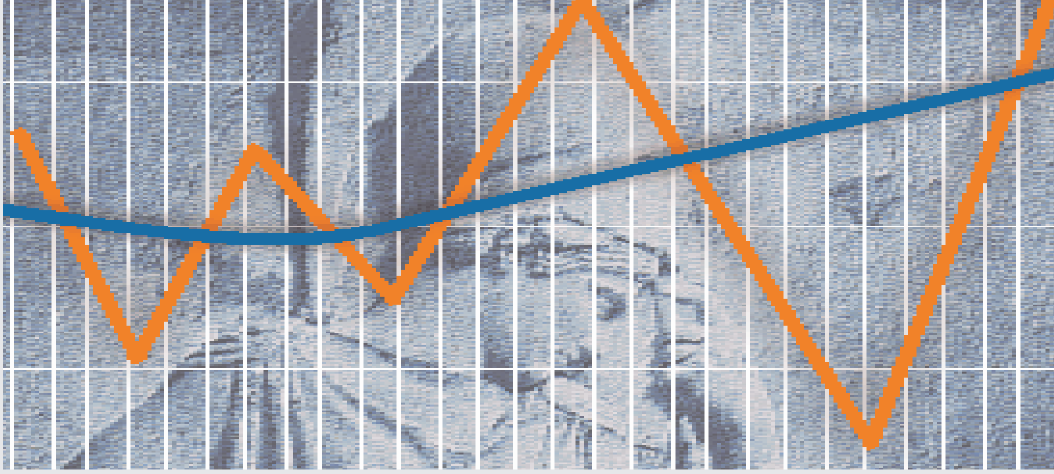


ifo Beiträge
zur Wirtschaftsforschung

65

Economic Growth and Business Cycle Forecasting at the Regional Level

Robert Lehmann

ifo Institut
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

Herausgeber der Reihe: Hans-Werner Sinn
Schriftleitung: Chang Woon Nam

ifo Beiträge zur Wirtschaftsforschung

65

Economic Growth and Business Cycle Forecasting at the Regional Level

Robert Lehmann

ifo Institut
Leibniz-Institut für Wirtschaftsforschung
an der Universität München e.V.

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-13: 978-3-95942-007-5

Druck: ifo Institut, München

ifo Institut im Internet:
http://www.cesifo-group.de

Vorwort

Die vorliegende Dissertationsschrift wurde von Robert Lehmann während seiner Tätigkeit an der Dresdner Niederlassung des ifo Instituts verfasst. Die Arbeit wurde im Mai 2015 abgeschlossen und im Oktober 2015 als Dissertationsschrift von der Wirtschaftswissenschaftlichen Fakultät der Technischen Universität Dresden akzeptiert.

Die Dissertationsschrift öffnet mit einer Einleitung und thematischen Einordnung der Arbeit (Kapitel 1). Im Fokus von Kapitel 2 steht die Frage, welche Rolle unterschiedliche Wirtschaftsstrukturen für das regionale Wirtschaftswachstum spielen. In der ökonomischen Literatur werden zwei konkurrierende Konzepte diskutiert: Spezialisierung und Diversifikation. Im Kontrast zur existierenden Literatur werden in dieser Arbeit die beiden Konzepte miteinander interagiert. Den Untersuchungsgegenstand bilden die 70 größten deutschen Städte. Es zeigen sich negative Interaktionseffekte im Verarbeitenden Gewerbe, Baugewerbe und für die höherwertigen Dienstleistungen.

Bei Kapitel 3 handelt es sich um eine systematische Studie der Prognosegüte einer Vielzahl regionaler, nationaler und internationaler Indikatoren für die Vorhersage des Bruttoinlandsprodukts (BIP) zweier deutscher Bundesländer sowie Ostdeutschlands. Besonders Prognose-Kombinationsstrategien und regionale Indikatoren, welche die regionale Wirtschaftsstruktur adäquat abbilden, zeigen die höchste Treffsicherheit.

Kapitel 4 geht einen Schritt weiter und stellt die Frage, ob sektoral disaggregierte Vorhersagen der Bruttowertschöpfung (BWS) das Prognoseergebnis der gesamten BWS verbessern. Besonders in der kurzen Frist (ein Quartal in die Zukunft) kann die Prognose für die gesamte BWS durch disaggregierte Vorhersagen deutlich verbessert werden.

Kapitel 5 fokussiert auf eine einzelne Komponente des BIP: die Exporte. Für eine Vielzahl europäischer Staaten wird der Frage nachgegangen, ob Befragungsdaten oder harte Fakten wie die preisliche Wettbewerbsfähigkeit die Exportprognosen verbessern können. Dabei zeigt sich, dass die Befragungsdaten klar im Vorteil sind. Jedoch weisen die Länder eine starke Heterogenität bei ihrer Prognosegüte auf. Diese Unterschiede können zu einem erheblichen Teil durch die Exportzusammensetzungen der einzelnen Staaten erklärt werden. Das letzte Kapitel fasst die Arbeit nochmals zusammen.

Stichworte: Spezialisierung, Diversifikation, Interaktionsmodelle, Regionale Bruttowertschöpfung, Regionale Konjunkturprognosen, Prognosekombination, Faktormodelle, Exportprognosen

JEL-Nr: C53, E37, F17, O18, R11.

Danksagung

Die vorliegende Dissertationsschrift ist in meiner mehr als fünfjährigen Tätigkeit als Doktorand am ifo Institut, Niederlassung Dresden entstanden. Während dieser Zeit haben meinen Weg eine Vielzahl von Menschen begleitet, ohne die die vorliegende Schrift in dieser Form nicht hätte entstehen können. Daher möchte ich die folgenden Zeilen nutzen, um diesen Menschen meinen besonderen Dank auszusprechen.

Ganz herzlicher Dank gebührt meinem Doktorvater Prof. Dr. Marcel Thum für seine hervorragende Betreuung dieser Dissertation. Seine anhaltende Unterstützung, Ermutigung und seine umfassende ökonomische Expertise waren wesentliche Bausteine, die zur Vollendung dieser Arbeit beigetragen haben. Zudem bewundere ich die von Prof. Dr. Marcel Thum gelebte fachliche Leidenschaft und seine Menschlichkeit, welche wesentliche Gründe für meine eigene Motivation waren.

Außerdem bedanke ich mich bei Prof. Dr. Joachim Ragnitz, von dem ich einen Großteil für die Bearbeitung von Drittmittelgutachten gelernt habe. Zudem verdanke ich ihm meinen aktuellen Wissensstand über die wirtschaftliche Entwicklung der ostdeutschen Bundesländer, welcher in diversen Publikationen von großem Nutzen für mich war.

Natürlich danke ich auch meinen Kolleginnen und Kollegen der Dresdner Niederlassung des ifo Instituts für zahlreiche Diskussionen und die angenehme Arbeitsatmosphäre. Ich danke ganz herzlich Michael Weber für die produktive Zusammenarbeit im Bereich Konjunktur und Wachstum. Besonderer Dank gebührt auch Wolfgang Nagl, welchen ich nunmehr zu einem wichtigen Freund zähle.

Ich danke auch ganz herzlich meinen zahlreichen Koautoren. Eine Vielzahl von Veröffentlichungen sind in fruchtbarer Koautorenschaft entstanden. Besonders bedanke ich mich bei Klaus Wohlrabe und Jan Kluge für die professionelle und produktive Zusammenarbeit. Die Ergebnisse aus diesen erfolgreichen Kooperationen spiegeln sich in den Kapiteln 2, 3 und 4 wider.

Nicht zuletzt gebührt großer Dank meinen Freunden und meiner Familie. Die immerwährenden Ermutigungen von meinen guten Freunden André Rolle, Peter Friebel und Roy Petran haben mir stets sehr geholfen. Großer Dank gebührt auch meinen Onkeln und Tanten, die den Verlauf dieser Arbeit immer mit einem wachen Ohr verfolgten. Besondere Kraft aber habe ich aus der Liebe meiner Eltern Karin und Martin gezogen. Ihr großes Interesse am Fortschritt dieser Arbeit gab mir das nötige Durchhaltevermögen, um diesen langen Weg bis zu Ende zu gehen.

Economic Growth and Business Cycle Forecasting at the Regional Level

Dissertationsschrift

zur Erlangung des akademischen Grades

Doctor rerum politicarum (Dr. rer. pol.)

der Wirtschaftswissenschaftlichen Fakultät

der Technischen Universität Dresden

2015

vorgelegt von

Robert Lehmann

Erster Gutachter: Prof. Dr. Marcel Thum

Zweiter Gutachter: Prof. Dr. Georg Hirte

Disputation: 11. Dezember 2015

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2SLS	two-stage least-squares
ADL	Autoregressive distributed lag
AG	Incorporated company (in German: Aktiengesellschaft)
AGFI	Agriculture, hunting and forestry; fishing
ARIMA	Autoregressive integrated moving average
AS	Advanced services
BBR	Federal Office for Building and Regional Planning
BIC	Bayesian or Schwarz Information Criterion
BS	Basic services
BW	Baden-Württemberg
CCOF	Consumer confidence indicator
CD	Cragg-Donald Wald F statistic
COF	Confidence indicator in manufacturing
CON	Construction
c.p.	ceteris paribus
DB	Central Bank of Germany (in German: Deutsche Bundesbank)
Div	Diversification
EA	Euro area
EC	European Commission

ECB European Central Bank

ECFIN Department of Economic and Financial Affairs

EG Eastern Germany

e.g. exempli gratia

EOBL Assessment of export order-book levels

Eq. Equation

ESA 1995 European System of Accounts of 1995

ESI Economic sentiment indicator

et al. et alii

etc. et cetera

EU European Union

e.V. Registered association (in German: eingetragener Verein)

EXEXP Export expectations

EXPI Price deflator for exports of goods and services

FE Fixed-effects

Fig. Figure

GDP Gross domestic product

GDPDEF GDP deflator

GfK Society for Consumer Research

GVA Gross value added

GVAR Global vector autoregression

HAC Heteroscedasticity-autocorrelation-consistent

HCPI Harmonized consumer price index

HHI	Hirschman-Herfindahl-Index
i.e.	id est
IfW	Kiel Institute for the World Economy
IND	Industry
IP	Industrial production
ISM	In-sample mean
IV	Instrumental variable
IWH	Halle Institute for Economic Research
MAR	Marshall-Arrow-Romer
max	Maximum
MCS	Model confidence set
MDM	Modified Diebold-Mariano test
med	Median
min	Minimum
MSFE	Mean squared forecast error
NUTS	Nomenclature des unités territoriales statistiques
OBL	Assessment of order-book levels
Obs.	Observations
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
PC	Principal components
PEXP	Production expectations
PMI	Purchasing Manager Index

PPS	Public and private services
QML	Quasi-maximum likelihood
qoq	quarter-on-quarter
REER	Real effective exchange rate
RMSFE	Root mean squared forecast error
RW	Random Walk
s.d.	Standard deviation
Sect.	Section
SFP	Assessment of stocks of finished products
SITC	Standard International Trade Classification
SPA	Superior predictive ability
Spec	Specialization
SX	Free State of Saxony
TFP	Total factor productivity
ULCTOT	Unit labor costs of the total economy
US	United States
UWCMAN	Nominal unit wage costs in manufacturing
VAR	Vector autoregression
VGRdL	National accounts for the German states
vs.	versus
yoy	year-on-year
ZEW	Centre for European Economic Research

1 Introduction

Many politicians from the national and sub-national level aim to ensure a sustainable and stable economic growth path in the mid- and long-term. What we observe are positive, but non-constant growth rates of total output for most economies in the world. Phases of increasing growth rates are followed by economic downturns. However, politicians do not prefer fluctuations in economic output and try to minimize these deviations from long-term growth with appropriate policy instruments. To achieve sustainable economic growth, macroeconomic research might help to pursue this goal and to find the determinants for a long-term economic growth path. Economic research can also concentrate on improving forecast mid-term business fluctuations so that politicians can identify effective instruments at an early stage. The instruments politicians base their decisions on vary strongly between mid- and long-term economic growth. Therefore, total output is usually decomposed into four different phenomena that overlap each other: a growth component, a business cycle component, a seasonal component and a stochastic component. The seasonal component captures short-term fluctuations of total output within a year. Stochastic components capture unsystematic interruptions due to events like strikes. The business cycle component represents mid-term fluctuations in total output, for example, through changing capital investments of firms. Finally, the growth component captures the long-term growth path of an economy, typically determined by the availability of its production factors. From these four components, politicians can actively influence the business cycle and growth component. Sustainable economic growth in the long-term can be achieved with a more structural policy such as the safeguarding of qualified employees, the strengthening of immigration, infrastructural improvements or the promotion of technological progress. To prevent the economy from drifting, for example, into a recession in the mid-term, politicians should base their decisions on policy instruments such as fiscal policy instead.

From an academic perspective a huge pool of knowledge exists for countries. Sub-national entities such as states or counties are, however, studied to a lesser extent. Since Paul R. Krugman's influential studies of the new economic geography, there is a renewed interest in studying regional relationships. The determinants of regional economic growth and spatial dependencies in particular are major research questions. But economists are not the only ones to focus on regional entities, political institutions at every governmental level also deal with regional economic problems. One prominent example is the Cohesion Policy of the European Union (EU), which, among other things, aims to reduce regional inequalities between regions. Most of these policies tend to focus on long-term regional developments. However, mid-term business fluctuations also play a crucial role for regional entities. In particular, regional economic structures cause different growth patterns and therefore business cycles of regional economies. An appropriate tool to reduce uncertainty for mid-term business cycle fluctuations are macroeconomic forecasts. Unbiased macroeconomic forecasts at the regional level are the basis for, for example, fiscal policy planning or the allocation of public expenditure to reduce existing labor market disparities. Regional macroeconomic forecasts can also be used to detect deviating developments of regional units from national growth paths or as a broadening information base of regional policy-makers. However, very little literature on regional macroeconomic forecasting exists to date.

This thesis concentrates on the growth and business cycle component from a regional point of view. First, it contributes to the discussion on the role of economic structures as determinants of long-term regional economic development (Chapter 2). Second, this thesis extends the literature on economic forecasting mainly from a regional economic point of view (Chapter 3, 4 and 5). The next Section 1.1 summarizes the main findings of this thesis and includes a discussion of how the different chapters are linked. Since three of the four main chapters are written together with co-authors, Section 1.2 identifies my own scientific contributions to these projects.

1.1 Summary and Thematic Coherence

Chapter 2 focuses on the long-term perspective and therefore on the growth component of total output. The chapter explicitly analyzes the role of sectoral structures for regional economic development. Academics see knowledge as a major driving force for regional economic growth. The free circulation of knowledge, for example, via social interaction, makes knowledge available to other firms, thus creating knowledge spillovers. In the economic literature two opposing theories exist as to how knowledge spillovers occur: either within an industry or between industries. The two theories result in different predictions about what regional sectoral structures look like. If knowledge spillovers occur within an industry, this should lead to a specialized sectoral structure. If the flow of knowledge emerges between industries, the result will be a highly diversified sectoral environment. The empirical literature has not reached a consensus yet as to which of these theories has more predictive power. In our view, this ambiguity is not surprising since these two theories cannot be treated as independent from each other because of, for example, the cross-fertilization of ideas. We therefore introduce interaction models between specialization and diversification to the empirical literature to measure the dependency between these theories.

The study of interaction effects between specialization and diversification is conducted for the 70 largest cities in Germany for the period from 1998 to 2008. We derive our empirical model from sector-specific production functions, where specialization and diversification as well as the interaction terms are modeled within the term of total factor productivity (TFP). The four sectors of investigation are manufacturing, construction, basic and advanced services. The empirical model is estimated via panel techniques and we explicitly account for spatial dependence. In the end, the study yields three major insights. First, we find that specialization in manufacturing and basic services has a positive impact. Sectoral diversification has only a minor influence on gross value added per employee in a specific sector. Second, the effects of sectoral specialization and diversification are non-linear, thus the signs of specialization and diversification change at certain thresholds. Third, we find a negative interaction effect between specialization and diversification in manufacturing, construction and for advanced services. Specialization thus has a negative impact on the effects of diversification and *vice versa*. This causes a trade-off between sectoral specialization and diversification. Whereas specialization fosters a positive development in a specific sector, a more diversified surrounding industrial structure reduces the positive effect of specialization.

While Chapter 2 studies the long-term perspective of regional economic development, we turn to the mid-term perspective in Chapter 3 and focus on the business cycle component of total output. Macroeconomic forecasts are powerful tools for reducing uncertainty in the near future. However, only a few studies exist that deal with the topic of macroeconomic forecasting at the regional level. The main reason for this is data unavailability. In our study, we draw on quarterly GDP data for the two German states Saxony and Baden-Württemberg, as well as for Eastern Germany for the period from the first quarter of 1996 to the fourth quarter of 2010. In Chapter 3 we systematically assess the forecast accuracy of regional, national or international indicators for the prediction of regional GDP. These three groups of indicators also enable us to evaluate the forecast performance of large data set methods such as pooling or factor models. These advanced methods have been proved to work well for national GDP forecasts.

It turns out that forecast combination strategies produce the best forecasts compared to all other models or strategies in terms of mean squared forecast errors. Turning to the regional level where the indicator is measured, we find that regional (either Saxon, Baden-Württemberg or Eastern German) and national (here: Germany) indicators have the highest forecast accuracy. International indicators only play a minor role in forecasting regional GDP. The best forecasts are produced by survey results or indicators that represent the specific regional economic structure.

Chapter 4 is closely related to the previous chapter. Instead of asking the general question of which indicators or strategies have the highest forecasting performance for regional GDP, Chapter 4 discusses the role of sectoral disaggregated forecasts. We think that some indicators are more closely linked to a specific industry (such as a business climate indicator for manufacturing) than to the whole economy. Therefore, the bottom-up aggregation of sector-specific GVA forecasts may result in a higher accuracy than the direct forecast of total GVA. Due to data unavailability of sectoral GVA, we exclusively focus on the Free State of Saxony. To answer the research question in Chapter 4, we draw on almost the same indicators and forecasting techniques as in Chapter 3.

We find that regional indicators in particular play a crucial role in forecasting sectoral GVA. As found in Chapter 3, forecast combination strategies can also improve the forecasting accuracy of sectoral GVA in Saxony compared to a simple benchmark model. We also find that disaggregated GVA forecasts produce lower forecasting errors for one quarter ahead predictions than a direct approach for total GVA. The forecasting error can be reduced by about 8% with a disaggregated approach for the forecast horizon of one quarter. The direct approach is preferable for longer forecast horizons of up to four quarter ahead. This leads to the conclusion that a forecaster should predict each single supply side component in the short-term to reduce the forecast error of total GVA.

Chapter 5 builds on the findings of Chapter 4. It also focuses on economic forecasting. We find in Chapter 4 that a disaggregated forecasting approach from the supply side calculation of total output (GVA for the sectors in an economy) lowers forecasting errors. In Chapter 5 we look at the calculation of GDP from the demand side approach. However, we do not ask whether GDP forecast accuracy can be increased by an aggregation of sub-components, but systematically analyze one single sub-component instead, namely exports. We ask whether export forecasts can be improved by either soft indicators like business survey results or by hard data such as price and cost competitive measures. Unfortunately, no quarterly data for exports exist at the regional level, which makes a forecasting exercise impossible. We therefore have to focus on the national level and evaluate the question for twenty European states plus the aggregates EA-18 and EU-28. The period of investigation runs from the first quarter of 1996 to the fourth quarter of 2013. We apply a standard autoregressive distributed lag (ADL) model and forecast encompassing tests to determine whether soft or hard indicators perform better to forecast export growth.

The results clearly show that survey-based indicators produce lower forecast errors than those obtained from hard data. We find the confidence indicator for the manufacturing sector and the production expectations among the best soft indicators. Among the hard indicators only the industrial production for the United States can continuously beat the benchmark model. However, we observe large country differences in the forecasting accuracy of soft and hard indicators. We therefore try to identify the reasons for these differences by running a standard regression analysis that attempts to explain the forecasting errors between the countries by looking at the export composition of the latter. This is done for soft and hard indicators separately. We find that the forecasting accuracy of soft indicators is lower in those countries that have a high export share in raw materials or in oil. The forecast performance of soft indicators becomes better the higher the share in machinery exports is. For hard indicators, the export share has little power to describe country differences in forecasting accuracy.

1.2 Own Contribution to Joint Research Projects

The works in Chapter 2, 3 and 4 are written together with co-authors. Chapter 5 comprises my single-authored paper.

Chapter 2 is a copy of the original article entitled: "Marshall or Jacobs? New insights from an interaction model", which was published in the journal *Jahrbuch für Regionalwissenschaft: Review of Regional Research*, **33** (2), 107-133 (Kluge and Lehmann, 2013). This paper was written together with Jan Kluge, junior economist and doctoral student at the Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich e.V., Dresden Branch. Jan Kluge's contribution lay in the processing of the underlying data set and the analysis of the empirical literature. He also introduced and calculated the interaction effects. My own contribution lay in the derivation of the empirical strategy and the examination of the corresponding literature. I also contributed the discussion section and dealt with the problem of endogeneity. We both worked on the motivation and the interpretation of the results. Jan Kluge and I also worked together on the comments put forward by two anonymous referees.

Chapter 3 is a copy of the article entitled: "Forecasting GDP at the Regional Level with Many Predictors", which was published in the journal *German Economic Review*, **16** (2), 226-254 (Lehmann and Wohlrabe, 2015). This paper is co-authored by Klaus Wohlrabe, economist at the Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich e.V. and Deputy Director of the Ifo Center for Business Cycle Analysis and Surveys. The contributions of Klaus Wohlrabe lay in the preparation of the forecasting approach and the corresponding programming. I compiled the relevant literature and worked on the motivation. Additionally, my contribution lay in the collection of regional variables for our forecasting exercise. We worked together on the interpretation of the results and prepared the revisions to the suggestions of two anonymous referees.

Chapter 4 is a copy of the original article entitled: "Forecasting gross value-added at the regional level: Are sectoral disaggregated predictions superior to direct ones?", which was published in the journal *Review of Regional Research: Jahrbuch für Regionalwissenschaft*, **34** (1), 61-90 (Lehmann and Wohlrabe, 2014). As in Chapter 3, this work was carried out together with Klaus Wohlrabe. The tasks are allocated just as in Lehmann and Wohlrabe (2015): Klaus Wohlrabe prepared the forecasting approach and did the programming, while I was responsible for the literature, motivation and data. We both worked on the interpretation of the results and revisions based on comments by two anonymous referees.

2 Marshall or Jacobs? New insights from an interaction model

With kind permission of Springer Science+Business Media, this chapter is the reprint of the original article by Kluge and Lehmann (2013), published in the journal *Jahrbuch für Regionalwissenschaft: Review of Regional Research*, October 2013, Volume 33, Issue 2, pp 107-133, © Springer-Verlag Berlin Heidelberg 2013.

2.1 Introduction

Knowledge is one of the major driving forces of economic growth (see, e.g., Romer, 1986). The social interaction of workers leads to the free circulation of knowledge and makes this knowledge available to other firms. This process has a clear impact on regional growth. Krugman (2011) even suggests placing more emphasis on such intangible factors when seeking explanations for the spatial distribution of economic activity.

Knowledge spillovers have been examined in a large number of studies. The literature explores where knowledge spillovers occur (within or between industries) and what the consequences of such spillovers are (specialization or diversification of industries). In particular, there are two major opposing concepts that describe how spillovers are responsible for the creation or diffusion of knowledge and thus foster economic growth: localization and urbanization economies.[1] Both concepts emphasize knowledge (besides other advantages that come with spatial proximity) as an important growth driver.

Marshall (1890) argues that companies surrounded by other companies in the same industry will grow faster, making the assumption that knowledge circulates primarily within industries. According to this viewpoint, companies benefit from being located closely to each other because they gain from localization economies. Arrow (1962) and Romer (1986) provide a formalized theory, leading Glaeser *et al.* (1992) to refer

[1] For recent surveys of the existing theoretical and empirical literature, see Rosenthal and Strange (2004) or Beaudry and Schiffauerova (2009).

to localization economies as MAR externalities. The empirical literature finds mixed results for localization economies (see, e.g., Henderson *et al.* (1995); Combes (2000); de Lucio *et al.* (2002); Dekle (2002); Blien and Suedekum (2005) or Dauth (2013) for different countries).

An opposing opinion was advanced by Jacobs (1970), who rejects the notion that knowledge flows within industries. According to her, companies gain from a diverse environment consisting of different types of industries. New ideas come not from within but from outside a firm's sector. The mechanisms by which diversity leads to economic growth are usually called urbanization economies. Empirical evidence for this type of externality is provided by Glaeser *et al.* (1992), Lee *et al.* (2005), Blien *et al.* (2006), Fuchs (2011) or Illy *et al.* (2011).

In spite of the rich body of literature, one must clearly state that there is no consensus about the effects of knowledge spillovers and in particular about the correctness of any of the two above mentioned concepts (Beaudry and Schiffauerova, 2009). This paper is aiming at finding reasons for this ambiguity and introducing new insights into the Marshall (1890) vs. Jacobs (1970) debate.

Most of the above mentioned and prominent studies do not take into account two relevant issues, which we directly address in our paper. First, these studies assume that localization and urbanization economies work independently of one another. Many studies find effects for both economies, but interaction effects are neglected.[2] Second, most of the literature does not consider non-linear effects of localization and urbanization economies. A few studies show U- or inverse U-relationships for these two externalities (see de Lucio *et al.*, 2002; Illy *et al.*, 2011). Additionally, Martin *et al.* (2011) argues that, e.g., congestion effects might foil the positive impacts of localization or urbanization economies. This leads to a threshold interpretation of agglomeration economies, where the effects of those two externalities may harm regional productivity.[3]

In our paper, we add to the existing literature by making the argument that localization and urbanization economies occur simultaneously and interact with each other. For example, a company can benefit from a more specialized environment (e.g., customers, suppliers, competitors) if enough service providers from other sectors (e.g., universities and research facilities) are available. Fahrhauer and Kröll (2012) argue that firms sim-

[2] One exception is the study of Brunow and Hirte (2009) who intend to study the interaction between agglomeration variables and production factors. Because of strong multicollinearity issues, they refrain from testing such interaction effects.

[3] Non-linear effects were studied in a few papers before (see de Lucio *et al.*, 2002; Illy *et al.*, 2011). Nevertheless, the authors who consider this issue do not interpret their results within the framework of interaction models.

ply profit from a common pool of knowledge, which is why new ideas can come from both within and outside the specific sector and are combined with each other, e.g., via cross-fertilization. This leads to an improvement of own ideas. Therefore, localization and urbanization economies can be even more pronounced when they interact with each other.[4] Nevertheless, to the best of our knowledge there is not much evidence, neither theoretical nor empirical, how interactions between local actors (e.g., firms, persons or institutions) work and what the mechanisms behind such interactions are. Additionally, a rethinking on economic geography is observable in the academic literature (see, e.g., Bathelt and Glückler, 2003). We therefore state that it is not possible for us to theoretically predict how interactions between localization and urbanization economies behave. Whereas the first argument implies a positive sign, an interaction could possibly have negative impacts. We therefore do not expect a specific sign of the interaction between localization and urbanization economies and leave this as an empirical matter.

Furthermore, our contribution accounts for non-linearities of localization and urbanization economies. We argue that the two externalities only occur if a sufficient number of knowledge carriers has assembled. Cities with few firms in a certain industry might therefore benefit more from further specialization than a city without any at all. However, there may also be disadvantages of further specialization, e.g., due to extensive competition or congestion costs. With interaction models, we are able to analyze non-linear and interaction effects graphically and give a threshold interpretation. This goes beyond the usual way of analyzing or interpreting point estimates or average effects instead of varying the variables of interest within their range of observations. When we present our empirical model, we elaborate more on this point.

Since existing studies, e.g., Blien *et al.* (2006), imply that the results of such analysis might differ between industries, we consider four different sectors at the 1-digit level using administrative German data. Another contribution of our paper is to study gross value added (GVA) per employee for different sectors.

We find negative interactions between localization and urbanization economies in most sectors. Furthermore, the effects often have a U-shaped form, meaning that localization (urbanization) economies become stronger with higher levels of specialization (diversification). We find some evidence for localization economies in some sectors. Urbanization economies do not play a major role for economic development in our sample. Besides searching for evidence of realized effects, we are also able to describe hypotheti-

[4] Fahrhauer and Kröll (2012) introduce the concept of diversified-specialization. In contrast to their study, we examine sector-specific effects and not overall effects on economic growth in cities. Furthermore, we introduce a different estimation approach which was not part of the literature yet.

cal constellations under which localization and urbanization economies could effectively be observed.

The paper is organized as follows. In section two, the estimation approach and data are described. Section three presents sector-specific regression results. The last section concludes.

2.2 Methods and Data

2.2.1 Estimation Approach

Knowledge spillovers influence the productivity of workers and thus the productivity of a firm. These spillovers foster economic development, which is why we start with a regional growth model. Since interaction terms have not been studied in this part of the economic literature before, we cannot refer to any theoretical model. Nevertheless, we make use of the existing literature and assume the following sector-specific Cobb-Douglas production function with constant returns to scale,

$$Y_{z,s,t} = A_{z,s,t}(\mathbf{X})\,(K_{z,s,t})^{\alpha}\,(L_{z,s,t})^{1-\alpha}\,h_{z,t}^{\eta}\ , \tag{2.1}$$

where $Y_{z,s,t}$ represents the sector-specific (s) real GVA in city z for a given year (t). Total factor productivity (TFP) is denoted with $A_{z,s,t}$ and depends on different shifting parameters ($\mathbf{X}$). GVA is produced with capital ($K_{z,s,t}$) and labor ($L_{z,s,t}$). α and $1-\alpha$ are the corresponding output elasticities. To account for human capital externalities, we include the average human capital level $h_{z,t}^{\eta}$ with the respective elasticity η (see Lucas, 1988).[5] Expressing (2.1) in terms of per employee gives

$$y_{z,s,t} = A_{z,s,t}(\mathbf{X})\,(k_{z,s,t})^{\alpha}\,h_{z,t}^{\eta}\ , \tag{2.2}$$

with $k_{z,s,t}$ representing capital intensity. We cannot observe TFP directly. However, following de Lucio *et al.* (2002), Brunow and Hirte (2009) and Martin *et al.* (2011), we assume that the functional form of TFP is given by

$$A_{z,s,t}(\mathbf{X}) = U_{z,s,t}e^{g(Spec_{z,s,t},\ Div_{z,s,t})}\ . \tag{2.3}$$

[5] Note that we have no data on sector-specific human capital due to data limitations.

TFP depends on specialization ($Spec_{z,s,t}$) and diversification ($Div_{z,s,t}$). $U_{z,s,t}$ represents sector- and city-specific components. Since (2.3) is a very general function, we assume that it takes the following form:

$$\begin{aligned} g(Spec_{z,s,t},\ Div_{z,s,t}) =& \beta \ln(Spec_{z,s,t}) + \gamma \ln(Div_{z,s,t}) + \delta \left(\ln(Spec_{z,s,t})\right)^2 \\ &+ \phi \left(\ln(Div_{z,s,t})\right)^2 + \chi \left(\ln(Spec_{z,s,t})\right) * \left(\ln(Div_{z,s,t})\right) \ . \end{aligned} \tag{2.4}$$

To account for non-linearities and possible interaction effects between localization and urbanization economies, we extend the standard model with several interaction terms. These interaction terms have the following form:

$$\begin{aligned} &\left(\ln(Spec_{z,s,t})\right)^2 \\ &\left(\ln(Div_{z,s,t})\right)^2 \\ &\left(\ln(Spec_{z,s,t})\right) * \left(\ln(Div_{z,s,t})\right) \ . \end{aligned}$$

The introduction of these interaction terms yields a fully specified interaction model (see Brambor *et al.*, 2006). We expect that the impacts of specialization and diversification are non-linear with $\delta, \phi \neq 0$. This results in a U-shaped or inverse U-shaped form of the effects of localization and urbanization economies. To measure a possible linkage between localization and urbanization economies, we use the product of the variables for specialization and diversification. Following the motivation in Sect. 2.1, we expect the interaction term to be non-zero ($\chi \neq 0$).

Inserting (2.3) into (2.2) and taking logs yields our empirical model,

$$\begin{aligned} \ln(y_{z,s,t}) = & \ c_s + \alpha \ln(k_{z,s,t}) + \beta \ln(Spec_{z,s,t}) + \gamma \ln(Div_{z,s,t}) \\ & + \delta \left(\ln(Spec_{z,s,t})\right)^2 + \phi \left(\ln(Div_{z,s,t})\right)^2 + \chi \left(\ln(Spec_{z,s,t})\right) * \left(\ln(Div_{z,s,t})\right) \\ & + \eta \ln(h_{z,t}) + controls + a_{zs} + v_t + \varepsilon_{z,s,t} \ . \end{aligned} \tag{2.5}$$

GVA per employee in real terms ($y_{z,s,t}$) is described by a sector-specific constant (c_s), the respective capital intensity ($k_{z,s,t}$) and the average human capital level ($h_{z,t}$). Furthermore, specialization ($Spec_{z,s,t}$) and diversification ($Div_{z,s,t}$) as well as our interaction terms play a role. Finally, we add several control variables: the balance of commuters to measure knowledge movements between a city and the surrounding periphery and the balance of migrants of a region to capture possible labor market effects as well as to mitigate the problem of correlation between regional units. Additionally, we assume that $\ln U_{z,s,t}$ is defined as $\ln U_{z,s,t} = c_s + controls + a_{z,s} + v_t + \varepsilon_{z,s,t}$. City- and industry-specific time-invariant fixed-effects are captured with $a_{z,s}$. Following Martin *et al.* (2011), these fixed-effects explicitly control for unobservable regional and sectoral characteristics such as the access to natural resources, infrastructure or local public services (static externalities). This empirical approach makes it possible to measure dynamic externalities such as localization and urbanization economies explicitly, controlling for static externalities and unobservable variables that do not vary over time. Furthermore, v_t are year dummies to capture business cycle effects that feed back to the region as well as temporal shocks to the economy. $\varepsilon_{z,s,t}$ is an error term.

The choice of appropriate indicators for localization and urbanization economies is critical. Beaudry and Schiffauerova (2009) recommend using a separate indicator for each of the two externalities, as both types can occur simultaneously.[6] They argue that otherwise it is impossible to distinguish between the two types of economies. Furthermore, we have to use one separate indicator for each concept to be able to consider interaction effects.

Following Glaeser *et al.* (1992), we use the relative location quotient,

$$Spec_{z,s,t} = \frac{labor_{z,s,t}/labor_{z,t}}{labor_{s,t}/labor_t}\ , \tag{2.6}$$

to measure specialization. Using employment figures ($labor_{z,s,t}$), this quotient compares the degree of specialization of an industry s in city z to the case where industry-specific employment is randomly distributed across Germany (see Glaeser *et al.*, 1992; Dekle, 2002). The relative location quotient takes a value greater than one if the share is above the German average. To ensure comparability with other studies, we use this relative location quotient since it is the most applied indicator in the existing literature (see Beaudry and Schiffauerova, 2009). Additionally, van Soest *et al.* (2002) have shown for the Netherlands that a relative version better captures localization economies in comparison to a simple location quotient.

[6] The alternative would be to define diversification as the absence (i.e. the inverse) of specialization.

Urbanization economies require a focus on the environment of a specific industry and the measurement of the degree of diversification of that environment. Following, e.g., Blien *et al.* (2006), we use an inverse Hirschman-Herfindahl-Index,

$$Div_{z,s,t} = \frac{1/\sum\limits_{s'=1,s'\neq s}^{S}(labor_{z,s',t}/(labor_{z,t} - labor_{z,s,t}))^2}{1/\sum\limits_{s'=1,s'\neq s}^{S}(labor_{s',t}/(labor_{t} - labor_{s,t}))^2} \,. \tag{2.7}$$

The index becomes larger as the diversity of the environment of industry s increases relative to the national average. To measure the diversification of the environment, we exclude the respective industry ($s' \neq s$) from the calculation. As before, we use regional employment figures to calculate $Div_{z,s,t}$. To be in line with most of the existing studies, we use the Hirschman-Herfindahl-Index to measure urbanization economies. Nevertheless, the choice of a suitable indicator may be one reason why the literature has not found a consensus regarding localization and urbanization economies yet (see Beaudry and Schiffauerova, 2009). We decided to use this indicator because it seems that this index does not distort the results in one direction. Positive, neutral and negative results for urbanization economies are published in several studies.

It is important to note that specialization and diversification do not exclude each other since the Hirschman-Herfindahl-Index does only take into account the surrounding industry of a particular sector. In fact, there are cities in our sample that are highly specialized but nonetheless characterized by a diversified surrounding (e.g., Wolfsburg).

Including interaction terms makes it possible to study the marginal effects of one variable (e.g., specialization), conditioned by the variation of another variable (e.g., diversification). Therefore, the marginal effects of specialization or diversification on GVA per employee in our model are as follows,

$$\frac{\partial y}{\partial \ln(Spec_{z,s,t})} = \beta + 2\delta \ln(Spec_{z,s,t}) + \chi \ln(Div_{z,s,t}) \,, \tag{2.8}$$

$$\frac{\partial y}{\partial \ln(Div_{z,s,t})} = \gamma + 2\phi \ln(Div_{z,s,t}) + \chi \ln(Spec_{z,s,t}) \,. \tag{2.9}$$

With these two expressions, it is possible to assess the strength and sign of localization and urbanization economies for different levels of specialization and diversification.[7] We measure the average effect if the two conditioning variables $\ln(Spec_{z,s,t})$ and $\ln(Div_{z,s,t})$ are zero, which, in our case, represents the German Average. Since specialization and diversification are not symmetrically distributed around the German average in our empirical analysis, interpreting simple coefficient estimates can result in misleading conclusions. We therefore interpret the marginal effects of our empirical model.

The coefficients of (2.5) are estimated using linear panel model techniques. All variables are stationary in levels or trend stationary. There is no problem with multicollinearity in our data.[8] We have tested, whether the assumptions for the residuals $\varepsilon_{z,s,t}$ hold: homoscedasticity, autocorrelation and cross-sectional or spatial dependence. Especially the problem of cross-sectional correlation in the error term causes the standard errors to be biased (see Hoechle, 2007). We employ the test developed by Pesaran (2004) to check whether cross-section dependence in the residuals is present. Whenever only heteroscedasticity and autocorrelation are present, we use a cluster-robust estimator to achieve unbiased standard errors. In the presence of very general forms of spatial correlation, we use the approach of Driscoll and Kraay (1998).[9] This nonparametric variance-covariance-matrix estimator produces consistent standard errors when all three before mentioned problems are present (see Hoechle, 2007) and is an extension to the HAC variance-covariance-matrix estimators proposed by Newey and West (1987) and Andrews (1991). The Driscoll and Kraay estimator has the advantage that it requires no knowledge about the form of cross-section dependence and puts only weak restrictions on it (see Driscoll and Kraay, 1998).

To determine whether to use the fixed-effects estimator or a random-effects estimator for (2.5) when cross-section dependence is absent, we apply a heteroscedasticity and cluster-robust form of the Hausman test.[10] The Driscoll and Kraay estimator is only applicable in a pooled or fixed-effects setup. A fixed-effects estimator has the before mentioned advantage: It captures time-invariant static externalities.

[7] In a model without interaction terms, the marginal effects are only β and γ. For further details on interaction models, see Brambor *et al.* (2006).

[8] One would suggest that the inclusion of interaction terms leads to strong multicollinearity problems. Brambor *et al.* (2006) state that dropping interaction terms would result in omitted variable bias, which is a much more striking problem in an empirical setup. Additionally, we have checked the multicollinearity issue using variance inflation factors and found that it is not a problem for our analysis.

[9] Hoechle (2007) suggests using the Driscoll and Kraay estimator in the presence of spatial dependence. Otherwise, he prefers a cluster-robust estimator.

[10] This module was developed by Schaffer and Stillman (2010).

Considering interaction terms let us suggest that our results may differ from those of previous studies because of two reasons. First, as it was addressed by Martin *et al.* (2011), we study the impacts of specialization and diversification in the short-term. In contrast to previous studies, which in most cases study agglomeration effects in a cross-section setup and therefore measure long-term impacts, we make use of panel data and the time variation therein. This short-term approach seems to be appropriate for detecting non-linearities in agglomeration effects (see Martin *et al.*, 2011) and we suggest to detect interactions between these two externalities. Second, introducing interaction effects may provide new insights into the force of agglomeration effects, since we do not only measure average effects.

2.2.2 Data

We use administrative data for the 70 largest German cities (in terms of population in 2008) with county status for the period between 1998 and 2008. The use of cities is advisable since the density of knowledge carriers and companies is greatest in large cities. Many researchers support the use of cities for such analyses, e.g., Feldman and Audretsch (1999).

To obtain the indices for specialization ($Spec_{z,s,t}$) and diversification ($Div_{z,s,t}$), we use all full-time employed persons subject to social security ($labor_{z,s,t}$).[11] These figures are provided by the German Federal Employment Agency (2010). The Working Group Regional Accounts VGRdL (2011b) provides data on nominal GVA for the German cities. Because of the lack of price indices at the regional level, we deflate GVA in nominal terms with the state-specific deflator.[12] We have to mention that this procedure only accounts for the annual state-specific inflation rate and not for differences in regional price levels.

We use data on GVA, even though most previous studies have used employment as the dependent variable. The use of employment for this purpose has often been criticized in the literature. First, labor must be a homogeneous input factor and regionally mobile (see Beaudry and Schiffauerova, 2009). This is not the case because migration costs are not equal across countries and not constant over time (see Almeida, 2007). Second, capital and labor are substitutes. Whenever technological progress is a labor-saving process, employment growth is not a good proxy for economic growth (see Paci and

[11] For example, civil servants and self-employed are not included in these figures.

[12] For example, for deflating nominal GVA of Frankfurt am Main, we use the deflator of the German state Hesse. For Munich, we use the deflator of Bavaria. The state-specific deflator is the ratio of unchained GVA measured in previous year prices and nominal GVA measured in actual prices. This approach is used for official statistics in Germany.

Usai, 2005). Third, using employment one has to assume that the capital stock remains constant over time (see Dekle, 2002). Because of these three issues, GVA might be the better indicator of productivity or economic growth. Following Dekle (2002), we therefore use GVA rather than employment.

Regional GVA figures are available for seven different sectors. We run our regression for only four (manufacturing, construction, basic and advanced services) of these sectors.[13] Basic services are, e.g., wholesale and retail trade or hotels. Advanced services comprises the sectors financial intermediation and real estate. Despite the exclusion of three sectors from our empirical analysis, we calculated the sector-specific Hirschman-Herfindahl-Index by taking the six remaining sectors into account. We know that our dependent variable is measured on a broad classification scheme. On the one hand, Beaudry and Schiffauerova (2009) mention that using data from the 1-digit or 2-digit level may result in an overestimation of localization economies. On the other hand, they state that it becomes even more difficult to distinguish between these two externalities by using data from a higher level of disaggregation (e.g., 3-digit level). Since we are probably the first who introduce interaction terms, we follow the before mentioned discussion on GVA and the paper by Dekle (2002) and use figures from the 1-digit level. Besides, no GVA data for a more detailed classification scheme is available at the regional level in Germany.

Another important point concerning the aggregation of sectors is the problem of intermediates. Despite the high level of aggregation it is still possible that, e.g., the sector of construction is an intermediate for manufacturing, which leads to misleadingly measuring urbanization economies as localization economies. However, our analysis shows that diversification may have an impact (positive or negative) for intermediates such as construction or basic services. Additionally, the literature shows that it is indispensable to analyze the whole range of economic activities (see, e.g., Combes, 2000; Rosenthal and Strange, 2004).

Since we know that knowledge spillovers are not the only component of agglomeration economies, we have to use several control variables in our fixed-effects estimation. Data on sector-specific capital stocks of all German states is available from the Working Group Regional Accounts VGRdL (2010). Data on capital stocks at the regional level

[13] We exclude the agriculture, mining, energy and water supply sectors as well as the public sector. It is hardly imaginable that, e.g., public service sectors gain from localization or urbanization economies. Public service providers cannot move freely between cities, which makes specialization nearly impossible. Furthermore, every city must provide a certain range of public goods. The effects from specialization in the mining sector are due to the initial wealth of natural resources and not to localization economies.

is missing. To avoid omitted variable bias, we need data for German cities. Therefore, viable weights to disaggregate the state-specific capital stocks are necessary. Assuming constant elasticities of substitution between capital and labor, we use employment shares as our weighting scheme. For example, x% of employees in the manufacturing sector of a German state i are working in city z. Then this city's share in the state-specific capital stock in manufacturing is x%. This approach should generate viable proxies for regional capital stocks. Average human capital is measured via the share of high-qualified employees in total employment in a city (Federal Office for Building and Regional Planning (BBR), 2010). Additionally, we include the balances of commuters (German Federal Employment Agency) and migrants (Federal Office for Building and Regional Planning (BBR), 2010). Descriptive statistics for all before mentioned variables can be found in the Appendix 2.A.

2.3 Results

The following section presents our regression results for manufacturing, advanced services, construction and basic services. When applying an interaction model, it is always preferable to concentrate on a graphical analysis in addition to the numeric regression results (see Brambor *et al.*, 2006). The results for all sectors are presented in Table 2.1. Note that the table only shows average effects, e.g., the marginal effect of variable $\ln(Spec_{z,s,t})$ given that the conditioning variable $\ln(Div_{z,s,t})$ in (2.8) is set to zero. A comfortable feature about the definition of our measures for $\ln(Div_{z,s,t})$ and $\ln(Spec_{z,s,t})$ is that a value of zero indicates the German average. This is not very helpful if we admit that the distribution of $\ln(Spec)$ and $\ln(Div)$ around the German average is not necessarily symmetric. Whenever most of the cities face, e.g., a specialization index smaller than the German average, the average effect would rather be hypothetical and therewith irrelevant for our sample. The pictures shown in the following subsections will make clear how the results behave if cities deviate from these averages. This will deliver an idea of how relevant the average results are for our data.

We discuss the figures for manufacturing and advanced services in greater detail. Abridged results for construction and basic services are shown afterward. We first present the figures for specialization and then show the results for diversification.

Table 2.1: OLS-Regression results

Sector	Manufacturing	Construction	Basic Services	Advanced Services
Preferred estimation technique	Driscoll& Kraay FE	FE cluster-robust	Driscoll& Kraay FE	Driscoll& Kraay FE
Variables:				
ln(Spec)	0.082	-0.056	0.230**	-0.597***
	(0.85)	(-0.85)	(2.90)	(-15.30)
ln(Div)	-0.436*	-0.368**	0.017	0.015
	(-1.97)	(-2.14)	(0.46)	(0.12)
Interactions:				
$\ln(\text{Spec})^2$	0.029	0.154**	0.020	0.274***
	(1.39)	(2.43)	(0.24)	(5.47)
$\ln(\text{Div})^2$	-0.355	0.675	0.304***	0.504***
	(-0.48)	(1.59)	(4.29)	(3.44)
ln(Spec)*ln(Div)	-0.556	-1.166***	0.081	-0.433***
	(-1.36)	(-4.00)	(0.78)	(-4.71)
Controls:				
ln(Capital intensity)	0.344***	0.360***	0.064*	0.673***
	(6.49)	(8.50)	(2.20)	(7.35)
ln(Share of high qualified)	-0.424***	0.082	0.180***	0.026**
	(-6.35)	(0.99)	(10.47)	(2.63)
Balance of commuters	2.00e-06***	-1.07e-06	-1.14e-06*	-4.77e-06***
	(3.67)	(-0.80)	(-2.22)	(-12.31)
Balance of migrants	3.04e-04	-0.001	0.001***	1.55e-04
	(0.88)	(-1.42)	(4.25)	(0.53)
Constant	7.675***	6.590***	9.362***	2.347*
	(12.45)	(12.80)	(30.34)	(1.94)
Time dummies	Yes	Yes	Yes	Yes
R^2	0.559	0.215	0.697	0.538
Obs.	770	770	770	770

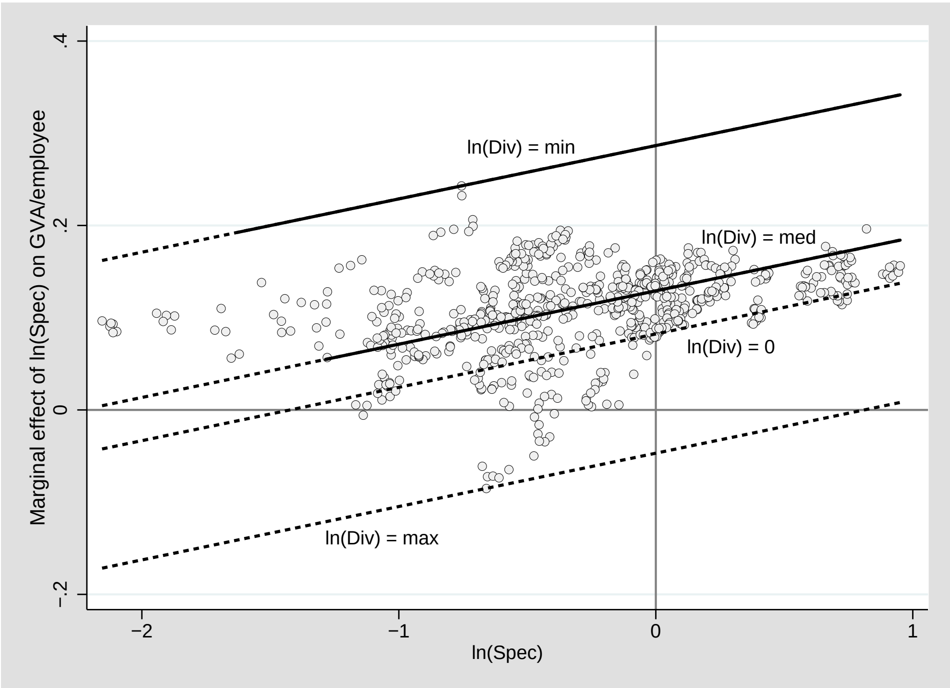
Note: Driscoll&Kraay FE...heteroskedasticity, auto- and spatial correlation robust fixed-effects estimation, FE cluster-robust...cluster-robust fixed-effects estimation, *t*-stats in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, 10% level. *Dependent variable:* logarithmic real GVA per employee. *Source:* authors' calculation.

2.3.1 Manufacturing

Specialization

Figure 2.1 shows the marginal effect of specialization $\ln(Spec_{z,s,t})$ on GVA per employee (y-axis) for different levels of diversification $\ln(Div_{z,s,t})$. This figure is the estimation outcome of Eq. (2.8). The x-axis of Fig. 2.1 presents the range of the degree of specialization that we observe in our data. The scatter plot depicts the marginal effect for each city. The lines show the hypothesized marginal effects, keeping the conditioning variable (in this case $\ln(Div_{z,s,t})$) constant at different levels. We choose four different values: the minimum, the median and the maximum level of diversification as well as $\ln(Div_{z,s,t}) = 0$. The last value represents the average marginal effect shown in Table 2.1. Whenever the interaction effect is positive, the line for $\ln(Div_{z,s,t}) = max$ will be located above $\ln(Div_{z,s,t}) = min$ and *vice versa*. A positive slope of the lines indicate a positive coefficient of $\ln(Spec_{z,s,t})^2$ and *vice versa*. Solid lines show significant marginal effects at the 90% confidence level while scattered areas indicate insignificant ones.

Figure 2.1: Effects of specialization in manufacturing



To combine Fig. 2.1 with Table 2.1, we first concentrate on the case in which the conditioning variable $\ln(Div_{z,s,t})$ takes a value of zero, i.e., is at the German average. In the case of manufacturing, the coefficient of $\ln(Spec_{z,s,t})$ is 0.082 which is the value at the y-axis where the line for $\ln(Div_{z,s,t}) = 0$ intercepts the zero line of $\ln(Spec_{z,s,t})$. The fact that the line is scattered at this point shows that the average effect of $\ln(Spec_{z,s,t})$ is insignificant. The coefficient of $\ln(Spec_{z,s,t})^2$ indicates a positive slope (2*0.029). The interaction effect is negative (-0.556) which is why the line for $\ln(Div_{z,s,t}) = min$ is to be found above the others. We observe, that the marginal effect of specialization is significantly positive for a large share of our observations within an area in the center of Fig. 2.1. The effects tend to be larger when the level of diversification gets smaller due to the negative interaction effect (i.e., further above in Fig. 2.1). Therefore, we can conclude that there are realized localization economies in manufacturing although the average effect is insignificant. It can be doubted that the average effect at $\ln(Div_{z,s,t}) = 0$ is the most relevant since most cities offer less diversified surroundings for their manufacturing industry than the German average suggests. The scatter plot shows the asymmetric distribution of $\ln(Div_{z,s,t})$, since the median line lies above the zero line.

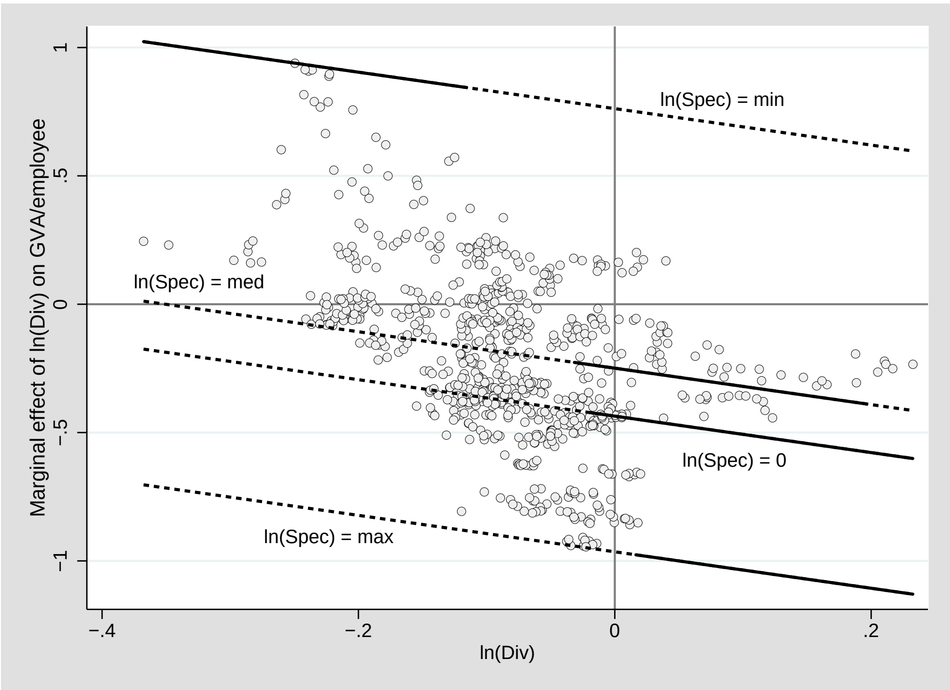
How do our results for specialization in the manufacturing sector fit into the existing literature? Compared with results for Germany, we are in line with most of the existing studies such as Blien *et al.* (2006), Brunow and Hirte (2009), Ehrl (2013) and Fahrhauer and Kröll (2012) which find positive effects of localization economies. The coefficient for specialization is fairly in line with those studies using output instead of employment growth. Nevertheless, we find opposite results in comparison to the studies by Blien and Suedekum (2005) or Illy *et al.* (2011) who find negative effects of specialization. There are several reasons why these differences emerge. First, we use data that differ in the composition of sectors, cities and years. Second, we estimate a panel with interaction effects, whereas these two studies carry out a cross-sectional analysis. We can explicitly account for time fixed-effects as well as short term variations. Finally, our dependent variable is GVA per employee and not employment growth. In Sect. 2.2.2 we mentioned the discussion on employment vs. economic growth. Some studies find opposite results for localization economies when using employment instead of productivity (see, e.g., Dekle, 2002; Almeida, 2007). The most striking reason, however, is the interpretation of interaction terms: If no interactions are used, researchers will only report average effects. Two authors using different data sets might find similar results (e.g., like in Fig. 2.1) but report completely different coefficients if the distributions of the two data sets differ strongly. This fact might explain why so many studies in this field stand in contrast to each other.

From an international point of view, our results are mostly in line with the existing literature (see Dekle, 2002; Mukkala, 2004; Martin *et al.*, 2011). The finding of a positive coefficient for squared terms is also in line with, e.g., de Lucio *et al.* (2002) or Illy *et al.* (2011). The positive coefficient of the squared term implies a U-shaped relationship of localization economies.

Diversification

The interpretation of Fig. 2.2 is basically the same as in Fig. 2.1. As presented in Eq. (2.9), it shows the marginal effect of diversification (y-axis) depending on its actual level (x-axis). Now, the conditioning variable is $\ln(Spec_{z,s,t})$.

Figure 2.2: Effects of diversification in manufacturing



On average, i.e. when $\ln(Spec_{z,s,t}) = 0$ and $\ln(Div_{z,s,t}) = 0$, diversification has a negative and significant impact on GVA per employee (-0.436). The squared term indicates a negative slope. This means that the effect becomes larger (in absolute terms) and even significant the higher the level of diversification is. The scatter plot shows, however, that this result only holds for few cities in our data. The majority does not face any significant effects. The level of specialization decreases the marginal effect of diversification due to the negative interaction effect.

For manufacturing, we must reject the proposal of Jacobs (1970). For the relevant ranges of parameters, the marginal effects of diversification are insignificant. Manufacturing companies in only few cities are facing an environment being so diversified that negative effects occur (further right in Fig. 2.2). Only few cities have, however, such low levels of specialization that positive effects could be observed (further above in Fig. 2.2).

Our results are in line with existing studies which use productivity or output measures for their analysis (see, e.g., Combes, 2000; Dekle, 2002; de Lucio *et al.*, 2002; Fahrhauer and Kröll, 2012). Nevertheless, they stay in contrast to the German studies by Blien and Suedekum (2005), Blien *et al.* (2006) and Brunow and Hirte (2009). The reasons for this difference may be the different dependent variable used in the two former studies as well as the sectoral classification scheme. Concerning our coefficient for urbanization economies, we find similar values in the literature; at least for studies which are comparable.

A first policy implication might be that specialization plays a major role for the sector of manufacturing as it increases GVA per employee. The level of diversification has no considerable impact per se but it harms the positive effects of specialization. Leaving the risk aside, that comes along with strongly specialized economies, specialization seems to be a growth driver in manufacturing.

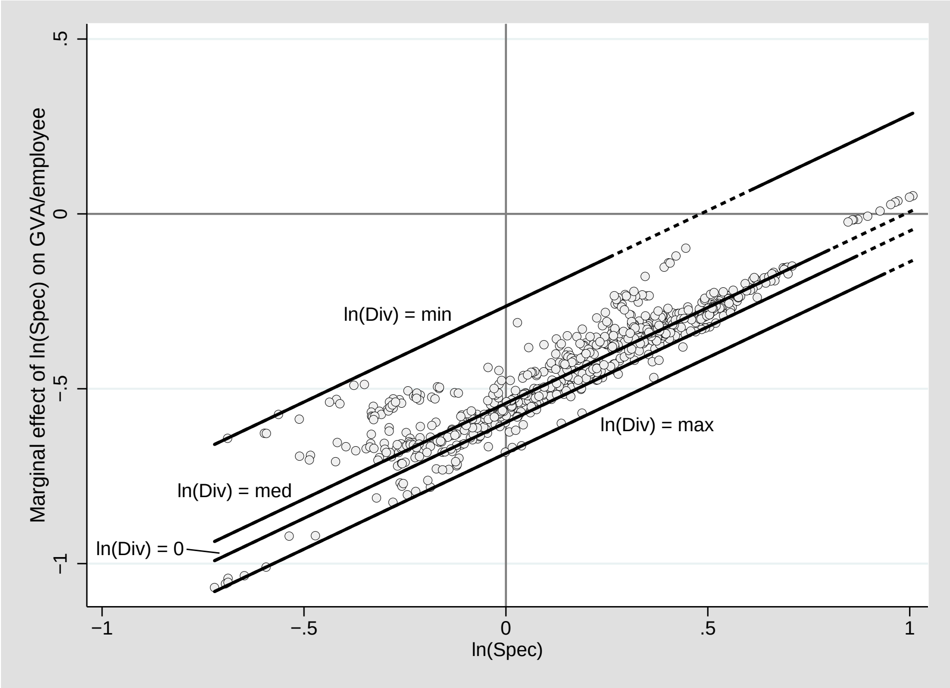
2.3.2 Advanced Services

Specialization

The results for specialization in the sector of advanced services are presented in Fig. 2.3. The coefficient of $\ln(Spec_{z,s,t})$ is significantly negative (-0.597) on average as shown in Table 2.1. The coefficient of the squared term is significantly positive (0.274) as indicated by the positive slope. Like in manufacturing, the interaction effect is negative.

As can be seen from the scatter plot, most cities face negative marginal effects from specialization in advanced services regardless of the level of diversification. Only those companies being located in cities with very high specialization and very low diversification levels (further to the upper right in Fig. 2.3) can possibly gain from localization economies. This might be the reason why advanced services tend to cluster excessively in certain cities like Frankfurt am Main or, from an international point of view, London or New York. We do not have sufficient evidence for realized positive localization economies in advanced services but we can state that they are generally possible.

Figure 2.3: Effects of specialization in advanced services



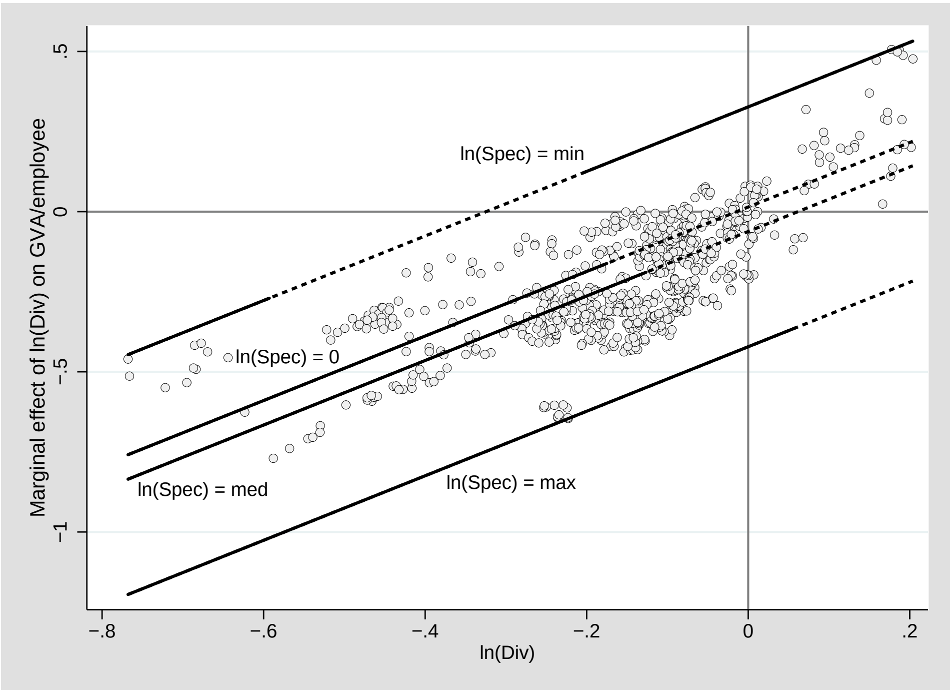
The comparison of our results for advanced services with other studies proves to be more difficult, because most of the before mentioned studies either compromise advanced and basic services within one sector or only analyze manufacturing. Blien and Suedekum (2005) find a positive effect of specialization for advanced services in Germany. Additionally, Dekle (2002) detects localization economies for advanced services in Japan. Our results stay in contrast to these two studies. Nevertheless, it is in line with the study by Combes (2000). An explanation, as it was stated in the study by Combes (2000), can be the linkage of localization economies to the business cycle. Whenever they do not work symmetrically, then negative effects can be observable. More observations are necessary to test such a hypothesis. A viable approach would be to interact the variable measuring specialization with time dummies.

Turning to the squared term of specialization, we find that our result stands in conflict to Illy *et al.* (2011). The fact that we are using GVA per employee instead of employment growth could be a reason why the results change.

Diversification

Turning to the average impact of diversification on GVA per employee in Fig. 2.4, we find a positive but insignificant effect (0.015). As shown by the positive slope, the marginal effect of diversification increases with higher levels of diversification. However, since most cities provide less diversified surroundings than the German average, some of them face significantly negative marginal effects. The highly significant interaction term again makes a difference: Companies in advanced services being located in considerably diversified cities with a low specialization level in this sector, may even gain significantly positive effects. However, these are realized by only a handful of observations.

Figure 2.4: Effects of diversification in advanced services



This completes the picture from Sect. 2.3.2: A specialization strategy in advanced services can only bring advantages in terms of localization economies when it builds upon a considerable already achieved share in this sector. Most cities are not in that position as shown by the scatter plot in Fig. 2.3. When a city chooses not to specialize, it should provide a sufficient level of diversification and hereby reach a situation in which it might gain from urbanization economies. This significant trade-off is much stronger than in manufacturing.

We conclude that there is little evidence for realized (positive) urbanization economies in our data. Our findings reflect the results of Dekle (2002) and Blien and Suedekum (2005), who find negative but insignificant effects of diversification in advanced services.

2.3.3 Construction

The results for the sector construction are shown in Fig. 2.5. The marginal effect of specialization on GVA per employee, depicted by the upper panel, is negative but insignificant (-0.056). The slope is positive, which indicates greater effects for higher levels of specialization but significantly negative ones for lower levels. As in manufacturing and advanced services, a negative interaction effect (-1.166) exists. If local construction companies would face an environment with a very low level of diversification, there would always be a significantly positive marginal effect from specialization for those companies. Again, there is a trade-off: The lower panel in Fig. 2.5 shows that diversification yields negative marginal effects in cities with higher specialization levels. It seems attractive for cities to keep both the specialization level in construction and the diversification level of the respective surrounding at the minimum level. This provides positive marginal effects from specialization while avoiding disadvantages from low diversification. We must therefore conclude, that we do not have persuasive evidence for neither (realized) localization nor (realized) urbanization economies in construction. To the best of our knowledge, no paper studied the effects in construction before. Hence, we are not able to compare our results to other studies.

2.3.4 Basic Services

The results for the sector of basic services are provided in Fig. 2.6. The average effect of specialization is positive and significant (0.230). The coefficient of the squared term is again positive which leads to a positive slope. In contrast to the other sectors, the interaction effect is positive. As the upper panel in Fig. 2.6 shows, positive marginal effects from specialization prevail for most of the cities except for very low levels of specialization or very low levels of diversification. We interpret this finding as considerable evidence for realized localization economies. The positive effect of specialization in basic services was also found by Dekle (2002).

While the average marginal effect of diversification is insignificant, it turns out that many of our observations face significantly negative effects regardless of the level of specialization (see lower panel in Fig. 2.6). This is in line with the findings of Dekle (2002). However, few cities with very high levels of diversification manage to realize positive urbanization economies for their basic service companies. Thus, the sectors of basic and advanced services as well as manufacturing realize positive urbanization economies for a very small share of the data set.

Figure 2.5: Effects of specialization (upper panel) and diversification (lower panel) in construction

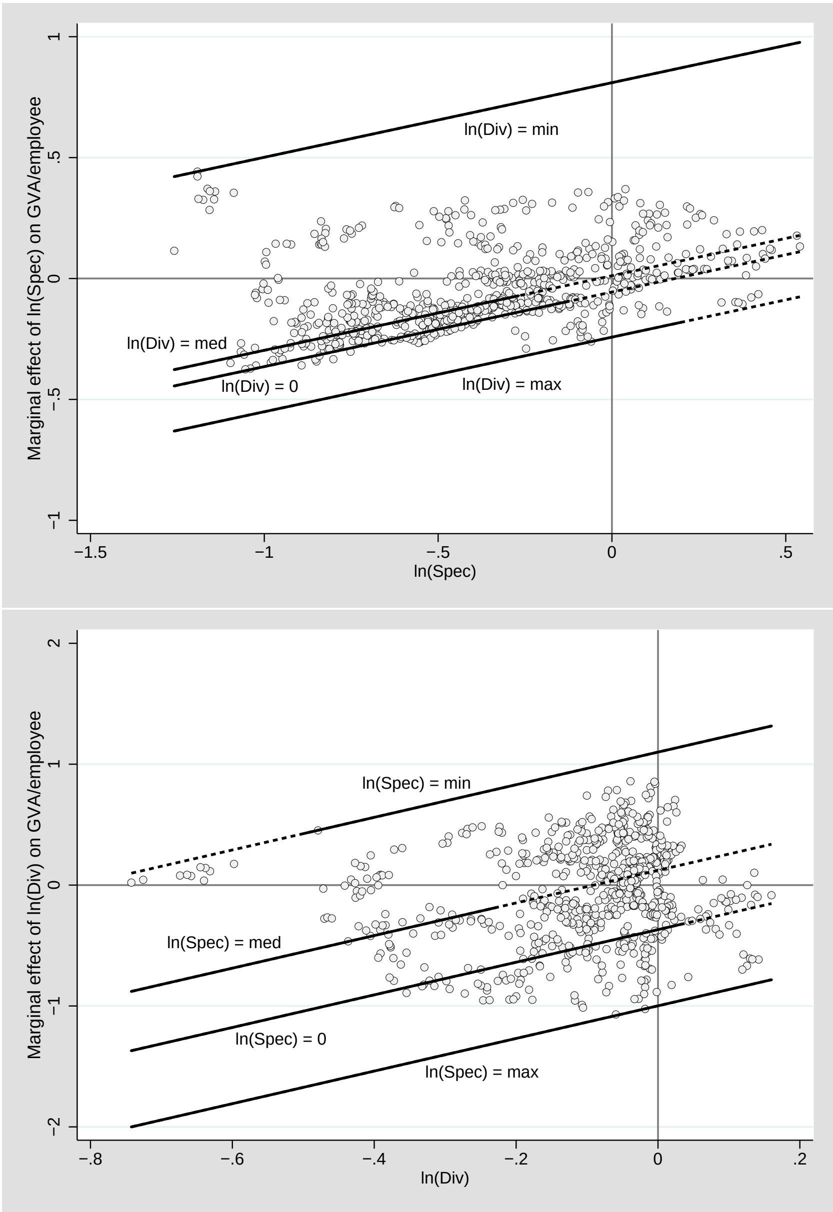
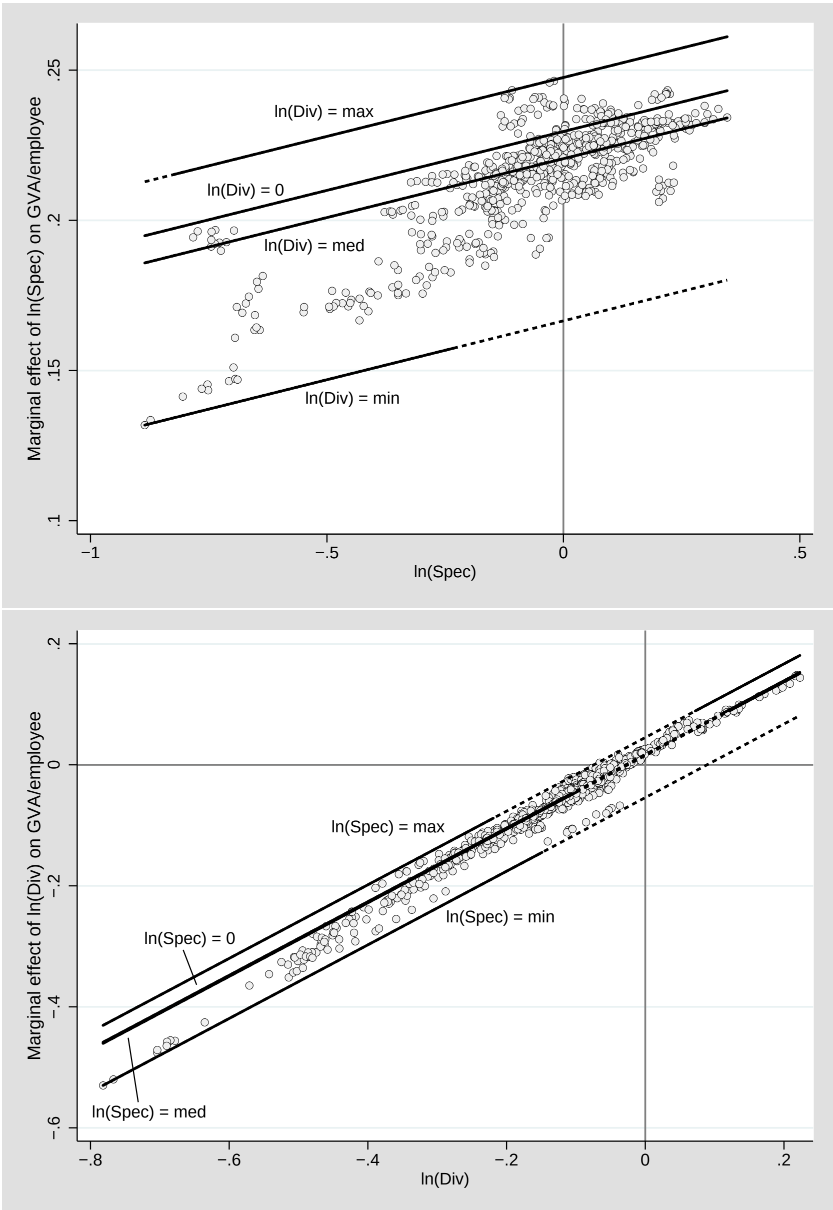


Figure 2.6: Effects of specialization (upper panel) and diversification (lower panel) in basic services



The implication for basic services would be to keep the degree of specialization at a high level. At the same time, there should be a highly diversified environment generating a wide range of demand for companies providing basic services.

2.3.5 Discussion and Endogeneity

One issue we have not discussed yet is endogeneity. Whenever it comes to analyze regional growth, endogeneity is a problem. Physical and human capital as well as commuters can be seen as highly endogenous variables. The capital stock raises the problem of reversed causality. One can argue whether a higher capital intensity rises productivity or a higher GVA per employee induces more investments in physical capital. Human capital, especially high-skilled labor, is very mobile which makes it highly endogenous (see Brunow and Hirte, 2009). Since we are focusing on cities and do not consider labor market regions, commuting may be endogenous, too. Additionally, specialization and diversification could be endogenous. The theoretical considerations of Marshall (1890) or Jacobs (1970) assume that knowledge spillovers foster economic growth. But whenever a market grows, there are rents which can be skimmed by firms. These rents may lead to market entries of new firms, hence the degree of, e.g., specialization could rise.

To handle these endogeneity issues, we have to find viable instruments. Finding such instruments is a hard task at the regional level, since we have to deal with data limitations. The capital stock is the best example. To the best of our knowledge, no official data on the capital stock at the regional level in Germany is available. Therefore, we construct a proxy variable, described in Sect. 2.2.2. In the end, we have to admit that most of the endogeneity problems remain unsolved in this paper. Nevertheless, we carried out some robustness checks and give a first attempt on an IV-estimation. First, we estimate the model by Fahrhauer and Kröll (2012), omitting capital intensity and including the number of employees. Second, we experiment with a reduced model and exclude capital intensity, human capital and commuters from our estimation. The estimation of these two models causes the problem of omitted variable bias. Nevertheless, our results stay fairly the same. Third, we apply a dynamic panel approach. The focus of our paper is not on identifying the timing of agglomeration effects, but rather on the interaction between localization and urbanization economies. As before, the results presented in the former section remain almost unchanged. Fourth, we apply a two-stage least-squares approach (2SLS) and instrument specialization and diversification.

A valid instrumental variable always requires two characteristics: relevance and exogeneity (Wooldridge, 2002). To check the relevance of our instruments and overcome the weak instrument problem (see Baum *et al.*, 2003), we use the Cragg-Donald Wald F statistic (CD) and the critical values from the Stock-Yogo weak ID test from the first stage of our regression. For a lag of up to two years for our specialization and diversification variable, we observe considerable correlations for most cities in our sample. For this reason, we use two lags of our specialization and diversification indicator as instruments. For exogeneity, we present the p-value for the Hansen J-statistic (Hp) to check the overall validity of our instruments.

Table 2.2 shows the results and the above mentioned test statistics of the instrumental variable (IV) approach. Applying a 2SLS approach imposes a problem: IV-estimates with interaction models require the manual estimation of the two stages. To check the statistical exogeneity of our two instruments, we show two columns in Table 2.2 for each sector. The first column presents the test statistics for $\ln(Spec_{z,s,t})$ and the second column for $\ln(Div_{z,s,t})$. We have to mention that not all results remain the same.

For manufacturing, nearly all variables of interest show the same signs as in the OLS-regression. One exception is the balance of migrants which has now a negative sign. However, the effect is insignificant. Turning to localization economies, the positive effect is significant. This confirms our results from the former section. Urbanization economies have no impact. The squared term for specialization is positive and the one for diversification is negative. When applying this IV-estimation, the interaction term in manufacturing remains negative. With these results we can conclude that specialization fosters productivity in the manufacturing sector.

For construction, we observe changes with regard to our variables of interest. First, the coefficient for specialization changes from negative to positive. This is not problematic at all, since the coefficient of the squared term stays positive and is higher in its magnitude. This makes the marginal effects curve of specialization even steeper, so that the threshold for reversed signs moves left on the x-axis. At the same time, this causes the average effect to get positive. The conclusion for specialization in the construction sector remains the same. Second, the coefficient of the squared term for diversification changes its sign. However, in all other specifications this term has a positive and nearly significant coefficient. In the IV-specification, the coefficient is negative but insignificant. One reason can be that the autocorrelations of our diversification index are not high enough, making the instrument not that relevant for this analysis. This causes the variable to send wrong signals concerning urbanization economies. Another explanation is that the effect of urbanization economies is maybe not U-shaped but rather bell-shaped or of higher order. We leave this discussion for further research.

Table 2.2: IV-Regression (2SLS) results

	Manufacturing		Construction		Basic Services		Advanced Services	
Variables:								
ln(Spec)	0.383***		**0.128**		0.224**		-0.253***	
	(4.630)		**(0.830)**		(2.450)		(-3.730)	
ln(Div)	-0.783		-0.235		**-0.389*****		**-0.482*****	
	(-1.530)		(-0.630)		**(-7.810)**		**(-5.100)**	
Interactions:								
$\ln(\text{Spec})^2$	0.100		0.297*		**-0.187**		0.321***	
	(1.250)		(1.860)		**(-1.340)**		(6.760)	
$\ln(\text{Div})^2$	-0.483*		**-0.200**		**-0.582*****		0.109	
	(-2.040)		**(-0.190)**		**(-8.350)**		(1.100)	
ln(Spec)*ln(Div)	-0.132		-0.100		0.262*		**0.147**	
	(-0.380)		(-0.130)		(0.087)		**(0.920)**	
Controls:								
ln(Capital intensity)	0.115		0.554***		**-0.056**		0.183*	
	(0.640)		(8.560)		**(-0.680)**		(2.100)	
ln(Share of high qualified)	-0.273***		0.082		0.151***		0.007	
	(-8.570)		(0.840)		(8.410)		(0.220)	
Balance of commuters	0.000***		-0.000		-0.000		-0.000***	
	(6.850)		(-1.220)		(-0.090)		(-4.620)	
Balance of migrants	**-0.000**		-0.001		0.001*		0.001*	
	(-0.080)		(-1.610)		(2.460)		(2.050)	
Constant	-		4.675***		10.838***		-	
	-		(6.460)		(12.650)		-	
CD	251.236***	58.236***	124.764***	129.292***	115.107***	114.929***	221.984***	82.058***
Hp	0.137	0.901	0.142	0.138	0.905	0.778	0.116	0.595
R^2	0.489		0.253		0.675		0.287	
Obs.	630		630		630		630	
Instrumented	*ln(Spec)*	*ln(Div)*	*ln(Spec)*	*ln(Div)*	*ln(Spec)*	*ln(Div)*	*ln(Spec)*	*ln(Div)*
Instruments	*t-1 to t-2*	*t-1 to t-2*	*t-1 to t-2*	*t-1 to t-2*	*t-1 to t-2*	*t-1 to t-2*	*t-1 to t-2*	*t-1 to t-2*

Note: An overview of the critical values for the CD test statistic is found in Stock and Yogo (2002), *t*-stats in parentheses. The results with reversed signs in comparison to the OLS estimates are printed in boldface. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level or the 10%, 15% and 20% level of maximal IV relative bias (size) for the CD statistic. *Dependent variable:* logarithmic real GVA per employee. *Source:* authors' calculation.

In the sector of advanced services, diversification has now a negative and significant impact on GVA per employee. Nevertheless, this result fits into our findings of the former section, because the slope of the marginal effect of diversification is much smoother. Additionally, we find for most of our cities negative urbanization economies when applying standard OLS-estimates. However, we have to mention that the interaction term is now positive but insignificant. This can be a hint on a possible non-linearity of the interaction between localization and urbanization economies. We leave this discussion as well as the study of endogeneity for future research, because several other robustness checks indicate a negative sign of the interaction term in advanced services.

The most serious changes are observable for basic services. Especially the coefficients for the squared terms of specialization and diversification change their signs. This casts some doubt on the validity of the 2SLS approach. Such a view is supported by the result for capital intensity. The coefficient for physical capital is negative when applying our IV-approach. This seems to be unrealistic, since theory suggests non-negative elasticities. The interaction effect remains positive and significant.

Despite that the Cragg-Donald Wald F statistics show no weak instrument problem in any of our regressions and the p-values for the Hansen J-statistic cannot reject exogeneity of our instruments, more research is necessary on endogeneity issues and agglomeration economies. Since we are probably the first ones who studied interaction effects of localization and urbanization economies, future research has to elaborate more on this point. However, the problem of endogeneity is mitigated to a small extent in our study. By using panel data, time-invariant factors which are correlated with the error term are absorbed within the time fixed-effects.

Another interesting point in this endogeneity debate is the question how to deal with cross-section or spatial dependence and interaction terms. To the best of our knowledge, standard techniques are not able to handle this, which is why we have to estimate the two stages separately. More research is necessary in this special field. Maybe the interaction effect is non-linear too, which makes the mechanism of localization and urbanization economies even more complex and sheds new light on this debate.

2.4 Conclusion

We have performed a panel regression analysis to find evidence of localization and urbanization economies. Our contribution is not only to reassess the old question of Marshall (1890) versus Jacobs (1970), but also to gain insights into the mechanisms by which the two externalities affect each other. We find that there are such interaction effects. The level of specialization has an impact on the strength of urbanization economies and *vice versa*. This interaction is positive in the sector of basic services but negative in manufacturing, construction as well as advanced services. Furthermore, we find that localization and urbanization economies depend crucially on the currently achieved levels of specialization and diversification (i.e., are non-linear). Furthermore, we close a gap in the German literature and evaluate the effects of localization and urbanization economies using GVA per employee.

If an author chooses not to use interaction terms, his or her results only represent the average effect. It is likely that these averages vary for different industrial and regional disaggregations, which may be the reason why the investigations carried out, present very different results. We show that an interaction approach provides more insight into the debate between the proposals of Marshall (1890) and Jacobs (1970).

Basically, we find a positive influence of specialization on GVA per employee (evidence for localization economies) in manufacturing and basic services. Diversification influences GVA per employee to a smaller extent, which is why we tend to reject the existence of urbanization economies in the majority of our sample. We are able to show, however, under which parameter constellations significant localization and urbanization effects can occur.

Regional growth models suffer from endogeneity whenever it comes to an empirical analysis. We therefore give some robustness checks and additionally apply a 2SLS estimation approach. Since our study is probably the first which evaluates interaction effects between localization and urbanization economies, we leave a more extensive elaboration for future research.

Because we find that localization and (in parts) urbanization economies exist, these findings might be helpful for regional policy makers. Our results show that industrial clusters, which are often pursued by decision makers, are favorable for economic growth in the sectors of manufacturing and basic services. However, the recent literature shows that cluster policy should be seen with caution, whenever policy hinders production factors to move freely across regions (see Brakman and van Marrewijk, 2013). Many authors advise against excessive specialization and see economic diversity as stabilizing element (for a survey, see Dissart, 2003). This paper does not intend to engage in this discussion but our findings suggest, that there can be a trade-off between specialization and diversification since the one reduces the advantages of the other. Taking these results seriously, it must be doubted that a combination of both can be a meaningful strategy, at least in the short-term.

What do our results imply for knowledge spillovers? Obviously, the matter is not as clear-cut as suggested by Marshall (1890) and Jacobs (1970). It seems as if specialization alone does not produce (growth enhancing) knowledge flows. The same holds for diversification. Both the relative size of the respective sector and the composition of the rest of the local economy must be considered in order to make meaningful statements. For example, if only one sector exists in a city, there can be considerable knowledge flows leading to economic growth. If, however, two different sectors are added to this economy, this effect can be reduced even if the absolute size of the first sector is left completely untouched. Suddenly, there are different kinds of knowledge present in the city. It is possible that these kinds of knowledge are compatible and create urbanization economies. This might, however, only happen if these sectors gain size cutting back on the first sector. These kinds of trade-offs could also be a subject of further research.

Appendix 2.A

Table A.3: Descriptive statistics – manufacturing

Variable	Mean	s.d.	Minimum	Maximum
GVA per employee (in Euro in real terms)	61,703	19,341	21,953 Gera (2003)	131,857 Ludwigshafen am Rhein (2007)
Spec	0.846	0.469	0.116 Potsdam (2008)	2.588 Wolfsburg (2006)
Div	0.924	0.077	0.692 Bonn (1998)	1.262 Herne (2006)
Capital intensity (in Euro)	118,165	22,008	79,028 Heilbronn (1999)	224,916 Berlin (2005)
Share of high qualified (in %)	11.460	4.309	4.2 Solingen (1998)	25.3 Erlangen (2008)
Balance of commuters	38,019	48,387	-9,556 Oberhausen (2008)	260,188 Frankfurt am Main (2001)
Balance of migrants (per 1.000 inhabitants)	1.005	6.785	-32.4 Cottbus (1999)	42.9 Mainz (2005)
Inhabitants	335,129	456,225	98,802 Trier (2000)	3,424,639 Berlin (2008)

Source: authors' calculation.

Table A.4: Descriptive statistics – advanced services

Variable	Mean	s.d.	Minimum	Maximum
GVA per employee (in Euro in real terms)	72,622	14,032	42,339 Heilbronn (2008)	117,736 Herne (1998)
Spec	1.231	0.359	0.486 Bottrop (2000)	2.739 Frankfurt am Main (1998)
Div	0.866	0.121	0.464 Wolfsburg (1998)	1.226 Bottrop (2002)
Capital intensity (in Euro)	693,161	141,062	422,774 Bottrop (2004)	1,209,478 Koblenz (1998)
Share of high qualified (in %)	11.460	4.309	4.2 Solingen (1998)	25.3 Erlangen (2008)
Balance of commuters	38,019	48,387	-9,556 Oberhausen (2008)	260,188 Frankfurt am Main (2001)
Balance of migrants (per 1.000 inhabitants)	1.005	6.785	-32.4 Cottbus (1999)	42.9 Mainz (2005)
Inhabitants	335,129	456,225	98,802 Trier (2000)	3,424,639 Berlin (2008)

Source: authors' calculation.

Table A.5: Descriptive statistics – construction

Variable	Mean	s.d.	Minimum	Maximum
GVA per employee (in Euro in real terms)	36,958	6,658	17,457 Gera (2000)	83,330 Herne (1999)
Spec	0.719	0.262	0.284 Wolfsburg (2008)	1.717 Herne (2005)
Div	0.912	0.110	0.476 Wolfsburg (1998)	1.174 Bottrop (2005)
Capital intensity (in Euro)	29,049	7,307	10,541 Rostock (1998)	45,112 Munich (1999)
Share of high qualified (in %)	11.460	4.309	4.2 Solingen (1998)	25.3 Erlangen (2008)
Balance of commuters	38,019	48,387	-9,556 Oberhausen (2008)	260,188 Frankfurt am Main (2001)
Balance of migrants (per 1.000 inhabitants)	1.005	6.785	-32.4 Cottbus (1999)	42.9 Mainz (2005)
Inhabitants	335,129	456,225	98,802 Trier (2000)	3,424,639 Berlin (2008)

Source: authors' calculation.

Table A.6: Descriptive statistics – basic services

Variable	Mean	s.d.	Minimum	Maximum
GVA per employee (in Euro in real terms)	37,767	10,304	21,122 Chemnitz (2000)	79,847 Fürth (2008)
Spec	0.971	0.184	0.413 Wolfsburg (1998)	1.414 Bremerhaven (1998)
Div	0.872	0.126	0.458 Wolfsburg (1998)	1.250 Herne (2006)
Capital intensity (in Euro)	62,497	11,723	42,390 Bonn (2002)	111,971 Berlin (2002)
Share of high qualified (in %)	11.460	4.309	4.2 Solingen (1998)	25.3 Erlangen (2008)
Balance of commuters	38,019	48,387	-9,556 Oberhausen (2008)	260,188 Frankfurt am Main (2001)
Balance of migrants (per 1.000 inhabitants)	1.005	6.785	-32.4 Cottbus (1999)	42.9 Mainz (2005)
Inhabitants	335,129	456,225	98,802 Trier (2000)	3,424,639 Berlin (2008)

Source: authors' calculation.

Figure A.7: Scatter plot for ln(Spec) and ln(Div)

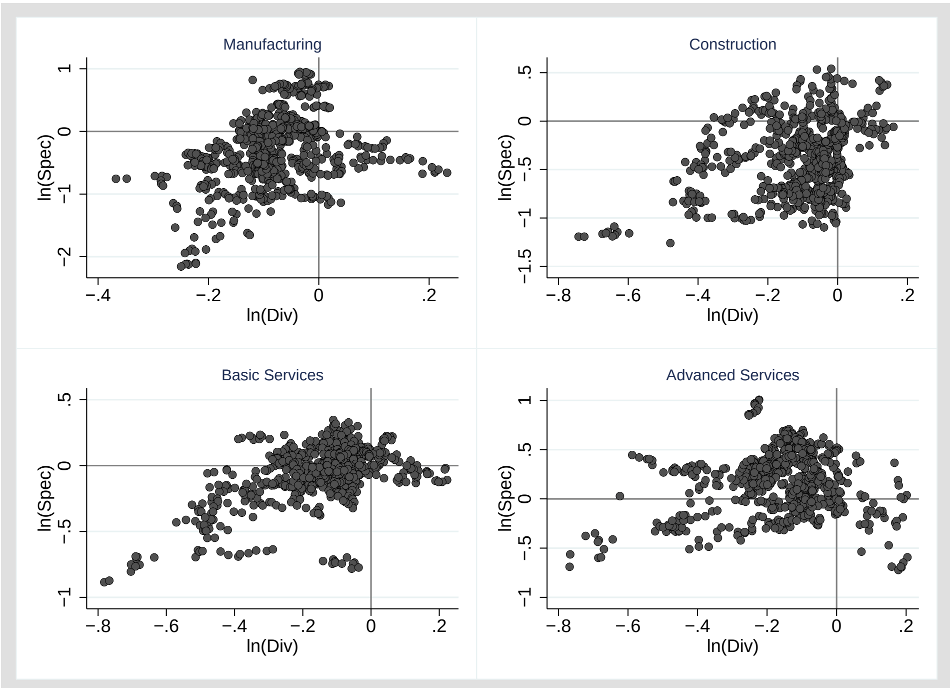
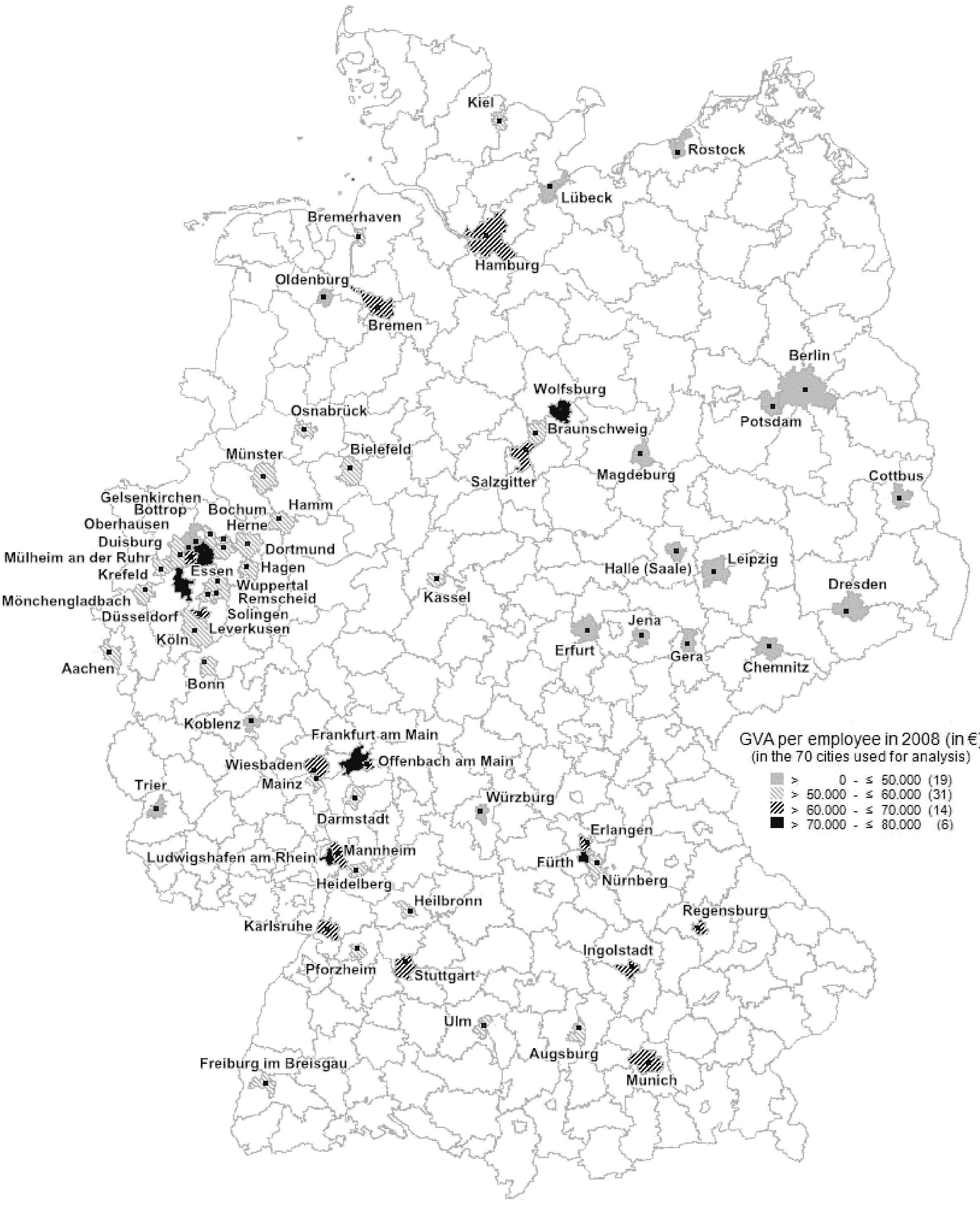


Figure A.8: GVA per employee for the cities in our sample



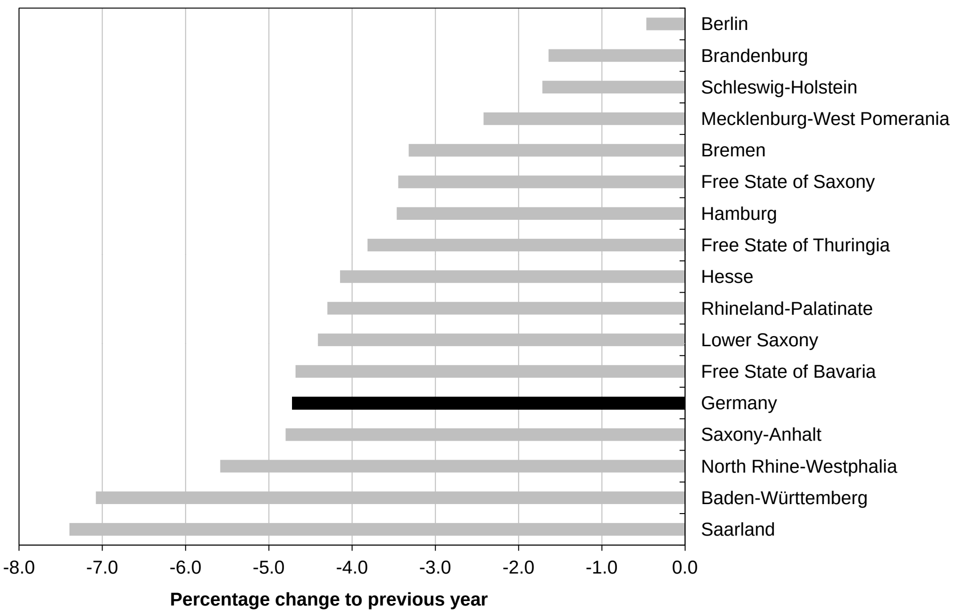
3 Forecasting GDP at the Regional Level with Many Predictors

With kind permission of John Wiley & Sons, this chapter is the reprint of the original article by Lehmann and Wohlrabe (2015), published in the journal *German Economic Review*, May 2015, Volume 16, Issue 2, pp 226-254, © 2013 German Economic Association (Verein für Socialpolitik).

3.1 Motivation

Regional policy-makers are increasingly interested in reliable forecasts of macroeconomic variables (e.g. gross domestic product – GDP) at the regional level. Such forecasts are important to the decision-making process (e.g. for fiscal policy planning). Assuming identical business cycles at the regional and national level, decision-makers can appraise future regional economic output with national forecasts. However, the use of national forecasts can lead to misestimates because of a high degree of regional heterogeneity (e.g. different economic structures). A high heterogeneity among regional units is observable for Germany. The 16 German states are characterized by high disparity in their economic structures. This disparity is explicitly reflected in annual growth rates for real GDP. Figure 3.1 shows the annual growth rates of real GDP in 2009, the year after the economic meltdown. This shock clearly illustrates how (regional) economies with different economic structures are affected by national or supra-national business fluctuations. A more open economy with higher export quotas can grow or shrink faster than an economy that focuses on domestic or regional markets. Whereas the economic output of a highly industrialized and export-dependent German state such as North Rhine-Westphalia shrinks by 5.6% in 2009, the GDP growth rate of Berlin, which is characterized by a large number of different services, lies at -0.5% for the same year. The economic recession of 2009 affected the regional units with different intensities. Obviously, the growth rate of Germany (-4.7%) does not appear to be a good approximation for a decrease in GDP for all sub-national German regions.[1] Regional macroeconomic aggregates are more difficult to forecast in comparison to national ones because of limited data availability and low publication frequency. In general, only annual information about regional GDP is provided by official statistics. For economic policy, it is crucial to know in what phase of the business cycle the whole economy actually is. The cyclical GDP movement, and therefore the knowledge of the current phase about the business cycle, can only be highlighted with quarterly data. More accurate predictions of regional GDP are only possible with such information. This information eventually reduces forecast errors and sends more accurate signals to regional policy-makers.

Figure 3.1: Percentage change in real GDP in 2009 for the German states



Source: Working Group Regional Accounts VGRdL (2011a).

The economic forecasting literature includes many studies on (supra-) national aggregates such as for the Euro Area (see e.g. Bodo *et al.*, 2000; Forni *et al.*, 2003; Carstensen *et al.*, 2011) and Germany (see e.g. Kholodilin and Siliverstovs, 2006; Breitung and Schumacher, 2008; Drechsel and Scheufele, 2012b); however, only a few attempts have been undertaken to predict economic output at the regional level.[2]

[1] Schirwitz *et al.* (2009) show that significant differences between regional business cycles in Germany exist.

[2] In his thesis, Vogt (2009) conducts a comprehensive survey of forecast activities for the German states.

Bandholz and Funke (2003) construct a leading indicator for Hamburg, notably to predict turning points of economic output. Dreger and Kholodilin (2007) use regional indicators to forecast the GDP of Berlin. A study by Kholodilin *et al.* (2008) employs dynamic panel techniques to forecast GDP on an annual basis for all German states at the same time, accounting for spatial effects. The study by Wenzel (2013) also studies the forecasting performance of business survey data for all German states within a panel framework. He found that business survey data are important for the prediction of regional economic growth. In addition, few studies forecast regional labor market indicators for Germany. First, Longhi and Nijkamp (2007) predict employment figures for all West German regions and particularly address the problem of spatial correlation. Second, Schanne *et al.* (2010) forecast unemployment rates for German labor market districts, using a global vector autoregression (GVAR) model with spatial interactions. All these studies employ different data frequencies. Whereas Bandholz and Funke (2003) and Dreger and Kholodilin (2007) use annual GDP information disaggregated into quarterly data, Kholodilin *et al.* (2008), Longhi and Nijkamp (2007) and Wenzel (2013) have only annual information. Schanne *et al.* (2010) instead use data on a monthly basis. To the best of our knowledge, there is only one international study that examines the forecasting performance of regional economic output. Kopoin *et al.* (2013) evaluate whether national and international indicators have information to forecast real GDP at the level of Canadian provinces.

Our study adds to these studies in several ways. First, we overcome the problem of data limitations at the regional level using a new data set with quarterly national accounts for Eastern Germany, the Free State of Saxony[3] and Baden-Württemberg. Altogether, we have 114 regional indicators, including the Ifo business climate for industry and trade in Saxony or new manufacturing orders for Baden-Württemberg. Second, we use regional, national and international indicators, and we assess their forecasting performance at the regional level. Most of the previously mentioned studies have only a few regional indicators and no national or international ones. Finally, our large data set enables us to study the forecasting accuracy of several pooling strategies and factor models. We are likely the first researchers to evaluate the properties of a large set of indicators and corresponding time series approaches at the regional level.

[3] Vogt (2010) studies the properties of a few indicators to forecast Saxon GDP on a quarterly basis. He combines forecasts from different VAR models.

We combine different strands of the economic forecasting literature. In particular, we attempt to determine which indicators are important in forecasting regional GDP. Does early information come from international (World or European Union) or national (Germany) indicators? Alternatively, does sub-national or regional information increase forecasting performance? Trading partners such as the United States and Europe (France, Poland, etc.), as well as the growing importance of Asian economies, create a stronger linkage between these countries and regional economies. These are two of several reasons why we include international indicators. Furthermore, shocks that hit the German economy are transmitted through different channels (e.g. the production of intermediate goods) to regional companies. Banerjee *et al.* (2005) construct a large data set containing leading indicators to forecast euro-area inflation and GDP growth and add comprehensive information from the US economy, and they find that a set of these variables improves forecasting performance. Banerjee *et al.* (2006) analyze the importance of euro-area indicators for the prediction of macroeconomic variables for five new Member States. Several studies analyze forecasting properties in a data-rich environment for different countries. Schumacher (2010) finds that international indicators do not deliver early information for forecasting German GDP if the data are not preselected. Otherwise, forecasting performance improves with international information. For the small and open economy of New Zealand, Eickmeier and Ng (2011) find that adding international data to nationwide information enhances the quality of economic forecasts. To improve forecasts of Canadian macroeconomic data (e.g. GDP and inflation), Brisson *et al.* (2003) use indicators from the United States and other countries. In our study, we use international and German indicators, as well as several variables from the sub-national (Eastern Germany) and regional levels (Saxony, Baden-Württemberg). To the best of our knowledge, our study is the first to evaluate this question from a regional perspective.

Furthermore, we add to the existing literature on forecast combinations. Since the seminal work by Bates and Granger (1969), it is known that combining forecast outputs from different models can lead to improved forecast accuracy in comparison to univariate benchmarks or predictions from a single model.[4] Several empirical contributions exist for different single countries (see e.g. Drechsel and Scheufele (2012a) and Drechsel and Scheufele (2012b) for Germany or Clements and Galvão (2009) for the United States) or for several states simultaneously (see e.g. Kuzin *et al.*, 2013; Stock and Watson, 2004). Studies at the regional level are absent. Given our large data set, we evaluate the fore-

[4] For recent surveys, see Timmermann (2006) and Stock and Watson (2006).

cast accuracy of different pooling strategies. Finally, our article studies the forecasting performance of several factor models. This class of models proved to enhance forecast accuracy at the national level (see e.g. Breitung and Schumacher (2008), Schumacher (2007) and Schumacher (2010) for Germany, or Stock and Watson (2002) for the United States). To the best of our knowledge, regional studies are missing.

The study is organized as follows. In section 3.2, we describe our data and empirical setup. The results are discussed in section 3.3. Section 3.4 offers a conclusion.

3.2 Data and Empirical Setup

3.2.1 Data

Gross Domestic Product at the Regional Level

The official statistics in Germany do not provide temporal disaggregated macroeconomic data (e.g. quarterly GDP) for regional units. Only annual information is available. Therefore, it is either problematic to find a suitable target variable to forecast or the number of observations is insufficient. In our study, we use a new data set that solves these two problems of availability and length of the time series.

To the best of our knowledge, three different sources currently exist that provide publicly available quarterly national accounts at the German regional or sub-national level. First, Nierhaus (2007) computes quarterly GDP for the German state Free State of Saxony. He applies the temporal disaggregation method of Chow and Lin (1971), which is also used for official statistics of the European Union. The method is based on a stable regression relationship between annual aggregates and indicators with a higher frequency (e.g. monthly). This relationship makes it possible to transform annual into quarterly data. For this transformation, Nierhaus (2007) uses official German statistics: regional turnovers for Saxony or quarterly data from national accounts for Germany (e.g. gross value added). Second, Vullhorst (2008) uses – like Nierhaus (2007) – the temporal disaggregation method of Chow and Lin (1971) to calculate quarterly national accounts for the state of Baden-Württemberg. For the temporal disaggregation of annual GDP for Baden-Württemberg, nearly the same indicators are used as for Saxony (e.g. regional turnovers for the manufacturing sector in Baden-Württemberg or quarterly gross value added from national accounts for Germany). Third, the Halle Institute for Economic Research (IWH) provides quarterly data on GDP for Eastern Germany (excluding Berlin). The quarterly data for Eastern Germany are not calculated with the method of Chow and Lin (1971), but with a so-called extrapolation method (see Brautzsch and Ludwig, 2002). Instead of using a stable regression relationship between the annual aggregate and an indicator, the extrapolation method applies quarterly shares in the annual aggregate.[5] The two methods (Chow-Lin and extrapolation) have in common that they use high-frequency indicators. If no regional indicators are available, the IWH also applies quarterly data from national accounts for Germany.

As one would suggest, regional indicators that are used for temporal disaggregation must perform well for forecasting regional GDP. To avoid such a bias, we do not consider such indicators for our analysis. These indicators are the following: turnovers in the manufacturing sector (Saxony and Eastern Germany), working hours (Eastern Germany) and turnovers in the construction sector (Saxony and Baden-Württemberg), as well as for the Saxon retail sale and wholesale trade.

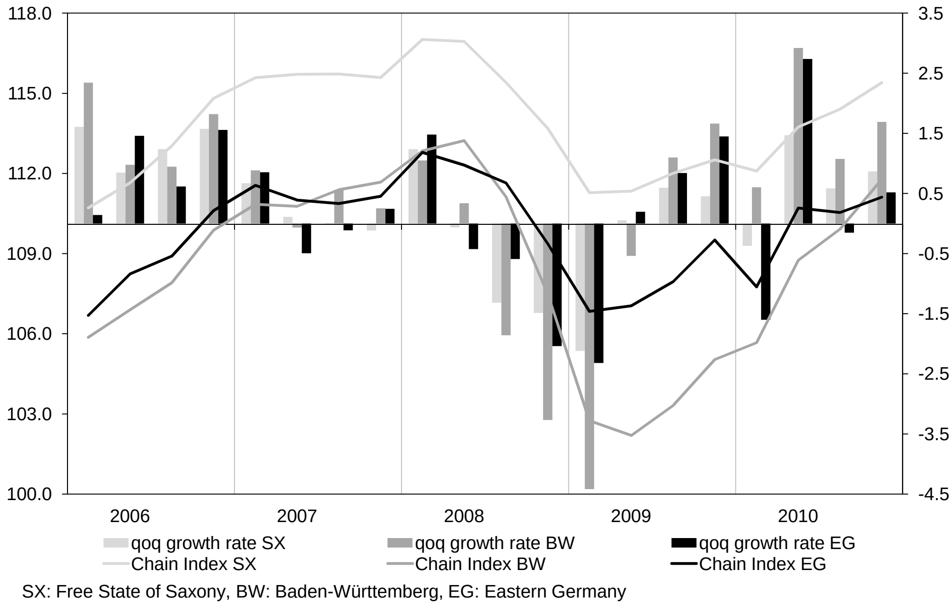
For all three GDP target variables, the time series are available for the period 1996:01 to 2010:04.[6] The data are provided in real terms, and we make a seasonal adjustment to calculate quarter-on-quarter (qoq) growth rates. Figure 3.2 shows the Chain Index, as well as qoq growth rates for the Saxon, Baden-Württemberg and Eastern German GDP from 2006:01 to 2010:04.

During that period, the movements of the two curves for the chain indices for Saxony and Eastern Germany are predominantly identical. Only the levels of qoq growth rates differ slightly for different points in time. The movement of the GDP for Baden-Württemberg is similar, but much more volatile than the output for Saxony and Eastern Germany.

[5] The extrapolation method becomes clearer using the example of manufacturing. Given that x% of all turnovers in the Eastern German manufacturing sector, which is the indicator used by the IWH for manufacturing, are gained in the first quarter of a given year, it is assumed that also x% of total gross value added in the manufacturing sector in that year is produced in the first quarter. Thus, the development of total gross value added in the manufacturing sector is identical to the development of total turnovers.

[6] The data are updated intermittently by the institutions. Quarterly national accounts for Saxony are available under *dresden@ifo.de*. The data are currently not available on the homepage of the Ifo Institute because they will be revised due to a change in the classification of economic activities in Germany. The data for Baden-Württemberg are available upon request from the regional Statistical Office of Baden-Württemberg under *vgr@stala.bwl.de*. For Eastern Germany, quarterly data can be downloaded from the homepage of the IWH (*http://www.iwh-halle.de/c/start/prognose/baro.asp*).

Figure 3.2: Real GDP for Saxony, Baden-Württemberg and Eastern Germany



Note: Chain Index 2010=100 (left sale), quarter-on-quarter growth rate (right scale, in %), seasonally adjusted with Census X-12-ARIMA. *Source:* Ifo Instiute, Statistical Office of Baden-Württemberg and IWH.

Set of Indicators

Our data set contains 361 indicators that can be used to assess their forecasting performance for our target variables. All indicators are from different sources and are grouped into seven different categories: macroeconomic variables (94), finance (31), prices (12), wages (4), surveys (74), international (32) and regional (114).[7] Macroeconomic variables contain industrial production measures, turnovers, new orders and employment figures, as well as data on foreign trade and government tax revenues. All these macroeconomic indicators are measured at the national level (here: Germany). The category of financial variables includes data on interest rates, government bond yields, exchange rates and stock indices. Furthermore, we have data on consumer and producer prices, as well as price indices for exports and imports. In addition to these quantitative data, we use qualitative information. Indicators from the category surveys are obtained from consumer, business and expert surveys (Ifo, ZEW, GfK and the European Commission). In addition, composite leading indicators for Germany (e.g. from the OECD) and the Early Bird index of the Commerzbank are grouped in this category. International data

[7] Under *http://www.cesifo-group.de/DocDL/Appendix_Lehmann_ Wohlrabe_ GEER_ 2013.pdf* the Appendix with a complete description of our data is provided.

cover a set of indicators for the European Union and the United States from the previously mentioned categories, for example, the Economic Sentiment Indicator for France and US industrial production. Finally, we add different regional indicators for Eastern Germany, the Free State of Saxony and Baden-Württemberg. The regional category covers quantitative (turnovers, prices and data on foreign trade) and qualitative information (Ifo and the business survey of the IWH). To avoid biased forecasts, we excluded potential regional indicators from our analysis that are used for temporal GDP disaggregation. In addition, we do not consider sectoral quarterly gross value added for Germany because this indicator, as mentioned in the previous section, is also used for temporal disaggregation.

The data set is predominantly the same one used by Drechsel and Scheufele (2012a), and we add regional indicators for Eastern Germany, the Free State of Saxony and Baden-Württemberg (38 indicators for every single region). Most of these indicators are available on a monthly basis. Hence, a transformation into quarterly data is necessary. First, we seasonally adjust the monthly indicators.[8] Second, we calculate a three-month average for each quarter. If necessary, we transform our data to obtain stationary time series. The external provided Appendix also contains information about the transformation of all indicators.

[8] We apply the Census X-12-ARIMA seasonal adjustment approach.

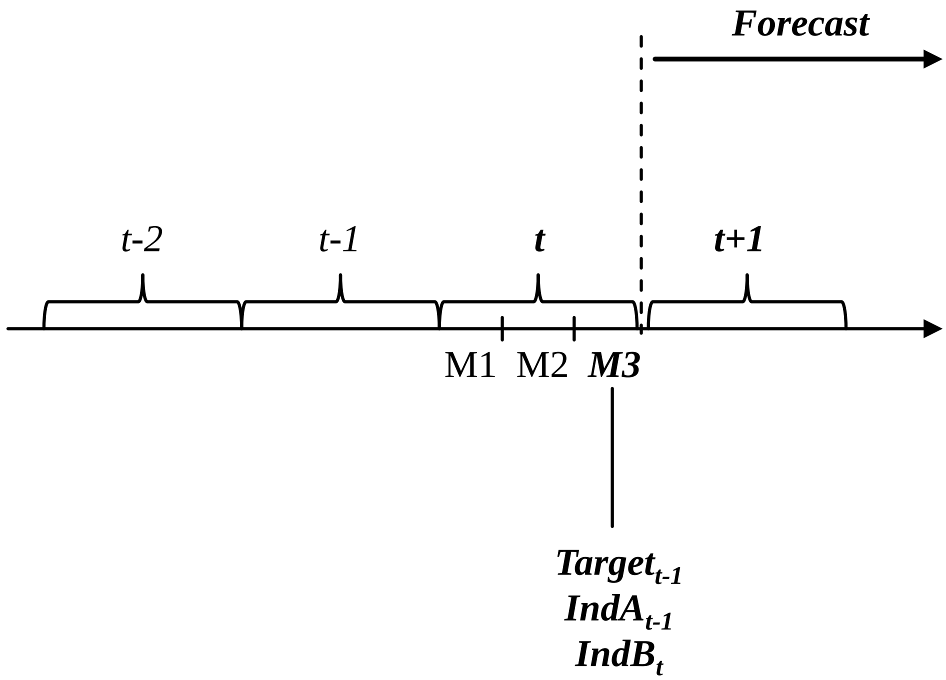
Publication Lags and real-time Aspect

Because official statistics have a substantial publication delay, we must account for this fact in our forecasting exercise. Hard indicators such as turnovers normally have a publication lag of several months. The same holds for regional GDP, which is also calculated with a substantial time lag. In contrast, soft indicators (e.g. survey results) are available immediately. The downloadable Appendix contains information about the publication lag (months) of each indicator and target variable.[9] Whereas real GDP for Saxony and Eastern Germany is available almost three months after the last month of the elapsed quarter, GDP for Baden-Württemberg has a publication lag of two months. The reason for this discrepancy is the fact that the data are available earlier for the Statistical Offices and need not be requested by the two research institutes. We presume that these lags are constant over time and have not changed since the first time the data were released.

Most of the macroeconomic indicators for Germany are available one and a half months later. The majority of financial variables are published with no lag. Nearly all survey-based or soft indicators have no publication lag and can be downloaded immediately at the end of each month. Regional indicators have some special characteristics in comparison to national or international data. Whereas the indicators from survey results have no publication delay, macroeconomic indicators are not available until two and a half months after the end of the quarter of interest. In particular, this circumstance must be considered when forecasting regional GDP. The timeline in Figure 3.3 shows exemplarily our forecasting approach for short-term forecasts (one quarter ahead). In this figure, t stands for the current quarter. $M1, M2$ and $M3$ denote the respective months of that quarter. We hold M3 in bold characters to symbolize that every forecast round is made at the last month of each quarter; for example, the forecast for the first quarter 2010 is calculated at the end of December 2009. With this assumption, we only have to distinguish between three publication lags. First, for our three GDP variables ($Target_{t-1}$), information is only available until the last quarter; thus, *Target* is indexed by *t-1*. Second, the set of indicators that have a publication lag is labeled by $IndA_{t-1}$; we use only the information with a time lag of one quarter. Finally, all remaining indicators with no publication delay are denoted by $IndB_t$. Therefore, our forecasting approach uses only information that is available at the point when a forecast is made.

[9] The time lag varies between 0 and 2.5 months. For each indicator with a publication lag, we assume a time lag of one quarter.

Figure 3.3: Timeline for short-term forecasts



When dealing with publication lags, we have to mention the real-time aspect of this analysis. Concerning our target variables, we are only able to model publication lags but no continuous data revisions. The reason is straightforward. Quarterly national accounts for Saxony were not available before 2007. Nierhaus (2007) first calculated quarterly real GDP for Saxony at the end of 2007 and provided the whole series from 1996 onwards. Thus, we are not able to observe substantial revisions of previous years. The same holds for Baden-Württemberg and Eastern Germany. Finally, for a consistent real-time analysis, the real-time data flow for all indicators would be necessary and preferable. Unfortunately, for such a large data set, such a data flow is currently unavailable. Thus, we refer to our analysis as "pseudo-real-time". How we implement the previously mentioned publication lags is described in the next section together with our empirical model.

3.2.2 Indicator Forecasts

To generate multiple-step ahead forecasts, we use the following autoregressive distributed lag (ADL) model,

$$y_{t+h}^{k} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=m}^{q} \gamma_j x_{t+1-j}^{k} + \varepsilon_t^{k} \ , \tag{3.1}$$

where y_{t+h}^{k} stands for the h-step ahead model k of the qoq growth rate of the Saxon, Baden-Württemberg or Eastern German real GDP and x_t^k denotes the exogenous indicator from the regional, national or international level. Because we use quarterly data, a maximum of four lags is allowed for both the lagged dependent and independent variables. The optimal lengths for p and q are determined by the Bayesian Information Criterion (BIC). To consider the availability of our indicators, m is introduced. The variable m takes a value of 1, whenever no publication delay exists. If a variable is not available immediately, m takes a value of 2.

We apply a recursive forecasting approach with a rolling estimation window. The initial estimation period ranging from 1996:01 to 2002:04 ($T = 28$) is moved forward successively by one quarter. In every step, the forecasting model of Equation (3.1) is newly specified. For each forecast horizon, the first forecast is calculated for 2003:01 and the last for 2010:04. Our forecast horizon h has four dimensions: $h \in \{1, 2, 3, 4\}$. Because we implement the ADL model as a direct-step forecast, we always produce $N = 32$ forecasts for $h = 1$ (short-term) or $h = 4$ (long-term) and every model k. As the benchmark, we choose the standard AR(p) process.[10]

There may be an information gain from applying a multi-indicator forecast model. Hence, combining regional with national indicators may reduce forecast errors due to a combination of different information sets; thus, we modify the model in Equation (3.1) by adding another indicator,

$$y_{t+h}^{k} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=m}^{q} \gamma_j r_{t+1-j}^{k} + \sum_{l=m}^{q} \eta_l z_{t+1-l}^{k} + \varepsilon_t^{k} \ . \tag{3.2}$$

We only estimate models for every regional indicator (r_t^k) in combination with an indicator from the national level (z_t^k).[11] Therefore, we have $38 \cdot 118 = 4,484$ extra models for all three regional units.

---

[10] We also tested the AR(1) process, the random-walk and an in-sample-mean forecast and found similar results.

[11] Because of computational reasons, we restrict the multi-indicator forecast approach to 118 national indicators, which include industrial production, new orders, new registrations of vehicles, exports, imports and surveys. All these indicators are labeled with an **X** in the Table in the downloadable Appendix (column Multi).

3.2.3 Combination Strategies

It is well known that an appropriate in-sample fitted model could have a bad out-of-sample performance, thus producing high forecast errors. Stock and Watson (2006) and Timmermann (2006) have shown the advantage of combining forecasting output from different models. This advantage has been confirmed in numerous empirical studies for different countries (see, e.g., Drechsel and Maurin, 2011; Eickmeier and Ziegler, 2008). Evidence for the advantage of pooling at the regional level is absent. With our study, we fill this gap.

A forecast obtained by pooling $\widehat{y}_{t+h}^{Pool}$ is based on the individual indicator forecasts $\widehat{y}_{t+h}^{k}$ and a weighting scheme w_{t+h}^{k}:

$$\widehat{y}_{t+h}^{Pool} = \sum_{k=1}^{K} w_{t+h}^{k} \widehat{y}_{t+h}^{k} \quad \text{with} \sum_{k=1}^{K} w_{t+h}^{k} = 1\,. \tag{3.3}$$

Because the weights are indexed by time, they are varying with every re-estimation of our ADL model and every forecasting horizon h. K represents the number of models we consider for pooling.

A very simple but empirically well-working scheme (see, e.g., Timmermann, 2006) is (i) equal weights: $w^k = 1/K$. The weights are not time varying and depend only on the number of included individual forecasting models K. In addition to a simple mean, we consider (ii) a median approach. This weighting scheme is time varying and more robust against outliers. In addition to these simple approaches, we can calculate different weights from two categories: in-sample and out-of-sample. We follow the studies by Drechsel and Scheufele (2012a) and Drechsel and Scheufele (2012b) and apply in-sample and out-of-sample weighting schemes. We use two in-sample measures for the calculation of our weights: (iii) BIC and (iv) R^2. The two schemes differ only slightly. Whereas the model with the lowest BIC gets the highest weight, the weight of a single model increases with higher R^2. The weights from these two schemes are time varying and have the following form:

$$w_{t+h}^{k,BIC} = \frac{\exp\left(-0.5 \cdot \Delta_k^{BIC}\right)}{\sum_{k=1}^{K} \exp\left(-0.5 \cdot \Delta_k^{BIC}\right)} \tag{3.4}$$

$$w_{t+h}^{k,R^2} = \frac{\exp\left(-0.5 \cdot \Delta_k^{R^2}\right)}{\sum_{k=1}^{K} \exp\left(-0.5 \cdot \Delta_k^{R^2}\right)}\,, \tag{3.5}$$

with $\Delta_k^{BIC} = BIC_{t+h}^k - BIC_{t+h,min}$ and $\Delta_k^{R^2} = R^2_{t+h,max} - R^2_{t+h,k}$.

When applying out-of-sample weights, it is appropriate to use the forecast errors of different models. First, we apply a (v) trimmed mean.[12] This weighting scheme filters indicators with bad performance and does not consider the forecasts of those models. Consistent with the literature, we use three different thresholds: 25%, 50% and 75% of all indicators in ranked order. If an indicator's performance lies within the worst (25%, 50% and 75%) performers, the outcome of that specific forecasting model is not considered for pooling. All other forecasts are combined with equal weights. Second, discounted mean squared forecast errors are used as weights (vi) to combine several model outcomes. This approach is based on Diebold and Pauly (1987) and is applied, for example, by Costantini and Pappalardo (2010) and Stock and Watson (2004). The weights from this approach have the following form:

$$w_{t+h}^k = \frac{\lambda_{t+h,k}^{-1}}{\sum_{k=1}^{K} \lambda_{t+h,k}^{-1}}\,. \tag{3.6}$$

$\lambda_{t+h,k} = \sum_{n=1}^{N} \delta^{t-h-n} \left(FE_{t+h,n}^k\right)^2$ represents the sum of discounted (δ) forecast errors of the single-indicator model k. The literature finds no consensus on how the discount rate δ should be chosen. We experimented with different values for δ, which show similar performances. In our setup, we use $\delta = 0.1$.

In this study, we will only combine forecasts that are calculated from regional indicators (either for Saxony, Baden-Württemberg or Eastern Germany) or the full sample excluding the other two regional units.[13]

[12] For the effectiveness of this approach, see, e.g. Drechsel and Scheufele (2012b) or Timmermann (2006).

[13] For example, for the Free State of Saxony, we use only the indicators for Saxony or all indicators excluding those from Eastern Germany and Baden-Württemberg.

3.2.4 Factor Models

When dealing with large data sets – where the cross-section dimension is large – standard econometric methodologies are not able to handle all available information. Next to the combination of forecast results (pooling), static and dynamic factor models yield good forecasting results (see Forni *et al.*, 2005; Marcellino *et al.*, 2003; Stock and Watson, 2002). The idea behind these models is to extract or summarize the inherent information of a large set of time series within some common factors. This approach allows us to specify a parsimonious model and thereby alleviate the uncertainty about parameter estimates (see Giannone *et al.*, 2008), which would be the case when estimating a model with nearly all available indicators. In this study, we apply three different approaches for estimating the common factors of the underlying series. To save space, we refer to the literature for further details on each approach. First, we use the standard principal components (PC) method to estimate the factors. Following Giannone *et al.* (2008), we apply the two-step estimator proposed by Doz *et al.* (2011). This two-step estimation procedure, which uses principal components and Kalman filtering (PCKF), has proven to provide some efficiency improvements in comparison with standard principal component methods. As a third approach, we estimate the common factors via quasi-maximum likelihood (QML) (see Doz *et al.*, 2012).[14]

In the next step, we must decide how many common factors shall be extracted from the data. We choose between one and three common factors. In addition, a decision must be made regarding which data source (cross-section and time dimension) should be used to estimate the factors. We have the choice of using either the full sample of indicators (FS) or only the information from regional ones (S, BW or EG). Furthermore, we can extract the factors from (i) monthly data and then aggregate these factors to quarterly information (M), or we aggregate the monthly indicators and then extract the factors from (ii) quarterly data (Q). In the end, we can use the extracted factors in two ways to generate forecasts for real GDP. First, we put the factors directly into the ADL model from Equation (3.1), such that lagged values from the dependent variable and the common factors are used to forecast real GDP. Second, we apply a standard OLS-estimate, where GDP is explained via a constant and the common factors available at time t (see Giannone *et al.*, 2008). The second method considers neither lagged values nor the dependent variable. To sum up, we test three different approaches with up to three common factors. We have two underlying databases from which the factors are

[14] We abstract from the ragged edge data problem (see Wallis, 1986) by extracting factors using only information up to $t-1$.

extracted, as well as two frequencies and forecasting approaches, which results in 72 factor models for each regional unit.

3.2.5 Forecast Evaluation

To analyze the forecast accuracy of different strategies (indicator models, factor models or pooling techniques), we first calculate forecast errors from our forecasting exercise. Let $\widehat{y}^k_{t+h}$ denote the h-step ahead forecast of model k; then, the resulting forecast error is as follows: $FE^k_{t+h} = y^k_{t+h} - \widehat{y}^k_{t+h}$. The forecast error for the AR(p) benchmark is FE^{AR}_{t+h}. In a second step, we use the root mean squared forecast error (RMSFE) as a loss function to assess the overall performance of a model. The RMSFE for the h-step ahead forecast is defined as:

$$RMSFE^k_h = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(FE^k_{t+h,n}\right)^2}\,. \tag{3.7}$$

The respective RMSFE for the autoregressive benchmark is $RMSFE^{AR}_h$. Finally, we construct a relative RMSFE (rRMSFE),

$$rRMSFE^k_h = \frac{RMSFE^k_h}{RMSFE^{AR}_h}\,, \tag{3.8}$$

to decide whether a model k is performing better or worse in comparison with the AR benchmark model. If this ratio is less than 1, the indicator model leads to smaller forecast errors for the respective horizon h. Otherwise, the simple autoregressive model is preferable. Because we have a large set of competing models, pairwise testing would result in the problem of data snooping. This problem means that pairwise tests signal a higher accuracy of one model just by chance.[15] To overcome this problem, we apply the superior predictive ability (SPA) test proposed by Hansen (2005). This test is based on the seminal paper by White (2000). The idea of the SPA test is to examine whether a benchmark model performs better in comparison with a whole set of competitors. Under the null hypothesis, no competing model should beat the benchmark model. Because the SPA test is a multiple test, the null hypothesis is formulated as follows,

[15] Imagine a set of repeated draws from a normal distribution. In some cases, this fact would result in values that lie near the critical values, whereby the null is rejected.

$$H_0 : E\left(d_{k,t+h}\right) \leq 0 \quad k = 1, ..., K\,. \tag{3.9}$$

The difference $d_{k,t+h}$ is defined as $d_{k,t+h} = \left(FE^0_{t+h}\right)^2 - \left(FE^k_{t+h}\right)^2$, whereas FE^0_{t+h} is the forecast error of the benchmark. Whenever the null is rejected, at least one competitor performs better than the chosen benchmark model. Every single-indicator, forecast combination approach and factor model serves as the benchmark. Thus, the corresponding benchmark errors $\left(FE^0_{t+h}\right)^2$ are used. However, because the expectations under the null are unknown, they can be estimated consistently by the sample mean $\overline{d}_{k,t+h}\ \forall i \in \{1, \ldots, k\}$. The original reality check test statistic was proposed by White (2000), but suffers from the inclusion of poor or irrelevant models. Thus, we use the modification proposed by Hansen (2005), which is stable against irrelevant or poor competitors. The corresponding p-values are calculated via bootstrap because the distribution under the null is not identified. The test by Hansen (2005) requires a rolling window approach. With the SPA test, we can decide whether at least one model outperforms the benchmark. However, we are not able to say that these models are the best ones (with some specific confidence). To find the best models, we apply the model confidence set (MCS) procedure proposed by Hansen *et al.* (2011). This procedure is closely related to the SPA test; however, we do not have to specify a benchmark model. The MCS procedure is a model selection algorithm, which filters a set of models from a given entirety of models. The resulting set contains the best models with a given confidence level (see Hansen *et al.*, 2011). Because we have a large set of indicators and therefore a large set of models, we can apply this procedure to find a set of superior models. The null hypothesis is defined as,

$$H^h_{0,M} : \mu^h_{ij} = 0 \quad \forall i, j \in M^h\,, \tag{3.10}$$

whereas $\mu^h_{ij} \equiv E(d^h_{ij,t}) \equiv E(RMSFE^h_{i,t} - RMSFE^h_{j,t})$ denotes the expected difference in the root mean squared forecast errors of models i and j $(i, j \subset k)$ for a given forecast horizon. The procedure tries to find the best set $M^{*,h}$ ($M^{*,h} \equiv \{i \in M^{0,h} : \mu^h_{ij} \leq 0\,\forall j \in M^{0,h}\}$), containing all models that are significantly superior to other models from a starting set $M^{0,h}$ (see Hansen *et al.*, 2011). Because our data set allows us to evaluate a large number of competing models with the MCS procedure, we must restrict the algorithm to a limited starting set.[16] The reason is that this procedure is computational

[16] If we would not restrict our starting set, then the MCS procedure must consider 4,862 different models. Among them, we have 4 benchmarks, 16 combination and 72 factor models, 286 single-indicator and 4,484 multi-indicator models.

very demanding.[17] Thus, our starting set $M^{0,h}$ always contains the best 250 models (from every category) in terms of RMSFE.

3.3 Results

This section presents the results for our three target variables. First, we discuss the general results of our forecasting exercise. Second, we present detailed and selected results that are consistent with the specific economic structures of our regional units.

3.3.1 General Results

The summary tables are divided into four quadrants, each representing one single forecast horizon (h). In the upper (lower) left, $h = 1$ ($h = 3$) is shown and the upper (lower) right presents $h = 2$ ($h = 4$). To obtain an impression about how well the several models are performing, we add the RMSFE of the autoregressive benchmark model (in %) for each forecast horizon and region. Every quadrant shows the top 20 models from our forecasting exercise due to the rRMSFE of Equation (3.8). These rRMSFE are presented in the column *Ratio*. The column *SPA p-value* shows the p-values from the test proposed by Hansen (2005). An **X** in column *MCS* indicates whether a model is included in the set of best models, based on the test by Hansen *et al.* (2011). To increase readability, we add one column with abbreviations for the different forecast models. National indicators are denoted with (N), whereas (I) represents international and (R) regional indicators. Combination strategies are denoted with (C). (M) stands for multi-indicator and (F) for factor models. Tables 3.1, 3.2 and 3.3 present the estimation results for our three regional units.

[17] For both tests (Hansen, 2005; Hansen *et al.*, 2011), we employ a block bootstrap approach with a block size of 12 and 2,500 replications.

Table 3.1: Results for the Free State of Saxony

Target variable: quarter-on-quarter growth rate GDP Free State of Saxony									
h=1		RMSFE AR(p): 0.993%			**h=2**		RMSFE AR(p): 0.992%		
Model	Abbreviation	Ratio	SPA p-value	MCS	Model	Abbreviation	Ratio	SPA p-value	MCS
MSFE weighted (FS)	(C)	0.582	1.000	X	MSFE weighted (FS)	(C)	0.620	1.000	X
IFOOOHCONSAX – GFKESE	(M)	0.730	0.439	X	MSFE weighted (S)	(C)	0.740	0.156	X
IFOOOHCONSAX – GFKCCC	(M)	0.737	0.448	X	PCWHSAX – GFKUE	(M)	0.773	0.002	
Trimmed 25 (FS)	(C)	0.740	0.000		PCWHSAX – EUCSUE	(M)	0.773	0.001	X
IFOOOHCONSAX – IFOEXEMAN	(M)	0.745	0.178	X	PCWHSAX – IFOBSCONNDUR	(M)	0.798	0.059	X
MSFE weighted (S)	(C)	0.754	0.018		PCWHSAX – GFKESE	(M)	0.800	0.025	
IFOOOHCONSAX – GFKCCIN	(M)	0.758	0.330		PCWHSAX – IFOBECONNDUR	(M)	0.803	0.008	
IFOOOHCONSAX – EUCSCCI	(M)	0.758	0.317	X	PCWHSAX – IFOBCCONNDUR	(M)	0.808	0.082	X
IFOBEMAN	(N)	0.763	0.002		PCWHSAX – GFKBCE	(M)	0.809	0.017	
IFOOOHCONSAX – IFOUNFWCON	(M)	0.764	0.348	X	Trimmed 25 (FS)	(C)	0.817	0.002	
IFOOOHCONSAX – GFKUE	(M)	0.766	0.304		PCWHSAX – GFKCCC	(M)	0.818	0.019	
IFOOOHCONSAX – EUCSUE	(M)	0.766	0.327		PCWHSAX – EUBSSPEIND	(M)	0.820	0.001	
IFOBECAP	(N)	0.766	0.092		PCWHSAX – EUBSRTCI	(M)	0.825	0.024	
IFOOOHCONSAX – EUBSPEIND	(M)	0.780	0.221		PCWHSAX – EUBSEMPEIND	(M)	0.825	0.004	
IFOBCMAN	(N)	0.785	0.038		PCWHSAX – IFOBEINT	(M)	0.830	0.022	

Results for the Free State of Saxony – continued

IFOOOHCONSAX – IFOBECONNDUR	(M)	0.792	0.043		HCNOSAX – IFOBEINT	(M)	0.832	0.090	
IFOOOHCONSAX – GFKIE	(M)	0.793	0.077		PCWHSAX – GFKCCIN	(M)	0.834	0.014	
IFOBCIT	(N)	0.793	0.022		PCWHSAX – EUCSCCI	(M)	0.834	0.013	
IFOOOHCONSAX – IFOBCCAP	(M)	0.794	0.121		CONFIRMSAX – EUBSSPEIND	(M)	0.834	0.000	
Trimmed 25 (S)	(C)	0.797	0.007		CONFIRMSAX – IFOBEINT	(M)	0.835	0.000	

h=3		RMSFE AR(p): 0.946%			**h=4**		RMSFE AR(p): 1.021%		
Model	Abbreviation	Ratio	SPA p-value	MCS	Model	Abbreviation	Ratio	SPA p-value	MCS
MSFE weighted (FS)	(C)	0.724	1.000	X	MSFE weighted (FS)	(C)	0.507	1.000	X
MSFE weighted (S)	(C)	0.816	0.123	X	MSFE weighted (S)	(C)	0.713	0.026	
PCWHSAX – IFOBCCONNDUR	(M)	0.850	0.636	X	Trimmed 25 (S)	(C)	0.805	0.000	
Trimmed 25 (FS)	(C)	0.852	0.000		Trimmed 25 (FS)	(C)	0.821	0.002	
PCWHSAX – EUBSPTIND	(M)	0.857	0.522	X	IFOBSITSAX – IFOEXEMAN	(M)	0.844	0.001	
PCWHSAX – IFOBCINT	(M)	0.863	0.572		ICWHSAX	(R)	0.844	0.000	
PCWHSAX – IFOBSCONDUR	(M)	0.899	0.328		IFOBCITSAX – IFOAOIWT	(M)	0.854	0.000	
Trimmed 25 (S)	(C)	0.907	0.000		IFOBEMANSAX – IFOEXEMAN	(M)	0.883	0.000	
HCNOSAX – EUBSOBLIND	(M)	0.908	0.424		IFOBEITSAX – IFOEXEMAN	(M)	0.895	0.001	
PCWHSAX – IFOBCCONDUR	(M)	0.910	0.158		HCWHSAX – IFOEXEMAN	(M)	0.900	0.000	
PCWHSAX – IFOBSCONNDUR	(M)	0.928	0.016		IFOBEITSAX – EUBSRTCI	(M)	0.900	0.003	
HCNOSAX – IFOBSINT	(M)	0.928	0.314		IFOBEMANSAX – EUBSRTCI	(M)	0.908	0.000	

Results for the Free State of Saxony – continued

HCNOSAX – IFOBEINT	(M)	0.929	0.293	HCTOSAX – GFKPL	(M)	0.908	0.000
PCWHSAX – EUBSPEIND	(M)	0.930	0.321	QML1QOLS (S)	(F)	0.909	0.000
CONEMPSAX – IFOBCINT	(M)	0.930	0.165	IFOBSMANSAX – IFOEXEMAN	(M)	0.914	0.002
Trimmed 50 (FS)	(C)	0.933	0.000	Trimmed 50 (FS)	(C)	0.916	0.002
CONWHSAX – EUBSPTIND	(M)	0.933	0.379	IFOCUCONSAX – IFOEXEMAN	(M)	0.919	0.000
IFOBSMANSAX – GFKFSE	(M)	0.935	0.001	Trimmed 50 (S)	(C)	0.922	0.001
HCNOSAX – ZEWES	(M)	0.935	0.166	EXVALUESAX – IFOBECAP	(M)	0.924	0.000
Trimmed 50 (S)	(C)	0.936	0.000	CONFIRMSAX – IFOEXEMAN	(M)	0.925	0.000

Note: The table reports the best 20 models with the smallest rRMSFE (column *Ratio*). The column *SPA p-value* presents the outcome of the SPA test by Hansen (2005). An **X** in column *MCS* denotes that this model is among the best ones, decided by the test of Hansen *et al.* (2011). The external Appendix shows the abbreviations used for the different indicators. (FS) Full Sample, (S) Saxony, (I) international, (N) national, (R) regional indicators, (C) combinations, (M) multi-indicator and (F) factor models.

Table 3.2: Results for Baden-Württemberg

Target variable: quarter-on-quarter growth rate GDP Baden-Württemberg									
h=1		RMSFE AR(p): 1.708%			**h=2**		RMSFE AR(p): 1.740%		
Model	Abbreviation	Ratio	SPA p-value	MCS	Model	Abbreviation	Ratio	SPA p-value	MCS
MSFE weighted (BW)	(C)	0.589	1.000	X	MSFE weighted (FS)	(C)	0.731	1.000	X
MSFE weighted (FS)	(C)	0.624	0.281	X	MSFE weighted (BW)	(C)	0.759	0.226	X
Trimmed 25 (BW)	(C)	0.681	0.062	X	IFOBSITBW – EUBSRTCI	(M)	0.781	0.198	X
Trimmed 25 (FS)	(C)	0.710	0.066	X	Trimmed 25 (FS)	(C)	0.787	0.053	
IFOBCITBW	(R)	0.725	0.000		IFOBCMANBW – EUBSRTCI	(M)	0.792	0.196	
IFOBCMANBW	(R)	0.747	0.006		IFOCUCONBW – GFKPL	(M)	0.799	0.002	
EUBSPEIND	(N)	0.755	0.028		IFOBCITBW – EUBSRTCI	(M)	0.808	0.030	
IFOIOFGMAN	(N)	0.758	0.000		IFOBEITBW – EUCSFSP	(M)	0.810	0.026	
Trimmed 50 (BW)	(C)	0.773	0.098		Trimmed 25 (BW)	(C)	0.815	0.024	
IFOBCCAP	(N)	0.774	0.078		IFOCUCONBW – GFKFSE	(M)	0.818	0.466	
IFOBSCAP	(N)	0.778	0.140		IFOBCWTBW – EUBSRTCI	(M)	0.822	0.010	
EUBSINDCI	(N)	0.779	0.002		IFOBSMANBW – EUBSRTCI	(M)	0.825	0.452	
EUBSSFGIND	(N)	0.781	0.000		KIBW – GFKWTB	(M)	0.825	0.026	
IFOBCMAN	(N)	0.784	0.003		IFOBCMANBW – GFKFSE	(M)	0.828	0.004	
Trimmed 50 (FS)	(C)	0.791	0.104		IFOBCMANBW – GFKCCC	(M)	0.830	0.076	
IFOBEINT	(N)	0.791	0.007		IFOBSMANBW – EUBSEMPEIND	(M)	0.830	0.419	
EUCSESI	(N)	0.793	0.000		IFOEMPECONBW – EUBSRTCI	(M)	0.830	0.024	

Results for Baden-Württemberg – continued

KIBW – COMBAEB	(M)	0.795	0.060		PCWHBW – EUBSRTCI	(M)	0.830	0.000	
KIBW – GFKCCC	(M)	0.800	0.125		IFOBSMANBW – GFKESE	(M)	0.831	0.008	
KIBW – GFKCCIN	(M)	0.800	0.154		IFOBCMANBW – GFKCCIN	(M)	0.831	0.067	

h=3		RMSFE AR(p): 1.645%			**h=4**		RMSFE AR(p): 1.664%		
Model	Abbreviation	Ratio	SPA p-value	MCS	Model	Abbreviation	Ratio	SPA p-value	MCS
MSFE weighted (FS)	(C)	0.714	1.000	X	MSFE weighted (FS)	(C)	0.649	1.000	X
ICNOBW – NRHT	(M)	0.802	0.182	X	MSFE weighted (BW)	(C)	0.705	0.243	X
CONNOBW – GFKBCE	(M)	0.828	0.297	X	PCNOBW – IFOBECONNDUR	(M)	0.816	0.000	
PCNOBW – IFOBCCONNDUR	(M)	0.832	0.068	X	Trimmed 25 (FS)	(C)	0.849	0.001	
PCNOBW – IFOOOHMAN	(M)	0.840	0.009		PCNOBW – GFKWTB	(M)	0.874	0.000	
IFOBEITBW – IFOEXEMAN	(M)	0.842	0.000		PCNOBW – GFKFSE	(M)	0.884	0.001	
Trimmed 25 (FS)	(C)	0.847	0.000		PCNOBW – GFKMPP	(M)	0.891	0.001	
PCNOBW – GFKUE	(M)	0.848	0.011		ICTOBW – EUBSSPEIND	(M)	0.891	0.013	
PCNOBW – EUCSUE	(M)	0.848	0.013		Trimmed 25 (BW)	(C)	0.894	0.006	
PCNOBW – IFOBSCONDUR	(M)	0.853	0.109		HCWHBW – GFKPL	(M)	0.899	0.000	
PCNOBW – IFOBSINT	(M)	0.853	0.023		PCNOBW – GFKFSL	(M)	0.904	0.000	
CONNOBW – GFKUE	(M)	0.858	0.104		ICWHBW – EUBSSPEIND	(M)	0.905	0.031	
CONNOBW – EUCSUE	(M)	0.858	0.106		PCNOBW – EUBSSPEIND	(M)	0.906	0.001	

Results for Baden-Württemberg – continued

HCWHBW – EUBSSSCI	(M)	0.858	0.244	IFOCUCONBW – GFKFSE	(M)	0.908	0.024
PCNOBW – EUBSOBLIND	(M)	0.870	0.006	IFOBSMANBW – EUBSSFGIND	(M)	0.910	0.033
PCNOBW – IFOBCCONDUR	(M)	0.873	0.020	IPMET	(N)	0.913	0.000
PCNOBW – EUBSSPEIND	(M)	0.874	0.017	HCWHBW – GFKWTB	(M)	0.915	0.003
KIBW – GFKWTB	(M)	0.876	0.079	ICNOBW – IFOAOIWT	(M)	0.922	0.008
HCNOBW – EUBSSSCI	(M)	0.876	0.207	ICTOBW – IFOBCINT	(M)	0.927	0.029
CONNOBW – IFOBSCONDUR	(M)	0.880	0.105	KIBW – GFKWTB	(M)	0.930	0.052

Note: The table reports the best 20 models with the smallest rRMSFE (column *Ratio*). The column *SPA p-value* presents the outcome of the SPA test by Hansen (2005). An **X** in column *MCS* denotes that this model is among the best ones, decided by the test of Hansen *et al.* (2011). The external Appendix shows the abbreviations used for the different indicators. (FS) Full Sample, (BW) Baden-Württemberg, (I) international, (N) national, (R) regional indicators, (C) combinations, (M) multi-indicator and (F) factor models.

Table 3.3: Results for Eastern Germany

Target variable: quarter-on-quarter growth rate GDP Eastern Germany									
h=1		RMSFE AR(p): 1.226%			**h=2**		RMSFE AR(p): 1.274%		
Model	Abbreviation	Ratio	SPA p-value	MCS	Model	Abbreviation	Ratio	SPA p-value	MCS
MSFE weighted (FS)	(C)	0.815	1.000	X	MSFE weighted (FS)	(C)	0.782	1.000	X
MSFE weighted (EG)	(C)	0.829	0.577	X	MSFE weighted (EG)	(C)	0.815	0.507	X
Trimmed 25 (EG)	(C)	0.864	0.285	X	IFOBSITEG – EUBSEMPEIND	(M)	0.876	0.414	X
Trimmed 25 (FS)	(C)	0.877	0.170	X	Trimmed 25 (FS)	(C)	0.879	0.032	X
EUCSITESI	(I)	0.907	0.002		IWHOLKMANEG – EUBSSSCI	(M)	0.882	0.518	X
IFOBSMANEG – ZEWES	(M)	0.916	0.324	X	IFOBCITEG – IFOAOIRS	(M)	0.882	0.148	X
IFOBSITEG – IFOIOFGMAN	(M)	0.917	0.007		IFOBSMANEG – GFKCCIN	(M)	0.887	0.430	
Trimmed 50 (EG)	(C)	0.925	0.008		IFOBSMANEG – EUCSCCI	(M)	0.887	0.433	X
Trimmed 50 (FS)	(C)	0.927	0.013		IWHOLKMANEG – GFKFSE	(M)	0.887	0.452	X
IWHOLKCONEG – IFOIOFGMAN	(M)	0.931	0.083		IFOBSMANEG – GFKCCC	(M)	0.889	0.416	
IFOBCITEG – IFOIOFGMAN	(M)	0.932	0.000		HCWHEG – IFOUNFWCON	(M)	0.890	0.239	
IFOBSMANEG	(R)	0.933	0.005		HCNOEG – IFOAOIRS	(M)	0.890	0.369	X
IFOBSITEG – IFOEOARS	(M)	0.933	0.002		IFOBSITEG – EUBSSFGIND	(M)	0.890	0.307	
IFOBSITEG – EUBSOBLIND	(M)	0.935	0.020		HCWHEG – EUBSINDCI	(M)	0.890	0.296	
IFOBSITEG – IFOOOHMAN	(M)	0.937	0.006		IFOBCITEG – EUBSSFGIND	(M)	0.891	0.175	
IFOBCMANEG – GFKESL	(M)	0.937	0.005		CONFIRMEG – IFOUNFWCON	(M)	0.891	0.188	

Results for Eastern Germany – continued

DREUROREPO	(N)	0.938	0.097		IFOBCWTEG – EUBSSFGIND	(M)	0.891	0.352	
IFOBCMANEG – IFOBEINT	(M)	0.938	0.005		IFOBCWTEG – IFOIOFGMAN	(M)	0.891	0.312	
IFOBSMANEG – GFKCCC	(M)	0.942	0.002		IFOBSMANEG – IFOBSINT	(M)	0.892	0.294	
IFOBCITEG – EUBSEMPEIND	(M)	0.942	0.014		IFOBSMANEG – IFOBCINT	(M)	0.893	0.293	
h=3		RMSFE AR(p): 1.319%			**h=4**		RMSFE AR(p): 1.223%		
Model	Abbreviation	Ratio	SPA p-value	MCS	Model	Abbreviation	Ratio	SPA p-value	MCS
MSFE weighted (FS)	(C)	0.742	1.000	X	MSFE weighted (FS)	(C)	0.654	1.000	X
IFOBSMANEG – ZEWPS	(M)	0.856	0.454	X	Trimmed 25 (FS)	(C)	0.836	0.010	X
Trimmed 25 (FS)	(C)	0.872	0.027	X	MSFE weighted (EG)	(C)	0.872	0.165	X
IFOBERSEG – IFOBECONNDUR	(M)	0.891	0.379	X	Trimmed 25 (EG)	(C)	0.888	0.126	X
IFOBCWTEG – IFOBECONNDUR	(M)	0.892	0.360	X	Trimmed 50 (FS)	(C)	0.901	0.000	
MSFE weighted (EG)	(C)	0.895	0.061	X	IFOBECONEG	(R)	0.921	0.017	
HCWHEG – IFOBECONNDUR	(M)	0.899	0.577		ZEWES	(N)	0.923	0.000	
HCNOEG – GFKUE	(M)	0.900	0.460		Trimmed 50 (EG)	(C)	0.929	0.003	
HCNOEG – EUCSUE	(M)	0.900	0.442		CONHW	(N)	0.934	0.002	
IFOBCRSEG – IFOBECONNDUR	(M)	0.900	0.482	X	CONTOEG – GFKPL	(M)	0.943	0.125	X
HCWHEG – IFOBCCONNDUR	(M)	0.901	0.556		DAXSPI	(N)	0.955	0.200	X
HCWHEG – GFKUE	(M)	0.901	0.585		Trimmed 75 (FS)	(C)	0.956	0.002	
HCWHEG – EUCSUE	(M)	0.901	0.590	X	TOVEMD	(N)	0.957	0.063	

Results for Eastern Germany – continued

IFOBSMANEG – GFKCCIN	(M)	0.902	0.413	NOCEOD	(N)	0.961	0.002
IFOBSMANEG – EUCSCCI	(M)	0.902	0.412	CONNREPE	(N)	0.970	0.057
IWHSITCONEG – IFOBCCONDUR	(M)	0.902	0.574	NOMECHD	(N)	0.972	0.043
HCWHEG – GFKCCIN	(M)	0.904	0.601	GFKIE	(N)	0.976	0.000
HCWHEG – EUCSCCI	(M)	0.904	0.592	PCWHEG	(R)	0.988	0.007
HCWHEG – EUBSPTIND	(M)	0.905	0.489	Trimmed 75 (EG)	(C)	0.989	0.002
HCWHEG – COMBAEB	(M)	0.906	0.568	HCTOEG	(R)	0.992	0.000

Note: The table reports the best 20 models with the smallest rRMSFE (column *Ratio*). The column *SPA p-value* presents the outcome of the SPA test by Hansen (2005). An **X** in column *MCS* denotes that this model is among the best ones, decided by the test of Hansen *et al.* (2011). The external Appendix shows the abbreviations used for the different indicators. (FS) Full Sample, (EG) Eastern Germany, (I) international, (N) national, (R) regional indicators, (C) combinations, (M) multi-indicator and (F) factor models.

For all three GDP target variables, the AR(p) benchmark model is significantly outperformed. This result holds true for all considered forecasting horizons. However, we must consider that forecast improvements in comparison to the autoregressive benchmark decrease with longer forecast horizons. It becomes even more difficult to predict regional GDP in the long-term. This fact is also indicated by the MCS test. With the exception of Eastern Germany, only few models are included in the set of best models in the long-term. Differences across the regions exist in the overall forecasting performance and the composition of indicators. The most accurate forecasts are observable for the Free State of Saxony and Baden-Württemberg. For Eastern Germany, the RMSFE is slightly higher in comparison with the other two regions. What we can see from the three tables is that pooling performs best for all three regional target variables. Next to MSFE weighted combination strategies, trimmed means in particular produce lower forecast errors than the benchmark models. As indicated by the tests, no competitor has a higher accuracy than pooling models. In addition, combination strategies are part of the set of best models.

Another interesting result is that in most cases multi-indicator models outperform single-indicator models. Adding another national indicator to a regional one clearly enhances the forecast accuracy of regional GDP. Single-indicator models perform well for Baden-Württemberg in the short-term ($h = 1$) and for Eastern Germany in the long-term ($h = 4$). We have to state that the most important forecasting signals come from regional and national indicators. International indicators do not play an important role in predicting regional GDP. Because we use a large data set, it is interesting to examine the differences between pooling and factor models. Whereas the combination of forecasts from different models performs quite well, the forecast improvement by factor models is not very impressive. We find rRMSFEs that are smaller than one; however, these models are not very competitive in comparison to pooling or multi-indicator forecasts in our case. With the exception of Saxony, no factor model is among the top 20.

3.3.2 Detailed Regional Results

Free State of Saxony

Pooling (MSFE weighted (FS), $rRMSFE = 0.582$) and multi-indicator models yield the best results for the Saxon GDP in our "pseudo-real-time" setting (see Table 3.1). The multi-indicator models are dominated by two regional indicators in the short- and mid-term: orders on hand in the Saxon construction sector (IFOOOHCONSAX) and working hours in the sector of public construction (PCWHSAX). These results are not surprising because construction traditionally plays an important role in Eastern German states. The MCS test also indicates that multi-indicator models are part of the best set of models in the short- and mid-term. In the long-term ($h = 4$), only the MSFE-weighted model is within the set of the best models. A closer look at the multi-indicator models reveals that surveys (consumer or business), in particular, produce lower forecast errors than our benchmark model and that regional indicators are essential when forecasting GDP. The Ifo business climate for industry and trade in Germany (IFOBCIT, $rRMSFE = 0.793$) in the short-term or in Saxony (IFOBCITSAX) in the long-term has a higher forecast accuracy than the autoregressive process. These results are consistent with forecasting literature for Germany. One of the most important leading indicators for German GDP is the Ifo business climate for industry and trade.[18] This phenomenon also applies to Saxony (Lehmann *et al.*, 2010). Turning to consumer surveys, Table 3.1 reveals that these indicators are very helpful in predicting Saxon GDP in the short- and mid-term. Particularly the consumer confidence climate (GFKCCC) significantly reduces forecast errors and, in combination with IFOOOHCONSAX, is part of the best set of models. This result is straightforward because Eastern German manufacturing firms mainly interact on domestic markets (see Ragnitz, 2009). Furthermore, exports (EXVALUE, $h = 4$) and export expectations in the manufacturing sector (IFOEXEMAN, short- and long-term) improve forecast accuracy. The latter indicator is also part of the set of best models in the short-term. Within the Eastern German states, the Saxon economy has the highest degree of openness (approximately 40% of all turnovers in the manufacturing sector come from abroad). Another highlight is the importance of business expectations from capital (IFOBECAP, $rRMSFE = 0.766$) and intermediate goods producers (IFOBEINT) in the medium- and long-term. This result is straightforward because the Saxon industry is predominantly described by these two sectors. Approximately 80% of all turnovers in 2011 come from intermediate and capital goods (e.g. vehicle manufacturing, which is the dominant sector in the Saxon industry) producers. Saxon firms are strongly linked to the Western German economy; therefore, national indicators are useful for predicting Saxon GDP. In comparison to the other regions, factor models belong to the top 20 only in Saxony (QML1QOLS, $rRMSFE = 0.909$, $h = 4$).

[18] For a recent survey, see Abberger and Wohlrabe (2006).

Baden-Württemberg

As we have seen from the results for Saxony, pooling of forecast outcomes also produces the lowest forecast errors in Baden-Württemberg. For all forecast horizons, pooling models dominate all other competitors and are always part of the best set of models. The best combination strategy predicts GDP one quarter ahead almost 40% more accurately than the AR benchmark (see MSFE weighted in Table 3.2). In contrast to Saxony, single-indicator models perform better than multi-indicator models in the short-term ($h = 1$). In particular, regional survey results such as the Ifo business climate for industry and trade in Baden-Württemberg (IFOBCITBW, $rRMSFE = 0.725$) and a regional business cycle indicator (KIBW) outperform the autoregressive benchmark. In addition, survey results from the manufacturing sector (IFOBCMANBW, $rRMSFE = 0.747$) and from capital goods producers (IFOBCCAP, $rRMSFE = 0.774$) provide important forecasting signals in our 'pseudo-real-time' setting. These results can be explained by the economic structure of Baden-Württemberg. Baden-Württemberg has the highest share of manufacturing among the German states; approximately 30% of nominal gross value added is generated in this sector. Manufacturing of motor vehicles (e.g. Daimler AG, which explains the performance of NRHT for $h = 3$), machinery and equipment, the fabrication of metal products and highly innovative capital goods producers such as the Bosch Group predominantly describe the industrial structure in manufacturing. As in Saxony, the multi-indicator models are dominated in the medium- and long-terms by two indicators: the Ifo business climate in manufacturing (IFOBCMANBW) and new orders in the public construction sector (PCNOBW). The latter indicator is indeed part of the best model set. Another interesting result is the importance of export expectations in the manufacturing sector (IFOEXEMAN) in the mid-term. Baden-Württemberg has one of the highest export quotas of the German states; more than 50% of all industrial turnovers are generated in foreign countries. The most important trading partners come from the Euro Area, followed by the United States. For companies such as Daimler AG and the Bosch Group, the United States is one of the most relevant markets.

Eastern Germany

Regional business surveys provided by the Ifo Institute (IFOBSMANEG, $rRMSFE$ 0.933) and the IWH (IWHOLKMANEG) are able to predict Eastern German GDP more accurately than the autoregressive benchmark in the short- and mid-term. Considering national variables, we also find results that are consistent with the Eastern

German economic structure. The Ifo business climate for intermediate goods producers (IFOBCINT, $h = 2$), macroeconomic variables for Germany (e.g. NOMECHD, $h = 4$) and the consumer sentiment indicator (GFKCCIN, mid-term) help for the prediction of Eastern German GDP. First, Eastern German firms interact mostly on domestic markets and have a lower export quota in comparison to their Western German counterparts (see Ragnitz, 2009). Second, the Eastern German industrial sector is mainly characterized by intermediate goods producers. Nearly 40% of all turnovers in 2011 were achieved in this industrial main group. Ragnitz (2009, p.55) states that most Eastern German firms are still so-called "extended workbenches" (in German: *verlängerte Werkbänke*) of Western German companies. Overall, Western German economic development is a crucial factor for quarter-on-quarter GDP growth in Eastern Germany. Another interesting result is that single-indicator models perform better in the long-term than multi-indicator models (see $h = 4$ in Table 3.3). In addition, the multi-indicator models are not dominated by a small number of indicators to the same extent as in the other two regions. Only the business situation for industry and trade in Eastern Germany (IFOBSITEG) in the short-term or the working hours for the Eastern German housing construction sector (HCWHEG) in the mid-term stand out from this overall picture. In line with the results for Saxony and Baden-Württemberg, pooling has the highest forecast accuracy in terms of RMSFE. This class of models dominate all competitors in the short- and long-term and are part of the model confidence set. In contrast to Saxony and Baden-Württemberg, a larger number of models are included in the set of best models in Eastern Germany.

3.4 Conclusion

This study analyzes the forecasting performance of single-indicator, multi-indicator, factor models and pooling techniques at the regional level. Our analysis is conducted in a "pseudo-real-time" setting, i.e., taking account of publication lags, though not of data revisions. We use a large data set with international, national and regional variables. As target variables, we use unique quarterly data for GDP that are provided by different sources for the period 1996:01 to 2010:04. Our study is the first to systematically use time series techniques to forecast regional GDP. Altogether, it is possible to predict GDP at the regional level at a quarterly frequency. A large number of indicators produce lower forecast errors than the benchmark model. The different results for our three target variables show that a high heterogeneity exists between regional units. An important reason for this heterogeneity is the regional economic structure, as the highlighted section shows. Furthermore, we can conclude that regional indicators have a high forecasting power. Whenever regional variables are available, these indicators are worth considering for forecasting. As our results show, regional variables deliver good forecasting signals or information. Because we use a large data set, pooling strategies can improve forecasting accuracy. For all three regional units, MSFE weights outperform all other weighting schemes, as well as single-indicator and multi-indicator forecasts. Hence, pooling in a regional context is just as important as on the national level. Another way to handle large data sets is to apply factor models. Despite the fact that this class of models improves forecast accuracy, which is in line with the existing literature, factor models are not that competitive compared to pooling or multi-indicator models in our case. Finally, we have shown that in most cases, multi-indicator models significantly improve forecast accuracy in comparison to single-indicator models. By adding national variables to regional indicators, forecasts become even better at the regional level. Regional policy-makers have to rely on accurate macroeconomic forecasts. With our exercise, we are able to reduce forecast errors significantly and therefore reduce uncertainty about future macroeconomic development at the regional level. This approach renders regional economic policy more assessable. Further research is necessary for different countries (e.g. the United States or EU) and aggregation levels. It would be interesting to know whether it is better to predict regional GDP directly or through its different components. This issue was analyzed for Germany as a whole by Drechsel and Scheufele (2012a); however, to date, no regional study exists.

4 Forecasting gross value added at the regional level: Are sectoral disaggregated predictions superior to direct ones?

With kind permission of Springer Science+Business Media, this chapter is the reprint of the original article by Lehmann and Wohlrabe (2014), published in the journal *Review of Regional Research: Jahrbuch für Regionalwissenschaft*, February 2014, Volume 34, Issue 1, pp 61-90, © Springer-Verlag Berlin Heidelberg 2013.

4.1 Motivation

Fiscal policy at the sub-national level is one of the major fields in the decision-making of policy makers. For this purpose, reliable forecasts of economic aggregates (as gross domestic product or gross value added) are necessary. At the regional level, e.g. states or counties, data limitations or a low publication frequency of national accounts make it difficult to predict macroeconomic aggregates and may cause higher forecast errors in comparison to countries' aggregate, e.g., total German gross domestic product (GDP). Additionally, the forecast for Germany may not be a good approximation for the economic development of sub-national (e.g., states) aggregates. The reasons are a high heterogeneity in regional economic structures and different regional business cycles. Whenever a shock such as the economic crisis of 2009 hits the German economy, not all states have to develop in the same way. Therefore, separate regional forecasts are needed. Only few attempts have been made to forecast regional macroeconomic aggregates. Bandholz and Funke (2003) predict turning points for the German state[1] Hamburg with a newly

[1] Germany consists of 16 different states which are categorized as NUTS-1 for statistics of the European Union. In comparison, Germany is classified as NUTS-0.

constructed leading indicator. The study by Dreger and Kholodilin (2007) employs a set of regional indicators to forecast the GDP of the German state Berlin. Kholodilin *et al.* (2008) predict the GDP of all German states simultaneously and account for spatial effects in a dynamic panel setup. Lehmann and Wohlrabe (2015) showed for three different regional units in Germany (the Free State of Saxony, Baden-Württemberg and Eastern Germany[2]) that forecast accuracy of GDP at the regional level can be improved with a huge data set of indicators in comparison to simple benchmark models. At the level of Canadian provinces, Kopoin *et al.* (2013) evaluate the forecasting information of national (Canadian) and international indicators.

While these few prominent studies focus on the prediction of aggregated GDP directly, this paper mainly concentrates, from a regional point of view, on the question whether it is possible to forecast gross value added (GVA) for different sectors (e.g., manufacturing, construction etc.). Regional policy makers or credit institutes (e.g., for granting of credits) are not only interested in the development of the economy as a whole but also in forecasts for different branches of the economy. From a practitioners point of view it is necessary to know which branches or aggregates drive future economic development, so that predicting sub-components makes the state of the economy more tangible. Another important point for disaggregated forecasts is the consideration that several indicators (e.g., the EU business survey for manufacturing) might be linked to sub-components even stronger than to macroeconomic aggregates (e.g., GDP or total GVA). As mentioned above, missing quarterly sectoral GVA data at the regional level makes such an analysis impossible until yet. But our data set enables us to carry out such an analysis, since we have quarterly GVA data for one German state (Free State of Saxony). To the best of our knowledge, this is the only German state where quarterly GVA data for different sectors is available.

Additionally, this paper evaluates whether it is preferable to forecast an aggregate directly (total GVA) or to sum up its weighted sub-components (sectoral GVA) at the regional level. Recently, this question has become more and more attractive in the field of economic forecasting. For the Euro Area as a whole, forecast performance for different sub-components of GDP is analyzed by Hahn and Skudelny (2008) and Angelini *et al.* (2010). Barhoumi *et al.* (2008) and Barhoumi *et al.* (2012) study this question for the French economy. A comparison of forecast accuracy of sub-components for Germany is made by Cors and Kuzin (2003) or Drechsel and Scheufele (2012a). Whereas the first

[2] Eastern Germany is the aggregation of five German states: Brandenburg, Mecklenburg-West Pomerania, the Free State of Saxony, Saxony-Anhalt and the Free State of Thuringia.

article only studies the production side (aggregation of sectoral GVA) of the German economy, the second study compares the different outcomes from the demand (e.g., private consumption, exports etc.) and supply side with those of aggregated German GDP. For the German labor market, the study by Weber and Zika (2013) finds an improvement of forecast accuracy for employment figures through disaggregation in the short-term. They show that the aggregation of forecasts for different branches of the economy can produce lower forecast errors in comparison to the prediction of total employment. Studies for regional units, which evaluate aggregate vs. disaggregate forecasts, are missing.

The contribution of our paper is manifold. First, we evaluate forecast accuracy of different indicators for several branches of the economy and forecast horizons (one up to four quarters). With such an analysis we make the state of the economy more tangible and can clearly specify what drives future economic development. Second, we apply different pooling strategies. It is well-known in the forecasting literature that the combination of forecasting output from competing models can yield lower forecast errors (Stock and Watson, 2006; Timmermann, 2006). In numerous studies, the advantage of pooling was confirmed (Drechsel and Maurin, 2011; Eickmeier and Ziegler, 2008). For three German regions, Lehmann and Wohlrabe (2015) find that pooling significantly produces lower forecast errors for regional GDP than a univariate benchmark model. Sub-national studies for different sectors are still missing. Third, this paper applies factor models as well. Several studies at the national level find significant improvements of forecast accuracy for this class of models (see, e.g., Schumacher (2007) and Schumacher (2010) for Germany, or Stock and Watson (2002) for the US). At the regional level, Lehmann and Wohlrabe (2015) find that factor models show no significant improvement for regional GDP in Germany. Finally, we compare direct and disaggregated forecasts of gross value added with each other and ask whether there is an information gain when predicting sub-components. To carry out this analysis we use a huge data set at the regional level which incorporates quarterly national accounts for one German state (Saxony). We have information on GDP, total GVA and its sub-components as well as 317 different indicators from the international (USA, EU etc.), national (Germany) and regional level (Saxony). This study is closely linked to the one by Lehmann and Wohlrabe (2015), since it focuses on regional forecasts. But in contrast, it studies sectoral forecasts instead of GDP and additionally asks whether it is preferable to predict sub-components instead of aggregates.

The paper is organized as follows. Section 4.2 describes our data, the aggregation method and our empirical setup. The results are discussed in Sect. 4.3. The last section summarizes our main findings.

4.2 Data and Methodology

4.2.1 Data

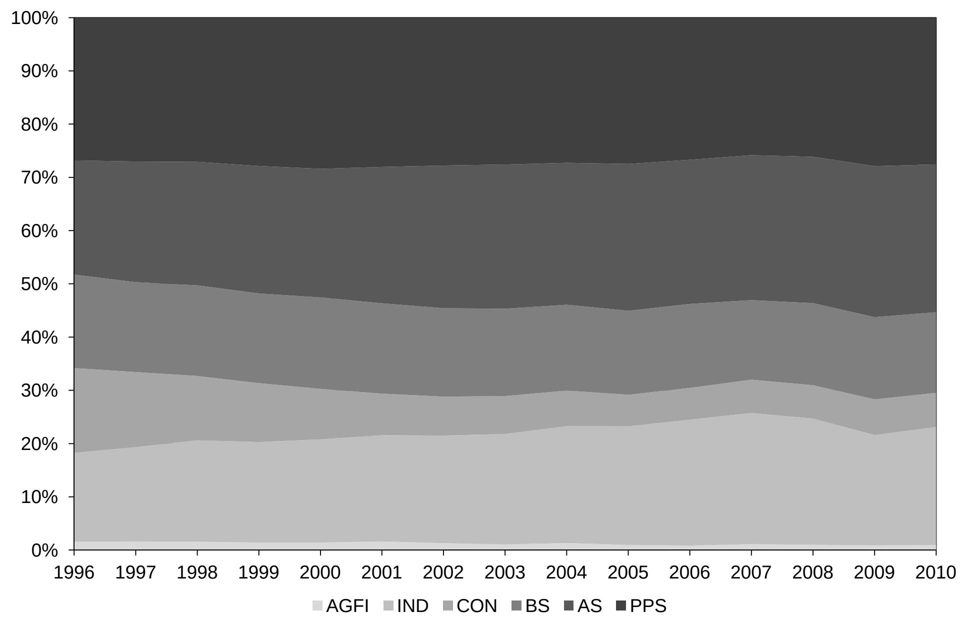
In general, there are no temporal disaggregated macroeconomic data (e.g., quarterly GVA) available at the regional level in Germany. It is possible to use annual information, but this causes the problem of an insufficient number of observations. To the best of our knowledge, only Nierhaus (2007) provides quarterly data on GVA for different sectors. He calculates national accounts for the German state Saxony, which we use in this paper.[3] Gross value added in real terms is available for six aggregated sectors: (i) agriculture, hunting and forestry; fishing (AGFI), (ii) mining and quarrying; manufacturing; electricity, gas and water supply (industry; IND), (iii) construction (CON), (iv) wholesale and retail trade; hotels and restaurants; transport (basic services; BS), (v) financial intermediation; real estate, renting and business activities (advanced services; AS), (vi) public administration; education; health and social work; private households (public and private services; PPS).[4] The methodological background for the computation of the quarterly data is the temporal disaggregation method developed by Chow and Lin (1971). They suggest to employ a stable regression relationship between annual aggregates and indicators with a higher frequency (e.g., quarterly data). With this relationship it is possible to convert annual into quarterly data. But these quarterly information have to fulfill two restrictions: horizontal and temporal aggregation (see Nierhaus, 2007). This means that first the sum of GVA of all sectors has to result in total GVA for every time period. Second, the average index of four quarterly data points has to equal the annual aggregate. We exclude those indicators from our analysis which were used for temporal disaggregation by Nierhaus (2007). These indicators have to perform well for predicting sector-specific GVA. To avoid such a bias, the following indicators for Saxony are not part of the analysis: turnovers in the manufacturing and construction sector, turnovers for retail sale and wholesale trade. All GVA target variables are available in real terms and for the period 1996:01 to 2010:04.

[3] The data are available upon request from *dresden@ifo.de*.
[4] These six sectors describe the whole economy so that the sum equals total GVA.

The data are seasonally adjusted with Census X-12-ARIMA and we transformed these into quarter-on-quarter (qoq) growth rates.

To get an impression on how the different sectors contribute to total GVA, Fig. 4.1 shows the sectoral structure of Saxony. The figure shows the share of our six sectors of interest in total GVA for the years 1996 to 2010. For all years, the share of agriculture, hunting and forestry; fishing (AGFI) is negligible (in 2010: 1%). The share of industry (IND) is approximately 22% of total GVA in 2010 (for comparison: Germany 24%). The construction sector (CON) is traditionally large in Eastern German states, because a building boom was initiated in Eastern Germany after reunification. Since the mid 1990s, the construction sector lost its importance for total GVA in Eastern Germany. The share of construction in Saxon GVA was 6.5% in 2010 (Germany: 4%). Basic services (BS) have a share in total GVA of about 15% (Germany: 17%). With a share of 28% of total GVA the sector advanced services (AS) is of a smaller magnitude than in Germany (30.5%). The public sector (PPS) is traditionally overrepresented in Eastern Germany (in comparison to Germany); the share of PPS in total GVA is 27.5% in Saxony and 24% in Germany.

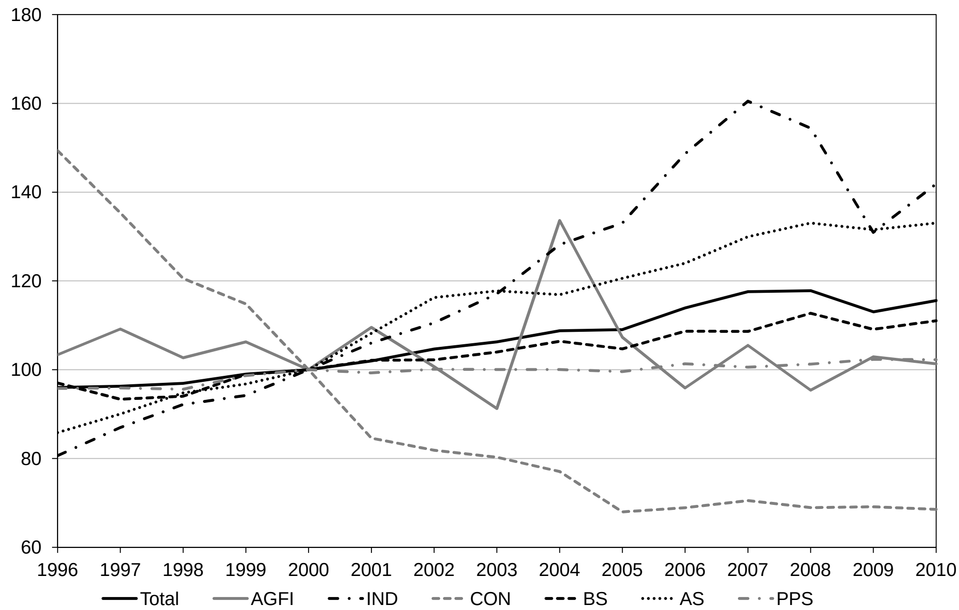
Figure 4.1: Sectoral shares in total GVA for Saxony



Abbreviations: AGFI...agriculture, hunting and forestry; fishing, IND...industry, CON...construction, BS...basic services, AS...advanced services, PPS...public and private services. *Source:* Working Group Regional Accounts VGRdL (2011a), authors' illustration.

After presenting the sectoral shares, Fig. 4.2 shows the development of total and sectoral real GVA for our period of investigation. The most volatile figure is the one for the primary sector (AGFI). Mainly special events drive real GVA growth in this branch of the economy. The public sector (PPS) is the branch with no dynamics at all. Real GVA in the Saxon construction sector (CON) shrinks throughout until the year 2005. Afterward this branch stabilizes and shows a lateral movement in growth rates. The two service sectors (basic services – BS and advanced services – AS) experienced a positive trend in real GVA growth for the whole period under observation. After the base year 2000, GVA in advanced services grew faster than value added for basic services. The industrial sector (IND) is the branch with the highest growth rates in real GVA. The reason is the high export dependence of this sector. But on the opposite, the export dependence makes the industrial sector prone to negative external shocks such as the one observed in the global downturn years 2008 and 2009. Total Saxon GVA is mainly driven by the development in the industrial sector.

Figure 4.2: Total and sectoral real GVA for the Saxon economy



Abbreviations: AGFI...agriculture, hunting and forestry; fishing, IND...industry, CON...construction, BS...basic services, AS...advanced services, PPS...public and private services. *Note:* The axis of ordinates shows the real Chain Index with the year 2000=100. *Source:* Working Group Regional Accounts VGRdL (2011a), authors' illustration.

To forecast sectoral GVA we use a huge data set containing 317 indicators which are grouped into seven categories: macroeconomic (95), finance (31), prices (12), wages (4), surveys (74), international (32) and regional (69). The category macroeconomic indicators contains German industrial production, new orders in manufacturing or foreign trade figures. Financial variables are, e.g., interest rates, exchange rates and government bond yields. Furthermore, we have price indices for exports and imports as well as consumer and producer prices. Qualitative measures are collected from different survey results. We have information from consumer surveys (Society for Consumer Research – GfK), business surveys (Ifo Institute or European Commission) or expert surveys (Centre for European Economic Research – ZEW). Additionally, we add composite leading indicators for Germany obtained from the OECD and the Early Bird of the Commerzbank to this group. International indicators cover a wide range of information from large economies (US, China, France or Italy). Finally, we have qualitative (Ifo business survey results) and quantitative indicators (e.g., new orders or prices) from the regional level. As mentioned before, we excluded regional indicators which were used for temporal disaggregation of sector-specific GVA.

Most of the indicators are available on a monthly basis. To obtain quarterly information, we first seasonally adjust the data with Census X-12-ARIMA and then calculate a three-month average. Stationarity is warranted through different transformations (either first differences or qoq growth rates), whenever the levels are nonstationary. For a complete description of our data set as well as the applied transformation for each indicator, see Table A.4 in the Appendix 4.A.

4.2.2 Aggregation of GVA Sub-components

National accounts provide two concepts for disaggregating GDP: (i) demand side and (ii) supply side. The first concept uses the identity that total production in an economy equals total domestic demand. So GDP is the sum of private and public consumption, investments, inventories and net exports (exports minus imports). The second concept looks at the production side of an economy. GDP is therefore the sum of gross value added of every industry plus taxes minus subsidies. In our data set no information about quarterly demand side variables are available. Therefore we can only look at the supply side. Since the aggregate taxes minus subsidies is difficult to forecast, we concentrate on GVA rather than GDP. The qoq growth rate of total Saxon GVA (y_t^{GVA}) could be expressed, for all $t = 1, 2, ..., T$, as:

$$
\begin{aligned}
y_t^{GVA} = {} & \omega_t^{AGFI} y_t^{AGFI} + \omega_t^{IND} y_t^{IND} + \omega_t^{CON} y_t^{CON} + \omega_t^{BS} y_t^{BS} \\
& + \omega_t^{AS} y_t^{AS} + \omega_t^{PPS} y_t^{PPS} .
\end{aligned} \tag{4.1}
$$

Therefore, the total growth rate is a sectoral-weighted sum of the single sectoral GVA growth rates (ω_t^s). As we can see from Eq. (4.1), the weights are time-varying and we assume that the sum of all weights has to equal unity. Whenever a forecast is made, the weights are ex ante unknown to the forecaster. In our forecasting exercise we assume that the weights in every forecasting period are constant with respect to the last known value.[5] For example, imagine we want to make a forecast for the first quarter of 2010 and information are available until 2009:04. Then we use the last known shares in total GVA from 2009:04 and apply them to aggregate sector-specific GVA forecasts in 2010:01.

4.2.3 Forecast Procedure

We employ the autoregressive distributed lag (ADL) model,

$$
y_{t+h}^{s,k} = \alpha + \sum_{i=1}^{p} \beta_i y_{t+1-i}^{s} + \sum_{j=1}^{q} \gamma_j x_{t+1-j}^{k} + \varepsilon_t^{s,k} , \tag{4.2}
$$

to generate our forecasts, where $y_{t+h}^{s,k}$ denotes the h-step ahead forecast of real GVA for sector s (including total) and x_t^k stands for one of our 317 exogeneous indicator. The variable k relates to one of these indicators. We allow a maximum of 4 lags, both for the endogeneous and exogeneous variables. The Schwarz Information Criteria (BIC) is used for the optimal lag length selection of p and q. Equation (4.2) is estimated in a recursive way and we use the data from 1996:01 to 2002:04 ($T_E = 28$) as the initial estimation period. Afterward we enlarge the estimation period successively by one quarter, at which the model of Eq. (4.2) is respecified. So we obtain for every forecast horizon h the first forecast for our target variables at 2003:01 and the last at 2010:04. h is defined as $\{1, 2, 3, 4\}$.[6] We apply a direct-step forecasting approach, so that for every forecasting

[5] Drechsel and Scheufele (2012a) state that in most cases simple averages are used for weighting subcomponents. In contrast, they use a moving average over the last four quarters to obtain their estimated weights. Since the shares in our sample are relatively persistent, the results should not differ dramatically by applying another approach.

[6] In this paper we denote one quarter ($h = 1$) as short-term, two and three quarters ($h = 2, 3$) as medium-term and four quarters ($h = 4$) as long-term. These definitions are in line with the forecasting literature and do not reflect time horizons in macroeconomic theory.

horizon and indicator $T_F = 32$ forecasts are generated. This is obtained by adjusting Eq. (4.2) in such a way that for each forecast horizon the first forecast is calculated for the first quarter 2003. Our benchmark model is a standard AR(p) process. We define $y^{agg,k}_{t+h}$ if the forecast is generated directly for total GVA and $y^{dis,k}_{t+h}$ for a weighted forecast from all sub-components.

4.2.4 Pooling

The outcome of a pooling-based forecast $\widehat{y}^{s,Pool}_{t+h}$ for sector s is the product of single indicator forecasts $\widehat{y}^{s,k}_{t+h}$ and a specific weighting scheme $w^{s,k}_{t+h}$:

$$\widehat{y}^{s,Pool}_{t+h} = \sum_{k=1}^{K} w^{s,k}_{t+h} \widehat{y}^{s,k}_{t+h} \quad \text{with} \sum_{k=1}^{K} w^{s,k}_{t+h} = 1 \,. \tag{4.3}$$

As Eq. (4.3) shows, the weights are indexed by time and thus varying with every estimation of our model. K stands for the number of models which are used for pooling.

We apply six different weighting schemes. A very simple scheme are (i) equal weights: $w^{s,k}_{t+h} = 1/K$. For this weighting scheme, the sheer number of models is important. To control for outliers, we additionally apply (ii) a median approach. We follow the studies by Drechsel and Scheufele (2012b) or Lehmann and Wohlrabe (2015) and calculate weights from two additional categories: in-sample and out-of-sample measures. Whereas weights from in-sample measures use criteria on how good the model fits the data, weights from out-of-sample measures are based on past forecast errors.

We apply two in-sample measures: (iii) BIC and (iv) R^2. The weights from these two measures are time-varying and have the following form:

$$w^{k,BIC}_{t+h} = \frac{\exp\left(-0.5 \cdot \Delta^{BIC}_k\right)}{\sum_{k=1}^{K} \exp\left(-0.5 \cdot \Delta^{BIC}_k\right)} \tag{4.4}$$

$$w^{k,R^2}_{t+h} = \frac{\exp\left(-0.5 \cdot \Delta^{R^2}_k\right)}{\sum_{k=1}^{K} \exp\left(-0.5 \cdot \Delta^{R^2}_k\right)}\,, \tag{4.5}$$

with $\Delta^{BIC}_k = BIC^k_{t+h} - BIC_{t+h,min}$ and $\Delta^{R^2}_k = R^2_{t+h,max} - R^2_{t+h,k}$. The difference between the two schemes is straightforward. Whereas a model with a lower BIC gets a higher weight, the importance of a single model for pooling increases with higher values of R^2.

For the application of out-of-sample weights, it is appropriate to use past forecast errors from different models. First, we apply a so called (v) trimmed mean. Indicators with a bad performance are filtered and not considered for pooling. In accordance with the existing literature, we include the best 25%, 50% or 75% performing indicators. The outcome of all remaining indicators are combined with equal weights. Second, (vi) discounted mean squared forecast errors (MSFE) are applied to calculate the weights, which have the following form:

$$w_{t+h}^{k} = \frac{\lambda_{t+h,k}^{-1}}{\sum_{k=1}^{K} \lambda_{t+h,k}^{-1}} \,. \tag{4.6}$$

$\lambda_{t+h,k} = \sum_{n=1}^{T_F} \delta^{t-h-n} \left(FE_{t+h,n}^{k}\right)^2$ represents the sum of discounted[7] (δ) forecast errors of the single-indicator model k. As the weighting scheme indicates, more recent forecast errors get a higher weight than older ones. Since the weighting schemes depend on the number of indicators considered for pooling, we either combine forecasts from all indicators of the full sample (FS) or only use indicators for Saxony (S).

4.2.5 Factor Models

Next to pooling, another way of dealing with large cross-sectional data sets are static and dynamic factor models. The literature finds that this class of models performs very well (see, e.g., Stock and Watson, 2002; Marcellino *et al.*, 2003; Forni *et al.*, 2005). The idea behind factor models is straightforward. Because standard econometric approaches cannot handle all available indicators (in our paper: 317) at the same time, factor models summarize the information of many time series in few common factors. With this approach we are able to specify a parsimonious model, thereby reducing the biases in parameter estimates (see Giannone *et al.*, 2008). In this paper, we apply three different methodologies to extract the common factors from our indicator series. For details, see the cited literature for each approach. First, the standard principal components (PC) method is the easiest way to extract the common factors. In line with Giannone *et al.* (2008), the second approach is the two-step estimator proposed by Doz *et al.* (2011). This procedure uses principal components and Kalman filtering (PCKF) and shows efficiency improvements over standard PC methods. Third, we extract the common factors via quasi-maximum likelihood (QML) estimation (see Doz *et al.*, 2012).[8]

[7] The literature has not found a consensus yet about the level of the discount rate. We apply different values ($\delta \in \{0, 0.1, 0.2, ..., 1\}$) and find similar results. Because of this and to avoid long tables, we only report the outcome for a discount rate equal to 0.1.

[8] We do not take into account the ragged edge problem (see Wallis, 1986) and extract the factors from the information set up to $t-1$.

For all three approaches we have to decide how many factors to extract from the series. We decide to choose a maximum of three common factors. The factors can either be estimated from the full sample of indicators (FS) of only extracting them from the regional series (S). Another decision has to be made according to the frequency. We extract the factors from the quarterly series (Q). To generate the forecasts for real GVA, we have another two possibilities. First, we can directly put the factors in the ADL model from Eq. (4.2), instead of using single indicators (ADL). Second, as proposed by Giannone *et al.* (2008), we can run a simple OLS-estimation, where real GVA is explained by a constant and the extracted factors available at different points in time (OLS). Whereas the first method considers lagged values of the dependent variable and the factors, the OLS-approach does not. In the end, this gives us 36 factor models for every Saxon branch of the economy as well as total GVA.[9]

4.2.6 Forecast Accuracy

To evaluate how good different indicators perform, we calculate forecast errors in a first step. The forecast of model k in sector s for the forecasting horizon h is denoted as $\widehat{y}^{s,k}_{t+h}$. The resulting forecast error is defined as $FE^{s,k}_{t+h} = y^{s,k}_{t+h} - \widehat{y}^{s,k}_{t+h}$ and $FE^{s,AR}_{t+h}$ is the forecast error from the autoregressive benchmark model. In a second step, we choose the root mean squared forecast error (RMSFE),

$$RMSFE^{s,k}_{h} = \sqrt{\frac{1}{T_F}\sum_{n=1}^{T_F}\left(FE^{s,k}_{t+h,n}\right)^2}\,, \tag{4.7}$$

as the loss function to get an assessment of the overall forecast accuracy of model k. The RMSFE for the AR(p) process is $RMSFE^{s,AR}_{h}$. With the ratio

$$rRMSFE^{s,k}_{h} = \frac{RMSFE^{s,k}_{h}}{RMSFE^{s,AR}_{h}}\,, \tag{4.8}$$

we can assess the performance of a single indicator forecast in comparison to the autoregressive benchmark. If the $rRMSFE$ is smaller than one, the specific indicator is performing better than the AR(p) process and therefore preferable.

[9] To understand the notation in the results section, the following example should make it clear. Imagine a factor model is abbreviated with QML1QSOLS. Then one common factor (1) is extracted via quasi-maximum likelihood (QML) from quarterly data (Q) and the forecast is generated from an OLS estimation. In this case, the factors are obtained from the set of Saxon indicators (S).

To test whether an indicator-based forecast produces lower forecast errors in comparison to the benchmark model, we apply the Diebold-Mariano test (Diebold and Mariano, 1995). Since we have a relatively small sample, we use the correction proposed by Harvey *et al.* (1997). The null hypothesis states the equality of expected forecast errors for two competing models. Or in other words, the expected difference between the forecast errors is zero,

$$H_0 : E\left[FE^{s,k}_{t+h} - FE^{s,AR}_{t+h}\right] = E\left[d^{s,k}_{t+h}\right] = 0\,. \tag{4.9}$$

Whenever the null can be rejected, the specific indicator or combination strategy produces smaller forecast errors than the autoregressive benchmark.

To conclude whether the direct or disaggregated approach performs better, we only consider the forecasts from our several pooling strategies. Therefore, we compare the forecast errors from the predictions $\widehat{y}^{agg,Pool}_{t+h}$ and $\widehat{y}^{dis,Pool}_{t+h}$ with each other. The modified Diebold-Mariano test (MDM) is used again for testing the difference in the produced forecast errors. Additionally, we apply a forecast encompassing test to check whether disaggregated forecasts have more information content than the direct approach. Granger and Newbold (1973) showed that it is insufficient to compare only the mean squared errors of competing forecasts. They suggest that a preferred forecast is not necessarily optimal and does not have to comprise all available information. This is known as "conditional efficiency". If a competing forecast has no more additional information, then the preferred forecast encompasses the competitor (see Clements and Hendry, 1993). In our setup, we examine whether the disaggregated approach ($\widehat{y}^{dis,Pool}_{t+h}$) contains more information than the direct one ($\widehat{y}^{agg,Pool}_{t+h}$). For this purpose we use a modified version proposed by Harvey *et al.* (1998). A regression of the form

$$FE^{agg,Pool}_{t+h} = \lambda\left(FE^{agg,Pool}_{t+h} - FE^{dis,Pool}_{t+h}\right) + \nu_t \tag{4.10}$$

is performed, using corrected standard errors with the method of Newey and West (1987). The null hypothesis of this test is than $H_0 : \lambda = 0$. If the tests rejects the null, the disaggregated approach contains more information beyond the direct one.

4.3 Results

We start by presenting our disaggregated results for the six different sectors: (i) agriculture, forestry and hunting; fishing, (ii) industry, (iii) construction, (iv) basic services, (v) advanced services as well as (vi) public and private services. Then we show the results for the aggregated forecasts of total GVA. Finally, we discuss the findings of the comparison between direct and disaggregated predictions.

4.3.1 Disaggregated Results

Table 4.1 shows the forecasting results for our six considered sectors. In order to show the results for our disaggregated forecasts in a compact way, we present the different sectors in one single table. We divide this table into sectoral parts, separated by new denotations of the target variable. We start with the results of agriculture, forestry and hunting; fishing. The last sector are public and private services. For every sector and forecast horizon (h) the table presents the top five indicators, pooling strategies or factor models. The $rRMSFE$ are presented in the column *Ratio*. Whenever the average forecasting errors differ significantly, asterisks are shown in the column *MDM*. To make the tables easier to read, we add abbreviations by the indicator categories, pooling strategies and factor models. Indicators from the national (German) level are denoted with (N). The abbreviations for international and regional indicators are (I) and (R) respectively. The combination strategies are indicated by (C) and factor models with (F). Abbreviations for the indicators can be found in Table A.4 in the Appendix 4.A.

Table 4.1: Disaggregated results

Target variable – qoq growth rate GVA: *Agriculture and Fishing*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
Trimmed 25 (FS)	(C)	0.986		MSFE weighted (FS)	(C)	0.953	
TRWIT	(N)	0.991		IFOBCBUENSAX	(R)	0.967	
Trimmed 25 (S)	(C)	0.991		Trimmed 25 (FS)	(C)	0.971	*
ICTOSAX	(R)	0.993	*	Trimmed 25 (S)	(C)	0.971	*
QML1QFSOLS	(F)	0.995		IFOBSBUENSAX	(R)	0.985	
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
WDAYS	(N)	0.988		IFOBECONDUR	(N)	0.958	
IFOBCCONSAX	(R)	0.988		Trimmed 25 (FS)	(C)	0.972	
IFOBCBUENSAX	(R)	0.993		MSFE weighted (FS)	(C)	0.980	

Disaggregated results – continued

IFOBSBUENSAX	(R)	0.994		Trimmed 25 (S)	(C)	0.981	
MSFE weighted (FS)	(C)	0.994		DREUROREPO	(N)	0.985	
Target variable – qoq growth rate GVA: *Industry*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
Trimmed 25 (FS)	(C)	0.849	**	WTCHEM	(N)	0.843	*
IFOBCMANSAX	(R)	0.849		Trimmed 25 (FS)	(C)	0.882	**
IFOBCCAPSAX	(R)	0.851		MSFE weighted (FS)	(C)	0.885	***
MSFE weighted (FS)	(C)	0.859	***	NOMANINTD	(N)	0.889	*
Trimmed 25 (S)	(C)	0.863	**	Trimmed 25 (S)	(C)	0.890	**
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
Trimmed 25 (FS)	(C)	0.909	**	IFOEOARS	(N)	0.888	*
IPCONG	(N)	0.909		MSFE weighted (FS)	(C)	0.912	***
Trimmed 25 (S)	(C)	0.919	**	Trimmed 25 (FS)	(C)	0.919	*
IFOBERS	(N)	0.921	*	IFOBERS	(N)	0.924	
MSFE weighted (FS)	(C)	0.922	***	YLFBOML	(N)	0.929	
Target variable – qoq growth rate GVA: *Construction*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
IFOEMPECONSAX	(R)	0.844		MSFE weighted (FS)	(C)	0.909	***
IFOBSCONSAX	(R)	0.867	*	Trimmed 25 (FS)	(C)	0.921	***
IFOBCBUENSAX	(R)	0.888		IFOBEFBTSAX	(R)	0.927	**
MSFE weighted (FS)	(C)	0.889	***	Trimmed 25 (S)	(C)	0.931	***
Trimmed 25 (FS)	(C)	0.900	***	HCTOSAX	(R)	0.958	*
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
MSFE weighted (FS)	(C)	0.892	***	MSFE weighted (FS)	(C)	0.931	**
Trimmed 25 (FS)	(C)	0.927	***	Trimmed 25 (FS)	(C)	0.943	*
Trimmed 25 (S)	(C)	0.943	***	WTSLGF	(N)	0.949	
TOCON	(N)	0.946		Trimmed 25 (S)	(C)	0.963	
GFKSE	(N)	0.948	**	TOCONNDURF	(N)	0.968	
Target variable – qoq growth rate GVA: *Basic Services*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
NOVEMF	(N)	0.947		MSFE weighted (FS)	(C)	0.880	**
MSFE weighted (FS)	(C)	0.949	***	Trimmed 25 (FS)	(C)	0.931	***
Trimmed 25 (FS)	(C)	0.950	***	Trimmed 25 (S)	(C)	0.939	***
PCNOSAX	(R)	0.958	**	EUBSSSCI	(N)	0.939	
Trimmed 25 (S)	(C)	0.965	***	IFOBCMOTSAX	(R)	0.946	
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
MSFE weighted (FS)	(C)	0.922	***	PCWHSAX	(R)	0.891	*
Trimmed 25 (FS)	(C)	0.824	**	MSFE weighted (FS)	(C)	0.918	***
EUBSSSCI	(N)	0.932		Trimmed 25 (FS)	(C)	0.945	***
Trimmed 25 (S)	(C)	0.936	**	Trimmed 25 (S)	(C)	0.951	***
IFOOOHCONSAX	(R)	0.954		NOMANCAPD	(N)	0.954	

Disaggregated results – continued

Target variable – qoq growth rate GVA: *Advanced Services*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
MSFE weighted (FS)	(C)	0.659	**	MSFE weighted (FS)	(C)	0.608	**
Trimmed 25 (FS)	(C)	0.826	**	Trimmed 25 (FS)	(C)	0.848	*
Trimmed 25 (S)	(C)	0.841	***	Trimmed 25 (S)	(C)	0.868	*
DJESI50	(I)	0.856	*	Trimmed 50 (FS)	(C)	0.902	*
SPUSSPI	(I)	0.884	*	SPUSSPI	(I)	0.916	
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
MSFE weighted (FS)	(C)	0.649	**	MSFE weighted (FS)	(C)	0.595	**
GFKSE	(N)	0.849		Trimmed 25 (FS)	(C)	0.839	
Trimmed 25 (FS)	(C)	0.863		Trimmed 25 (S)	(C)	0.857	
GFKIE	(N)	0.866		IFOBCCONNDURSAX	(R)	0.882	
Trimmed 25 (S)	(C)	0.890		ZEWES	(N)	0.885	**

Target variable – qoq growth rate GVA: *Public and Private Services*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
QML1QFSOLS	(F)	0.943		MSFE weighted (FS)	(C)	0.796	***
Trimmed 25 (FS)	(C)	0.958	***	Trimmed 25 (FS)	(C)	0.880	***
MSFE weighted (FS)	(C)	0.960		Trimmed 25 (S)	(C)	0.890	***
Trimmed 25 (S)	(C)	0.963	***	Trimmed 50 (FS)	(C)	0.931	***
M2MS	(N)	0.982		QML1QSOLS	(F)	0.932	
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
MSFE weighted (FS)	(C)	0.672	***	MSFE weighted (FS)	(C)	0.657	***
Trimmed 25 (FS)	(C)	0.835	***	Trimmed 25 (FS)	(C)	0.852	***
Trimmed 25 (S)	(C)	0.839	***	Trimmed 25 (S)	(C)	0.856	***
Trimmed 50 (S)	(C)	0.890	**	MSFE weighted (S)	(C)	0.896	**
Trimmed 50 (FS)	(C)	0.898	**	Trimmed 50 (FS)	(C)	0.898	**

Note: This Table reports the best five indicators due to the smallest rRMSFE for single indicator forecasts, pooling or factor model for every sector and horizon. MDM presents significance due to the modified Diebold-Mariano test. *Abbreviations:* FS: Full Sample, S: Saxony and GVA: gross value added. (I) international, (N) national, (R) regional indicators, (C) combinations and (F) factor models. Table A.4 in the Appendix 4.A shows the abbreviations used for the different indicators. ***, ** and * indicates significant smaller forecast errors at the 1%, 5% and 10% level. *Source:* authors' calculations.

In general, it is possible to forecast GVA more accurately than the autoregressive benchmark model. This holds for every forecasting horizon. But there exists a large heterogeneity in forecast accuracy between the sectors. Indicators from each level (international, national and regional) are able to beat the AR process. In the short-term ($h = 1$), forecasting signals predominantly come from regional (R) or international (I) indicators, whereas national (N) ones are important for medium- and long-term predictions (see $h = 2, 3, 4$). As we can conclude from the table, the forecasting performance of different pooling strategies is overwhelming. For all sectors and forecasting horizons, at least one forecast outcome from pooling is within the top five. Mainly MSFE weights or trimming (25% or 50% either with the full sample or only with regional indicators) produce significantly lower forecast errors than the autoregressive benchmark. In comparison to that, factor models are not that competitive at all. This class of models produce lower forecast errors than the benchmark only in some cases, but are not able, with some exceptions, to reach a higher forecast accuracy than indicator models or pooling. Since the results differ notably between the sectors, we will briefly discuss sectoral results subsequently.[10]

The improvement of forecast accuracy with indicator-based models for the Saxon *Agricultural Sector* is only minor, as the results for GVA in Table 4.1 suggest. We have ratios which are smaller than one, but in most cases, forecast errors from indicators or pooling are not statistically different from those of the autoregressive benchmark. International indicators are negligible for this sector. The best performance have regional indicators or pooling strategies (MSFE weighted or trimming). Factor models are only in the top five in the short forecasting horizon. However, the improvement against the AR process is not very large.

For the Saxon *Industrial Sector*, regional and national indicators are important for predicting GVA one quarter ahead (see $h = 1$ for GVA industry). International indicators are able to forecast industrial GVA in Saxony for all forecasting horizons better than the benchmark. Considering pooling, we see that trimming (25%) and MSFE weights significantly beat the AR(p) process. Factor models show no significant improvement at all. A closer look reveals that regional surveys send important forecast signals. For example, the Ifo business climate for Saxon manufacturing (IFOBCMANSAX, $rRMSFE = 0.849$) or the Ifo business expectations in the manufacturing sector (IFOBEMANSAX, $rRMSFE = 0.889$) produce lower forecast errors in comparison to the autoregressive benchmark. Macroeconomic variables such as domestic new orders of German intermediate good producers (NOMANINTD) or domestic turnovers from German capital goods producers significantly improve forecast accuracy. These results are straightforward, because the Saxon manufacturing sector is dominated by intermediate and capital goods producers. Approximately 82% of total turnovers in 2011 were achieved by firms from these two main groups, whereas capital goods producer have the highest share (45%) of total turnovers.

[10] Detailed results for all sectors are available upon request.

The third part of Table 4.1 shows the results for the Saxon *Construction Sector*. As for the agricultural sector, regional and national indicators yield the best forecasting results for construction. In the short-term, regional indicators produce the lowest forecast errors. National indicators are more important for long-term predictions. In contrast, international indicators are more or less negligible. This result is not surprising, because construction firms mainly operate on domestic markets. As we could see from the manufacturing sector, pooling (trimming 25% and MSFE weights) is also favorable to forecast GVA for the Saxon construction sector. In addition to these more general results, there are some specific indicators that have to be highlighted. Regional survey indicators such as the Ifo assessment of the business situation for the Saxon construction sector (IFOBSCONSAX, $rRMSFE = 0.867$) or the Ifo business climate either for building engineering or civil engineering (IFOBCBUENSAX, IFOBCCIENSAX) have a higher forecast accuracy than the autoregressive benchmark model. Turnovers from housing construction in Saxony, with a share of approximately 9% of all regional turnovers, significantly produce lower forecast errors.

As for construction, regional and national indicators produce the lowest forecast errors in *Basic Services*; international indicators do not play a role. These results are in line with the focus of this sector, because basic services are predominantly traded in a certain region. Gross value added in retail trade, tourism or restaurants is mainly generated by regional demand. Survey indicators obtained from regional or national business surveys (Ifo and European Commission) are again important for the prediction of GVA in this aggregated sector (see, e.g., IFOBCMOTSAX). These findings are also reflected in forecast accuracy of macroeconomic variables. For example, new orders from public (PCNOSAX) and industrial construction in Saxony or domestic new orders from German capital goods producers (NOMANCAPD) produce lower forecast errors in comparison to the autoregressive benchmark. Wholesale and retail trade as well as the transport sector react with a time lag to the development in manufacturing and construction. Since GVA in basic services is mainly generated by regional demand, consumer surveys should perform really well. The national indicators obtained by the GfK significantly beat the autoregressive benchmark.

Advanced Services comprise the sectors financial intermediation, real estate, renting and business activities. Therefore, credit institutes as well as research and development are part of this aggregate. The best forecasting results are observed for advanced services. Here, we are able to produce approximately 40% lower forecast errors than the autoregressive benchmark model. These results are obtained with MSFE-weighted combination approaches. Another result is the importance of international and national indicators for this sector. This importance is described by two reasons. First, regional credit institutes and other services highly depend on decisions of the European Central Bank (ECB) or the Central Bank of Germany (DB). This is why, e.g., financial indicators such as money supply produce lower forecast errors than the AR(p) process. Second, regional indicators for different sub-sectors are missing. However, regional survey results from the Saxon manufacturing sector have a good forecasting performance. Since business activities such as tax or business consultancy depend on the development in the manufacturing sector with a specific time lag, indicators from the industrial sector have important forecasting signals. In addition, consumer surveys have good forecasting properties. Saving or income expectations of private households can significantly increase forecast accuracy. A reason for this result is the fact that regional credit institutes (e.g., saving banks) mostly lend money to private persons, *inter alia* (see German Council of Economic Experts, 2008).

Our last aggregate is *Public and Private Services*. This is the only sector in our sample, where factor models show the lowest forecast errors in comparison to the benchmark. But this result only holds for the short-term. Forecast accuracy for this sector can also significantly be improved by pooling. Almost all weighting schemes, either for the full sample or only with Saxon indicators, produce lower forecast errors than the autoregressive benchmark model. There is no indicator (international, national or regional) which beats the forecasting outcome of pooling. Especially in the medium- and long-term ($h = 3, 4$), no indicator is in the top 10. The reason for this is that there are no indicators available for this sector. Only consumer surveys produce lower forecast errors than the autoregressive process for public and private services. This result is straightforward because GVA of clubs, culture, sports and education are part of this sector and demand for these services is mainly generated by private households.

4.3.2 Aggregated Results

Our results for total GVA are presented in Table 4.2. The structure of this table is the same as for our disaggregated results.

Table 4.2: Aggregated results

Target variable – qoq growth rate GVA: *total*							
h=1				**h=2**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
IFOBEWTSAX	(R)	0.858	*	GOVBY	(N)	0.912	
MSFE weighted (FS)	(C)	0.869	***	YLFBOML	(N)	0.919	*
Trimmed 25 (FS)	(C)	0.886	**	IFOEOARS	(N)	0.922	
Trimmed 25 (S)	(C)	0.889	**	MSFE weighted (FS)	(C)	0.924	***
IFOBCITSAX	(R)	0.921		WTCHEM	(N)	0.933	
h=3				**h=4**			
Indicator or strategy	Abbrev.	Ratio	MDM	Indicator or strategy	Abbrev.	Ratio	MDM
MSFE weighted (FS)	(C)	0.902	**	MSFE weighted (FS)	(C)	0.895	*
IFOEOARS	(N)	0.928		Trimmed 25 (FS)	(C)	0.943	***
Trimmed 25 (FS)	(C)	0.935	***	IFOBERSSAX	(R)	0.951	
Trimmed 25 (S)	(C)	0.949	**	ICTOSAX	(R)	0.956	
GOVBY	(N)	0.968		Trimmed 25 (S)	(C)	0.961	*

Note: This Table reports the best five indicators due to the smallest rRMSFE for single indicator forecasts, pooling or factor models for total GVA and every horizon. MDM presents significance due to the modified Diebold-Mariano test. *Abbreviations:* FS: Full Sample, S: Saxony and GVA: gross value added. (I) international, (N) national, (R) regional indicators, (C) combinations and (F) factor models. Table A.4 in the Appendix 4.A shows the abbreviations used for the different indicators. ***, ** and * indicates significant smaller forecast errors at the 1%, 5% and 10% level. *Source:* authors' calculations.

We are able to beat a simple autoregressive benchmark model for all forecast horizons. In the short- and long-term, especially regional indicators and pooling lead to a higher forecast accuracy than the AR(p) process. The medium-term is dominated by national indicators and combination strategies. An important leading indicator,[11] namely the Ifo business climate for industry and trade in Saxony (IFOBCITSAX), is within the top five in the short-term forecasts. As for the disaggregated results, MSFE-weights or trimming (25% and 50%), either for the full set of indicators or the Saxon sample, perform best within our considered pooling strategies. Our results are in line with the existing pooling literature. The improvement of factor models is negligible. In the end, the aggregated results are perfectly in line with those of Lehmann and Wohlrabe (2015). The top five indicators shown in Table 4.2 can be found within the top 20 for Saxon GDP. Whereas the ranking and the ratios of the indicators or combination strategies differ between the two studies, all qualitative results (e.g., that factor models are not that competitive) remain the same. The differences between our results and those found by Lehmann and Wohlrabe (2015) are explained by the fact that we consider gross value added instead of gross domestic product.

[11] See Abberger and Wohlrabe (2006) for a recent survey for Germany. For an analysis for Saxony, see Lehmann *et al.* (2010).

4.3.3 Comparison of the two Approaches

This section presents the comparison of our results from the aggregated and the disaggregated approach. Table 4.3 shows the $rRMSFE$ of $\widehat{y}_{t+h}^{dis,Pool}$ and $\widehat{y}_{t+h}^{agg,Pool}$ for our different forecast horizons and pooling techniques. The structure of Table 4.3 differs in several ways from the tables shown in the former sections. First, we present the ratios for all considered combination approaches either for the whole sample of indicators (FS) or for the Saxon indicators (S) only. This means that we combine either the forecast outcomes of all indicators with each other or use forecasts produced with Saxon indicators. Second, columns two till four present the results for each of our four forecasting horizons. Third, the presented $rRMSFE$ are always calculated as follows: $RMSFE^{dis,Pool}/RMSFE^{agg,Pool}$. So we always make a pairwise comparison (e.g., $RMSFE^{dis,Mean}/RMSFE^{agg,Mean}$). A ratio smaller than one means that the disaggregated approach is favorable in comparison to a direct forecast of Saxon GVA. Fourth, significance due to the MDM and the forecast encompassing test is separated by asterisks ($*$) and daggers (†). Asterisks indicate that a disaggregated forecast produce lower forecast errors then an aggregated one and daggers show that disaggregated predictions comprise more information beyond a direct forecast of total GVA.

As our forecast outcome shows, a disaggregated approach is preferable for short-term predictions. Nearly all combination strategies (with all indicators as well as only with Saxon ones) significantly beat the direct approach. For medium- and long-term predictions, a direct approach produces lower forecast errors in comparison to disaggregated predictions. However, the ratios are not statistically significant. The forecast encompassing tests clearly state that there is an information gain from disaggregated forecasts in comparison to direct ones for all considered pooling techniques in the short-term. We can conclude that direct predictions of GVA significantly neglect information. Our results are in line with the existing literature. Drechsel and Scheufele (2012b) find for Germany that the supply-side approach produces in some cases lower forecasts errors. This holds especially for the short-term. We think that the disaggregated approach loses its power against the direct one in the medium- and long-term since many indicators (e.g., surveys or new orders) only have a lead of up to three months or provide forecasting signals contemporaneously. Whenever the forecast horizon becomes larger, the performance of those indicators for sector-specific forecasts is negligible. We leave this for future research.

Table 4.3: Comparison of aggregated and disaggregated results

Target variable – qoq growth rate GVA: *total*				
Strategy	**h=1**	**h=2**	**h=3**	**h=4**
Mean (FS)	0.948*,††	1.029	1.039	1.035
Median (FS)	0.948*,††	1.040	1.044	1.045
BIC (FS)	0.947*,††	1.028	1.039	1.033
R^2 (FS)	0.947*,††	1.029	1.039	1.034
Trimmed 25 (FS)	0.918**,†††	1.025	1.036	1.028
Trimmed 50 (FS)	0.926**,††	1.038	1.080	1.041
Trimmed 75 (FS)	0.937*,††	1.039	1.082	1.046
MSFE weighted (FS)	0.948††	1.026	1.081	1.040
Mean (S)	0.943*,††	1.036	1.046	1.048
Median (S)	0.958†	1.048	1.058	1.063
BIC (S)	0.942*,††	1.037	1.044	1.048
R^2 (S)	0.943*,††	1.036	1.045	1.048
Trimmed 25 (S)	0.928**,††	1.023	1.038	1.024
Trimmed 50 (S)	0.928**,††	1.023	1.037	1.038
Trimmed 75 (S)	0.939*,††	1.026	1.038	1.042
MSFE weighted (S)	0.949†	1.034	1.044	1.038

Note: This table compares the disaggregated results of our different combination strategies with those of the aggregated ones. *Abbreviations:* FS: Full Sample, S: Saxony and GVA: gross value added. ***, ** and * indicates significance (MDM) at the 1%, 5% and 10% level. †††, †† and † indicates significance due to the forecast encompassing test at the 1%, 5% and 10% level. *Source:* authors' calculations.

The pooling results suggest that it makes no difference whether to use the whole set of indicators (FS) or just the one restricted to Saxon indicators (S). We find no systematic pattern so that either FS or S lead to a higher forecast accuracy for the disaggregated approach. This holds for all combination strategies and forecast horizons. However, out-of-sample weighted combination strategies perform better than in-sample weights or simple averages. Using a trimmed mean for the 25% best performing indicators in the full sample, a disaggregated approach produces on average nearly 8% smaller forecast errors than the direct approach (Trimmed 25 (FS), $rMSFE = 0.918$).

For short-term predictions we can conclude that disaggregated forecasts have a higher forecast accuracy than direct ones. Since we are able to predict sectoral GVA with different indicators better than an autoregressive benchmark model, practitioners and forecasters should use the available information to forecast the state of the economy in the short-term. For long-term predictions, they should predict the whole aggregate directly in addition to sectoral forecasts.

4.4 Conclusion

With our empirical setup, we are able to predict sectoral GVA (e.g., for manufacturing) more accurately than a benchmark model. But forecast accuracy significantly differs between different sectors of the economy. These results are important for regional policy-makers, practitioners or regional credit institutes. We are able to make the state of the economy more tangible. If external shocks only hit a few sectors, regional policy-makers can systematically align their future policy. For credit institutes it is important to know how different sectors will develop in the near future. Especially for granting credit, such information are necessary.

All in all, we find that for short-term predictions (one quarter ahead) disaggregated forecasts for GVA are preferable in comparison to direct ones. The resulting forecast errors could be reduced by about 8% on average. This outcome is straightforward, because we find that different indicators are linked to sectoral GVA even stronger than to total outcome. To predict GVA in the medium- (two and three quarters) and long-term (four quarters), a direct approach for total GVA produces lower forecast errors.

Regional indicators (e.g., business surveys) produce significantly lower forecast errors than the benchmark, especially in the short-term. This result may explain, why the weighted sum of disaggregated predictions is more accurate than a direct forecast of total GVA, since the information surplus of these regional indicators is most present in the short-term. National and international indicators are more important in the medium- and long-term. Whenever it is possible to use regional indicators, forecasters should include those information in their analysis. Pooling performs really well for the different sectors and total GVA too. Factor models are not that competitive at the regional level.

Our analysis has shown that indicator-based sectoral forecasts produce smaller forecast errors and that forecast accuracy of total GVA can be improved by disaggregated forecasts. This gives a more detailed picture of the development of the economy and makes economic policy more assessable. Due to data limitations, our paper focuses exclusively on the Free State of Saxony. To the best of our knowledge, this is the only German state for which quarterly national accounts for different sectors are available. However, we think that such forecast improvements can be found for other German states or other regions too. If official statistics are able to provide quarterly data at the regional level, then such an analysis could be extended to other regional units. In the end, the results of our analysis suggest that forecasts for total German GDP could be improved by aggregation of state level GDP predictions. We leave this for future research.

Appendix 4.A

Table A.4: Indicators, abbreviations and transformations

Abbreviation	Indicator	Transform.
Dependent Variables		
GVAAGFISAX	gross value added (GVA): agriculture, hunting and forestry; fishing, Saxony	1
GVAINDSAX	GVA: industry, Saxony	1
GVACONSAX	GVA: construction, Saxony	1
GVABSSAX	GVA: basic services, Saxony	1
GVAASSAX	GVA: advanced services, Saxony	1
GVAPPSSAX	GVA: public and private services, Saxony	1
Macroeconomic Variables		
IPTOT	industrial production (IP): total (incl. construction)	1
IPCON	IP construction: total	1
IPENY	IP energy supply: total	1
IPMQU	IP manufacturing: mining and quarrying	1
IPMAN	IP manufacturing: total	1
IPCAP	IP manufacturing: capital goods	1
IPCONDUR	IP manufacturing: consumer durables	1
IPCONNDUR	IP manufacturing: consumer non-durables	1
IPINT	IP manufacturing: intermediate goods	1
IPCONG	IP manufacturing: consumer goods	1
IPCHEM	IP manufacturing: chemicals	1
IPMET	IP manufacturing: basic metals	1
IPMECH	IP manufacturing: mechanical engineering	1
IPMOT	IP manufacturing: motor vehicles, trailers	1
IPEGS	IP manufacturing: energy, gas etc. supply	1
IPVEM	IP manufacturing: motor vehicles, trailers etc.	1
TOCON	turn over (TO): construction	1
TOMQD	TO: mining and quarrying, domestic	1
TOMQF	TO: mining and quarrying, foreign	1
TOMAND	TO: manufacturing total, domestic	1
TOMANF	TO: manufacturing total, foreign	1
TOCAPD	TO: capital goods, domestic	1
TOCAPF	TO: capital goods, foreign	1
TOCONDURD	TO: consumer durables, domestic	1
TOCONDURF	TO: consumer durables, foreign	1
TOCONNDURD	TO: consumer non-durables, domestic	1
TOCONNDURF	TO: consumer non-durables, foreign	1
TOINTD	TO: intermediate goods, domestic	1

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	**Indicator**	**Transform.**
TOINTF	TO: intermediate goods, foreign	1
TOCONGD	TO: consumer goods, domestic	1
TOCONGF	TO: consumer goods, foreign	1
TOCEOD	TO: computer, electronic and optical products, domestic	1
TOCEOF	TO: computer, electronic and optical products, foreign	1
TOCHEMD	TO: chemicals, domestic	1
TOCHEMF	TO: chemicals, foreign	1
TOMECHD	TO: mechanical engineering, domestic	1
TOMECHF	TO: mechanical engineering, foreign	1
TOVEMD	TO: motor vehicles, trailers etc., domestic	1
TOVEMF	TO: motor vehicles, trailers etc., foreign	1
TOEGSD	TO: energy, gas etc. supply, domestic	1
TOEGSF	TO: energy, gas etc. supply, foreign	1
NOCON	new orders (NO): construction	1
NOMANTOT	NO: manufacturing total	1
NOMANTOTD	NO: manufacturing total, domestic	1
NOMANTOTF	NO: manufacturing total, foreign	1
NOMANCAP	NO: capital goods	1
NOMANCAPD	NO: capital goods, domestic	1
NOMANCAPF	NO: capital goods, foreign	1
NOMANCONG	NO: consumer goods	1
NOMANCONGD	NO: consumer goods, domestic	1
NOMANCONGF	NO: consumer goods, foreign	1
NOMANINT	NO: intermediate goods	1
NOMANINTD	NO: intermediate goods, domestic	1
NOMANINTF	NO: intermediate goods, foreign	1
NOCHEMD	NO: chemicals, domestic	1
NOCHEMF	NO: chemicals, foreign	1
NOMECHD	NO: mechanical engineering, domestic	1
NOMECHF	NO: mechanical engineering, foreign	1
NOVEMD	NO: motor vehicles, trailers etc., domestic	1
NOVEMF	NO: motor vehicles, trailers etc., foreign	1
NOCEOD	NO: computer, electronic and optical products, domestic	1
NOCEOF	NO: computer, electronic and optical products, foreign	1
CONEMPL	construction: total employment	1
CONTOT	construction: permits issued, total	1
CONHOPE	construction: housing permits issued for building	1
CONNREPE	construction: non-residential permits	1
CONBPGTOT	construction: building permits granted, total	1
CONBPGHO	construction: building permits granted, new homes	1
CONBPGNRE	construction: building permits granted, non-residentials	1
CONHW	construction: hours worked	1
WTEXMV	wholesale trade (WT): total (excl. motor vehicles)	1
WTCLFW	WT: clothing and footwear	1
WTCHEM	WT: chemicals	1
WTCONMA	WT: construction machinery	1
WTSLGF	WT: solid, liquid, gaseous fuels etc.	1
WTEMPL	WT: total employment	1

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	Indicator	Transform.
RSEXC	retail sales (RS): total (excl. cars)	1
NRTOT	new registrations (NR): all vehicles	1
NRCARS	NR: cars	1
NRHT	NR: heavy trucks	1
EXVOL	exports: volume index, basis 2005	1
IMVOL	imports: volume index, basis 2005	1
UNPTOT	unemployed persons (UNP): total, % of civilian labor	2
EMPLRCTOT	employed persons (EMPL): residence concept, total	1
EMPLWPCTOT	EMPL: work-place concept, total	1
WDAYS	working days: total	1
VACTOT	vacancies: total	1
MANHW	manufacturing: hours worked (excl. construction)	1
TREUCD	tax revenues (TR): EU customs duties	1
TRITTOT	TR: income taxes, total	1
TRVAT	TR: value added tax	1
TRVATIM	TR: value added tax on imports	1
TRVATTOT	TR: value added tax, total	1
TRWIT	TR: wage income tax	1
Finance		
MMRDTD	money market rate (MMR): day-to-day, monthly average	2
MMRTM	MMR: three-month, monthly average	2
DREUROREPO	discount rate - short term euro repo rate	2
GOVBY	long term government bond yield, 9-10 years	2
YFTBOPB	yields on fully taxed bonds outstanding (YFTBO): public bonds	2
YFTBOCB	YFTBO: corporate bonds	2
YLFBOMS	yields on listed fed. bonds outstand. mat. (YLFBOM): 3-5 years	2
YLFBOML	yields on listed fed. bonds outstand. mat. (YLFBOM): 5-8 years	2
TSPI	term spread (TS): 10 years, policy inst	0
TSDAY	TS: 10 years, 1Day	0
TSMTH	TS: 10 years, 3Month	0
SPRDAYPR	1Day - policy rates	0
SPRCTB	corporate - treasury bond	0
GPC23CPI	german price competition: 23 industrialized countries, basis: cpi	1
DAXSPI	DAX share price index	1
NEER	nominal effective exchange rate	1
VDAXNVI	VDAX: new volatility index, price index	2
VDAXOVI	VDAX: old volatility index, price index	2
M1OD	M1, overnight deposits	1
M2MS	M2, money supply	1
M3MS	M3, money supply	1
EMMSM1EP	EM money supply: M1, ep	1
EMMSM1F	EM money supply: M1, flows	2
EMMSM2M1I	EM money supply: M2-M1, index	1
EMMSM2M1F	EM money supply: M2-M1, flows	2
EMMSM3M2EP	EM money supply: M3-M2, ep	1
EMMSM3M2F	EM money supply: M3-M2, flows	2
BLDNB	bank lending to domestic non-banks, short term	1

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	Indicator	Transform.
BLDEI	banl lending to enterprises and individuals, short term	1
TDDE	time deposits of domestic enterprises	1
SDDE	saving deposits of domestic enterprises	1
Prices		
CPI	consumer price index	1
CPIEE	consumer price index (excl. energy)	1
HWWAPITOT	HWWA index of world market prices: eurozone, total	1
HWWAPIEY	HWWA index of world market prices: eurozone, energy	1
HWWAPIEEY	HWWA index of world market prices: eurozone, excl. energy	1
OIL	oil prices, euro per barrel	1
OILUK	brent oil price, UK average	1
LGP	London gold price, per US $	1
IMPI	import price index	1
EXPI	export price index	1
WTPI	wholesale trade price index, 1975=100	1
PPI	producer price index	1
Wages		
WSLTOTHOU	wage and salary level (WSL): overall economy, basis: hours	1
WSLTOTMTH	WSL: overall economy, basis: monthly	1
WSLMANHOU	WSL: manufacturing, basis: hours	1
WSLMANMTH	WSL: manufacturing, basis: monthly	1
Surveys		
ZEWPS	ZEW: present economic situation	0
ZEWES	ZEW: economic sentiment indicator	0
IFOBCIT	ifo business climate industry and trade, index	0
IFOBEIT	ifo: business expextations industry and trade, index	0
IFOBSIT	ifo: assessment of business situation industry and trade, index	0
IFOBCMAN	ifo: business climate manufacturing, index	0
IFOBEMAN	ifo: business expextations manufacturing, index	0
IFOBSMAN	ifo: assessment of business situation manufacturing, index	0
IFOEXEMAN	ifo: export expectations next 3 months manufacturing, balance	0
IFOOOHMAN	ifo: orders on hand manufacturing, balance	0
IFOFOOHMAN	ifo: foreign orders on hand manufacturing, balance	0
IFOIOFGMAN	ifo: inventory of finished goods manufacturing, balance	0
IFOBCCAP	ifo: business climate capital goods, balance	0
IFOBECAP	ifo: business expectations capital goods, balance	0
IFOBSCAP	ifo: assessment of business situation capital goods, balance	0
IFOBCCONDUR	ifo: business climate consumer durables, balance	0
IFOBECONDUR	ifo: business expectations consumer durables, balance	0
IFOBSCONDUR	ifo: assessment of business situation consumer durables, balance	0
IFOBCCONNDUR	ifo: business climate consumer non-durables, balance	0
IFOBECONNDUR	ifo: business expectations consumer non-durables, balance	0
IFOBSCONNDUR	ifo: assessment of business sit. consumer non-durables, balance	0
IFOBCINT	ifo: business climate intermediate goods, balance	0
IFOBEINT	ifo: business expectations intermediate goods, balance	0
IFOBSINT	ifo: assessment of business situation intermediate goods, balance	0

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	Indicator	Transform.
IFOBCCONG	ifo: business climate consumer goods, balance	0
IFOBECONG	ifo: business expectations consumer goods, balance	0
IFOBSCONG	ifo: assessment of business situation consumer goods, balance	0
IFOBCCON	ifo: business climate construction, index	0
IFOBECON	ifo: business expectations construction, index	0
IFOBSCON	ifo: assessment of business situation construction, index	0
IFOOOHCON	ifo: orders on hand construction, balacne	0
IFOUNFWCON	ifo: unfavourable weather situation	0
IFOBCWT	ifo business climate wholesale trade, index	0
IFOBEWT	ifo: business expextations wholesale trade, index	0
IFOBSWT	ifo: assessment of business situation wholesale trade, index	0
IFOAOIWT	ifo: assessment of inventories wholesale trade, balance	0
IFOEOAWT	ifo: expect. with regard to order activity next 3 months WT, balance	0
IFOBCRS	ifo business climate retail sales, index	0
IFOBERS	ifo: business expextations retail sales, index	0
IFOAOIRS	ifo: assessment of inventories retail sales, balance	0
IFOEOARS	ifo: expect. with regard to order activity next 3 months RS, balance	0
GFKBCE	GfK consumer survey (GfK): business cycle expectations	0
GFKIE	GfK: income expectations	0
GFKWTB	GfK: willingness to buy	0
GFKPL	GfK: prices over the last 12 months	0
GFKPE	GfK: prices over the next 12 months	0
GFKUE	GfK: unemployment situation over next 12 months	0
GFKFSL	GfK: financial situation over the last 12 months	0
GFKFSE	GfK: financial situation over the next 12 months	0
GFKESL	GfK: economic situation over the last 12 months	0
GFKESE	GfK: economic situation over the next 12 months	0
GFKMPP	GfK: major purchases at present	0
GFKMPE	GfK: major purchases over the next 12 months	0
GFKSP	GfK: savings at present	0
GFKSE	GfK: savings over the next 12 months	0
GFKCCI	GfK: consumer confidence, index	0
GFKCCC	GfK: consumer confidence climate, balance	0
GFKCCIN	GfK: consumer confidence indicator	0
EUCSUE	EU consumer survey (EUCS): unemploy. expect. over next 12 months	0
EUCSFSP	EUCS: statement on financial situation	0
EUCSCCI	EUCS: consumer confidence indicator	0
EUCSESI	EUCS: economic sentiment indicator	0
EUBSPTIND	EU business survey (EUBS): prod. trends recent month, industry	0
EUBSOBLIND	EUBS: assessment of order-book levels, industry	0
EUBSEXOBLIND	EUBS: assessment of export oder-books level, industry	0
EUBSSFGIND	EUBS: assessment of stocks of finished products, industry	0
EUBSPEIND	EUBS: production expectations for the month ahead, industry	0
EUBSSPEIND	EUBS: selling price expectations for the month ahead, industry	0
EUBSEMPEIND	EUBS: employment expectations for the month ahead, industry	0
EUBSINDCI	EUBS: industrial confidence indicator	0
EUBSSSCI	EUBS: service sector confidence indicator	0
EUBSRTCI	EUBS: retail trade confidence indicator	0

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	Indicator	Transform.
EUBSCONCI	EUBS: construction confidence indicator	0
COMBAEB	Commerzbank EarlyBird	0
International		
BGBIS	Belgium business indicator survey, whole economy	0
BGBISMAN	Belgium business indicator survey, manufacturing (not smoothed)	0
UMCS	University of Michigan US consumer sentiment, expectations	0
USISMP	US ISM production	0
EUCSFRESI	EUCS: economic sentiment indicator, France	0
EUCSESESI	EUCS: economic sentiment indicator, Spain	0
EUCSPOESI	EUCS: economic sentiment indicator, Poland	0
EUCSCZESI	EUCS: economic sentiment indicator, Czech Republic	0
EUCSITESI	EUCS: economic sentiment indicator, Italy	0
EUCSUKESI	EUCS: economic sentiment indicator, United Kingdom	0
DJESI50	EM Dow Jones EUROSTOXX index, benchmark 50	1
DJIPRI	Dow Jones industrials, price index	1
SPUSSPI	Standard & Poor´s 500 stock price index	1
GOVBYUK	government bond yield long term, United Kingdom	2
GOVBYUS	government bond yield long term, United States	2
USIPTOT	IP: United States, total	1
CLIAA	OECD Composite Leading Indicator (CLI): OECD, amplitude adjusted	0
CLITR	CLI: OECD, trend restored	1
CLINORM	CLI: OECD, normalised	0
CLIASAA	CLI: Asia, amplitude adjusted	0
CLIASTR	CLI: Asia, trend restored	1
CLIASNORM	CLI: Asia, normalised	0
CLICAA	CLI: China, amplitude adjusted	0
CLICTR	CLI: China, trend restored	1
CLICNORM	CLI: China, normalised	0
CLIEUAA	CLI: Euro Area, amplitude adjusted	0
CLIEUTR	CLI: Euro Area, trend restored	1
CLIEUNORM	CLI: Euro Area, normalised	0
CLIUSAA	CLI: United States, amplitude adjusted	0
CLIUSTR	CLI: United States, trend restored	1
CLIUSNORM	CLI: United States, normalised	0
ECRTE	Euro-Coin real time estimates	0
Regional – Free State of Saxony		
IFOBCITSAX	ifo business climate industry and trade Saxony, balance	0
IFOBEITSAX	ifo: business expextations industry and trade Saxony, balance	0
IFOBSITSAX	ifo: assessment of business sit. indus. and trade Saxony, balance	0
IFOBCMANSAX	ifo: business climate manufacturing Saxony, balance	0
IFOBEMANSAX	ifo: business expextations manufacturing Saxony, balance	0
IFOBSMANSAX	ifo: assessment of business sit. manufacturing Saxony, balance	0
IFOBCCONSAX	ifo: business climate construction Saxony, balance	0
IFOBECONSAX	ifo: business expectations construction Saxony, balance	0
IFOBSCONSAX	ifo: assessment of business situation construction Saxony, balance	0
IFOEMPECONSAX	ifo: employment expect. over next 3 months constr. Saxony, balance	0
IFOBCWTSAX	ifo business climate wholesale trade Saxony, balance	0

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	Indicator	Transform.
IFOBEWTSAX	ifo: business expextations wholesale trade Saxony, balance	0
IFOBSWTSAX	ifo: assessment of business situation wholesale trade Saxony, balance	0
IFOEMPEWTSAX	ifo: employment expect. over next 3 months WT Saxony, balance	0
IFOBCRSSAX	ifo business climate retail sales Saxony, balance	0
IFOBERSSAX	ifo: business expect. retail sales Saxony, balance	0
IFOBSRSSAX	ifo: assessment of business situation retail sales Saxony, balance	0
IFOEMPERSSAX	ifo: employment expect. over next 3 months RS Saxony, balance	0
IFOBCINTSAX	ifo business climate intermediate goods Saxony, balance	0
IFOBEINTSAX	ifo: business expextations intermediate goods Saxony, balance	0
IFOBSINTSAX	ifo: assess. of busin. sit. intermediate goods Saxony, balance	0
IFOBCCAPSAX	ifo: business climate capital goods Saxony, balance	0
IFOBECAPSAX	ifo: business expextations capital goods Saxony, balance	0
IFOBSCAPSAX	ifo: assessment of busin. sit. capital goods Saxony, balance	0
IFOBCCONDURSAX	ifo: business climate consumer durables Saxony, balance	0
IFOBECONDURSAX	ifo: business expectations consumer durables Saxony, balance	0
IFOBSCONDURSAX	ifo: assessment of business sit. consumer durables Saxony, balance	0
IFOBCCONGSAX	ifo business climate consumer goods Saxony, balance	0
IFOBECONGSAX	ifo: business expextations consumer goods Saxony, balance	0
IFOBSCONGSAX	ifo: assessment of business situation consumer goods Saxony, balance	0
IFOBCFBTSAX	ifo business climate food, beverage and tobacco Saxony, balance	0
IFOBEFBTSAX	ifo: business expect. food, beverage and tobacco Saxony, balance	0
IFOBSFBTSAX	ifo: assessment of business situation FBT Saxony, balance	0
IFOBCCHEMSAX	ifo business climate chemicals Saxony, balance	0
IFOBECHEMSAX	ifo: business expextations chemicals Saxony, balance	0
IFOBSCHEMSAX	ifo: assessment of business situation chemicals Saxony, balance	0
IFOBCMECHSAX	ifo business climate mechanical engineering Saxony, balance	0
IFOBEMECHSAX	ifo: business expextations mechanical engineering Saxony, balance	0
IFOBSMECHSAX	ifo: assessment of busin. sit. mechanical engineering Saxony, balance	0
IFOBCMOTSAX	ifo business climate motor vehicles Saxony, balance	0
IFOBEMOTSAX	ifo: business expextations motor vehicles Saxony, balance	0
IFOBSMOTSAX	ifo: assessment of business sit. motor vehicles Saxony, balance	0
IFOBCBUENSAX	ifo business climate building engineering Saxony, balance	0
IFOBEBUENSAX	ifo: business expextations building engineering Saxony, balance	0
IFOBSBUENSAX	ifo: assessment of busin. sit. building engineering Saxony, balance	0
IFOBCCIENSAX	ifo business climate civil engineering Saxony, balance	0
IFOBECIENSAX	ifo: business expextations civil engineering Saxony, balance	0
IFOBSCIENSAX	ifo: assessment of busin. sit. civil engineering Saxony, balance	0
NOMANSAXTOT	NO: manufacturing Saxony, total	1
HCNOSAX	housing construction (HC): new orders Saxony	1
HCWHSAX	HC: working hours Saxony	1
HCTOSAX	HC: turnover Saxony	1
ICNOSAX	industry construction (IC): new orders Saxony	1
ICWHSAX	IC: working hours Saxony	1
ICTOSAX	IC: turn over Saxony	1
PCNOSAX	public construction (PC): new orders Saxony	1
PCWHSAX	PC: working hours Saxony	1
PCTOSAX	PC: turn over Saxony	1
CONNOSAX	construction: new orders Saxony	1

Continued on next page...

Indicators, abbreviations and transformations – continued

Abbreviation	Indicator	Transform.
CONWHSAX	construction: working hours Saxony	1
CONFIRMSAX	construction: firms Saxony	1
CONEMPSAX	construction: employed people Saxony	1
CONFEESAX	construction: fees Saxony	1
IFOCUCONSAX	ifo: capacity utilization construction, Saxony	2
IFOOOHCONSAX	ifo: orders on hand construction, Saxony	0
TOHRSAX	TO: hotels and restaurants Saxony, total	1
CPISAX	consumer price index, Saxony	1
EXVALUESAX	exports: value, Saxony	1
IMVALUESAX	imports: value, Saxony	1

Note: 0 = three-month average in levels; 1 = three-month average and qoq growth rate; 2 = three-month average and Δ. Industry: Mining and quarrying; manufacturing; electricity, gas and water supply. Basic services: Wholesale and retail trade; hotels and restaurants; transport. Advanced services: Financial intermediation; real estate, renting and business activities. Public and private services: public administration; education; health and social work; private households. *Source:* Drechsel and Scheufele (2012b), authors' extensions and calculations.

5 Survey-based indicators vs. hard data: What improves export forecasts in Europe?

This chapter is based on the working paper by Lehmann (2015).

5.1 Motivation

When it comes to macroeconomic forecasting, the main figure noticed by the public is gross domestic product (GDP). However, from a practical point of view, economic forecasts are more than just the prediction of a single number. Most forecast suppliers, such as supra-national organizations, research institutes or banks, predict each single component of GDP (e.g. private consumption or exports) separately and merge these components together to form, from their point of view, a plausible and most likely forecast of total output. Such a disaggregated approach of forecasting GDP is also found to be preferable compared to a direct approach by the academic literature (see, among others, Angelini *et al.*, 2010; Drechsel and Scheufele, 2012b). The forecast error for GDP can thus significantly be reduced by forecasting single components such as private consumption or exports. Academics have studied forecasts of private consumption in particular (see, among others, Vosen and Schmidt, 2011). The other components are not as explicitly studied as consumption. In this paper, we focus on exports and ask whether export forecasts for a multitude of European states can be improved by either hard data, such as price and cost competitiveness measures, or by qualitative information gained from surveys.

From the demand-side calculation of GDP, exports are one of the major components. Considering that the share of exports of goods and services in total GDP rose from almost 30% in 1995 to 45% in 2013 for the EU-15, exports are one major source of the creation of business cycles, since they transfer international shocks into the domestic economy.[1] Thus, especially unbiased export forecasts can, c.p., significantly reduce forecast errors of GDP.

Only a few studies exist that focus on the improvement of export forecasts. An early attempt has been made by Baghestani (1994). He finds that survey results obtained from professional forecasters improve predictions for US net exports. In the case of Portugal, Cardoso and Duarte (2006) find that business surveys improve the forecasts for export growth. For Taiwan, standard autoregressive integrated moving average (ARIMA) models are able to improve export forecasts compared to heuristic methods (Wang *et al.*, 2011). Additionally, two German studies exist. Jannsen and Richter (2012) use a capacity utilization weighted indicator obtained from major export partners to forecast German capital goods exports. Elstner *et al.* (2013) use hard data (e.g. foreign new orders in manufacturing) as well as indicators from the Ifo business survey (e.g. Ifo export expectations) to improve forecasts for German exports. Overall, Elstner *et al.* (2013) find that survey indicators produce lower forecast errors than hard indicators do. Finally, Hanslin and Scheufele (2014) show that a weighted Purchasing Manager Index (PMI) from major trading partners improves Swiss exports more than other indicators.

Next to these country-specific studies, some contributions focus on country-aggregates. Keck *et al.* (2009) show that trade forecasts for the OECD25 can be improved by applying standard time series models in comparison to a 'naïve' prediction based on a deterministic trend. Economic theory names two major drivers of exports: relative prices and domestic demand of the importing trading partners. Thus, Ca'Zorzi and Schnatz (2010) use different measures of price and cost competitiveness to forecast extra euro-area exports and find that for a recursive estimation approach the real effective exchange rate based on the export price index outperforms the other measures as well as a 'random walk' benchmark. For the Euro area, Frale *et al.* (2010) find that survey results play an important role for export forecasts. From a global perspective, Guichard and Rusticelli (2011) show that the industrial production (IP) and Purchasing Manager Indices are able to improve world trade forecasts.

We contribute to this existing literature by creating a forecasting competition between indicators gained from hard data and different survey-based indicators for a multitude of European countries. We do not focus solely on one indicator or state, but rather analyze sixteen indicators for twenty European states and the aggregates EA-18 and

[1] Fiorito and Kollintzas (1994) find for the G7 that exports are procyclical and coincide with the business cycle of total output. Additionally, trade is an important pillar for the economic development of countries, as the empirical literature shows (see Frankel and Romer, 1999).

EU-28 in the period from 1996 to 2013. From a pseudo out-of-sample analysis and forecast encompassing tests we can conclude that survey-based indicators produce the most accurate export forecasts and cannot be beaten by hard indicators.

In general, it is common knowledge that business and consumer surveys are powerful tools for macroeconomic forecasting. However, business surveys are not free of criticism. Croux *et al.* (2005) mention that surveys are very expensive and time-consuming for both the enterprise and the consumer. This expense, in terms both time and money, should result in significant improvements of forecasts of different macroeconomic aggregates by the questions asked in the specific survey. The study by Croux *et al.* (2005) finds an improvement in industrial production forecasts through the usage of production expectations expressed by European firms. Despite the forecasting power of a survey indicator for European industrial production, the results for different macroeconomic aggregates are mixed. This leads to the conclusion by Claveria *et al.* (2007) that we actually have no definite idea why some qualitative indicators work for specific macroeconomic variables, whereas others do not. With this paper, we ask whether survey-based indicators are able to predict export growth for a multitude of European states. Additionally, our paper searches for the reasons of country differences in the forecasting performance of survey-based indicators. We find with standard regression techniques that in particular the composition of exports plays a crucial role for the forecast accuracy of soft indicators. The forecast accuracy of survey-based indicators worsens in countries with a high share in raw materials or oil exports. The opposite holds for countries with a high share in machinery and transport equipment exports. These results are underpinned by studying the impact of export diversity. It turns out that survey indicators produce, on average, lower forecast errors in countries with a higher degree of export diversification.

To evaluate the competition between soft and hard indicators to forecast export growth, the paper is organized as follows. In Section 5.2, we present the data and our empirical setup. Section 5.3 discusses our results in detail. Section 5.4 offers a conclusion.

5.2 Data and Empirical Setup

5.2.1 Data

Target Variable

Eurostat supplies comprehensive export data on a quarterly basis for all member states of the European Union plus Switzerland and Norway. These figures are comparable to each other, since they are based on consistent standards within national accounts. We use total exports, which is the sum of exports of goods and services.[2] These total export figures are measured in real terms and are seasonally adjusted by the Census X-12-ARIMA procedure. Since we are interested in growth forecasts rather than levels, we transform the export figures into year-on-year (yoy) growth rates. Our forecast experiment relies on quarterly data from 1996:01 to 2013:04 for a large sample of European states. Due to some data restrictions (e.g. missing export data or survey results), we eliminate some countries, leaving us with the following 20 European states in the sample: Austria, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Italy, Latvia, Lithuania, Luxemburg, the Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden and the United Kingdom. Additionally, we test the indicators for the aggregates EA-18 and EU-28. Descriptive statistics are available upon request.

Indicators

The European Commission (EC) provides both survey indicators and hard data. The survey-based indicators are collected within the *Joint Harmonised EU Programme of Business and Consumer Surveys* on behalf of the European Commission. The survey is harmonized across all European states. The samples in each country are representative. For the business survey, the sample comprises firms from different sectors (industry, construction, retail trade and services).[3] We concentrate on the survey results obtained from the manufacturing sector for two reasons. First, the majority of exports are goods produced in the manufacturing sector. Second, the survey in the service sectors was first

[2] The code of the corresponding time series is: *namq_exi_k*. All the data can be downloaded free of charge under *http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home*. Note that the figures used here are based on the European System of Accounts of 1995 (ESA 1995).

[3] The European Commission wants to keep the sample representative for each month. To ensure this, sample updates are necessary on occasion due to (for example) start-ups or bankruptcies. However, the samples for the business survey are very stable in each state. Additional details on the sample composition can be found in European Commission Economic and Financial Affairs (2014).

conducted in the mid-2000s, so the time series is too short for our purposes. The survey program in manufacturing is divided into monthly and quarterly questions. The most intuitive candidate to predict future export growth in a specific country is the following question, which we call export expectations (*EXEXP*): 'How do you expect your export orders to develop over the next three months?' The respondents can answer this question in three ways: (+) increase, (=) remain unchanged or (-) decrease. Since *EXEXP* is measured on a quarterly basis (as exports), no transformation is necessary. In line with the literature, we assess the forecasting power of "balances". These balances are expressed as differences between the weighted share of firms whose exports will increase and the weighted share of those that expect a decrease. The weights are based on the size of the firms (see European Commission Economic and Financial Affairs, 2014). All firms with a response "remain unchanged" are not considered.[4]

Bearing in mind that other survey indicators may also deliver important information to forecast export growth, we evaluate the following monthly indicators as well: (i) the confidence indicator in manufacturing (COF), (ii) the assessment of export order-book levels (EOBL), (iii) the assessment of order-book levels (OBL), (iv) production expectations for the month ahead (PEXP), (v) the assessment of stocks of finished products (SFP), (vi) a self-constructed capacity-based indicator in the style of the Kiel Institute for the World Economy (IfW; see Jannsen and Richter, 2012) and (vii) the economic sentiment indicator (ESI) of the whole economy. In addition, we use the consumer confidence indicator (CCOF) as a possible predictor. Since the balances of these eight additional indicators are on a monthly basis, we transform these balances with a simple three-month average to obtain quarterly data. All survey results are seasonally adjusted by the provider via the procedure DAINTIES.[5] All in all we end up with nine survey-based indicators.

Since the purpose of the paper is to create a "horse race" between survey-based indicators and hard data, we have to specify which variables are found in the category of hard data. One major driver for exports is the price and cost competitiveness of a specific country. The Department of Economic and Financial Affairs (ECFIN) at the European Commission provides price and cost competitiveness measures based on different price weights. We choose the quarterly real effective exchange rate (REER) against 37 indus-

[4] Balances are not indisputable in the existing literature; for a critical discussion, see Croux *et al.* (2005); Claveria *et al.* (2007) and the references therein.

[5] We are aware of the fact that an intensive discussion about seasonal adjustment and the forecasting properties of survey indicators exists in the academic literature. However, this issue is beyond the scope of this paper.

trial countries for each specific state in our sample. The ECFIN provides REER data based on five different price weights[6]: (i) harmonized consumer price index (HCPI), (ii) nominal unit labor costs of the total economy (ULCTOT), (iii) nominal unit wage costs in manufacturing (UWCMAN), (iv) the GDP deflator (GDPDEF) and (v) the price deflator for exports of goods and services (EXPI).[7] The discussion in Ca'Zorzi and Schnatz (2010) reveals different advantages and shortcomings of each of these five indicators (see Table 5.1 for an overview). The EXPI, in particular, has some remarkable disadvantages, such as heavy data revisions. We test the forecasting performance of each indicator and evaluate which of them works best.

Table 5.1: Advantages (+) and shortcomings (–) of different REER measures

Price Weights	(+)	(–)
HCPI	• homogeneity across countries	• non-tradable goods included • no capital or intermediate goods included • distortions through subsidies and taxes
ULCTOT	• whole economy considered	• non-tradable goods included • only a fraction of the firm's costs considered • measurement problems
UWCMAN	• focus on cost side • labor productivity included	• only manufacturing considered
GDPDEF	• services included	• no complete comparability across countries • distortions through subsidies and taxes
EXPI	• direct prices of exports	• endogenous to exchange rate changes • if measured in values per physical unit, then export composition unfortunately changes competitiveness • publication lags and heavy revisions • no complete comparability across countries

Source: Authors' illustration based on Ca'Zorzi and Schnatz (2010).

As for the soft indicators, we do not only test these price and cost competitiveness measures as hard indicators exclusively. Thus, we decide to add two additional indicators to the horse race: the specific national industrial production index (PIPROD) and the industrial production index of the United States (PIPRODUS).[8] It could be argued that the national production index partially reflects foreign demand and should therefore be a good predictor for national exports. Additionally, PIPROD is a widely accepted business cycle indicator with a high forecasting power. We choose PIPRODUS since the US is one of the most important export partners for a multitude of European states.

[6] More information can be found at: *http://ec.europa.eu/economy_finance/db_indicators/competitiveness/index_en.htm*.

[7] However, there is no standard indicator that measures price and cost competitiveness best (see Ca'Zorzi and Schnatz, 2010).

[8] One important indicator in the existing literature is foreign new orders in the manufacturing sector (see Elstner *et al.*, 2013). Unfortunately, to the best of our knowledge, Eurostat stopped reporting this indicator in 2012, so we cannot use it as a hard indicator.

5.2.2 Empirical Setting

Forecast Model

We generate our pseudo out-of-sample forecasts by employing the following autoregressive distributed lag (ADL) model:

$$y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=m}^{q} \gamma_j x_{t+1-j} + \varepsilon_{t+h} \,, \tag{5.1}$$

where y_{t+h} is the h-step ahead forecast for export growth and x_t represents one of the single indicators. The forecast horizon h is defined in the range of $h \in \{1, 2\}$ quarters since survey-based indicators are usually applied for short-term forecasts (see, among others, Gayer, 2005). We allow a maximum of four lags for our target variable and each single indicator: $p, q \leq 4$. The optimal lag length is determined by the Bayesian Information Criterion (BIC). The initial estimation period for Equation (5.1) ranges from 1996:01 to 2004:03 ($T_E = 35$). The period is then expanded successively by one quarter with a new specification of the model; the first forecast for y_t is calculated for 2004:04 and the last for 2013:04. We implement the ADL model in a direct-step fashion. This means that y_{t+h} is directly explained with lagged values of the dependent variable and the indicator. This results in the same number of forecasts ($T_F = 37$) for every forecast horizon h. More details on direct-step forecasting can be found in Robinzonov and Wohlrabe (2010). As the benchmark model we chose a common AR(p) process.

Forecast Evaluation

To evaluate the forecast accuracy of our different models, we calculate forecast errors. Let $\widehat{y}_{t+h}$ denote the h-step ahead forecast produced at time t. Then the resulting forecast error is defined as $FE_{t+h} = y_{t+h} - \widehat{y}_{t+h}$. The corresponding forecast error of our AR(p) benchmark model is FE_{t+h}^{ARp}. To assess the performance of an indicator-based model, we calculate the root mean squared forecast error (RMSFE) as the loss function. For the h-step ahead indicator-based forecast, the RMSFE is:

$$RMSFE_h = \sqrt{\frac{1}{T_F} \sum_{n=1}^{T_F} \left(FE_{t+h,n}\right)^2} \,. \tag{5.2}$$

The RMSFE for the benchmark model is $RMSFE_h^{ARp}$. To decide whether one indicator performs, on average, better than the autoregressive process, we calculate the relative RMSFE between the indicator model and the benchmark:

$$rRMSFE_h = \frac{RMSFE_h}{RMSFE_h^{ARp}} \,. \tag{5.3}$$

Whenever this ratio is smaller than one, the indicator-based model performs better than the benchmark. Otherwise, the AR(p) process is preferable. Nonetheless, calculating this ratio does not clarify whether the forecast errors of the indicator-based model and the benchmark are statistically different from each other. To check this, we apply the test proposed by Diebold and Mariano (1995). Under the null hypothesis, the test states that the expected difference in the MSFE equals zero. With our notation this gives:

$$H_0 : E\left[\left(FE_{t+h}^{ARp}\right)^2 - \left(FE_{t+h}\right)^2\right] = E\left[MSFE_{t+h}^{ARp} - MSFE_{t+h}\right] = 0 \,. \tag{5.4}$$

The null hypothesis states that the AR(p) is the data generating process. Adding an indicator to this process can then cause a typical problem of nested models. The larger model – with each of our single indicators – introduces a bias through estimating model parameters that are zero within the population. Thus, the AR(p) process nests the indicator model by setting the parameters of the indicator to zero. As stated by Clark and West (2007), this causes the MSFE of the larger model to be biased upwards since redundant parameters have to be estimated. As a result, standard tests, such as the one proposed by Diebold and Mariano (1995), lose their power. On this account, we follow the literature (see, among others, Weber and Zika, 2013; Lehmann and Weyh, 2014) and apply the adjusted test statistic by Clark and West (2007):

$$CW_h = \sqrt{\frac{1}{\widehat{V}(a_{t+h})T_F}} \sum_{t=1}^{T_F} \left(\underbrace{MSFE_{t+h}^{ARp} - \left[MSFE_{t+h} - \left(FE_{t+h} - FE_{t+h}^{ARp}\right)^2\right]}_{a_{t+h}} \right) \,, \tag{5.5}$$

with $\widehat{V}(a_{t+h})$ as the sample variance of a_{t+h} and $\left(FE_{t+h} - FE_{t+h}^{ARp}\right)^2$ as the adjustment term. After this adjustment, standard critical values from the Student's t-distribution with $T_F - 1$ degrees of freedom can be used to decide whether forecast errors are statistically significant from each other.

Forecast Encompassing Test

In order to give a formal statement whether survey-based indicators or hard data perform better, we apply a standard forecast encompassing test. To keep it simple and to not test each possible indicator pair, we separately averaged the forecast errors from all soft (FE_{t+h}^{soft}) and all hard (FE_{t+h}^{hard}) indicators. With a forecast encompassing test, we can easily answer the question of whether a group of indicators (here: soft indicators) has more information content to forecast a target variable in comparison to the other group (here: hard data). We do not want to give a statement which indicator is the best one but rather to answer the question whether soft indicators are better than hard indicators. We therefore test the before mentioned groups against each other. The encompassing test follows the idea of Granger and Newbold (1973), who state that it is insufficient to compare only mean squared forecast errors between competing models. Their suggestion deals with the optimality of a forecast. The preferred forecast does not necessarily comprise all available information and is thus not optimal. This principle is known as "conditional efficiency". The preferred forecast encompasses the competitor, if the competing forecast has no more additional information (see Clements and Hendry, 1993), thus, a combination of both forecasts would not increase forecast accuracy. In our export case, we examine whether soft indicators (FE_{t+h}^{soft}) contain additional information compared to hard data (FE_{t+h}^{hard}). This can simply be answered with the following regression:

$$FE_{t+h}^{hard} = \lambda \left(FE_{t+h}^{hard} - FE_{t+h}^{soft}\right) + \varepsilon_{t+h} \,. \tag{5.6}$$

We apply standard ordinary-least-squares (OLS) with corrected standard errors in the style of Newey and West (1987). We test the null hypothesis $H_0 : \lambda = 0$. Whenever the test rejects the null, soft indicators contain more information than their competitors based on hard data.

5.3 Results

5.3.1 Pseudo Out-of-sample Analysis

Do soft or hard indicators best improve export growth forecasts? The very simple answer is that survey-based indicators do. Table 5.2 shows the pseudo out-of-sample results for all twenty European states in our study and the aggregates EA-18 and EU-28. The target variables are the growth rates of total exports yoy; an expanding window is applied (see *yoy, expanding* in the caption of Table 5.2). The table is divided into the two forecasting horizons ($h = 1, 2$). For every country and forecast horizon, the performance of each soft and hard indicator is presented. Whenever a cell is shown in gray, the specific indicator significantly outperforms the autoregressive benchmark model, thus, the relative root mean squared forecast error (rRMSFE) is smaller than one. A white-colored cell shows that the specific indicator has no higher forecast accuracy than the AR(p) process. Whenever an indicator series was too short for our purposes, a dash ("–") appears in the specific cell. Detailed results can be found in Table A.7 in Appendix 5.A.[9]

To summarize the large amount of information from Table 5.2, we compare the results in two different ways. First, we discuss performance differences across indicators. In a second step, we discuss country differences. Survey-based indicators beat the benchmark model quite often compared to hard indicators, since more cells for soft indicators are shown in gray. Turning to the indicator comparison, it is favorable to work with simple ranks. Therefore, we first assign country-specific ranks to each indicator. Then, we calculate average ranks for each indicator over all countries. This has been done for the two forecast horizons separately. For $h = 1$ the best indicator is the confidence indicator for the manufacturing sector (COF), followed by the specific economic sentiment indicator (ESI) and the production expectations (PEXP). For the larger forecast horizon ($h = 2$), COF and ESI change their positions. Again, production expectations are ranked in the third place. But how large are the forecast improvements of these indicators? We only discuss the results for $h = 1$. The results for $h = 2$ can also be found in Table A.7 in Appendix 5.A. For the COF, the improvement over the benchmark model ranges from 40% for Spain to more than 3% for Poland. In the case of ESI, the range runs from 35% for the EA-18 to 4% in Italy. The PEXP indicator outperforms the benchmark model of

[9] The results table in the appendix presents the rRMSFE for all soft and hard indicators plus three additional benchmark models. One exception is the number for the AR(p) process: here we present the forecast errors in percentage points. Asterisks denote significant differences between the forecast errors based on the outcome of the Clark-West test.

almost 35% for the EA-18 and nearly 8% for the Netherlands. The largest improvement can obviously be found for the aggregate EA-18. One reason might be that the combination of state-specific survey indicators end up in a higher forecast power for total exports of the aggregate. The overall performance of the export expectations (EXEXP) indicator is rather poor in comparison to the three best indicators. From sixteen possible indicators, EXEXP ranks sixth for $h = 1$ and ninth for $h = 2$. The improvement of EXEXP ranges from 32% in Denmark to more than 5% in Sweden for the shorter forecast horizon.

The worst hard indicator is the real effective exchange rate based on unit wage costs in the manufacturing sector (UWCMAN). This indicator is ranked in sixteenth place for the shorter forecast horizon and in fifteenth place for $h = 2$. We ascertain that the hard indicators have in general a poorer forecast performance than the soft ones. However, there is one exception: the US industrial production. For $h = 1$ and $h = 2$, the average rank for PIPRODUS is four. The good forecast performance of PIPRODUS is clearly indicated by the gray-colored boxes in the last column of Table 5.2. By focusing only on the price and cost competitiveness measures, the most intuitive candidate, a real effective exchange rate (REER) based on export price indices (EXPI), is the "less worse" one.

Now we deal with observable country differences. Since we have argued before that US industrial production performs well, we base our country comparison on the performance between soft indicators and the different price and cost competitiveness measures. For this purpose, we can summarize the countries in four possible groups: (i) only soft indicators can beat the benchmark model; (ii) only real effective exchange rates (REER) are better than the autoregressive process; (iii) at least one indicator from both groups works; (iv) no indicator delivers better results at all. Most of the countries fall into the first group. In eleven countries (or country aggregates), only soft indicators beat the benchmark model (see Table 5.3). There is no case where only the price and cost competitiveness measures are better than the autoregressive process. The third group consists of eight countries. In this group soft indicators as well as price and cost competitiveness measures beat the benchmark. There are three countries (Bulgaria, Latvia and Lithuania) where almost no indicator works at all. Especially for these Eastern Europe countries we find no improvement, with the exception of the industrial production of the United States (PIPRODUS) and $h = 2$ for Latvia, through any of our considered indicators. We have to conclude that especially in those three countries, the AR(p) process is a hard-to-beat benchmark model.

Table 5.2: Pseudo out-of-sample results for export growth (yoy, expanding)

h=1																
	Soft indicator									Hard indicator						
Country	EXEXP	COF	EOBL	OBL	PEXP	SFP	IfW	ESI	CCOF	HCPI	ULCTOT	UWCMAN	GDPDEF	EXPI	PIPROD	PIPRODUS
Austria																
Bulgaria									–						–	
CZ																
Denmark																
Estonia																
Finland																
France																
Germany																
Italy																
Latvia															–	
Lithuania									–						–	
Luxemburg																
Netherlands																
Poland									–							
Portugal																
Slovakia									–							
Slovenia																
Spain																
Sweden																
UK																
EA-18																
EU-28																
h=2																
	Soft indicator									Hard indicator						
Country	EXEXP	COF	EOBL	OBL	PEXP	SFP	IfW	ESI	CCOF	HCPI	ULCTOT	UWCMAN	GDPDEF	EXPI	PIPROD	PIPRODUS
Austria																
Bulgaria									–						–	
CZ																
Denmark																
Estonia																
Finland																
France																
Germany																
Italy																
Latvia															–	
Lithuania									–						–	
Luxemburg																
Netherlands																
Poland									–							
Portugal																
Slovakia									–							
Slovenia																
Spain																
Sweden																
UK																
EA-18																
EU-28																

Note: A gray-colored box indicates that the indicator significantly improves the autoregressive benchmark model. A white-colored box denotes no improvement. – not available.

As the analysis of the ranks revealed, we observe a high heterogeneity in the forecasting performance of soft and hard indicators between countries. In Section 5.3.4 we apply standard regression techniques to explain these differences. We especially ask whether the country-specific export composition is able to give some deeper insights into why certain groups of indicators work, while others do not.

Table 5.3: Country differences between soft and hard indicators

Group	Countries
(i): only soft	Austria, Czech Republic, Denmark, France, Italy, Netherlands, Poland, Slovenia, Spain, EA-18, EU-28
(ii): only REER	–
(iii): soft and REER	Estonia, Finland, Germany, Luxemburg, Portugal, Slovakia, Sweden, United Kingdom
(iv): no indicator	Bulgaria, Latvia, Lithuania

5.3.2 Encompassing Test

Before we present some robustness checks as well as a discussion of why the forecast performance of indicators varies between countries, we formally show that soft indicators perform better than hard data. Table 5.4 presents the forecast encompassing test results from Equation (5.6) for the two forecast horizons $h = 1$ and $h = 2$. Asterisks for the standard significance levels 1%, 5% and 10% indicate that soft indicators have significantly more information to forecast export growth in comparison to their hard counterparts.

The table clearly underlines that soft indicators produce lower forecast errors than hard data for almost all of the countries in the sample. However, we observe some exceptions from this clear pattern. For Lithuania, Slovakia and the United Kingdom we find no significant differences between soft and hard indicators. Table 5.3 shows that Lithuania falls into group (iv), where no indicator works, and Slovakia and the United Kingdom are in group (iii), where soft as well as hard indicators work. For these three countries, regardless of whether the indicators improve forecast accuracy or not, no information advance of soft indicators exist. This is also the case for Estonia and Latvia by looking at the shorter forecast horizon ($h = 1$) and for Bulgaria for $h = 2$. All in all, the results of the encompassing test strengthen the findings from the previous subsection. Whenever it comes to a practical application of export predictions, the forecaster should rely on soft indicators, especially the three mentioned above: the confidence indicator for the manufacturing sector (COF), the country-specific economic sentiment indicator (ESI) and the production expectations (PEXP).

Table 5.4: Encompassing results (yoy, expanding)

Country	h=1	h=2
Austria	***	***
Bulgaria	***	
Czech Republic	***	***
Denmark	***	***
Estonia		**
Finland	***	***
France	***	***
Germany	***	***
Italy	**	***
Latvia		***
Lithuania		
Luxemburg	***	***
Netherlands	***	***
Poland	***	***
Portugal	***	***
Slovakia		
Slovenia	***	***
Spain	***	***
Sweden	**	***
United Kingdom		
EA-18	***	***
EU-28	***	***

Note: Estimation with robust standard errors. ***, **, * indicate a p-value below the 1%, 5% or 10% level.

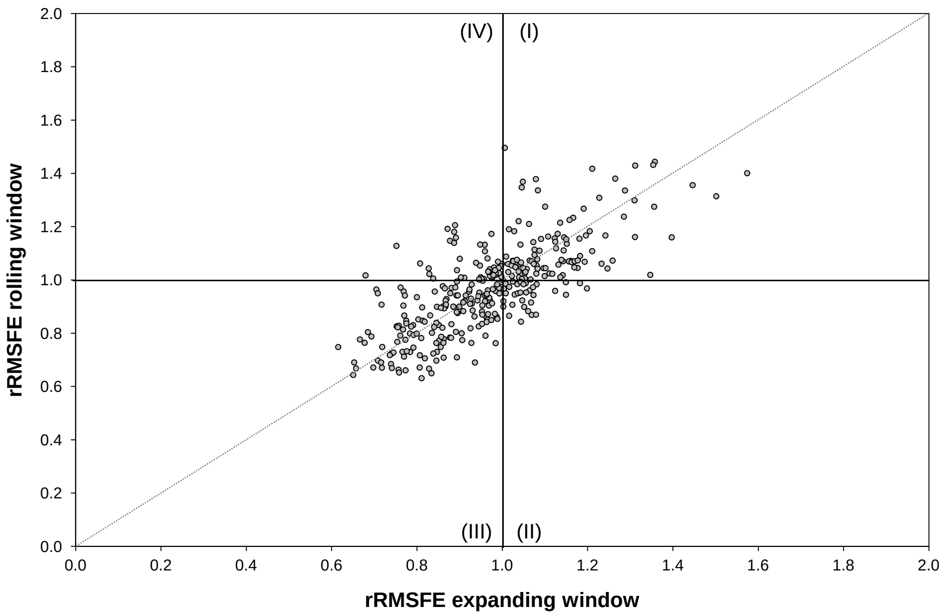
5.3.3 Robustness Checks

To check the validity of our results, we present two types of robustness checks. First, we use a rolling window instead of applying an expanding window approach. This means that the initial estimation window for Equation (5.1) is not successively enlarged by one quarter but is rather fixed and moved forward by one quarter in each single step. Especially if breaks are present in the time series of export growth, the rolling window approach is more suitable. The advantage of the expanding window approach is its ability to capture the whole cyclicality of the underlying time series. In our second robustness check, we apply a different transformation of the target variable. Instead of using year-on-year growth rates, we calculate quarter-on-quarter (qoq) growth rates. Such a transformation captures the cyclical movement of the target variable during the year. In practice, forecasts of macroeconomic aggregates are usually based on the qoq transformation. We use this transformation as the second robustness check.

Let us first stick to the rolling window approach. Figure 5.1 shows a comparison of the relative root mean squared forecast errors (rRMSFE) for the short forecast horizon ($h = 1$); the target variables are yoy growth rates (yoy in the caption of Figure 5.1). Detailed results are available upon request. The rRMSFEs from the rolling window approach are drawn on the y-axis. The rRMSFEs from the expanding window approach can be found on the x-axis. Each dot represents an x-y-pair of an indicator for a specific country (e.g., performance EXEXP for Germany). To ease interpretation of the figure, we add the 45° line as well as a horizontal and vertical line, which both cross the value of the rRMSFE of one, thus, indicating whether an indicator performs better or worse compared to the specific benchmark model. Each dot below the 45° line means that the rRMSFE of the rolling window approach is lower than the one from an expanding window. The opposite holds for values above the 45° line. The horizontal and vertical lines divide the figure into four quadrants. The interpretations for quadrant (I) and (III) are straightforward. A dot in quadrant (I) stands for an indicator that produces a higher root mean squared forecast error (RMSFE) in comparison to the benchmark within the expanding as well as rolling window approach. The opposite holds for an indicator lying in quadrant (III), thus, producing a lower RMSFE in both approaches. Whenever an indicator enters quadrant (IV) its performance becomes worse in a rolling window approach compared to an expanding window. For quadrant (II) the indicator beats the benchmark in a rolling setup, whereas it fails to do so in the expanding window approach.

The results would be perfectly robust to the applied window if all dots lay on the 45° line. Figure 5.1 reveals that this is not the case. The results do not vary much between the two approaches, however, since the dots are located close to the 45° line. Only 24% of all indicators either become better or worse in the rolling window setup compared to their performance in the expanding window approach. However, most of these differences are not statistically significant. The remaining 76% of all indicators remain either in quadrant (I) or (III), thus, their relative performance does not vary between the different forecast approaches. We conclude that the results are fairly robust for the shorter forecast horizon. The figure for the larger forecast horizon ($h = 2$) can be found in the Appendix (see Figure A.3 in Section 5.A). In that case, 30% of all results lie in either quadrant (II) or (IV). Still, 70% of all indicators stay robust in their relative performance. This is a confirmation of the results from the expanding window approach in Section 5.3.1.

Figure 5.1: Relative forecast errors in expanding vs. rolling window (yoy, $h = 1$)

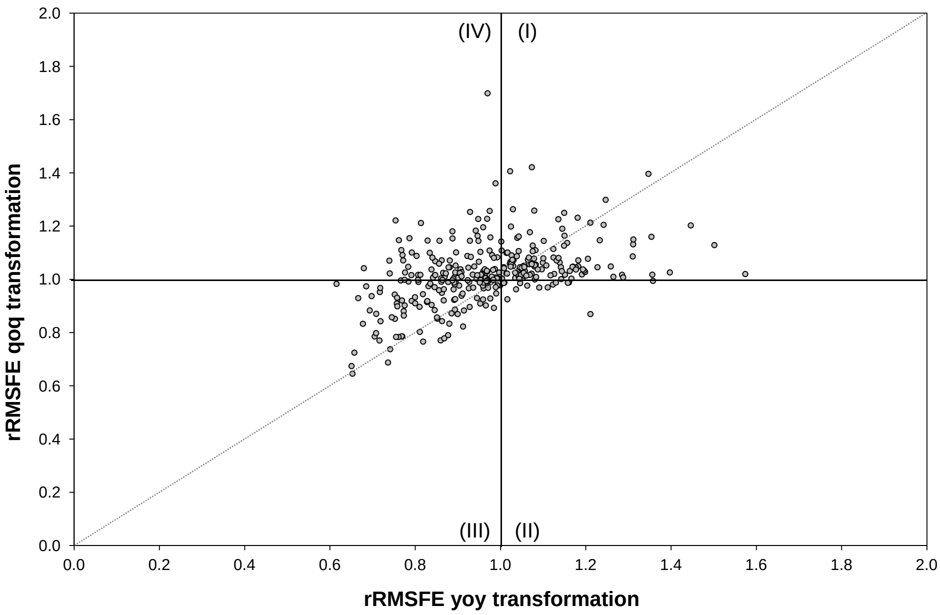


The second robustness check is based on an alternative transformation of our target variable: qoq growth rates.[10] As for the rolling window, we present a similar figure as for the alternative transformation. For $h = 1$, Figure 5.2 compares the relative performance of the indicators in both transformations; the expanding window approach is applied (*expanding* in the caption of Figure 5.2). The results are not as robust as for the rolling window. 32% of all indicators change their relative performance for $h = 1$ by applying qoq instead of yoy growth rates. The bulk of these indicators are located in quadrant (IV), thus, the relative performance worsens. For the larger forecast horizon ($h = 2$) even more indicators can be found in quadrant (II) or (IV). Nearly 42% change their relative performance between the two transformations (see Figure A.4 in Appendix 5.A). Almost 68% ($h = 1$) and 58% ($h = 2$) of all indicators keep their relative performance; thus, most of the findings remain the same. Additionally, qoq growth rates show a higher volatility compared to their yoy counterparts and are thus not that persistent. This fact makes them harder to predict. Gayer (2005) recommends clarifying to which reference series different survey indicators refer. This statement is directly transferable to our

[10] All numerical results are available upon request.

question. Do our indicators refer to yoy or qoq export growth rates? From the previous findings we suggest that most of the indicators clearly refer to yoy export growth rates. Whenever it comes to predicting exports of goods and services, the forecaster should rely on yoy instead of qoq growth rates.

Figure 5.2: Relative forecast errors yoy vs. qoq transformation (expanding, $h = 1$)



5.3.4 Discussion of the Results

In the final step, we try to find some explanations for the high heterogeneity in performance between countries. Why do soft indicators work better in country A compared to country B? A similar question can be raised for hard indicators. To answer these questions, we run the following regression:

$$\overline{rRMSFE}_i^k = c^k + \beta_1 East_i + \beta_2 Service_i + \sum_{j=1}^{7} \beta_j SITC_i + \beta_8 HHI_i + \varepsilon_i^k \,. \tag{5.7}$$

First, we calculate the average $\overline{rRMSFE}$ of all soft (hard) indicators, here abbreviated with $k \in \{soft, hard\}$, for each country (i). Second, we ask the question of which variables may explain the differences in relative forecast errors. Since the sample is not too large, we end up with the composition of total exports. Therefore we use the average share of service exports in total exports (*Service*) between 2005 and 2013. Additionally, we add the shares of different product groups. For instance, Germany exports more cars, whereas the United Kingdom has a higher share in oil exports. Maybe it is easier for firm A to expect what their car exports will be instead of the highly uncertain or more volatile exports of oil from firm B. Thus, the performance of soft and hard indicators may crucially depend on the composition of exports and therefore the possibility of a firm to correctly anticipate future developments in foreign markets. In the end, we add average shares of seven different product groups based on the Standard International Trade Classification (*SITC*) between 2005 and 2013. The codes as well as the corresponding product groups can be found in Table A.9 in the Appendix 5.A. Instead of using each single product group in the regression, we calculate a standard Hirschman-Herfindahl-Index (*HHI*) to measure the diversification of exports. At last, we add a Dummy for Eastern Europe countries in the sample (*East*). This dummy accounts for the observed differences in forecast performance between Eastern and Western Europe countries. We focus on the short forecast horizon $h = 1$, yoy export growth rates and an expanding window (see the caption of the following tables). Equation (5.7) is estimated with OLS and robust standard errors based on the Huber-White-Sandwich-Estimator.

Table 5.5 presents the regression results for the soft indicators. It should be notified that we use the average values of the rRMSFE for each country so that we end up with twenty observations, one for each country, in the regression. All these results thus should be interpreted with caution since the number of observations is rather small. In the end, we estimate the model with only one SITC variable, in order not to stress the few degrees of freedom. Therefore, the output tables contain eight columns, each for one single SITC group plus the HHI. We find for soft indicators that the average rRMSFE is higher for Eastern European states compared to non-Eastern European countries. We find no statistically significant correlation between the share of service exports and the relative performance of soft indicators. Hence, we expect that the performance of soft indicators is almost independent of the target variable. It seems to make no difference whether we forecast exports of goods, exports of services or the sum of both.[11]

[11] We run our forecasting exercise for the two components of total exports as well. On average, we find no large difference, which explains the insignificant *Service* coefficient. Some performance differences do exist, but we do not want to discuss these results in detail; these results are available upon request.

Table 5.5: Composition of exports and performance of soft indicators (yoy, expanding, $h = 1$)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
East	0.094**	0.078*	0.080**	0.132**	0.095**	0.076**	0.100	0.108**
	(0.038)	(0.040)	(0.036)	(0.051)	(0.044)	(0.029)	(0.058)	(0.038)
Service	0.048	0.035	0.077	0.118	0.050	0.001	0.059	0.078
	(0.102)	(0.096)	(0.084)	(0.115)	(0.145)	(0.061)	(0.125)	(0.077)
SITC01	0.430							
	(0.407)							
SITC24		0.674*						
		(0.354)						
SITC3			0.791***					
			(0.212)					
SITC5				0.732				
				(0.494)				
SITC68					0.047			
					(0.365)			
SITC7						-0.471***		
						(0.092)		
SITC9							0.133	
							(2.273)	
HHI								-0.630**
								(0.228)
c	0.832***	0.846***	0.805***	0.756***	0.853***	1.057***	0.861***	1.028***
	(0.049)	(0.042)	(0.046)	(0.086)	(0.099)	(0.048)	(0.070)	(0.060)
R^2	0.302	0.313	0.467	0.330	0.258	0.527	0.258	0.411
Obs.	20	20	20	20	20	20	20	20

Note: Robust standard errors are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, 10% level.

Now let us turn to the SITC variables. Obviously the share of three product groups correlates with the relative performance of our soft indicators. These are: SITC24 – raw materials etc., SITC3 – mineral fuels etc. and SITC7 – machinery and transport equipment. Whenever a country has a higher export share in raw materials or, for example, oil, the relative forecast performance of soft indicators worsens (see the coefficients 0.674 and 0.791 in Table 5.5). Thus, it seems either harder for the firms to really anticipate future developments of exports or confidence indicators are not able, from a time series perspective, to grab export growth in a meaningful way. On the other hand, a higher share of machinery goods leads to a significant improvement in the forecasting performance of survey-based indicators (see the coefficient -0.471 in Table 5.5). These three results are underpinned by the significant negative coefficient for the HHI. Since the HHI is coded in a way that a larger number represents a lower degree of diversification, the negative coefficient is interpreted as follows: the more diversified the exports of a country are, the better the performance of soft indicators.

The same exercise can be done for the average performance of our hard indicators; Table 5.6 shows the corresponding results. We find no significant difference for Eastern Europe countries and no impact of the share in service exports. The composition of goods exports seems to matter only in a minor way. Only a higher share of products in the group SITC01 (food, beverages and tobacco) seems to worsen the relative performance of hard indicators.

Table 5.6: Composition of exports and performance of hard indicators (yoy, expanding, $h = 1$)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
East	-0.033	-0.014	-0.024	0.027	-0.016	-0.033	-0.070	-0.018
	(0.048)	(0.056)	(0.052)	(0.047)	(0.044)	(0.051)	(0.072)	(0.051)
Service	-0.156	-0.120	-0.132	-0.045	-0.104	-0.156	-0.142	-0.122
	(0.119)	(0.102)	(0.112)	(0.124)	(0.107)	(0.108)	(0.103)	(0.119)
SITC01	1.002**							
	(0.426)							
SITC24		-0.353						
		(0.508)						
SITC3			-0.018					
			(0.349)					
SITC5				1.072				
				(0.754)				
SITC68					-0.162			
					(0.387)			
SITC7						-0.193		
						(0.212)		
SITC9							-2.370	
							(1.394)	
HHI								-0.322
								(0.321)
c	1.035***	1.120***	1.112***	0.952***	1.150***	1.190***	1.170***	1.195***
	(0.052)	(0.061)	(0.071)	(0.103)	(0.127)	(0.094)	(0.079)	(0.097)
R^2	0.246	0.050	0.037	0.172	0.046	0.076	0.102	0.072
Obs.	20	20	20	20	20	20	20	20

Note: Robust standard errors are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, 10% level.

All in all the composition of exports seem to matter for the relative performance of indicators. However, we suspect that firm characteristics in particular explain these observed country differences, i.e., firm samples of each country over time would offer a rich source of variation. With this information, future research activities could either analyze the number of exporting firms or their corresponding characteristics could explain our observed differences in forecasting performance. To the best of our knowledge, no European study exists that links firm-level information to the macroeconomic forecasting performance of survey-based indicators. However, there is some literature which links so called non-responses of firms to the accuracy of survey-based indicators (for Germany see Seiler, 2014).

Another explanation for the country differences could lie in the aggregation of firm responses. In this paper, we use a rather standard measure: the balances of positive and negative responses. However, there is a broad discussion on the usefulness of balances (see Croux *et al.*, 2005; Claveria *et al.*, 2007). Future research activities could focus on a sensitivity analysis with respect to different aggregation methods.

Finally, we return to the discussion brought forward by Gayer (2005). He asks which survey indicator refers to which specific reference series. The European Commission also provides survey indicators for different sub-sectors in manufacturing. Since the discussion before reveals the fact that the export composition matters for the relative performance of soft indicators, maybe sectoral results are more closely linked to total export growth. We leave all these issues for follow-up studies.

5.4 Summary and Conclusion

Macroeconomic forecasts consist of more than the prediction of a single number, namely gross domestic product (GDP). In practice it is standard to forecast each single component (e.g. exports) of total output. Disaggregated GDP forecasts are also seen as more accurate than direct predictions in the academic literature. Thus, better forecasts on each single component lead, c.p., to lower forecast errors for GDP. In this paper we concentrate on one major aggregate in total output: exports of goods and services. In conclusion, do soft or hard indicators have better predictive power for export growth? This paper evaluates this question with pseudo out-of-sample techniques and forecast encompassing tests for twenty single European states and the aggregates EA-18 and EU-28. Our period of investigation runs from the first quarter of 1996 to the fourth quarter of 2013 and therefore covers more than one business cycle. For most of our countries we find a significant improvement in forecast accuracy through survey-based indicators. Hard indicators such as price and cost competitiveness measures are only in a few cases able to beat the benchmark model. One exception of a hard indicator is US industrial production, which is a tough competitor compared to the soft indicators. Two robustness checks confirm our results.

All in all, we find remarkable differences in forecast accuracy between the countries in the sample. We therefore ask: what are the reasons for these country differences? It turns out that the export composition in particular has an impact on the forecast accuracy of survey-based indicators. The relative performance of soft indicators is lower in countries with a higher share in exports of raw materials or oil. The opposite holds for countries with a higher share in machinery exports. For hard indicators, we find only weak results for the export composition.

This paper expands the discussion on export forecasts in several ways. First, we use a multitude of indicators for the forecasting exercise and employ a competition between soft and hard data. Second, we analyze this competition for a multitude of European states, thus broadening the picture of the usefulness of indicators for export forecasts. Third, we implicitly stick to the discussion by Claveria *et al.* (2007) by searching for the reasons for observed country differences. We find that the accuracy of soft indicators depends on export composition. However, further investigation of this result is needed. Finally, this paper gives some suggestions for future research activities to develop a broader understanding of the forecasting power of survey results for exports in particular and different macroeconomic variables in general.

Appendix 5.A

Table A.7: Detailed out-of-sample results for export growth (yoy, expanding)

Model	h=1	h=2	Model	h=1	h=2
Austria			**Bulgaria**		
AR(p) in %	3.899	6.001	AR(p) in %	14.621	14.879
AR(1)	1.692	1.421	AR(1)	0.875	1.067
ISM	2.053	1.351	ISM	0.871	0.871
RW	1.560	1.414	RW	1.053	1.321
EXEXP	0.925*	0.990**	EXEXP	1.160	1.142
COF	0.852**	0.863**	COF	1.009	1.587
EOBL	0.960**	1.070	EOBL	1.073	1.113
OBL	0.984**	1.108	OBL	1.034	1.079
PEXP	0.769**	0.824**	PEXP	1.080	1.159
SFP	0.908**	1.044	SFP	1.006	1.192
IfW	1.211	1.395	IfW	1.242	1.289
ESI	0.851**	0.888**	ESI	1.264	1.656
CCOF	1.016	1.041	CCOF	–	–
HCPI	1.051	1.052	HCPI	1.151	1.075
ULCTOT	1.064	1.052	ULCTOT	1.166	1.178
UWCMAN	1.054	1.061	UWCMAN	1.357	1.442
GDPDEF	1.044	1.053	GDPDEF	1.157	1.215
EXPI	1.067	1.087	EXPI	1.191	1.264
PIPROD	0.999	1.011	PIPROD	–	–
PIPRODUS	0.997	1.017	PIPRODUS	1.198	1.374
Czech Republic			**Denmark**		
AR(p) in %	8.793	9.322	AR(p) in %	5.363	6.279
AR(1)	1.010	1.088	AR(1)	1.015	1.023
ISM	1.122	1.070	ISM	1.229	1.064
RW	1.031	1.276	RW	1.038	1.153
EXEXP	0.847*	0.976**	EXEXP	0.677*	0.654*

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	Model	h=1	h=2
COF	0.863*	0.963**	COF	0.845*	0.841*
EOBL	0.986	1.067	EOBL	0.800*	0.763*
OBL	0.982*	1.112	OBL	0.773*	0.751*
PEXP	0.791*	0.871**	PEXP	0.708*	0.685*
SFP	1.002	1.075	SFP	1.052	1.078
IfW	1.022	1.153	IfW	1.044	1.120
ESI	0.693*	0.818**	ESI	0.862**	0.802**
CCOF	0.960*	1.069	CCOF	0.976	0.908**
HCPI	1.007	1.153	HCPI	1.082	1.002
ULCTOT	1.123	1.113	ULCTOT	1.064	1.029
UWCMAN	1.142	1.092	UWCMAN	1.156	1.245
GDPDEF	1.111	1.093	GDPDEF	1.124	1.161
EXPI	1.130	1.069	EXPI	1.066	1.236
PIPROD	0.740**	0.998*	PIPROD	0.976	0.949**
PIPRODUS	0.959*	1.068	PIPRODUS	0.757*	0.742*
	Estonia			**Finland**	
AR(p) in %	10.628	14.532	AR(p) in %	10.461	11.457
AR(1)	1.285	1.165	AR(1)	1.053	1.059
ISM	1.311	1.155	ISM	1.130	1.041
RW	1.310	1.299	RW	1.122	1.290
EXEXP	0.968	1.024	EXEXP	0.867*	0.937**
COF	0.942	0.835**	COF	0.698**	0.760**
EOBL	0.807***	0.913**	EOBL	0.806*	0.820*
OBL	0.948**	0.991	OBL	0.807**	0.841*
PEXP	1.001	1.030	PEXP	0.784**	0.903*
SFP	0.911*	0.935**	SFP	0.903**	1.114
IfW	0.856**	0.966**	IfW	0.783	0.981
ESI	0.881*	0.824**	ESI	0.766**	0.783**
CCOF	1.024	0.956*	CCOF	0.775*	0.774**
HCPI	0.984**	0.824***	HCPI	1.030	1.057
ULCTOT	0.980	1.069	ULCTOT	0.938*	1.001

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
UWCMAN	0.970	0.947*	UWCMAN	0.812**	0.928**
GDPDEF	1.052	1.035	GDPDEF	1.053	1.023
EXPI	0.895***	0.940**	EXPI	1.057	1.085
PIPROD	1.039	0.890***	PIPROD	0.906*	0.956**
PIPRODUS	0.834*	0.883*	PIPRODUS	0.792*	0.920*
	France			**Germany**	
AR(p) in %	4.100	5.335	AR(p) in %	7.180	8.137
AR(1)	1.256	1.187	AR(1)	1.095	1.144
ISM	1.523	1.184	ISM	1.200	1.070
RW	1.278	1.319	RW	1.100	1.283
EXEXP	0.926*	1.097	EXEXP	0.718**	0.917**
COF	0.705**	0.745*	COF	0.716**	0.805**
EOBL	0.945**	0.947*	EOBL	0.773**	0.848*
OBL	0.810**	0.850*	OBL	0.752**	0.844*
PEXP	0.768**	0.749*	PEXP	0.811**	0.802**
SFP	0.769**	0.864**	SFP	0.745**	0.862**
IfW	1.347	1.319	IfW	0.857*	1.009
ESI	0.708**	0.693*	ESI	0.845**	0.919*
CCOF	0.827**	0.817*	CCOF	0.963	0.983*
HCPI	1.285	1.180	HCPI	1.083	1.077
ULCTOT	1.356	1.264	ULCTOT	1.035	1.052
UWCMAN	1.574	1.429	UWCMAN	0.924**	0.998*
GDPDEF	1.287	1.206	GDPDEF	1.097	1.095
EXPI	1.083	1.216	EXPI	1.141	1.109
PIPROD	1.311	1.048	PIPROD	1.149	1.140
PIPRODUS	0.862**	1.003	PIPRODUS	0.776**	0.916*
	Italy			**Latvia**	
AR(p) in %	5.923	7.323	AR(p) in %	6.894	8.660
AR(1)	1.254	1.235	AR(1)	1.191	1.136
ISM	1.463	1.197	ISM	1.416	1.142
RW	1.272	1.384	RW	1.246	1.305

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
EXEXP	0.994	1.052	EXEXP	1.124	1.035
COF	0.899*	0.925*	COF	0.990	0.957
EOBL	0.891*	1.026	EOBL	0.935	0.868**
OBL	0.952*	0.982*	OBL	1.091	1.044
PEXP	0.914*	0.953*	PEXP	1.069	1.029
SFP	0.971	0.966	SFP	1.046	0.994*
IfW	1.149	1.380	IfW	1.013	0.964**
ESI	0.959*	0.926*	ESI	1.007	1.002
CCOF	0.992	0.990	CCOF	1.100	1.036
HCPI	1.080	1.060	HCPI	1.023	1.031
ULCTOT	1.055	1.029	ULCTOT	1.022	1.039
UWCMAN	1.069	1.039	UWCMAN	1.049	1.047
GDPDEF	1.080	1.039	GDPDEF	0.997	1.009
EXPI	1.136	1.144	EXPI	0.998	0.998
PIPROD	0.974*	0.915*	PIPROD	–	–
PIPRODUS	0.953*	0.987	PIPRODUS	0.946*	0.963**
Lithuania			**Luxemburg**		
AR(p) in %	7.705	9.590	AR(p) in %	6.891	8.154
AR(1)	1.248	1.163	AR(1)	1.067	1.050
ISM	1.338	1.084	ISM	1.230	1.051
RW	1.384	1.489	RW	1.086	1.189
EXEXP	1.030	1.111	EXEXP	0.869**	0.936**
COF	1.145	1.241	COF	0.860**	0.945**
EOBL	1.181	1.285	EOBL	0.845***	0.868***
OBL	1.150	1.179	OBL	0.861***	0.928**
PEXP	1.075	1.214	PEXP	0.893**	1.022
SFP	1.075	1.094	SFP	0.865***	0.918**
IfW	1.003	0.991	IfW	0.813	1.001
ESI	1.072	1.229	ESI	0.791**	0.918**
CCOF	–	–	ESI	0.791**	0.918**

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
HCPI	1.015	1.036	HCPI	1.047	1.014
ULCTOT	1.024	1.033	ULCTOT	0.900**	0.924*
UWCMAN	1.045	1.024	UWCMAN	1.205	1.226
GDPDEF	1.014	1.039	GDPDEF	0.965*	0.965
EXPI	1.042	1.076	EXPI	0.959**	0.979
PIPROD	–	–	PIPROD	0.863*	0.920*
PIPRODUS	1.232	1.167	PIPRODUS	0.922	0.973
Netherlands			**Poland**		
AR(p) in %	4.024	5.202	AR(p) in %	7.307	7.803
AR(1)	1.299	1.198	AR(1)	0.957*	0.967***
ISM	1.443	1.128	ISM	1.055	0.999*
RW	1.351	1.418	RW	1.011	1.244
EXEXP	0.944***	0.978**	EXEXP	0.754***	0.881**
COF	0.912**	0.990*	COF	0.966*	1.132
EOBL	0.965**	0.990*	EOBL	0.963*	1.069
OBL	0.892**	1.041	OBL	0.960**	1.069
PEXP	0.923**	1.014	PEXP	0.772***	0.891**
SFP	0.893**	0.904*	SFP	0.761**	0.943**
IfW	1.135	1.266	IfW	0.969	1.381
ESI	0.775**	0.837**	ESI	1.029	1.167
CCOF	0.841***	0.929***	CCOF	–	–
HCPI	1.144	1.068	HCPI	1.038	1.175
ULCTOT	1.126	1.084	ULCTOT	1.211	1.208
UWCMAN	1.196	1.043	UWCMAN	1.397	1.438
GDPDEF	1.310	1.153	GDPDEF	1.227	1.174
EXPI	1.062	1.026	EXPI	1.048	1.154
PIPROD	1.043	0.979	PIPROD	0.818*	1.012
PIPRODUS	0.896**	1.039	PIPRODUS	1.023	1.113
Portugal			**Slovakia**		
AR(p) in %	6.592	7.129	AR(p) in %	11.158	12.949
AR(1)	1.042	1.102	AR(1)	0.986	0.995

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
ISM	1.130	1.056	ISM	1.105	0.965
RW	1.090	1.310	RW	1.010	1.126
EXEXP	0.887*	1.054	EXEXP	0.948**	0.979
COF	0.799*	0.926	COF	0.767**	0.901**
EOBL	0.739**	0.890*	EOBL	1.038	1.095
OBL	0.786**	0.893*	OBL	1.026	1.045
PEXP	0.831*	1.042	PEXP	0.685**	0.812**
SFP	0.963	1.052	SFP	0.872	0.886
IfW	1.073	1.191	IfW	0.887*	1.236
ESI	0.867*	1.003	ESI	0.930***	0.940**
CCOF	0.992	1.155	CCOF	–	–
HCPI	0.993	1.077	HCPI	0.877	0.898*
ULCTOT	0.966	1.047	ULCTOT	0.889	0.908
UWCMAN	1.054	1.061	UWCMAN	1.078	1.150
GDPDEF	0.964	1.053	GDPDEF	0.891	0.898
EXPI	0.953	1.043	EXPI	0.887	0.905
PIPROD	0.908	1.030	PIPROD	0.894	0.911
PIPRODUS	0.928*	1.045	PIPRODUS	0.829**	0.837**
	Slovenia			**Spain**	
AR(p) in %	7.247	9.321	AR(p) in %	6.044	7.267
AR(1)	1.197	1.124	AR(1)	1.022	1.027
ISM	1.387	1.092	ISM	1.192	1.003
RW	1.156	1.203	RW	1.044	1.114
EXEXP	0.984*	1.041	EXEXP	0.894*	0.928*
COF	0.828**	0.822***	COF	0.615**	0.549**
EOBL	0.893***	0.887***	EOBL	0.717**	0.659**
OBL	0.758***	0.885***	OBL	0.752**	0.699**
PEXP	0.847**	0.973**	PEXP	0.770**	0.682**
SFP	0.989**	0.968**	SFP	0.666**	0.619**
IfW	1.079	1.187	IfW	0.928**	0.962*
ESI	0.936***	0.923***	ESI	0.679**	0.568**

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
CCOF	1.036	1.090	CCOF	0.838**	0.695*
HCPI	1.050	1.047	HCPI	1.099	0.987*
ULCTOT	1.002	1.004	ULCTOT	1.182	1.001
UWCMAN	1.073	1.065	UWCMAN	1.182	1.022
GDPDEF	1.056	1.027	GDPDEF	1.139	1.033
EXPI	1.008	1.036	EXPI	1.131	1.045
PIPROD	1.101	1.207	PIPROD	0.988*	0.936*
PIPRODUS	0.974	1.026	PIPRODUS	0.803**	0.928*
	Sweden			**United Kingdom**	
AR(p) in %	4.665	6.407	AR(p) in %	8.428	7.967
AR(1)	1.355	1.268	AR(1)	0.953*	1.064
ISM	1.667	1.229	ISM	0.944*	1.006
RW	1.307	1.305	RW	1.042	1.301
EXEXP	0.949***	0.933**	EXEXP	0.988	1.065
COF	0.905**	0.959**	COF	1.034	1.050
EOBL	1.060	1.079	EOBL	1.107	1.063
OBL	0.757**	0.872**	OBL	1.087	1.079
PEXP	0.953	0.951*	PEXP	0.991	1.032
SFP	1.016	1.018	SFP	1.044	1.058
IfW	0.976*	1.033	IfW	1.028	1.195
ESI	0.835**	0.838**	ESI	0.979	0.997
CCOF	0.881**	0.870**	CCOF	0.890**	0.865*
HCPI	0.969*	1.010	HCPI	0.970*	0.978
ULCTOT	0.957*	1.010	ULCTOT	0.969**	0.998
UWCMAN	0.984	1.029	UWCMAN	0.951**	0.978
GDPDEF	0.968*	1.009	GDPDEF	0.967**	0.991
EXPI	0.954**	1.007	EXPI	0.983*	0.991
PIPROD	0.978*	0.995	PIPROD	0.945*	1.010
PIPRODUS	0.878**	0.907**	PIPRODUS	0.908*	0.950

Continued on next page...

Detailed out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
	EA-18			**EU-28**	
AR(p) in %	4.892	6.317	AR(p) in %	4.953	6.160
AR(1)	1.325	1.239	AR(1)	1.274	1.223
ISM	1.488	1.166	ISM	1.427	1.160
RW	1.309	1.368	RW	1.267	1.366
EXEXP	0.718**	0.983*	EXEXP	0.838**	0.920**
COF	0.736**	0.759**	COF	0.818*	0.800*
EOBL	0.877*	0.937*	EOBL	0.927*	0.992
OBL	0.859**	0.839*	OBL	0.880*	0.882*
PEXP	0.653**	0.682**	PEXP	0.741*	0.717*
SFP	0.868**	0.870**	SFP	0.885*	0.885*
IfW	1.246	1.438	IfW	1.042	1.174
ESI	0.650**	0.627**	ESI	0.657**	0.632**
CCOF	0.761**	0.750**	CCOF	0.755**	0.715**
HCPI	1.176	1.095	HCPI	1.354	1.341
ULCTOT	1.162	1.084	ULCTOT	1.258	1.125
UWCMAN	1.117	1.050	UWCMAN	1.193	1.110
GDPDEF	1.170	1.102	GDPDEF	1.311	1.406
EXPI	1.169	1.111	EXPI	1.177	1.105
PIPROD	1.501	1.315	PIPROD	1.446	1.233
PIPRODUS	0.856**	0.954*	PIPRODUS	0.840**	0.943**

Note: The table presents the relative root mean squared forecast errors (*rRMSFE*) of the different models and the benchmark. The row AR(p) in % shows the *RMSFE* for the benchmark model. *ISM*, in-sample mean; *RW*, Random-Walk. Asterisks show significant differences between forecast errors due to the Clark-West test. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

Table A.8: Encompassing results (yoy, rolling)

Country	**h=1**	**h=2**
Austria	***	***
Bulgaria	**	
Czech Republic	***	***
Denmark	**	**
Estonia		***
Finland	***	***
France	***	***
Germany	***	***
Italy		***
Latvia		***
Lithuania	*	
Luxemburg	***	***
Netherlands	***	***
Poland	***	**
Portugal	***	*
Slovakia	***	***
Slovenia	**	***
Spain	***	***
Sweden	***	***
United Kingdom		
EA-18	***	***
EU-28	***	***

Note: Estimation with robust standard errors. ***, **, * indicate a p-value below the 1%, 5% or 10% level.

Table A.9: SITC Codes and product groups

Code	**Product group**
SITC01	Food and live animals, beverages and tobacco
SITC24	Crude materials, inedible, except fuels, animal and vegetable oils, fats and waxes
SITC3	Mineral fuels, lubricants and related materials
SITC5	Chemicals and related products, n.e.s.
SITC68	Manufactured goods classified chiefly by material, miscellaneous manufactured articles
SITC7	Machinery and transport equipment
SITC9	Commodities and transactions not classified elsewhere in the SITC

Source: Department of Economic and Social Affairs of the United Nations (2006).

Figure A.3: Relative forecast errors in expanding vs. rolling window (yoy, $h = 2$)

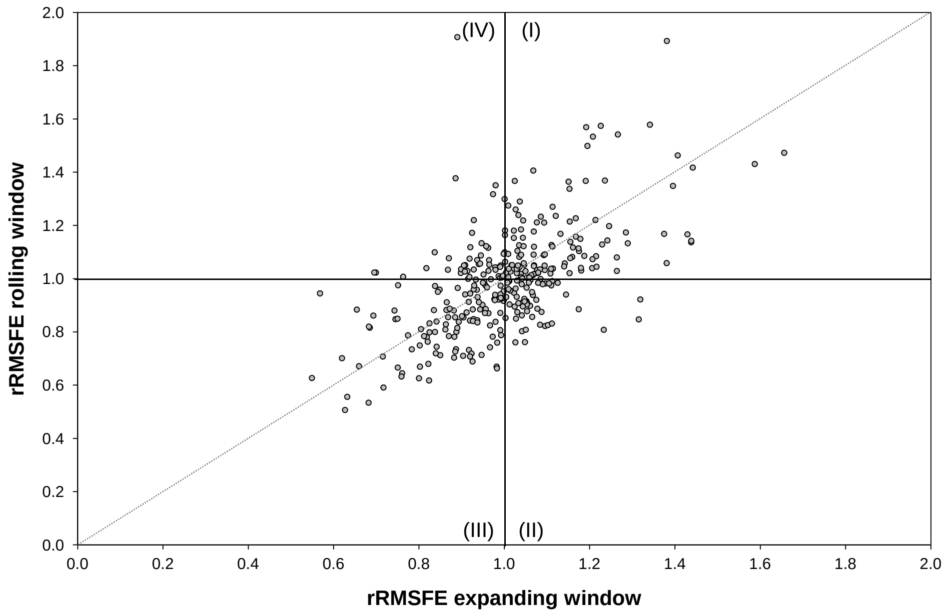
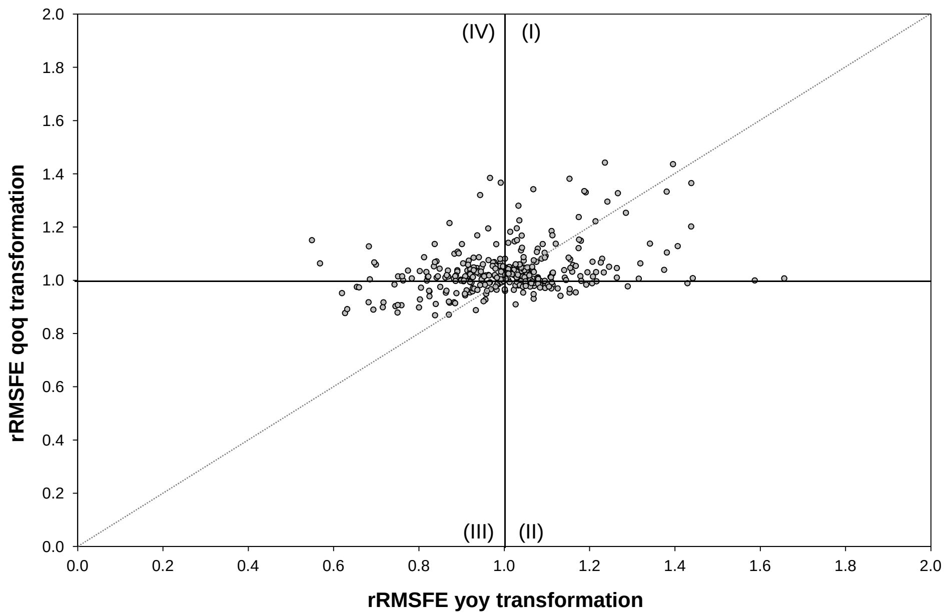


Figure A.4: Relative forecast errors yoy vs. qoq transformation (expanding, $h = 2$)



6 Concluding Remarks

This thesis deals with the topics of economic growth and business cycle forecasting for sub-national entities. Chapter 2 contributes to the discussion of the role of sectoral structures for regional economic development. Economic theory views either specialization or diversification in regional economic structures as the main driver of sectoral development. However, empirical evidence is mixed, with either specialization or diversification identified as the main driver. Existing studies mainly neglect interaction effects between specialization and diversification. It turns out from our study of German cities that a negative interaction effect exists in manufacturing, advanced services and the construction sector. We instead find a positive interaction between specialization and diversification for basic services. Chapters 3 and 4 deal with the topic of regional economic forecasting. Whereas Chapter 3 generally asks which indicator or forecast strategy works best to forecast regional gross domestic product, Chapter 4 investigates whether a disaggregated forecasting approach (each sectoral gross value added) is preferable to a direct forecast (total gross value added). From the exercise in Chapter 3 we can conclude that regional indicators in particular (either for the two German states Saxony and Baden-Württemberg or Eastern Germany) and combination strategies can significantly improve the forecasting accuracy of regional gross domestic product. The main result in Chapter 4 is that one quarter ahead forecasts can significantly be improved by a disaggregated forecasting approach of Saxon gross value added. Finally, Chapter 5 deals with the forecasting of export growth. The main focus of this chapter is on the question of whether survey-based indicators or hard data (for example, price and cost competitiveness measures) can improve export forecasts to a greater degree. For most of the European states we find that survey-based indicators beat hard data. However, major country differences in the forecast performance of soft and hard indicators exist. These differences are described by the export composition of a specific country.

But what follows from the results obtained in this book? And what are possible future research questions? From Chapter 2 we can cautiously state that industrial clusters promote regional economic development in a specific sector. However, the literature names diversification as an insurance against negative economic shocks that hit regional economies. Bearing our results in mind, a more diversified surrounding sectoral structure harms the positive effects of specialization in most of our analyzed sectors. These findings should clearly be more investigated in future research activities. Interaction models are a useful tool to describe different constellations of specialization and diversification, so that future studies can elaborate on this point. In our study we use a rather crude classification of sectoral structures (1-digit level). Our empirical approach can easily be applied to a broader classification scheme. Future research activities can also concentrate on a deeper theoretical understanding of interaction effects between specialization and diversification. With the approaches applied in Chapter 3 and 4, we can significantly reduce forecast errors for regional entities, making the future state of the regional economy more tangible. Regional policy-makers can therefore use the forecasts either for a broader information base or as an indicator of, for instance, fiscal policy planning. In the German case, regional economic forecasts may also become important in the discussion of the debt brake. Based on the insights and the existing literature in Chapter 3 and 4, we can identify some gaps where future research activities may set in. Our approaches are easily applicable to other regions, variables (like employment) and time periods. Another interesting research question is also whether a forecast of German gross domestic product can be improved by an aggregation of forecasts for each single German state. We think the main reason is that regional indicators better capture region-specific business cycles, and are thus more closely linked to regional economic development. Such a closer linkage may reduce forecasting errors for total German GDP. Chapter 5 sheds more light on the question of why survey-based indicators work for some countries, whereas they are not able to improve export growth forecasts in other countries. One reason is the export composition of a specific state. This is one finding that future research activities could expand upon. For example, future studies could use firm characteristics or firm responses in the survey to potentially describe country differences in the forecasting accuracy of survey-based indicators. Such an exercise should not exclusively focus on exports, but could be applied to a large amount of economic series to be forecasted. Finally, to reintroduce the regional focus of this thesis, the study of single demand side components of gross domestic product is also a very interesting field for gaining insights into how regional indicators can help to predict these components.

Literature

ABBERGER, K. and WOHLRABE, K. (2006). Einige Prognoseeigenschaften des ifo Geschäftsklimas – Ein Überblick über die neuere wissenschaftliche Literatur. *ifo Schnelldienst*, **59** (22), 19–26.

ALMEIDA, R. (2007). Local Economic Structure and Growth. *Spatial Economic Analysis*, **2** (1), 65–90.

ANDREWS, D. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, **59** (3), 817–858.

ANGELINI, E., BANBURA, M. and RÜNSTLER, G. (2010). Estimating and Forecasting the Euro Area Monthly National Accounts from a Dynamic Factor Model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, **2010** (1), 5–26.

ARROW, K. (1962). The Economic Implications of Learning by Doing. *The Review of Economic Studies*, **29** (3), 155–173.

BAGHESTANI, H. (1994). Evaluating Multiperiod Survey Forecasts of Real Net Exports. *Economics Letters*, **44** (3), 267–272.

BANDHOLZ, H. and FUNKE, M. (2003). Die Konstruktion und Schätzung eines Konjunkturfrühindikators für Hamburg. *Wirtschaftsdienst*, **83** (8), 540–548.

BANERJEE, A., MARCELLINO, M. and MASTEN, I. (2005). Leading Indicators for Euro-area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics*, **67** (Supplement s1), 785–813.

—, — and — (2006). Forecasting macroeconomic variables for the new member states. In M. Artis, A. Banerjee and M. Marcellino (eds.), *The Central and Eastern European Countries and the European Union*, Cambridge University Press, Cambridge, pp. 108–134.

Barhoumi, K., Brunhes-Lesage, V., Darné, O., Ferrara, L., Pluyaud, B. and Rouvreau, B. (2008). *Monthly Forecasting of French GDP: A Revised Version of the OPTIM Model*. Banque de France Working Papers No. 222.

—, Darné, O., Ferrara, L. and Pluyaud, B. (2012). Monthly GDP Forecasting using Bridge Models: Application for the French Economy. *Bulletin of Economic Research*, **64** (Supplement s1), s53–s70.

Bates, J. and Granger, C. (1969). The Combination of Forecasts. *Operational Research Quarterly*, **20** (4), 451–468.

Bathelt, H. and Glückler, J. (2003). Toward a relational economic geography. *Journal of Economic Geography*, **3** (2), 117–144.

Baum, C., Schaffer, M. and Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, **3** (1), 1–31.

Beaudry, C. and Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, **38** (2), 318–337.

Blien, U. and Suedekum, J. (2005). Local Economic Structure and Industry Development in Germany, 1993-2001. *Economics Bulletin*, **15** (17), 1–8.

—, — and Wolf, K. (2006). Local employment growth in West Germany: A dynamic panel approach. *Labour Economics*, **13** (4), 445–458.

Bodo, G., Golinelli, G. and Parigi, G. (2000). Forecasting Industrial Production in the Euro Area. *Empirical Economics*, **25** (4), 541–561.

Brakman, S. and van Marrewijk, C. (2013). Reflections on cluster policies. *Cambridge Journal of Regions, Economy and Society*, **6** (2), 217–231.

Brambor, T., Clark, W. and Golder, M. (2006). Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*, **14** (1), 63–82.

Brautzsch, H. and Ludwig, U. (2002). *Vierteljährliche Entstehungsrechnung des Bruttoinlandsprodukts für Ostdeutschland: Sektorale Bruttowertschöpfung*. IWH Discussion Papers No. 164.

Breitung, J. and Schumacher, C. (2008). Real-time Forecasting of German GDP Based on a Large Factor Model with Monthly and Quarterly Data. *International Journal of Forecasting*, **24** (3), 386–398.

Brisson, M., Campbell, B. and Galbraith, J. (2003). Forecasting some Low-predictability Time Series using Diffusion Indices. *Journal of Forecasting*, **22** (6-7), 515–531.

Brunow, S. and Hirte, G. (2009). The age pattern of human capital and regional productivity: A spatial econometric study on german regions. *Papers in Regional Science*, **88** (4), 799–823.

Cardoso, F. and Duarte, C. (2006). *The Use of Qualitative Information for Forecasting Exports*. Banco de Portugal Economic Bulletin Winter 2006.

Carstensen, K., Wohlrabe, K. and Ziegler, C. (2011). Predictive Ability of Business Cycle Indicators under Test: A Case Study for the Euro Area Industrial Production. *Journal of Economics and Statistics*, **231** (1), 82–106.

Ca'Zorzi, M. and Schnatz, B. (2010). Explaining and Forecasting Euro Area Exports: Which Competitiveness Indicator Performs Best? In P. de Grauwe (ed.), *Dimensions of Competitiveness*, CESifo Seminar Series September 2010, MIT Press, Cambridge, pp. 121–148.

Chow, G. and Lin, A. (1971). Best Linear Unbiased Interpolation, Distribution and Exploration of Time Series by Related Series. *The Review of Economics and Statistics*, **53** (4), 372–375.

Clark, T. and West, K. (2007). Approximately normal tests for equal predicitve accuracy in nested models. *Journal of Econometrics*, **138** (1), 291–311.

Claveria, O., Pons, E. and Ramos, R. (2007). Business and Consumer Expectations and Macroeconomic Forecasts. *International Journal of Forecasting*, **23** (1), 47–69.

Clements, H. and Hendry, D. (1993). On the Limitations of Comparing Mean Squared Forecast Errors. *Journal of Forecasting*, **12** (8), 617–637.

Clements, M. and Galvão, A. (2009). Forecasting US Output Growth using Leading Indicators: An Appraisal using MIDAS Models. *Journal of Applied Econometrics*, **24** (7), 1187–1206.

Combes, P. (2000). Economic Structure and Local Growth: France, 1984-1993. *Journal of Urban Economics*, **47** (3), 329–355.

Cors, A. and Kuzin, V. (2003). An Approach for Timely Estimations of the German GDP. *AStA Advances in Statistical Analysis*, **87** (2), 201–220.

Costantini, M. and Pappalardo, C. (2010). A Hierarchical Procedure for the Combination of Forecasts. *International Journal of Forecasting*, **26** (4), 725–743.

Croux, C., Dekimpe, M. and Lemmens, A. (2005). On the Predictive Content of Production Surveys: A Pan-European Study. *International Journal of Forecasting*, **21** (2), 363–375.

Dauth, W. (2013). Agglomeration and regional employment dynamics. *Papers in Regional Science*, **92** (2), 419–435.

de Lucio, J. J., Herce, J. A. and Goicolea, A. (2002). The effects of externalities on productivity growth in Spanish industry. *Regional Science and Urban Economics*, **32** (2), 241–258.

Dekle, R. (2002). Industrial Concentration and Regional Growth: Evidence from the Prefectures. *The Review of Economics and Statistics*, **84** (2), 310–315.

Department of Economic and Social Affairs of the United Nations (2006). *Standard International Trade Classification, Revision 4*. Statistical Papers Series M No. 34/Rev. 4.

Diebold, F. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, **13** (3), 253–263.

— and Pauly, P. (1987). Structural Change and the Combination of Forecasts. *Journal of Forecasting*, **6** (1), 21–40.

Dissart, J. (2003). Regional Economic Diversity and Regional Economic Stability: Research Results and Agenda. *International Regional Science Review*, **26** (4), 423–446.

Doz, C., Giannone, D. and Reichlin, L. (2011). A Two-step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering. *Journal of Econometrics*, **164** (1), 188–205.

—, — and — (2012). A Quasi-Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models. *The Review of Economics and Statistics*, **94** (4), 1014–1024.

DRECHSEL, K. and MAURIN, L. (2011). Flow of Conjunctural Information and Forecast of Euro Area Economic Activity. *Journal of Forecasting*, **30** (3), 336–354.

— and SCHEUFELE, R. (2012a). *Bottom-up or Direct? Forecasting German GDP in a Data-rich Environment.* Swiss National Bank Working Papers 2012-16.

— and — (2012b). The Performance of Short-term Forecasts of the German Economy Before and During the 2008/2009 Recession. *International Journal of Forecasting*, **28** (2), 428–445.

DREGER, C. and KHOLODILIN, K. (2007). Prognosen der regionalen Konjunkturentwicklung. *Quarterly Journal of Economic Research*, **76** (4), 47–55.

DRISCOLL, J. and KRAAY, A. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics*, **80** (4), 549–560.

EHRL, P. (2013). Agglomeration economies with consistent productivity estimates. *Regional Science and Urban Economics*, **43** (5), 751–763.

EICKMEIER, S. and NG, T. (2011). Forecasting National Activity using Lots of International Predictors: An Application to New Zealand. *International Journal of Forecasting*, **27** (2), 496–511.

— and ZIEGLER, C. (2008). How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach. *Journal of Forecasting*, **27** (3), 237–265.

ELSTNER, S., GRIMME, C. and HASKAMP, U. (2013). Das ifo Exportklima – ein Frühindikator für die deutsche Exportprognose. *ifo Schnelldienst*, **66** (4), 36–43.

EUROPEAN COMMISSION ECONOMIC AND FINANCIAL AFFAIRS (2014). *The Joint Harmonised EU Programme of Business and Consumer Surveys – User Guide.* Directorate General for Economic and Financial Affairs.

FAHRHAUER, O. and KRÖLL, A. (2012). Diversified specialisation – going one step beyond regional economics' specialisation-diversification concept. *Jahrbuch für Regionalwissenschaft*, **32** (1), 63–84.

Federal Office for Building and Regional Planning (BBR) (2010). *Indikatoren und Karten zur Raum- und Stadtentwicklung, INKAR, Ausgabe 2010*. The Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR) within the Federal Office for Building and Regional Planning (BBR).

Feldman, M. and Audretsch, D. (1999). Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, **43** (2), 409–429.

Fiorito, R. and Kollintzas, T. (1994). Stylized Facts of Business Cycles in the G7 from a Real Business Cycles Perspective. *European Economic Review*, **38** (2), 235–269.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2003). Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area? *Journal of Monetary Economics*, **50** (6), 1243–1255.

—, —, — and — (2005). The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting. *Journal of the American Statistical Association*, **100** (471), 830–840.

Frale, C., Marcellino, M., Mazzi, G. and Proietti, T. (2010). Survey Data as Coincident or Leading Indicators. *Journal of Forecasting*, **29** (1-2), 109–131.

Frankel, J. and Romer, D. (1999). Does Trade Cause Growth? *The American Economic Review*, **89** (3), 379–399.

Fuchs, M. (2011). The determinants of local employment dynamics in Western Germany. *Empirical Economics*, **40** (1), 177–203.

Gayer, C. (2005). Forecast Evaluation of European Commission Survey Indicators. *Journal of Business Cycle Measurement and Analysis*, **2005** (2), 157–183.

German Council of Economic Experts (2008). *Das deutsche Finanzsystem: Effizienz steigern – Stabilität erhöhen*. Expertise commissioned by the Federal Government, German Council of Economic Experts.

German Federal Employment Agency (2010). *Statistics of the Federal Employment Agency – Employed persons subject to social security according to selected branches of the economy, selected regions (place of work)*. Data upon Request, Nuremberg 2010.

Giannone, D., Reichlin, L. and Small, D. (2008). Nowcasting: The Real-time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, **55** (4), 665–676.

Glaeser, E., Kallal, H., Scheinkman, J. and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, **100** (6), 1126–1152.

Granger, C. and Newbold, P. (1973). Some Comments on the Evaluation of Economic Forecasts. *Applied Economics*, **5** (1), 35–47.

Guichard, S. and Rusticelli, E. (2011). *A Dynamic Factor Model for World Trade Growth*. OECD Economics Department Working Papers No. 874.

Hahn, E. and Skudelny, F. (2008). *Early Estimates of Euro Area Real GDP Growth – A Bottom up Approach from the Production Side*. ECB Working Paper Series No. 975.

Hansen, P. (2005). A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics*, **23** (4), 365–380.

—, Lunde, A. and Nason, J. (2011). The Model Confidence Set. *Econometrica*, **79** (2), 453–497.

Hanslin, S. and Scheufele, R. (2014). *Foreign PMIs: A Reliable Indicator for Swiss Exports*. Paper presented at the 2014 IWH Workshop Nowcasting and Forecasting, mimeo.

Harvey, D., Leybourne, S. and Newbold, P. (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, **13** (2), 281–291.

—, — and — (1998). Tests for Forecast Encompassing. *Journal of Business and Economic Statistics*, **16** (2), 254–259.

Henderson, V., Kuncoro, A. and Turner, M. (1995). Industrial Development in Cities. *Journal of Political Economy*, **103** (5), 1067–1090.

Hoechle, D. (2007). Robust standard errors for panel regressions with cross-sectional dependence. *The Stata Journal*, **7** (3), 281–312.

Illy, A., Schwartz, M., Hornych, C. and Rosenfeld, M. (2011). Local economic structure and sectoral employment growth in German cities. *Journal of Economic and Social Geography*, **102** (5), 582–593.

Jacobs, J. (1970). *The Economy of Cities*. Vintage, New York.

Jannsen, N. and Richter, J. (2012). Kapazitätsauslastung im Ausland als Indikator für die deutschen Investitionsgüterexporte. *Wirtschaftsdienst*, **92** (12), 833–837.

Keck, A., Raubold, A. and Truppia, A. (2009). Forecasting International Trade: A Time Series Approach. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, **2009** (2), 157–176.

Kholodilin, K., Kooths, S. and Siliverstovs, B. (2008). A Dynamic Panel Data Approach to the Forecasting of the GDP of German Länder. *Spatial Economic Analysis*, **3** (2), 195–207.

— and Siliverstovs, B. (2006). On the Forecasting Properties of the Alternative Leading Indicators for the German GDP: Recent Evidence. *Journal of Economics and Statistics*, **226** (3), 234–259.

Kluge, J. and Lehmann, R. (2013). Marshall or Jacobs? New insights from an interaction model. *Jahrbuch für Regionalwissenschaft: Review of Regional Research*, **33** (2), 107–133.

Kopoin, A., Moran, K. and Paré, J. (2013). Forecasting Regional GDP with Many Factor Models: How Useful are National and International Data? *Economics Letters*, **121** (2), 267–270.

Krugman, P. (2011). The New Economic Geography, Now Middle-aged. *Regional Studies*, **45** (1), 1–7.

Kuzin, V., Marcellino, M. and Schumacher, C. (2013). Pooling versus Model Selection for Nowcasting GDP with Many Predictors: Empirical Evidence for Six Industrialized Countries. *Journal of Applied Econometrics*, **28** (3), 392–411.

Lee, B., Sosin, K. and Hong, S. (2005). Sectoral Manufacturing Productivity Growth in Korean Regions. *Urban Studies*, **42** (7), 1201–1219.

Lehmann, R. (2015). *Survey-based indicators vs. hard data: What improves export forecasts in Europe?* Ifo Working Paper No. 196.

—, Speich, W., Straube, R. and Vogt, G. (2010). Funktioniert der ifo Konjunkturtest auch wirtschaftlichen Krisenzeiten? Eine Analyse der Zusammenhänge zwischen ifo Geschäftsklima und amtlichen Konjunkturdaten für Sachsen. *ifo Dresden berichtet*, **17** (3), 8–14.

— and Weyh, A. (2014). *Forecasting Employment in Europe: Are Survey Results Helpful?* Ifo Working Paper No. 182.

— and Wohlrabe, K. (2014). Forecasting gross value-added at the regional level: Are sectoral disaggregated predictions superior to direct ones? *Review of Regional Research: Jahrbuch für Regionalwissenschaft*, **34** (1), 61–90.

— and — (2015). Forecasting GDP at the Regional Level with Many Predictors. *German Economic Review*, **16** (2), 226–254.

Longhi, S. and Nijkamp, P. (2007). Forecasting Regional Labor Market Developments under Spatial Autocorrelation. *International Regional Science Review*, **30** (2), 100–119.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, **22** (1), 3–42.

Marcellino, M., Stock, J. and Watson, M. (2003). Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-wide Information. *European Economic Review*, **47** (1), 1–18.

Marshall, A. (1890). *Principles of Economics*. MacMillan, London.

Martin, P., Mayer, T. and Mayneris, F. (2011). Spatial concentration and plant-level productivity in France. *Journal of Urban Economics*, **69** (2), 182–195.

Mukkala, K. (2004). Agglomeration economies in the finnish manufacturing sector. *Applied Economics*, **36** (21), 2419–2427.

Newey, W. and West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55** (3), 703–708.

Nierhaus, W. (2007). Vierteljährliche Volkswirtschaftliche Gesamtrechnungen für Sachsen mit Hilfe temporaler Disaggregation. *ifo Dresden berichtet*, **14** (4), 24–36.

Paci, R. and Usai, S. (2005). *Agglomeration Economies and Growth in Italian Local Labour Systems 1991-2001*. ERSA Conference Papers.

Pesaran, M. (2004). *General Diagnostic Tests for Cross Section Dependence in Panels.* CESifo Working Paper Series No. 1229.

Ragnitz, J. (2009). East Germany Today: Successes and Failures. *CESifo Dice Report*, **7** (4), 51–58.

Robinzonov, N. and Wohlrabe, K. (2010). Freedom of Choice in Macroeconomic Forecasting. *CESifo Economic Studies*, **56** (2), 192–220.

Romer, P. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, **94** (5), 1002–1037.

Rosenthal, S. and Strange, W. (2004). Evidence on the nature and sources of agglomeration economies. In J. Henderson and J. Thisse (eds.), *Handbook of Regional and Urban Economics: Cities and Geography*, Elsevier, pp. 2119–2171.

Schaffer, M. and Stillman, S. (2010). *xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2, xthtaylor.* http://ideas.repec.org/c/boc/bocode/s456779.html.

Schanne, N., Wapler, R. and Weyh, A. (2010). Regional Unemployment Forecasts with Spatial Interdependencies. *International Journal of Forecasting*, **26** (4), 908–926.

Schirwitz, B., Seiler, C. and Wohlrabe, K. (2009). Regionale Konjunkturzyklen in Deutschland – Teil II: Die Zyklendatierung. *ifo Schnelldienst*, **62** (14), 24–31.

Schumacher, C. (2007). Forecasting German GDP Using Alternative Factor Models Based on Large Datasets. *Journal of Forecasting*, **26** (4), 271–302.

— (2010). Factor Forecasting using International Targeted Predictors: The Case of German GDP. *Economics Letters*, **107** (2), 95–98.

Seiler, C. (2014). On the Robustness of the Balance Statistics with Respect to Nonresponse. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, **forthcoming**.

Stock, J. and Watson, M. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, **20** (2), 147–162.

— and — (2004). Combination Forecasts of Output Growth in a Seven-country Data Set. *Journal of Forecasting*, **23** (6), 405–430.

— and — (2006). Forecasting with Many Predictors. In G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, *10*, Elsevier, Amsterdam, pp. 515–554.

— and Yogo, M. (2002). *Testing for Weak Instruments in Linear IV Regression.* NBER Technical Working Paper 284.

Timmermann, A. (2006). Forecast Combinations. In G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, *4*, Elsevier, Amsterdam, pp. 135–196.

van Soest, D., Gerking, S. and van Oort, F. (2002). *Knowledge Externalities, Agglomeration Economies, and Employment Growth in Dutch Cities.* CentER Discussion Paper No. 2002-41.

Vogt, G. (2009). *Konjunkturprognose in Deutschland – Ein Beitrag zur Prognose der gesamtwirtschaftlichen Entwicklung auf Bundes- und Länderebene.* ifo Beiträge zur Wirtschaftsforschung Nr. 36, Ifo Institute – Leibniz-Institute for Economic Research at the University of Munich e.V.

— (2010). VAR-Prognose-Pooling: Ein Ansatz zur Verbesserung der Informationsgrundlage der ifo Dresden Konjunkturprognosen. *ifo Dresden berichtet*, **17** (2), 32–40.

Vosen, S. and Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, **30** (6), 565–578.

Vullhorst, U. (2008). Zur indikatorgestützten Berechnung des vierteljährlichen Bruttoinlandsprodukts für Baden-Württemberg. *Statistisches Monatsheft Baden-Württemberg*, **6** (9), 32–35.

Wallis, K. (1986). Forecasting with an Econometric Model: The 'ragged edge' Problem. *Journal of Forecasting*, **5** (1), 1–13.

Wang, C., Hsu, Y. and Liou, C. (2011). A Comparison of ARIMA Forecasting and Heuristic Modelling. *Applied Financial Economics*, **21** (15), 1095–1102.

Weber, E. and Zika, G. (2013). *Labour Market Forecasting: Is Disaggregation Useful?* IAB-Discussion Paper 14/2013.

Wenzel, L. (2013). *Forecasting Regional Growth in Germany: A Panel Approach using Business Survey Data.* HWWI Research Paper 133.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica*, **68** (5), 1097–1126.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge.

Working Group Regional Accounts VGRdL (2010). *Stock of fixed assets in Germany by Bundesland and East-West-Regions 1991 to 2008*. Working Group Regional Accounts VGRdL, Series 1, Regional results Volume 4, Date of calculation: August 2010, Stuttgart 2011.

Working Group Regional Accounts VGRdL (2011a). *Gross domestic product, gross value added in Germany by Bundesland and East-West-Regions*. Working Group Regional Accounts VGRdL, Series 1, State results Volume 1, Date of calculation: August 2010 / February 2011, Stuttgart 2011.

Working Group Regional Accounts VGRdL (2011b). *Gross domestic product, gross value added in Germany on NUTS-3 Level 1992, 1994 to 2009*. Working Group Regional Accounts VGRdL, Series 2, Regional results Volume 1, Date of calculation: August 2010, Stuttgart 2011.

Curriculum Vitae
Robert Lehmann

seit 01/2016	Wissenschaftlicher Mitarbeiter, ifo Zentrum für Konjunkturforschung und Befragungen
01/2010 - 12/2015	Doktorand, ifo Institut Niederlassung Dresden
04/2014 - 09/2014	Lehrbeauftragter, Technische Universität Dresden
09/2013 - 10/2013	Gastwissenschaftler, BI Norwegian Business School
10/2005 - 01/2010	Studium der Volkswirtschaftslehre, Technische Universität Dresden Abschluss: Diplom-Volkswirt
08/2002 - 07/2004	Sekundarstufe II, Bergstadtgymnasium „Glück auf“ Altenberg Abschluss: Abitur
1. April 1986	Geboren in Dresden