds.), *Expanding Membership of the European Union*. Cambridge, University Press.
- Baldwin, R. (2006), Multilateralising regionalism: Spaghetti bowls as building blocs on the path to global free trade. *World Economy* 29(11): 1451–1518.
- Baldwin, R., D. Jaimovich (2012), Are free trade agreements contagious? *Journal of International Economics* 88: 1–16.
- Baldwin, R., R. Rieder (2007), A test of endogenous trade bloc formation theory on EU data. CEPR Discussion Papers 6389.
- Bhagwati, J. (1991), *The world trading system at risk*. Princeton, NJ: Princeton University Press.
- Bhagwati, J., A. Panagariya (1996), Preferential trading areas and multilateralism: strangers, friends, or foes? Pp. 1–78 in: J. Bhagwati, A. Panagariya (eds.), *The economics of preferential trade agreements*. Columbia University: College Park.
- Bhagwati, J. N. (2008), *Termites in the trading system: how preferential agreements undermine free trade*. Oxford: Oxford University Press.
- Dumke, R. H. (1981), Die wirtschaftlichen Folgen des Zollvereins. Pp. 241–273 in: W. Abels-hauser, D. Petzina (eds.), *Deutsche Wirtschaftsgeschichte im Industriezeitalter*. Königstein/Ts.: Athenäum.
- Dumke, R. H. (1984), Der Deutsche Zollverein als Modell ökonomischer Integration. Pp. 71–101 in: H. H. Berding (ed.) *Wirtschaftliche und politische Integration in Europa im 19. und 20. Jahrhundert. Geschichte und Gesellschaft: Sonderheft 10*. Göttingen: Vandenhoeck & Ruprecht. (A revised and shortened version is published as: Tariffs and market structure. The German Zollverein as a model for economic integration. Pp. 77–115 in: W. R. Lee (ed.), *German industry and German industrialization. Essays in German economic and business history in the nineteenth and twentieth centuries*. London, Routledge.)
- Dumke, R. H. (1994), German economic unification in the 19th Century: the political economy of the Zollverein.
- Egger, P., M. Larch (2008), Interdependent preferential trade agreement memberships: An empirical analysis. *Journal of International Economics* 76(2) 384–399.
- Eisenhart Rothe, W. v., A. Ritthaler, H. Oncken (eds.) (1934), *Vorgeschichte und Begründung des deutschen Zollvereins*. Veröffentlichungen der Friedrich-List-Gesellschaft. Berlin: Hobb-ing.
- Fischer, G. (1857), Ueber die Errichtung staatswissenschaftlicher Seminarien auf den deutschen Universitäten nebst einem Bericht über das staatswissenschaftliche Seminar zu Jena. F. Mauke, Jena.
- Fischer, G. (1864), Ueber das Wesen und die Bedingungen eines Zollvereins: Die Idee eines deutschen Zollvereins und ihre Ausführung geschichtlich entwickelt (On the nature and the conditions of a customs union: The idea of a German customs union and its implementation considered historically). *Jahrbücher für Nationalökonomie und Statistik* 2: 317–385 and 397–432.
- Fischer, G. (1865), *Litteratur: Aus der Vorzeit des Zollvereins: Beitrag zur deutschen Geschichte*. Boyes & Geisler, Hamburg, 1865, von L. K. Aegidi und *Die Idee eines Zollvereins*, 1865, von L. K. Aegidi. *Jahrbücher für Nationalökonomie und Statistik* 5: 361–401.



- Fischer, G. (1866), Ueber das Wesen und die Bedingungen eines Zollvereins: Das Wesen eines Zollvereins und der Unterschied desselben von anderen Arten der Zollgemeinschaft unter mehreren Staaten, Teil 1 (On the nature and the conditions of a customs union: The nature of a customs union and the difference to other kinds of customs communities among several states, part 1). *Jahrbücher für Nationalökonomie und Statistik* 7: 225–304.
- Fischer, G. (1867), Ueber das Wesen und die Bedingungen eines Zollvereins: Das Wesen eines Zollvereins und der Unterschied desselben von anderen Arten der Zollgemeinschaft unter mehreren Staaten, Teil 2 (On the nature and the conditions of a customs union: The nature of a customs union and the difference to other kinds of customs communities among several states, part 2). *Jahrbücher für Nationalökonomie und Statistik* 8: 252–350.
- Fischer, G. (1868), Ueber die Bundeszollgemeinschaft Nordamerikas (On the federal customs community of North America). *Jahrbücher für Nationalökonomie und Statistik* 11: 314–327.
- Fischer, W. (1960), The German Zollverein: a case study in customs union. *Kyklos* 13(1): 65–86.
- Grossman, G. M., E. Helpman (1994), Protection for sale. *American Economic Review* 84(4): 833–850.
- Grunzel, J. (1893), Die Meistbegünstigungsklausel in der modernen Handelspolitik. *Jahrbücher für Nationalökonomie und Statistik* 60: 561–569.
- Haberler, G. v. (1933), *Der internationale Handel*. Berlin: Springer. English translation: *The theory of international trade*, London: Hodge, 1936.
- Hahn, H.-W. (1982), Wirtschaftliche Integration im 19. Jahrhundert. No. 52 in: *Kritische Studien zur Geschichtswissenschaft* Göttingen: Vandenhoeck und Ruprecht.
- Hahn, H.-W. (1984), Geschichte des Deutschen Zollvereins. No. 1502 in: *Kleine Vandenhoeck-Reihe*. Göttingen: Vandenhoeck & Ruprecht.
- Henderson, W. O. (1939), *The Zollverein*. Cambridge: Cass. 2. edition 1959, 3. edition 1984.
- Henderson, W. O. (1966), Mitteleuropäische Zollvereinspläne 1840–1940. *Zeitschrift für die gesamte Staatswissenschaft (JITE)* 122(1): 130–162.
- Henderson, W. O. (1981), The German Zollverein and the European Economic Community. *Zeitschrift für die gesamte Staatswissenschaft (JITE)* 137(3): 491–507.
- Jahrbücher für Nationalökonomie und Statistik* (1868), Litteratur über den Zollverein. Vol.10: 343–347.
- Jahrbücher für Nationalökonomie und Statistik* (1870), Neueste Literatur zur Geschichte des deutschen Zollvereins. Vol.14: 226–230.
- Krauss, M. B. (1972), Recent developments in customs union theory: An interpretive survey. *Journal of Economic Literature* 10(2): 413–436.
- Kreinin, M. E. (1960), The “Outer-Seven” and European Integration. *The American Economic Review* 50(3): 370–386.
- Kreinin, M. E. (1964), On the dynamic effects of a customs union. *Journal of Political Economy* 72: 193–195.
- Krugman, P. R. (1991), Increasing returns and economic geography, *Journal of Political Economy* 99(3): 483–499.
- Lampe, M. (2009), Effects of bilateralism and the MFN clause on international trade: Evidence for the Cobden-Chevalier network, 1860–1875. *The Journal of Economic History* 69(4): 1012–1040.
- Lampe, M. (2011), Explaining nineteenth-century bilateralism: economic and political determinants of the Cobden-Chevalier network. *The Economic History Review* 64(2), 644–668.
- List, F. (1841), *Das nationale System der politischen Oekonomie*. J. G. Cotta’scher Verlag. The English translation cited by Viner (1950, page 94 f.) is: *The national system of political economy*, translated by S. S. Lloyd, New York (Longmans, Green and Company), 1904.
- Machlup, F. (1977), *A history of thought on economic integration*. New York: Columbia University Press.
- Meade, J. E. (1955), The theory of customs unions. No. [1] in: *Professor Dr. F. de Vries lectures*. Amsterdam: North-Holland.
- Moura Filho, H. (2006), *Pioneering multilateralism: the sugar agreements 1864–1914*. MPRA Paper No. 5208.

- Nebenius, C. (1833), *Denkschrift für den Beitritt Badens zu dem zwischen Preussen, Bayern, Württemberg, den beiden Hessen und mehren andern deutschen Staaten abgeschlossenen Zollverein*. Verlag Müller, Karlsruhe.
- O'Brien, D. (1976), Customs unions: Trade creation and trade diversion in historical perspective. *History of Political Economy* 8(4): 540–563.
- Pahre, R. (2008), *Politics and trade cooperation in the nineteenth century: the 'agreeable customs' of 1815–1914*. Cambridge University Press.
- Panagariya, A. (2000), Preferential trade liberalization: The traditional theory and new developments. *Journal of Economic Literature* 38(2): 287–331.
- Pentmann, J. (1917), Die Zollunionsidee und ihre Wandlungen im Rahmen der wirtschaftspolitischen Ideen und der Wirtschaftspolitik des 19. Jahrhunderts bis zur Gegenwart. No. 27 in: *Probleme der Weltwirtschaft* 27. Jena: Fischer.
- Ploeckl, F. (2010), *The Zollverein and the formation of a customs union*. Oxford University Economic and Social History Series 84.
- Pomfret, R. (1988), *Unequal trade: The economics of discriminatory international trade policies*. Oxford and New York: Blackwell.
- Posse, H. E. (1928), Meistbegünstigung. *Jahrbücher für Nationalökonomie und Statistik* 129(5): 672–698.
- Roscher, W. (1870), Zur Gründungsgeschichte des deutschen Zollvereins. Deutschland (143–211), The article was also published as a separata by the publisher Stilke & Muyder.
- Sapir, A. (2001), Domino effects in Western European regional trade, 1960–1992. *European Journal of Political Economy* 17(2): 377–388.
- Schmölders, G. (1954), Der deutsche Zollverein als Vorbild einer wirtschaftlichen Integration in Europa. Pp. 137–148 in: *Aspects financiers et fiscaux de l'integration economique internationale*. Cambridge and London: Van Stockum.
- Tena-Junguito, A., M. Lampe, F. T. Fernandes (2012), How much trade liberalization was there in the world before and after Cobden-Chevalier? *The Journal of Economic History* 72(3): 708–739.
- Treitschke, H. (1897), *Deutsche Geschichte im 19. Jahrhundert*, vol. IV. Berlin.
- United Nations (1947), Customs Unions. No. 1948, 3 in *United Nations Publications II, D: Trade, financial and commercial studies*. Lake Success, NY: United Nations.
- Viner, J. (1931), The most-favored-nation clause. Index, VI, 61. Published by Svenska Handelsbanken. Pp. 262–273 in: K. Anderson, B. Hoekman (1994), *The global trading system*. Volume 2, Core rules and procedures. Reprint.
- Viner, J. (1950), The customs union issue. No. 10 in *Studies in the administration of international law and organization*. Anderson Kramer Associates.
- Weber, W. (1869), *Der deutsche Zollverein: Geschichte seiner Entstehung u. Entwicklung*. Veit & Comp, Leipzig.
- WTO (2011), *World Trade Report 2011 – The WTO and Preferential Trade Agreements: From Co-existence to Coherence*.
- Zollvereinsblatt (1846), Discussion zwischen dem Rheinischen Beobachter und der Allgemeinen Zeitung über die Frage: ob bei der Auflösung des Zollvereins die süddeutschen Staaten gewinnen oder verlieren würden? Vol. 4: 74–77.
- Zwiedineck-Südenhorst, O. v. (1926), Das Problem eines deutsch-österreichischen Zollvereins. *Jahrbücher für Nationalökonomie und Statistik* 124(5): 385–402.
- Zwiedineck-Südenhorst, O. v. (1935), Zur Entstehungsgeschichte des Deutschen Zollvereins. *Jahrbücher für Nationalökonomie und Statistik* 142(1): 25–34.

Prof. Dr. Switgard Feuerstein, Department of Economics, University of Heidelberg, Bergheimer Straße 58, 69115 Heidelberg, Germany, and GER, University of Nottingham.  
Switgard.Feuerstein@awi.uni-heidelberg.de

# The Relevance of Census Results for a Modern Society

Roderich Egeler, Natalie Dinsenhacher, and Birgit Kleber\*

Statistisches Bundesamt, Wiesbaden

JEL J10

Census; register-based census; population; system of national accounts; gross domestic product (GDP); fiscal policy; administrative register.

Received: 22.03.2012

Revision received: 27.11.2012

Accepted: 22.01.2013

---

## Summary

Population censuses have been conducted for thousands of years and they are one of the bases of any official statistics. By giving a detailed overview of the structure of a country's population at a specific point in time, they provide one of the most important parameters for political action.

In addition, in modern statistical systems, population censuses are the data basis for a variety of further statistical information, such as intercensal updates of the number of inhabitants. They provide the basis for the statistical fields which either describe the population or in which population data are integrated. Furthermore they allow drawing reliable population samples and performing qualified extrapolation of sample results.

This article focuses on the relevance of censuses in general and, more specifically, of the current 2011 Census in Germany. Following a historical overview in the first part of the article, the second part describes the various definitions of "population", the relevance and issues of defining the survey unit and of operationalising concepts of variables. In the third part, some examples illustrate various areas that are relevant for statistics and social policy and for which the 2011 Census will provide information, while the fourth part provides an outlook on how the census can be developed further, both in terms of methodology and contents.

## 1 Census: yesterday and today

Ascertaining the number of inhabitants or population data by population censuses has a long tradition. As early as in 2255 BC, a population census was held in China – the oldest census known today (Michel 1985: 79). Counts of the population are also documented for the ancient world<sup>1</sup> (Michel 1985: 79). However, the motivation underlying those censuses was much more specific than it is today. From demographic aspects, the surveys

---

\* We are particularly grateful to Dr. Sabine Bechtold and Dr. Joachim Schmidt. Their suggestions and ideas have truly enriched this manuscript.

We thank all our colleagues of the federal statistical office who provided us with their professional advice to individual chapters of our contribution.

<sup>1</sup> E.g. Persia, Egypt and Greece.

conducted at the time were incomplete because they were not systematically designed and often covered only part of the population; censuses in antiquity were conducted specifically for purposes of military planning and taxation (Daszynska 1896: 486 ff.), e.g. the count in the gospel according to Luke (Luke 2, verses 1–5) was conducted for taxation purposes in contrast to the scope today. Today's censuses count all people living in a defined regional area for general political, administrative and scientific purposes, while data are made fully anonymous and are subject to statistical confidentiality.

Modern population statistics as a separate science emerged in the 19<sup>th</sup> century (Michel 1985: 79). At the time, the first censuses were conducted in nearly all European countries. One of the main factors here was technological progress. For example, punch card technology was applied in data processing for the first time in the late 19<sup>th</sup> century and remained common practice in Germany until the 1970 population census (Grohmann 2008: 79). Since 1816, there have been population censuses at more or less regular intervals in the German states. When the German *Reich* was founded in 1871, they were largely standardised – at first based on the provisions of the *Deutscher Zollverein*<sup>2</sup> – and were held at five-year intervals. In the 20<sup>th</sup> century, population censuses in Germany were conducted only irregularly, which in the first half of the century was due to the two world wars.<sup>3</sup> In the second half of that century the period between two censuses increased from just five or six years to almost a quarter of a century in the old *Länder* and to 30 years in the new *Länder*.<sup>4</sup>

The official population figures were based on intercensal updates. These were performed since the 1987 population census for the territory of the old *Länder* and Berlin-West. Intercensal updates for the new *Länder* and Berlin-East were based on the counts of the central population register of the former GDR. The years between the latest population censuses and the 2011 Census were however characterised by serious political change – German unification, civil wars in the Balkan countries and ensuing flows of refugees, and the EU enlargement involving freedom of movement. These are just a few keywords in this context. All those events led to considerable population flows, which were represented by intercensal population updates.

However, the method of continuously updating the current population of every single municipality by means of intercensal population updates requires regular adjustments. Experience shows that occasional errors occur in processing the monthly reports on natural population change and migration which the statistical offices receive from the residents' registration offices. Over the years, these errors may add up to considerable inaccuracies.<sup>5</sup> It is therefore standard practice at the international level to conduct a census every ten years in order to put intercensal population updates on a new basis and, consequently, to assure the quality of population data in the long term.

<sup>2</sup> *Deutscher Zollverein*: Federation of German states for customs and trade purposes.

<sup>3</sup> In the Federal Republic of Germany, four population censuses were conducted until 2010, that is, in 1950, 1961, 1970 and 1987, see *Statistisches Bundesamt* (1990: 41).

<sup>4</sup> In the German Democratic Republic, a total of four population censuses were conducted at quite irregular intervals: 1950, 1964, 1971 and 1981, see *Staatliche Zentralverwaltung für Statistik* (1955: 8); *Staatliche Zentralverwaltung für Statistik* (1968: 3); *Staatliche Zentralverwaltung für Statistik* (1975: 1); *Staatliche Zentralverwaltung für Statistik* (1982: 1).

<sup>5</sup> Errors may already occur in the original data, for example, in the population register data as a result of the registration behaviour of the population, e.g. when people do not de-register after moving away, or during processing in the administration. They may also arise from the specific conditions of intercensal population updates, for example, in the course of eliminating negative cases.

A characteristic feature of a census is that, while the number of variables covered for the survey units is very limited, they are covered in a detailed regional breakdown. The main criteria applied to select the survey variables are relevance, effectiveness, and efficiency (*Zensuskommission* 2009: 3). As regards relevance, the variables to be covered in a census must be defined by way of legislation. The decisive question here is whether they are directly required for government action by the central, regional or local authorities or whether they are part of the statistical infrastructure, which is crucial for the tasks of other institutions. The criterion of effectiveness serves to check whether a variable classified as relevant should really be covered in a census – that is, in a detailed regional breakdown – or whether it would be more appropriate to use a different tool. Finally, it has to be verified for every variable whether coverage in a census is efficient, that is, whether there is a reasonable relation between the benefit of the information content of the census variable and the costs. The German legislative bodies have decided that the 2011 Census should go just slightly beyond the European Union data requirements (OJ L 218, 14). The variables additionally included in the list of variables defined by the EU are “legal affiliation with a religious society under public law” and “adherence to religions, persuasions or beliefs”, and the coverage of the variable “migrant background” is based on a definition that is broader than required by EU provisions.

For the 2011 Census – following a methodological test (*Statistische Ämter des Bundes und der Länder* 2004: 813 ff.) – a new, register-based method for the population census is applied for the first time in Germany: existing register data are used in combination with results of various primary surveys (*Statistisches Bundesamt* 2011: 4). While the housing census has also been modified but conducted as a full enumeration, the census of non-agricultural local units, which had generally been a component of previous population censuses in Germany, is nowadays conducted within the scope of business statistics. To be able to implement the complex mixed-mode method of the 2011 Census, a register of addresses and buildings was set up as a first step. It contains all addresses in Germany at which residential space is available. It is used, first, as the statistical population underlying the census of buildings and housing and, second, as a sampling frame from which the sample for the household survey was drawn. In addition, the register of addresses and buildings is used as a central basis for organising all survey components of the 2011 Census. As described, the data of the 2011 Census are obtained from the following components (cf. Chart 1):

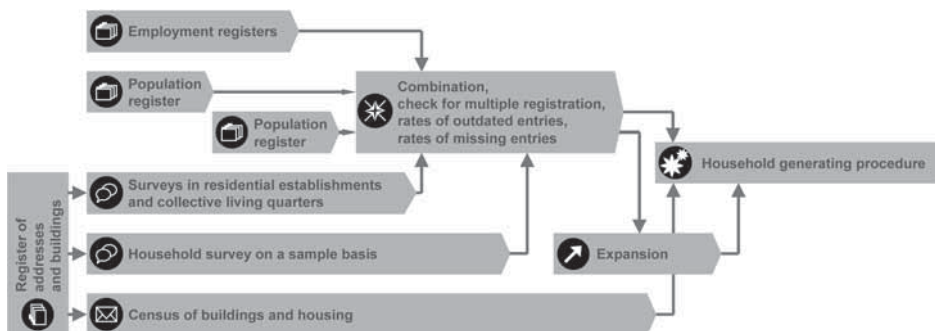


Chart 1 The census model

1. Register data of the residents' registration offices are the core data used to ascertain both the number of inhabitants and the data on the demographic structure of the population living in private households. They are combined into a data set for the whole of Germany and are checked for any incorrect duplicate cases<sup>6</sup>. To meet the requirements, the residents' registration offices had to deliver their data three times. It is important to mention that the checked data are only used for statistical purposes – they will not be retransmitted to any public authority.
2. For residents of collective living quarters, institutions, residential establishments and similar facilities, population register data involve too many errors, so that data on these people are ascertained through a complete enumeration.
3. At the level of individuals, register data of the Federal Employment Agency and register data on public service personnel complement the demographic variables obtained from the population registers and from the survey conducted at special facilities. For most of the people covered by the register data from the residents' registration offices, data are available on the place of work, employment status, economic branch and occupation; however, information on self-employment cannot be derived from the registers used. This information was collected in the sample survey described in the following.
4. A sample survey conducted among just under 10 percent of the population assures the quality of the population register data by ascertaining, for every municipality, the rates of outdated entries and missing entries and by taking them into account when determining the number of inhabitants of the municipalities.<sup>7</sup> The survey also collects additional data on the population that cannot be obtained from register data.<sup>8</sup> This includes information on employment, education, migrant background and religious affiliation. The data are collected directly by interview.
5. Data on the number and structure of residential buildings and dwellings are collected by conducting a postal survey among all owners because the exact number of residential buildings is not known in Germany. To reduce the burden on the population, only the owners, managers or other parties entitled to use or dispose of a building or dwelling are questioned.
6. Finally, the household generating procedure allows combining personal data into residential households by linking the population register data, which include information on family relationships, to dwellings data from the census of buildings and housing.

This census method ensures complete coverage of the target population<sup>9</sup>.

---

<sup>6</sup> Incorrect duplicate cases may refer to persons registered with their sole residence in several municipalities.

<sup>7</sup> That method of assuring the quality of population register data is applied for municipalities with 10,000 inhabitants and over. In municipalities with fewer than 10,000 inhabitants, quality assurance is achieved by conducting a survey to clarify discrepancies and a survey to clarify the residence.

<sup>8</sup> The sample design and the extrapolation model for the household survey were developed in a specific project by Trier University and GESIS – Leibniz Institute for the Social Sciences in Mannheim. Münnich et al. (2012).

<sup>9</sup> Covered are inhabitants at their main or second residence; however, the official number of inhabitants is calculated only on the basis of the population at the main or sole residence.

## 2 Problems of adequation and operationalisation in a census

In German, the term adequation refers to a number of different issues dealing with the correct delimitation of the survey population and the statistical variables describing it as well as with the determination of suitable concepts of measurement. It also refers to the challenges (faced by a data producer) regarding the planning and conduct of statistics and the comparability of the data, that is, the exact definition and the correct collection of the data.

*Adequation* and *operationalisation* are terms which have had a lasting impact on German official statistics. The challenges involved are illustrated in the following, taking as examples the term “population” and the employment concept.<sup>10</sup>

### 2.1 The term “population”

Population data for today’s Germany (as a whole) – based on an all-German census and subsequent updates – were not available until the 2011 Census. However, the boundaries of constituencies, the number of seats of any *Land* in the *Bundesrat*, the horizontal, vertical and municipal equalisation of revenue<sup>11</sup> – this list mentions just a few major examples – are strongly affected by reliable data on the population and its spatial distribution. The mission of official statistics is to provide and disseminate statistical information which is used e.g. for decision-making regarding political and economic problems. But what is *the* population?

The question of who should be counted as part of the population of a country or of another regional unit cannot necessarily be answered through the term “population” or “inhabitant”. The term “population” has a functional meaning and, consequently, depends on the concrete task at hand. When, in the years before the foundation of the German *Reich* in 1871, the *Deutscher Zollverein* attempted to set up a single market among the German states, the customs revenue had to be distributed among these states, which was done on the basis of the number of inhabitants. For the purpose, the *Deutscher Zollverein* created the term *Zollabrechnungsbevölkerung*, thus establishing a first generally accepted definition and initiating the efforts to harmonise the methodology of population censuses (Michel 1985: 82). The *Zollabrechnungsbevölkerung* comprises all persons staying permanently in the country (and, consequently, being relevant for the consumption of goods). To ascertain the *Zollabrechnungsbevölkerung*, a population census was conducted in the German states every three years. In 1863 the term was defined more precisely, specifying that persons who were absent from their place of usual residence for more than a year did not belong to the *Zollabrechnungsbevölkerung*. Later the *Deutscher Zollverein* extended its survey of the population to cover the *ortsanwesende Bevölkerung* (Scheel 1869: 157) on the census reference date because they participate in economic life, too (Michel 1985: 83 ff.). According to the concept of *ortsanwesende Bevölkerung*, the inhabitants are counted at, and allocated to, the place at which they are present on census day.

The question of what population censuses should be used for is decisive for defining the universe of such a survey. This becomes apparent when looking at the early volumes of

<sup>10</sup> For more information on the issue of adequation see also: Flaskämper (1931: 379 ff.); Egeler et al. (2012: 269 ff.); Grohmann (1988: 25 ff.).

<sup>11</sup> Equalisation of tax revenue between the German *Länder* and the Federation. This fiscal strategy is a fundamental contribution to greater fairness in Germany.



the Journal of Economics and Statistics. Two articles have been chosen here to illustrate the issue when talking about the population of a country. In the sixth volume of the Journal of Economics and Statistics published in 1866, Meyer<sup>12</sup> and Fabricius<sup>13</sup> explain what definition of the term “population” is best suited for what type of evaluation. In his paper on *Das Princip der rechtlichen und der factischen Bevölkerung*, Meyer advocates ascertaining the population according to three definitions:

- de jure population – people having the right of residence (*Indigenat*);
- de facto population – people present at the relevant place at the time of the census;
- population with permanent residence – people spending most of the time at the relevant place (Meyer 1866: 97 ff.)

The decision as to which definition is taken as a basis of a population census has an impact on the analytical potential of the survey. The intention behind a definition of population was – and still is – to provide a solid basis for decision-making, which means it is necessary to count the population staying at a specific place at a given point in time, irrespective of the duration of stay (Meyer 1866: 98 ff.). When opting for that definition, however, the question arises who should be included. Especially where that term does not include a time dimension regarding the people’s presence, the question arises how “present” is defined: permanent and regular or simple presence at the time of a census?

Meyer, and also Fabricius, summarise that, depending on the statistical issue to be examined, one of the three definitions of population is to be preferred. However, to cover the population and its structure and to derive political decisions, as is allowed by a population census, the definition of the population with permanent residence is to be used. This definition covers persons with residence according to a fixed criterion such as at least six months at the relevant place: It is the only basis allowing planning and calculation (Meyer 1866: 106 ff.). Nowadays this criterion is met by the use of the administrative register, which is based on the German Basic Registration Law (*Melder-rechtsrahmengesetz*). Here some aspects are taken as examples to illustrate that the argumentation used at the time is still topical and consistent today.

The application of the definition of the permanently present population is reflected also today in the 2011 Census. In its requirements for the 2011 Census and the coming census rounds, the EU has agreed on the definition of the “population at its usual residence at the reference date” (OJ L 218, 14 [15]). Even today, agreeing on this definition of the population is not a matter of course. Even the Conference of European Statisticians, at its meeting organised by the United Nations Economic Commission for Europe (UN ECE) in 2006, and in the recommendations subsequently adopted for the census round 2010/11, did not succeed in agreeing on one concept of “usual residence” (United Nations 2006: 35). The “usual residence” has been defined by the EU to be the “place where a person normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage.” (OJ L 218, 14 [15]). In short, the place of usual residence for the EU population is where a person’s own bed is.<sup>14</sup>

<sup>12</sup> Georg Meyer (1841-1900), lawyer and politician, public official at the statistical office in Jena.

<sup>13</sup> Karl (Carl) August Fabricius (1825-1890), member of the founding body of the *Hessische Zentral-stelle für Landesstatistik* in Darmstadt in 1861.

<sup>14</sup> For further information on statistical innovations especially information for a better understanding of the development in economic statistics and in this context the relevance of the definition of the term “population” in Germany in the first half of the 20<sup>th</sup> century, see also: Tooze, J. A. (2001), *Statistics and the German State, 1900-1945*, Cambridge.

The requirements regarding the operationalisation of the definition specify that those persons have to be counted as inhabitants of the relevant place or the relevant member state who have permanently lived at that “usual residence” for at least twelve months prior to the survey date or who arrive there with the intention of staying there for at least twelve months. Due to the difficulties of operationalising<sup>15</sup> that definition, Germany has made use of the derogation mentioned in Regulation (EC) No 763/2008, Article 2 (d). It specifies that the legal or registered residence can be considered as “usual residence”. For the 2011 Census, the information on the housing status has been used as shown in the population register – sole residence, main residence or second residence.<sup>16</sup> Applying that variable ensures that all persons participating in the economic, social and political life of the society are taken into account, similar to the definition of permanent residence.

Apart from defining the term “population”, other issues of adequation and operationalisation have to be settled for every census. What definitions should be taken as a basis to describe the population in terms of employment statistics or education statistics? What concepts are used to show the migrant background and the religious affiliation, which have been covered in the 2011 Census in Germany? As an example of the questions to be answered to solve such problems, the employment concept underlying the 2011 Census will be explained here.

## 2.2 Employment concept

In the 2011 Census, the employment data had to be collected according to the concept of the International Labour Organization (ILO). This means that all persons aged 15 or over are considered as employed if they worked for remuneration or as self-employed or family workers for at least one hour in the reference week (*Statistisches Bundesamt* 2008: 110).

Employment data are obtained from three different sources in the 2011 Census: As mentioned above, the employment status of employees subject to social insurance contributions and of all people registered as unemployed or as seeking a job is obtained from register data of the Federal Employment Agency. As regards public officials, judges and soldiers, register data of the public employers are used in addition. For the remaining persons employed, employment status data have to be collected in the household survey on a sample basis, and extrapolated subsequently. For that specific purpose, a decision had to be taken on the concept of questions to cover the employment status according to the ILO concept.

Generally, the questions are designed in line with the microcensus to ensure that the census results are comparable with that set of statistics (Gauckler 2011). However, experience has shown that covering smaller and marginal activities performed for

<sup>15</sup> Operationalisation problems consist, first, in the fact that quality losses have to be expected when covering “intentions of planned duration of stay”. Second, it is difficult to ascertain the place of usual residence of a person living for less than twelve months at the place. This is due to the registers used and their structure.

<sup>16</sup> In section 12 (2) of the German Basic Registration Law, to ascertain the main residence, a distinction is made between married persons/persons living in a registered same-sex partnership who do not permanently live separated from the family or the partner, and any other persons. For the first group of persons, the main residence is the dwelling mainly used by the family or the partners. For the second group of persons, the main residence is the dwelling mainly used by the relevant person.

remuneration is difficult in household surveys.<sup>17</sup> Based on a methodological study<sup>18</sup> carried out in 2008 on the quality of employment data obtained from the microcensus<sup>19</sup>, the microcensus concept of questions was revised in 2011 to achieve better coverage of smaller jobs (Gauckler 2011: 720 ff.). After adjustment to the specific framework conditions, the new concept of questions was applied also in the 2011 Census.<sup>20</sup>

The problem of covering smaller jobs consists mainly in the difference between the respondents' common understanding of specific terms and the definitions of the ILO concept. Tests have shown that in particular the above-mentioned limit of one working hour per week does not coincide with the respondents' common understanding of "employment". The methodological microcensus studies have clearly shown that many respondents take their main social status as a basis for orientation. This means that, for example, students consider especially their main status and, consequently, they do not realise that a small part-time job they may have is also to be regarded as employment (Köhne-Finster/Körner 2008: 2; Gauckler/Körner 2011: 197 ff.).

Consequently, the 2011 Census asked first of all for the main employment status, while smaller activities and side jobs were then covered through specific follow-up questions. A quantitative pretest of the household survey questionnaire demonstrated that this allowed better representation of the number of persons in employment than the concept of microcensus questions applied until 2010. However, even with the new concept of questions, there is still a difference compared with the employment registers of the Federal Employment Agency (Gauckler/Körner 2011: 196 ff.; Körner et al. 2011: 1072 ff.).

### 3 Current relevance of the census – Knowing what will count tomorrow

Since the times of the *Deutscher Zollverein*, the methodology, the relevance and the content of population censuses in Germany have changed. The usability of the results has become much more varied. It ranges from policy issues and economic aspects to social themes. In the statistical field, too, census results are used in manifold ways.

#### 3.1 Providing neutral and robust data

##### • Intercensal population updates

Between two censuses, the official numbers of inhabitants in Germany are updated on the basis of the official number of inhabitants ascertained in the latest census. Such intercensal population updates are compiled for every single municipality and based mainly on the numbers reported on births and deaths as well as on arrivals and departures across municipal borders. Over the longer term, this leads to increasing discrepancies, especially on the local level, so that it is very important to regularly

<sup>17</sup> See the overview in Körner et al. (2011: 1065–1085).

<sup>18</sup> For a systematic study of the problems of covering the ILO employment concept in the microcensus see also Köhne-Finster/Körner (2008: 3 ff.).

<sup>19</sup> The microcensus provides official representative statistics of the population and the labour market in Germany, thereby bridging the data gap between two population censuses. The microcensus is based on a sampling fraction of 1 % of the population and is designed as a multi-topic survey.

<sup>20</sup> For example, laptop interviews – which are common in the microcensus – were not carried out. Therefore, due to the different framework conditions, it was not possible to implement the microcensus questions on a one-to-one basis.

put the existing population figures on an up-to-date basis by conducting a census. This becomes obvious when looking at the cohorts of people aged 90 years or over, which have been updated on an intercensal basis since the 1987 population census. In official statistics, those cohorts are strongly overestimated where only intercensal updates are used. This is because, over a longer period, the inaccuracies in the population registers accumulate, especially those regarding arrivals and departures. People who moved away without de-registering with their registration office continue to be covered by intercensal updates. Such inaccuracies become obvious especially for older cohorts with smaller numbers of people. Using pension insurance data can yield higher accuracy (Scholz/Jdanov 2006). When new results of the 2011 Census are available, it will be possible again to show age groups of over 90 years without having to apply estimation methods and to compile a general life table (Eisenmenger/Emmerling 2011: 226).

In the 1987 population census, a difference of “just” 75,000<sup>21</sup> people between census results and intercensal updates was ascertained for the whole Federal Republic of Germany. A look at the regional distribution of the result shows, however, that the population in five of the eleven *Länder* was higher, and in six *Länder* lower than the figures of intercensal updates. The differences ranged between –2.3 % and +6.5 % (*Statistisches Bundesamt* 1995: 23). Adding up those differences separately at the municipality level shows that the intercensal updates overestimated the number of inhabitants by some 900,000 and underestimated them by about 830,000 people. This is a total difference of some 1.7 million people.

The census test carried out in 2001/2002 indicated that intercensal population updates overestimated the number of inhabitants in Germany at the time by some 1.3 million, with the extent of the error differing considerably between regions (*Statistische Ämter des Bundes und der Länder* 2004: 813 ff.). The question of how the assumed overestimation of the number of inhabitants developed until 2011 – for example, whether the tax identification number introduced since 2007 has led to adjustments in the population registers, which would then have improved intercensal population updates – can be answered only when the 2011 Census results have become available.

#### • Basic data for sample surveys

For all kinds of statistics, census results are used as a frame for both sampling and extrapolation in sample surveys. Only updated census results make it possible to extrapolate results of sample surveys and to obtain data on the universe (Wagner 2010: 4). This applies to all sample surveys based on the population distribution and structure, irrespective of whether they are conducted by statistical offices, scientific or commercial institutions.

When examined from a statistical-methodological aspect, census results play a major role in two respects. First, the updated figures lead to an adjustment of the sample systems. For example, sample designs have to be revised to adapt them to changes regarding the regional distribution, the age and sex structure as well as further demographic or other variables. Second, future extrapolations and weightings will then have to be based on the new data. Such adjustments ensure the representativeness and reliability of future sample surveys.

One of the central sample surveys for official statistics on the population and the labour market is the microcensus. It both bridges the data gap between two population censuses

<sup>21</sup> Values rounded.

and provides a wealth of information that goes beyond census results.<sup>22</sup> As in any sample survey, the quality of the microcensus results is mainly determined by the quality of the sampling frame. The results of the 2011 Census in Germany will provide new sampling frames not only for population samples but also for samples of buildings and housing.

### • Monitoring the economic development

The national accounting system of the Federal Statistical Office, which describes in quantitative terms the economic development in Germany for a past period, generally uses the entire range of official statistical data as a basis. Economic statistics are used most often, but population statistics play a major part, too. However, national accounts apply some specific definitions, for example, of the term “population”. According to the international concept of national accounts, the population comprises all persons – citizens or foreigners – who are resident in a country’s economic territory, in this case the Federal Republic of Germany. People are considered as resident in the country if they have had their permanent residence in Germany for at least a year. This includes people who are resident in the country but are staying abroad temporarily, i.e. for a period of less than a year. Not included are persons staying temporarily in the country such as non-residents working in the country or tourists and foreign students.<sup>23</sup> The stock of the population defined in this way is shown in national accounts as an annual and quarterly average. Correct “per inhabitant” or “per capita” reference values can be calculated. The quarterly and annual averages of the population in national accounts are based on continuous intercensal population updates, which cover the population in Germany according to the concept of the population at the place of sole or main residence.<sup>24</sup> Consequently, readjusting the population figures by current data obtained from a census has a considerable impact on the quarterly and annual averages of national accounts. This applies in particular to the per capita GDP, which is one of the most important indicators and often directly used for international comparisons of economic performance. A reliable per capita GDP is also of major importance for compiling composite indicators such as the Human Development Index of the United Nations (United Nations 2011) as part of the international discussion on measuring well-being.

If the number of inhabitants is considered a central criterion for measuring differences in size between regions, per capita GDP data are generally well suited as a structural indicator. The Statistical Office of the European Communities (Eurostat) includes that indicator<sup>25</sup> when evaluating to what extent the Lisbon objectives adopted by the European Council in 2000 have been achieved (Eurostat 2011). However, a methodo-

<sup>22</sup> The microcensus results are used for government reports, the annual reports of the Council of Economic Experts and the annual pension insurance report of the Federal Government, and they are the basis for continuous employment research, etc.

<sup>23</sup> In his definition published in 1866, Meyer subsumes that group of persons under the term “population with permanent residence”, defining it as the group of persons suitable for deriving the country’s production capacity and the people’s wealth (Meyer 1866: 107). In addition, according to Meyer, that definition is the only one suitable for the statistical calculations regarding the production capacity and wealth of a society. This is due to the fact that the work of persons not staying permanently in Germany is not taken into account when calculating production figures.

<sup>24</sup> However, the population concept applied in German national accounts differs from the provisions of the European System of National and Regional Accounts (ESA) 1995, especially as regards the inclusion of persons staying temporarily in the country and of “residents” staying temporarily abroad (short-time migrants, students, etc.).

<sup>25</sup> It is measured in purchasing power standards to compensate for price differences between countries.

logical shortcoming should be taken into account especially for small-area comparisons: the calculation of the GDP is based on the domestic concept, while the resident population is based on the national concept. Allocation to the place of work or the place of residence is done accordingly. Due to the surplus of people commuting to conurbations, the per capita GDP tends to be too high in such areas, whereas it tends to be too low in regions where those people live.<sup>26</sup> A regional indicator taking account of the “bias” caused by commuter flows is the GDP per capita of the *Wirtschaftsbevölkerung*<sup>27, 28</sup>. Up to the late 1980s, the *Wirtschaftsbevölkerung* was used quite frequently in the Federal Republic of Germany to standardise GDP results by *Kreis* (administrative district).

Changes in the description of the economic development that are caused by updated population figures will also have an impact on political action based on that description. To what extent that may happen is shown by the reactions in various political fields to the 1987 population census.

### 3.2 Impact of data changes on individual political fields

- Fiscal policy and equalisation of revenue

Changes in population data will change the financial flows of the equalisation-of-revenue system. For the compensation in 2010, the equalisation of revenue between the *Länder* led to equalisation payments made/received of Euro 7.0 billion (Section 2 of the 2nd Ordinance implementing the Financial Equalisation Act in 2010). The extent to which that can change as a result of updated figures is shown by the adjustments made after the 1987 population census. At the time, the total of equalisation payments was corrected by about DM 908 million<sup>29</sup> (*Deutscher Bundestag* 1990). For example, after the correction Schleswig-Holstein received DM 159 million less, while Nordrhein-Westfalen was granted DM 100 million more.

At the EU level, too, the number of inhabitants of the member states has an influence on economic and fiscal policies. The European Union pursues the goal of reducing differences in development between the member states in order to strengthen the social coherence and economic competitiveness of the EU and to balance the living conditions in the individual states. Between 2007 and 2013 the EU made an amount of Euro 347 billion available to the regions of the so-called NUTS-2 level<sup>30</sup> (*Europäische Gemeinschaft* 2007: 24). Distribution of the funds for “eligibility under the Convergence objective” is based on the gross domestic product (GDP) per capita of that regional

<sup>26</sup> Apart from the commuter problem, an issue to be taken into account at the regional level when comparing intercensal population updates over time is the changes that occur due to territorial changes. For example, the comparability between results by municipality or *Kreis* (administrative district) in a *Land* and data of previous years is limited where smaller municipalities were incorporated into bigger ones or where a territorial reform was performed in the reference year. As changes generally are rare for larger regional units (e.g. *Länder*), comparability over time is good at that level.

<sup>27</sup> *Wirtschaftsbevölkerung*: resident population adjusted for two-way commuting.

<sup>28</sup> In contrast to the resident population according to intercensal updates in a region, the statistical artefact of *Wirtschaftsbevölkerung* allocates the persons in employment (including family members) to the region of their place of work, irrespective of their actual place of residence.

<sup>29</sup> About Euro 464 million.

<sup>30</sup> NUTS = nomenclature des unités territoriales statistiques; in Germany, this corresponds to *Regierungsbezirke* (administrative regions).

classification. The number of inhabitants in a detailed regional breakdown is highly important also in this context (*Europäische Gemeinschaft* 2007: 6 ff.).

#### • Labour market policy

As described in section 2.2, the 2011 Census collects data on the people's labour force participation according to the international labour force concept. The census results thus allow representing the labour market in Germany, including structures of persons in employment. What corrections may consequently become necessary after the 2011 Census is illustrated by the decisions taken on the basis of the 1987 population census results.

In 1987, in national accounts, the population census results led to an upward correction of the number of persons in employment by about one million and, consequently, to a revision of the assessment of the labour market trends in the 1970s and 1980s (*Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung* 1990: 104). Based on the corrected figures, the reference data required for unemployment rates were recalculated. Especially the regional unemployment rates changed considerably. In about a third of the labour office districts of the former territory of the Federal Republic, they were corrected downwards by 20 percent and more (*Statistisches Bundesamt* 1995: 24 ff.).

An analysis of the numbers of persons in employment by economic sector for 1989 illustrated the ongoing structural change. The sectors of industry, trade and transport, service enterprises, general government, households, etc. recorded an increase in the number of persons in employment, with the increase in industry remaining below the average for the overall economy. The service enterprises sector accounted for almost half of the volume of corrections required. In agriculture and forestry, the number of persons in employment had decreased (Schoer/Hanefeld 1990: 251 ff.).

#### • Housing policy

Other information to be collected by the 2011 Census in Germany, apart from the official number of inhabitants and the population structure, is data on the stock and structure of dwellings and residential buildings. This is done in a special census component, the census of buildings and housing, which is designed as a traditional census, with one important difference: the questionnaire was only sent to the owners of buildings, instead of the usual approach of directly questioning every household. As in the past, the results of the census of buildings and housing will provide current information on the distribution and structure of residential space and vacant dwellings as well as on the supplier structure. In addition, they will be an important basis for decision-making in terms of housing policy and regional planning. The results of the 1987 census of buildings and housing showed that there were no reserves in terms of housing supply in the Federal Republic of Germany at the time because the number of dwellings had been overestimated by about 1 million through the intercensal dwelling updates. Due to the results of the 1987 census of buildings and housing, the situation had to be reassessed by politicians and the construction industry.

As results of the census of buildings and housing are linked to other census variables on persons and households, municipalities are able to compile solid forecasts of housing demand – based on the specific demographic structure of each municipality. So the results of the census of buildings and housing provide both important information for highly specific market monitoring and an updated basis for small-area urban monitoring.



### • Shaping legislative structures

The Federal Elections Act requires that the number of inhabitants<sup>31</sup> of a constituency should not differ by more than 15 % upwards or downwards from the average number of inhabitants of all constituencies. Where the difference exceeds 25 %, the constituency has to be reshaped (Section 3 (1) no. 3 of the Federal Elections Act).

Up to now, the delimitation of constituency boundaries has been based on intercensal updates of the 1987 population census and on the data of the central population register of the GDR, whose basis was the population census of 1981. Readjusting the official number of inhabitants by the 2011 Census will provide the basis for the activities of the Constituency Commission. At the beginning of any electoral period, the commission has to report on how the population in the electoral territory has changed, and it explains whether and, if so, what changes in the delimitation of constituency boundaries are required. The information obtained from the 2011 Census will be used in the delimitation of constituencies at the federal level for the first time for the election to the 19<sup>th</sup> German *Bundestag* in 2017.

The number of inhabitants also has an impact on the distribution of seats in the *Bundesrat*. Each *Land* is represented there and has a specific number of votes according to its number of inhabitants. When the official number of inhabitants has been ascertained by the 2011 Census, it will reveal whether a *Land* will lose in importance in the *Bundesrat*. Only for *Hessen*, the number of inhabitants as obtained by intercensal updates is currently close to one of the thresholds defined in Article 51 (2) of the Basic Law<sup>32</sup> (the German constitution), which are decisive for the number of votes in the *Bundesrat*.

For the distribution of seats in the European Parliament, too, the number of inhabitants is relevant. Here, the “degressive proportionality” principle applies, meaning that more populous countries have more seats, although the number of seats granted is relatively smaller than the number of inhabitants. Smaller countries benefit from the provision that every member state must have at least six members in the European Parliament (*Europäische Gemeinschaft* 2007a). A German member of the European Parliament currently represents a good 825,000 citizens, but one from Malta only just under 70,000.

The 2011 Census will provide updated numbers of inhabitants for all European member states. Due to the new voting rules of the European Council applying from 2014 onwards, they will be particularly relevant. As from 2014, European Council decisions will require a qualified majority, that is, *both* the majority of member states *and* the majority of European Union citizens. The double majority is considered to be achieved if a decision is supported by 55 percent of the member states and if, at the same time, that majority represents at least 65 percent of the EU population. Every member state will then have one vote as a state and, at the same time, its number of inhabitants will be taken into account as a weight (*Europäische Union* 2011). This example shows not only how

<sup>31</sup> Foreigners are not included when the number of inhabitants is ascertained for the delimitation of constituency boundaries according to section 3 (1), seventh sentence of the Federal Elections Act. According to the decision of the Federal Constitutional Court of 31 January 2012, legislation will in the future have to take account also of the proportion of minor Germans in the German population when delimiting the constituencies.

<sup>32</sup> “Each Land shall have at least three votes; Länder with more than two million inhabitants shall have four, Länder with more than six million inhabitants five, and Länder with more than seven million inhabitants six votes.”

important an up-to-date and exact number of inhabitants is but also that it is necessary in the overall European system to apply technical terms – such as the official number of inhabitants – that are based on standard definitions.

#### **4 Challenges for future censuses**

The methodological reorientation of the 2011 Census in Germany, abandoning the traditional population census as a primary complete enumeration and introducing the large-scale use of administrative data, was a challenge for all parties involved. Although major issues had been settled in advance through a census test, actual implementation was done in large part without any tests. Experience has now shown what will have to be discussed before the next census in order to obtain the data more efficiently and to make wider use of the information contained in the census results.

A major finding is that setting up a permanent register of buildings and dwellings, combined with information from the population register, has to be the choice of the future. Also, a decision has to be taken on whether such a register should be designed as an administrative register instead of just a statistical register. The register of addresses and buildings of the 2011 Census is only used to organise the 2011 Census and will have to be deleted when those activities are finished. That approach had been planned under the assumption that permanently maintaining a register of addresses and buildings until it is used for the next census would be much more costly and time-consuming than setting up a new register immediately before the census. Judging by the experience now available, that assumption cannot be maintained. The administrative data used to set up the register of addresses and buildings were not very well suited for the purpose. The administrative data are available in exactly the way they are needed by the relevant administrative authority. Operationalising the data for statistical purposes requires enormous efforts in terms of technology, manpower and especially time. In contrast, a permanent register maintained centrally could support many standardisation processes – especially if it were maintained as an administrative register outside the secure area of the statistical offices, so that it could be used by the administrative authorities, too. This would also support other e-government projects. For the statistical system itself, such a register would allow the flexible provision of current population and housing data and be available as an up-to-date sampling frame.

The use of administrative data, which will further intensify in the future, in combination with collected data leads to highly demanding problems of adequation and methodologically very complex estimation models which create new challenges not only for the sample designs. Using such data sets is more demanding, too, because choosing suitable methods of analysis requires detailed knowledge of the sampling methodology underlying the data sets. For the societal acceptance of the results of official statistics – and this applies particularly to official numbers of inhabitants ascertained by a census – the paradigm shift from traditional population census results to census results whose accuracy depends on their sampling error poses a major difficulty. The basic principle of a traditional population census is easy to understand even for statistical laypersons, whereas a complex survey design that is demanding in terms of sampling methodology and consists of various data sources means that most users of official numbers of inhabitants are no longer able to understand how the results have been produced. The basic situation for the use of administrative registers could considerably be improved by standardisation processes. This would not only improve efficiency and create savings in the production

of statistical results from such registers but it would also reduce the complexity of surveys and, consequently, enhance their societal acceptance. In 1983, the Federal Constitutional Court demanded that this should be a permanent goal for official statistics (*BVerfGE* 65, 1 [55]).

A second major finding is that the methodological and technological possibilities available today to evaluate census data on a geo-referenced basis and, consequently, to make use of the information potential of those data according to the state of the art have not been exploited for the 2011 Census. The data of the 2011 Census will have the same spatial reference as those of the 1987 population census. Although many of the results will be available down to municipality level, it will not be possible to show all variables at the same detailed regional level, which is due to the methodology.<sup>33</sup> For all municipalities with safe statistical units, many results can be disaggregated down to block sides. Comprehensive use of geo-coordinates (geographical longitude and latitude) for spatial data analysis, which is legally regulated in other countries, is currently not legally allowed for census data in Germany. Pilot projects and checks for data usability and provision have been carried out for a few years already by the statistical offices of the Federation and the *Länder* and they have been supported by the Statistical Advisory Committee, a body giving advice in matters of federal statistics. Flexible data processing based on coordinates, however, requires either a specific legal basis for every set of statistics or the inclusion of a general provision in the Federal Statistics Law. Preparatory work is currently done here, with the future users of such data being involved. There is still an urgent need felt in this context that legal provisions be found which can still be applied to the data of the 2011 Census.

## References

- Daszynska, Z. (1896), Stoff und Methode der historischen Bevölkerungsstatistik. *Jahrbücher für Nationalökonomie und Statistik* 66: 481–506.
- Deutscher Bundestag (1990), Drucksache 11/6621, 1990.
- Egeler, R., T. Wöll, M. Zwick (2012), Perspektiven für die amtliche Statistik. *AStA Wirtschafts- und Sozialstatistisches Archiv* 5: 269–284.
- Europäische Gemeinschaft (2007), Die Kohäsionspolitik 2007–2013 Erläuterungen und offizielle Texte. Luxemburg.
- Europäische Union (2011), Die Beschlussfassungsverfahren der EU. Available at: [http://europa.eu/scadplus/constitution/doublemajority\\_de.htm](http://europa.eu/scadplus/constitution/doublemajority_de.htm) (last accessed 23 December 2011).
- Eurostat (2011), Strukturindikatoren. Available at: [http://epp.eurostat.ec.europa.eu/portal/page/portal/structural\\_indicators/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/structural_indicators/introduction) (14 December 2011, last accessed 16 January 2012).
- Eisenmenger, M., D. Emmerling (2011), Amtliche Sterbetafeln und Entwicklung der Sterblichkeit. S. 219–238 in: Statistisches Bundesamt (Hrsg.), *Wirtschaft und Statistik März 2011*. Wiesbaden.
- Fabricius, K.(C.)A. (1866), Ueber Volkszählungen. *Jahrbücher für Nationalökonomie und Statistik* 6: 305–323.

<sup>33</sup> The variables covered only in the household survey on a sample basis, such as education and employment variables, can be shown at municipality level only for municipalities with 10,000 or more inhabitants. For those variables, results for municipalities with fewer than 10,000 inhabitants and results below municipality level will not be available, except for cities with 400,000 or more inhabitants.

- Flaskämper, P. (1931), Beitrag zur Logik der statistischen Mittelwerte. AStA Wirtschafts- und Sozialstatistisches Archiv 21: 379–404.
- Gauckler, B. (2011), Die Entwicklung des Fragebogens zur Haushaltebefragung des Zensus 2011. S. 718–734 in: Statistisches Bundesamt (Hrsg.), Wirtschaft und Statistik, August 2011. Wiesbaden.
- Gauckler, B., T. Körner (2011), Measuring the Employment Status in the Labour Force Survey and the German Census 2011. S. 181–205 in: GESIS (Hrsg.), Methoden – Daten – Analysen, 2011, Jg. 5, Heft 2. Mannheim.
- Grohmann, H. (1988), Die statistische Adäquation als Postulat einer sachgerechten Abstimmung zwischen Theorie und Empirie. S. 25–42 in: K.D. Freimann, A.E. Ott (Hrsg.), Theorie und Empirie in der Wirtschaftsforschung, Schriftenreihe/Institut für Angewandte Wirtschaftsforschung Tübingen, Band 44.
- Grohmann, H. (2008), Zensus im Wandel. S. 77–102 in: G. Rolf, M. Zwick, G.G. Wagner (Hrsg.), Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag. Baden-Baden.
- Köhne-Finster, S., T. Körner (2008), Nachbefragung im Mikrozensus – ein Bund-Länder-Projekt zur Messung des ILO-Erwerbsstatus. S. 3–6 in: Statistisches Bundesamt (Hrsg.), Methoden ... Verfahren ... Entwicklungen. Wiesbaden.
- Körner, T., K. Puch, T. Frank, H. Meinken (2011), Geringfügige Beschäftigung in Mikrozensus und Beschäftigungsstatistik. S. 1065–1085 in: Statistisches Bundesamt (Hrsg.), Wirtschaft und Statistik, November 2011. Wiesbaden.
- Meyer, G. (1866), Das Princip der rechtlichen und der factischen Bevölkerung. Jahrbücher für Nationalökonomie und Statistik 6: 97–110 sowie 425–432.
- Michel, H. (1985), Volkszählungen in Deutschland. Jahrbuch für Wirtschaftsgeschichte 1985/2: 79–91.
- Münnich, R., S. Gabler, M. Ganninger, J.P. Burgard, J.-P. Kolb (2012), Stichprobenoptimierung und Schätzung im Zensus 2011. Statistisches Bundesamt (Hrsg.), Statistik und Wissenschaft, Band 21, Juli 2012. Wiesbaden.
- Sachverständigenrat zur Begutachtung der wirtschaftlichen Entwicklung (1990), Jahresgutachten 1990/91 des Sachverständigenrates zur Begutachtung der wirtschaftlichen Entwicklung. S. 104 in: Deutscher Bundestag (Hrsg.), Drucksache 11/8472. Berlin.
- Scheel, H. von (1869), Zur Technik der Volkszählungen. Jahrbücher für Nationalökonomie und Statistik 12: 156–172.
- Schoer, K., U. Hanefeld (1990), Erwerbstätige 1970 bis 1989. S. 245–257 in: Statistisches Bundesamt (Hrsg.), Wirtschaft und Statistik, April 1990. Wiesbaden.
- Scholz, R.D., D.A. Jdanov (2007), Nutzung der Daten des Forschungsdatenzentrums der Rentenversicherung zur wissenschaftlichen Mortalitätsanalyse – Verfahren zur Korrektur der Bevölkerungsbestände der amtlichen Statistik im hohen Alter in Deutschland. S. 200–211 in: Deutsche Rentenversicherung Bund (Hrsg.), Erfahrungen und Perspektiven: Bericht vom dritten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 26. bis 28. Juni 2006 in Bensheim, DRV-Schriften 55. Bad Homburg.
- Staatliche Zentralverwaltung für Statistik (1955), Gebiet und Bevölkerung, Berufsgliederung. Statistisches Jahrbuch der Deutschen Demokratischen Republik 1955: 8–33.
- Staatliche Zentralverwaltung für Statistik (1968), Übersicht über Gebiet und Bevölkerung. Statistisches Jahrbuch der Deutschen Demokratischen Republik 1968: 1–14.
- Staatliche Zentralverwaltung für Statistik (1975), Übersicht über Gebiet und Bevölkerung. Statistisches Jahrbuch der Deutschen Demokratischen Republik 1975: 1–12.
- Staatliche Zentralverwaltung für Statistik (1982), Übersicht über Gebiet und Bevölkerung. Statistisches Jahrbuch der Deutschen Demokratischen Republik 27. Jahrgang: 1–12.
- Statistische Ämter des Bundes und der Länder (2004), Ergebnisse des Zensus 2011. S. 813–833 in: Statistisches Bundesamt (Hrsg.), Wirtschaft und Statistik, August 2004. Wiesbaden.
- Statistisches Bundesamt (1990), Bevölkerung. Statistisches Jahrbuch 1990 für die Bundesrepublik Deutschland 1990: 40–75.
- Statistisches Bundesamt (1995), Künftige Zensen in Deutschland. Ergebnisse der Untersuchungen der Arbeitsgruppe Künftige Zensen zu Inhalt und Methode. Wiesbaden 1995.

- Statistisches Bundesamt (2008), Arbeitsmarkt und Erwerbstätigkeit. S. 109–144 in: Statistisches Bundesamt, Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen, Mannheim, Zentrum für Sozialindikatorenforschung, Heinz-Herbert Noll, Wissenschaftszentrum Berlin für Sozialforschung, Zentrales Datenmanagement, Roland Habich (Hrsg.), Datenreport 2008. Bonn.
- Statistisches Bundesamt (2011), Das registergestützte Verfahren beim Zensus 2011. Wiesbaden 2011.
- United Nations Economic Commission for Europe (2006), Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. New York and Geneva 2006
- United Nations (2011), International Human Development Indicators. Available at: <http://hdr.undp.org/en/statistics> (2011, last accessed 28 February 2012).
- Wagner, G. G. (2010), Volkszählung bitte nicht überfrachten ... vor allem nicht hinsichtlich der Erwartungen. Available at: [http://www.ratswd.de/download/atSWD\\_WP\\_2010/RatSWD\\_WP\\_134.pdf](http://www.ratswd.de/download/atSWD_WP_2010/RatSWD_WP_134.pdf) (January 2010, last accessed 14 February 2012).
- Zensuskommission (2009), Stellungnahme der Zensuskommission zum Merkmalskatalog im Kabinettsentwurf des Zensusanordnungsgesetzes (ZensusG2011). Available at: [http://www.zensuskommission.de/docs/Stellungnahme\\_2009\\_01\\_22.pdf](http://www.zensuskommission.de/docs/Stellungnahme_2009_01_22.pdf) (22 January 2009, last accessed 08 December 2011).
- Roderich Egeler, Statistisches Bundesamt, 65180 Wiesbaden, Germany.  
roderich.egeler@destatis.de
- Natalie Dinsenhacher, Statistisches Bundesamt, 65180 Wiesbaden, Germany.  
natalie.dinsenhacher@destatis.de
- Birgit Kleber, Statistisches Bundesamt, 65180 Wiesbaden, Germany.  
birgit.kleber@destatis.de

# Celebrating 150 Years of Analyzing Fertility Trends in Germany

**Michael J. Kendzia and Klaus F. Zimmermann**  
Institute for the Study of Labor GmbH (IZA), Bonn

JEL J10; J11; J13

Fertility; Malthus, Q-Q trade-off, self-regulation; total fertility rate.

Received: 07.02.2012

Revision received: 17.10.2012

Accepted: 30.10.2012

---

## Summary

Ever since the very beginning of the *Journal of Economics and Statistics* population economics has featured prominently in the *Journal*. Fertility naturally plays an important role in population economics. However, the level of German fertility has decreased significantly from the 1900s. The paper documents and analyzes the long-term development trends. We identify three different explanatory approaches for the decline in fertility according to which the various articles of the *Journal* related to this area are categorized. The paper also investigates fertility studies published by the *Journal* since the beginning. It points out that several articles anticipated subsequent research directions in the area of population economics at an early stage. In addition, significant contributions were made to improve and develop existing knowledge and understanding. Thus, the *Journal* has helped to expand the research area of population economics.

## 1 Introduction

“People will always have babies” was what German chancellor Konrad Adenauer said in 1957 when he defended the reform of the German pension system to a pay-as-you-go scheme. With that judgment he destroyed the arguments of those who doubted that the system – based on the so-called generation contract – would last forever. In the same year the total fertility rate in West Germany was 2.36 children per woman, but by 2010 that rate had fallen to only 1.39 (Destatis 2012b). While Adenauer did not consider this a problem, demography has proven him wrong. This is why a proper understanding of population economics is so important.

Since its beginning in 1863 the *Journal of Economics and Statistics*, founded by Bruno Hildebrand (1812–1878), has taken this research topic comprehensively into account. This article seeks to take stock of the development of population economics focusing on the discussion on fertility, in particular with regard to early articles of the *Journal*. To do this, we aim to examine both the explanatory variables and the development of the understanding of the population process after what Walt Rostow (1969: 4ff.) refers to as the take-off of industrialization in Germany which took place between 1850 and 1873.

Fertility and mortality undoubtedly constitute the essential growth components of the population development. In the paper we focus primarily on the development of fertility in Germany without neglecting to investigate the mortality rates.

The outline of the paper is as follows. The next section provides an overview of the time-series and trends of fertility in the long run. The analysis includes a discussion of the recent situation in Germany. Thereafter, we present and discuss the respective articles of the *Journal* in chronological order taking into account the three explanatory approaches or theories. The paper then ends with conclusions.

## 2 The long-term demographic development

This section analyzes the economic factors influencing the changes of fertility over time. Many European countries including Germany experienced a phase of demographic transition from 1800 to 1900. This phenomenon is usually defined as the transition process from (both) high to (both) low mortality and birth rates (Thompson 1929; Notestein 1945). This transition usually starts with a decreasing mortality rate which after some time is followed by a decline in the fertility rate.

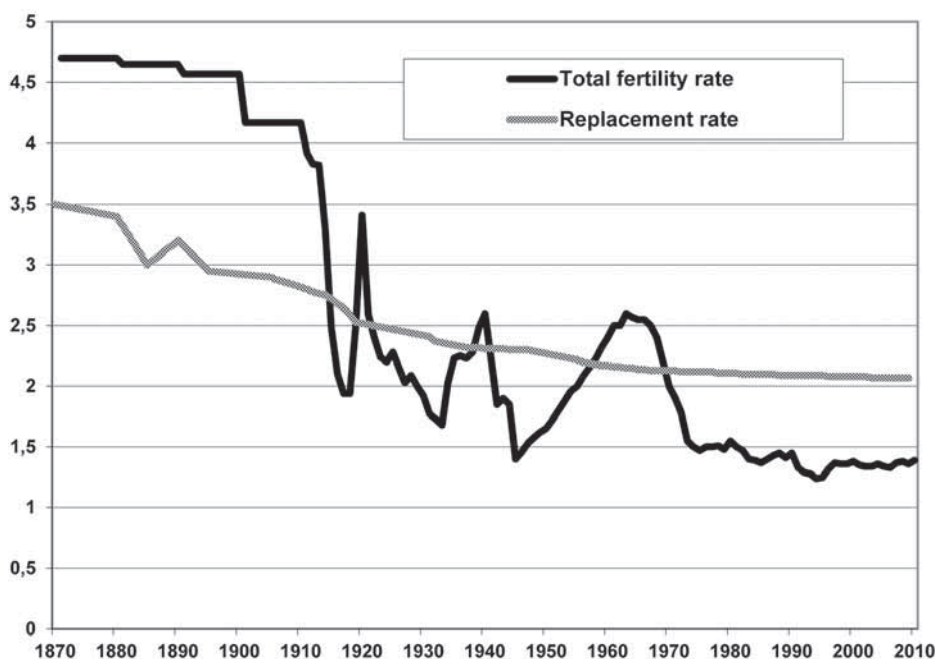
In order to study the long time-series, we present the development of the total fertility rate from 1870 to 2010. As a reference, we also measure and present the replacement rate over time which is the level of fertility required to ensure a constant population. In this calculation we do not consider migration. As Figure 1 shows, about 40 years after the beginning of the *Journal* in 1863 the total fertility rate started to decrease significantly. The first drastic low occurred during World War I (1914–1918). Afterwards, the total fertility rate recovered again peaking around 1920 at a little less than 3.5 but reached another low only a few years later during the Great Depression. Thereafter, the rate recovered again quickly before diminishing sharply with the beginning of World War II to a new record low by the end of the war. After a rise in the birth rate between 1950 and 1960 the rate increased strikingly before reaching a new low in the early 1970s and remained relatively stable up to 2010.

Figure 1 shows that the fluctuations around the trend diminished over time. Since the middle of the 1970s the course of the total fertility rate seems to have remained relatively stable. Shortly after the first publication of the *Journal* the average life expectancy at birth was approximately 34 years for men and 37 for women (Ehmer 2004: 34), whereas a man born in 2008/2010 is likely to live for about 77.5 years and a woman 82.6 (Destatis 2012a). In 2010 the total fertility rate in Germany was roughly 1.4 on average (Destatis 2012b). Moreover, since 1983 the rate has remained below 1.5 children per woman (OECD 2011). During the first years of the *Journal*, the average total fertility rate after the formal reunification of Germany (*Reichseinigung*) in 1871 was roughly 4.7 and thus significantly higher compared to 2010.

## 3 The contributions of the *Journal* to population economics

During our investigations we identified three different explanatory approaches for the decline in fertility according to which we will categorize the various articles of the *Journal* related to this research area. Additional to the three explanatory approaches identified in the *Journal*, there are several other theories and contributions in the area of the economic theory of fertility. To provide an overview, the main thoughts will be mentioned at this point. These include, for example, contributions to the infant mortality





Source: Bundesinstitut für Bevölkerungsforschung (2011).

Note: For several years no data were available. To create a long time-series, intermediate years were estimated.

**Figure 1** The long decline of the German total fertility rate from 1871–2010

rate of (Rosenzweig/Schultz 1982). Thereafter, families desire to have a certain number of children. An increasing income goes in line with a decrease in the death rate and an incline as regards the probability of surviving of children. Consequently, fewer births are sought by the individuals. This is shown by Army and May (1968) within their article about the desired number of children in India within the series Population Studies. A similar argument is used, inter alia, by Repetto (1972) and Leibenstein (1974: 464).

Furthermore, the socioeconomic status is suggested to be another sociological and economic theory of fertility. Depending on the different status, people tend to have different preferences concerning children or material goods. According to this theory, the higher the social status is the stronger the preferences for material goods are pronounced. In this context, it is important to mention the contributions made by Easterlin (1969) and Freedman (1963). Similarly, some authors emphasize (Easterlin 1978; Encarnación 1973) the existence of so-called thresholds, both in terms of income and education, according to which the inverse relations between fertility, income or rather education can be observed. Also, demographic researchers (Freedman 1963) argue that changes in norms and institutions, through e.g. a process of modernization, have an influence on the size of the family. This might encompass group membership such as a peer group. Easterlin (1978) develops an idea of Duesenberry (1960), hereafter an individual seeks to maintain a certain target standard of living and excessive numbers of children worked against this (Leibenstein 1974: 465).

**Table 1** Relation between population and food growth according to Malthus

	Annual intervals								
	0	25	50	75	100	125	150	175	200
<b>Total population</b>	1	2	4	8	16	32	64	128	256
<b>Means of sustenance</b>	1	2	3	4	5	6	7	8	9

Source: Esenwein-Rothe (1978: 58).

The first approach was provided by the founder of population economics, Thomas R. Malthus (1766–1834). The English economist was rather concerned by the near future of the population. His ‘Essay on the Principle of Population’ (Malthus 1798) was one of the first on population economics which received a great deal of attention.

Malthus assumed that the total number of the population would grow in a geometric series which meant the number would double every 25 years. On the other hand, the means of sustenance could only be enhanced by arithmetical progression (see Table 1). The population would thus grow more rapidly than food production. Further growth in the population could not be supported by the food supply and would lead to impoverishment or death from starvation (Esenwein-Rothe 1978: 58). Thus, the fear of overpopulation and food scarcity, introduced by Malthus, constitutes the first explanatory approach in the area of population economics.

The second approach was developed by Lujo Brentano, Gary Becker, and Jacob Mincer (Brentano 1909; Becker 1960; Mincer 1963). This approach is closely related to the rational choice theory dealing with trade-offs between goods and opportunity costs. This theory is based on utility maximization and takes into account in various forms the costs of raising children, the value of time, the costs for education, and the participation of women in the labor market as well as the impact of uncertainties on the labor market. It creates a potential for a non-linear budget constraint that generates an impact of income on fertility that switches from positive to negative with rising income. Rising living expenses and female wages are negatively related with fertility.

The third explanatory approach is circumscribed by self-regulation. Ronald Freedman (1975) offers the explanation that fertility reacted to the increase of child survival in a more or less homoeostatic way. This means that self-regulation played a strong role during 1875 and 1960. According to this approach, the phenomenon of the decline in fertility can be attributed to a regulatory circuit. Hence, fertility adapts to a decline in the child mortality rate and a natural number of children is reached. In turn, when child mortality is high, more children are raised. This approach is not linked with the rational choice theory, but it rather contains a leveling off in terms of fertility.

Ever since the very beginning of classical economics, population problems have played an important role in this emerging research area. At the time of mercantilism, which represented the dominant economic doctrine across Europe from the 16<sup>th</sup> to the 18<sup>th</sup> century, people assumed that the strength of a country largely relied on the number of soldiers. To obtain a large army, a country had to strive for a large population. The German pastor, statistician, and demographer Johann P. Süßmilch (1707–1767) believed population growth to be the ‘happiness of a country’. The prosperity of a nation would, according to the mercantilist conception, benefit enormously from an increasing population by stimulating social and economic development. It was assumed that demographic growth would contribute to consolidating the position of the ruling class (Overbeek 1974: 30).

In order to paint a realistic picture of the past, it should be noted that at the beginning of the 19<sup>th</sup> century most women in Germany died at an early age. Although lifespans were shorter, there was an average of five births per married woman. However, during that time we have to differentiate between the number of births and the number of children. In the middle of the 19<sup>th</sup> century only half of all children born in Germany reached adulthood. Not until the course of the 20<sup>th</sup> century did the number of births and the number of children start to converge. It was at the turn of this century when generations of women stood on the verge of a decline in the fertility rate. Thereafter, the average number of children per woman decreased dramatically from nearly 5 to around 2 (Ehmer 2004: 34-44). With the growing acceptance of how germs spread disease, more emphasis was put on hygiene which constrained the spread of diseases and contributed to a rise in life expectancy. Factors contributing to lower mortality rates amongst children included the declining price of food relative to wages and better quality and diversity of food. This was especially true of food bought from urban markets. Moreover, better housing together with the expansion of welfare systems greatly improved the medical and nutritional conditions for the vast majority of people (Schultz 2009).

At the end of the 19<sup>th</sup> century Reich Chancellor Otto von Bismarck gradually introduced social legislation in Germany starting with accident insurance in 1884. This insurance was created in order to protect workers against basic risks. The emergence of the social security system at the end of the 19<sup>th</sup> century established a welfare system which began to expand in the following years and provided a safety net from risks such as accidents, poverty, aging, and illness (Kendzia 2010a). However, this was not the first legal provision of this kind. The “Regulativ über die Beschäftigung jugendlicher Arbeiter in Fabriken” was one the first legal measures which protected workers and came into force in 1839. The result was a major step towards a ban in Prussia on child labor (aged younger than 10). Several other factors also played a role, including the decreasing demand for unskilled workers as a consequence of rapid technical progress and a stronger enforcement of the compulsory school attendance which had existed throughout Prussia since 1717 (Kendzia/Pierenkemper 2010). Through the stronger enforcement of the compulsory school attendance, fewer individuals were joining the workforce.

This, in turn, resulted in a relaxed situation in the labor market which went hand in hand with a rise in living standards across Germany (Zimmermann 1988). Most parts of Europe during the 19<sup>th</sup> century were shaped by a new and irreversible process which had begun and would lead to the single most sectoral structural change in the economic landscape: industrialization. The share of people working in agriculture decreased, and more and more people began to work in new industries such as mining and steel manufacturing or later in the chemical and electrical industries, which soon started to gain international reputation.

#### **4 The first years of the *Journal* until the seizure of power by the National Socialists**

In the first years of the *Journal* the economist and statistician Hans von Scheel (1839–1901), an assistant of Hildebrand at the statistical office in Jena, showed that the wealth and growth of the population would be closely related to the quantity of available resources, such as means of sustenance (e.g. food). According to Scheel, it would be the task of statistics to recognize and verify the degree of dependence the population had on comestible goods. He referred to an early study carried out by the French demographer

Louis Messance (1734–1796) on the relation between grain prices and changes in the mortality rate.<sup>1</sup> Messance (1767) had shown a time-series including the average grain prices and compared those with the mortality rate over time. He found that the mortality rate was lower when grain was cheaper. After presenting further studies on this issue he concluded that there was little point investigating the manner further because showing that grain prices influence the mortality rate could be displayed by following simple common sense – without any help from statistics. Scheel then investigated the prices of grain, potatoes, and meat in the small duchy of Altenburg from 1835 to 1864. However, since the consumption of meat – relative to grain or potatoes – was small for large parts of the population, he decided to concentrate on the development of the price of grain and compared its course with the number of births. He showed a reverse development of the price of grain and fertility, i.e. that a high price for grain resulted in lower births in following years (Scheel 1866: 176ff.).

A few years after Scheel's analysis of the price of grain and its close relation to the mortality rate the economist, and later rector of the University of Strasbourg, Georg Friedrich Knapp (1842–1926) investigated the mortality rate for the city of Leipzig from 1751 to 1870. He found that the mortality rate of children 12 months and younger was 0.223, which meant that from 1000 children approximately 223 had died during their first year. By contrast, this rate was 0.144 for children between 1 and 10. Knapp argued that compared to earlier years both rates had already decreased and concluded that this could in particular be attributed to an improvement in health care, better hospitals, an approved type of construction of cities, and better household furnishings (Knapp 1873a).

Thus, in the literature there already existed agreement over improvements in terms of urban public health infrastructure leading to increased living standards. For instance, investments and regulations concerning water supply and sanitation had had a huge impact on the quality of life. The growing standard of living can be interpreted as a consequence of a general process that had been occurring during the 19<sup>th</sup> century: urbanization. It describes the migration of large parts of the population from the countryside to steadily growing cities. As a result, more and more of those agglomerations attracted an increasing number of workers. Around 1913 the general employment in industries in Germany exceeded the employment in agriculture for the first time (Kendzia 2010a: 6).

The process of urbanization was also documented by several authors of the *Journal* (e.g. N.N. 1871: 135ff.; Seutemann 1906: 88). Scheel (1874: 1ff.) indicated that the urbanization went hand in hand with an increasing rate of women's employment. Moreover, he denoted that they were becoming less active in the field of domestic work. As Scheel put it, women were being driven out of their traditional sphere of work. A later article examined the further growth of the population and in particular considered the high fertility rate in industrial areas which was attributed to the large number of children working class families had (Rohr 1891: 118ff.).

Prinzing (1899: 581) gave a truly comprehensive account of child mortality across Germany. He analyzed the development of the child mortality rate within European countries over time. According to him, favorable economic conditions would result in higher fertility rates, but he failed to substantiate this. Prinzing's elaborations and assumptions

<sup>1</sup> Recherches sur la population des généralités d'Auvergne, de Lyon, de Rouen et de quelques province et villes du royaume avec des reflexions sur la valeur du bled tant en France qu'en Angleterre depuis 1674 jusqu'en 1764 par M. Messance (Louis Messance), Paris 1767.

were closely related to Malthusian concerns and can thus be attributed to the first category of our explanatory approach. Similar to Malthus, his assumptions would have led to an enormous rise in population by an increasing national income per capita. However, he was proved to be wrong shortly after the turn of the 20<sup>th</sup> century.

A later author of articles in the *Journal* contributing to the discussion on the development of fertility, Karl Seutemann (1907: 293ff.), observed a decline in the fertility rate between 1890 and 1905. He denoted that a decline in the mortality rate would lead to two growing parts of the population: on the one hand the non-fertile cohort, such as the very young, and on the other hand the old. He went on later to discard this notion. He pointed out that a certain mechanism of self-regulation would imply decreasing fertility rates. He concluded that a decline of the fertility rate was mostly associated with food shortages resulting from higher life expectancies as well as the decreasing mortality rate. Consequently, people would raise fewer children than in the past (Seutemann 1907: 293ff.). With this, Seutemann (1907) relates to the first category of the three identified explanatory approaches.

The German economist and social reformer Lujo Brentano (1844–1931) found a direct and negative correlation between welfare and matrimonial fertility. In addition to Malthus, who argued that sexual abstinence, vice, and misery all impeded fertility, Brentano (1909) suggested that physiological reasons as well as the separation between sexual drive and reproductive behavior would also affect fertility. However, we will not elaborate on the physiological aspect in this paper since we argue that the decisive factors influencing fertility have always been primarily of an economic nature. According to Brentano (1909), the decline of the willingness to procreate was due to different activities from which the household could choose and to the enhanced quality of child care. The concept of choice between different activities by respecting the utility maximization, the time costs of bringing up children as well as the quality of life for children can all be interpreted as an important step into the area of modern population economics (Zimmermann 1989: 474ff.). Since Brentano (1909) stresses household choice, he should be considered as a representative of the rational choice theory and belongs to the second category of the explanatory approaches.

In 1913 Henriette Fürth (1861–1938), who later became the first woman to join the German Society for Sociology, noted in the *Journal* that a decline in fertility would constitute a major social problem and indicated that with increasing costs of the standard of living – and growing expectations as a result – fewer children would be born. With regard to higher fertility within working-class families compared to other social classes, she argued that a rise of food prices would also contribute to higher costs of living. Since fertility rates had decreased among both the poor and the rich, she also traced the decline back to more sophisticated individual, cultural demands (Fürth 1913: 721ff.). However, Fürth follows the second explanatory approach by exhibiting higher demands which would lead to higher costs to keep the standard of living. With that, not only Seutemann (1907) but also Fürth (1913) had already identified key elements of the quality-quantity trade-off which is later further discussed by Becker. Interestingly, Fürth (1913) showed that the administration was deliberately not spreading contraception among the vast majority of the population but instead distributing it solely to members of the German navy. The government wanted unrestrained population growth in order to strengthen military power (Fürth 1913: 747ff.). A possible reason why sailors are given contraception is to protect its sailors from possible diseases. However, this is not speculated by Fürth.

During World War I fertility declined, but Fürth did not consider reasons for this phenomenon in her article (Guradze 1916: 550ff.). As P. Kollmann (1915: 251f.) had argued one year earlier, the only cogent reason for the further decline in fertility could be associated with the continued reduction in child mortality. If we follow his argument, then most people in society had started to anticipate the converging of the number of births and the number of children which was due to a decline in child mortality rates. Similar deliberations were shared by Karl Oldenberg (1923: 315ff.) who dealt explicitly with the connection between fertility and child mortality. Both deliberations dealing with anticipation can be seen in connection with the third explanatory approach, 'self-regulation'. In the 19<sup>th</sup> century the number of births and the number of children were divergent. Due to a decline in the mortality rate in the 20<sup>th</sup> century, those numbers converged. Self-regulation thus meant that people anticipated these changes and adapted the number of children accordingly.

Ernst Günther dealt with the issue of declining fertility and its potential impact on the unemployment rate in an article in 1931. He examined a period of around 15 to 20 years during the demographic transition. The paradox he formulated was that fewer births – and with that fewer workers – would lead to higher unemployment (Günther 1931). A later author of the *Journal*, Adolf Wagner, adopted this idea in his article on the 'Günther paradox'. He believed that if the dependence were to exist there would be something – in the style of the demographic transition – like demographic unemployment (Wagner 1980).

## 5 The *Journal* in the time of National Socialism

Following the rise to power by the National Socialists in 1933 their ideology, including the role of women as mothers and housewives, could soon be observed across society. After the Great Depression, during the early 1930s, the regime started replacing female employment with male employment and promoted female activities in the area of private households and agriculture. According to Mackenroth (1934: 205), this measure was aimed at stabilizing the wage level. As we can see from Figure 1, there was a noticeable increase in fertility. The incline of the fertility rate across Germany, despite a high participation rate of women on the labour market, was described in the *Journal* by the same author (Mackenroth: 1934: 445). Michalke (1935: 438ff.) noted that during the Great Depression the spread of female employment was due to the fact that female labor was cheaper. In the following years possible female employment was considered as a 'reserve'.

In 1936 the official office responsible for the use of female workers – *Frauenamt der Deutschen Arbeitsfront* – prepared women for their later roles as mothers and housewives rather than for the labor market (Willeke 1936: 197-208). This bears witness to the fact that research documented the impact of the ideology of National Socialism which included the role of women with care responsibilities in the household, rather than economically independent individuals.

After reaching full employment in 1936 the regime first tolerated and afterwards supported female workers. When labor shortages occurred, the interest of the National Socialist regime to promote female employment grew rapidly from 1942 on. This resulted in obligatory work for women in 1943 (Kendzia 2010b). Interestingly, in 1937 Ungern-Sternberg (1937: 471ff.) examined the development of the economic situation in parallel with fertility. He presented a theory which compared the relatively low fertility with the 'state of emergency' in which individuals found themselves. The reason for this assumed



state was that most people thought they could not afford to have children. Furthermore, in his investigation he suggested that the reason for the decline in fertility would stem from self-imposed ‘artificial birth restriction’. This behavior was, according to Ungern-Sternberg, due to the intention of many people in society to sustain prosperity for the later generation. Hence, he belongs to the second category within this paper. Birth control in order to ensure a higher living standard for future generations reflects a trade-off which can be interpreted as a rational choice. His explanations are early indications of an idea which would later be described as ‘fertility control’ by the French demographer Louis Henry in the 1950s. Fleury and Henry (1956) stressed that fertility control would result in fewer children because households would avoid further births after a certain number of children had been reached (Coale/Watkins 1986: 9). And indeed, this argumentation again raises the importance of the formulated trade-off in a household which states that human beings should behave or act rationally. Accordingly, the strong desire for restricted births corresponded directly to the social and economic change taking place at that time (Pollack/Watkins 1993: 468ff.).

As shown above, the research on fertility trends in the *Journal* documented the developments across Germany during the time under the National Socialist regime. Only one author used the then relatively national wording ‘Das deutsche Volk’ for the German population and somewhat exaggerated the increasing fertility rate as achievement of the ruling regime. Its policies in view of fertility included direct financial aids aiming at increasing marriages and reducing the costs of rearing children. Moreover, the same group was supported through privileges and advancements. And lastly, the abortion law was enforced by the state (Taeuber/Taeuber 1940: 150). The German demographer Friedrich Burgdörfer (1890–1967), member of the then ruling party NSDAP<sup>2</sup> since 1937, stated referring to the increase of births during the years 1934–1937, ‘The German mothers presented the National Socialist Reich within four years a full birth cohort better than expected’ (Burgdörfer 1938: 300).<sup>3</sup> However, as Taeuber and Taeuber (1940) put it, ‘the extent to which Germany has raised the birth rate should not be overestimated’ (Taeuber/Taeuber 1940: 163). That is, the impact of the National Socialist policies on the fertility rate remains controversial in population economics. In Germany, the industrial production reached its low around the year 1932. Afterwards, the economy within the leading industrialized countries such as the United States, the United Kingdom, France, Germany, and Japan recovered noticeably. Thus, a general economic upswing across the mentioned countries took place soon after the seizure of power by the National Socialists in 1933 (Romer 1993: 21).

The wording suggests that the author could not entirely escape the then dominating national atmosphere. Nevertheless, the author described correctly that the fertility rate during that time increased, as Figure 1 shows (Ungern-Sternberg 1937: 484). However, other authors without any article in the *Journal* followed a rather dubious approach during the National Socialist regime such as Karl V. Müller (1935) investigating the core of the racial hygiene in order to create a sustainable population and race policy (Ferdinand 2006: 217). Nonetheless, it is true to say that population economics dealing with fertility in Germany did not suddenly end up in the year 1933, the seizure of power through the

<sup>2</sup> The shortcut stands for the ‘Nationalsozialistische Deutsche Arbeiterpartei’ (National Socialist German Workers’ Party).

<sup>3</sup> The original quote is as follows: “Die deutschen Mütter haben dem Nationalsozialistischen Reich in vier Jahren sozusagen einen vollen Geburtsjahrgang überplanmäßig geschenkt”.



**Table 2** Articles in the *Journal* on fertility during 1863 and 2013 (15-year period)

1863– 1877	1878– 1892	1893– 1907	1908– 1922	1923– 1937	1938– 1952	1953– 1967	1968– 1982	1983– 1997	1998– 2012
1	1	4	4	4	1	0	3	4	2

National Socialists. Similarly, we can state that there did not exist a ‘zero hour’ concerning population economics in the year 1945 and after (Gutberger 2006: 162).

In 1941 Zwiedineck-Südenhorst (1871–1957) (1941: 161ff.) noted that in previous centuries the fear of overpopulation often occurred and thus rather Malthusian concerns were shared. However, in recent years, as he observed, the opposing fear of depopulation would arise. Similarly, he reported on the then alleged demographic policy claiming at least three children per woman (Zwiedineck-Südenhorst 1941: 168). With Zwiedineck-Südenhorst the *Journal* experienced a caesura in terms of publishing articles on fertility issues. As Table 1 displays, the *Journal*’s interest in population economics decreased soon after World War II. It is true to say that during this period more emphasis was generally put on theoretical developments in economics, but discussions on fertility vanished from the scene. According to Kurz, World War II and a massive increase in the fertility rate in the following years led to this standstill (Kurz 1982: 235f.).

## 6 The *Journal* from the post-war period to today

This changed with stagnating and later decreasing fertility rates in the beginning of the 1970s in many industrialized countries. Jacob Mincer (1922–2006) (1963) further developed the decision-making process within households by introducing two important factors. Firstly, he introduced the time costs of raising children, and secondly, he argued that the female labor force supply and the wish to have children are both part of the decision-making process within the family. In doing so, Mincer applied the concept of opportunity costs to the issue of raising children (Zimmermann 1988: 123, 1989: 469).<sup>4</sup> In the sequel, it was possible to predict a decline in fertility and a rise in female work participation as the outcome of joint decision making driven by a relative increase of female wages in comparison to male wages.

The next groundbreaking contribution to model the decline in fertility was Becker’s attempt to model home production within the so-called quantity–quality (Q–Q) trade-off (Becker 1981/1991). This approach pointed out that the Malthusian model had missed the connection that with increasing income more and more emphasis was put on child quality. The child Q–Q trade-off between fertility and education is generally described as the choice between the quantity of children and the quality of human capital invested in each child. Becker showed that the Q–Q trade-off is closely related to income and prices as well as tastes (Becker et al. 2010). Zimmermann (1985, 1989) has shown that the complex quantity-quality approach of Becker can be replaced by simply assuming that child material expenditures (“child living expenses”) are rising with the consump-

<sup>4</sup> As Zimmermann (1988, 1989) has pointed out, it was Mincer (1963) and not Becker (1960) who developed the most important elements of the neoclassical theory of fertility first before Becker (1981/1991) established and dominated the field. A full discussion of the international literature can also be found in Zimmermann (2005) which also contains a comprehensive reference list.

tion level of the parents. This child living expenses approach leads to the same conclusions concerning a negative impact of rising income on fertility.

As another representative of the rational choice theory, Ingeborg Esenwein-Rothe (1911–2002), dealt with population forecasting models and their limits in 1978. She examined the different developments concerning fertility across the world. She investigated not only the issue of overpopulation in developing countries but also the decreasing fertility and wrote of the fear of the death of an entire nation. Esenwein-Rothe (1978) also described the reasons for the increasing standard of living during industrialization by highlighting how the railway and steamships had expanded means of sustenance. At the same time the transparency of the market had improved through modern telecommunications. She spoke of a cultural change in the awareness of future generations. Fertility would go hand in hand with the increasing standard of living which meant that the size of a family would adapt to the financial and time resources of the individuals according to their social conditions and desires (Esenwein-Rothe 1978).

Meanwhile, the family of four with two parents and two children had become the dominant family structure in society. During the time when national ideas were sweeping across Germany, this dramatic decrease was heavily debated and criticized as the possible death of the nation and later, assuming an ever more dramatic scenario during the years of the National Socialist regime, the death of the German race. By World War I the total fertility rate had reached an all-time low which recovered afterwards but fell once again to a new low during the Great Depression and at the end of World War II. As Figure 1 shows, the 1950s and 1960s witnessed a baby boom with the total fertility rate rising to nearly 2.5. This can be linked with the German economic miracle which took place until the first recession after World War II in 1966 to 1967. After that period a short and dramatic decline in fertility occurred until the early 1970s when the decline stabilized. The development of the total fertility rate in West and East Germany until 1970 remained more or less the same. Only in the second half of the 1970s did the rate increase in East Germany, whereas the rate in the West remained stable at 1.4 (Ehmer 2004: 44–46).

Another contribution in the *Journal*, which was made by Adolf Wagner, concerned a book dealing with the decline in fertility (Dinkel 1985). Wagner pointed out that most people worked in agriculture prior to industrialization. During that time children were often employed as workers on the family farm (Wagner 1985: 544). Thus, child work diminished during industrialization and fewer children were born. The Malthusian concerns regarding overpopulation did not materialize in developed countries because one of his key assumptions proved to be wrong. Malthus failed to consider that technical progress would increase farming yields so greatly which meant that the increase in food supply outstripped the demand.

Furthermore, Malthus suggested a positive correlation between income and fertility (Zimmermann 1988: 121f.). Due to the implementation of a social security system, the social benefit of children decreased. At the same time, the costs of raising children increased. Matthias Bletzinger and Uwe Walz (1989) extended the theoretical model of fertility by Zimmermann (1985) to explain the empirically observable negative relationship between income and fertility. They showed that in particular the social status of the parents determines the essential inputs for the children. Olaf Hübler (1991) added that women would earn on average less than men. He dealt with the causes of gender wage discrimination and gender dependent earnings. The article did not discuss the economic factors influencing fertility, however.

Assenmacher and Wenke (1993) referred to Becker's model of the household which not only consumes but also produces. Both authors reported that the service sector has risen constantly across industrialized countries. To explain this growth they indicate the shift from home production to market production in terms of services as a consequence of increasing labor market participation by women. This trend can be described as follows: In the beginning of the 1960s the female employment rate accounted for less than 50 percent. However, the same rate increased significantly during the last 50 years and peaked in 2010 at nearly 70 percent (Statistisches Bundesamt 2012).

More than thirty years after Mincer's contribution on the family, Galor and Weil (1996) created a model which was based on the idea that increasing capital per worker would imply a rise of the relative wage of female workers. This, in turn, would lead to a reduction in the size of the family which would imply that child-rearing activities and female labor supply would substitute each other (Galor 2005). Concerning the development of women's income as a percent of men's, similar to the participation rate of women in the labor market, women's relative income has risen relatively stably since the 1960s. Latest Figures show that the percentage of women's income by men's income now clearly exceeds 50 percent, whereas the same Figure amounted only to less than 40 percent in 1960 (BFSFJ 2005: 219).<sup>5</sup>

Althammer and Wenzler (1996) exhibited that the decision on the allocation of working time would take place within households and was dependent on the husband's wage. The total fertility rate in East Germany was the lowest in the world at 0.77. Eberhard Schaich (1998: 94-105) stressed that from 1991 to 1995 the rate remained constantly under 1. He offered several explanations for this, such as a pessimistic assessment of the future of young women because of the new political situation and many job losses during that period. In addition to new alternatives to the traditional family pattern, new job opportunities and consumption possibilities had emerged. After some time, an alignment to the demographic situation in West Germany began to occur.

Kleinhenz (2004) dealt with the consequences of the demographic change. He stressed the impact of the sudden drop in birth rates due to the birth control pill. Further explanations are associated with balancing the different demands between the workplace and family life since most women tend to have fewer children at a later age (Kohler et al. 2002). The increase in life expectancy and the technological change taking place is accompanied by higher wages. This, in turn, leads to more educational attainments across the population in order to acquire time-intensive human capital (Cervellati/Sunde 2007).

According to recent research in this area, the rational choice approach forms a conceptual framework which is likely to be consistent and gives an account of both the economic and fertility transition (Cervellati/Sunde 2007). It appears that from today's point of view, the basic changes in mortality and fertility are dependent on the reallocation of family resources (Schultz 2009). Furthermore, the latest research shows that even prior to the demographic transition, the Q-Q trade-off had existed (Becker et al. 2010). There has been an extensive empirical literature on both human capital and family economics in recent years in the area of population economics. The majority of that literature concludes that factors such as marriage, work, wages, and schooling as well as a high labor

<sup>5</sup> Percentage of women's income by men's income. The income is cumulated on the basis of the average annual earning points and the average insurance years when retiring. From 1990 on it is weighted by the proportion of East (about 20 %) and West Germany (about 80 %).

force participation rate of women all affect fertility to a great extent (Browning et al. 2011).

## 7 Conclusions

To sum up, as Table 1 shows, right after the introduction of the *Journal* and within the very first 30 years of its existence, the *Journal* only rarely addressed issues regarding fertility. The reason might be that before the turn of the century, a relatively high fertility rate was observed and there was no need to put fertility on the agenda. Later, during 1893–1937, the situation changed. Most attention throughout the existence of the *Journal* has been paid to fertility between 1893 and 1937. This might be explained by the dramatic decline in fertility during that time. However, in the following years, from 1938 to 1967, the topic received very little interest from the *Journal*. An explanation could be on the one hand, that owing to the war the topic became less important and, on the other hand, Germany faced a post-war baby boom during the economic miracle which also might have resulted in less interest and thus publications in the area of fertility. Nevertheless, during 1968 and 2012 more attention has been paid to fertility in the *Journal*, which might refer to the relatively stable and low level of fertility since then.

Deliberations on population economics in Germany have lost nothing of their importance as a closer look at the trends and the number of works in the *Journal* regarding fertility confirms. As we have found, the three explanatory approaches (i.e. Malthus' early contributions, the rational choice theory and self-regulation) have all appeared in the *Journal* in some way over time. Whereas Malthusian concerns of food scarcity due to overpopulation were shared 1899 for the last time, treatises dealing with the explanatory approach of self-regulation appeared even in 1915 (Kollmann) and 1923 (Oldenberg). Nevertheless, the concept of self-regulation was not noticed in the *Journal* after 1923. As already explained, the German economist Brentano in 1909 was the first to point out the direct and negative correlation between welfare and matrimonial fertility. Thereby, he laid the groundwork for Becker (1960, 1981/1991) and Mincer (1963). Brentano (1909) suggested, in essence, that the decline in fertility was related to various activities from which the household could choose. By doing so, Brentano was a fore-runner of modern-day population economics.

Subsequently, the rational choice approach became the dominant stream to explain the course of fertility in the area of population economics. The approach which is based on the rational choice theory seems to provide a plausible answer to many questions raised in this article. It appears that the benefit of children was higher before the demographic transition. After the end of the transition process, the decline in the child mortality rate and the wide establishment of a social security system led to a diminishing benefit of children. At the same time, to ensure children the same standard of living, parents faced increasing costs of rearing children. However, since the 1970s the total fertility rate has remained relatively stable – but at a much lower level.

The displayed explanatory approaches have contributed to the development of theory in the area of population economics. The Q-Q trade-off seems to be an appropriate method of investigation. However, it remains a theory of limited scope. The authors neglected the influence of social groups on fertility. All in all, the determining factors concerning the development of fertility appear to be diverse. Some may be due to unique historical circumstances, such as the two World Wars, others depend on the changing socio-economic conditions and cultural factors. Therefore, the development seems to depend on various

issues such as economic changes, social circumstances, and a comprehensive process of modernization (Leibenstein 1974: 468-471). One question that remains unanswered is, as (Leibenstein 1974: 471) puts it: "A sensible theory must take into account competition between the budgetary demands for expenditures on children versus pressures for other expenditures as the social and economic circumstances change in the course of economic development. Where do these pressures for other expenditures come from?"

## References

- Althammer, J., S. Wenzler (1996), Intrafamiliale Zeitallokation, Haushaltsproduktion und Frauenerwerbstätigkeit. *Jahrbücher für Nationalökonomie und Statistik* 215: 398-418.
- Assenmacher, W., M. Wenke (1993), Haushaltsproduktion, Frauenerwerbstätigkeit und Dienstleistungsnachfrage privater Haushalte in der Bundesrepublik Deutschland. *Jahrbücher für Nationalökonomie und Statistik* 211: 22-41.
- Becker, G.S. (1960), An Economic Analysis of Fertility. *Demographic and Economic Change in Developed Countries*. NBER, Princeton: 209-231.
- Becker, G.S., K.M. Murphy, R. Tamura (1990), Human Capital, Fertility and Economic Growth. *Journal of Political Economy* 98: 12-37.
- Becker, G.S. (1981/1991), *A Treatise on the Family*. Harvard University Press.
- Browning, M., P.A. Chiappori, Y. Weiss (2011), *Family Economics*, Unpublished.
- Becker, S., F. Cinnirella, L. Woessmann (2010), The Trade-off between Fertility and Education: Evidence from before the Demographic Transition. *Journal of Economic Growth* 15(3): 177-204.
- BFSFJ (2005), *Gender-Datenreport – 1. Datenreport zur Gleichstellung von Frauen und Männern in der Bundesrepublik Deutschland*. München: Pp. 219. VDR-Statistik Rentenzugang. Available at: <http://www.bmfsfj.de>
- Bletzinger, M., U. Walz (1989), Zum Schattenpreis- und Einkommensbegriff in der ökonomischen Theorie der Fertilität. *Jahrbücher für Nationalökonomie und Statistik* 206: 591-598
- Brentano, L. (1909), Die Malthus'sche Lehre und die Bevölkerungsbewegung der letzten Dezenen. Pp. 565-625 in: *Abhandlungen der historischen Klasse der Königlich-Bayerischen Akademie der Wissenschaften*, Bd. 24, München.
- Bundesinstitut für Bevölkerungsforschung 2011, *Zusammengefasste Geburtenziffer in Deutschland, 1871 bis 2009*. Available at: <http://www.bib-demographie.de>
- Cervellati, M., U. Sunde (2007), Human Capital, Mortality and Fertility: A Unified Theory of the Economic and Demographic Transition. IZA Discussion Paper No. 2905.
- Coale, A.J., S.C. Watkins (1986), *The Decline of Fertility in Europe*. Princeton.
- Destatis (2012a), *Durchschnittliche und fernere Lebenserwartung nach ausgewählten Altersstufen*. <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Bevoelkerung/GeburtenSterbefaelle/Tabellen/Content50/LebenserwartungDeutschland,templateId=renderPrint.psml>
- Destatis (2012b), *Durchschnittliche Kinderzahl je Frau, Zusammengefasste Geburtenziffer*, <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Bevoelkerung/GeburtenSterbefaelle/Tabellen/Content50/GeburtenZiffer,templateId=renderPrint.psml>
- Dinkel, R. (1985), *Nur Transfers zur Anpassung an den Bevölkerungsrückgang?* Stuttgart.
- Duesenberry, J.S. (1960), An Economic Analysis of Fertility: Comment, *Demographic and Economic Change in Developed Countries*. Princeton.
- Easterlin, R.A. (1969), Towards a Socioeconomic Theory of Fertility: A Survey of Recent Research on Economic Factors in American Fertility. In: S.J. Behrman, L.Corsa, R.Freedman (eds.), *Fertility and Family Planning: A World View*. Ann Arbor, University of Michigan Press.
- Easterlin, R.A. (1978), The Economics and Sociology of Fertility: A Synthesis. Pp. 57-133 in: Ch. Tilly (ed.), *Historical Studies of Changing Fertility*. Princeton.

- Ehmer, J. (2004), *Bevölkerungsgeschichte und historische Demographie 1800-2000*. Enzyklopädie deutscher Geschichte, Bd. 71, München.
- Esenwein-Rothe, I. (1978), Modelle für eine Bevölkerungsprojektion und die Grenzen ihrer Aussagekraft. *Jahrbücher für Nationalökonomie und Statistik* 193: 54-83.
- Ferdinand, U. (2006), Historische Argumentationen in den deutschen Debatten zu Geburtenrückgang und differentieller Fruchtbarkeit. Fallbeispiel Karl Valentin Müller (1896-1963). *Historical Social Research* 31(4): 208-235.
- Fleury, M., L. Henry (1956), *Des registres paroissiaux à l'histoire de la population*. Manuel de dépouillement et d'exploitation de l'état civil ancien. Paris.
- Freedman, D.S. (1963), The Relation of Economic Status to Fertility. *American Economic Review* 53: 414-426.
- Freedman, R. (1964), Norms for Family Size in Underdeveloped Areas. *Proceedings of the Royal Society of London* 159: 220-245.
- Freedman, R. (1975), *The Sociology of Human Fertility*. New York.
- Fürth, H. (1913), Der Rückgang der Geburten als soziales Problem. *Jahrbücher für Nationalökonomie und Statistik* 100: 721-760.
- Galor, O. (2005), From Stagnation to Growth: Unified Growth Theory. Pp. 171-293 in: P. Aghion, S. Durlauf (eds.), *Handbook of Economic Growth*, vol. 1A. Amsterdam.
- Galor, O., D. Weil (1996), The Gender Gap, Fertility and Growth. *American Economic Review* 86: 374-387.
- Günther, E. (1931), Der Geburtenrückgang als Ursache der Arbeitslosigkeit? Untersuchung einiger Zusammenhänge zwischen Wirtschaft und Bevölkerungsbewegung. *Jahrbücher für Nationalökonomie und Statistik* 134: 921-973.
- Guradze, H. (1916), Säuglingssterblichkeit, Geburtenhäufigkeit, Eheschließungen und Gesamtsterblichkeit in Berlin während des Krieges. *Jahrbücher für Nationalökonomie und Statistik* 52: 548-554.
- Gutberger, H. (2006), Demographie und Sozialstrukturforschung. Überlegungen zu einem Vergleich zwischen amerikanischer und deutscher Sozialdemographie 1930-1960. *Historical Social Research* 31(4): 155-182.
- Heer, D. M., D. A. May (1968), Son Survivorship Motivation and Family Size in India: A Computer Simulation. *Population Studies* 22: 199-210.
- Hübler, O. (1991), Einkommensdiskriminierung von Frauen und geschlechtsabhängige Einkommensdeterminanten. *Jahrbücher für Nationalökonomie und Statistik* 208: 607-624.
- Hoffman, W. G., J. H. Müller (1959), *Das deutsche Volkseinkommen 1851-1957*. Tübingen.
- Kendzia, M. J., T. Pierenkemper (2010), Der vormoderne Allokationsprozess von Arbeit in Deutschland. IZA Discussion Paper No. 4962.
- Kendzia, M. J. (2010a), *Herausbildung erster Wesenszüge des Normalarbeitsverhältnisses in Deutschland*. IZA Discussion Paper No. 5107.
- Kendzia, M. J. (2010b), *Der Institutionalisierungsprozess des Lohnarbeitsverhältnisses vom Ersten bis zum Zweiten Weltkrieg in Deutschland*. IZA Discussion Paper No. 5231.
- Kleinhenz, G. D. (2004), Bevölkerung und Wachstum: Die Bevölkerungsentwicklung in Deutschland als Herausforderung für die Wirtschafts- und Sozialpolitik. *Jahrbücher für Nationalökonomie und Statistik* 224: 74-90.
- Knapp, G. F. (1873a), Aenderungen der Sterblichkeit vom Jahre 1751 bis 1870. *Jahrbücher für Nationalökonomie und Statistik* 22: 141-144.
- Kohler, H. P., F. C. Billari, J. A. Ortega (2002), The Emergence of Lowest-Low Fertility in Europe During the 1990s. *Population and Development Review* 28(4): 599-639.
- Kollmann, P. (1915), Zur Geburtenfrage. *Jahrbücher für Nationalökonomie und Statistik* 104: 251-253.
- Kurz, R. (1982), Wirtschaftswachstum bei stagnierender und schrumpfender Bevölkerung. *Jahrbücher für Nationalökonomie und Statistik* 197: 235-250.
- Mackenroth, G. (1934), *Deutsche Industriepolitik 1933*. *Jahrbücher für Nationalökonomie und Statistik* 140: 205-224.
- Malthus, Th. (1798), *An Essay on the Principle of Population*. London.



- Messance, L. (1767), *Recherches sur la population des généralités d'Auvergne, de Lyon, de Rouen et de quelques province et villes du royaume avec des reflexions sur la valeur du bled tant en France qu'en Angleterre depuis 1674 jusqu'en 1764* par M. Messance. Paris.
- Mincer, J. (1963), Market Prices, Opportunity Costs and Income Effects. Pp. 67-82 in: C.F. Christ (ed.), *Measurement in Economics*. Stanford.
- Michalke, O. (1935), Die Frauenarbeit. *Jahrbücher für Nationalökonomie und Statistik* 142: 435-448.
- Müller, K.V. (1935), *Der Aufstieg des Arbeiters durch Rasse und Meisterschaft*. München.
- N.N. (1871), Die Bevölkerung des deutschen Reiches nach der Volkszählung vom 1. December 1871 verglichen mit der Bevölkerung von 1867. *Jahrbücher für Nationalökonomie und Statistik* 19: 135-137.
- Notestein, F.W. (1945), Population – The Long View. Pp. 36-57 in: T.W. Schultz (ed.), *Food for the World*. Chicago.
- Oldenberg, K. (1923), Zusammenhang zwischen Geburtenhäufigkeit und Säuglingssterblichkeit. *Jahrbücher für Nationalökonomie und Statistik* 120: 351-353.
- OECD (2011), *OECD Family database, Doing Better for Families Germany*. Paris.
- Overbeek, J. (1974), *History of Population Theories*. Rotterdam.
- Pollack, R.A., S.C. Watkins (1993), Cultural and Economic Approach to Fertility: Proper Marriage or Mésalliance? *PDR* 19: 467-496.
- Prinzing, F. (1899), Die Entwicklung der Kindersterblichkeit in den europäischen Staaten. *Jahrbücher für Nationalökonomie und Statistik* 72: 577-635.
- Repetto, R. (1972), Son Preference and Fertility Behavior in Developing Countries. *Studies in Family Planning* 3(4): 70-76.
- Rohr (1891), Die Bevölkerung der deutschen Groß- und Mittelstädte. *Jahrbücher für Nationalökonomie und Statistik* 57: 111-127.
- Romer, C.D. (1993), The Nation in Depression. *The Journal of Economic Perspectives* 7(2): 19-39.
- Rosenzweig, M.R., T.P. Schultz (1982), Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India. *The American Economic Review* 27(4): 803-815.
- Rostow, W.W. (1960), *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge.
- Schaich, E. (1998), Der Geburteneinbruch in den neuen Bundesländern seit 1990 und einige Hypothesen zu seiner Erklärung. *Jahrbücher für Nationalökonomie und Statistik* 217: 93-107.
- Schultz, T.P. (2009), *Population and Health Policies*. IZA Discussion Paper No. 4340, Bonn.
- Seutemann, K. (1906), Die Hauptergebnisse der Volkszählung im Deutschen Reiche vom 1. Dezember 1905. *Jahrbücher für Nationalökonomie und Statistik* 87: 81-91.
- Seutemann, K. (1907), Der Stand der Statistik der Bevölkerungsbewegung im Deutschen Reich und die Hauptzüge der Bevölkerungsentwicklung in den letzten 15 Jahren. *Jahrbücher für Nationalökonomie und Statistik* 88: 289-313.
- Statistisches Bundesamt (2012), *Ergebnisse des Mikrozensus und der Arbeitskräfteerhebung, Wiesbaden*. <https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/Arbeitsmarkt/Erwerbsstaetigkeit/Arbeitskraefteerhebung/ArbeitskraefteerhebungMikrozensus.html>
- Taeuber C., I.B. Taeuber (1940), German Fertility Trends, 1933-1939. *American Journal of Sociology* 46(2): 150-167.
- Thompson, W.S. (1929), Population, *American Journal of Sociology* 34(6): 959-975.
- von Scheel, H. (1866), Untersuchungen über den Einfluss der Fruchtpreise auf die Bevölkerungsbewegung. *Jahrbücher für Nationalökonomie und Statistik* 6: 161-185.
- von Scheel, H. (1874), Frauenfrage und Frauenstudium. *Jahrbücher für Nationalökonomie und Statistik* 22: 1-17.
- von Ungern-Sternberg, R. (1937), Wirtschaftliche Konjunktur- und Geburtenfrequenz. *Jahrbücher für Nationalökonomie und Statistik* 145: 471-488.
- von Zwiedineck-Südenhorst, O. (1941), Psychologie und Biologie in der Ausdeutung der Geburtenstatistik. *Jahrbücher für Nationalökonomie und Statistik* 145: 160-176.



- Wagner, A. (1980), Der Geburtenrückgang als Ursache von Arbeitslosigkeit? Einige Bemerkungen zum Günther-Paradoxon. *Jahrbücher für Nationalökonomie und Statistik* 195: 261-269.
- Wagner, A. (1985), Nur Transfers zur Anpassung an den Bevölkerungsrückgang? *Jahrbücher für Nationalökonomie und Statistik* 200: 542-550.
- Zimmermann, K.F. (1985), *Familienökonomie*. Berlin u. a. O.
- Zimmermann, K.F. (1988), Wurzeln der modernen ökonomischen Bevölkerungstheorie in der deutschen Forschung. *Jahrbücher für Nationalökonomie und Statistik* 205: 116-130.
- Zimmermann, K.F. (1989), Die Konkurrenz der Genüsse: Ein Brentano-Modell des Geburtenrückgangs. *Zeitschrift für Wirtschafts- und Sozialwissenschaften (ZWS)* 109: 467-483.

Dr. Michael J. Kendzia, Institute for the Study of Labor GmbH (IZA), Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany.  
kendzia@iza.org

Prof. Dr. Klaus F. Zimmermann, Institute for the Study of Labor GmbH (IZA), Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany.  
director@iza.org

# Monetary Theory and Monetary Policy: Reflections on the Development over the last 150 Years

Otmar Issing and Volker Wieland\*

Goethe University of Frankfurt

JEL B22; E42; E52

Monetary policy; history of economic thought; monetary policy regimes; German monetary history.

Received: 27.3.2012

Revision received: 22.10.2012

Accepted: 15.11.2012

---

## Summary

In this paper, we provide some reflections on the development of monetary theory and monetary policy over the last 150 years. Rather than presenting an encompassing overview, which would be overambitious, we simply concentrate on a few selected aspects that we view as milestones in the development of this subject. We also try to illustrate some of the interactions with the political and financial system, academic discussion and the views and actions of central banks.

## 1 The early days

The birth of the “Journal of Economics and Statistics” in 1863 sets the start for our reflections on the development of monetary theory and monetary policy over the last 150 years. It is obvious that it would be overambitious to try to cover this period in an all-encompassing manner. Therefore we will concentrate on a number of aspects which we see as milestones in the development of this subject. In this context we will also refer to some articles published in the Journal which reflect these developments.

Like hardly any other field of economics the development of monetary theory and monetary policy in the course of time reflects the influence and interaction between the political and financial system, academic discussion, and the views and actions of central banks. In the words of Wicksell (1906: 3/4): “...the choice of a measure of value, of a monetary system, of currency and credit legislation – all are in the hands of society, and natural conditions (e.g. the scarcity or abundance of the metals employed in the currency, their chemical properties, etc.) are relatively unimportant. Here, then, the

---

\* This paper has been prepared as a contribution to the Special Issue on “150 years Journal of Economics and Statistics” of the Journal of Economics and Statistics/Jahrbücher für Nationalökonomie und Statistik. We would like to thank Peter Bernholz, Knut Borchardt and Heinz Rieter, the editor, Peter Winker, and three anonymous referees for valuable suggestions on the development of monetary theory in Germany. Thanks to Florian Hense for assistance in the literature review. All remaining errors are our own.

rulers of society have an opportunity of showing their economic wisdom or folly. Monetary history reveals the fact that folly has frequently been paramount; for it describes many fateful mistakes.”

At the time when the *Journal* was first published, the state of monetary theory in “Germany” was anything but well developed (Rieter 2002). To a large extent this was due to the fact that Germany as a state was created only later (in 1871) implying that a large number of currencies existed and that the financial system was very fragmented and hardly developed (Häuser 1998). This is a special case of a general observation: The development of monetary theory through the centuries is strongly related to the state of the financial and monetary system of the time (Born 1981). This is e.g. true for the “discovery” of the quantity theory of money at a time when large price increases were observed in the 16<sup>th</sup> century, and Bodin (1577) and others identified the increase in money resulting from the Spanish conquests in America and the following silver and gold imports in Europe as the cause for rising prices throughout Europe.

Later centuries saw a disastrous experience with paper money (John Law), the problem of convertibility triggered by wars, the choice of gold versus silver or of both (bimetallism), and the debate on what should be counted as money (see e.g. Rist 1938). These phenomena in real economies were reflected in theoretical discussions in the context of the Bullion Report and the currency-banking controversy. During that period the discussions in the German economic literature remained on the sidelines. However, this changed after the middle of the 19<sup>th</sup> century. Although his work is widely forgotten Carl Knies (1873, 1876, 1879) was an outstanding figure in his time and the thorough analysis of the functions of money remains exemplary (Häuser 1996).

The debate between supporters of the banking and currency theory had immediate practical consequences in the form of the organization of central banks. The foundation of the Reichsbank in 1876 reflects the dominance of the reception of the banking theory in Germany (Holtfrerich 1988): the circulation of money was seen as being dominated by developments in the real sector. Adolph Wagner, who had already triggered a discussion in Germany on the Banking Act of 1844, is the main representative for the dissemination of the banking theory (Wagner 1873).

Some articles in the early years of the *Journal* reflect the discussions in England, however, contributions on the situation in Germany dominated. Theoretical papers do not deserve further comment from today’s perspective as theory remained without any significant impact on the evolution of monetary theory. The publication by Laspeyres (1871) on the proper calculation of price developments represents an early highlight in the field of measurement.

## **2 Wicksell and Hayek – a neglected dissent**

Before continuing the line of connecting changes in the real world of money and finance with developments in theory we would like to highlight a dissent between two eminent researchers in this field. Knut Wicksell’s “Geldzins und Güterpreise” (1898) is a milestone in monetary theory. As German was the publishing language of many Scandinavian economists at that time, it is not surprising that Wicksell expressed his ideas first in an article in the “*Jahrbücher*” (Wicksell 1897).

Wicksell’s seminal contribution is based on his distinction between the natural and the nominal interest rate and the consequences for the development of prices once

the nominal interest rate differs from the natural rate. Hayek (1931) explicitly praises Wicksell for this innovative idea. However, he comes to the conclusion that with all merits of Wicksell's approach he makes a mistake by claiming that the coincidence in equilibrium of the natural and the nominal interest rate represents always a constellation which keeps also the price level constant (Hayek 1931: 22). The consequences of Hayek's correct critique are far reaching. If we define neutrality of money as a situation in which the interest rate is in equilibrium, a situation of a stable price level is not necessarily consistent with this condition (Hayek 1933).

For Hayek, not a constant price level, but the neutrality of money, i.e., the idea that money does not influence, that is to say distort relative prices, is the benchmark for the conduct of monetary policy. The implicit challenge is obvious: What are the consequences for the economy if monetary policy follows the goal of price level stability (or low and stable inflation)? Is the implicit "non-neutrality" relevant? Is "neutrality" a concept of any relevance for the conduct of monetary policy? Hayek's discussion of a constant money supply sounds even odder today than 80 years ago. What is the relevance of the "optimal money" developed by Friedman (1956)?

For his well known book Woodford (2003) not only uses deliberately the same title as Wicksell, but also takes the same position on this issue. Interestingly, Hayek and his critique are not even mentioned. As Woodford's approach had such a big influence on monetary theory and policy recently, it might be interesting to resume this debate.

### 3 From gold to paper

With respect to the origin of the "Jahrbücher", we have tried to give a very short summary of the state of monetary theory in Germany at that time. The journal also contains a number of reviews of important books. Although this overview is still very rudimentary, due to lack of time and space we will now concentrate on main developments. This selection reflects our priorities and is anything but encompassing.

The 19<sup>th</sup> century finally saw the triumph of the gold standard following the British example. Monetary policy was dominated by the principle of gold convertibility, i.e., it was constrained by the balance of payments<sup>1</sup> and monetary theory explained how gold movements influenced the quantity of money and thereby the price level in individual countries. In Germany this period was characterized by the coincidence of several fundamental factors, namely a new state (1871), a new currency (1873), and central bank (1875), as well as deep changes in the economy (Borchardt 1976).

The collapse of the gold standard in the context of World War I caused the deepest change of the monetary system in the history of mankind. Before, with a few exceptions money either consisted of physical entities of scarce goods or was "backed" by gold or silver. In theory a debate had started on the "nature" of money and its functions (see e.g. Menger 1970). A special German contribution by Knapp (1905) defined money by its legal tender character (for a discussion see Ellis 1934 and Krohn 1986) – money was a "creation of the state", an idea which obviously was not helpful to explain the emergence of inflation and to develop a solution for ending the disaster.

<sup>1</sup> Bloomfield (1959) argued that there was a habitude in managing the currency. For a critique, see Issing (1965) who argues that the principles of the gold standard dominated over "management". On the currency reforms after the founding of the Deutsche Reich and the "automaticity" of the gold standard see also Borchardt (1976).

Conceptually, paper money (to be correct: a paper standard), i.e., a currency with de facto zero production costs, “represents the logical culmination of the history of the development of money” (Helfferich 1923: 665). With the outbreak of the First World War in 1914, Germany was de facto under a paper standard and the government (Reichsregierung) had direct access to central bank credit. Unfortunately, the experience in Germany at that time revealed the rather dire side of paper money. After the lost war a return to the gold standard was not an option. To conduct monetary policy under these circumstances was obviously a tremendous challenge.

The interaction between the economic situation and the development of monetary theory became most intensive during the period of hyperinflation in Germany after the First World War ending in the collapse of the currency (Mark) in November 1923. Over this debate one should not overlook the fact that monetary theory before had made substantial progress. Already in the first edition of his most influential book von Mises (1912) had warned against any interference of the state into the value of money. Schumpeter (1917) starts from an “income approach to the value of money” (Rieter 1971) and then analyses the quantity theory of money in all its aspects, and finally comes to the conclusion that the general level of prices is determined by the quantity of money. One might think that this should have equipped the central bank with a solid theoretical fundament.<sup>2</sup> As will be shown below the opposite was true.

Inflation had started already during the war and finally accelerated to record levels in the years 1922/23.<sup>3</sup> The discussion on the explanation of inflation went along the lines of the old “Bullion versus Banking School” controversy (Holtfrerich 1988; Rieter 1971).

The Reichsbank supported the “balance-of-payments theory”. As knowledge of this approach might have disappeared, Figure 1 shows its main elements (Eucken 1923).<sup>4</sup>

The initial cause for inflation is the deficit in the balance of payments, due to the reparation payments which via the transmission mechanism pictured in Figure 1 finally ends in inflation. As a consequence of this interpretation, the Reichsbank rejected any responsibility for inflation even when inflation reached unprecedented levels. President Havenstein defended the Reichsbank against its critics, e.g. in a meeting of the Zentralausschuss on 25 August 1923 arguing: “There is no doubt that by giving credit, a central bank increases the circulation of banknotes; however, insofar the central bank gives credit which is economically justified and necessary, as it serves production and turnover of goods, it does not create artificial purchasing power” (Protocol, p. 2; translation by Issing). At the same time the Reichsbank apologized for not being able to satisfy fully the demand for banknotes and justified the decision to keep the discount rate constant (since 1914) at 5 % until 28 July 1922.

A prominent opponent of this explanation of the causes of inflation by the Reichsbank and its monetary policy was Eucken. He rejected the balance of payments theory, and on the basis of his quantity theory approach saw the cause in the increase in the quantity of money triggered itself by the public deficit and the artificially low central bank interest

---

<sup>2</sup> An article by Bendixen (1919) published in this journal might be mentioned where the author is criticizing the concept of “metallism” as the fundament of money and develops arguments in favour of Knapp’s theory.

<sup>3</sup> It is interesting to note that it were foreign authors like Bresciani-Turroni (1937) who published the first authoritative book on the German hyperinflation.

<sup>4</sup> The English version is taken from Holtfrerich (1986).



**Figure 1** Balance of payments theory

rate. Following his analysis, inflation could only be stopped by closing these sources with the consequence of restoring the needed scarcity of money.

For Eucken the interpretation of the causes of inflation by the Reichsbank was totally flawed, and for modern readers it looks just odd, and is of interest only because it inevitably led to hyperinflation. So it is no surprise that this theory has totally disappeared (however, not the notion of the endogeneity of money creation).

So, the Reichsbank has to take the full responsibility for the inflation because it was misguided by a wrong theory?<sup>5</sup> Borchardt (1972) has a rather sobering argument on this debate. In short, in his opinion the war with all its consequences, especially the fact that it was lost left hardly any option to contain inflation. Therefore, any theory of inflation not respecting the political circumstances cannot explain properly what happened in Germany in these years.<sup>6</sup>

Von Mises can be seen as the leading expert of monetary theory in this period. It would take much more space to do justice to all the aspects analysed in his impressive work "The Theory of Money and Credit" (1980), the English translation of the second edition (1924) of the book already mentioned above. He already develops the purchasing power theory. He strictly rejects Knapp's legal concept of money, and systematically applies marginal utility theory to explain the value of money. Whereas this approach, adopted by Wieser (1927), had no lasting impact on monetary theory, his analysis of the detrimental effect of the influence of the government on the value of money was validated since time and again. His conclusion is short and strict: "Sound money still means what it meant in the nineteenth century: the gold standard" (von Mises 1980: 480). In his pleading for currency competition Hayek (1990) refers repeatedly to the work of von Mises and his argument that it was not "capitalism" but government intervention which has

<sup>5</sup> Initially, the Reichsbank saw its policy being supported by an international consensus of central bankers. However, this changed in the course of 1923 (James 1999).

<sup>6</sup> For a short survey of the political circumstances see e.g. Rieter (1971). Cagan (1956) in his seminal work shows that an extreme rise in prices depends almost entirely on changes in the quantity of money. However, a precise analysis of the determinants of money supply cannot be purely mechanistic, but would involve the motives of governments, "with whom the authority to open and close the spigot of the note issue ultimately lies".

been responsible for the recurrent crises of the past. But for Hayek, currency competition rather than the gold standard is the solution.

Overall, these and other publications not mentioned here indicate that German speaking economists had closed the gap relative to the international standard of monetary theory which had existed before. This is e.g. demonstrated by a secret conference in the middle of the crisis in September 1931 which brought together important political actors and leading academics which by itself is a remarkable and rare event. At this gathering the “*crème de la crème*” (Borchardt/Schötz 1991) discussed whether and how the Reichsbank could and should help to finance measures to create jobs. Although the discussion was focused on a plan to stimulate the economy via deficit spending by the “German Keynes”, Wilhelm Lautenbach, important arguments were based on monetary theory. The quality of the exchange of views in these two days at the conference is also an indication of the loss in intellectual capacity thereafter due to the brain drain and the isolation during the Nazi regime.

#### **4 Currency reform in Germany 1948**

Only one generation and another lost war later, Germany was again confronted with the challenge of a currency ruined by inflation. This time the destruction of the value of money was not evident in the inflation rate as prices and wages were controlled (the so-called “stopped inflation”). As a consequence, a huge monetary overhang had to be dealt with. Numerous plans were developed how to solve the problem (Möller 1961). The currency reform of 1948 was encompassing as it addressed not only monetary and banking issues but also matters beyond money like distributional aspects (Lastenausgleich). On 20 June 1948 the new currency was introduced – an act which also implied monetary division between West Germany and the East. The impact of the introduction of the Deutsche Mark on the (West) German economy has to be seen in context with the measures to liberalize the economy for which Ludwig Erhard was responsible (Möller 1976; Buchheim 1999).

As it turned out with the new currency (West) Germans for the first time in two generations experienced a stable currency. Indeed, the Deutsche Mark became one of the most stable currencies in the world. Over its 50 years existence the average annual inflation rate was 2,6 %. What is of particular interest in this context is to what extent this was due to following the right monetary theory in the conduct of monetary policy. Two aspects are key:

The first is the institutional arrangement initially established by the allies and later included in the law on the Deutsche Bundesbank (1957) which gave the central bank independence and a mandate for maintaining price stability.<sup>7</sup> For the development of monetary theory and policy this is a remarkable fact as the independence of the central bank was at that time hardly an issue in the international discussion, and when it was raised there was not much sympathy for such a statute. It was only after the “great inflation” of the seventies when institutional aspects were intensively discussed and empirical studies revealed a strong correlation between central bank independence and inflation (Issing 1993).

---

<sup>7</sup> Article 3 of the Bundesbank act defined the aim as “safeguarding the currency” which included both external and internal stability – a combination impossible to be realized in a world of inflationary developments. However, this legal goal was more and more interpreted as stability of domestic prices.



It is true that the success of the monetary policy of the Bank deutscher Länder and then the Deutsche Bundesbank did not only rely on its legal status but was also firmly based on the support by a population which had lost its wealth invested in financial assets twice in a generation. But there is no doubt that the example of the Bundesbank played also a role in the development of institutional aspects of monetary policy.

The second issue concerns the theoretical fundament of the German central bank's monetary policy (Neumann 1999). The early years were dominated by a pragmatic approach within the constraints of a fixed exchange rate (Emminger 1976; Schlesinger 1976). Monetary targeting was the new strategy announced in 1975 for the first time.

## **5 After World War II: Paper standard under fixed exchange rates**

In contrast to the situation after World War I the idea of a return to the gold standard was totally discarded. Paper money standard remained the only option. As a consequence of the Bretton Woods Conference in 1944 and the following statute for the International Monetary Fund, the Western world entered a period of fixed exchange rates. Once foreign exchange controls were abolished the idea that under such a regime monetary policy could be directed towards domestic goals like price stability or full employment turned out to be an illusion. A fixed exchange rate, free capital flows and a sovereign monetary policy form an “impossible trinity” (Mundell 1973) which means that only two of the three goals are mutually consistent.

The collapse of the regime of fixed exchange rates in 1973 opened a new era for the conduct of monetary policy on the basis of a paper standard. Once again, a new period started in which the interaction between economic events, not least the “great inflation” of the seventies, and the development of monetary theory opened a new chapter.

This is also true for the “opposite direction” most notably to be observed in Europe where first a regional regime of fixed exchange rates (European Monetary System) was established before the ultimate solution of creating a single currency shared by initially 11 and later (2012) 17 countries was adopted.

## **6 Keynesianism versus monetarism and the impact on monetary policy**

The development of monetary theory since the end of World War II is marked by a series of traditional topics as well as by new aspects. On the neutrality and value of money Patinkin (1965), for example, must be mentioned (also in the context of the real-balance-effect). For a while the controversy on inside-outside money triggered a lively debate (Gurley/Shaw 1976; Pesek/Saving 1969). There is a legion of publications. For an early survey see Johnson (1962); encompassing Friedman and Woodford (2011).

Although these theoretical developments and the accompanying empirical work on, e. g., demand and supply of money are also of relevance for monetary policy, we will concentrate here on the main controversy which had a direct impact. This controversy can be summarized under the headline of this paragraph: Keynesianism versus monetarism.

The 1950s and 60s were the time when Keynesianism dominated worldwide. Keynesianism here means the monetary theory originating in Keynes' “General Theory”, developed further by Hicks, Modigliani, and many other proponents which Herbert Stein (1990) coined “simple-minded Keynesianism”. The view that there was a permanent and stable

trade-off between inflation and unemployment was propagated by an influential study on the Phillips-Curve by Samuelson and Solow (1960). As a consequence society seemed to be able to choose according to the “menu”. Fiscal policy was the powerful instrument to guarantee full employment. Monetary policy had no major role to play and should be coordinated with the government’s policy. Tobin’s (1965) contribution worked in the same direction. The view that “money” was of minor (if any) importance was a main element of the “liquidity theory” for which the Radcliffe Report (Committee 1959) can be seen as a representative survey.

As far as monetary policy was based on theory central banks world-wide followed these ideas. In the first place this is true for the Fed as it is analysed in detail in Allan Meltzer’s authoritative history (2009). The philosophy of “money does not matter” is encapsulated in a remark by the former chairman William McChesney Martin, Jr. “They don’t really know what the money supply is now, even today. They print some figures ... but a lot of it is just about superstition.”<sup>8</sup>

The reception of Keynesianism in Germany after 1945 happened step by step reflecting a dispute between the older rather negative and the younger, supportive generation (Richter 1999a,b). A bundle of objectives was the orientation for the central bank’s monetary policy. Bank liquidity was the main target of the Bundesbank’s actions. Over time the concept of free liquid reserves was the guideline for the conduct of monetary policy.

However, under the regime of a fixed exchange rate, even during the years of foreign exchange controls, the conduct of monetary policy was confronted with the balance of payments restriction (Emminger 1976; Richter 1999b). Having been a strict supporter of a fixed exchange rate for the DM against the US dollar, after the introduction of convertibility the Bundesbank’s policy was more and more marked by its fight against the impact of the consequences of surpluses in the balance of payments. To what extent external considerations dominated monetary policy decisions is, for example, demonstrated by the fact that in November 1960 the Bundesbank lowered the discount rate from 5 % to 4 %, in order to reduce the interest rate spread to foreign markets, notwithstanding the fact that the domestic situation would have required the opposite. Over the period from 1951-1973 the foreign component dominated the source of the money base (see e.g. Issing 1996). Under these circumstances the Bundesbank moved gradually in its position from supporting appreciations of the exchange rate to later favouring a regime of a flexible exchange rate. Eventually, price stability had turned into the single monetary policy objective.

While the Keynesian doctrine was still dominating, an intellectual challenge was emerging which was later called the monetarist counterrevolution. Milton Friedman was the key figure publishing a series of influential papers. His voluminous “A Monetary History of the United States” (1963) written with Anna Schwartz became sort of the benchmark of how empirical studies should be conducted. In his “A Program for Monetary Stability” (1960) he proposed a fundamental change of the financial system (100 % minimum reserves) and his famous k-percent rule, according to which money supply should increase steadily by 4 % year per year. This extreme approach was never considered as a practical advice for the conduct of monetary policy and later also rejected by Friedman himself.

---

<sup>8</sup> Quoted in Meltzer (2009: 267).

However, the main message of monetarism (Friedman 1968; Brunner/Meltzer 1989) had a lasting influence on monetary theory and policy (Laidler 1981).

1. Money demand is a stable function of a few key variables.
2. Discretionary monetary policy – not least because of long and variable time lags – causes volatility in output and employment and has no permanent impact.
3. The Phillips trade-off cannot be exploited, unemployment is determined by the natural rate.

As a consequence monetary policy should follow strict rules for the control of money geared towards price stability. Tobin (1980) coined the term Monetarism Mark I in order to distinguish it from Mark II, which included Robert Lucas's (1972) contribution that markets forming rational expectations make any attempt of a discretionary systematic monetary policy ineffective. The so-called Lucas critique (1976) explained why the structural parameters of existing models change under the influence of policies and could therefore not be used for the simulation of outcomes.

The emergence of monetarism triggered a debate which can be seen as one of the most productive developments not only in monetary theory but for macroeconomics as a whole. Two new journals were founded which soon became a platform for a worldwide discussion (*Journal of Money, Credit and Banking*; *Journal of Monetary Economics*). In Germany Manfred Neumann (1971) became the leading German monetarist (for his assessment of the Bundesbank's policy, for example, see Neumann 1999).

These developments in monetary theory were a challenge for central banks. The Deutsche Bundesbank can be seen as an outstanding example of a timely, but reflected reaction to new research. The move from a fixed exchange rate to floating on March 19, 1973 created the fundament for the choice of a monetary policy strategy geared to a domestic goal which was price stability. As the first central bank in the world the Bundesbank in December 1974 announced a growth target for the money stock in 1975. The choice of a monetary target signalled a fundamental regime shift. This decision was based on two arguments. First, and foremost, was the adoption of an intermediate target, i.e., the intention of controlling inflation through the control of monetary expansion. Second, the Bundesbank tried to provide a guidance for agents', especially wage bargainers' expectations through the announcement of a quantified objective for monetary growth (Schlesinger 1983).

The Bundesbank was convinced that, while monetary policy maintaining price stability in the longer run would exert a positive impact on economic growth, fostering potential growth in the economy should be considered a task of fiscal structural policies, while employment was a responsibility of the social partners conducting wage negotiations.

The new strategy was seen as an experiment and the first experiences with monetary targeting were not particularly encouraging. However, the Bundesbank had made it clear from the beginning that it could not and would not promise to reach the monetary target with any degree of precision.

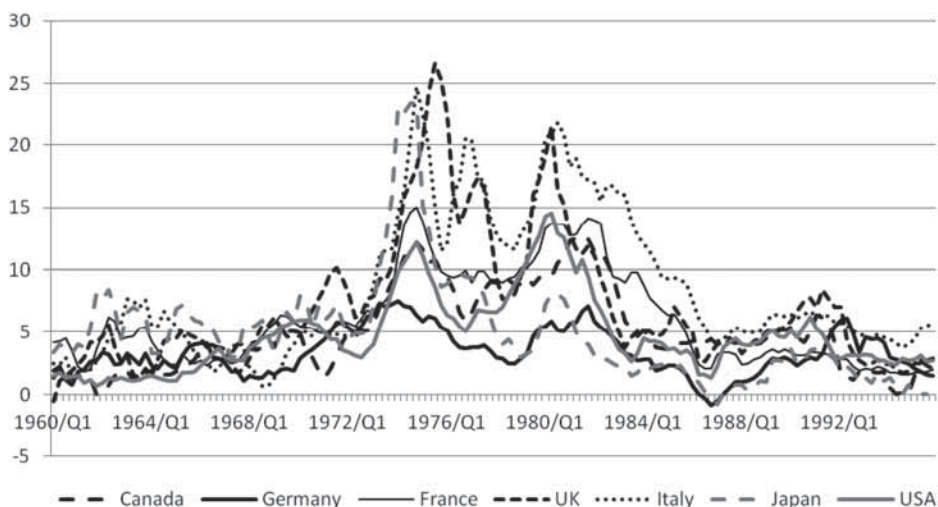
After the first years of experience the Bundesbank enhanced the monetary target concept from its experimental stage into a fully fledged strategy (Issing 1992, 1997; Neumann 1999). While details of the concept were adapted from time to time, monetary targeting remained the strategy until the end of the DM in 1998. The Bundesbank on the one hand based its strategy and the consequent monetary policy to a large part on the monetarist theory, the flexibility which it demonstrated in the implementation, on the other hand however, raised critique by the monetarists right from the beginning.

“Pragmatic monetarism” as this policy has been called – which was in line with the central bank’s own interpretation – met its first hard test in the seventies, and successfully stood the test. Germany avoided the “great inflation” which hit other countries (Issing 2005; Beyer et. al. 2008). Confidence in the DM as a stable currency was maintained also throughout the nineties during the turbulences of unification.

## 7 The “great inflation” and rational expectations theory

This is not the occasion for a comparative study. However, it is interesting to briefly contrast monetary policy and its results in Germany relative to other countries during the seventies. Figure 2 shows inflation in the G7 economies from 1960 to 1995 measured by the rate of change of the Consumer Price Index. It is no surprise that commentators have dubbed the seventies the period of the “Great Inflation”. Inflation reached double digits in 1974/75 in all of them but Germany – and a second time around in 1980/81 except for Germany and Japan. Initial price rises emanated from oil prices, but central banks were responsible for the lasting impact.

The most important reasons for the U.S. Federal Open Market Committee’s disappointing performance during this period can be seen in the continuation of a discretionary monetary policy that was focussed on two targets – employment and inflation – and was misguided due to unreliable indicators like the output gap (Orphanides 2002), reliance on the Phillips curve trade-off and neglect of money (Meltzer 2009). When this policy ended in stagflation, the FOMC under Chairman Volcker reoriented its policy in the direction of pragmatic monetarism (Meltzer 2009). While the Bundesbank was confronted with similarly biased output gap estimates (Beck/Wieland 2008), its commitment to the primacy of the price stability objective and the monetary strategy helped it avoid the inflation excesses experienced in the United States and elsewhere.



Notes: 4-quarter average rates of change of the Consumer Price Index (CPI) in percentage points. Source: OECD.

**Figure 2** Inflation in the G 7 Economies – 1960 to 1995

The “great inflation” had a huge impact on monetary theory. Against this background, rational expectations theory quickly came to dominate macroeconomic research. The expectations-augmented Phillips curve (Phelps 1967) implied that only inflation surprises would have an impact on real GDP and employment. Thus, anticipated changes in monetary policy would not affect the real economy (Sargent and Wallace 1975, Taylor 1975). Kydland and Prescott (1977) showed that a central bank attempting to exploit the inflation-unemployment tradeoff would only induce greater inflation without raising employment – the so-called time inconsistency problem. Since this outcome may be cast as a Nash equilibrium in a dynamic game between the central bank and private sector price setters, a substantial literature ensued that applied game theory to monetary policy analysis. It emphasized the importance of reputation, credibility and transparency in policy making (cf. Barro/Gordon 1983, Cukierman/Meltzer 1986) and provided theoretical foundations for the institutional independence of central banks (Cukierman 1992). Research in this vein also indicated that the time inconsistency problem can be alleviated and inflation reduced, if central banks are assigned the pursuit of price stability as their primary objective and use explicit monetary targets (Rogoff 1985; Garfinkel/Oh 1993), as in the case of the Bundesbank.

## 8 From the Lucas critique to the Taylor rule

While monetary policy modeling based on the natural rate hypothesis and rational expectations helped clarify how better institutional design can improve inflation performance, its implication that anticipated policy decisions had absolutely no effect on the real economy did not agree with practitioners perceptions and empirical readings. New Keynesian contributions such as Fischer (1977), Phelps and Taylor (1977), Taylor (1979a) and Calvo (1983) used long-term nominal contracts, for example staggered wage contracts, to explain how monetary shifts could cause real fluctuations even if expectations are rational and the shifts are anticipated. Models with such contracts were estimated empirically and found to fit the persistent fluctuations in output and inflation.

As a consequence of the nominal rigidities, monetary policy is faced with a trade-off between inflation and output volatility. This trade-off or Taylor curve is stable in the sense that it is possible for the central bank to pick a particular combination of output and inflation volatilities depending on its preferences. However, this finding did not provide a new argument for discretionary policy. Rather, it strengthened the case for choosing a policy rule that would serve to stabilize macroeconomic outcomes, including private sector expectations. Using an estimated model of the U.S. economy Taylor (1979b) showed that Friedman’s constant money growth rule would have achieved a lower degree of price level and output variability than actual U.S. policy. Furthermore, he derived optimal feedback rules for real money balances. The purpose of such analyses was to present new econometric methods for selecting macroeconomic policy when expectations are formed rationally, thus taking into account the famous Lucas critique.

The new methods caught on and induced a wave of New Keynesian modeling. Models with nominal rigidities and rational expectations were extended and enlarged and eventually applied rather intensively in practical monetary policy analysis at central banks and institutions such as the International Monetary Fund. Models of this type that featured in U.S. policy discussions in the early 1990s include Taylor’s (1993a)

multi-country model, Fuhrer and Moore's (1995) model with staggered real wage contracts and the Federal Reserve's FRB-US model (see Reifschneider et al. 1999 for a description). A version of the latter is still used at the Fed today.

New Keynesian modeling made more direct contact with the practical conduct of monetary policy, by specifying rules for the central bank's main policy instrument – the interest rate on central bank money such as the federal funds rate in the United States. For example, a large-scale international model comparison exercise conducted at the Brookings Institution (Bryant et al. 1993) evaluated the performance of such interest rate rules across a range of models. Interestingly, Taylor (1993a) credits this comparison project as the crucial testing ground for what later became known as the Taylor rule. Thus, he saw the rule as a useful normative guideline. It implied a quantitative prescription for increasing (lowering) the policy rate in response to higher (lower) inflation and deviations of GDP above (below) trend. Such systematic policy could be anticipated by market participants and help induce stabilizing expectations.

Interest in this rule quickly reached far beyond academic and central banking circles, because it matched the Fed's interest rate decisions between 1987 and 1993 surprisingly well, as Taylor had shown. Indeed, after the unavoidable, but costly disinflation policy administered under Chairman Paul Volcker, the Greenspan Fed succeeded in maintaining low inflation without much output volatility during this period. Later on, the period from the mid-1980s to the early 2000s became known as the "Great Moderation" because of the sustained reduction in business cycle fluctuations in output and employment along with low and stable inflation relative to the preceding post-war period. A number of studies such as Giannone et al. (2008) have rejected a "good luck" explanation and scholar such as Taylor (2011) attribute it to a switch from discretionary to rule-based policy making. While no central bank has formally adopted Taylor's rule as its policy strategy, nowadays Taylor-style rules are standard fare in the briefing materials of policy makers and the analysis of central bank watchers.

Before we pick up this thread again, we take a look at developments in monetary policy across the Atlantic and related issues in monetary theory. Following the breakup of the Bretton-Woods system, European governments were heading in a different direction than the United States and started to explore new monetary arrangements hoping to limit exchange rate volatility.

## **9 Europe and the Bundesbank: Via the EMS crisis to European monetary union**

A concrete first step had already been taken at the European Economic Community summit in The Hague on 1 and 2 December 1969. Heads of state or government had agreed that a plan should be drawn up with a view to the creation of an economic and monetary union. Even so, the 1970 plan presented by the "Werner Group" – named after the then Prime Minister of Luxembourg – was not followed by further steps, because of disagreement over appropriate exchange rate policies. A Franco-German initiative eventually helped break the deadlock and the European Monetary System (EMS) came into existence in March 1979 (see Issing 2008).

In the EMS, exchange rates were set between the member currencies and interventions were compulsory if the parities threatened to breach certain bands. Here, we cannot do justice to the substantial literature on the economics of exchange rate target zones that flourished during this period (cf. Krugman/Miller 1992). However, it is important to

note that the European Currency Unit (ECU), though formally at the system's core, only played a limited role as unit of account. Instead, the EMS soon functioned as a Deutsche Mark area, centered on the strongest currency in the system (McKinnon 1993). Under this arrangement and with free capital flows, member countries could either align monetary policy with the Bundesbank, or re-align the parity. Typically this meant devaluing their currencies relative to the Deutsche Mark.

Tensions in the EMS escalated into full-out crisis in 1992/93. As a consequence of re-unification the German economy experienced a boom and rising inflation. The German boom implied a large, asymmetric shock to the system (see Wieland 1996). True to its mandate, the Bundesbank tightened policy and brought inflation back under control. Faced with the choice to raise the interest rate and accept its negative impact on domestic economic activity or to depreciate the currency, the United Kingdom dropped out of the system in 1992 and Italy stopped interventions. The permissible bands between the currencies remaining in the system were substantially widened in 1993. Notwithstanding substantial macroeconomic tensions in the context of re-unification the Bundesbank continued its strategy of monetary targeting. To signal its determination to keep inflation under control it derived the target on the basis of an unchanged number of two per cent for the price norm, but accepted the following overshooting (Issing 1992).<sup>9</sup> It succeeded finally in bringing down inflation.

The woes of the EMS did not distract but rather hasten the political process towards European monetary union (EMU).<sup>10</sup> The Maastricht treaty of December 1991 set out criteria for economic convergence and fixed the latest starting date for EMU at January 1, 1999. And indeed, by that date a new currency and a new central bank had been created. Eleven sovereign countries adopted the Euro and delegated monetary policy to the European Central Bank. Principles of monetary theory, such as central bank independence, primacy of the goal of price stability, and the prohibition of monetary financing were enshrined in treaties and European law. Fiscal authorities' tendencies towards excessive deficits and debt were meant to remain constrained even after convergence by explicit limits to be enforced by a stability and growth pact.

The young central bank aimed to gain credibility by announcing a numerical objective – HICP inflation below two percent over the medium term – as well as a policy strategy. The ECB's strategy became known as the two-pillar strategy, resting on an economic analysis of short- to medium-run inflation pressures, and a separate monitoring of longer-run inflation risks emanating from monetary trends (see Issing et al. 2001; Binder/Wieland 2006). Thus, the ECB neither stuck to Bundesbank-style monetary targeting, nor did it adopt a Taylor rule or direct inflation targeting, but instead created a new policy framework. Its main competitor – inflation targeting – had become rather popular initially among newly-independent central banks in several small open economies since the 1990s. Theoretical foundations for this strategy will become clearer in the course of the next sections.

<sup>9</sup> For a discussion of target misses as the central bank learned about the possibility of shifts in money demand see Wieland (2000). The price norm was the Bundesbank's version of a numerical price stability objective.

<sup>10</sup> For a thorough review see Issing (2008).



## 10 Real business cycle theory versus the New Keynesians

Lucas (1976) had asked for taking full account of the forward-looking and optimizing reactions of businesses and households in the formulation of macroeconomic policies. Ultimately, this approach required building macro models bottom up from the micro-economic utility- and profit-maximizing decisions of households and firms. Following this prescription, Kydland and Prescott (1982) proposed a new theory of macroeconomic fluctuations that became known as the real business cycle (RBC) approach. Their extension of the neoclassical growth model to study the real (rather than monetary) sources of business cycles delivered a modeling approach that stringently enforced all the restrictions following from the utility maximization of representative households and the profit maximization of representative firms on the dynamics of macroeconomic variables. At the same time they put forth technological innovations as the main drivers of business cycles. As to the monetary side, they returned to the policy ineffectiveness proposition of early natural rate theories. In perfect markets, stabilization policy was of no concern.

In the ensuing academic debate with RBC researchers, New Keynesians emphasized the need for including nominal rigidities as well as real market imperfections such as imperfect competition or imperfect information (cf. Mankiw/Romer 1991). Goodfriend and King (1997) and Rotemberg and Woodford (1997) presented a first monetary business cycle model using the approach to microeconomic foundation practiced in RBC research but also including imperfect competition and overlapping nominal contracts. They incorporated key features of earlier New Keynesian research in the dynamic general equilibrium framework used in the RBC literature. Thus, Goodfriend and King named it the New Neoclassical Synthesis model. In recent years it has been commonly referred to as the benchmark simple New Keynesian model.

A linearized version of this model consists of three simple equations, a forward-looking Phillips curve, a forward looking IS curve and a Taylor-style interest rate rule. The New Keynesian Phillips curve relates current inflation to expected future inflation and the deviation of output from a new measure of potential, the level of output that would occur under flexible prices. The IS curve or aggregate demand equation is derived from the consumer Euler equation. Thus, current output depends on expected future output and the expected short-term real interest rate.

The small-scale New Keynesian or New Neoclassical Synthesis model was quickly extended with additional decision aspects and constraints following the contribution of Christiano et al. (2005). Their model was estimated with euro area data by Smets and Wouters (2003). These models are typically referred to as New Keynesian dynamic stochastic general equilibrium (DSGE) models.<sup>11</sup> Nowadays, medium- to large-scale DSGE models are routinely used by economists at central banks and international institutions to evaluate monetary and fiscal stabilization policies. In contrast to the wide array of statistical nowcasting and forecasting models also employed at these institutions, DSGE models can be used for counterfactual policy simulations that respect the Lucas critique.

---

<sup>11</sup> See Taylor and Wieland (2012) for a comparison of earlier New Keynesian models with state-of-the-art DSGE models using a new monetary model database that provides interested researchers with easy access to such models.

## 11 Inflation targeting and the role of money in monetary policy

In the early 1990s more and more countries took seriously the lessons of the “great inflation” and made it possible for their central banks to pursue price stability. To this end, they had to grant central banks a certain degree of operational independence and achieve control of public finances without taking recourse to seigniorage. The rational expectations literature as well as empirical experience in some countries served to highlight the potential for reducing the costs of disinflation by credibly signaling policy objectives to wage and price setters. In 1990, New Zealand and Chile tasked their central banks with a primary responsibility for price stability, and the central banks published official numerical targets for the rate of inflation. According to Bernanke et al. (1999) the responsibility for price stability and the numerical target are the two key elements that characterize the framework for monetary policy termed “inflation targeting”. Soon, other countries followed including Canada, the United Kingdom, Sweden, Norway and Australia.

Inflation forecasts play a central role at inflation-targeting central banks in terms of signaling how they plan to achieve their target in the future, or on average. Some inflation targeting central banks have used Taylor-style rules with forecasts to characterize their strategy in model evaluations (cf. Batini/Haldane 1999). Alternatively, inflation targeting has been described as an optimal control problem with the inflation forecast as intermediate target (Svensson 1997). In this regard, Clarida et al. (1999) proved influential by deriving the implications of the simple benchmark New Keynesian model of the preceding section for the theory of monetary policy. Subsequent New Keynesian contributions have tended to consider inflation targeting an optimal approach to monetary policy (see, for example, Woodford 2007 and Wieland 2009).

The literature on inflation targeting is vast and many aspects of practical policy making have been integrated in its formal optimal control representation. For a recent exposition the reader is referred to Svensson (2010). In the remainder, we only put the spotlight on one aspect, namely the role envisioned for monetary aggregates, which was debated rather vigorously and involved somewhat extreme positions. Woodford (2008), for example, saw no reason to continue assigning a prominent role to monetary aggregates in policy deliberations. In his judgment, research on more refined models of money supply and money demand would not help understanding the kinds of uncertainty about the effects of monetary policy that are the actual obstacles to the development of more effective policy. These conclusions were consistent with the implications of the benchmark New Keynesian model for the relevant transmission channels of monetary policy.

Not surprisingly, the European Central Bank’s continued adherence to a strategy with a prominent role for monetary aggregates triggered much criticism. This debate had important implications for practical policy at the time, because of its relevance to the question whether or not the sustained increase in money and credit growth from 2003 onwards should be a cause for concern.

Other central banks had largely de-emphasized money in the context of their strategies. Few studies argued in support of the ECB’s position. Lucas (2007) expressed skepticism regarding the New-Keynesian model’s ability to explain monetary trends, and proposed that central banks should continue to use monetary information as a kind of add-on or cross-check. Along these lines, Beck and Wieland (2008) suggested that cross-checking with monetary trends would help reduce the negative impact of output gap mispercep-

tions. Others focused on the use of money as a commitment device (Beyer et al. 2008; Christiano et al. 2008). A fundamental critique of inflation targeting was presented by Issing (2011). However, the global financial crisis that broke out shortly thereafter provided ample reason for re-considering the role of money and credit growth in monetary policy.

## **12 In lieu of a conclusion: Some thoughts on monetary policy and the global financial crisis**

Having reflected on almost 150 years of the history of monetary theory and policy, we have returned to the present. The global financial crisis has been going on for almost five years. While it is too soon to summarize its impact on monetary thought and policy in a definitive manner, it is useful to highlight some developments and their likely impact. In doing so, we restrict attention to issues concerning monetary policy. While the financial crisis offers important lessons for financial regulation and supervision, including the responsibilities of central banks in these fields, these policy areas are beyond the topic of our review. Similarly, we refrain from discussing lessons for central banks' role of lender of last resort to financial intermediaries.

Monetary policy is always conducted in an environment with many sources of uncertainty. Unforeseen macroeconomic shocks, imperfectly understood and imprecisely estimated channels of policy transmission and macroeconomic propagation, and noisy empirical measurement render monetary policy design and implementation a challenging task. They were all part of the cast of the financial crisis. The collapse of housing prices and financing in the United States caused substantial and (at least partly) unexpected macroeconomic disruptions. The extent of financial risk, the fragility of financial intermediaries and the interdependence of the exposures of these institutions around the world were not well understood by decision makers in markets and policy institutions. To a significant extent, these sources of risk and disruption arose as unintended negative consequences of regulatory, supervisory, fiscal and monetary policies practiced prior to the crisis. The magnitude of the impact of financial disturbances on the other sectors of the economy was greatly underestimated. To give an example, while the Federal Reserve worried about a potential housing crash, its empirical models derived from historical data underestimated the impact of such a crash on the real economy. Finally, the course of events required policy makers to take recourse to unusual policy measures whose impact was subject to considerable uncertainty.

Certainly, all central banks have to reconsider the policy strategies they pursued prior to the crisis and assess whether these strategies are still appropriate for the future. As to the Federal Reserve's policy prior to the crisis, Taylor (2007) has pointed out that the federal funds rate remained too low for too long relative to the prescriptions of the Taylor rule. He provides empirical evidence that the deviation in interest rates was sufficient to account for at least part of the housing boom. From the Federal Reserve's perspective, the low interest rate policy prior to the crisis was justified by the outlook on inflation and economic activity (Bernanke 2010). Indeed, a Taylor-style interest rate rule, which includes FOMC forecasts of inflation and unemployment rather than recent outcomes, matches Fed policy very well (see Orphanides/Wieland 2008). The same rule, however, would have implied higher interest rates with Blue Chip CPI forecasts instead of the FOMC's prediction regarding inflation in the personal consumption expenditures

(PCE) deflator. Ex-post data revisions indicate that the FOMC under-predicted PCE inflation at the time.

As a consequence of its strategy and in contrast to other central banks, the ECB had to worry about the sustained increase in trend money growth prior to the crisis. At least ex-post, it is understood as a signal of the credit boom that led to inflated asset prices and financial risks around the world. Furthermore, it did have an influence on ECB decision making. Trichet (2008), for example, emphasized that the ECB decided to increase the policy rate in December 2005 against the advice of the IMF and OECD because the ECB's monetary analysis particularly strongly suggested that they should. Despite the policy tightening, money growth stayed high. Thus, it is a fair question whether the ECB should not have given more weight to its monetary pillar in setting policy rates prior to the crisis.

The role of asset prices in monetary policy has also been the subject of much research over the years, specifically whether monetary policy should involve a sort of "leaning against the wind" with regard to the development of asset prices. One view is to abstain from any reaction during the built-up of an asset boom, because of the difficulty involved in defining what is a bubble and what not, but to essentially pre-announce a rescue with ample liquidity provision once the bubble bursts. This view has been dominating for some time, particularly in the United States, given the seeming success of the Greenspan Fed in handling the 1986 stock market crash and other such events later on. This view also explains the reluctance of the Greenspan Fed to tighten policy more quickly during the period of the so-called "dot-com bubble" of the late 1990s and its aggressive easing following the correction in 2001. However, Issing (2009), who refers to this approach as the "Jackson Hole Consensus", points out that such an asymmetric approach may create moral hazard and encourage behavior that induces ever greater asset price bubbles. Instead, a strategy that includes cross-checks against monetary trends would implicitly lean against asset booms. As long as money and credit remain broadly controlled, the scope for financing unsustainable runs in asset prices should be limited. Similarly, Taylor (2007) implies that if the Fed would not have deviated from his rule, housing price increases would have been much more limited (see also Jarocinski/Smets 2008). While these arguments support the conclusion that monetary policy can act to stabilize financial markets, be it with communication or even moderate interest rate adjustments, monetary policy should not be left alone in this task. There are a number of tools, also of a regulatory nature, that can be deployed to ensure financial stability.

Having started this section on the financial crisis by acknowledging several failures and omissions of monetary policy, it is only fair to continue with highlighting some successes. In August 2007, when banks turned reluctant to lend funds to fellow banks and a 60 basis points premium emerged in the inter-bank money market, central banks stepped in and immediately increased liquidity provision. This response is essentially automatic in the standard policy framework that uses the price of central bank funds rather than its quantity as an operating target.

In the fall of 2008, when the ongoing recession became apparent and inflation rates dropped, that is after the Lehman collapse, central banks in leading industrial economies responded by aggressively lowering interest rates. As short-term money market rates started to approach the zero lower bound, central banks resorted to additional measures. Generally, these measures focused on reducing premia associated with longer-term interest rates and riskier assets and the quantitative expansion of base money. While the U.S. Fed focused more on direct asset purchases, the ECB relied to a greater extend

on longer-term repo operations with the banks. Thus, monetary policy helped averting a longer drawn-out recession and avoided any significant deflation. In the euro area, monthly HICP inflation rates dipped only shortly into negative territory, reaching a trough of -0.6 percent (annual percentage change) in July 2009, and quickly returned close to 2 percent by summer 2010. In the United States, monthly CPI inflation rates hovered near zero in the first few months of 2009, dropped to a trough of -2.0 percent in July but then quickly rose above 2 percent by December.

Of course, one may question whether the measures taken were excessive and less would have been sufficient. Also, the proper balancing of deflation scenarios with medium-term inflation risks is rightly debated. For the purposes of this paper, however, we want to emphasize that central banks were not surprised by the need for additional instruments with near-zero nominal interest rates. Ever since the late 1990s, when Japan started to experience near zero interest rates, low growth and slow deflation, monetary economists and central bank researchers have worried about how to conduct monetary policy with near-zero policy rates.

The zero-nominal interest rate floor arises because of the availability of cash as an asset that pays a zero nominal interest rate. Thus, savers need not accept less. In 1998/99 Federal Reserve economists investigated the role of a credible objective with a low but positive target rate for inflation in minimizing the risk of reaching the zero-interest-rate floor (see Orphanides/Wieland 1998) and expositied a wide range of policy tools that remain available once this constraint becomes binding (cf. Clouse et al. 2000 and Orphanides/Wieland 2000), including measures to influence longer-term premia and outright asset purchases with the aim of quantitative easing. The ensuing literature remained a niche for monetary policy experts for many years but yielded very useful insights for policy makers in 2008/2009.

When the European Central Bank conducted a mid-term review of its policy strategy in 2003, it concluded that “There are a number of well-grounded arguments for tolerating a low rate of inflation, and not aiming at zero inflation. The major concern is the need for a safety margin against potential risks of deflation. In a context of strong deflationary pressures, monetary policy may become less effective if central bank interest rate management is constrained by a liquidity trap, or a zero bound problem” (ECB 2003). Consequently, it affirmed that a quantitative definition of its price stability objective was a successful contribution to anchoring medium and longer-term inflation expectations and clarified that it meant to keep the rate of increase of the HICP below but close to two percent. The “close to” helped create a safety margin against potential risk of deflation, while the public commitment ensured that long-term inflation expectations remained positive and near two percent throughout the global financial crisis. Against this background, recent calls for giving up on past commitments to low and stable inflation in order to gain a much greater safety margin against deflation would appear to be misguided and possibly de-stabilizing (see Blanchard et al. 2010 for such a proposal and Issing 2011 for a rebuttal).

At the time of writing of this paper, the euro area remains mired in its own sovereign debt crisis. While euro area sovereigns’ finances have certainly been impacted by the banking rescues triggered by the global financial crisis, the seeds of the euro debt crisis also lie in the failure to enforce the fiscal rules that were meant to secure a stable monetary union. A thorough discussion of the appropriate policy responses is best reserved for a separate article. Suffice it to say that the concerns of the founders of European Monetary Union regarding the independence of the central bank and the need for a

sustainable fiscal policy have been thoroughly validated. In this context, the violation or revision of the Maastricht fiscal criteria on Franco-German initiative a few years prior to the start of the financial crisis proved disastrous.

## References

- Barro, R., D.B. Gordon (1983), Rules, Discretion and Reputation in a Model of Monetary Policy. *Journal of Monetary Economics* 12: 101-121.
- Batini, N., A. G. Haldane (1999), Forward-Looking Rules for Monetary Policy. Pp. 157-202 in: J.B. Taylor (ed.), *Monetary Policy Rules*. University of Chicago Press, Chicago.
- Bendixen, F. (1919), Vom theoretischen Metallismus. *Jahrbücher für Nationalökonomie und Statistik* 5: 497-534.
- Beck, G., V. Wieland (2008), Central Bank Misperceptions and the Role of Money in Interest Rate Rules. *Journal of Monetary Economics* 55(1): S1-S17.
- Beyer, A., V. Gaspar, Ch. Gerberding, O. Issing, (2008), Opting Out of the Great Inflation: German Monetary Policy after the Break Down of Bretton Woods. NBER Working Paper 14596, December.
- Bernanke, B., Th. Laubach, F. Mishkin, A. Posen (1999), *Inflation Targeting: Lessons from the International Experience*. Princeton University Press, Princeton, New Jersey, U.S.A.
- Bernanke, B. (2010), Monetary policy and the housing bubble. Speech at the American Economic Association Annual Meeting in Atlanta, U.S.A., January 3, 2010.
- Binder, M., V. Wieland (2006), The European Central Bank (2006). Pp. 54-58 in: S.N. Durlauf, L.E. Blume (eds.), *The New Palgrave Dictionary of Economics*. Second edition, Volume 3, Macmillan, London.
- Blanchard, O., G. dell'Arricia, P. Mauro (2010), Rethinking Macroeconomic Policy. IMF Staff Position Note 10/03.
- Bloomfield, A.I. (1959), *Monetary Policy under the International Gold Standard: 1880-1914*. Federal Reserve Bank of New York.
- Bodin, J. (1577), *Les six livres de la République*. Paris.
- Borchardt, K. (1972), Erfahrungen mit Inflationen in Deutschland. Pp. 9-22 in: J. Schlemmer (ed.), *Enteignung durch Inflation? Fragen der Geldwertstabilität*, München.
- Borchardt, K. (1976), *Währung und Wirtschaft*. Pp. 3-55 in: Deutsche Bundesbank (ed.), *Währung und Wirtschaft in Deutschland 1876-1976*, Frankfurt.
- Borchardt, K., O. Schötz (eds.) (1991), *Wirtschaftspolitik in der Krise*. Baden-Baden.
- Born, K.E. (1981), Geldtheorie und Geldpolitik. Pp. 360-374 in: *Handwörterbuch der Wirtschaftswissenschaft*. Dritter Band, Stuttgart.
- Bresciani-Turroni, C. (1937), *The Economics of Inflation. A Study of Currency Depreciation in Post War Germany*, London.
- Brunner, K., A.H. Meltzer (1989), *Monetary Economics*. Oxford.
- Bryant, R., P. Hooper, C. Mann (eds.) (1993), *Evaluating Policy Regimes: New Research in Empirical Macroeconomics*. The Brookings Institution, Washington, D.C.
- Buchheim, Ch. (1999), The Establishment of the Bank deutscher Länder and the West German Currency Reform. Pp. 55-100 in: Deutsche Bundesbank (ed.), *Fifty Years of the Deutsche Mark*. Oxford.
- Cagan, Ph. (1956), Studies in the Quantity Theory of Money. Pp. 25-117 in: M. Friedman (ed.), *The Monetary Dynamics of Hyperinflation*. Chicago.
- Calvo, G. (1983), Staggered Prices in a utility-maximising framework. *Journal of Monetary Economics* 12(3): 383-398.
- Christiano, L., M. Eichenbaum, Ch. Evans (2005), Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113(1): 1-45.
- Christiano, L., R. Motto, M. Rostagno (2008), Two Reasons Why Money and Credit May be Useful in Monetary Policy. Pp. 28-55 in: A. Beyer, L. Reichlin (eds.), *The Role of Money – Money and Monetary Policy in the Twenty-First Century*. European Central Bank, Frankfurt.



- Committee on the Working of the Monetary System (1959), Report. Her Majesty's Stationary Office, London
- Clarida, R., J. Gali, M. Gertler (1999), The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature* 37: 1661-1707.
- Clouse, J., D. Henderson, A. Orphanides, D. Small, P. Tinsley (2000), Monetary Policy when the Nominal Short-Term Interest Rate is Zero. Finance and Economics Discussion Series 2000-51. Board of Governors of the Federal Reserve System (U.S.).
- Cukierman, A., A. Meltzer (1986), A Theory of Ambiguity, Credibility and Inflation under Discretion and Asymmetric Information. *Econometrica* 54: 1099-1128.
- Cukierman, A. (1992), Central Bank Strategy, Credibility and Independence: Theory and Evidence. MIT Press, Cambridge, MA.
- Ellis, H.S. (1934), German Monetary Theory 1905-1933. Harvard.
- Emminger, O. (1976), Deutsche Geld- und Währungspolitik im Spannungsfeld zwischen innerem und äußerem Gleichgewicht (1948-1975). Pp. 485-554 in: Deutsche Bundesbank (ed.), *Währung und Wirtschaft in Deutschland 1876-1976*, Frankfurt.
- Eucken, W. (1923), *Kritische Betrachtungen zum deutschen Geldproblem*. Jena.
- European Central Bank (2003), Background Studies for the ECB's Evaluation of its Monetary Policy Strategy. European Central Bank, Frankfurt.
- Fischer, St. (1977), Long-term contracts, rational expectations and the optimal money supply rule. *Journal of Political Economy* 85(1): 191-205.
- Friedman, B.M., M. Woodford (eds.) (2011), *Handbook of Monetary Economics*. Vol. 3, Amsterdam.
- Friedman, M. (1956), The Quantity Theory of Money – A Restatement. Pp. 3-21 in: M. Friedman (ed.), *Studies in the Quantity Theory of Money*. Chicago.
- Friedman, M. (1960), *A Program for Monetary Stability*. New York.
- Friedman, M. (1968), The Role of Monetary Policy. *American Economic Review* 58(1): 1-10.
- Friedman, M., A.J. Schwartz (1963), *A Monetary History of the United States*. Princeton.
- Fuhrer, J., G. Moore (1995), Monetary Policy Trade-offs and the Correlation between Nominal Interest Rates and Real Output. *American Economic Review* 85(1): 219-239.
- Garfinkel, M., S. Oh (1993), Strategic Discipline in Monetary with Private Information: Optimal Targeting Horizons. *American Economic Review* 83(1): 99-117.
- Giannone, D., M. Lenza, L. Reichlin (2008). Explaining the great moderation: It is not the shocks. *Journal of the European Economic Association* 6(2-3): 621-633.
- Goodfriend, M., R. King (1997), The New Neoclassical Synthesis and the Role of Monetary Policy. NBER Macroeconomics Annual 231-282.
- Gurley, J.G., E.S. Shaw (1976), *Money in a Theory of Finance*. 8<sup>th</sup> ed., Washington.
- Haller, H. (1976), Die Rolle der Staatsfinanzen für den Inflationsprozess. Pp. 115-156 in: Deutsche Bundesbank (ed.), *Währung und Wirtschaft in Deutschland 1876-1976*. Frankfurt.
- Häuser, K. (1996), Knies als Geldtheoretiker. Pp. 31-51 in: K. Häuser, G. Eisermann, K. Yagi, K. (eds.), *Carl Knies' Das Geld: Vademecum zu einem deutschen Klassiker der Geldtheorie*. Verlag Wirtschaft & Finanzen, Düsseldorf.
- Häuser, K. (1998), Dogmengeschichtliche Betrachtungen zur Währungsunion von 1871-1876. Pp. 67-94 in: E.W. Streissler (ed.), *Studien zur Entwicklung der ökonomischen Theorie XVII: Die Umsetzung wirtschaftspolitischer Grundkonzeptionen in die kontinentaleuropäische Praxis des 19. und 20. Jahrhunderts. II. Teil*, Berlin.
- Hayek, F.A. (1931), *Preise und Produktion*. Wien.
- Hayek, F.A. (1933), Über neutrales Geld. *Zeitschrift für Nationalökonomie* 4: 659-661.
- Hayek, F.A. (1990), *Denationalisation of Money: The Argument Refined*. The Institute of Economic Affairs. 3<sup>rd</sup> edition, London.
- Helfferich, K. (1923), *Das Geld*. 6. Auflage, Leipzig.
- Holtfrerich, C.-L. (1986), *The German Inflation 1914-1923*. Berlin.
- Holtfrerich, C.-L. (1988), Zur Rezeption der Bullion und der Banking-Currency-School-Kontroverse in Deutschland. Pp. 9-27 in: H. Scherf (eds.), *Studien zur Entwicklung der ökonomischen Theorie VI*. Berlin.
- Issing, O. (1965), *Leitwährung und internationale Währungsordnung*. Berlin.



- Issing, O. (1992), Theoretical and Empirical Foundations of the Deutsche Bundesbank's Monetary Targeting. *Intereconomics* 27(6): 289-300.
- Issing, O. (1993), Central Bank Independence and Monetary Stability. Institute of Economic Affairs, Occasional Paper No. 89, London.
- Issing, O. (1996), Einführung in die Geldpolitik. 6. Auflage, München.
- Issing, O. (1997), Monetary Targeting in Germany: The Stability of Monetary Policy and of the Monetary System. *Journal of Monetary Economics* 39: 67-70.
- Issing, O., V. Gaspar, I. Angeloni, O. Tristani (2001), Monetary Policy in the Euro Area: Strategy and Decision-Making at the European Central Bank. Cambridge University Press, Cambridge, U.K..
- Issing, O. (2005), Why did the Great Inflation not happen in Germany? Federal Reserve Bank of St. Louis. Review, March/April.
- Issing, O. (2008), The Birth of the Euro. Cambridge University Press, Cambridge U.K.
- Issing, O. (2009), Asset Prices and Monetary Policy. *The Cato Journal*, Winter.
- Issing, O. (2011), Lessons for Monetary Policy: What Should the Consensus be? IMF Working Paper 11/97, April.
- James, H. (1999), The Reichsbank 1876-1945. Pp. 3-53 in: Deutsche Bundesbank (ed.), Fifty Years of the Deutsche Mark, Oxford.
- Jarocinski, M., F.R. Smets (2008), House prices and the stance of monetary policy. Pp. 339-366 in: Review, Federal Reserve Bank of St. Louis, issue July.
- Johnson, H. G. (1962), Monetary Theory and Policy. *American Economic Review*, June.
- Knapp, G.F. (1905), Staatliche Theorie des Geldes. München.
- Knies, C. (1873, 1876, 1879), Geld und Kredit. Berlin.
- Krohn, C.-D. (1986), Geldtheorien während der Inflation 1914-1924. Pp. 3-45 in: G.D. Feldman et al., Die Anpassung an die Inflation. Berlin.
- Krugman, P., M. Miller (eds.) (1992), Exchange Rate and Currency Bands. Cambridge University Press. Cambridge, U.K.
- Kydland, F.E., E.C. Prescott (1977), Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy* 85(3): 473-492.
- Kydland, F.E., E.C. Prescott (1982), Time to Build and Aggregate Fluctuations. *Econometrica* 50(6): 1345-1370.
- Laspeyres, E. (1871), Die Berechnung einer mittleren Warenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik* 16: 296-314.
- Laidler, D. (1981), Monetarism: An Interpretation and an Assessment. *Economic Journal* 91: 1-28.
- Lucas, R. E., Jr. (1972), Expectations and the Neutrality of Money. *Journal of Economic Theory* 4: 103-124.
- Lucas, R. E., Jr. (1976), Economic Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1): 19-46.
- Lucas, R. E., Jr. (2007), Central Banking: Is science replacing art? Pp. 168-171 in *Monetary Policy: A Journal From Theory to Practice*, ECB.
- Mankiw, G., D. Romer (eds.) (1991), *New Keynesian Economics*. Vol. 1 & 2, MIT Press, Cambridge, MA.
- McKinnon, R. (1993), The Rules of the Game: International Money in Historical Perspective. *Journal of Economic Literature* 31(1): 1-44.
- Meltzer, A. H. (2009), A History of the Federal Reserve. Volume 2, Book 2, 1970-1986, Chicago.
- Menger, C. (1970), Geld. Pp. 1-116 in: F.A. Hayek (ed.), *Gesammelte Werke*. Band IV, Tübingen.
- Möller, H. (1961), Zur Vorgeschichte der Deutschen Mark. Basel/Tübingen.
- Möller, H. (1976), Die westdeutsche Währungsreform von 1948. Pp. 433-483 in: Deutsche Bundesbank (ed.), *Währung und Wirtschaft in Deutschland 1876-1976*. Frankfurt.
- Mundell, R. A. (1973), Uncommon Arguments for Common Currencies. Pp. 114-132 in: H. G. Johnson, A. K. Swoboda (eds.), *The Economics of Common Currencies*. Harvard University Press, Cambridge.
- Neumann, M. J. M. (1971), Zwischenziele und Indikatoren der Geldpolitik. *Kredit und Kapital* 4: 398-420.

- Neumann, M.J.M. (1999), Monetary Stability: Threat and Proven Response. Pp. 269-306 in: Deutsche Bundesbank (ed.), Fifty Years of the Deutsche Mark. Frankfurt.
- Orphanides, A. (2002), Monetary Policy Rules and the Great Inflation. *American Economic Review* 84(4): 569-583
- Orphanides, A., V. Wieland (1998), Price Stability and Monetary Policy Effectiveness when Nominal Interest Rates are Bounded at Zero. Finance and Economics Discussion Series, 1998-35 (later published with Gunter Coenen as co-author in *Advances in Macroeconomics* 4(1): January 2004.).
- Orphanides, A., V. Wieland (2000), Efficient Monetary Policy Design Near Price Stability. *Journal of the Japanese and International Economies* 14: 327-365.
- Orphanides, A., V. Wieland (2008), Economic Projections and Rules-of-Thumb for Monetary Policy. *Federal Reserve Bank of St. Louis Review* 90(4): 307-324.
- Patinkin, D. (1965), Money, Interest, and Prices. 2<sup>nd</sup> ed., New York.
- Pesek, B.P., Th.R. Saving (1969), Money, Wealth and Economic Theory. 3<sup>rd</sup> ed., New York.
- Phelps, E. (1967), Phillips Curves, Expectations of Inflation and Optimal Unemployment over Time. *Economica* 34: 254-81.
- Phelps, E., J.B. Taylor (1977), Stabilizing powers of monetary policy with rational expectations. *Journal of Political Economy* 85: 163-190.
- Reifschneider, D., R. Tetlow, J. Williams (1999), Aggregate Disturbances, Monetary Policy and the Macroeconomy: The FRB/US Perspective. *Federal Reserve Bulletin* 85(1): 1-19.
- Richter, R. (1999a), German Monetary Policy as Reflected in the Academic Debate. Pp. 525-571 in: Deutsche Bundesbank (ed.), Fifty Years of the Deutsche Mark. Frankfurt.
- Richter, R. (1999b), Deutsche Geldpolitik 1948-1998. Tübingen.
- Rieter, H. (1971), Die gegenwärtige Inflationstheorie und ihre Ansätze im Werk von Thomas Tooke. Berlin.
- Rieter, H. (2002), Deutsche Geldtheorie im 19. Jahrhundert – mehr als nur ein Echo englischer Debatten? Pp. 13-55 in: B. Schefold (ed.), Exogeneity and Endogeneity. Marburg 2002.
- Rist, Ch. (1938), Histoires des doctrines relatives au crédit et à la monnaie. Paris.
- Rogoff, K. (1985), The Optimal Commitment to an Immediate Monetary Target. *Journal of Economics* 100: 1169-1189
- Rotemberg, J., M. Woodford (1997), An optimization-based econometric framework for the evaluation of monetary policy. NBER Macroeconomics Annual 12.
- Samuelson, P.A., R.M. Solow (1960), Analytical Aspects of Anti-Inflation Policy. *American Economic Review* 50(2): 177-194.
- Sargent, Th., N. Wallace (1975), Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule. *Journal of Political Economy* 83: 241-254.
- Schlesinger, H. (1976), Geldpolitik in der Phase des Wiederaufbaus (1950-1958). Pp. 555-607 in: Deutsche Bundesbank (ed.), Währung und Wirtschaft in Deutschland 1876-1976. Frankfurt.
- Schlesinger, H. (1983), The Setting of Monetary Objectives in Germany. Pp. 6-17 in: P. Meek (ed.), Central Bank Views of Monetary Targeting. Federal Reserve Bank of New York.
- Schumpeter, J.A. (1917), Das Sozialprodukt und die Rechenpfennige. *Archiv für Sozialwissenschaft und Sozialpolitik* 44: 627-715.
- Smets, F., R. Wouters (2003), An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association* 1: 1123-1175.
- Stein, H. (1990), Presidential Economics 2<sup>nd</sup> ed., Washington.
- Svensson, L.E.O. (1997), Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets. *European Economic Review* 41(6): 1111-1146.
- Svensson, L.E.O. (2010), Inflation Targeting. Pp. 1237-1308 in: B.M. Friedman, M. Woodford (eds.), *Handbook of Monetary Economics*. Volume 3b, Elsevier.
- Taylor, J.B. (1975), Monetary Policy During a Transition to Rational Expectations. *Journal of Political Economy* 84(5): 1009-1022.
- Taylor, J.B. (1979a), Staggered wage setting in a macro model. *American Economic Review* 69(2): 108-113.
- Taylor J.B. (1979b), Estimation and Control of a Macroeconomic Model with Rational Expectations. *Econometrica* 47(5) 1267-1286.

- Taylor, J. B. (1993a), *Macroeconomic Policy in a World Economy*. W. W. Norton, New York.
- Taylor, J. B. (1993b), Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy* 39: 195-214.
- Taylor, J. B. (2007), Housing and Monetary Policy. Pp. 463-476 in: *Housing, Housing Finance, and Monetary Policy*. Proceedings of FRB of Kansas City Symposium, Jackson Hole, WY.
- Taylor, J. B. (2011), The Cycle of Rules and Discretion in Economic Policy. *National Affairs* (7): 55-65.
- Taylor, J. B., V. Wieland (2012), Surprising Comparative Properties of Monetary Models: Results from a New Monetary Model Base. *Review of Economics and Statistics* 94 (3): 800-816.
- Tobin, J. (1958), Liquidity Preferences as Behavior towards Risk. *Review of Economic Studies* 25: 65-86.
- Tobin, J. (1965), Money and Economic Growth. *Econometrica* 33(4): 671-684.
- Tobin, J. (1980), *Asset accumulation and economic activity*. Oxford.
- Trichet, J.-C. (2008), Interview with *Le Figaro* and *Frankfurter Allgemeine Zeitung*. 11. July 2008.
- von Mises, L. (1912), *Theorie des Geldes und der Umlaufsmittel*. München – Leipzig.
- von Mises, L. (1980), *The Theory of Money and Credit*, Indianapolis. (Reprint of the edition 1934, translation of the edition of his “*Theorie des Geldes und der Umlaufsmittel*”, München 1924).
- Wagner, A. (1873), *System der Zettelbankpolitik, mit besonderer Rücksicht auf das geltende Recht und auf deutsche Verhältnisse*. Ein Handbuch des Zettelbankwesens, 2. Auflage, Freiburg.
- Wieser, F. (1927), *Geld: Theorie des Geldes*. Allgemeine Theorie des Geldes, Handwörterbuch der Staatswissenschaften. 4. Auflage, Jena.
- Wieland, V. (1996), Monetary Policy Targets and the Stabilization Objective: A Source of Tension in the EMS. *Journal of International Money and Finance* 15(1): 95-116.
- Wieland, V. (2000), Monetary policy, parameter uncertainty and optimal learning. *Journal of Monetary Economics* 46: 199-228.
- Wieland, V. (ed.) (2009), *The Science and Practice of Monetary Policy Today*. Springer Science, Berlin.
- Wicksell, K. (1897), *Der Bankzins als Regulator der Warenpreise*. *Jahrbücher für Nationalökonomie und Statistik*.
- Wicksell, K. (1898), *Geldzins und Güterpreise*. Jena.
- Wicksell, K. (1906), *Lectures on Political Economy*. Vol. 2, Money, London.
- Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, New Jersey, U. S. A.
- Woodford, M. (2007), The Case for Forecast Targeting as a Monetary Policy Strategy. *Journal of Economic Perspectives* 21: 3-27.
- Woodford, M. (2008), How Important is Money in the Conduct of Monetary Policy? *Journal of Money, Credit and Banking* 40: 1561-1598.

Prof. Dr. Otmar Issing, Center for Financial Studies, Goethe University of Frankfurt, Gruenburgplatz 1, 60323 Frankfurt a.M., Germany.  
 issing@ifk-cfs.de

Prof. Volker Wieland, Ph.D., Institute for Monetary and Financial Stability, Goethe University of Frankfurt, Gruenburgplatz 1, 60323 Frankfurt a.M., Germany  
 wieland@wiwi.uni-frankfurt.de; <http://www.volkerwieland.com>

## Buchbesprechung / Book Review

**Holub, Hans-Werner**, Eine Einführung in die Geschichte des Ökonomischen Denkens. Band V: Die Ökonomik des 20. Jahrhunderts, Teil 2: Englische und amerikanische Ökonomen. Reihe: Einführungen Wirtschaft, Band 17. Wien (Lit Verlag) 2012, 256 S., € 29,90, ISBN 978-3-643-50423-4.

Das ist der siebte und letzte Band einer groß angelegten Dogmengeschichte. Behandelt werden acht Autoren jeweils nach einem weitgehend gleichen Schema, und es werden abschließend noch einmal einige neuere Schulen bzw. Richtungen in der Ökonomie zusammenfassend dargestellt, wie z. B. die post-keynesianische und neue keynesianische Makroökonomik und die „neue klassische“ Makroökonomik (NKM). Die Abschnitte über die einzelnen Autoren beginnen immer mit einem recht detaillierten Lebenslauf, der nicht selten auch sehr ausführliche und lesenswerte Hinweise zu bedeutenden Ereignissen und Aktivitäten im Leben der Autoren enthält und der sich stets hinsichtlich des Layouts deutlich vom übrigen Buch unterscheidet (Rahmen und grauer Hintergrund). Es folgen dann jeweils Hinweise zu Publikationen der Autoren, zu ihren wichtigsten Thesen sowie zur „Rezeption“, also zu ihrer Wirkung in Wissenschaft und Politik und zur Kritik ihrer Thesen in Fachkreisen. Der Aufbau des Buches ist also primär an Personen, nicht an Gegenständen orientiert.

Die acht z. T. sehr ausführlich behandelten Ökonomen sind Irving Fisher, Arthur Cecil Pigou, John Maynard Keynes, John Richard Hicks, Joan Violet Robinson, Piero Sraffa, Paul Anthony Samuelson und Milton Friedman. Etwas verkürzt wird am Ende des Bandes auch noch Robert E. Lucas behandelt.

Es liegt in der Natur der Sache, dass es schwierig ist, eine Auswahl zu treffen ist, und dass man natürlich endlos darüber streiten kann, warum es gerade der Ökonom oder die Ökonomin (wie im Falle von Joan Robinson) verdient hat, in einem längeren Abschnitt gewürdigt zu werden, und warum andererseits diesem oder dieser eine solche Ehre verweigert wurde. Entscheidungen dieser Art sind sicher immer etwas subjektiv und auch nicht wenig von den eigenen Forschungsinteressen oder auch Sympathien zu entsprechenden Thesen beeinflusst. Längere Diskussionen darüber, wem wie viel Ehre gebührt und wie unterschiedlich prominente Ökonomen hinsichtlich Ruhm und Verdienst sind, dürften nicht sehr fruchtbar sein. Ich könnte mir zwar vorstellen, dass die Namen Piero Sraffa und Joan Robinson kommenden Generationen von Ökonomen immer weniger sagen werden und demgegenüber vielleicht andere Namen wie Harrod, Tobin oder Sen dauerhafter von Bedeutung sein könnten. Aber sei's drum. Die Auswahl von Ökonomen, die in einem solchen Band behandelt werden sollten ist quasi „naturgemäß“ nie unproblematisch und man kann wohl schon deshalb H. W. Holub kaum einen Vorwurf machen, was seine in diesem Punkte getroffenen Entscheidungen betrifft.

In diesem Zusammenhang ist es auch sehr zu loben, dass Holub es nicht selten explizit anmerkt, wenn eine gerade gemachte Aussage ein Werturteil seinerseits darstellt. Der betreffenden Aussage selbst, bei der es meist um Methodologisches oder wirtschaftspolitische Erfahrungen mit entsprechenden Thesen der behandelten Ökonomen geht, kann fast immer zugestimmt werden. Die gelegentlich erkennbaren Ansichten Holubs über die Rolle der Mathematik in der Wirtschaftstheorie, über Lenkung vs. Markt oder über sozialpolitische Implikationen scheinen mir durchwegs moderate und vernünftige Positionen zu sein.

Auf zwei ebenfalls in der Natur der Sache liegende Probleme bei einer so breit angelegten und an Personen entwickelten Dogmengeschichte soll im Folgenden eingegangen werden.

Die erste, nicht ganz einfache Herausforderung besteht darin, einerseits die Positionen der ausgewählten Ökonomen klar und ausführlich genug zu präsentieren, aber andererseits dabei auch nicht der Versuchung zu erliegen, ein weiteres, diesmal etwas historisch orientiertes Lehrbuch der Volkswirtschaftslehre zu schreiben. Und ein zweites Problem ist, dass man sich heutzutage bei so einem Vorhaben fragen muss, ob ein entsprechend an Wirtschaftstheorie interessierter potentieller Leser nicht genauso weit kommt, wenn er einfach die entsprechenden Namen „googelt“ und dann die jeweiligen Wikipedia Artikel „downloadet“.

Was die Annäherung des Textes an ein Buch mit Lehrbuchcharakter betrifft, so gibt es Passagen von deutlich unterschiedlicher Qualität. Was m. E. recht gut herausgearbeitet wurde ist – um nur ein paar Beispiele zu nennen – der „Pigou-Effekt“ und die Kritik an seinen Voraussetzungen und seiner (mangelnden) empirischen Relevanz (S. 41 f.), oder auch die Bestimmungsfaktoren der Geldnachfrage nach Friedman (S. 212 f.) und überhaupt die Grundüberzeugungen des Monetarismus (S. 218 ff.) in Gestalt einer langen, wenngleich nicht sehr systematischen und auch etwas redundanten Auflistung. Positiv fällt auch auf, dass z. B. über Samuelsons Verdienste in der Ökonomie deutlich mehr gesagt wird als in manchen anderen Schriften, bei denen sein berühmtes Lehrbuch alles dominiert. Andererseits ist kritisch zu vermerken, dass es – um auch hier wieder nur einige Beispiele zu nennen – sehr nebulös bleibt, was mit dem Aggregationsproblem im Zusammenhang mit der mikroökonomischen Fundierung der Makroökonomie (S. 52 f.) oder mit Samuelsons „Korrespondenzprinzip“ (S. 190f) gemeint ist und z. B. auch was die Kapitaltheorie von Hicks eigentlich beinhaltet (S. 132). Obgleich an zwei Stellen behandelt (S. 162 und 175) wird auch die Cambridge-Cambridge-Kontroverse oder das „reswitching“ nicht wirklich klar. Auf die Gründe für die oft geringe Tauglichkeit des Textes im Sinne eines Lehrbuchs (also als eine einführende Information) kommen wir noch zurück.

Was hat man mehr, als wenn man nur nach den Namen „googelt“? Es mag sonderbar erscheinen und ist sicher auch etwas subjektiv, aber für mich waren die Hinweise zur Persönlichkeit, zum politischen Wirken oder auch zu anderen nicht im engeren Sinne wirtschaftstheoretischen Überlegungen der genannten Ökonomen oft ganz besonders interessant und wertvoll. Sehr lesenswert dürfte vor allem – nicht nur für Historiker, sondern gerade auch für Ökonomen in unseren Tagen – die ausführliche Darstellung der Kritik von Keynes am Versailler Vertrag sein (S. 69 ff.). Ähnlich wertvoll sind z. B. auch Passagen über die Hintergründe bei der Konferenz von Bretton Woods 1944.

Die zahlreichen interessanten Hinweise dieser Art, die auch weit über die wirtschaftstheoretischen Lehren der besprochenen Ökonomen hinausgehen zeigen deutlich die große Belesenheit des Autors Holub. Diese Belesenheit scheint aber auch eine Schattenseite zu haben. Denn sehr auffallend und kennzeichnend, aber leider auch ziemlich negativ für das Werk ist, dass es gut und gerne zu einem Drittel aus wörtlichen Zitaten besteht und dass Holub offenbar eine Scheu hat, etwas in eigenen Worten auszudrücken. Oft liest man in einem Zitat einer deutschsprachigen Veröffentlichung noch einmal ziemlich genau das, oder zumindest etwas sehr Ähnliches wie das, was man kurz zuvor schon in einem englischen Zitat gelesen hat. Die Aneinanderreihung von Zitaten verhindert auch oft einen sinnvollen zusammenhängenden Gedankengang. Nicht selten enthalten die Zitate auch sehr allgemein gehaltene und umfassende Würdigungen eines Autors oder einer ganzen Theorie, was für einen Leser meist nicht sehr hilfreich ist. Was nützt einem die Vogelperspektive, wenn man sich am Boden mit den elementaren Aussagen einer Theorie beschäftigt?

Damit hängt auch zusammen, dass – wie oben bereits angedeutet – manche Stellen ziemlich unverständlich sind, zumindest für solche Leser, die bei dem jeweils angesprochenen Gegenstand das nötige Lehrbuchwissen nicht parat haben. Sehr oft werden Begriffe eingeführt, die für die Theorien des jeweils besprochenen Autors als fundamental herausgestellt werden, die aber gleichzeitig dem Leser nicht richtig klar erklärt werden. Denn bevor Holub den betreffenden Begriff in eigenen Worten klar definiert präsentiert er bereits eine Reihe von Zitaten, in denen diese Begriffe benutzt werden. Ein typisches Beispiel hierfür ist S. 134. Die ganz Seite dreht sich um die „Traversen“ von Hicks. Man liest in fünf Zitaten über eine erste und zweite Traversenanalyse von Hicks und ob sein steady state model dynamisch oder doch nur statisch ist, aber von „Traversen“ selbst wird zu Beginn nur gesagt „das sind die Wege von einem Wachstumsgleichgewicht (steady state) zum nächsten“, was deutlich zu wenig sein dürfte um z. B. zu verstehen, warum die erste Traversenanalyse nach Aussage von Zamagni 1997 „gescheitert“ sein soll und warum laut Bliss 1997, die Traversenanalyse „ein kluges Konstrukt“ sein soll. Und im Endeffekt, wenn man alles gelesen hat, hat einem die ganze Seite wenig oder nichts gebracht.

Hinsichtlich der zahlreichen Zitate ist leider auch noch zu sagen, dass es mir nicht gelungen ist, die konkreteren Angaben (Fundstellen) zu finden, die genau gemeint sind bei Schefold 1991d, Schefold 1997e, Raab 1983, Cansier 1989, Linß 2007, Rohwetter 1999, Keynes (deutsch) 2004,

Zank 1992 und den vielen (sehr vielen) anderen derartigen Literaturangaben. Ich vermute, man findet die entsprechenden Angaben im ersten Band dieses mehrbändigen Werks. Aber ich weiß es nicht, und es konnte mir auch keiner sagen, wo man sie findet.

Obgleich ich sehr positiv denke über viele Darstellungen in dem Buch (einschließlich über die darin durchschimmernden Positionen und Meinungen von Holub) und auch über das Projekt einer derartig breit angelegten Dogmengeschichte muss ich leider noch zwei kritische Anmerkungen machen.

Zum einen hat der Autor nicht selten für eine wenig relevante Mitteilung ungewöhnlich viel Platz reserviert und zum anderen vermisst man gelegentlich etwas Sorgfalt beim Editieren. An vielen Stellen wird beispielsweise eine ausführliche Kapitelgliederung eines Buches oder Aufsatzes präsentiert (etwa auf S. 72, 79, 102, 125, 145, 189) und man erfährt z.B. in welche fünf Kapitel sich ein Aufsatz von Sraffa aus dem Jahre 1925 gliedert (S. 172), oder dass Joan Robinson mindestens sieben „edelmetallene Zeitalter“, wie u.a. das hinkende goldene, das bleierne, oder das galoppierende Platin-Zeitalter usw. unterschied (S. 162), aber zu Fisher's Making of Index Numbers liest man gerade mal zehn Zeilen, in denen die „Tests“ und der „Idealindex“ nur kurz (dem Namen nach) erwähnt werden (was sehr wenig ist, auch im Verhältnis z. B. zum unbedeutenden Treatise on Probability von Keynes, der auf S. 67 und 85 behandelt wird).

Was das Editieren betrifft, so hätte auffallen müssen, dass der die NKM betreffende Text auf den Seiten 112 f. und 233-235 zu einem großen Teil sogar Wort für Wort identisch ist. Nicht sehr vertrauenswürdig ist es auch, wenn ein und die gleiche Gesellschaft „Kardex Rent Company“ und ein paar Zeilen weiter „Kardex Rand Company“ genannt wird (S. 12). Das offensichtlich als bekannt vorausgesetzte IS-LM Modell erscheint auch in folgender Schreibweise: I/S-L/M Modell (oder Diagramm, Schema, System) oder I/S/LM-Modell (auf S. 100 -102 findet man alle drei Schreibweisen) und es wird auch als „Festpreismodell“ bezeichnet. Etwas irritierend ist es auch, wenn eine Graphik als „Keynesianisches Kreuz“ bezeichnet wird und nur in Worten beschrieben wird (S. 99), wo doch eine entsprechende Abbildung sehr viel mehr Klarheit geschafft hätte. An dieser Stelle war übrigens auch die m.W. einzige vollständige Literaturangabe (zumindest die einzige, die ich entdeckt habe), nämlich: Schelling Th., Income Determination: A Graphical Solution, in Review of Economics and Statistics 1948 (30), S. 227 – 229.

Es werden, wie gesagt, sowohl deutsche als auch englische Texte im Original zitiert. Dabei ist z. T. die zitierte deutsche Übersetzung einer im Original in englischer Sprache verfassten Schrift erkennbar unglücklich. Die General Theory von Keynes wird in deutscher Übersetzung zitiert als Keynes 1936 und es heißt dort über die marginale Konsumquote, dass sie positiv und „weniger als die Einheit“ ist (S. 82). Offenbar war dem Übersetzer nicht klar, dass mit „unity“ auch die Zahl 1 gemeint sein kann und dass es damit eigentlich „positiv und kleiner als eins“ heißen müsste. In der gleichen Übersetzung ist offenbar wiederholt von einer „nachgiebigen Lohnpolitik“ die Rede (S. 92). Man könnte das mit Nachgiebigkeit gegenüber den Lohnforderungen von Gewerkschaften assoziieren. In dem hier zitierten Kapitel 19 der General Theory heißt es aber „reduction in money-wages“ und *in diesem Sinne* am Ende des Kapitels auch wiederholt „flexible wage- policy“; es geht also um die Kritik von Keynes an einer Politik, die durch Lohnsenkungen wieder Vollbeschäftigung herstellen möchte. Dass hier mit der deutschen Übersetzung etwas nicht ganz stimmen kann, hätte einem auffallen müssen.