Institut für Arbeitsmarkt- und Berufsforschung

Die Forschungseinrichtung der Bundesagentur für Arbeit

IAB

IAB·Bibliothek

354

Die Buchreihe des Instituts für Arbeitsmarkt- und Berufsforschung

Spatial Dependence and Heterogeneity in Empirical Analyses of Regional Labour Market Dynamics

Norbert Schanne

Dissertationen

wbv

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit

IAB

IAB·Bibliothek 354
Die Buchreihe des Instituts für Arbeitsmarkt- und Berufsforschung

Spatial Dependence and Heterogeneity in Empirical Analyses of Regional Labour Market Dynamics

Norbert Schanne

Dissertationen

wbv

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) eingereicht an der Fakultät für Wirtschaftswissenschaften der Universität Regensburg

Berichterstatter:	Prof. Dr. Dr. h.c. Joachim Möller
	Prof. Dr. Rolf Tschernig
Tag der Disputation:	24.03.2014

Herausgeber der Reihe IAB-Bibliothek: Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit (IAB), Regensburger Straße 104, 90478 Nürnberg, Telefon (09 11) 179-0 ■ **Redaktion:** Martina Dorsch, Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit, 90327 Nürnberg, Telefon (09 11) 179-32 06, E-Mail: martina.dorsch@iab.de ■ **Gesamtherstellung:** W. Bertelsmann Verlag, Bielefeld (wbv.de) ■ **Rechte:** Kein Teil dieses Werkes darf ohne vorherige Genehmigung des IAB in irgendeiner Form (unter Verwendung elektronischer Systeme oder als Ausdruck, Fotokopie oder Nutzung eines anderen Vervielfältigungsverfahrens) über den persönlichen Gebrauch hinaus verarbeitet oder verbreitet werden.

ISBN 978-3-7639-4095-0 (Print)
ISBN 978-3-7639-4096-7 (E-Book)
Best.-Nr. 300882 www.iabshop.de www.iab.de

Contents

Preface

This book is a slightly revised version of my doctoral thesis which was accepted by the University of Regensburg in March 2014. I worked on it while being employed at the Institute of Employment Research (IAB) at Nuremberg.

It would not have been possible to finalize this thesis without the support of many people. First of all, I would like to thank my companion PhD students at the chair for empirical macroeconomics and Rolf Tschernig at Regensburg University, as well as numerous colleagues at IAB Nuremberg and IAB's Regional Research Network for helpful comments and discussions on the papers in this thesis, and their mental support in general. Here, I would like to mention in particular the regional forecasting group: Alfred Garloff, Anne Otto, Rüdiger Wapler, Antje Weyh, and Manja Zillmann. I'm specially grateful to Stefan Bender, Stefan Fuchs, and Anette Haas who gave me the chance to complete this thesis while being employed at IAB. Last but not least I thank Antje König and my daughter Laetitia who made my smile and sing even in the desperate moments known to any PhD student.

List of Figures

List of Tables

Chapter 1

Introduction

One of the major issues of economic research on macroeconomic real quantities, and even more of economic policy, is to deal with the disparities that arise between regional labour markets. To name just a few: be it the rural-urban wage differential or the inter-state income inequalities, be it the unemployment rates known to be uneven between North and South Italy, between Flanders and Wallonia in Belgium, between the pre-1990 Federal Republic of Germany and the Neue Länder, or even within West Germany. Disparities arise not only in the levels of these variables, they are even observed in the dynamics of regional labour markets. For example, unemployment rates have been more or less persistent in the West German high-unemployment regions (e.g. the Ruhr area, the state of Bremen, or the Saar area) for three decades. For East Germany, in contrast, unemployment increased sharply after re-unification in the early 1990s till its first peak around 1997/98, remained until 2004/05 roughly ten percentage points above the West German average, and then declined in most regions (exceptions are the city of Berlin, the eastern part of Mecklenburg-Western Pomerania, the triangle Dessau-Halle-Leipzig and southern Brandenburg around Cottbus) to values not too far above the national unemployment rate. Disparities in the economic and the labour market performance can be observed not only along the national gradients, they arise even at a smaller regional scale. Moreover, the phenomenon of similarly prosperous regions clustering together is observable at various levels of spatial disaggregation as well.

The geographical allocation of labour market activity in the long run can be attributed to a large extent to the endowment of regions with natural resources and amenities (e.g. Marston, 1985; Moretti, 2011). In the response to shocks, however, it is not clear why endowment-based models should not behave like an archipelago of 'island economies' as sketched by Lucas/Prescott (1974) where the unmatched workers are randomly redistributed across the islands. To date, the geographically clustered dynamics in the labour markets can be explained only by models accounting for spatial economics; two major strands in this literature, New Economic Geography (e.g. Fujita/Krugman/Venables, 1999) and Urban Labour Economics (e.g. Zenou, 2010), will be introduced briefly below. Both model families have two features in common which seem to be particularly important for the present work: First, a shock in a single location and the corresponding shift in relative prices for domestic vs. non-domestic goods (or for land, respectively) will cause adjustment processes in all regions as long as distance-related costs are neither negligible nor extremely large. The adjustment dynamics are interdependent across the regions; they form geographical patterns which are, in accordance with the transportation network, spatially structured. Second, neither randomly generated shocks nor targeted policy interventions can be expected to have a homogeneous effect on all regions.

Empirical economic research is typically interested in identifying and quantitatively assessing a single (homogeneous or average) parameter and its distribution (e.g. a model parameter like an elasticity, an average treatment effect and its confidence bands, or mean and variance of the one-year ahead growth rate forecasts). In addition, the properties of most statistical and econometric models hinge on some cross-sectional independence assumptions. However, taking the previously described features seriously, any stochastic model for multi-regional data (or a data-generating process, DGP) should account for the cross-sectional heterogeneity in causal relationships and for cross-sectional interdependence, or at least test for both. Early attempts like Granger (1969) or Bronars/Jansen (1987) are still close to time-series analysis in their design. Nowadays state of the art is to employ spatially autoregressive models in which the cross-sectional interdependence is accounted for by a weighted average over adjacent or otherwise close regions. E.g., Molho (1995) and Niebuhr (2003) use spatial lags of unemployment and employment to model the propagation of labour market shocks over space; Aldashev (2012) estimates a regional search-matching model which allows for commuting and in which vacancies in neighbouring regions affect the local labour-market tightness; looking beyond unemployment or employment, Schanne/Weyh (2009) use a spatial lag in the start-up rate to explicitly model collocation effects of new firms while controlling for other spatially interdependent agglomeration effects. Many studies, however, consider spatial autocorrelation solely because of some specification test suggesting any unexplained spatial patterns, or because of plain intuition that something spatially interdependent (like agglomeration economies) might be inherent in the data. Neither do they provide an exact definition of what the effect should account for and what not, nor do they consider alternative strategies how to model these spatial effects more precisely, how to identify them. Hence, these studies may be affected by under-identification, with the consequence that their estimates as a whole are in question, that their analysis becomes "mostly pointless" (Gibbons/Overman, 2012) due to spatial econometrics.

What I intend to do throughout the subsequent studies is to analyse certain aspects of regional labour market dynamics with a particular focus to the spatial heterogeneity of effects and the spatial interdependence. Though each paper pursues a distinct economic question, there exist some aspects regarding the econometric design fundamental to all:

- Which components of a certain process require exact specification in order to find an answer to the question under investigation, and which can be dealt with in a flexible, little restrictive fashion?
- What can I do to warrant identification of this process and to establish robustness of the identification strategy?

- Does the information on the spatial location of observations (and the collocation of others) add anything to what I can learn without this information?

In answering these questions, the thesis proceeds as follows. Chapter 2 sketches briefly the central arguments of spatial economic theory regarding the labour-market allocation in space. Its aim is to give support to the two fundamental hypotheses of this work that economic dynamics are spatially heterogeneous and that the evolution of several quantities (and prices) is interdependent across space.

Chapter 3 surveys a number of econometric methods which account for either cross-sectional dependence or cross-sectional heterogeneity. The models have in common the utilization of some information on the collocation of observations in (typically geographic) space. This information, which is collected in a *spatial weights matrix*, allows me to impose structure on both the interdependence relations between observations at different locations and the parameters characterizing relations to explanatory variables. I first provide a classification of cross-sectional interdependencies possibly arising in the data, and a classification of various types of parameter heterogeneity. I sketch in brief the three workhorses of spatial econometrics, the *spatial lag*, the *spatial error* and the *spatial Durbin model*, and alternative methods to estimate the respective spatial autocorrelation parameters. However, as I will argue, these mainstream models have two major disadvantages: first, the estimated parameters and, to a smaller extent, even the estimated partial effects depend on the choice of the spatial weights matrix (and its correct definition). Second, identification hinges crucially on restrictions which are rarely made explicit. Hence, we discuss some non-parametric methods which require weaker assumptions, spatial filtering and a spatial analogue to HAC variance-covariance estimation. They may be considered a robust alternative as long as the focus is not on the identification of the spatial autocorrelation parameter or on calculating spatial interdependence effects. Thereafter, I present some methods which allow for spatially structured heterogenous regression parameters. The chapter closes with some comments on detecting and distinguishing spatial heterogeneity and dependence jointly.

Chapter 4, "Persistence of Regional Unemployment: Application of a Spatial Filtering Approach to Local Labour Markets in Germany", presents a study carried out in cooperation with Roberto Patuelli, Daniel A. Griffith and Peter Nijkamp. We are interested in the degree of persistence of shocks on local unemployment, that is on hysteresis, con- and divergence.[1] In our investigation we follow the

1 A similar question regarding the persistence of unemployment differentials in German regions is dealt with by Kunz (2012: Ch. 3) he investigates local deviations from the national unemployment rate with homogeneous panel unit-root tests without cross-sectional dependence.

standard approach, testing for stationarity of the time series. The established procedures – e.g. assuming one homogeneous coefficient (e.g. Levin/Lin/Chu, 2002), a small number of individual coefficients (e.g. Sarno/Taylor, 1998), or a combination of independent individual test statistics (e.g. Im/Pesaran/Shin, 2003; Maddala/Wu, 1999) – do not account for the previously mentioned characteristics of regional dynamics. Hence, we develop a spatial-filter geographically-weighted regression (SF-GWR) equivalent (Griffith, 2008) for serially correlated processes by interacting an eigenfunction-decomposition spatial filter with an AR(1) process. This accounts for spatial heterogeneity in the spatiotemporal adjustment process without losing too many degrees of freedom. In addition, it has the advantage to provide information on the spatial patterns within the adjustment to local shocks since the eigenvectors reflect mapping characteristics. Results are compared with a homogenous parameter estimate, individual coefficients and regime-specific coefficients (differentiated according to the BBSR agglomeration typology). When analysing local German unemployment rates at quarterly frequency from 1996 to 2004, we find generally high though widely heterogeneous persistence in local unemployment rates. Moreover, the heterogeneity in the adjustment speed after shocks can be assigned to a large extent to eigenvectors – they can be interpreted as factor loadings relating spatial structures with the data – which show three to four peaks and three to four troughs when they are plotted in a map. This hints at persistence as a regional and not as a national or local phenomenon.

In Chapter 5, "Forecasting Regional Labour Markets with GVAR Models and Indicators", I investigate whether accounting for the joint development of close regions provides information sufficiently rich to improve the forecast accuracy. The model used in the paper builds strongly upon Pesaran/Schuermann/Weiner (2004) and Chudik/Pesaran (2011); it deals with the necessary and adequate aggregation or disaggregation of regions in order to make a multiregional VAR econometrically feasible.[2] Important econometric aspects are the distinction between semi-strong and strong cross-sectional dependence, the construction of spatial aggregates, as well as the handling of the non-stationarity inherent in the data. The predictive content of the spatially interacted variables is compared with the information content of various leading indicators. Evaluation criterion is the relative forecast error in simulated out-of-sample forecasts for the Federal Employment Agencies Regional Divisions (RDs) at the 3- and 12-month horizon. Germany turns out to have no economically dominant region (which reflects the polycentric structure of the country), and the RDs do not follow a joint stable long run trend which

2 An unrestricted Vector Autoregressive (VAR) model is not tractable because of the number of paramaters growing quadratically in the number of regions; a spatially autoregressive Panel-VAR as provided by Mutl (2005) seems too restrictive since it does not allow for heterogeneous parameters.

could be used to implement cointegration. Accounting for spatial dependence improves the forecast accuracy compared to a model without spatial linkages while using the same leading indicator. Amongst the tested leading indicators, only few produce more accurate forecasts when included in a GVAR model, than the GVAR without indicator.

In Chapter 6, "Do they Run with the Pack? The Formation of Experts' Expectations on Labour Markets", I investigate if cross-sectional dependence arises even in economic expectations. The spatial information of an economic tendency survey collected amongst local labour market experts allows in principle identification of the endogenous peer effect as a measure for herding. Hence, the focus here is indeed on quantitatively assessing the parameter of cross-sectional dependence which requires exact specification of the spatial autocorrelation process. The model includes, in addition to local unemployment and vacancies, the spatial lags of unemployment and vacancies as regressors representing the contextual peer effect. I evaluate robustness with regard to the measure for spatial (or administrative) relatedness and alternative error-component structures, and establish the validity of my instrumentation strategy. Considering hence the herding effect as identified, I employ the same estimation strategy to analyse several aspects of expectation formation and social learning: The size of the herding effect is robust if I account for the available short-run expectations of forecasting experts. The effect collapses at the time when these experts agree on a trend reversal in their published forecasts, and re-establishes thereafter; this hints at an informational cascade. Constructing counterfactual sentiments without herding, I find that, despite the threat of informational cascading, economic expectations potentially benefit from social learning.

I close this work with a short resume in Chapter 7.

Chapter 2

Geography Matters

The Impact of Distance in Regional Labour Market Models

2.1 Spatial structure by workplace – New Economic Geography

Within the last twenty years, New Trade Theory (NTT) and New Economic Geography (NEG) have overcome the spatial impossibility theorem

> *"Consider an economy with a finite number of locations and a finite number of consumers and firms. If space is homogeneous, transport is costly and preferences are locally nonsatuated, then there is no competitive equilibrium involving transportation."*
>
> Starrett's 1978 theorem, according to Ottaviano/Thisse (2004).

by introducing distance and transport costs, combined with local increasing returns to scale into neoclassical models; in this, they contributed essentially to a theoretically closed explanation of a number of observed phenomena, like the spatial distribution of production or the formation of cities. The theoretical models generate economic equilibria where industrial production is allowed to be endogenously concentrated or distributed over space; the number and the location of the industrial cores is determined by transparent, elaborated centrifugal and centripetal forces (see, for example, Fujita/Krugman/Venables, 1999). These mechanisms determine not only the equilibrium allocation of production and population, they are also at work in the stylized ad-hoc dynamics towards the equilibrium. If the economy in a region deviates from its equilibrium allocation, this affects not only the dynamics within this region – it changes also the price vector to which other regions adapt. In a simple formulation this can be seen from equation 5.2 in Fujita/Krugman/Venables (1999), where the change of region r's share on the worlds manufacturing labour force λ_r depends on the difference between the current real wage w_r in region r and the average real wage $\bar{\omega} = \sum_r \lambda_r \omega_r$:

$$\dot{\lambda}_r = \gamma \left(\omega_r - \bar{\omega} \right) \lambda_r \tag{2.1}$$

The real wage

$$\omega_r = w_r G_r^{-\mu} = \left[\sum_s Y_s T_{rs}^{1-\sigma} G_s^{1-\sigma} \right]^{1/\sigma} G_r^{-\mu} \tag{2.2}$$

itself is a function of the interregional transport costs T_{rs}, other regions production Y_s and their price vectors G_s. Thus, both the level of regional employment and its dynamics depend on the spatial connectivity between regions. However, the focus of most NEG models is on the location of production. They in general assume full employment; another, more realistic situation on labour markets (where involuntary unemployment exists) has been left aside for a long time.

2.2 Structure by residential location – Urban Labour Economics

To my knowledge, explicitly spatial aspects within modern labour-market theory – in which regionally varying unemployment may exist in equilibrium – have first been considered by Zenou/Smith (1995), van Ommeren/Rietveld/Nijkamp (1997), Wasmer/Zenou (1997) and, thereafter, by several succession papers, see Zenou (2009). In this strand of literature, search-matching models (or, less often, other models with equilibrium unemployment, e.g. efficiency wages) are augmented by an Alonso-Mills-Muth housing market (for a general presentation of these urban models without labour market, see Brueckner, 1987). Firms are situated in the Central Business District (CBD), and persons have to commute from their residential location to the CBD as their working place or the place where they get information about vacant jobs. Commuting is related with costs and reduces the net benefit of a job. Thus, the further away from the CBD a person lives the less likely she is to find a profitable job match and the more likely she is (long-term) unemployed. However, it becomes easier to find a job the closer to the centre an unemployed resides, thus it might be interesting for an unemployed person to move towards the CBD and to outbid employed persons on the housing market (in particular, if the quantity of housing consumed by an unemployed is smaller than that consumed by an employed person). In this basic setting, it is possible to identify (with relatively simple frameworks) three different stable location patterns (see Zenou, 2009: p. 48–49):

1. A central core of employed surrounded by a peripheral ring of unemployed with low search intensity *(segregated equilibrium)*,
2. A central core of (short-term) unemployed with high search intensity surrounded by a peripheral ring of employed *(integrated equilibrium)*,
3. Both a central core and a peripheral ring of unemployed separated by an intermediate ring of employed *(Core-Periphery equilibrium)*.

By this, even though still requiring a given central business district (i.e. an exogenously determined location of firms), a closed theoretical framework explaining the intra-urban "spatial mismatch hypothesis" as an economic equilibrium is established. Extensions of these models have been developed, considering rural-urban migration (in addition to the intra-urban commuting), the existence of (and thus, the choice of workplace between) multiple CBD's or Suburban Business Districts or the choice between occupational and regional mobility. The perspective in Zenou (2009) is mostly on the employee and the non-coincidence (spatial divergence) of her places of work and residence. All these models have in common to result in an equilibrium (reservation/non-shirking/market) wage which is related to the unemployment benefit w_u plus the wage mark-up typical for the regarding model family and, in addition to this, a spatial mark-up which depends on the search intensity s, the costs per unit of distance τ and the space occupied by the employed persons $L=\left(1-u\right)N$ (with N the population size and u the unemployment rate) and thus on the maximum distance an employee has to commute[3]. As a consequence, the average unemployment rate depends on these spatial costs as well. For example, in the standard search-matching model in the line of Mortensen/Pissarides (1999) and Mortensen (2003), the dynamics equation of unemployment (as stock, with constant population) is determined by the inflows into unemployment (the number of destructed jobs, where δ is the job destruction rate) minus the outflows out of unemployment (the newly created job matches, with a_U the job offer rate and $F\left(w_L^r\right)$ the cumulated density function of reservation wages):

$$\frac{du(t)N}{dt}=\delta\left[1-u\left(t\right)\right]N-sa_U u\left(t\right)\left[1-F\left(w_L^r\right)\right]N \tag{2.6}$$

The steady-state (equilibrium) unemployment rate follows from setting this expression for the change in unemployment equal to zero. As easily can be seen, there is a (space-cost dependent) reservation wage – and thus a "critical distance" – which satisfies this condition.

Due to perfect intra-region housing market with bid rents, the models generate stable unemployment disparities across different locations. At the extremum, when the housing market within a region is perfect and reacts instantly, newly unemployed move immediately according to their bid-rent. Then, two (or more) zones form with homogenous employment status (employed or unemployed).

Models with multiple regions, distance costs (e.g. due to trade or commuting), equilibrium unemployment and (exogenous or endogenous) differences in

---

3 For example, in the urban-rural two-region search model with wage-bargaining of Zenou (2009: Ch. 3.2), the wage in the city C is, with β the bargaining power, c the cost for posting a vacancy and θ the labour market tightness,

$$w_L^{C*}=\beta\left[y^C+cs\theta^{C*}\right]+\left(1-\beta\right)\left[w_U^C+\left(1-s\right)\tau L^{C*}\right] \tag{2.3}$$

in the two-city efficiency wage model provided by Zenou/Smith (1995) (see also Zenou, 2009) for city k it is, with e the effort, $m(\cdot)$ the detection function (which describes the probability to be detected while shirking), δ the job destruction rate, and r the discount rate (the interest rate),

$$w_L^k=w_u+e+\frac{e}{m\left(\frac{L^k}{M}\right)}\left(\frac{\delta N^k}{N^k-L^k}+r\right)+\left(1-s\right)\tau L^k \tag{2.4}$$

In the single-region search model with wage-posting and heterogeneous workers of Zenou (2011), the reservation wage follows as

$$w_L^r=w_U+\left(1-s\right)\tau L+\frac{sa_U}{r+\delta}\int_{w_L^r}^{\infty}\left(w_L-w_L^r\right)dF\left(w_L\right) \tag{2.5}$$

productivity and mobile population (e.g. Zenou/Smith, 1995, Epifani/Gancia, 2005 and Suedekum, 2005) allow in addition to analyse the effects between unemployment and migration – which for long time has been seen as adjustment mechanism balancing unemployment differentials across the country although the empirical evidence was not too clear (e.g. Pissarides/Wadsworth, 1989). If there are differences in productivity (i.e. there is no symmetric equilibrium), and if there exist congestion costs sufficiently high such that population in the less productive region does not completely move to the higher productive region, equilibrium unemployment will diverge between regions. Furthermore, it is unclear whether a decline in distance-related costs leads to convergence in the unemployment rates or to increasing differentials across regions.

The strands of argumentation in the literature cover on the one hand the endogenous arising of bifurcating core/periphery structures in the distribution of wages, unemployment rates, employment opportunities and so on due to agglomeration effects (Epifani/Gancia, 2005; Suedekum, 2005; vom Berge, 2012; Zierahn, 2012, with mechanisms similar to the initially described Krugman model), and on the other hand the genesis of the land distribution between employed and unemployed persons for given city centres (Zenou/Smith, 1995). Supposedly, the findings should hold in a world where firms are initially continuously (or randomly) distributed over space, where each employer has a certain "labour shed" (a circle from which its employees commute) around his firm and each employee has an "employment field" (her search/commuting radius) around her residence, when there is on-the-job search, and when shocks to unemployment (negative or positive) spread along a vacancy chain over space, just to mention some of the mechanisms which Morrision (2005) describes as essential for adjustment at the sub-national level.

The spatial pattern in the adaption to local shocks is demonstrated in an easily understandable fashion in Moretti (2011: Ch. 3.1). He shows that in a two-city economy the benefit from a positive productivity shock in region b accrues to workers in region b only if labor is completely immobile (that is with enormously high commuting costs). Otherwise, shares of the productivity gain can be appropriated by land-owners in region a, land-owners in region b or even workers in region a, depending on the mobility of workers and the elasticity of housing in the two regions.

At the end of the day, we have seen that the regional distribution of employment and unemployment may be affected by space and through space (in the sense of geographical location and distance between locations); this applies to the labour-market equilibrium, the current off-equilibrium distribution, and to the dynamics as well. In particular adjustment processes are likely spatially interdependent. Furthermore, these effects of space may result in heterogenous economic relations or heterogenous impact of economic policy across regions. In the following I argue that accounting for these spatial relations may require econometric models distinct from the most frequently applied linear regression model.

Chapter 3

Spatial Econometrics

An Introduction in the Econometrics of Cross-sectional Dependence and Cross-sectional Heterogeneity

3.1 Introduction

I have argued in the previous chapter that a typical regional economics setting is characterized by two elements: heterogeneity of effects and interdependence of variables across observations. Let $y_{i,t}$ be a cross-sectionally interdependent double-index process with $i=\{1,\ldots,n\}$ and $t=\{1,\ldots,T\}$ supposedly affected by up to n contemporaneous observations of k exogenous variables $x_{i,t,k}$. To account for the previously suggested spatially-economic structures, it would be adequate to specify a separate equation for each cross-sectional unit i and, for estimation, stack these n equations simultaneously in a system:

$$\Pi Y_{(n)t} = \sum_{k=1}^{K} B_k X_{(n)t,k} + U_{(n)t} \tag{3.1}$$

with $U_{(n)t} \sim (0,\Sigma)$, $Y_{(n)t} = (y_{1,t},\ldots,y_{n,t})'$, $X_{(n)t,k} = (x_{1,t,k},\ldots,x_{n,t,k})'$ and $U_{(n)t} = (u_{1,t},\ldots,u_{n,t})'$ as $n\times 1$ vectors. The system is in general estimated in reduced form, that is with Π^{-1} pre-multiplied on both sides of the equation; parameters are identified from restrictions on the Cholesky-decomposition $\Sigma^{\frac{1}{2}}$ and B_k.[4] Note that the matrices of slope parameters B_k and the symmetric disturbances' covariance matrix Σ have dimension $n\times n$. Hence, the parameters in eq. (3.1) are identified only if T increases at least at a rate of n^2. That is, fully flexible estimation requires the time dimension large relative to the cross-sectional dimension – a condition typically not satisfied with most regional data.

If the parameter matrices $A_k = \Pi^{-1}B_k$ are diagonal (with off-diagonal elements restricted to equal zero), eq. (3.1) describes a system of seemingly unrelated regressions (SURE). This reduces the number of parameters enormously; because of the non-zero off-diagonal elements in Σ, however, the request on the time dimension remains. A number of estimators have been developed with weaker requirements on the time dimension: e.g., the Dynamic Factor Models (DFM) discussed discussed by Forni et al. (2000, 2004), Stock/Watson (2002, 2011), Bai (2004), the Correlated Common Effects (CCE) estimator developed by Pesaran (2006), or Factor-Augmented VARs and VECs (Bernanke/Boivin/Eliasz, 2005; Banerjee/Marcellino, 2009; Banerjee/Marcellino/Masten, 2009). In DFMs the system's cross-sectionally correlated development is due to the dependence on joint latent variables described by the measurement equations in a State-Space model, or to the loadings of principal components. State-space models typically demand T to increase to infinity, n to be fixed, and additionally independence of

4 For an introduction, see see Mariano (2001). Identification demands (orthogonality or exclusion) restrictions, e.g. the utilization of instruments in three-stage least squares (3SLS). For dynamic systems, see the discussion on identification of VARs and Structural VARs, e.g. in Lütkepohl (2005: Ch. 9, 10).

the idiosyncratic components in the DFM. Approximate or principal-component based factor models require $n \to \infty$ and $T \to \infty$ (see e.g. Bai, 2004). The CCE combines the individual slopes of a SURE approach with cross-sectional factor dependence; the distribution of the slopes is established under the condition that n and T increase at the same rate.

The data employed in most regional studies at most do not suggest applicability (or adequacy) of T asymptotics: frequently, the observation period is short, and the cross sectional dimension exceeds the time dimension by far. At the extremum, models have to be estimated with data observed in a single period. The strategies for restricting cross-sectional dependence and heterogeneity in the system estimators hint already at the direction which is likewise pursued in the literature on spatial econometrics: reduced-form estimation or orthogonalization of the disturbances, zero-restrictions on the off-diagonal slopes, linear restrictions on slopes, eigendecomposition of the covariance matrix etc.

Because of the short data frequently available for applied research, it makes sense for the following discussion to focus on methods which are applicable in cross-sections. I will use the linear regression model to ground the subsequent discussion on; because of its desirable properties and the well-established necessary and sufficient conditions, other estimators try to show how similar features can be generated. Hence, a systematic classification of the dependence and parameter structures in a spatial setting might allow to identify the relevant estimation issues.

3.2 The classical linear regression model – and looking beyond

Let y_i denote observation i's realization of the dependent variable y in a regression model which has to be explained by a set of covariates x collected for observation i in a vector $x_i = (x_{i,1}, \ldots, x_{i,K})$; x_i is defined such that the relation between y_i and x_i is linear in the parameter vector $\beta = (\beta_1, \ldots, \beta_K)'$. Furthermore, let the regression model enclose an additive stochastic disturbance u for which the realization corresponding to observation i is denoted with u_i. If y is continuous and observable over the entire real space $\mathbb{R}$, we can write the multivariate linear regression model for a single cross-section of data as

$$y_i = x_i\beta + u_i \tag{3.2}$$

or, in matrix notation, as

$$Y_{(n)} = X_{(n)}\beta + U_{(n)}$$

where $Y_{(n)} = (y_1, \ldots, y_n)'$, $X_{(n)} = (x_1', \ldots, x_n')'$ and $U_{(n)} = (u_1, \ldots, u_n)'$. The Ordinary Least Squares (OLS) estimator for the parameters in eq. (3.2) is, in matrix and summation notation:

$$\hat{\beta}_{OLS} = (X_{(n)}'X_{(n)})^{-1}(X_{(n)}'Y_{(n)}) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) \tag{3.3}$$

Besides implicitly (or, sometimes, even explicitely) stating that
i) the observation range of the dependent variable is not limited: $y \in (-\infty, \infty)$,
ii) the model is complete in the set of relevant covariates,
iii) the model is linear in the parameters,
iv) there exists only a single set of (true) parameters that are homogenous across observations,

introductory econometric textbooks refer to a set of convenient assumptions (e.g. Greene, 2000: Ch. 6 and Ch. 9, Ass. 1–6). These allow to establish easily desired properties of OLS estimates. Ass. 1(a-c) below are known as *Gauss-Markov* assumptions warranting that OLS is the best linear unbiased (BLU) estimator.

Assumption 1 (Classic linear regression model). *For a linear regression as given in eq. (3.2):*
a) *The explanatory variables x are strictly exogenous (i.e. non-stochastic). The $n \times K$ matrix entailing these covariates $X_{(n)}$ has full rank, i.e. $rk(X_{(n)}) = K$.*
b) *The stochastic disturbance is distributed independently to the covariates. It has an expectation conditionally on x equal to zero, i.e. $E(u) = E(u\,|\,x) = 0$.*
c) *The disturbance is both homoscedastic (constant variance) and serially/cross-sectionally uncorrelated (zero auto-covariance), i.e. $V(u) = E(uu') = \sigma_u^2 I_{(n)}$ (where $I_{(n)}$ denotes an identity matrix of dimension n).*
d) *The disturbance has a gaussian distribution.*

Ass. 1a ensures that the covariates' $K \times K$ cross product matrix $X_{(n)}'X_{(n)}$ is nonsingular and finite; thus, its inverse exists. Ass. 1b warrants that OLS is unbiased. Ass. 1c simplifies the variance-covariance matrix of the OLS estimator. And under 1d, OLS estimates have a normal distribution even in finite samples.

Weaker assumptions are however sufficient to ensure (for $n \to \infty$) consistency and asymptotic normality: Davidson (1994) or Pötscher/Prucha (2001) provide various (Weak) Law of Large Numbers [(W)LLN] and Central Limit Theorems (CLT) and their respective conditions. Ass. 1a can be weakened such that x is a vector of (independently distributed) random variables with finite second population moments $E(xx') = C$ where C a positive definite and henceforth invertible matrix. Then, a WLLN applied to their sample covariance matrix states that $E\left(\frac{1}{n}X'_{(n)}X_{(n)}\right)_{n\to\infty} \xrightarrow{p} E(xx') = C$. Depending on the framework, the appropriate CLT and LLN, Ass. 1b can be relaxed, e.g. to u being a mean-zero martingale difference sequence, or an independently distributed mean-zero heteroscedastic random variable which is uncorrelated with x and for which some moments higher than the second exist (the Lyapunov condition), or to u being deducable to an independent process by a transformation such that convergence can be derived from a Cramér–Wold device.

Table 3.1: Typology: Cross-sectional relations and dependencies

Structure	Assumptions	Consequence
■ Correlation of regressors: $E(x_i \cdot x_j) \neq 0$	Implicit ass. hold, for exogenous x Ass. 1a–d hold. The weak analogue to 1a may be violated if some (spatially) covariance-instationary elements in x exist.	At most, OLS will be unbiased, consistent and efficient.
■ Cross-dependence on regressors: $\exists x_j : x_j \to y_i$	Model in eq. (3.2) incomplete (implicit Ass. ii violated).	Omitted variable bias?
$\Rightarrow E(y_i \mid x_i) \neq E(y_i \mid x)$	If x_j (or $\Sigma_j a_{ij} x_j$, with exogenous real-valued weights a_{ij}) is included in the regression, multi-collinearity may arise (Violation of Ass. 1a).	Parameter identification not warranted, depending on weights a_{ij}
■ Correlated disturbances: $\exists y_j : \{y_j \to y_i \wedge y_j \nrightarrow E(y_i \mid x)\}$ $\Leftrightarrow E(u_i \cdot u_j) \neq 0$	Violation of Ass. 1c	OLS unbiased and consistent but not efficient. Correlation between y_i and y_j but not between y_i and $E(y_j \mid x)$
■ Regressand's simultaneity: $\exists y_j : \{y_j \to y_i \wedge y_j \to E(y_i \mid x)\}$	Model in eq. (3.2) incomplete (implicit Ass. ii violated).	Omitted variable bias?
	If y_j or $\Sigma_j a_{ij} y_j$ is included as variable on the right hand, Ass. 1b is violated.	OLS is biased. If dependence is not unidirectional ($y_j \to y_i$ and $y_i \to y_j$), OLS is even inconsistent because of endogeneity/simultaneity.

In this table an arrow $a \to b$ denotes the existence of an effect from a on b.

I provide an overview on various forms of cross-sectional dependence between pairs of observations in Table 3.1. Dependence structures where only (to the model) exogenous variables x are involved typically won't result in serious statistical problems with the estimation as long as they are accounted for (i.e. if they do not cause an omitted variable bias) and as long as the parameters in the model are still identified. Thus, I will focus later on interdependencies arising in the endogenous components, u and y. In general, cross-sectional dependencies are bi-directional: if each y_i (or u_i, respectively) for $i=\{1,\ldots,n\}$ depends on y_j (or u_j) for $j=\{1,\ldots,i-1,i+1,n\}$, y_j depends as well on y_i. The number of dependence relations increases quadratically in the sample size. Hence, restrictions will be necessary to reduce the complexity of the system and to overcome the curse of dimensionality.

In some frameworks, a single leading unit (region/individual) is modeled to affect all other units:

$$y_i = \alpha c_{i0} y_0 + \nu_i \qquad \forall i = \{1,\ldots,n\} \tag{3.4}$$

I.e., this unit works as an observable factor which has to be exogenous with respect to all other observations (see e.g. Pesaran/Schuermann/Weiner, 2004 and Holly/Pesaran/Yamagata, 2011); this is contradictory to our previous statement that spatial relations are typically bidirectional. Chudik/Pesaran/Tosetti (2011) denote this case as cross-sectionally strong dependent (CSD), with the dominant unit indexed with 0. In cross-sectional data, CSD sometimes is dealt with by restricting both α and c_{i0} to one and to investigate only the differential of region i to the dominant region. Alternative restricting weights c_{i0} (e.g. the physical distance to the dominant observation) may even be imposed as to get variation in the impact of y_0; otherwise, the influence of y_0 is not separable from the constant and α is not identified. Thus, cross-sectionally strong dependence is in general considered only with panel data.

More frequently it is assumed that y_i depends on a weighted average over all y_j and a cross-sectionally independent remainder υ_i:

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \upsilon_i \tag{3.5}$$

Models like this are denoted as Cliff-Ord type, acknowledging their early discussion by Cliff/Ord (1972, 1973). Herein, the weights w_{ij} typically reflect some kind of exogenous relatedness or distance (see for example Conley/Topa, 2002, Haining, 2003, Pesaran/Schuermann/Weiner, 2004 and associated comments in the same JBES volume, or Corrado/Fingleton, 2012). Metrics like commuting or trade flows are in general closer to the underlying economic theory than geographical metrics, but are at best weakly exogenous. Frequently employed weighting-scheme families in regional studies rely either on a distance-decay function (e.g. inverse distance between regions i and j, eventually with a certain threshold distance beyond which a dependence relation is set to zero) or on contiguity (adjacent regions are first-order neighbours, regions which can be entered after passing through an adjacent region are second-order neighbours). For various estimators presented in subsequent sections, consistency and asymptotic distribution have been analysed under a unifying set of assumptions regarding these spatial weights:

Assumption 2 (Spatial weights). *Let $W_{(n)}$ denote the (positive definite) $n \times n$ matrix of triangular sequences of non-stochastic spatial weights w_{ij} satisfying:*

a) $w_{ii}=0 \quad \forall i \in \{1,\ldots,n\}$ *, i.e. the main-diagonal elements of W are zero.*
b) $w_{ij} \geq 0 \quad \forall i,j \in \{1,\ldots,n\}$ *, i.e. there are no negative spatial weights.*
c) *The row norm and column norm of $W_{(n)}$ are bounded in absolute value.* [5]

Sometimes, it is further assumed that $w_{ij} = O\left(\frac{1}{n}\right)$, i.e. that the weight of a single (additional) regions becomes negligible as $n \rightarrow \infty$. This can be achieved by adequate standardization (e.g. row, eigenvalue, or population-share standardization; see Mutl, 2006) of any matrix satisfying Ass. 2a–c.

Remark 1. The choice of weights determines the interpretation and the identification of the spatial autocorrelation parameter in eq. (3.5); I will demonstrate the sensitivity with regard to the weighting scheme with two examples. First, let $W_{c(n)}$ a n-dimensional first order contiguity matrix. Eq. (3.5) in matrix notation is

$$Y_{(n)} = \rho W_{c(n)} Y_{(n)} + U_{(n)} = (I_n - \rho W_{c(n)})^{-1} U_{(n)} \tag{3.6}$$

where the second equality stands for the infinite spatial moving average [SMA(∞)] representation. The expression $\rho W_{c(n)}$ describes the aggregate influence from adjacent regions. $\rho W_{c(n)}$ can be used to construct a spatial effect $\tilde{\rho} W_{d(n)} = (I_n - \rho W_{c(n)})^{-1} \rho W_{c(n)}$. The latter is by nature related to the further; however, it describes an exponentially declining distance-decay, similar to the Iceberg transportation costs as used in NEG models. Both $W_{c(n)}$ and $W_{d(n)}$ can satisfy Ass. 2 at the same time and describe the same geographical space (such that there is no formal rational to rely more on one). Though, the corresponding parameters represent different economic relations. As a second example, consider a weights matrix $W_{a(n)} = \frac{1}{a} W_{c(n)}$ with constant a chosen such that $W_{a(n)}$ still satisfies Ass. 2. Then the parameter $\breve{\rho} = a\rho$ is for any $a \neq 1$

5 The column norm of matrix $W_{(n)}$ is defined as $\left\| W_{(n)} \right\|_1 = \max_{i \in \{1,\ldots,n\}} \sum_{j=1}^{n} \left| w_{ij} \right|$, the row norm as $\left\| W_{(n)} \right\|_\infty = \max_{j \in \{1,\ldots,n\}} \sum_{i=1}^{n} \left| w_{ij} \right|$ and the spectral norm (the analogue to the Euclidean norm for square matrices) as $\left\| W_{(n)} \right\| = \left[\max \left\{ \lambda_{\left(W'_{(n)} W_{(n)}\right)} \right\} \right]^{\frac{1}{2}}$, where $\max\left\{\lambda_{(A)}\right\}$ is the largest eigenvalue of a symmetric positive-semidefinite matrix A. See Chudik/Pesaran/Tosetti (2011).

distinct from ρ although both weighting schemes $\breve{\rho} W_{a(n)}$ and $\rho W_{c(n)}$ result in the same averages.

As it is custom in spatial econometrics, I adapt the terminology from time-series analysis to describe cross-sectional dependence structures. I will denote the weighted average over all other observations as cross-sectional or spatial lag – depending on the model component we say *spatially lagged dependent variable*, *spatially lagged exogenous variable*, *spatially lagged error* (or *spatial error*). The spatially lagged dependent variable is sometimes denoted in short as *spatial lag* or *local average*. Various approaches to estimate models with cross-sectional dependencies are described in Sections 3.3 and 3.4; the first of the two sections focuses on the spatial econometric mainstream using either Method-of-Moments estimators or Maximum-Likelihood estimators relying on the multivariate Gaussian distribution. Section 3.4 focuses on nonparametric approaches like eigenvector-based methods or Newey-West-type covariance estimators.

Cross-sectional heterogeneity in parameters (or in partial effects/impacts) is scarcely surveyed; a rare exception is is Hastie/Tibshirani (1993). I try to establish a typology with a particular focus on spatially structured heterogeneity in Table 3.2 below. However, I leave heterogeneity in the variance parameter aside. Heteroscedasticity arises frequently in regional empirical models. Though, it receives only little attention in spatial econometric modeling[6] since the adequate treatment of heteroscedasticity is standard, discussed in introductory econometric textbook. Reliable standard-errors of the parameters can be derived from the Sandwich form (see Eicker, 1967; Huber, 1967; White, 1980) if it is impossible, superfluous or not desired to impose further structure. If the variance heterogeneity is assumed to follow a parametric form estimable (and predictable) with an equation like $\sigma_i^2 = \sigma^2 \cdot f(z_i; \alpha)$, it is even possible to employ a Feasible Generalized Least Squares (FGLS) estimator.

Thus, a discussion of heterogeneity in the slope parameters – which describe the relation between some explanatory variables x and the dependent variable y – seems more relevant. A particular focus is on geographically varying slopes $\beta_i = \beta_{(o_i, a_i)}$ where o_i is observation i's longitude and a_i her latitude (as general representation of geographical locations).

6 To my knowledge, there exists only a small number of studies on heteroscedasticity in a spatial econometrics context most of which combining it with spatial autoregressive processes. Bera/Simlai (2005) establish a spatially autoregressive conditional heteroscedastic (SpARCH) estimator, Kelejian/Prucha (2007a) a heteroscedasticity-robust version of their moments estimator as described in Section 3.3.2. The Sandwich estimator of Conley (1999), see Section 3.4.2 aims at cross-sectional dependence but deals with heteroscedasticity as well.

Table 3.2: Typology: Heterogeneity in relation parameters (slopes, β)

Description/Structure	Hastie/Tibshirani (1993)	Examples from other fields, Spatial counterpart
Stochastic h.: $\beta_i \sim i.i.d.\left(\beta, \sigma_\beta^2\right)$		Random coefficients
Discrete h.: $\beta_i = \beta + I_{\{i \in A\}} \Delta\beta$	(d) [discrete w.r.t. X], (e) [possible]	Structural breaks, Spatial regimes
Functional h.: $\beta_i = Z_i \gamma$	(c), eventually (e)	Hierarchical (multilevel) regression, Smooth transition, Spatial expansion, Spatial smooth transition
Data-related h.: $\beta_i = \beta_{(X_i = x)}$	(f), (g), (h), eventually (e)	Local-polynomial regression, Flexible functional (polynomial) forms, Geographically weighted regression
Models (a) and (b) in Hastie/Tibshirani (1993), Sect. 2, describe homogenoeus coefficients.		

In fact, most approaches to generate cross-sectional variation in the slopes are a-spatial in the sense that they do not employ information on space. Discrete heterogeneity in regression parameters according to geographical location, frequently employed not only in regional studies but also beyond, is denoted as *spatial regimes* (e.g. Anselin, 1988, 1991). The *spatial expansion* method (see Casetti, 1972) models parameters to depend on functions of third-party variables which are not directly included in the economic model. These variables are considered to describe geographical, geological or environmental characteristics. A third family of heterogeneous parameter methods is constituted by non- or semiparametric methods which allow the slopes to vary over the distribution of the covariates (like local-polynomial regressions with Kernel-based weights, see e.g. Pagan/Ullah, 1999: Ch. 3, and Robinson, 2011). Truly spatial approaches arise by simply incorporating spatial information in the aforementioned approaches. E.g. Geographically Weighted Regression (GWR, see Brunsdon/Fotheringham/Charlton, 1998) uses the rows of a spatial weights matrix $W_{(n)}$ in a locally weighted regression, rather than some Kernel density estimates or Spline smoothers. Spatial regimes, spatial expansion and some estimators which indeed employ information on space (as incorporated in $W_{(n)}$) are discussed in Section 3.5.

Stochastic heterogeneity may arise separately or in combination with forms of structural heterogeneity. If the stochastic part can be modeled as a mean-zero variation around a structural parameter component satisfying $E(\eta_i \mid x_i) = 0$, it can be attached also to the disturbance of the regression equation:

$$y_i = x_i\left(\beta + \eta_i\right) + \varepsilon_i = x_i\beta + \underbrace{\left(x_i\eta_i\right) + \varepsilon_i}_{=u_i} \tag{3.7}$$

Here, the disturbance $u_i \sim \left(0, x_i^2\sigma_\eta^2 + \sigma_\varepsilon^2\right)$; hence, stochastic parameter heterogeneity results in structural heteroscedasticity. This result is well known for the random coefficient panel estimator (Swamy, 1970) but can be generalized. In our case, it seems be particularly important because many other forms of parameter heterogeneity can be modeled such that there is a stochastic (unobserved) remainder.

Nevertheless, before we continue, some clarification with regard to the terminology seems necessary: (Inter-)dependence refers to structures which are present in *data*, that is, which arise among observations drawn from a multivariate distribution. Heterogeneity in contrast refers to *parameters* describing a data generating process (DGP) or a statistical model. The distinction is, however, diffuse since data at the aggregate level – for example the average regional wage level – is equal to parameters describing distributions (or DGPs) at a finer observational unit – such as the mean of a regional wage distribution.

3.3 Spatial autocorrelation: Mainstream models

3.3.1 Spatial lag model

When considering models with spatial autocorrelation, I start with the so-called *spatial lag model*, a Cliff-Ord type model in the dependent variable:

$$y_i = \rho\sum_{j=1}^{n} w_{ij} y_j + x_i\beta + u_i \tag{3.8}$$

with $u_i \sim i.i.d.\left(0, \sigma_U^2\right)$, which in matrix notation becomes

$$Y_{(n)} = \rho W_{(n)} Y_{(n)} + X_{(n)}\beta + U_{(n)} \tag{3.9}$$

or, rewritten in reduced form,

$$Y_{(n)} = (I_{(n)} - \rho W_{(n)})^{-1} X_{(n)}\beta + (I_{(n)} - \rho W_{(n)})^{-1} U_{(n)} \tag{3.10}$$

From the reduced form it becomes obvious that the mean of Y[7] conditional on W and X is $\mu_{Y|X,W} = (I - \rho W)^{-1} X\beta$. The conditional variance of Y is $\Sigma_{Y|X,W} = V\left(\left(I - \rho W\right)^{-1} U\right) = \sigma_U^2\left(I - \rho W\right)^{-1}\left(\mathrm{I} - \rho W\right)^{-1\prime}$.

7 Subsequently, I omit the subscript (n) for notational simplicity.

The OLS estimator of the parameter ρ deduced from eq. (3.9) is, with projector $M_X = (I - X\,(X'X)^{-1}X')$:

$$\begin{aligned}\hat{\rho} &= \left(Y'W'M_X WY\right)^{-1}\left(Y'W'M_X\left[\rho WY + X\beta + U\right]\right)\\ &= \rho + (Y'W'M_X WY)^{-1}\,(Y'W'M_X U)\end{aligned} \tag{3.11}$$

By substiting $E[(Y'W'M_X WY)^{-1}\,|\,WX] = \psi \neq 0$ and inserting eq. (3.10) in the expectation of eq. (3.11), it is easy to show that

$$E\left[(Y'W'M_X WY)^{-1}Y'W'M_X U\right] = \psi\, E(U'(I-\rho W)^{-1'}W'U) \neq 0 \tag{3.12}$$

because any region i is a second-order neighbour to itself. That is, the OLS estimator for ρ will be biased and inconsistent, and the OLS estimator for β as well (for the latter result, see Appendix 3.A. The endogeneity (or simultaneity, respectively) of the spatially lagged dependent variable can be accounted for either by Maximum Likelihood (ML) estimation or by an instrumental variable (IV) approach. The further requires the additional assumption that Y is distributed n-dimensional multivariate gaussian, the latter a valid instrumentation strategy for WY.

The ML estimator for the single-period spatial lag model follows from maximization of the log-likelihood

$$\begin{aligned}\ln\mathcal{L} &= \ln\left((2\pi)^{-\frac{n}{2}}\,|\Sigma|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(Y-\mu)'\Sigma^{-1}(Y-\mu)\right\}\right)\\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_U^2 + \ln\left|(I-\rho W)\right|\\ &\quad - \frac{1}{2\sigma_U^2}\sum_{i=1}^{n}(y_i - \rho\sum_{j=1}^{n} w_{ij}y_j - x_i\beta)^2\end{aligned} \tag{3.13}$$

which, for exogenous X, will be unbiased if Y has a joint gaussian distribution (Anselin, 1988, 2001). The asymptotic behaviour for the (Quasi-)ML estimator has been investigated by Lee (2004) who established normality of estimates for a variety of spatially autocorrelated models under conditions analogue to Ass. 1a–c and 2.

With regard to the IV With regard to the IV estimator, Kelejian/Robinson (1993) have been the first to suggest utilization of internal instruments (short time before, Anselin, 1988 considered the search for instruments difficult). The bias of $\hat{\rho}$ results from $(I-\rho W)^{-1}Y$, or precisely from $\sum_{q=0}^{\infty}\rho^q W^q U$, in eq. (3.12). The linear projection of the polynomial $\sum_{q=0}^{\infty}\rho^q W^q X\beta$ or of any component of this polynomial on $(I-\rho W)^{-1}Y$ will be exogenous however if the regressors X are exogenous themselves. Thus, spatial lags of the exogenous variables W^qX (at any exponent $q>0$) may serve as excluded instruments (see as well Kelejian/Prucha, 1998; Anselin, 2001); typically, the first and second power are sufficient.

Remark 2. A general feature of the *spatial lag model* is that an estimated scalar parameter β_k can not be interpreted as variable's x_k economic effect, as can be seen easily from eq. (3.10); β_k measures only the direct effect of a marginal change in the variable's observation $x_{i,k}$ on y_i. This effect spills over to other observations j and from them back to observation i; the total effect of a marginal change in $x_{i,k}$ on the entire vector Y (the impact from observation i) follows then from the ith column of the spatial multiplier

$$\frac{\partial Y}{\partial x_{i,k}} = \left[(I-\rho W)^{-1} \beta_k \right]_{\cdot,i} \tag{3.14}$$

the effects on observation i, $\frac{\partial y_i}{\partial x_k}$ follows likewise from the ith row of the multiplier. The diagonal of the spatial multiplier constitutes the vector of local impacts

$$\left\{ \frac{\partial y_i}{\partial x_{i,k}} \right\}_{(n)} = vecdiag\left((I-\rho W)^{-1} \right) \beta_k \tag{3.15}$$

Frequently applied impact measures are, according to the nomenclature of LeSage/Pace (2009), the *average local impact* (which aggregates direct and feedback effects)

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial y_i}{\partial x_{i,k}} = \frac{1}{n} tr\left[(I-\rho W)^{-1} \right] \beta_k \tag{3.16}$$

the *average total impact* (for row-standardized W)

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial y_i}{\partial x_{j,k}} = \frac{1}{n} \iota_n' \left(I-\rho W \right)^{-1} \iota_n \beta_k \tag{3.17}$$

and the *average indirect impact* as the difference between average total and local impact.

3.3.2 Spatial error model

The *spatial error model* is a linear model where the disturbance follows a Cliff-Ord type process:

$$y_i = x_i \beta + e_i \quad with \quad e_i = \lambda \sum_{j=1}^{n} w_{ij} e_j + u_i \tag{3.18}$$

where again $u_i \sim i.i.d.\left(0,\sigma_U^2\right)$. In matrix notation, the regression residuals are distributed with mean $E\left(e_{(n)}\right) = E\left(e_{(n)} \mid X_{(n)}\right) = \left(I_{(n)} - \lambda W_{(n)}\right)^{-1} E\left(U_{(n)}\right) = 0$ and variance $E\left(e_{(n)} e'_{(n)}\right) = \sigma_U^2 \left(I_{(n)} - \lambda W_{(n)}\right)^{-1} \left(I_{(n)} - \lambda W_{(n)}\right)^{-1'}$. For exogenous x_i, the OLS estimator $\hat{\beta}$ corresponding to eq. (3.18) will be unbiased and consistent. A Newey-West HAC-type sandwich estimator may provide robust parameter variance estimates (as discussed below), in contrast to the inconsistent form $\left(X'_{(n)} X_{(n)}\right)^{-1} \sigma_e^2$ or purely heteroscedasticity-consistent standard errors. However, deriving the asymptotic distributions of these HAC estimators is rather complex since the OLS residuals e_i are neither independently distributed nor necessarily regularly spaced. Properties of parametric estimators are more easily to establish.

Similar to the *spatial lag model*, two estimation strategies have a prominent role in the literature: Assuming a joint multivariate gaussian distribution $Y_{(n)} \sim \mathcal{N}\left(X_{(n)}\beta, \left[\sigma_U^2 \left(I_{(n)} - \lambda W_{(n)}\right)^{-1} \left(I_{(n)} - \lambda W_{(n)}\right)^{-1'}\right]\right)$, Anselin (1988) pursues a ML approach (once again suppressing subscript (n)):

$$\begin{aligned} \ln L = &-\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\sigma_U^2 + \ln\left|\left(I - \lambda W\right)\right| \\ &- \frac{1}{2\sigma_U^2}\sum_{i=1}^{n}\left(y_i - x_i\beta - \sum_{j=1}^{n} w_{ij} y_j \lambda + \sum_{j=1}^{n} w_{ij} x_j \beta\lambda\right)^2 \end{aligned} \tag{3.19}$$

The second (derived by Kelejian/Prucha, 1998, 1999) pursues uses a Feasible Generalized Least Squares approach where moment conditions on the residuals' variance and auto-covariance are employed to estimate the parameters σ_U^2 and λ. The moment conditions[8]

$$E\left(\frac{1}{n}U'U\right) = E\left(\frac{1}{n}e'\left[I - \lambda W\right]'\left[I - \lambda W\right]\mathrm{e}\right) = \sigma_U^2 \tag{3.20a}$$

$$E\left(\frac{1}{n}U'W'U\right) = E\left(\frac{1}{n}e'\left[I - \lambda W\right]'W'\left[I - \lambda W\right]e\right) = 0 \tag{3.20b}$$

$$E\left(\frac{1}{n}U'W'WU\right) = E\left(\frac{1}{n}e'\left[I - \lambda W\right]'W'W\left[I - \lambda W\right]e\right) = tr\left(W'W\right)\sigma_U^2 \tag{3.20c}$$

lead, when replacing the theoretical disturbance e by empirical residuals $\hat{e}$ estimated consistently (e.g. with OLS), to a system of three equations with unknown variables $\sigma_U^2, \lambda, \lambda^2$. The estimates $\sigma_U^2, \hat{\lambda}$ allow the construction of an estimate for Σ_e and a corresponding FGLS estimator for which asymptotic normality can be shown. Additional (Quasi-)ML procedures for panel data (for fixed and random effects, serially correlated disturbances etc.) have been established amongst

8 The first conditions states the variance of the (i.i.d. mean zero) innovations; the second states zero correlation of the innovation with a weighted average over all other innovations; and the third states the variance of these weighted averages.

others by Anselin (1988), Baltagi et al. (2007) and Lee/Yu (2010). Similar moment conditions have been derived for systems of equations by Kelejian/Prucha (2004), random-effect panels by Kapoor/Kelejian/Prucha (2007), fixed-effects panels by Mutl/Pfaffermayr (2011), and for dynamic panels by Mutl (2005) and Hujer/Rodrigues/Wolf (2008).

3.3.3 Spatial Durbin model and SARAR-model

So far, we have left spatially lagged exogenous variables aside since their effect is estimable unbiasedly and consistently with OLS if there is no other source of cross-sectional dependence. This looks quite different if the regression model includes both a spatially lagged dependent variable and spatially lagged exogenous variables, that is, if we have to estimate a *spatial Durbin model:*

$$y_i = \rho\sum_{j=1}^{n} w_{(1)ij}\, y_j + x_i\beta + \sum_{j=1}^{n} w_{(2)ij}\, x_j\gamma + u_i \tag{3.21}$$

with $u_i \sim i.i.d.(0,\sigma_U^2)$ and γ a $K\times 1$ parameter vector. This model is favored by some authors (e.g. LeSage/Pace, 2009) because of its generality. It incorporates not only the *spatial lag model* which holds under the restriction $\gamma=0$ and the spatially lagged regressors model under restriction $\rho=0$, but also under restrictions $W_1 = W_2$ and $\gamma = -\rho\cdot\beta$ the *spatial error model*, and under restrictions $W_1 = W_2$ and $\gamma = -(\lambda)\cdot\beta$ the *spatially autoregressive model with autoregressive disturbances* (SARAR) model (Kelejian/Prucha, 1998) in which spatial lag and spatial error are combined.

Thus, if the models are estimated with ML, they differ only with regard to the restrictions imposed and the weights matrices employed for the three spatially autocorrelated components. This generalization hints already at the crucial factor and the major complication in these models (see e.g. Gibbons/Overman, 2012). To make this explicit, we write a model incorporating the three spatial lag, spatially lagged regressors, and spatial error all constructed with the same weights matrix W; we state it in matrix notation and in terms of orthogonalized (i.i.d.) disturbances:

$$\begin{aligned} U &= (I-\lambda W)(I-\rho W)Y-(I-\lambda W)(X\beta+WX\gamma) \\ &= \left(I-(\lambda+\rho)W+\lambda\rho W^2\right)Y-X\beta-WX(\gamma-\lambda\beta)+W^2X\lambda\gamma \end{aligned} \tag{3.22}$$

In contrast to the spatial lag and spatial error model, ρ and λ are not identifyable from the concentrated likelihood (which is symmetric in both parameters). It is easy to see that in one-step estimation of eq. (3.22) only β is clearly identified.

Identification of γ, λ (and thus also of ρ) hinges on the terms $\gamma-\lambda\beta$ and $\lambda\gamma$ associated with the first and second order spatial lag of the explanatory variables. In general, it might be possible to solve the (nonlinear) system in the parameters λ and γ; however, it is necessary to care about values associated with local rather than global maxima of the likelihood function. Furthermore, collinearity may arise if W^2X is extremely smooth.

A solution may lie in the use of two-step or three-step procedures. Kelejian/Prucha (1998), adding a spatial error to a spatial lag model, show that both spatial autocorrelation parameters can be estimated consistently (the further by their moments procedure, the latter by IV); moreover, they establish the asymptotic normality for the spatial error parameter, the spatial lag parameter and the structural parameters β. Here, the complication is in the choice of adequate instruments. Note that some of the instruments suggested by Kelejian/Prucha (1998), the first-order spatially lagged regressors, are here in eq. (3.22) employed as regressors – and there is, ex-ante, no reason why higher order spatial lags should not be regressors, too. Standard tests for the validity of the instrumentation strategy may provide some guidance. For example, rejection of an overidentifying-restrictions hypothesis (indicating correlation of the disturbance with excluded instruments) may be due to omitted variables in the main equation. Controlling for some of these instruments in the main equation rather than only in the first stage might contribute to non-rejection of the tests. Another problem might occur if the W are too dense or if they consist of mutually (almost) exclusive blocks: Then, the W^qX used as instruments show only little variation for different q, with the consequence that they are weak instruments which might be insufficient for identification in finite samples.

The spatial econometric mainstream demonstrated so far is able to deal with the two features which, in our view, charactarize regional models: spatially structured cross-sectional dependence and, albeit in a little flexible way, spatial heterogeneity of relations between variables. Estimation of *spatial lag, spatial error* and *spatial Durbin models* will be consistent and efficient if they are correctly specified. However, identification builds to a large extent on presumptions most of which are often neither made explicit nor tested for. First, there is the choice of the spatial weights matrices and the normalization scheme. Another issue is the selection of the model: at least in cross-sectional data it is tested for no spatial autocorrelation vs. spatial lag and spatial error, and the two against each other, with Lagrange-multiplier tests proposed by Anselin (1988) and several succeeding studies (e.g. Florax/Folmer/Rey, 2003; Mur/Angulo, 2009). However, these tests do not allow for spatially lagged exogenous variables (or, in fact, test one restriction on them against another) which can be seen as a major disadvantage: as discussed before, only these spatially lagged exogenous variables enable identification in an unrestricted model. Finally, LeSage/Pace (2009: Ch. 2) provide a large array of motivation for spatially autocorrelated models (besides estimating a correctly specified DGP or a model with economically motivated spatial externalities, it may serve as an approximation to space-time dependence, account for omitted variables, for spatial heterogeneity, or for model uncertainty) – the interpretation of the estimated parameters is however distinct. Thus, one needs to think about what a spatial-autocorrelation parameter estimate identifies economically in the specific context (and how to rule out other motivations and interpretations of the relevant parameters), and not only if it is identified technically. If identification of a certain parameter is not crucial for the investigated task, less sensitive nonparametric approaches might be an alternative when accounting for spatial dependencies.

3.4 Spatial dependence beyond the mainstream

3.4.1 Spatial filter procedures

Spatial filtering encloses a number of procedures which aim to remove spatial (auto-)correlation from the disturbance in a regression. Getis/Griffith (2002) provide an overview across the two major techniques developed by Getis (1990) and Griffith (1996); note that, in addition to these, the FGLS estimator relying on Cochrane-Orcutt transformed data as suggested suggested by Kelejian/Prucha (1998) can as well be understood as a filtering procedure. Getis' approach aims at explicitly removing local similarities between spatially adjacent observations from the data. Using the Getis-Ord statistics[9], y and x are decomposed into a spatially correlated part and a non-spatial remainder. The non-spatial parts are then used in further regression to determine the regression parameters and their distribution.

Griffith's method likewise refers to a statistics for spatial correlation, Moran's Index.[10] Here, the auto-covariance generating matrix (with $I_{(n)}$ the $n \times n$ identity matrix, $\iota_{(n)}$ a $n \times 1$ vector of ones, and $W_{c(n)}$ an unstandardized (symmetric) contiguity-based first-order neighborhood matrix[11],

9 The Getis-Ord statistics $G_i = \frac{\sum_{j=1}^{n} w_{ij} x_j}{-x_i + \sum_{j=1}^{n} x_j}$ is generally considered a metric for local autocorrelation. The statistics is however defined only for variables on a positive scale (in $\mathbb{R}^+$).

10 The index developed by Moran (1950) computes the ratio of the spatial auto-covariance, measured by a contiguity-based first-order neighbourhood matrix, of a random variable y over the variance of y.

11 A more general eigenvector filter can be achieved by replacing $W_{c(n)}$ with $\frac{1}{2}\left(W_{(n)} + W'_{(n)}\right)$ where $W_{(n)}$ represents any spatial relation matrix satisfying Ass. 2.

$$C_{(n)} = \left(I_{(n)} - \frac{1}{n}\iota_{(n)}\iota'_{(n)}\right) W_{c(n)} \left(I_{(n)} - \frac{1}{n}\iota_{(n)}\iota'_{(n)}\right) \tag{3.23}$$

is decomposed into its $n-1$ eigenvectors $E_{(W)}$ and eigenvalues. The first eigenvector of $W_{c(n)}$ is $\frac{1}{\sqrt{n}}\iota_{(n)}$ which is eliminated by filtering with $\left(I_{(n)} - \frac{1}{n}\iota_{(n)}\iota'_{(n)}\right)$, a matrix with $\frac{n-1}{n}$ on the diagonal and $\frac{1}{n}$ off the diagonal. The further eigenvectors of $W_{(n)}$ are identical to those of $C_{(n)}$. Note that the eigenvectors are an orthogonal standardized decomposition of the spatial weights matrix, i.e.:

$$W_{(n)} = \left(\frac{1}{\sqrt{n}}\iota_{(n)}, E_{(W)}\right) \Lambda \left(\frac{1}{\sqrt{n}}\iota_{(n)}, E_{(W)}\right)' \tag{3.24}$$

For columns r; s of the matrix of eigenvectors it holds that $E'_{(W)r} E_{(W)s} = 0 \; \forall r \neq s$ and $E'_{(W)r} E_{(W)r} = 1$. Note that the eigenvectors can be understood as orthogonal mapping patterns of the geographic space, similar to the loadings corresponding to principal components. A major difference to principal component analysis is however that all eigenvectors $E_{(W)}$ are associated with bounded eigenvalues (which are $O(1)$) whereas the first principal components frequently correspond to eigenvalues diverging at an order close to $O(n)$.

The variables in $(Y_{(n)}, X_{(n)})$ are first regressed on the eigenvectors (or a subset of candidate eigenvectors) in order to find those which explain the spatial patterns in the data best (in terms of information criteria or significance). The κ stepwise selected eigenvectors $E_{(W)SF}$ are then included as additional information in the regression of $Y_{(n)}$ on $X_{(n)}$ [again suppressing subscript (n)]:

$$Y = X\beta + E_{(W)SF}\delta + U \tag{3.25}$$

In order to understand the spatial filter's effect, let the spatial-filter projector matrix be $M_{(W)SF} = \left[I_{(n)} - E_{(W)SF}\left(E'_{(W)SF}E_{(W)SF}\right)^{-1}E'_{(W)SF}\right]$. The inner matrix cross-product of eigenvectors selected in the spatial filter becomes an identity matrix, $E'_{(W)SF}E_{(W)SF} = I_{\kappa}$. Furthermore, note that the set of eigenvectors of $W_{c(n)}$ which have been eliminated by the stepwise selection procedure – the complement to the spatial filter $E^{c}_{(W)SF}$ – are approximately orthogonal to Y and X, and should thus be approximately orthogonal to the disturbance U, too. Hence, a model with eventually spatially autocorrelated disturbance $e = (I - \rho W)^{-1}U$ [such as the one described by eq. (3.18)] which is augmented by a spatial filter would be transformed into a model with independent disturbances U since

$$M_{(W)SF}\left[I-\rho W\right]^{-1}U = M_{(W)SF}\sum_{q=0}^{\infty}\rho^{q}W^{q}U \\ = U+\sum_{q=1}^{\infty}\rho^{q}M_{(W)SF}W^{q}U \approx U \tag{3.26}$$

for any $|\rho| < 1$ and W satisfying Ass. 2. Thus, the spatial filter indeed eliminates spatial autocorrelation in the residual. It is easy to prove that OLS estimates for β from eq. (3.25) are unbiased under Ass. 1a and b. These assumptions and results, together with the independence of the disturbances in eq. (3.25), allow application of the Lindeberg-Levy CLT (Pötscher/Prucha, 2001) to prove asymptotic normality of the *spatial-filter* estimator. Hence, standard inference regarding the parameters β becomes valid. Additionally, under Lemma 7a (referring to Case 3) of Rao (1967), the *spatial-filter* estimator may have smaller parameter variance-covariance than the corresponding OLS estimator (without eigenvectors) of eq. (3.21). Furthermore, note that the same set of eigenvectors can represent any set of weights constructed from a transformation of $W_{(n)}$; this makes eigendecomposition spatial filtering more robust against misspecification of the weights matrix, compared to a conventional parametric *spatial error model.*

Remark 3. An interesting feature of eigenvector-based spatial filtering arises when the true model should describe simultaneity of the dependent variable (as the spatial lag model) and not only correlation of disturbances (see Tiefelsdorf/Griffith, 2007). Regression of Y on X as in the linear regression model should yield biased parameters β if the true model is a spatial lag model and the spatial lag is omitted. If a spatial filter is included in the regression model, estimation of eq. (3.25) will yield nevertheless asymptotically unbiased estimates $\hat{\beta}$ although the model will be misspecified:

$$\begin{aligned}\hat{\beta}_{SF} &= \left(X'M_{W(SF)}X\right)^{-1}X'M_{W(SF)}Y \\ &= \left(X'M_{W(SF)}X\right)^{-1}X'M_{W(SF)}\left[\left(I-\rho W\right)^{-1}X\beta+\left(I-\rho W\right)^{-1}U\right] \\ &= (X'M_{W(SF)}X)^{-1}X'M_{W(SF)}X\beta+(X'M_{W(SF)}X)^{-1}X'M_{W(SF)}U \\ &= \beta+\left(\frac{1}{n}X'M_{W(SF)}X\right)^{-1}\frac{1}{n}X'M_{W(SF)}U \xrightarrow{p} \beta\end{aligned}$$

However, this does not allow to compute the correct impact measures as discussed in Subsection 3.3.1. I do not receive any parameter estimate for the spatially related processes; knowledge of them is required for the calculation of $\frac{\partial y_i}{\partial x_{j,k}}$.

Remark 4. Without loss of generality, any variance-covariance matrix can be decomposed by eigenfunction statistics in $\Sigma_Y = V\Lambda V'$ with $\Lambda = diag_{k=1}^{n}\lambda_k$ the diagonal matrix of eigenvalues (or Principal Components) and V the matrix of eigenvectors (or loadings); the columns in V are orthogonal to each other. The selection of (in n) diverging eigenvalues (in square roots) are even denoted as (principal) factors f, the corresponding selection of eigenvectors as factor loadings, and $\Sigma_f = \Gamma ff'\Gamma'$ is the part of the variance covariance matrix due to common factors. The complementary eigenvectors and eigenvalues can be aggregated to a cross-sectionally approximately independent idiosyncratic component Σ_υ; $\Sigma_Y = \Sigma_f + \Sigma_\upsilon$. Principal factor decomposition of Σ_Y requires in general two dimensions (see e.g. Forni et al., 2000, 2004 or Bai/Ng, 2011): the eigenvalues converge for $n \to \infty$ and the eigenvectors for $T \to \infty$.

Note that the variance-covariance matrix of a spatially autocorrelated process (in a cross-section) is given by $\Sigma_Y = \left[I_{(n)} - \rho W_{(n)}\right]^{-1}\left[I_{(n)} - \rho W_{(n)}\right]^{-1'}\sigma_U^2$ with exogeneous X and exogenous $W_{(n)}$ satisfying Ass. 2. The candidate eigenvectors of a Griffith-type spatial filter are an orthogonal decomposition of $W_{(n)}$; they are exogenous themselves, i.e. they do not demand T to increase. Thus, the eigenfunction spatial filter can be understood as a small-T approximation of a less-than strongly dependent factor model (with at maximum one strong factor represented by the constant). All that needs to be estimated in the spatial filter is the coefficient vector δ as an analogue to the principal factors.

3.4.2 Newey-West type variance-covariance estimators

Surprisingly, robust nonparametric variance-covariance estimations plays hardly a role in the spatial econometrics literature whereas Heteroscedasticity-and-Autocorrelation Consistent (HAC) estimators as introduced by Newey/West (1987) are fairly common in both time-series analysis and microeconometrics; the subsequent discussion always refer to a consistent estimator distributed as

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, C^{-1}\tilde{V}C^{-1}\right)$$

with $C = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i x_i'$. The crucial part is to provide a feasible estimator for

$$\tilde{V} = plim_{n\to\infty} Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i u_i\right)$$

A standard HAC estimators (following Bester/Conley/Hansen, 2011) is given by:

$$\hat{V}_{HAC} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} K_n\left(d_{i,j}\right) x_i u_i x_j' u_j \tag{3.27}$$

where $d_{i,j}$ is a measure for the distance between observations i, j and $K_n(d_{i,j})$ are Kernel weights in the bi-dimensional plane (rather than along a line as in time series) which satisfy standard conditions (see Conley, 1999: cond. C13): $K(\cdot)$ is a bounded, continuous function on the rectangle $[-1, 1] \times [-1, 1]$ with absolutely summable Fourier coefficients; $K(0)=1$, $K_n(d_{i,j}) \to 1$ as $n \to \infty$, and $K_n(d_{i,j})=0$ if $d_{i,j} > d_n^*$ (a threshold distance). Conley (1999) suggests a Kernel analogue to the Bartlett window (with $d_{i,j}$ defined as a *taxicab metric*, see Davidson, 1994: Ch. 5): $K_{L,M}(l,m) = \left[\max\left(\left(1-\frac{|l|}{L}\right), 0 \right) \times \max\left(\left(1-\frac{|m|}{M}\right), 0 \right) \right]$ where l, m measure the distance between observations i and j along the lattice's cartesian coordinate axes, and L, M are bounded by the cubic root of the samples expansion along these axes. The resulting weights behave like a rook-type neighbourhood scheme up to neighbourhood order $L \times M$ (see Haining, 2003). Conley (1999) furthermore argues that it is possible to impose a regularly spaced grid over the observations (with cell size bounded by the infimum of the distances between observations) even if their true locations are irregularly spaced (or observed only imprecisely). This allows to deal with a certain amount of measurement error or with a small degree of endogeneity of the spatial weights. Statistical inference achieved with the HAC estimator as in eq. (3.27) is thus less dependent on the choice of the weights (e.g. road distance vs. traveling distance), compared to the ML estimator and GMM estimator provided in Section 3.3.2; though, it is still sensitive with regard to strongly different spaces (e.g. those represented by geographic vs. trade weights) which may not be accounted for by the artificially regularised lattice.

An alternative sandwich estimator relying on covariance clusters has been proposed by Bester/Conley/Hansen (2011):

$$\hat{V}_{Clus} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{I}\left(i, j \in \mathcal{G}_g\right) x_i u_i x_j' u_j = \frac{1}{n}\sum_{g=1}^{G_n} X'_{(g)} U_{(g)} U'_{(g)} X_{(g)}$$

where G_g denotes the index set of group/cluster g, $X_{(g)}$ and $U_{(g)}$ are the group-specific partitions of $X_{(n)}$ and $U_{(n)}$. The number of groups is fixed, and each group has to enclose approximately the same number of elements (a certain fraction Cn). Groups are assumed to be contiguous, i.e. the elements within a group cover an area with a single boundary; these areas are disjoint, such that the groups are mutually exclusive. If the sample size increases, the number of elements in the interior of the group (surrounded by elements belonging to the same group) increases; different clusters g, h become asymptotically independent.

This cluster procedure does not estimate the variance-covariance consistently; it converges in distribution but not in probability; Bester/Conley/Hansen (2011) demonstrate that the resulting inference is reliable despite the fact that the typical forms of dependence do not coincide with the employed grouped correlation structure. Its main advantage compared to the Conley (1999) HAC estimator is that the clustered structure is much easier to implement in econometric practice.

3.5 Spatial heterogeneity

3.5.1 Impact measures from spatial lag and Durbin model

As noted in Subsection 3.3.1, *spatial lag models* and likewise *spatial Durbin models* generate spatially varying im pact measures. They can be considered insofar as a method to account for spatial heterogeneity in the relation between x and y. However, there is a major discrepancy between them and the heterogenous-effect models discussed sofar: parameter heterogeneity refers in general to the effect from x_i on y_i. In contrast, most impact measures describe heterogeneous effects from x_i on y_i, only the local impacts (the main diagonal of the spatial multiplier matrix) measure an effect with origin at the same location/observation.

A big disadvantage of using spatial lags for modelling effect heterogeneity is that the resulting impact measures are fairly smooth across space by nature; they are not able to represent spatial heterogeneity at a fine spatial scale. Impact measures calculated from spatial Durbin models are in general less smooth than those from spatial lag models; nevertheless, even they are restricted to the limited degree of heterogeneity which the weights and the two to four relevant parameters ($\beta_k, \gamma_k, \lambda, \rho$) in eq. (3.22) allow.

3.5.2 Spatial expansion

Modelling of regression coefficients as functions of geographic information Z_i in order to let them vary across locations is denoted as *spatial expansion* (see e.g. Casetti, 1972; Anselin, 1991):

$$y_i = x_i\beta_{(p_i)} + u_i \quad with \quad \beta_{(p_i)} = p\left(Z_i;\gamma\right) \tag{3.28}$$

Linearly expanded models can be estimated by interaction terms between the variable of interest and the expansion variables, that is by inserting the parameter function in the regression:

$$y_i = \left(x_i \odot Z_i\right)\gamma + u_i$$

where $\odot$ denotes Hadamard multiplication of matrices. γ can be estimated by OLS (or, to account for potential heteroscedasticity, GLS). In my view, identification of the (average) effect of interest still requires a careful design, analogously to the identification of a covariate-dependent treatment effect (ATE_x) as discussed in Wooldridge (2002: Ch. 18, p. 613ff.). For example, if eq. (3.28) includes a constant, it might make sense to demean all elements in Z_i before computing the interaction effects.

Nonlinear expansion is rather complex. It might be possible to understand the observation-specific values of the parameter functions as latent variables which may be estimated in a State-Space model; Other nonlinear likelihood methods such as spatial smooth-transition estimation (see Pede/Florax/Holt, 2009) may be employed as well. Panel data allow, as an alternative, to first estimate individual slopes which are then regressed in a second step (the surface estimation) on the expansion variables, see e.g. Herwartz/Niebuhr (2011); inference in the surface estimation is however only valid if it accounts for previous estimation of the slopes.

3.5.3 Spatial regimes

Spatial regimes, that is the heterogeneity of regression parameters across groups g, have been discussed manyfold; most prominently by Anselin (1988, 1990). Regime-switching parameters are modeled by interaction terms between the explanatory variables under research and (binary) group identifiers, i.e. by expansion with dummies:

$$y_i = x_i \mathbf{I}_{\mathrm{i}} \left(i \in \mathcal{G}_g \right) \beta_g + u_i \tag{3.29}$$

The association of an observation with a group is often rather heuristical, e.g. an East-West, North-South or rural-urban divide; it is rarely tested for the adequacy of the imposed regimes (in contrast to structural break tests in time-series analysis). Estimation of spatial regimes is straightforward such that I would like to highlight only two aspects. First, it is not possible to identify an average national (or global) parameter vector and regime-specific parameters jointly; either the national slope or a reference group has to be omitted (or otherwise restricted). However, the national slope does not follow immediately from averaging regime-specific slopes. The latter have to be re-weighted by the ratio between the regime-specific and the national covariances of the covariates, $\left(\frac{1}{n} X'_{(n)} X_{(n)}\right)^{-1} \left(\frac{1}{|\mathcal{G}_g|} X'_{(g)} X_{(g)}\right)$ with $|\mathcal{G}_g|$ the number of elements in index set G_g. Second, for consistent estimation it is necessary that both $\left|\mathcal{G}_g\right|T$ and $\left(n-\left|\mathcal{G}_g\right|\right)T$ will increase if n or T increases, depending on the employed asymptotics; otherwise, an incidental parameter problem may arise (Neyman/Scott, 1948; Lancaster, 2000). As a consequence, the number of regimes is typically held constant. A specific regime per region is only feasible if $T \rightarrow \infty$ for n fixed.

3.5.4 Geographically weighted regression

On the one hand, the variables and the functional form employed in spatial expansion, or the selection of groups in a spatial-regimes approach might be insufficient to describe the true heterogeneity in space; on the other hand, (nonlinear) functional forms might be difficult to fit. *Geographically weighted regression* (GWR, see Fotheringham/Charlton/Brunsdon, 1997; Brunsdon/Fotheringham/Charlton, 1998) follows the intuition that in such cases a nonparametric assessment of the local relation between x_i and y_i can be helpful. The estimation is carried out as locally constant regression similar to a Nadaraya-Watson estimator (see Pagan/Ullah, 1999); as kernel, GWR employs the rows $\tilde{w}_i$ of a (modified) spatial weights matrix $\tilde{W}_{(n)} = \left(I_{(n)} + W_{(n)}\right)$ (or a bi-dimensional Bartlett kernel as described in Subsection 3.4.2) rather than a density estimate of the distribution of covariates $K_f\left(\frac{X_i - x}{h}\right)$ used in Nadaraya-Watson. In general, the kernel $\tilde{w}_i\left(d_{ij}\right)$ depending on the distance between observations i and j has to satisfy (a) $\tilde{w}_i\left(0\right) = 1$, (b) $\lim_{d_{ij}\rightarrow\infty}\left\{\tilde{w}_i\right\} = 0$ and (c) $w_i\left(d_{ij}\right)$ is a monotone decreasing function for positive real numbers. GWR allows even for weights $\tilde{\tilde{w}}_i\left(d_{ij}, h\right)$ adapting to the data insofar that an optimal (MSE-minimizing) bandwidth h can be determined. In this, cross-validation is necessary such that the previous condition (a) is replaced by $\tilde{\tilde{w}}_i\left(0\right) = 0$, see Brunsdon/Fotheringham/Charlton (1998); i.e. the cross-validation procedure draws its information directly from $W_{(n)}$. The estimated local parameter is given by

$$\beta_i = \left(X'_{(n)}\, diag\left(w_i\right) X_{(n)}\right)^{-1} \left(X'_{(n)}\, diag\left(w_i\right) Y_{(n)}\right) \tag{3.30}$$

Brunsdon/Fotheringham/Charlton (2000) claim that GWR 'parameters' are unbiased (without formally proving evidence); this seems somewhat bold in the light of the bias known for other kernel-based non-parametric estimators. In the spatial statistics literature, however, GWR is criticised in particular for consuming too many degrees-of-freedom and for being sensitive with regard to multi-collinearity.

3.5.5 Coefficient expansion with eigenvector spatial filters

Starting from this critique, Griffith (2008) argues that GWR parameters, reflecting any spatially dispersed relation on a certain map described by a spatial weights matrix $W_{(n)}$, can be replicated by a eigendecomposition spatial filter, that is by a linear combination of an orthogonal representation of mapping patterns in $W_{(n)}$:

$$\beta_i = \overline{\beta} + \left(E_{(W)SF_i} \otimes \iota'_{(k)}\right)\delta \tag{3.31}$$

where $k+1$ is the number of variables (including the constant), p is the number of relevant eigenvectors, $E_{(W)SF_i}$ is the ith row of the spatial filter matrix, and δ is the $(k+1)\,p \times 1$ vector of meta-parameters (δ_0 denotes the $(p \times 1)$-partition of δ associated with the constant). Inserting these in a regression equation (with $\odot$ as Hadamard operator for element-wise multiplication)

$$\begin{aligned} Y_{(n)} &= \overline{\beta}_0 + E_{(W)SF(n)}\delta_0 + X_{(n)}\overline{\beta}_{1..k} \\ &\quad + \left[\left(\iota'_{(p)} \otimes X_{(n)}\right) \odot \left(E_{(W)SF(n)} \otimes \iota'_{(k)}\right)\right]\delta_{1..k} + U_{(n)} \end{aligned} \tag{3.32}$$

shows that the Spatial-filter analogue to Geographically Weighted Regression (SF-GWR) can be even understood as a Casetti-type linear spatial expansion model (see Section 3.5.2 using the eigenvectors extracted from a spatial weights matrix as geographical information. SF-GWR has the advantage compared to conventional GWR to be computationally less demanding (once the eigenvectors have been extracted).

3.6 Comments and concluding remarks

In this chapter I have discussed several empirical approaches which allow for cross-sectional dependence and regional heterogeneity in parameters/effects and which are at the same time feasible for estimation in data with short time dimension. I have used the single-period (cross-sectional) version for demonstration; adaption of the methods for panel data is in general possible.

In the parametric mainstream of spatial econometrics exogenous (fixed) spatial weights are employed to restrict the enormous number of possible parameters. Restrictions are imposed on the parameters describing the simultaneity in the dependent variable [matrix Π in eq. (3.1)] in the *spatial lag model*, on the slopes which describe the relation between x_i and y_i or y_j, respectively, [in eq. (3.1) the elements of B_k] by *spatially lagged exogenous variables*, as well as on the correlation of residuals (Σ) in the *spatial error model*. However, these models have two major disadvantages: The parameters in the spatially-autoregressive models show up to be sensitive with regard to the spatial weights, and identification of these parameters depends on further restrictions (as it does in system estimation).

A variety of alternative (albeit less popular) methods is available for both the estimation of (or correction for) error correlation in Σ and the estimation of heterogeneous slopes on the main diagonal of B_k. Some employ other (non-spatial) restrictions (e.g. equality of parameters in *spatial regimes*, or third-variable dependence in the *spatial expansion* method). *Geographically weighted regression* (for slope estimation) and *spatial Bartlett kernels* (for Newey-West type standard errors) use the information on the distance between observations to construct weights for nonparametric smoothing. *Eigendecomposition spatial filtering* and *Spatial-F ilter-analogue Geographically Weighted Regression* employ eigenvectors extracted from a spatial weights matrix in order to correct or to model some common features. I.e., the latter approaches still utilize information provided by the spatial weights, albeit in a more flexible and robust fashion. The cost for this increase in robustness is in the consumption of degrees of freedom: eigenvector-based methods (not to speak of remaining degrees of freedom in kernel smoothing) require estimation of more parameters than spatially autoregressive models.

I would like to close this chapter with a short remark on combining models for spatially autoregressive and spatially heterogeneous processes, or trading one against the other. Anselin (1990) has demonstrated that in data generated by a spatial error model it is possible to erroneously identify spatial regimes in the intercept. Likewise, we have argued that spatially autocorrelated processes may be used to address effect heterogeneity. In fact, the difference between the two is in the non-zero off-diagonal elements if the autoregressive process describes the true model; if spatial heterogeneity arises in addition to spatial autocorrelation it is hardly be possible to distinguish between them with statistical methods.

3.A Consistency and bias of estimators for β

Suppose the spatial lag model (as in Section 3.3.1) as true model. The following list provides the limiting parameter estimates corresponding to the exogenous covariates achieved by various estimation methods.

Plain OLS (spatial lag omitted)

$$\begin{aligned}\hat{\beta}_{OLS,0} &= (X'X)^{-1}X'Y = (X'X)^{-1}X'\left[(I-\rho W)^{-1}X\beta + (I-\rho W)^{-1}U\right] \\ &\xrightarrow{p} \beta + (X'X)^{-1}X'(I-\rho W)^{-1}\rho WX\beta\end{aligned}$$

Spatial lag included, estimation with OLS, with biased OLS estimate $\hat{\rho} = \rho + \delta_\rho$

$$\begin{aligned}\hat{\beta}_{OLS,B} &= (X'X)^{-1} X'(I - (\rho + \delta_\rho)W)Y \\ &= \beta - \delta_\rho (X'X)^{-1} X'W(I - \rho W)^{-1} X\beta + (X'X)^{-1} X'U \\ &\xrightarrow{p} \beta - \delta_\rho (X'X)^{-1} X'W(I - \rho W)^{-1} X\beta\end{aligned}$$

Spatial lag included, estimation with ML with $\sqrt{n}$-consistent ML estimate $\hat{\rho}$

$$\begin{aligned}\hat{\beta}_{ML} &= (X'X)^{-1} X'(I - \hat{\rho}W)Y \\ &= (X'X)^{-1} X'(I - \hat{\rho}W)\left[(I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} U\right] \\ &= \beta + (X'X)^{-1} X'\left[\frac{1}{\sqrt{n}} \mathcal{N}\left(0, \sigma^2_{\rho,ML}\right)(I - \rho W)^{-1} X\beta + U\right] \xrightarrow{p} \beta\end{aligned}$$

Spatial lag included, estimation with 2SLS with instruments Z, $\sqrt{n}$-consistent IV estimate $\hat{\rho}$ and $H = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z')$

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'(I - \hat{\rho}W)Y) \\ &= \beta + H \frac{1}{\sqrt{n}} \mathcal{N}\left(0, \sigma^2_{\rho,IV}\right)(I - \rho W)^{-1} X\beta + HU \xrightarrow{p} \beta\end{aligned}$$

Spatial filter (Griffith's method), estimation with OLS

$$\begin{aligned}\hat{\beta}_{SF} &= (X'M_{W(SF)}X)^{-1} X'M_{W(SF)}Y \\ &= \beta + (X'M_{W(SF)}X)^{-1} X'M_{W(SF)}U \xrightarrow{p} \beta\end{aligned}$$

Chapter 4

Persistence of Regional Unemployment

Application of a Spatial Filtering Approach to Local Labour Markets in Germany

By Roberto Patuelli, Norbert Schanne, Daniel A. Griffith and Peter Nijkamp[12]

Condensed version published at

Journal of Regional Science, 2012, Vol. 52, pp. 300–323

12 We would like to thank Paolo Foschi and Simone Giannerini (University of Bologna, Italy), as well as Cristian Bianchi, Alberto Ferrante and Alex Villazon (University of Lugano, Switzerland), for their help in executing the computational problems encountered during the preparation of this paper. We thank, for useful comments and suggestions, the Editor, three anonymous referees, Daniel Werner, seminar participants at Free University (VU) Amsterdam, the University of Lugano (USI) and the University of Regensburg, and session participants to the 55th Annual North American Meetings of the Regional Science Association International (New York, NY), the 49th European Congress of the Regional Science Association International (Łódź), and the ZEW/IAB Workshop on Spatial Dimensions of the Labour Market (Mannheim), The SSES Annual Meeting 2010 (Fribourg), and the 16th International Conference on Panel Data (Amsterdam).

Abstract

The geographical distribution and persistence of regional/local unemployment rates in heterogeneous economies (such as Germany) have been, in recent years, the subject of various theoretical and empirical studies. Several researchers have shown an interest in analysing the dynamic adjustment processes of unemployment and the average degree of dependence of the current unemployment rates or gross domestic product from the ones observed in the past. In this paper, we present a new econometric approach to the study of regional unemployment persistence, in order to account for spatial heterogeneity and/or spatial autocorrelation in both the levels and the dynamics of unemployment. First, we propose an econometric procedure suggesting the use of spatial filtering techniques as a substitute for fixed effects in a panel estimation framework. The spatial filter computed here is a proxy for spatially distributed region-specific information (e.g., the endowment of natural resources, or the size of the 'home market') that is usually incorporated in the fixed effects parameters. The same argument applies for the spatial filter modelling of the heterogenous dynamics. The advantages of our proposed procedure are that the spatial filter, by incorporating region-specific information that generates spatial autocorrelation, frees up degrees of freedom, simultaneously corrects for time-stable spatial autocorrelation in the residuals, and provides insights about the spatial patterns in regional adjustment processes. We present several experiments in order to investigate the spatial pattern of the heterogeneous autoregressive parameters estimated for unemployment data for German NUTS-3 regions. We find widely heterogeneous but generally high persistence in regional unemployment rates.

Keywords: Unemployment persistence; Dynamic panel; Hysteresis; Spatial filtering; Fixed effects

Jel: C 31; E 24; E 27; R 11

4.1 Introduction

Regional labour market developments mirror the spatial socio-economic dynamics of the economy. Therefore, timely information on the functioning of these markets is of critical importance for regional policy. In particular, panel-type information on the social economic labour markets may be an important sign post for effective policy, as the spatial-temporal evolution of these markets is critical for understanding the emergency and persistence of spatial disparities among regions. Disparities in economic development and welfare within countries (at the regional level) are often bigger than between countries (Elhorst, 1995; Ertur/Le Gallo, 2003; Patuelli, 2007; Taylor/Bradley, 1997; see, for example, the cases of Germany and Italy), and they often show typical geographical/spatial structures. Consequently, spatial disparities have for decades been a source of policy concern and applied research (for a recent overview of this field, see Kochendörfer-Lucius/Pleskovic, 2009). Spatial disparities occur in both developed and developing countries; their genesis may date back far in history, while their removal may take generations.

For example, Germany faced, in the first semi-decade after reunification, an increase in unemployment, from 2.6 million people in 1991 to 4.3 million people in 1997 – or, including the hidden reserve, from 3 millions to 5.6 millions (Fuchs et al., 2010). Unemployment remained, with only slight movements, at the same level for roughly 10 years, until the rapid decline after the 2005 reforms. In the period from 2006 to 2010 unemployment dropped again to the level of the early 1990s, despite the credit crunch. Throughout the high-unemployment period from 1995 to 2005, the unemployment rate in East Germany was 9 to 11 percentage points higher than the unemployment rate in West Germany; however, as we show later in the paper, there were large disparities within West German unemployment rates as well. In particular, in the two most recent years, the East-West disparities in the unemployment rates have diminished.

Underperforming regions imply, for a (redistributive) state, the need to allocate a higher share of public spending to those regions, eventually creating distortions in the redistribution of tax revenues and increasing conflicts with local policy makers and the public. Additionally, high unemployment has historically been linked to a number of socioeconomic problems, such as single-parent households, underperformance of students in school, truancy rates, and more (Armstrong/Taylor, 2000). Persistently high unemployment rates have been shown to be correlated with high shares of long-term unemployment and outmigration (for example, recent data for Southern Italy show an increase in the outmigration – toward the North – of the top university graduates; SVIMEZ, 2009).

With regard to regional unemployment disparities, policy makers need, in order to correctly target their actions and policies, to understand two aspects of such disparities: (a) the determinants of 'equilibrium' unemployment and its variation; and, (b) the region-specific and the cross-regional dynamics of unemployment. The determinants of unemployment have been studied extensively in the regional economic literature (Aragon et al., 2003; Badinger/Url, 2002; Basile/De Benedictis, 2008; Elhorst, 2003; Moretti, 2011; Niebuhr, 2003; Nijkamp, 2009; Oud et al., 2012; Taylor/Bradley, 1997; Zenou, 2009). Some attention has been as well devoted to the internal dynamics of regional unemployment, and to each region's sensitivity to shocks, seasonal factors, and persistence of unemployment. The available literature is mostly focusing on a macroeconomic setting, such as in a 'non-accelerating inflation rate of unemployment (NAIRU)' or in a (conditional/unconditional) 'convergence towards a natural rate of unemployment' perspective (following the approach of Blanchard/Summers, 1986; see, for example, Bayer/Juessen, 2007; Decressin/Fatás, 1995; Graciadel Barrio/Gil-Alana, 2009; Song/Wu, 1997; Tyrowicz/Wójcik, 2010a,b, 2011). From a technical perspective, these studies generally test for unit roots in the unemployment series.[13] However, they suffer from the major drawbacks of treating regions as homogeneous and/or cross-sectionally independent: they consider neither spatial correlation of shocks nor spatially structured heterogeneity in the adjustment process.

Similarly, the correlation of unemployment rates in space – that is, between neighbouring regions – has been studied both in an exploratory/descriptive fashion (Cracolici/Cuffaro/Nijkamp, 2007; López-Bazo/del Barrio/Artis, 2002; Mayor/López, 2008; Molho, 1995; Patuelli et al., 2011), and with regard to the determinants of unemployment (Aldashev, 2012; Elhorst, 1995; Kosfeld/Dreger, 2006; Mitchell/Bill, 2004; Patacchini/Zenou, 2007), using spatial-econometric techniques. However, little effort has been made, aside from in a time series/forecasting context (Schanne/Wapler/Weyh, 2010), to decompose the spatial dynamics of unemployment, so that region-specific autoregressive processes (responses to shocks), or region-specific seasonal characteristics can be traced. However, besides the old and general story that regions are not isolated islands, some specific arguments – such as commuting and internal migration, the spatial diffusion of information on vacancies, the (limited) search radius of unemployed persons, which affect the duration (and persistence) of individual unemployment – exist for spatially structured regional interdependence in the development of

13 Stationarity implies that a series has a distribution with finite variance and that it converges towards its long-run expectation. Convergence between the regions arises only if the regional series have the same long-run expectation. In contrast, non-stationary regional series imply that shocks persist and that in the long-run the cross-regional distribution depends completely on accumulated (random) events.

aggregate unemployment. In other words, regions are expected to differ in their degree of persistence, and this heterogeneity is likely to show a spatial pattern.

Policy makers who understand the specific characteristics of a region and of interregional dependencies are able to tackle problems more effectively and to anticipate more accurately necessary responses to aggregate and local shocks. Likewise, a group of (contiguous) regions that share common characteristics has the opportunity to develop common strategies (for example, within a single macro-region, such as a German Bundesland). We stress the need to investigate (break down) the components of region-specific dynamics, from an autoregressive (or reaction-to-shocks) viewpoint, so as to identify spatial patterns of common characteristics. A similar view was recently expressed by Partridge/Rickman (2010) in their review and discussion of (desirable) developments in CGE modelling.

The empirical research in our study will address the development of regional labour markets over a longer period in Germany. This country offers a unique natural experiment for our purposes, as – in addition to the regular spatial dynamics of an advanced industrial economy – the post-reunification effects appear to play a prominent role in the initial distribution of unemployment and the subsequent evolution of spatial disparities in the country, generating a certain amount of regional dynamics. Nevertheless, since unit-root tests are sensitive to structural breaks, it is important to deal properly with the direct impact of reunification. This paper aims to develop a number of autoregressive models for analysing regional unemployment between 1996 and 2004, that is, the period after the direct effect of reunification has fully realized, and before the major labour market reforms, in the 439 German NUTS-3 regions (Kreise). These administrative regions can be considered an ideal unit of analysis, because they directly relate to local policy-making choices, for example in public welfare[14], in terms of attracting capital- or labour-intensive industries through the provision of a productive environment, infrastructure, enterprise zones, or by subsidizing desired economic activities.[15] We estimate autoregressive effects specific to both each administrative region and different urbanization and agglomeration degrees of regions. In addition to a standard fixed effects (FE)/individual slopes estimation, we propose an econometric procedure suggesting the use of spatial filtering (SF) techniques as a substitute for region-specific parameters in a panel estimation framework. The spatial filter is a proxy for spatially distributed region-specific information (e.g., the endowment of natural resources or the size of the

14 Until 2004, two parallel benefit systems for long-term unemployed coexisted. The ‘Arbeitslosenhilfe’ was administered by the local departments of the Federal Employment Agency, while the ‘Sozialhilfe’ was under the responsibility of the NUTS-3 authorities (Kreise).

15 Although the major part of subsidies is distributed by the federal states, the national government or the European Union, many programmes require co-funding from the local authorities, and availability depends on criteria often calculated at the NUTS-3 level.

'home market') that is usually incorporated in the FE or in region-specific slope parameters. The approach presented here is beneficial, because it allows considerable savings in terms of degrees of freedom. Most importantly, the spatial filter provides a straightforward interpretation – as the linear combination of orthogonal spatial patterns – of the FE components surrogate. By incorporating region-specific information that generates spatial autocorrelation and dynamics, our procedure provides new insights about the spatial patterns that make it interesting to adopt the approach also for the analysis of other spatiotemporal processes, such as GDP growth/convergence, house price diffusion, and spread of diseases.

In this paper, we present several experiments investigating the spatial patterns of autoregressive parameters estimated for the unemployment rates of German NUTS-3 regions. Our findings show that – on average – unemployment rates are rather persistent and that the levels of persistence have an identifiable spatial structure. The proposed methodological approach also shows to be a promising tool for the analysis of regional dynamics. Additionally, we propose a model based on spatial regimes, which allows to decompose the dynamic processes of regional unemployment rates according to agglomeration/urbanization criteria, rather than to the well-known – but oversimplifying – East-West Germany division. The remaining part of the paper is structured as follows. Section 4.2 describes the analytical design of the model used in our study. Sections 3 and 4 present the dataset used and the results obtained, respectively. Finally, Section 5 provides a rejoinder and conclusive remarks.

4.2 Analytical design of the model

4.2.1 The traditional approach

The current standard approach to analyse the persistence of unemployment or, in a multi-region context, its convergence speed (see, for a recent overview, Lee/Chang, 2008) is to estimate a system of AR(1) processes, and to test each single equation as well as the entire system of equations for unit roots. Here, the basic equation for unemployment u in region i is given by eq. (4.1):

$$u_{i,t} = \alpha_i u_{i,t-1} + \mu_i + s_{i,t} + \varepsilon_{i,t} \qquad (4.1)$$

where μ_i denotes the average unemployment[16], $s_{i,t}$ its seasonal component, and $\varepsilon_{i,t}$ an i.i.d. mean-zero random disturbance. Stacked over all regions, this set can be written as the following system of equations:

16 We assume that unemployment does not have a deterministic trend.

$$\mathbf{U}_t \iota_n = \mathbf{U}_{t-1}\mathbf{A}_n + \mathbf{M}_n + \mathbf{S}_t + \varepsilon_t \tag{4.2}$$

where $\mathbf{U}_t = diag_{i=1}^{N} u_{i,t}$ is the $n \times n$ diagonal matrix of unemployment rates at time t, $\mathbf{A}_n = (\alpha_1, \dots, \alpha_N)'$ and $\mathbf{M}_n = (\mu_1, \dots, \mu_N)'$ are $n \times 1$ column vectors of parameters, $\mathbf{S}_t = (s_{1,t}, \dots, s_{N,t})'$ is an $n \times 1$ column vector (generated from the $n \times 3$ matrix of parameters corresponding to the seasonal dummies, multiplied by the 3×1 matrix containing the seasonal dummies), $\iota_n = (1, \dots, 1)'$ is a unit vector of length n, and $\varepsilon_t = (\varepsilon_{1,t}, \dots, \varepsilon_{N,t})'$ is the $n \times 1$ vector of residuals. The subscript n in A_n and M_n denotes the length of the parameter vectors. Vectors and matrices with subscript t always have length n. M_n is equivalent to FE in a panel framework.[17]

If the autoregressive parameter α_i is smaller than 1 in absolute value, the impact of a "shock" $\varepsilon_{i,t}$ will vanish over time, and the series will converge to its long-run expectation. In contrast, if α_i equals one, the process in region i has a unit root. A single equation is tested for stationarity by augmented Dickey-Fuller (ADF) tests, or by Phillips-Perron (PP) tests; likewise, various tests derived for panels or systems that rely as well on subtracting lagged unemployment from both sides of eq. (4.2) require the following form of eq. (4.2):

$$(\mathbf{U}_t - \mathbf{U}_{t-1})\iota_n = \mathbf{U}_{t-1}(\mathbf{A}_n - \iota_n) + \mathbf{M}_n + \mathbf{S}_t + \varepsilon_t \tag{4.3}$$

Next, we may test if the elements of $(\mathbf{A}_n - \iota_n)$ are, individually or jointly, significantly less than zero.[18] Some procedures test the entire set of parameters directly (for example, Sarno/Taylor, 1998), whereas others combine the individual t-statistics to form a joint test statistic (see Maddala/Wu, 1999 or Im/Pesaran/Shin, 2003). As an alternative, restrictions may be imposed on the parameter, enabling a test only for stationarity of the average autoregressive process, as in Levin/Lin/Chu (2002), or for the stationarity of a limited number of regimespecific processes (also referred to as the 'convergence clubs' hypothesis).

Regarding the validity of panel unit-root tests, most of these procedures require the time dimension to be sufficiently large in order to converge and not to be plagued by the so-called Nickell bias arising in panels with a small time dimension (Nickel,1981) or by the Hurvicz bias in short times series. Moreover, eqs. (4.2) and (4.3) are only estimable in a seemingly unrelated regression (SURE) form (that is, in a specification that allows for simultaneously correlated errors)

17 For small time dimensions, the estimates of the autoregressive parameters are typically downward biased. With individual parameters, the Hurwicz bias is $\hat{\alpha}_i - \alpha_i = \frac{-(1+3\alpha_i)}{T}$. The Nickell bias, $\hat{\alpha} - \alpha = \frac{-(1+\alpha)}{T-1}$, for a common parameter across the regions $\alpha_1 = \ldots = \alpha_N = \alpha$ has a smaller size than the Hurwicz bias (Nickell, 1981). However, it can be seen that both converge towards zero when T goes to infinity.

18 The parameters $(\alpha_i - 1)$ follow, under the null hypothesis of a non-stationary process, a non-normal degenerate distribution, typically a Wiener process (also denoted as Brownian motion).

when the number of regions is small. Else one has to assume independence of the regions, resulting in equation-wise unit-root tests with low efficiency/power. Nonetheless, cross-sectional correlation seems rather plausible, in particular when considering small spatial units, and therefore taking this structure into account in the error term ε_t is preferable.

Cross-sectional (spatial) correlation arises not only in contemporaneous shocks, but also in levels and trends (see Table 4.1), in seasonal patterns, or in the adjustment speed. On the one hand, these spatial patterns or correlations could likewise be utilized to get better – more efficient, more powerful, less demanding in terms of degrees of freedom, and large-N, small-T consistent – estimates of the average convergence speed. On the other hand, knowledge about spatial interdependence between the structures of a time-series – average/trend, seasonality, autoregressive properties – may be of direct interest as well.

In the following subsection, we propose an alternative approach to estimating eq. (4.2), which decomposes the autoregressive processes according to exogenous spatial patterns that are representative of accessibility/contiguity relations between the regions studied. The benefit is twofold: (a) we obtain an explicit model of the spatial patterns in unemployment without being over-restrictive by imposing (probably erroneous) regime-specific constraints; and, (b) we are able to estimate more parsimoniously while covering the most relevant spatial structures.[19]

4.2.2 Spatial filtering

A wide array of methods, as well as several dedicated 'spatial' econometric procedures, for the statistical analysis of georeferenced data is available in the literature. Most commonly employed, spatial autoregressive techniques (see, for example, Anselin, 1988) model interregional dependence explicitly by means of spatial weights matrices that provide measures of the spatial linkages between values of georeferenced variables, with a structure similar to serial correlation in timeseries econometrics.

An alternative approach to spatial autoregression, modelling spatial autocorrelation in the mean response rather than in the variance, is the use of spatial filtering (SF) techniques (Getis/Griffith, 2002). Their advantage is that the studied variables (which are initially spatially correlated) are split into spatial and nonspatial components. Then these components can be employed in a linear

19 This claim clearly needs to be further explored by simulation evidence showing that SF is a suitable substitute/approximation of the fixed effects. Preliminary simulation results by the authors suggest that the SF and SFGWR are fully competitive – unless N or T tend to infinite – with mainstream econometrics methods such as bias-corrected LSDV (Bun/Carree, 2005) or Blundell/Bond (1998) in terms of parameter estimate bias.

regression framework. This conversion procedure requires the computation of a 'spatial filter'.

Table 4.1: Descriptive statistics of regional unemployment

Unemployment Rates (Levels), in percentage points						
Region	**Mean**	**St.Dev.**	**1st quartile**	**Median**	**3rd quartile**	**Moran's I**
Germany	11.8	5.5	7.6	10.1	15.4	0.903
East	19.4	3.5	17.0	19.3	21.8	
North	11.1	2.8	9.0	10.7	13.0	
South	8.1	2.5	6.2	7.7	9.5	
First Differences (in percentage points)						
Region	**Mean**	**St.Dev.**	**1st quartile**	**Median**	**3rd quartile**	**Moran's I**
Germany	0.01	1.21	-0.43	0.11	0.59	0.623
East	0.06	1.76	-0.88	0.30	1.22	
North	-0.01	0.89	-0.34	0.06	0.40	
South	-0.06	0.88	-0.72	-0.07	0.60	

The SF technique introduced by Griffith (2003) is based on the computational formula of Moran's I (MI) statistic.[20] This eigenvector decomposition technique extracts n orthogonal, as well as uncorrelated, numerical components from the $n \times n$ modified spatial weights matrix:

$$\mathbf{W}_n = \left(\mathbf{I}_n - \frac{1}{n}\iota_n\iota'_n\right)\mathbf{C}_n\left(\mathbf{I}_n - \frac{1}{n}\iota_n\iota'_n\right) \tag{4.4}$$

where I_n is an identity matrix of dimension n, ι_n is an $n \times 1$ unit vector, and C_n is a spatial weights matrix representing the spatial relation between each pair of regions; here we use a binary first-order contiguity (C-coding rook) matrix[21] where element c_{ij}

20 Moran's I is calculated as follows:

$$I = \frac{N\sum_{i=1}^{N}\sum_{j=1}^{N} w_{it}\,(x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\right)\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

where, in the case of a set of n regions, x_i is the value of the generic variable x in region i, and w_{ij} is the cell (i, j) of a spatial weights matrix W, indicating the proximity of each pair of regions i and j.

21 For a discussion of coding schemes and proximity definitions, see, with regard to the German NUTS-3 case, Patuelli et al. (2010), and more generally Griffith/Peres-Neto (2006). However, across most definitions for spatial weights matrices, the weights corresponding to element (i, j) are highly positively correlated. The results in spatial filtering hardly depend on the matrix from which the eigenvectors are extracted, thus the choice of the weights matrix is of little importance (see Getis/Griffith, 2002, Griffith, 2000). This is due to the fact that eigenvectors extracted from one (geographical) matrix can almost surely be generated by a linear combination of eigenvectors extracted from any other (geographical) matrix. For example, the matrix $\mathbf{I}_n - \rho\frac{1}{2}(\mathbf{W}_n + \mathbf{W}'_n)$ and its inverse $\left[\mathbf{I}_n - \rho\frac{1}{2}(\mathbf{W}_n + \mathbf{W}'_n)\right]^{-1}$ have the same eigenvectors, although the first may represent just a weighted average across the direct neighbours, whereas the latter represents an (infinite) distance-decay scheme.

equals 1 if regions i and j have a common border, and 0 otherwise. Matrix $\left(\mathbf{I}_n - \frac{1}{n}\iota_n\iota'_n\right)$ is the standard projection matrix found in the multivariate statistics and regression literature. Because matrix C_n is pre- and post-multiplied by the projection matrix [see eq. (4.4)], these eigenvectors are centred at zero. The eigenvectors extracted are in a decreasing order of spatial autocorrelation, and the first corresponds to the largest eigenvalue of W_n. Thus, the first two eigenvectors computed (E_1 and E_2) often identify map patterns along the cardinal points (that is, some rotated version of the major North-South and East-West patterns). Eigenvectors with intermediate values of MI display regional map patterns, whereas eigenvectors with smaller values of MI display local map patterns. The set of relevant eigenvectors – those explaining the spatial pattern in the variable of interest – can be found by regressing the dependent variable on the eigenvectors in a stepwise fashion, retaining the significant eigenvectors (or eliminating the insignificant ones). The linear combination of selected eigenvectors and their corresponding parameter estimates define the spatial filter for the variable of interest. In an autoregressive setting (where no covariates are employed), residuals obtained with stepwise regression constitute the spatially filtered component of the georeferenced variable examined (see Griffith, 2000). The eigenvectors can be seen as independent map patterns that coincide with the latent spatial autocorrelation of a given georeferenced variable, according to a given spatial weights matrix. Moreover, they can work as proxies for omitted variables that show a certain coincidence or similarity regarding their spatial distribution.

In this regard, Griffith's SF approach works differently from differently from Getis (1990, 1995), which decomposes each involved variable into a spatial and a nonspatial component, and requires the use of non-negative variables. Moreover, differently from mainstream spatial econometric models, such as spatial lag or spatial error models, which are developed mostly in a linear estimation framework, the SF approach can be applied to any functional form. Additionally, the tools necessary for implementing the technique – eigenvector decomposition and stepwise regression – are available in all statistical software packages.

Griffith (2008) shows that SF not only refers to the unobserved spatial correlation of a variable, but also contributes to the explanation of spatial heterogeneity in the parameters. An equivalent to the parameters of a geographically weighted regression (GWR, Brunsdon/Fotheringham/Charlton, 1998) can be computed by introducing interaction terms between the exogenous variables of an equation and the eigenvectors extracted from a spatial weights matrix into a model specification. The possibility to combine the SF approach with a panel estimation framework and with geographically heterogeneous regression parameters (SFGWR) constitutes an additional advantage over existing methods. The next section details the functioning of the SFGWR approach.

4.2.3 An adjustment-process spatial filter

The parameters α_i and μ_i in eqs. (4.2) and (4.3) can be expected to show spatial heterogeneity,[22] that is, a pattern in space that may be related to the structure of a spatial weights matrix, and for which they could be tested, for example, by computing these parameters' MI. These spatial patterns can be and preferably should be considered explicitly instead of in the parameter-intensive formulation of heterogeneity given in eqs. (4.2) and (4.3). We introduce spatial patterns by decomposing the terms A_n and/or M_n into a spatial and a non-spatial part, setting $\mathbf{A}_n = \omega\tilde{\mathbf{A}}_k + \eta_n$ and $\mathbf{M}_n = \omega\tilde{\mathbf{M}}_k + \nu_n$ where ω is an $n \times k$ matrix of eigenvectors E_k extracted from the normalized spatial weights matrix given in eq. (4.4) (Griffith, 2003). ω collects the constant (that is, ι_i) as well, because $\frac{1}{\sqrt{n}}\boldsymbol{\iota}_n$ is also an eigenvector of matrix W_n. η_i and ν_i contain only non-spatial patterns within the individual parameters – hence they have zero mean and are orthogonal to the spatial process – and can thus move to the residuals. As we can substitute both the level and the dynamic adjustment in a process by their spatial counterparts, three alternative specifications to eq. (4.2) yield:

$$\mathbf{U}_t\iota_n = \mathbf{U}_{t-1}\mathbf{A}_n + \omega\tilde{\mathbf{M}}_k + \nu_n + \mathbf{S}_t + \varepsilon_t \tag{4.5}$$

$$\mathbf{U}_t\iota_n = \left(\mathbf{U}_{t-1}\omega\right)\tilde{\mathbf{A}}_k + \mathbf{U}_{t-1}\eta_n + \mathbf{M}_n + \mathbf{S}_t + \varepsilon_t \tag{4.6}$$

$$\mathbf{U}_t\iota_\mathrm{n} = \left(\mathbf{U}_{t-1}\omega\right)\tilde{\mathbf{A}}_k + \mathbf{U}_{t-1}\eta_n + \omega\tilde{\mathbf{M}}_k + \nu_n + \mathbf{S}_t + \varepsilon_t \tag{4.7}$$

Eq. (4.5) is the SF equivalent to the FE panel estimation [see eq. (4.2)]. In contrast, eqs. (4.6) and (4.7) show similarities with the SF representation of GWR (Griffith, 2008). $\tilde{\alpha}_1$, the first element of the parameters vector $\tilde{\mathbf{A}}_k$, and the one linked to the constant, estimates the average adjustment speed. The further autoregressive parameters specify regional patterns in the adjustment speed: for example, the parameters for the interaction terms between lagged unemployment and eigenvectors E_1 and E_2 reflect regional deviations from the average adjustment speed along the cardinal coordinates, similarly to the patterns that the eigenvectors themselves represent for the levels. Similarly, the parameters for the subsequent eigenvector interactions reflect how the above deviations can be attributed to more composite spatial patterns: first global, then regional, and finally local.

---

22 By the term spatial heterogeneity we refer to spatial structure in the parameters (i.e., the effects of variables), and by the term spatial correlation to spatial structure in variables. However, these terms are insofar related, as on the one hand, spatial correlation (e.g., in a spatial lag or spatial Durbin model) results in spatially heterogeneous marginal impacts (e.g., see LeSage/Pace, 2009: Ch. 2.7), and on the other hand, regression parameters can be considered as moments of (multivariate) distributions (in our case, the parameters μ_i represent the region-specific in-sample expectations of the unemployment rate) which may themselves be used as variables.

The new residuals vector – for example, defined as $\zeta_t = \mathbf{U}_{t-1}\eta_n + \nu_n + \varepsilon_t$ in eq. (4.7) – may exhibit either a panel-specific mean-zero component (a random effect, when $\sigma_\nu^2 > 0$, or panel-specific serial correlation in the residuals (when $\sigma_\eta^2 > 0$). Nonetheless, the orthogonality between the spatial eigenvectors and the non-spatial time-constant component suffices to guarantee orthogonality between the regressors $\mathbf{U}_{t-1}\omega, \mathbf{U}_{t-1}$ and ζ_t; that is, consistency of the estimation of eqs. (4.5), (4.6) and (4.7). However, the overall variance of these equations is inflated by the variance of ν_n and/or $\mathbf{U}_{t-1}\eta_n$ with respect to eq. (4.2).

4.2.4 Spatial regimes

An alternative approach to studying spatial heterogeneity in parameters is the introduction of explicit spatial regimes that, for example, distinguish between urban and rural economies, or to have one regime for each federal state (covering all districts within a single state). Because discrete schemes – in contrast to continuous parameter heterogeneity – allow results to be interpreted as a structural break (Anselin, 1990), a common choice in applied work is to use just two regimes: typically, North versus South for Europe (Ertur/Le Gallo/Baumont, 2006), or East versus West for Germany. In this paper, we apply a classification of regions by the German Federal Institute for Research on Building, Urban Affairs and Regional Development (Bundesinstitut für Bau-, Stadt- und Raumforschung, BBSR), which identifies nine different degrees of urbanization and agglomeration.[23] The number of spatial regimes to use is rather heuristic, since the classification of districts is due to population density, and is not directly linked to labour market considerations. The intuition is that cities or agglomerations – which have a different industrial and firm structure, different information channels, and populations with different preferences than rural areas – adjust to shocks differently.

In our analysis, we differentiate the (serial) autoregressive parameters (and seasonal effects) according to $r=9$ discrete spatial regimes, and follow the previous estimation approaches for the region-specific levels (by FE or SF). Thus, let $\mathbf{D}_{class}$ denote the $n \times r$ matrix that assigns a certain urbanization/agglomeration class to each region. In order to avoid perfect multicollinearity, there is no average autoregressive effect included in the equation system. ξ_n is the part of spatial heterogeneity in the autoregressive process that is not covered by the regimes, and that is considered unobservable. Then, the two spatial-regimes specifications are given by:

23 The nine classes are: (1) central cities in regions with urban agglomerations; (2) highly-urbanized districts in regions with urban agglomerations; (3) urbanized districts in regions with urban agglomerations; (4) rural districts in regions with urban agglomerations; (5) central cities in regions with tendencies towards agglomeration; (6) highly-urbanized districts in regions with tendencies towards agglomeration; (7) rural districts in regions with tendencies towards agglomeration; (8) urbanized districts in regions with rural features; and (9) rural districts in regions with rural features.

$$\mathbf{U}_t \iota_n = \mathbf{U}_{t-1} \mathbf{D}_{class} \breve{\mathbf{A}}_r + \mathbf{U}_{t-1} \xi_n + \mathbf{M}_n + \mathbf{S}_t + \varepsilon_t \tag{4.8}$$

$$\mathbf{U}_t \iota_n = \mathbf{U}_{t-1} \mathbf{D}_{class} \breve{\mathbf{A}}_r + \mathbf{U}_{t-1} \xi_n + \omega \tilde{\mathbf{M}}_k + \nu_n + \mathbf{S}_t + \varepsilon_t \tag{4.9}$$

In summary, we present three different approaches to model spatially heterogeneous autoregressive processes: by individual, spatial-filtering, and spatial-regimes parameters. In addition, we can estimate a homogeneous parameter as well, as in a standard dynamic panel. The length of the parameter vector $\tilde{\mathbf{A}}_k$ in the SF autoregressive model is $1 < k \leq n$; that is, more parameters need to be estimated than in the homogeneous model (with $\alpha_i = \alpha$) and, typically, much less than in the heterogeneous model of eq. (4.2). Likewise, the number of spatial-regimes autoregressive parameters is $1 < r \leq n$. Thus, both the SF and the spatial-regimes autoregressive models are more parsimonious than the individual model.

Theoretically, all other model components are possible to modulate – deterministic mean and seasonal effects – according to the same four schemes. Instead of considering all 64 possible models, in this paper we analyse only specifications where the deterministic mean is represented by FE or the spatial filter, and with homogeneous versus individual (region-specific) autoregressive and seasonal effects.

4.3 Data

Germany has shown in the past two decades the emergence of interesting dynamics on its regional labour markets and is therefore, for our purposes, a good case study. Analyses in this paper employ data about German regional unemployment rates, at the NUTS-3 level of geographical aggregation (Kreise, denominated 'districts' henceforth). The data are available for all 439 districts, on a quarterly basis, for the years 1996 to 2004.[24]

Summary statistics for the data at hand are presented in Table 4.1. The table results confirm that high and low (regional) unemployment rates are not randomly distributed across Germany. A first examination of the data suggests an asymmetric distribution, which is skewed toward high unemployment rates (the difference between the median and the third quartile is almost one standard deviation). When inspected spatially, the data show marked spatial autocorrelation (Moran's I (MI) for the districts' average unemployment is 0.878), which is further confirmed

24 The recently formed East German district of Eisenach (ID 16056) belonged to the Wartburgkreis district (ID 16063) until the end of 1997. Thus, unemployment rates for Eisenach before 1998 are not available, and we set them equal to the ones of Wartburgkreis. Also, in the first quarter of 1996, labour force figures are not available for five East German regions. In order to compute unemployment rates, we set the labour force (the denominator of the rate) equal to the labour force reported in the subsequent four quarters (as it is determined only once per year by micro-census data).

by descriptive statistics calculated for macro-regional subsets, and by the map in Figure 4.1a. While the former East Germany shows persistently high unemployment rates (averaging 19.4 per cent) with (apparently) little variation (the first quartile is 17 per cent), the former West Germany shows low-to-moderate rates in the North (North Rhine-Westphalia, Lower Saxony, Schleswig-Holstein, and the city-states of Bremen and Hamburg) and in the South (Bavaria, Baden-Wurttemberg, Hesse, Rhineland-Palatinate, and the Saarland). When differencing the data, one can note that a certain amount of spatial autocorrelation remains (MI = 0.531), suggesting that not only the levels of unemployment, but also the dynamics, are spatially correlated. Again, this feature is evident in Figure 4.1b. This first finding implies that, when estimating a simple AR(1) panel model, one should expect spatial autocorrelation, as well as group-specific serial correlation, in the residuals.

A further visualization of the data, following Peng (2008), allows a plot of all data (15,804 records) simultaneously, providing a bird's eye view over regional disparities and trends. Figure 4.2a shows the unemployment rates of all German districts, by using a common colour scheme, where the different shadings are based on quantiles of the pooled data, and darker shades indicate higher unemployment. The graph (and the accompanying box plots) clearly shows that East German districts (in the bottom rows of each graph) have significantly higher unemployment. Seasonal effects are visible in the background, as the winter quarters show consistently higher unemployment (regularly occurring darker columns). It is also possible to identify some lightly coloured rows among the West German districts (in the left panel roughly at the top quarter, shortly below the first half of the rows for West Germany and little above the thick line separating East and West German districts; these rows indicate heterogeneity in the time-series characteristics within West German local unemployment rates, suggesting the inappropriateness of a homogeneous estimation approach.

Assigning to each district its own colour scheme (based on each time series' quantiles), renders Figure 4.2b. Although most West German districts appear to have had their best performance (that is, lowest unemployment rates) between 2000 and 2002, this is not the case for the East German districts. Instead, they seem to have had lower unemployment in 1996.[25]

25 In this regard, it should be recalled that no NUTS-3-level unemployment data are available for East Germany before 1996.

4.4 Empirical application

4.4.1 Fixed Effects and Spatial Filter estimation

In the preceding discussion, we presented a class of dynamic panel models, ranging from standard FE estimation [eq. (4.2)] to an alternative approach based on surrogating the FE by means of a spatial filter [eq. (4.5)], to GWR-type spatial filter and spatial regimes models. This subsection presents and compares results obtained for the first (FE and SF) approaches mentioned for a class of models with homogeneous and/or heterogeneous estimates of AR(1) parameters and seasonal effects. In particular, in Table 4.2, we compare summary results such as measures of fit (R^2 and RMSE), (average) autoregressive parameters estimated by the two approaches, and spatial autocorrelation in regression residuals.

The top left panel of Table 4.2 compares the most basic model specifications in terms of autoregressive parameters, in which just one (homogeneous) AR(1) parameter is estimated, assuming $\alpha_1 = \alpha_2 = \ldots = \alpha_N$. The FE and SF approaches are then compared. We find that the computed AR(1) parameters differ between the two approaches. The FE estimation with common seasonal dummies yields a homogeneous AR(1) parameter of 0.766, and with region-specific seasonal dummies an AR(1) parameter of 0.901. The corresponding (not reported) bias-adjusted parameters – obtained applying a correction according to the formula for the Nickell bias (see Footnote 17) – would be approximately 0.815 (and 0.955 in case of heterogeneous seasonality). The SF estimations give slightly higher parameters of 0.945 and 0.957, respectively. In anticipation of our further results, the two (corrected) parameter estimates from the FE specifications with homogeneous AR terms are insofar interesting, that they define (approximately) the range in which all other estimates for the average AR parameter fall, that is, the interval running from 0.81 to 0.96. The difference between the parameters does not seem to be high at first glance. However, the degree of persistence – measured as the half-life of a shock given by $\frac{\ln \frac{1}{2}}{\ln \alpha_i}$ – varies from 3.25 quarters (corresponding to an AR parameter of 0.81) to approximately 17 quarters for an AR parameter of 0.96.

Figure 4.1: Quantile maps of average unemployment rates: in levels (a) and in one-year differences (b)

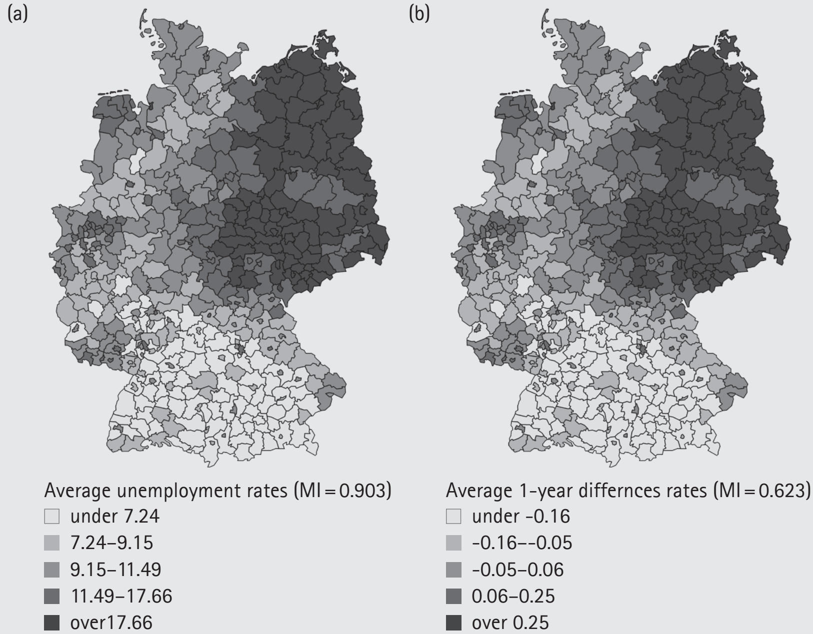
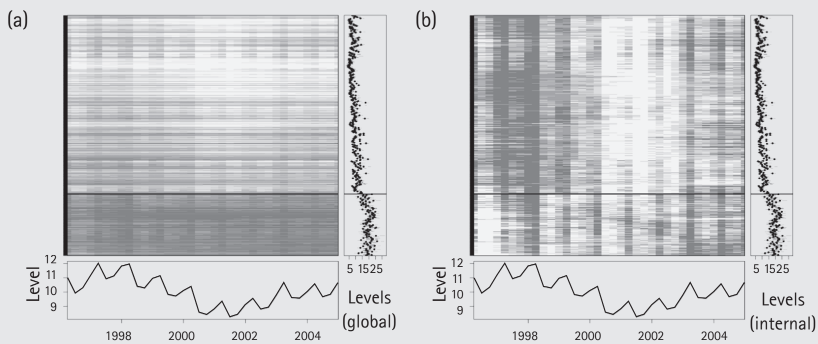


Figure 4.2: Visual representation of German regional unemployment rates



In the left graph, the colour scheme is common, in right graphs it is regionspecific. The thick line separates West German (above) and East German (below) districts. The right margin shows box plots for each district's time series. The bottom margin shows median features.

Table 4.2: Selected results for the homogeneous and heterogeneous AR process models[26]

Level	Homogeneous seasonality		Heterogeneous seasonal effects	
	FE	SF	FE	SF
Homogeneous AR(1) process: $\alpha_i = \alpha$				
AR(1) coeff.	0.766	0.945	0.901	0.957
Av. residuals MI	0.489	0.482	0.357	0.317
Min. residuals MI	0.195	0.204	0.142	0.038
Max residuals MI	0.775	0.734	0.754	0.767
R^2	0.977	0.975	0.992	0.991
RMSE	0.827	0.872	0.504	0.530
Res. Dfs	14,922	15,321	13,608	13,979
Heterogeneous AR(1) process: $\alpha_i = \mathbf{A}_{ni}$				
Av. AR(1) coeff.	0.833	0.823	0.906	0.914
Min. AR(1) coeff.	0.135 (3,462)	0.113 (9,271)	0.485 (14,181)	0.594 (14,188)
Max. AR(1) coeff.	1.120 (5,382)	1.275 (5,162)	1.035 (5,711)	1.137 (9,677)
No. of AR(1) $\geq$ 1	72/439	79/439	6/439	48/439
No. of AR(1) < 1 (ADF, 5% sign.)	156/439	284/439	97/439	264/439
Av. residuals MI	0.486	0.478	0.369	0.365
Min. residuals MI	0.169	0.094	0.143	0.128
Max residuals MI	0.787	0.804	0.782	0.805
R^2	0.981	0.980	0.992	0.992
RMSE	0.753	0.777	0.493	0.500
Res. Dfs	14,484	14,865	13,170	13,564

In terms of model fit, the SF estimate provides a fit to the data – in terms of R^2 – very similar to the one for the FE estimate (0.975 versus 0.977), while saving about 400 degrees of freedom. As stated in Section 4.2.3, the variance of the SF estimation is deemed to be (slightly) inflated with respect to the FE variance, which is also suggested by the computation of the RMSE (this is true for all estimations presented in Table 4.2). Meanwhile, in Figure 4.3 we can see how the SF computed (as the linear combination of the 39 eigenvectors selected) approximates the spatial patterns shown in the FE parameters. The spatial patterns shown in the two maps may be expected to include both region-specific variations from the average (homogeneous) AR(1) parameter and seasonal effects, as well as unobserved variables (such as, for example, other lags of the

26 The (upward biased) autoregressive parameter estimated with a pooled OLS and homogeneous seasonal dummies is 0.993 (with a regionally clustered standard error of 0.0014), the asymptotically consistent Blundell-Bond estimator with homogeneous seasonal dummies is 0.902 (with a standard error of 0.0028).

unemployment rate). Not surprisingly, the eigenvector contributing most to the SF is E_2, which shows a clear NE-SW pattern, although it should be kept in mind that the amount of variance explained by this top eigenvector, in this dynamic panel framework, is less than 0.7 per cent of the one explained, for example, by the seasonal dummies. Subsequent eigenvectors are at least three times less informative than E_2.

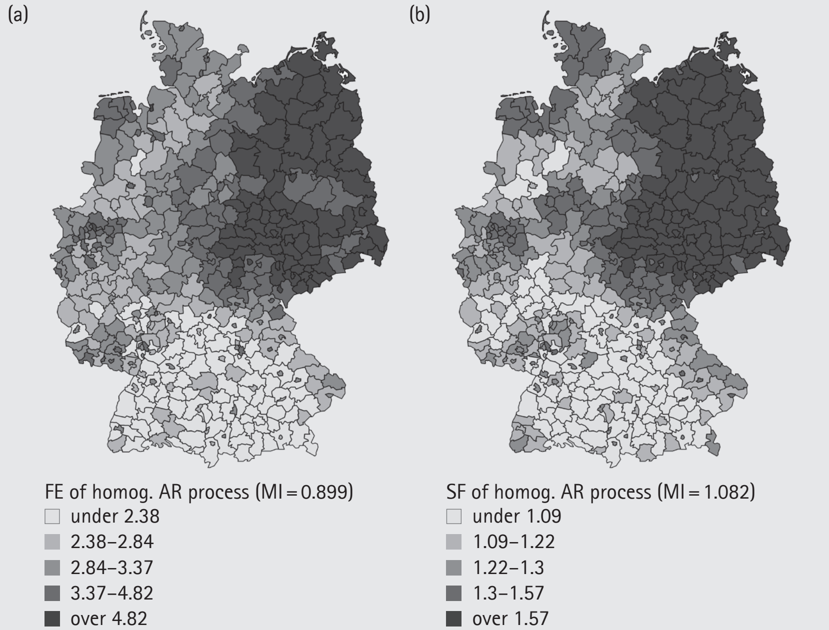
Finally, the levels of residual spatial autocorrelation appear to be similar for the FE and SF approaches, with a tendency for the SF approach to obtain residuals slightly less correlated in space. The time-averaged residual per region is zero or very close to zero, and spatial autocorrelation is absent. Consequently, quarter-specific spatial autocorrelation can be related directly to each quarter's specific shocks or unobserved characteristics (beyond direct seasonal effects, which are included in the model), and no recurring pattern exists over time.

Subsequently, the bottom left panel of Table 4.2 provides summary results for estimation of the models presented in eqs. (4.2) and (4.5), estimating heterogeneous AR(1) parameters according to the FE and SF approaches, respectively. In contrast with the homogeneous case, where the estimated AR(1) parameter differed markedly between the two models, the estimates obtained here are rather similar on average, although the number of estimated parameters greater than or equal to 1 is slightly different: 72 and 79 for the FE and SF approaches, respectively. However, tests on the Dickey-Fuller transformation of the system suggest that unit roots can be *excluded* (at the 95 per cent critical value of a student-t distribution) for 156 districts in the FE approach and for 284 districts in the SF approach.

Once again, eigenvector E_2 is the most informative one, but in this occasion also eigenvector E_1 emerges amongst the main ones. The quantity of variance explained by the top eigenvector (E_2) is now greater in relative terms, for example if compared to the one of the seasonal dummies (4 per cent rather than the previous 0.7 per cent).

A certain level of numerical differences may be expected between the two vectors of AR(1) parameters (given in Figure 4.4). Indeed, the number of eigenvectors selected is distinct between a direct extraction of the SF (the procedure followed in this paper) and an indirect procedure, where FE are computed first, and an SF is extracted from the FE parameters vector. In the former case, fewer eigenvectors are selected, most likely because of the error component ε_t [see eq. (4.2)] not being considered in the indirect procedure. In contrast, a number of eigenvectors are selected only in the direct procedure, suggesting a correlation between these eigenvectors and the covariates (for example, U_{t-2} is *not* assumed to be orthogonal to the eigenvectors). Consequently, possible differences exist between the AR(1) vectors of parameters for eqs. (4.2) and (4.5). The extent of these differences depends on each specific case, and their direction remains to be fully inspected with a simulation experiment. With regard to the present analysis, clear differences appear to be mostly in the extremes, as shown by the similar quantiles and geographical patterns appearing in Figure 4.4. Both maps indicate higher first-quarter autoregressive effects in the western urbanized areas going (South to North) from Munich to the Stuttgart and Mannheim areas, to the Ruhr and Rhine areas, to Bremen, patterns that generally resemble the spatial distribution of population density in Germany.

Figure 4.3: Quantile maps of the FE (a) and SF (b) computed for the homogeneous AR(1) process



Conceivably, once we let the autoregressive parameter vary over the cross-section of districts, the measures of fit of the models (R^2 and RMSE) improve, while 438 (that is, $n-1$) additional degrees of freedom are consumed. Again, the SF estimation allows us to save about 420 degrees of freedom, while approximating closely the spatial patterns included in the FE parameters (Figure 4.5). Finally, residual spatial autocorrelation is the same – on average – in both the homogeneous and heterogeneous AR(1) parameter estimates, with the SF exhibiting lower minima in this regard.

Figure 4.4: Quantile maps of estimated heterogeneous AR(1) parameters: FE (a) and SF (b) approaches [parameters α_i according to equations (4.2) and (4.5)]

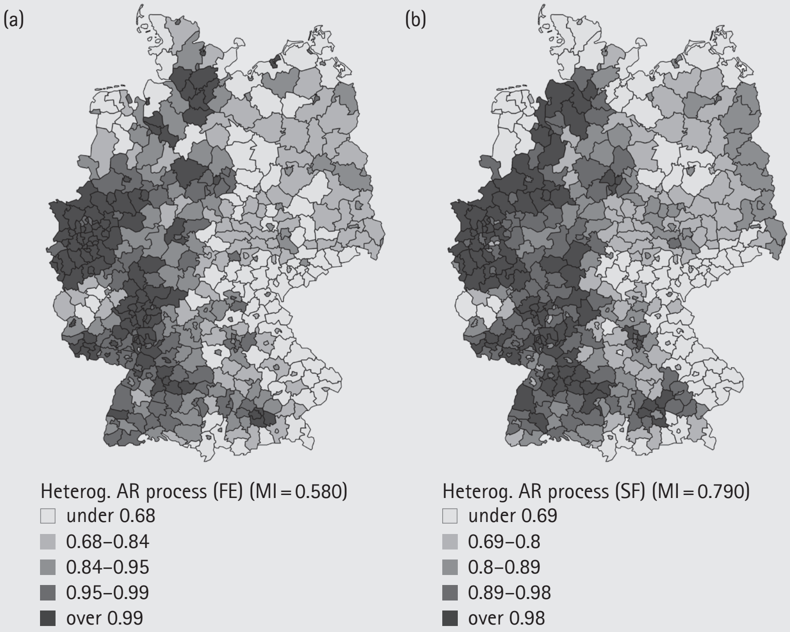
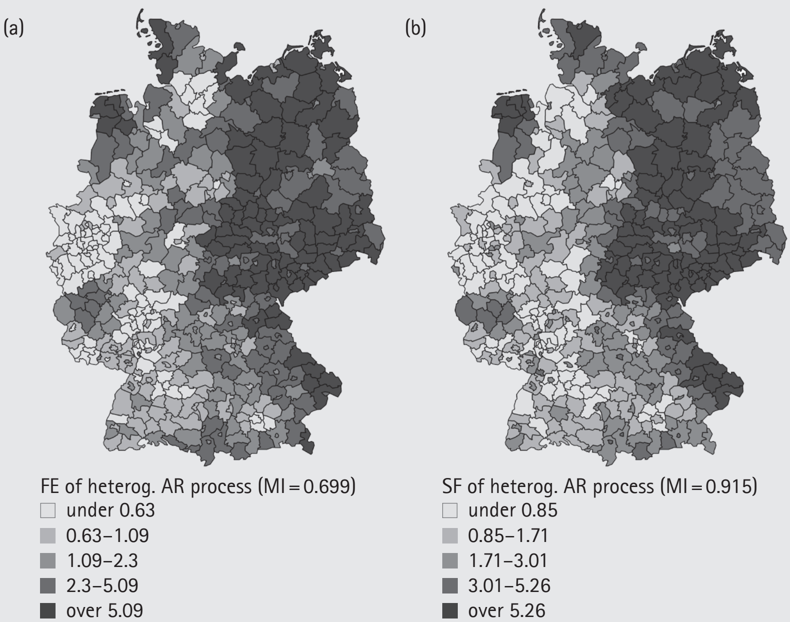


Figure 4.5: Quantile maps of the FE (a) and SF (b) computed for the heterogeneous AR(1) process [levels in equations (4.2) and (4.5)]



Finally, the right-hand panels of Table 4.2 provide additional empirical results, as the above models are extended to include individual (heterogeneous) seasonal effects. This extension implies computing (439 * 3 =) 1,317 regression parameters rather than the three previously computed seasonal parameters (for spring, summer and fall, while winter is used as the reference category). In the case in which both the autoregressive and seasonal effects are computed for each district, which we use as our example in the following discussion, (439 * 4 + 1 =) 1,757 parameters are computed, which increase to (439 * 5 =) 2,195 in the FE case.[27] As a result, an improved fit (higher R^2 and lower RMSE) as well as a diminished spatial autocorrelation in the residuals may be expected, which is confirmed by the summary statistics reported in Table 4.2. In addition, higher average AR(1) parameters are found, though with comparable results in terms of unit roots, as suggested by the ADF test results. Noteworthy are the changes in the spatial distribution of the AR(1) parameters and of the FE estimates, as shown in Figure 4.6. Figure 4.6a, referring to the AR(1) parameters, portrays patterns appearing in Figure 4.4 that are more sparse, as the result of individual seasonal effects having been filtered out. Meanwhile, Figure 4.6b, appears more similar to Figure 4.5, although it is slightly smoother.

The analyses presented above suggest that SF may be used to approximate the standard FE estimation for the study of unemployment persistence. Each of the two approaches appears to have specific advantages, allowing a researcher to choose freely between them on the basis of his/her needs. However, further approaches to decomposing region-specific autoregressive effects can be employed, as suggested in Sections 4.2.3 and 4.2.4. Results obtained for these additional classes of models are presented next.

4.4.2 Spatial Filter/Fixed Effects in the autoregressive component

The maps of the AR(1) parameters appearing in Figure 4.4 and the related MI scores highlight that autoregressive parameters are indeed strongly spatially correlated. As proposed in Section 4.2.3, the spatial patterns obtained according to eq. (4.5), by computing n autoregressive parameters, may be approximated by parameter expansion in a spatial-filter GWR-fashion. Eqs. (4.6) and (4.7) give the FE and SF specifications, respectively, implying that, for the latter, two spatial filters are computed (or, more generally, one for each SFGWR-type regressor,

27 Needless to say, the increase in computational load leads to a much slower stepwise selection of the SF, which on the other hand may be improved by the use of faster CPUs, by implementing stepwise solutions suitable for multi-core computers or clusters, or by resorting to different types of model selection procedures (see, for example, Miller, 2002).

plus the SF substituting the FE). In our specific case, substituting An by its SF representation implies saving 392 degrees of freedom (47 versus 439 AR-related regressors), while extending the SFGWR-type approach to seasonal effects allows us to save 1,602 degrees of freedom (154 versus 1,756 = 439 * 4), although at the (opportunity) cost of running extensive stepwise regression in order to select the relevant eigenvectors.[28] The relevance of such a huge saving in terms of degrees of freedom becomes evident when considering panels with large N and small T. In addition, the computational intensity of the spatial filter construction only applies to the first estimation of the model, while subsequent estimations – for example, for forecasting purposes – are faster than in the respective cases of eqs. (4.2) and (4.5), because the relevant eigenvectors already have been selected.

Table 4.3 reports summary statistics for the aforementioned model specifications. The mean, minimum and maximum AR(1) parameters reported for the SFGWR model (left panel) appear to provide a picture similar to the one found in Table 4.2 for the case of the heterogeneous AR(1) process, with the exception of a higher average parameter in the SF case. The inferential advantage with regard to unit-root testing becomes evident: while above the SF model with heterogeneous AR(1) process allows to reject – at a 5 per cent significance level – 264 to 284 unit roots and the FE model with heterogeneous seasonality and AR(1) process has a unit-root rejection rate of less than one quarter of the regions, the SFGWR model leads to a further increase of the rejection rate, reaching 337 unit-root rejections for the SFGWR model with heterogeneous seasonality and fixed effects (third column of Table 4.3).[29] Additionally, we can observe that the GWR models using FEs have roughly the same rejection frequency as the models using SF for the levels (274 vs 270, 337 vs 317) although the estimated average adjustment parameters are smaller in value – that is, the models using SF for the levels seem to be more efficient.

28 Given our starting set of 98 candidate eigenvectors, a backward stepwise regression identifying a SFGWR representation of both the AR(1) parameters and the seasonal effects evaluates, in the first step, (98 * 4 =) 392 models in the FE case, and (98 * 5 =) 490 models in the SF case.

29 For the GWR-type models, the vector of AR(1) parameters is obtained as the linear combination of the related eigenvectors, using as weights the regression parameters computed for the interactions terms between the lagged unemployment rates and the eigenvectors themselves $\alpha_i = \omega_i \tilde{\mathbf{A}}_k$. Seasonal parameters for each season, when included, are computed in a similar fashion. Because of this construction, unit-root tests are computed as t-tests, where the variance of each region's autoregressive parameter α_i is computed as $Var(\alpha_i) = \sum_k \omega_{ki}^2 \sigma_k^2$ and σ_k^2 is the kth diagonal element of the variance-covariance (sub)matrix of the K eigenvectors selected.

Table 4.3: Selected results for the spatial-filter-GWR (SFGWR) AR process models

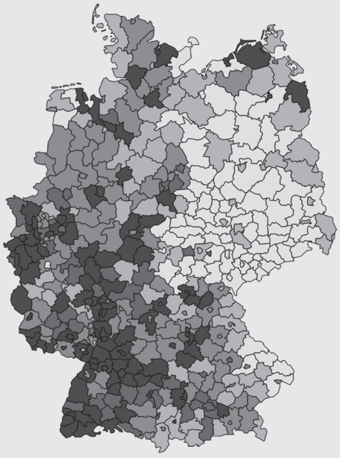
Level	Heterogeneous AR(1) process		Heterogeneous AR(1) process & seasonal effects	
	FE	SF	FE	SF
	Spatial filter AR(1) process: $\alpha_i = \sum_k \mathbf{E}_{ki}\tilde{\mathbf{A}}_k$			
Av. AR(1) coeff.	0.853	0.935	0.882	0.961
Min. AR(1) coeff.	0.162 (9,276)	0.276 (9,271)	0.530 (14,188)	0.697 (9,271)
Max. AR(1) coeff.	1.238 (7,338)	1.211 (5,374)	1.163 (9,274)	1.140 (5,374)
No. of AR(1) ≥ 1	94/439	136/439	44/439	94/439
No. of AR(1) < 1 (ADF, 5% sign.)	274/439	270/439	337/439	317/439
Av. residuals MI	0.481	0.440	0.333	0.176
Min. residuals MI	0.139	0.129	0.012	-0.016
Max residuals MI	0.817	0.730	0.803	0.704
R^2	0.980	0.978	0.985	0.986
RMSE	0.776	0.824	0.666	0.650
Res. Dfs	14,876	15,227	14,772	15,064
Selected eigenvecs for SFGWR-AR(1)	46	64	27	46

Once again, the levels of spatial autocorrelation in the residuals vary greatly, depending on quarter-specific noise, and are comparable but slightly lower than the earlier ones. RMSE increases moderately, as expected, but is being balanced out by the aforementioned huge savings in terms of degrees of freedom. These results are confirmed by extending the SFGWR specification to seasonal effects (right panel).

In terms of the spatial autocorrelation observed in the AR(1) parameters resulting from eqs. (4.6) and (4.7), Figure 4.7 confirms the similarities with the spatial distribution of population density. The spatial distribution of the estimated FE and SF (plotted in Figure 4.8) again is consistent pairwise, showing higher unexplained variation in the levels for East German districts. Not surprisingly, the light-shaded areas of Figure 4.7 appear to match the dark-shaded areas of Figure 4.8, as greater relative stability in the East German unemployment rates due to time-constant unobserved regional characteristics (or just lower dependence from their one-quarter lag) is reflected in the FE or in the SF. Similar observations can be made by comparing Figure 4.4 and 4.5, or the two maps in Figure 4.6.

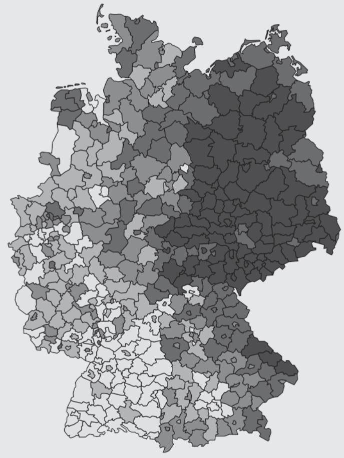
Figure 4.6: Quantile maps of the AR(1) (a) and FE (b) parameters computed for the heterogeneous AR(1) and seasonal process (FE estimation)

(a)



AR of heterog. AR & seas. process (FE)
(MI = 0.686)
- under 0.85
- 0.85–0.92
- 0.92–0.95
- 0.95–0.96
- over 0.96

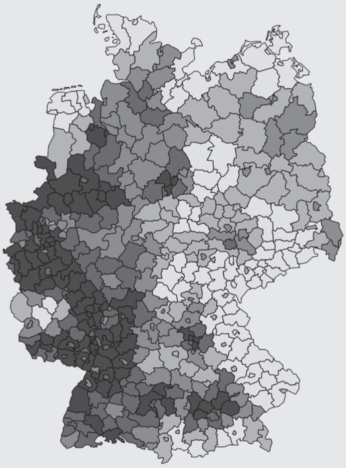
(b)



FE of heterog. AR & seas. process
(MI = 0.853)
- under 0.53
- 0.53–0.81
- 0.81–1.26
- 1.26–3.77
- over 3.77

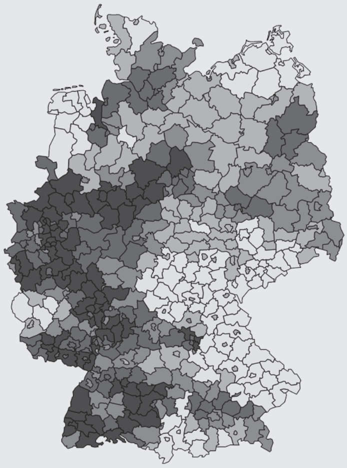
Figure 4.7: Quantile maps of estimated spatial-filter-GWR (SFGWR) AR(1) parameters: FE (a) and SF(b) approaches

(a)



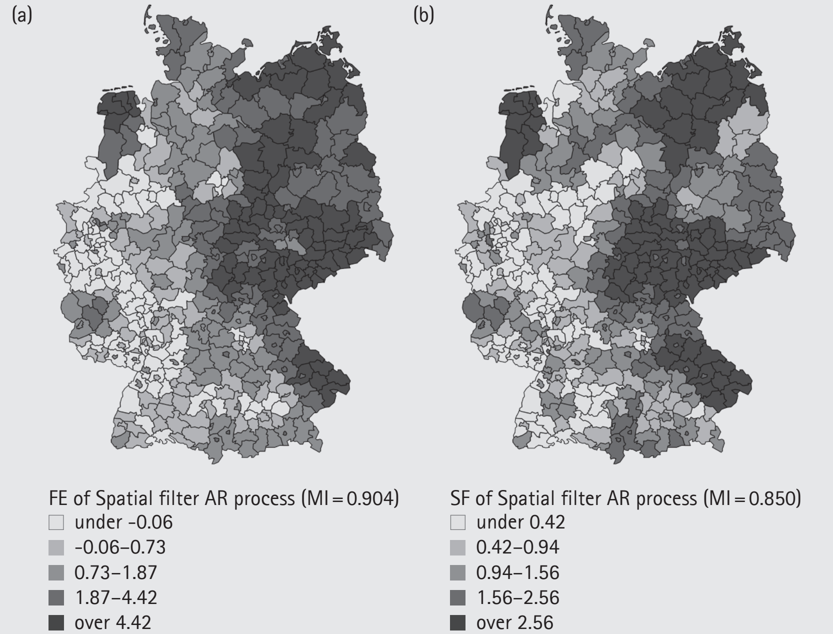
Spatial filter AR process (FE) (MI = 0.866)
- under 0.71
- 0.71–0.82
- 0.82–0.92
- 0.92–1
- over 1

(b)



Spatial filter AR process (SF) (MI = 0.842)
- under 0.87
- 0.87–0.93
- 0.93–0.98
- 0.98–1.03
- over 1.03

Figure 4.8: Quantile maps of the FE (a) and SF (b) computed for the spatial filter AR(1) process



As we already noted, the spatial-filter GWR surrogate for the region-specific autoregressive parameters allows identification of the spatial structure underlying the heterogeneity of the dynamic labour market process. Amongst the selected eigenvectors in the SFGWR specification with a spatial filter for the level component and homogeneous seasonal figures (Figure 4.7b and Figure 4.8b), there are four (of the five) eigenvectors associated with global patterns – that is, eigenvectors which, when the values are plotted into a map, show one or two large 'peaks' and one or two big 'valleys' spreading out over a large areas. 40 selected eigenvectors can be associated with regional, and 20 with local patterns.[30] Since all eigenvectors have the same scale (their values have an identical standard deviation), the partial contribution of each eigenvector to the overall autoregressive process is sized proportionately to the absolute value of the corresponding parameter. However, amongst the 15 eigenvectors with the highest parameter in absolute value, only two are global and two are local (the first local is at position 13), but 11 eigenvectors reflect regional patterns. In the other specifications, we find a similar selection of eigenvectors (the same four global, and roughly twice as many

30 The classification of global, regional and local eigenvectors is according to the table for 98 candidate eigenvectors extracted from a rook C-coding matrix given by Patuelli et al. (2012). Eigenvectors 1 to 5 are considered global, 6 to 66 regional and 67 to 98 local.

regional as local). However, in the corresponding SFGWR estimation using fixed effects (i.e., when the levels are forced to show maximum heterogeneity), all four global eigenvectors are amongst the 15 most influential eigenvectors.

More interestingly, there is a negative relation between the parameters associated with the (common) eigenvectors selected for modelling serial dependence and for the levels, as suggested by Figure 4.9. Additionally, eigenvectors which are selected only in one case (for which we include a value of zero in case of non-selection) have parameter values closer to zero even when significant, showing that the common eigenvectors are the ones with the greatest importance in both filters. On the other hand, the negative Pearson correlation of -0.89 (-0.93 for the common subset) between the two sets of parameters suggests that the SF in the levels behaves in the opposite way than the SF for the AR(1) parameters.[31] This indicates a trade-off between the level of persistence (i.e., serial dependence) and the influence of the (deterministic) level showing the spatial pattern modelled by the filter: unemployment is then represented as a weighted average of (more or less) persistent random elements (with a set of weight a) and deterministic elements [with weights $(1-a)$]. The more unemployment in a certain number of contingent regions (described by the mapping pattern of the eigenvectors) is driven by persistent shocks, the less important are the deterministic components in these regions – and vice versa, the lower the persistence, the faster regions adjust towards their initial (or natural) levels which become more important. This finding calls for further analytical investigation, which goes beyond this paper's objective.

Finally, the residual variance and the number of parameters of the models presented above can be combined to compute various information criteria (see Table 4.5, in the Annex). The Akaike information criterion (AIC) suggests that the SFGWR specification for the autoregressive process uses the information best, when compared to other model specifications, and that FE in the levels are superior to the SF. However, the AIC is often considered not adequate (or weak) for finite samples, and other criteria may be more reliable. The Schwartz Bayesian information criterion (BIC), which is often found to be over-selective, indicates superiority of the SF in the levels compared to the FE, and superiority of the SF AR process as well, because of the greater importance given to the degrees of freedom saved. The advantage of spatial filters in modelling both levels and autoregressive processes is confirmed by the Hannan-Quinn information criterion (HQ).

31 A similar finding is obtained when both the AR(1) and the seasonal parameters are computed by means of the SFGWR approximation. A Pearson correlation of -0.83 is obtained the two sets, and -0.91 is found for the common sets.

4.4.3 Adjustment to shocks according to the spatial regimes

In our final analysis, we present, in Table 4.4, summary statistics for the spatial regimes specification introduced in eqs. (4.8) and (4.9). In these specifications, heterogeneity of the autoregressive parameters is introduced by distinguishing between districts with different levels of agglomeration and urbanization. Consequently, instead of n AR(1) parameters, only nine are computed, corresponding to the specific classes introduced in Section 4.2.4. This approach makes identification of (average) autoregressive (and seasonal) effects possible for classes such as city-districts in agglomerated areas, or rural districts belonging to rural areas. The results obtained by applying the spatial regimes decomposition to the AR(1) process alone are shown in the left panel of Table 4.4.

Table 4.4: Selected results for the spatial-regimes AR(1) process models

Level	Heterogeneous AR(1) process		Heterogeneous AR(1) process & seasonal effects	
	FE	SF	FE	SF
Spatial-regimes AR(1) process: $\alpha_i = D_{class}(i \in r)\check{\mathbf{A}}_r$				
Av. AR(1) coeff.	0.808	0.937	0.812	0.946
Min. AR(1) coeff.	0.613 (type 9)	0.927 (type 9)	0.670 (type 3)	0.916 (type 2)
Max. AR(1) coeff.	0.984 (type 1)	0.949 (type 5)	0.934 (type 1)	0.960 (type 9)
No. of AR(1) ≥ 1	0/9	0/9	0/9	0/9
No. of AR(1) < 1	8/9	9/9	9/9	9/9
(ADF, 5% sign.)				
Av. residuals MI	0.485	0.476	0.425	0.417
Min. residuals MI	0.195	0.198	0.167	0.178
Max residuals MI	0.769	0.746	0.747	0.729
R^2	0.978	0.975	0.981	0.979
RMSE	0.810	0.869	0.754	0.798
Res. Dfs	14,914	15,306	14,890	15,291

We obtain nine AR(1) parameters ranging from 0.613 to 0.984 in the FE case, and from 0.927 to 0.949 in the SF case. These results are consistent with our previous findings (see Table 4.2). It turns out that the average AR parameters are higher for the SF approach, but when employing ADF tests only the FE case presents a unit root. This single unit root (which is not confirmed when decomposing seasonal effects as well) is found for districts of type 1 (that is, 'central cities in regions with urban agglomerations'). Our findings confirm the tendency of the AR(1) parameters to resemble the spatial distribution of population density, and of the central business districts (CBDs) of dense regions to show the highest parameters. Figure 4.10 maps the values found for the spatial regimes AR(1) parameters (SF estimation with homogeneous seasonal effects), and clearly shows that this approach provides a rough approximation of the parameter estimates obtained above, while showing – within a general picture of high persistence – some core-periphery patterns between the 'central cities' (type 1 and 5 districts, with higher persistence) and their surroundings; equality of all nine AR parameters is rejected both in the FE estimations and in the SF estimations. However, the regimes approach associates also a high degree of persistence to agglomerated areas in Eastern Germany (e.g., Dresden, Berlin or Chemnitz) which has not been found when using individual parameters (see Figure 4.4), that is, this rough approximation may indeed be missing some pattern. There are pros and cons to using spatial regimes, and this preliminary finding may deserve further investigation in the future research.

Figure 4.9: Correlation between the parameters of the eigenvectors selected for the SFGWR AR interpolation and for the levels (with homogeneous seasonal effects)

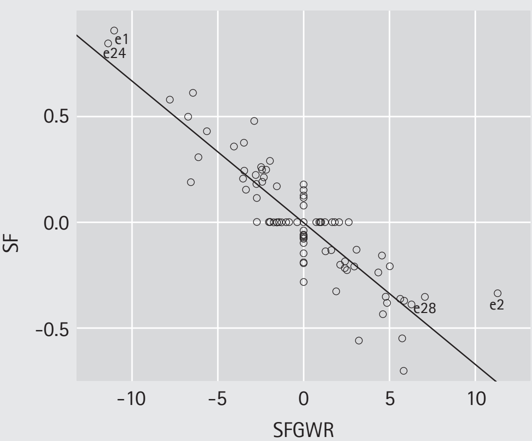
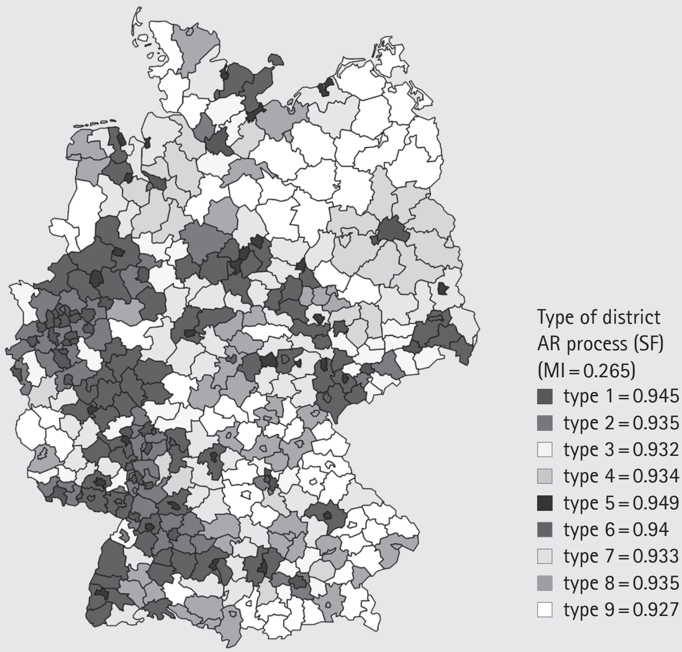


Figure 4.10: Map of estimated spatial-regimes AR(1) parameters: SF approach [parameters $\breve{\alpha}_r$ according to equation (4.9)]



4.4.4 Concluding remarks: Persistence of unemployment

The empirical findings presented in this section give a clear picture of unemployment persistence in Germany. We find the adjustment speed of regional unemployment to shocks to be extremely heterogeneous, which makes estimation of a single AR-parameter look unreasonable and supports our call for regionally disaggregated estimations. Modelling the heterogeneity by SFGWR seems to capture most of this heterogeneity, but spatial regimes do surprisingly well too. The averages over the AR parameters – and the majority of them – throughout the various specifications lie between 0.76 and 0.96, that is, close to 1. Thus, shocks to unemployment may be expected to be persistent, or at least to have a long half-life in most regions. For example, an AR parameter of 0.8 is equivalent to a half-life of more than three quarters, or the effect of the shock vanishing after eight years (10 times the half-life); an AR parameter of 0.9 corresponds to a half-life of 6.6 quarters, and a parameter of 0.95 to a half-life of 13.5 quarters. When using Dickey-Fuller equivalent transformations of the models, we can reject the hypothesis that the difference of the average autoregressive parameter minus one – the average of this distance is between -0.24 and -0.04 – is greater than or equal to zero. At least on average, unemployment is stationary – a necessary condition for the existence of (conditional) convergence – although non-stationarity can hardly be rejected for a large fraction of regions. Thus, unemployment adjusts very slowly – if ever – toward a kind of natural rate; it behaves (in particular in the agglomerated districts along the river Rhine) more like a random walk. Saying that there is clear evidence of (cross-sectional) convergence among the rates would be an excessive statement.

Our findings are particularly significant with regard to exogenous shocks: positive, in the case of active labour market policy interventions; negative, as in the case of the recent global economic crisis. Strong persistence of the regional unemployment rates suggests that a negative shock, due for example to a sudden increase in labour supply, to not-anticipated deflation, or to economic catastrophes, would take a rather long time to be absorbed. We can think, for example, of new labour regulations for foreign workers (the enlargement of the European Union from EU-15 to EU-25), of the collapse of the states/markets belonging to the socialist Council for Mutual Economic Aid (Comecon) in the late 1980s/early 1990s (affecting the former German Democratic Republic), or of political events as in Card (1990). In this regard, there is potential in expanding the above analyses to the analysis of relative unemployment, which appears to have different persistence dynamics than the absolute levels [see, e.g., Jimeno/Bentolila, 1998, where the determinants of unemployment persistence are also discussed].

4.5 Conclusions

Studies about the convergence or persistence of unemployment typically employ univariate autoregressive equations and test them for stationarity. This procedure is straightforward and computationally simple, but can hardly account for cross-sectional heterogeneity and dependence – thus, in the best case, it is statistically inefficient (imprecise) or, in the worst case, misspecified. Derived conclusions may then be misleading.

In this paper, we have focused on two questions. First, starting with a system of AR(1) equations, we aimed to show the substitutability of fixed effects (FE) and spatial filters and, analogously for autoregressive processes, the one between individual autoregressive parameters and SFGWR-type estimation. The SF surrogates [which allow to decompose the FE into a spatially structured and a spatially unstructured (random) part] are more parsimonious with regard to the number of parameters, and use, instead of region-specific parameters, a set of parameters defined and computed over all regions. Second, we applied SF methods when analysing the dynamics of quarterly regional unemployment rates for Germany from 1996 to 2004. Because the eigenvectors employed in an SF represent map patterns, one advantage of this approach is that the heterogeneous autoregressive adjustment parameters of the GWR-type models have a geographical interpretation. For comparison, we also provided estimates of a homogeneous autoregressive process, and of one approach differentiated according to nine urbanization/agglomeration regimes.

Indeed, when comparing pairwise the individual and SF specifications for the process component (AR or level), keeping everything else equal, we found that the SF approach provides a gain in residual degrees of freedom, without losing much estimation accuracy, measured, for example, in terms of goodness-of-fit (R^2) or root mean squared error (RMSE). We found, for the SF AR specification, some gain in precision when compared with the homogeneous and spatial regime specifications. Summary diagnostics for all models, based on information criteria, provided a confirmation of the potential of the proposed SF-based models. The residuals from individually-specified models and of their corresponding SF equivalents are highly correlated, and the error distributions are quite similar pairwise. The estimates for the average autoregressive parameter vary, in particular, between the FE estimation with homogeneous seasonal effects (0.76 – 0.85) and the remaining level/seasonality combinations (0.90 – 0.96). Consequently, a potential bias in the autoregressive parameter does not seem to depend on the way in which the autoregressive process is specified. However, obtaining exact evidence about the consistency of the AR estimates is only possible by means of Monte Carlo simulation. This aspect will be the subject of future research, since here we limit ourselves to showcasing the practical relevance of the proposed approaches. A further aspect that may be expected to be investigated in future research is the extension of the proposed models to the estimation of nonlinear regression models (e.g., in the case of unemployment rates, the logistic regression), for which panel models are generally less popular in the econometric literature and competition with other applied statistics fields is stronger (e.g., generalized linear mixed models).

4.A Comparative model overview and further statistics

Table 4.5: Information criteria results

AR process	Levels	Seasonality	Av.AR	R²	RMSE	Res. Dfs.	Nr. Coeffici-ents	AIC	BIC	HQ
Homogenous	FE	Homogenous	0.766	0.996	0.827	14,922	443	-0.321	-0.095	-0.246
Homogenous	SF	Homogenous	0.945	0.975	0.872	15,321	44	-0.268	-0.246	-0.261
Homogenous	FE	Heterogeneous	0.901	0.992	0.504	13,608	1,757	-1.112	-0.141	-0.789
Homogenous	SF	Heterogeneous	0.957	0.991	0.530	13,979	1,386	-1.071	-0.323	-0.822
Heterogeneous	FE	Homogenous	0.833	0.981	0.753	14,484	881	-0.446	0.015	-0.292
Heterogeneous	SF	Homogenous	0.823	0.980	0.777	14,865	500	-0.437	-0.181	-0.352
Heterogeneous	FE	Heterogeneous	0.906	0.992	0.493	13,170	2,195	-1.081	0.166	-0.665
Heterogeneous	SF	Heterogeneous	0.914	0.992	0.500	13,564	1,801	-1.121	-0.123	-0.788
SFGWR	FE	Homogenous	0.853	0.980	0.849	14,876	489	-0.262	-0.012	-0.179
SFGWR	SF	Homogenous	0.935	0.978	0.824	15,227	138	-0.369	-0.300	-0.346
SFGWR	FE	Heterogeneous	0.882	0.985	0.666	14,772	593	-0.733	-0.428	-0.631
SFGWR	SF	Heterogeneous	0.961	0.986	0.650	15,064	301	-0.822	-0.669	-0.771
Spatial regimes	FE	Homogenous	0.808	0.978	0.810	14,914	451	-0.361	-0.131	-0.285
Spatial regimes	SF	Homogenous	0.937	0.975	0.869	15,306	59	-0.273	-0.244	-0.263
Spatial regimes	FE	Heterogeneous	0.812	0.714	0.754	14,890	475	-0.501	-0.258	-0.420
Spatial regimes	SF	Heterogeneous	0.946	0.979	0.798	15,291	74	-0.442	-0.405	-0.429

Chapter 5

Forecasting Regional Labour Markets with GVAR Models and Indicators

By Norbert Schanne[32]

32 I greatfully acknowledge research assistance by Marco Weissler in the very early stages of this paper. I would like to thank Sabine Klinger, Alexander Vosseler, Rüdiger Wapler and Antje Weyh, Sergio Rey and three anonymous referees for helpful comments. Thanks also to seminar participants at IAB Nuremberg and Regensburg University, as well as to conference participants at the SEA 2010 in Chicago, the ISF 2010 in San Diego and the ERSA 2010 in Jönköpping. Remaining errors are to my fault.

Abstract

It is broadly accepted that two aspects regarding the modeling strategy are essential for the accuracy of forecast: a parsimonious model focusing on the important structures, and the quality of prospective information. Here, we establish a Global VAR framework, a technique that considers a variety of spatio-temporal dynamics in a multivariate setting, that allows for spatially heterogeneous slope coefficients, and that is nevertheless feasible for data without extremely long time dimension. Second, we use this framework to analyse the prospective information regarding the economy due to spatial co-development of regional labour markets in Germany. The predictive content of the spatially interdependent variables is compared with the information content of various leading indicators which describe the general economic situation, the tightness of labour markets and environmental impacts like weather. The forecasting accuracy of these indicators is investigated for German regional labour-market data in simulated forecasts at different horizons and for several periods.

Germany turns out to have no economically dominant region (which reflects the polycentric structure of the country). The regions do not follow a joint stable long run trend which could be used to implement cointegration. Accounting for spatial dependence improves the forecast accuracy compared to a model without spatial linkages while using the same leading indicator. Amongst the tested leading indicators, only few produce more accurate forecasts when included in a GVAR model, than the GVAR without indicator.

Keywords: Global VAR; Labour-market forecasting; Leading indicators; Regional forecasting; Space-time dynamic model

Jel: C 23; E 24; E 27; R 12

5.1 Introduction

Making predictions on the aggregate development of quantities and prices in the markets – e.g. GDP, inflation, liquidity demand, or as in this paper, unemployment and employment growth – is one of the most important tasks of the economic profession. Much of the recent criticism the discipline still has to face is due to the fact that neither the credit-crunch crisis (2008/2009) nor the Euro-zone crisis (2010 till present) were foreseen by a notable fraction of economists. Moreover, many of the GDP and labour-market forecast revisions made throughout the crisis turned out to be wrong as well, regardless of the statistical or economical model behind the forecasts.[33] Alike most parts of the Western hemisphere, the German economy suffered a strong decline in real GDP within one year[34] – but, in contrast to many countries, did not show a strong reaction in unemployment, a phenomenon denoted as the *German job miracle* (Möller, 2010). A weakening of the relation between GDP and unemployment is even observed in other countries, albeit at most as a jobless recovery. For example, the Chairwoman of the US 'Board of Economic Advisors' Christina Romer has remarked in the New York Times on Feb. 11, 2011:[35] "The usual relationship between GDP growth and the unemployment rate has broken down somewhat"; that is, labour market's dependence on expected production and the business cycle – stated by Okun's Law – may have relaxed. This encourages us to re-think the economic relations and indicators we previously employed for forecasting labour markets at the regional (and even at the national) level.

It is broadly accepted that two aspects regarding the modeling strategy are essential for the accuracy of forecasts: on the one hand a parsimonious model focussing on the important dependency structures and simplifying or omitting the less relevant, on the other hand the utilization of prospective information with a high information content. This paper deals with both issues in the context of forecasting regional labour markets; thus, its contribution to the literature is twofold. First, we establish a framework that considers spatiotemporal dynamics within the labour market in a multivariate setting. The model deals with the dimensionality problem of large heterogeneous spatial systems. It has the advantage of, in principle, allowing for both weak (spatially declining) cross-

33 For example, the unemployment forecasts of the Institute of Employment Research (IAB) and the German Federal Employment Agency shifted from the pre-crisis prediction for 2009 (Aug. 2008) of 3.16 mio over 3.3 mio in the early crisis (Oct. 2008) to 3.6 mio (Feb. 2009), and saw unemployment reaching 4.1 mio in 2010 (Summer 2009). Roughly at the same time (June 15 2009), 'Deutsche Bank'-economist Norbert Walter predicted unemployment to exceed 5 mio in 2010, to our knowledge the most pessimistic forecast for Germany.

34 The growth rate according to the German Federal Statistical Office published on January 13th 2010 mounts to -5%.

35 http://www.nytimes.com/2010/02/12/business/economy/12usecon.html, accessed Nov. 16th, 2012.

sectional dependence and strong dependence on a dominant region (which is rejected later from the data at the level of aggregation used in the analysis). Second, applying this framework to estimate and forecast the spatial co-development of regional labour markets, we examine the information content of several economically prospective as well as non-economical indicators.

Recent research has found improvements in univariate forecasts on regional labour market quantities when accounting for spacetime dynamics. E.g., Longhi/Nijkamp (2007) forecast local employment under consideration of contemporaneous spatial dependence. In Hampel et al. (2008) and Schanne/Wapler/Weyh (2010) serially lagged spatial dependence contributes to a higher forecast accuracy in employment and unemployment, respectively. Mayor/Patuelli (2012) employ a Spatial VAR (SpVAR) and a Spatial-Filter based dynamic heterogeneous coefficient model (SFGWR) to predict unemployment; they argue that the SpVAR is advantageous when forecasting in long data with a small cross-sectional dimension, and that the SFGWR forecast performance becomes relatively better with a shorter observation period or increasing number of regions. However, recently techniques became available that allow on the one hand for a certain variability of other variables' impacts across the regions, and on the other hand provide a more distinct view on the cross-dependence of regions by discussing the conditions for either strong or weak cross-sectional dependence (Pesaran/Tosetti, 2007; Chudik/Pesaran, 2011; Chudik/Pesaran/Tosetti, 2011). Cross-sectionally weak dependence is defined as correlation patterns which arise in a certain group of series and which do not extend towards series not included in this group. An example is spatial autocorrelations where the degree of cross-sectional dependence declines over space. In contrast, cross-sectionally strong dependence requires the existence of a series (a region) that is correlated with all other series in the system; these are considered to depend on this dominating series. This method has been developed for modeling a multinational monetary system (Pesaran/Schuermann/Weiner, 2004, Dees et al., 2007, Pesaran/Schuermann/Smith, 2009), hence it is denoted as Global VAR (GVAR). It is also employed to model the spatio-temporal diffusion of shocks on housing prices across regions and to forecast real-estate markets in the UK (Holly/Pesaran/Yamagata, 2011). To our knowledge, we are the first to adopt this method in a multivariate model of regional labour market development.

Once the baseline model has been established, further investigating the forecast content of the usual leading indicators is quite a natural exercise, of interest for two reasons. On the one hand, it might be possible that these indicators entail only information which develops simultaneously and which is already incorporated in the (local and spatially interdependent) labour market history itself: e.g., consumption or wholesale sentiments will be affected by aggregate disposable income which, in turn, will be affected by recent unemployment. If an indicator for consumption provides more prospective information than the recent development within the labour market is thus an empirical question, but so is the question whether small independent single region indicator models are more precise in forecasting than complex models with interdependent regions. On the other hand, most frequently employed leading indicators in both national and regional forecasting refer to sentiments or register-based information on the development of production or financial markets: Economic tendency surveys, stock-market indices, wholesale, new orders, etc. Further investigation of their forecast content regarding employment and unemployment stands to reason, given the somewhat broken relationship between production and labour markets. With regard to the information content of the usual leading indicators several questions may be raised: If (regional) labour markets have detached from GDP (or product markets), is the relation between labour and product markets weakened only temporarily or regionally, and has a new relation already formed? If the relation is broken, can expectations regarding product market developments still contribute to improve the forecast accuracy in labour markets? And, are other indicators, beyond business expectations, available providing information regarding the future development of regional labour markets? We thus compare standard production and financial indicators with supposedly prospective indicators immanent to the labour market. Furthermore, we test another set of potential indicators which describe climate data. The idea for this results from the observation that both local employment and unemployment show extraordinary persistent shifts in particular in periods with abnormal weather phenomena, e.g. the mild winters in 2006/07 and 2007/08 – in these, unemployment did not show the usual seasonal increase at the end of the year, but a spring decline with normal size.

In this paper, we focus on the development of regional labour market quantities, (log) employment and (log) unemployment, at a monthly frequency, for which data is introduced in Section 5.2. Section 5.3 sketches theoretically the economic intuition behind the empirical model developed in Section 5.4. Here, the focus is on describing the Global VAR, an econometric approach that makes the two-variables multi-regional system of time series tractable. The identification of joint developments of the non-stationary series and the identification of a dominant region (or our failure to identify it) are of particular interest. Section 5.5 introduces the various indicators tested out in the subsequent forecasting exercise. The forecast accuracy of the models is discussed in Section 5.6 by evaluation of simulated out-of-sample forecasts for the ten regional subdivisions of the Federal Employment Agency (in size roughly equivalent to NUTS-1 regions). Indeed, in our setting, the prospective information regarding the labour market which is provided by business-cycle indicators turns out to be extremely limited, hardly exceeding the contribution some climate series can make to labour-market forecasting. In contrast, accounting for cross-sectional dependence improves the forecast accuracy in most of the tested specifications.

5.2 The data and their statistical properties

Information on labour-market quantities is provided at various regionally disaggregated levels by the German Federal Employment Agency (FEA, Bundesagentur für Arbeit). Our monthly series on unemployment and employment stem from register data, begin in January 1996 and are not seasonally adjusted. Unemployment covers all persons officially registered as unemployed: they receive unemployment benefits from the FEA, look for a job and are ready to take on a job. Employment covers all employees in full- and part-time jobs liable to social security contributions, reported at their workplace. The analysis is carried out at the level of the Federal Employment Agency's Regional Divisions (RD). These are equivalent or slightly larger than the German federal states, often entailing two smaller states. Some descriptive statistics and a stationarity analysis are provided in Table 5.1.

To highlight just a few details in the data, we first observe that the difference between the largest region, North Rhine-Westphalia, and the smallest regions, Saxony and Rhineland-Palatinate/Saarland, mounts to less than 1.5 log points – equivalent to NRW being approximately 4.5 times as large as the smallest regions, and not 20 times as we find it at the level of Federal States. That is, the RD series reflect indeed a rather homogeneous number of persons. Second, the average monthly change of log employment and log unemployment is almost zero in any region whereas the standard deviations of both the levels and the monthly differences are larger. Hence, we carry out the subsequently presented analyses without considering a deterministic linear trend. Third, furthermore, the reported ADF tests and HEGY tests always do not reject unit roots in the first lag (the zero frequency) at the 99% significance level (and only in a few regions at the 95% level), whereas non-stationarity at the seasonal frequencies and in the monthly-differentiated series are rejected; this finding is supported by other (not reported) unit-root tests. Thus, we consider all series to be integrated of order one, $I(1)$. Shocks in the regional labour markets can be considered as persistent. Information on additional features of the data will be provided throughout the following sections, after the corresponding description of the estimation technique.

Table 5.1: Descriptive statistics and unit-root tests

	Levels (Y)		Differences (ΔY)		HEGY-tests		ADF-tests	
RD	Mean	St.Dev.	Mean	St.Dev.	$t_{(\pi_1)}$	$F_{(\pi_2-\pi_{12})}$	$t_Y^{(DF)}$	$t_{\Delta Y}^{(DF)}$
				Log unemployment, 1/1996–12/2011				
Nord	12.78	0.14	-0.00	0.04	-0.89	577.18	-1.42	-8.11
BB	13.07	0.14	-0.00	0.03	-0.25	607.79	-1.04	-9.02
Sat	12.90	0.24	-0.00	0.05	0.54	369.57	-0.39	-8.42
S	12.75	0.20	-0.00	0.04	0.38	494.11	-0.53	-8.08
BY	12.82	0.22	-0.00	0.07	-1.22	824.59	-1.64	-8.27
BW	12.63	0.17	-0.00	0.03	-2.15	1660.31	-0.55	-9.64
RPS	12.14	0.14	-0.00	0.04	-1.01	555.83	-1.20	-8.81
H	12.35	0.14	-0.00	0.03	-1.63	831.98	-0.89	-9.43
NRW	13.64	0.10	-0.00	0.02	-1.87	851.34	-1.04	-8.86
NSB	12.90	0.14	-0.00	0.04	-1.01	461.93	-1.24	-9.07
				Log employment, 1/1996–12/2011				
Nord	14.57	0.03	0.00	0.01	-2.26	1308.77	-1.47	-7.24
BB	14.44	0.05	-0.00	0.01	-1.72	1285.16	-1.89	-8.02
Sat	14.26	0.07	-0.00	0.01	-1.86	495.83	-1.87	-7.35
S	14.18	0.06	-0.00	0.01	-2.01	610.08	-1.88	-7.20
BY	15.29	0.04	0.00	0.01	-1.16	1,040.44	-0.96	-8.90
BW	15.15	0.03	0.00	0.01	-1.54	1,454.40	-0.55	-11.73
RPS	14.24	0.03	0.00	0.01	-1.74	682.84	-1.19	-9.15
H	14.58	0.03	0.00	0.01	-2.12	696.32	-0.98	-9.75
NRW	15.57	0.02	0.00	0.01	-2.21	981.57	-1.05	-9.76
NSB	14.80	0.03	0.00	0.01	-1.09	417.34	-1.17	-9.26

BB: Berlin & Brandenburg – BW: Baden-Wurttemberg – BY: Bavaria – H: Hesse – Nord: City of Hamburg, Mecklenburg-Western Pommerania, Schleswig-Holstein – NRW: North Rhine-Westphalia – NSB: City of Bremen & Lower Saxony – RPS: Rhineland-Palatinate & Saarland – S: Saxony – SAT: Saxony-Anhalt & Thuringia

HEGY-tests are carried o ut with seasonal dummies and a constant (without deterministic trend), see Beaulieu/ Miron (1993). The critical values at the 5% level are -2.760 for the zero frequency and 4.490 for the joint test on the seasonal frequencies.

The ADF tests refer to deseasoned data. The critical value at the 5% level is -2.889, at the 1% level -3.507.

5.3 Regional labour-market dynamics: A sketch

The standard dynamics in a search-matching framework (see, for example, the textbook version of Cahuc/Zylberberg, 2004: Ch. 9.3) can be adapted in a multi-region model such that the unemployment change equation is

$$\begin{aligned}\Delta U_{i,t} = \Delta N_{i,t} - \left[m(\theta_i)\,(1-a_i) + m(\theta_i^*)\,a_i \right] U_{i,t-1} \\ + \delta_i (1-a_i) L_{i,t-1} + \delta_i^* a_i L_{i,t-1}^* \end{aligned} \tag{5.1}$$

where $U_{i,t}$ denotes unemployment in region i at time t. $N_{i,t}$ is the labour force (for simplicity, all entering persons are assumed to start as unemployed job-searchers, all retiring persons to leave from unemployment). Job-creation depends on the matching function $m(\theta)$ with $\theta = \frac{vacancies}{unemployment}$ the labour-market tightness observed at home θ_i or abroad θ_i^* and weighted with the probability to work at home $1-a_i$ or to commute a_i. Likewise, $L_{i,t}$ is employment and the parameters δ_i the job-separation rate; an asterisk marks variables abroad. Analogously, the employment change equation is

$$\Delta L_{i,t} = -\delta_i L_{i,t-1} + m(\theta_i)(1-a_i) U_{i,t-1} + m(\theta_i)\,a_i^* U_{i,t-1}^* \tag{5.2}$$

Stacking both equations, this can be written as

$$\begin{aligned}\begin{pmatrix} \Delta U_{i,t} \\ \Delta L_{i,t} \end{pmatrix} = \begin{bmatrix} -m(\theta_i) - \left[m(\theta_i^*) - m(\theta_i) \right] a_i & \delta_i (1-a_i) \\ m(\theta_i)\,(1-a_i) & -\delta_i \end{bmatrix} \begin{pmatrix} U_{i,t-1} \\ L_{i,t-1} \end{pmatrix} \\ + \begin{bmatrix} 0 & \delta_i^* a_i \\ m(\theta_i)\,a_i^* & 0 \end{bmatrix} \begin{pmatrix} U_{i,t-1}^* \\ L_{i,t-1}^* \end{pmatrix} + \begin{pmatrix} \Delta N_{i,t} \\ 0 \end{pmatrix} \end{aligned} \tag{5.3}$$

Note that the structure of this model is similar to a first order VEC model. Post-multiplying (5.3) with $\begin{pmatrix} U_{i,t-1}^{-1} & 0 \\ 0 & L_{i,t-1}^{-1} \end{pmatrix}$ results in a model for the log growth rates (or the difference of the logs) depending on the ratio of previous unemployment at home and abroad, the matching rates and the job-separation rates at home and abroad which can be assumed to be affected by, for example, the business cycle or labour market policy. The model in logs would moreover contain an approximately linear relationship $\ln L_{it}^* - \ln L_{it}$ included in the unemployment equation (as long as job separation is considered exogenous); supposedly, it won't be possible to derive a linear relationship containing unemployment because it is included also in the definition of labour-market tightness.

5.4 Specifying a system of regional labour markets

5.4.1 The Global VAR formulation

Vector Autoregressions (VAR) are the starting point for forecasting multiple interdependent time-series in a Global VAR model. Let $y_{i,t}$ denote the $m \times 1$ vector of target variables for region $i \in \{0, \ldots, n\}$ (here, $y_{i,t} = (\ln U_{i,t}, \ln L_{i,t})'$ with $m \times 2$). The vector $\xi_{it} = B_i x_{i,t}$ contains the contribution regarding unemployment's and employment's development provided by indicators $x_{i,t}$ available at time t (observed before t), and B_i the matrix of parameters corresponding to these indicators. For notational simplicity, $\xi_{i,t}$ entails the deterministic mean (modeled by a constant plus seasonal dummies and, for the estimations with a sample ending after June 2005, additionally by a dummy variable for the pre-2005 period in order to account for the structural break due to the 2004/2005 labour-market reforms in Germany) as well. Let $\mathbf{Y}_t = (y'_{0,t}, \ldots, y'_{n,t})'$, $\Xi_t = (\xi'_{0,t}, \ldots, \xi'_{n,t})'$, and Υ_t the random error vector. ϕ_ℓ is the coefficient matrix corresponding to lag $\ell = \{1, 2\}$; lag order 2 corresponds to the lag order determined to be optimal in region-specific VARs with seasonal dummies coincidentally according to the three information criteria of Akaike, Schwartz, and Hannan/Quinn (AIC, BIC, and HQIC), which should be sufficiently large even for a multi-regional VAR. Then, a VAR over all regions can be written as

$$\mathbf{Y}_t = \phi_1 \mathbf{Y}_{t-1} + \phi_2 \mathbf{Y}_{t-2} + \Xi_t + \Upsilon_t \tag{5.4}$$

or, rewritten in VEC form with $\Pi = (\phi_1 + \phi_2 - I_{m\times n})$ and $\Gamma = -\phi_2$, as

$$\Delta \mathbf{Y}_t = \Pi \mathbf{Y}_{t-1} + \Gamma \Delta \mathbf{Y}_{t-1} + \Xi_t + \Upsilon_t \tag{5.5}$$

These equation systems are not estimable unrestrictedly unless the number of regions is extremely small since the number of the coefficients in the square matrices ϕ_1, ϕ_2 and the residuals' covariance matrix Σ_Υ grows quadratically, with a rate of n^2 (and m^2). The idea of a GVAR is the following: To impose restrictions, we use the location and the corresponding information on geographical proximity between regions. This information allows the aggregation of the observable or predetermined information. Moreover, it allows to aggregate most of the unobservable simultaneous movement in the system (the correlated residuals) to a component which, under some additional assumptions discussed below, converges towards zero. Then, the system (5.4) or (5.5), respectively, can be split into partitions which may be considered independent from each other in an econometric sense and, hence, may be estimated partition-by-partition.

We assume analogously to Pesaran/Schuermann/Weiner (2004) that most regions contribute only little to explaining labour market development in other regions, relative to the joint influence of all other regions; as an aggregate, however, they may have a non-negligible impact. The labour market in region i can be considered to depend on the one hand on a important, dominant or leading region whose influence should be modelled explicitely[36], and on the other hand on a weighted average over the non-dominant regions instead of the particular development of each region $j = \{0, 1, \ldots, i-1, i+1, \ldots, n\}$. Variation in the strength of dependence across regions can be modeled by various predetermined or exogenous metrics for proximity between regions i and j. These weights $w_{ij,k}$ may reflect geographical, cultural, social or economical distance (see Conley/Topa, 2002, Corrado/Fingleton, 2012 and, for the pros and cons of different weights, the comments on Pesaran/Schuermann/Weiner, 2004 in the respective volume of the JBES).

Assumption 3 (Spatial weights in a GVAR). *Let matrix $W_{(N)}$ entail the sequences of weights in $w_{ij,k}$ (combined across variables $k = 1, \ldots, m$ and regions $i = (0, \ldots, n)$. $W_{(n)}$ satisfies a number of "smallness" or "granularity" conditions (see Chudik/Pesaran/Tosetti, 2011): that its spectral norm (the Euclidean matrix norm) is bounded by a sequence converging with rate $\frac{1}{\sqrt{n}}$ or faster to a constant, i.e. $\| W_{(n)} \|_2 = [max_k \lambda_{(W_{(n)} W'_{(n)})}]^{1/2} = O\left(\frac{1}{\sqrt{n}}\right)$, and that $\frac{w_{ij,k}}{\| W_{(n)} \|_2} = O\left(\frac{1}{\sqrt{n}}\right)$. These conditions hold if the row and column norms, i.e. $\| W_{(n)} \|_1 = sup \sum_{i=0}^{n} \left| w_{ij,k} \right| \leq c$ and $\| W_{(n)} \|_\infty \; sup \sum_{j=0}^{n} \left| w_{ij,k} \right| \leq c$, are bounded in absolute value (a standard assumption in spatial econometrics).*

Different weights can be employed for different variables (indexed with k), although typically the same weights are applied for all m elements of the vector $y_{i,t}$. We define the $m \times m$ matrix block $w_{ij} = w_{ij,k} I_m$, using the elements of a row-standardized contiguity matrix as weights in our empirical application (due to standardisation the row sums are always unity). These weights are used to construct the *local average* corresponding to region i (i.e. the weighted average

36 Examples for dominant units are London for the UK or Paris for France. Other countries like the US have a multi-core structure without a region (state, Metropolitan Area) that dominates the country as a whole. For Germany, ex ante, North Rhine-Westphalia (with the Ruhr area, Germany's largest agglomeration, and the one-million-inhabitant city of Cologne) could be considered a natural candidate for being dominant; however, other regions have similar size and economic power, hence a multipolar structure (without clear dominance structure) is possible as well.

over all "non-domestic" regions or, in other words, the spatial lag), subsequently denoted with $y^*_{i,t} = \sum_{j=0}^{n} w_{ij} y_{j,t}$. With granular weights, the local average can not be implicitly dominated by any single region.

The dominant region (and its history) drives the development of all other regions (series). It behaves, as it has been discussed by Pesaran/Tosetti (2007) and Chudik/Pesaran/Tosetti (2011), similar to a factor f_t in a dynamic factor model with mutual cross-sectional dependence (see also Stock/Watson, 2011, for an overview, and for factor methods using (dynamic) principal components Forni et al., 2000, 2005, Bai/Ng, 2002, Peña/Poncela, 2004):

$$y_{i,m,t} = \theta_{i,m} f_t + \upsilon_{i,t} \tag{5.6}$$

for $i = 1,\ldots,n$, with $\theta_{i,m}$ the vector of factor loadings which relate the common factor (or the dominant region) to the dependent variable and $u_{i,t}$ the idiosyncratic component. Non-dominant regions may show cross-sectional correlation with regard to their idiosyncratic part; however, this mutual dependence is too weak to form a distinct factor pattern which loads on all regions.

Let c_i denote the parameter matrix describing the contemporaneous dependence of region i on the innovation in the dominant region 0, i.e. the 'factor loadings'; $C_0 = \left(0_{m\times m}, c_1', \ldots, c_n'\right)'$ is the $\left[m\left(n+1\right)\times m\right]$ matrix of loadings. Then, we can assume for the error covariance matrix that

$$\Sigma_{\Upsilon_t} = R^{-1\prime} \begin{pmatrix} \sigma_0^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ 0 & \sigma_{2,1} & \sigma_2^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \sigma_{n-1,n} \\ 0 & \sigma_{n,1} & \cdots & \sigma_{n,n-1} & \sigma_{\mathrm{n}}^2 \end{pmatrix} R^{-1} \tag{5.7}$$

with $R = \left[I_{m(n+1)} - \left(C_0, 0_{m(n+1)\times mn}\right)\right]$, $\sigma_{ij} = E\left(\varepsilon_i \varepsilon_j'\right)$ the system cross-covariance between regions i and j, σ_i^2 the variance-covariance of the system within region i. Whereas for all non-dominant regions the errors may be interdependent, the dominant region is considered to be stochastically independent from the other regions. Thus, its development can be included contemporaneously as (weakly) exogenous variable in partial systems regarding other regions.

Granularity of the weights ensures that the covariance between the disturbance and its local average is bounded as $n \to \infty$:

$$
\begin{aligned}
E\left(\varepsilon_{it}\varepsilon_{it}^{*}{}'\right) &= E\left(\sum_{j=0}^{n}\varepsilon_{it}\varepsilon'_{jt}\,w'_{ij}\right) = \sum_{j=0}^{n}E\left(\varepsilon_{it}\varepsilon'_{jt}\right)\|W_n\|\frac{w'_{ij}}{\|W_n\|} \\
&= \sum_{j=0}^{n}\sigma_{ij}O\left(\frac{1}{\sqrt{n}}\right)O\left(\frac{1}{\sqrt{n}}\right) \le \sum_{j=0}^{n}\frac{1}{N}\sigma_{ij}
\end{aligned}
\tag{5.8}
$$

The variance-covariance of the local average converges towards zero as $n \to \infty$:

$$
E\left(\varepsilon_{it}^{*}\varepsilon_{it}^{*}{}'\right) = E\left(\sum_{j=0}^{N}w_{ij}\varepsilon_{jt}\varepsilon'_{jt}\,w'_{ij}\right) \le \frac{1}{N^2}\sum_{j=0}^{N}O(1)\sigma_j^2 \to 0 \tag{5.9}
$$

Hence, under granularity of weights, the local average can be considered as asymptotically (weakly) exogenous.

Utilizing weak exogeneity of the local averages, the cross-regional equation system (5.5) – the VEC form is used here for convenience – is divided into only weakly dependent blocks of equations. The vector $y_{0,t}$ is included separately in the systems for all other regions, whereas the non-dominant units are accounted for through $y_{i,t}^{*}$. Then, the system for the labour market (unemployment and employment) in a single region becomes

$$
\begin{aligned}
\Delta y_{i,t} &= h_i y_{i,t-1} + h_i^{*} y_{i,t-1}^{*} + h_i^{0} y_{0,t-1} + g_i \Delta y_{i,t-1} + g_i^{0}\Delta y_{0,t-1} + g_i^{*}\Delta y_{i,t-1}^{*} \\
&\quad + \mu_{i,t} + c_i \Delta y_{0,t} + \varepsilon_{i,t}
\end{aligned}
\tag{5.10}
$$

For the dominant region itself (for $i=0$) or, if there is no dominant unit for all regions, a region-specific equation system can be extracted as

$$
\Delta y_{i,t} = h_i y_{i,t-1} + h_i^{*} y_{i,t-1}^{*} + g_i \Delta y_{i,t-1} + g_i^{*}\Delta y_{i,t-1}^{*} + \mu_{i,t} + \varepsilon_{i,t} \tag{5.11}
$$

By defining $\mathbf{w}_i = \left(w_{i0},\ldots,w_{iN}\right)'$ and

$$
G = \begin{pmatrix}
g_0 & 0 & 0 & \cdots & 0 \\
g_1^0 & g_1 & 0 & \cdots & 0 \\
g_2^0 & 0 & g_2 & \ddots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
g_N^0 & 0 & \cdots & 0 & g_N
\end{pmatrix}
+
\begin{pmatrix}
g_0^{*}\mathbf{w}'_0 \\
g_1^{*}\mathbf{w}'_1 \\
g_2^{*}\mathbf{w}'_2 \\
\vdots \\
g_N^{*}\mathbf{w}'_N
\end{pmatrix}
$$

$$
H = \begin{pmatrix}
h_0 & 0 & 0 & \cdots & 0 \\
h_1^0 & h_1 & 0 & \cdots & 0 \\
h_2^0 & 0 & h_2 & \ddots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
h_N^0 & 0 & \cdots & 0 & h_N
\end{pmatrix}
+
\begin{pmatrix}
h_0^{*}\mathbf{w}'_0 \\
h_1^{*}\mathbf{w}'_1 \\
h_2^{*}\mathbf{w}'_2 \\
\vdots \\
h_N^{*}\mathbf{w}'_N
\end{pmatrix}
$$

stacking the region specific systems (5.11) and (5.10) gives, here again for the case of a single dominant region, a structural VEC over all regions:

$$R\Delta\mathbf{Y_t} = H\mathbf{Y_{t-1}} + G\Delta\mathbf{Y_{t-1}} + \mu_\mathbf{t} + \varepsilon_\mathbf{t} \tag{5.12}$$

System (5.5) follows from $\Upsilon_\mathbf{t} = R^{-1}\varepsilon_\mathbf{t}$, $\Pi = R^{-1}H$, $\Gamma = R^{-1}G$ and $\Xi_\mathbf{t} = R^{-1}\mu_\mathbf{t}$. Solve backward with $\phi_1 = \left(\Pi + \Gamma - I_{m(n+1)}\right)$ and $\phi_2 = -\Gamma$ to get the VAR described in system (5.4). The unit-specific systems given by eqs. (5.10) and (5.11) can be estimated region-by-region. Since the number of parameters per region-specific partial system is limited, the estimations are computationally tractable. What needs to be answered ex-ante is on the one hand which regions share common stochastic trends, on the other hand whether there is one (or more than one) dominating region, and if so, which regions are dominant.

5.4.2 Regional dominance in cross-sectional dependence

We start the discussion on the strength of cross-sectional dependence from a dynamic-factor perspective since a dominant region can be considered as a strong factor. Suppose [as in eq. (5.6)] that each series $y_{i,m,t}$ in Y_t can be separated additively into a component $\chi_{i,m,t} = \theta_{i,m} f_t$ which entails the co-evolution of the series due to a small number of common factors f_t and an idiosyncratic component $\upsilon_{i,m,t}$. The factor space is spanned by the (dynamic) principal components of the systems variance-covariance matrix $\Sigma_{(Y)t}$. Principal Component Analysis (PCA) allows to determine a standardized orthogonalization of the factor space; the factors themselves might be represented by any rotation of the factor space (Forni et al., 2000). The dimension of the factor space (the number of factors r) equals the number of diverging eigenvalues of $\Sigma_{(Y)t}$. Research using PCA frequently employs the number of eigenvalues exceeding one in absolute value as a criterion. Several studies (e.g. Bai/Ng, 2002) establish information criteria penalizing each additional factor in order to test for the number of factors. Here, we use the procedure presented by Onatski (2010) which, though still overestimating the number of factors in small cross-sections, tends to perform better than many information criteria. The number of factors according to standard PCA and to the Onatski-Criterion are, besides other statistics described below, reported in Table 5.2. We show numbers for the original $I(1)$ series (Y), for series filtered for deterministic seasonal means ($Y - \overline{Y}_s$), for the (stationary) first differences of the seasonally filtered series ($\Delta\left[Y - \overline{Y}_s\right]$) and for series filtered for spatial auto-correlation (that is, weak dependence) ($\left[I - \rho W\right]Y$), because the findings may be sensitive to joint deterministic components, non-stationarity and mutual correlation of the idiosyncratic components.

In the traditional factor literature, the diverging (factor-related) eigenvalues increase linearly in the cross-sectional dimension, that is with a rate equal to n, whereas the remaining eigenvalues are bounded and independent from the cross-sectional dimension. As a consequence, statistical criteria for the identification of factor structures test for divergence vs. non-divergence of the eigenvalues while neglecting their actual rate of divergence (e.g. Bai/Ng, 2002). Chudik/Pesaran/Tosetti (2011) have introduced a concept of semi-strong (or semi-weak) factors associated with eigenvalues that increase less than linearly, at a rate of $O\left(n^{1-\varepsilon}\right)$ with $\varepsilon \in (0,1)$. Semi-strong and semi-weak factors affect only a limited number of regions/series but not all, or are related with every region with a strength of relation declining at a rate faster than $\frac{1}{n}$.[37] Semi-strong or weaker factors generate only cross-sectionally weak dependence – similarly to those correlation patterns that are considered $O(1)$, that is non-increasing with the cross-sectional dimension. For the existence of a dominant region in a GVAR model it must however hold – as for any model with at least one strong factor – that the first (maximum) eigenvalue diverges linearly, i.e. that $\max\lambda_{\left(\Sigma Y_t|\mathcal{I}_t, w_{(n)}\right)} = O(n)$. Hence, to identify or reject regional dominance, we have to determine the exponent $(1-\varepsilon)$ in the order of divergence.

Bailey/Kapetanios/Pesaran (2012) propose estimators to determine the order of divergence of the largest eigenvalue (here denoted with $O(n^{\alpha})$) in a system of equations; we employ the first-order bias corrected version $\tilde{\alpha}$. The procedure follows the intuition that, within a system of n equations, αn series are strongly cross-sectionally dependent whereas the remaining are independent (or at most weakly dependent) from the others; strong cross-dependence results from factor loadings (with a single factor) which are bounded away from zero. In this setting, the largest eigenvalue of the system's variance-covariance matrix has to increase at order n^{α}. The estimator has the main advantage that its distribution has been established analytically; this allows to test the order of divergence against lower and upper thresholds, albeit at a very low rate of convergence ($\ln n$). However, it does not fit perfectly to our analysis: On the one hand, the procedure employs only the overall system covariance matrix and the proportion of this covariance assessable to the first factor. On the other hand, those regions which are not strongly dependent are considered as uncorrelated (in contrast to being weakly

37 In the words of Chudik/Pesaran (2013), p. 15: "… that affects only a subset of the units and the number of affected units rise more slowly than the total number of units." Formally, a factor is strong if all factor loadings are bounded away from zero, $|\theta_{i,m}| > 0$. A semi-strong factor requires that, for some i, m, the loadings are not bounded away from zero.

dependent). A fraction of strongly correlated units and a complementary fraction of independent units do not match the spatial structure where the degree of dependence declines continuously in the distance. Finally, the maximum number of neighbours which a single region has in our case is six (of nine possible regions); hence, the lower threshold which describes only weak (spatial) correlation is at least in the univariate partial systems (only unemployment, only employment) extremely high.

Thus, as an alternative we try to estimate the exponent in $O(n^{1-\varepsilon}) = n^{1-\varepsilon}O(1)$ from randomly determined partial systems with varying size, an approach which has not been pursued before and of which the statistical properties have not been proven yet. For each possible number of regions $n \in 2,\dots,n$, we randomly draw 50 subsystems (without replacement, i.e. without including the same series twice). We retain the largest eigenvalue of the corresponding variance-covariance matrix $\lambda_{max,(n),j} = \max\lambda_{(\Sigma)}$ and the size of the system n_j for each iteration j. Then, we estimate the equation

$$\ln\left(\lambda_{max,(n),j}\right) - \ln(n)_j = -\varepsilon \ln(n)_j + c + \nu_j \tag{5.13}$$

to determine the parameter ε in the eigenvalues' order of divergence. If the estimate $-\hat{\varepsilon}$ is significantly[38] negative, we can conclude that the eigenvalues diverge at a less than linear rate.

Results for the cross-regional systems of unemployment, employment and the two variables jointly together are provided in Table 5.2, differentiated by the two sampling periods and the various filters. PCA suggests throughout all filters and sampling periods three to four factors in the whole system of twenty equations. Two factors are inherent to the employment subsystem and one to three factors exist in the unemployment subsystem. That is, according to PCA it is likely that one factor affects both variables jointly whereas there exist also factors specific to unemployment and employment. The Onatski criterion is much more restrictive: when taking the unfiltered, the seasonally filtered and the monthly-differentiated series into consideration, it suggests that only one factor exists which affects both employment and unemployment together. If we filter the data for spatial autocorrelation, it rejects the existence of a factor in the unemployment subsystem and, in the longer observation period, even in the complete system.

---

38 The 'population' of series in our case is small, and so is the number of candidate eigenvalues. Eigenvalues are Tracy-Widom distributed (Tracy/Widom, 1994; Onatski, 2009). Thus, since inference may be non-standard, critical values have been derived by simulation. Critical values for the hypothesis $H_0: -\varepsilon \geq 0$ are provided in Table 5.10 in the appendix to this chapter, together with the code for the simulation.

Table 5.2: Rate of divergence and factor structure in spatially filtered series

Statistic	Filter	$U_1,\ldots,U_N$	$L_1,\ldots,L_N$	$U_1,\ldots,U_N$		$L_1,\ldots,L_N$	
		2005	2011	2005	2011	2005	2011
r (PCA)	Y	3	3	2	2	2	2
r (Onatski)		1	1	1	1	1	1
$\hat{\varepsilon}$		0.099	0.062	0.116	0.043	0.181	0.172
		(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
$\tilde{\alpha}$		0.653	0.755	0.958	0.986	0.930	0.938
		(0.92)	(0.33)	(0.14)	(0.11)	(0.08)	(0.09)
r (PCA)	$Y-\bar{Y}_s$	3	3	2	1	2	2
r (Onatski)		2	1	1	1	2	2
$\hat{\varepsilon}$		0.116	0.078	0.117	0.041	0.199	0.193
		(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
$\tilde{\alpha}$		0.639	0.758	0.949	0.985	0.927	0.933
		(1.84)	(0.31)	(0.18)	(0.46)	(0.08)	(0.10)
r (PCA)	$\Delta(Y-\bar{Y}_s)$	4	4	2	2	2	2
r (Onatski)		1	1	1	1	2	1
$\hat{\varepsilon}$		0.155	0.157	0.108	0.100	0.206	0.169
		(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
$\tilde{\alpha}$		0.825	0.837	0.962	0.969	0.921	0.940
		(0.23)	(0.34)	(0.04)	(0.03)	(0.09)	(0.06)
r (PCA)	$(I-\rho W)\ Y$	3	3	3	3	2	2
r (Onatski)		2	0	0	0	1	1
$\hat{\varepsilon}$		0.218	0.168	0.257	0.155	0.119	0.121
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\tilde{\alpha}$		0.623	0.691	0.883	0.945	0.827	0.848
		(0.34)	(0.37)	(0.13)	(0.11)	(1.47)	(0.54)

Data filtered for spatial autocorrelation using Kapoor/Kelejian/Prucha (2007) estimates applied after a dynamic panel regression with homogeneous AR(1); for unemployment $\hat{\rho} = .5450$, for employment $\hat{\rho} = .6593$.

All estimates of the exponent in the order of divergence achieved in the subsystems $U_1,\ldots,U_N$ and $L_1,\ldots,L_N$ reject that the series are independent: $\tilde{\alpha}$ is always significantly above 0.5; $\hat{\varepsilon}$ is in all tests smaller than one. Moreover, the estimates $\tilde{\alpha}$ suggest an exponent larger than 0.9 in the two subsystems with all filters except the one eliminating spatial correlation; an exponent of one, i.e. cross-sectionally strong dependence, is not rejected in any of these specifications. In contrast, $\tilde{\alpha}$ is much lower when estimated from the complete system, with standard errors so large that neither strong dependence nor independence can be rejected. The estimates $\hat{\varepsilon}$ in the unemployment subsystem hint in the same direction as the estimates $\tilde{\alpha}$, rejecting strong dependence only if we eliminate weak dependence before. In the employment subsystems, perfectly strong dependence can be rejected at the usual significance levels. The estimates for ε achieved when sampling from all regional variables are in general in between the estimates achieved in the employment and unemployment subsystems; the values for $-\hat{\varepsilon}$ are significantly smaller than the 10% critical value for the case of strong dependence but not smaller than the 10% value under semi-strong dependence in Table 5.10. Hence, estimation of ε likewise suggests only semi-strong but not strong cross-sectional dependence.

5.4.3 Common trends: Cointegration and nonstationary common factors

We have seen in Table 5.1 that regional employment and unemployment (in logs) can be considered as nonstationary series, integrated of first order $I(1)$.[39] In the following, we analyse if two (or more) series – unemployment and employment within a region, or the same variable across regions – have a joint nonstationary stochastic trend describing the long-run relationship between the variables; the adjustment to deviations from such a trend can be employed in order to improve the forecasts. Joint trends across the regions can arise in the relationship to a possible dominant region or be due to correlated persistent shocks on neighbouring regions. They can be incorporated in system (5.10) by restricting

$$\begin{aligned} h_i y_{i,t-1} + h_i^* y_{i,t-1}^* + h_i^0 y_{0,t-1} &= \left(h_i - \eta_i^0 - \eta_i^*\right) y_{i,t-1} \\ &+ \eta_i^0 \left(y_{i,t-1} + \left[\eta_i^{0+} h_i^0\right] y_{0,t-1}\right) + \eta_i^* \left(y_{i,t-1} + \left[\eta_i^{*+} h_i^*\right] y_{i,t-1}^*\right) \end{aligned} \tag{5.14}$$

where A^+ denotes the generalized inverse of reduced-rank matrix A. $\left(h_i - \eta_i^0 - \eta_i^*\right)$ describes a linear relationship between the series of region i in levels; that is, intra-region cointegration. $\left[\eta_i^{0+} h_i^0\right]$ describes a linear stationary relationship between region i and a potentially dominant region, $\left[\eta_i^{*+} h_i^*\right]$ is the linear stationary relationship between region i and the corresponding local average; η_i^0 and η_i^* are the corresponding loading matrices.[40] For cointegration between the series, the traditional approach to model common trends (Engle/Granger, 1987, Lütkepohl, 2005), the parameter matrices have to satisfy either $rk\left\{\left(h_i - \eta_i^0 - \eta_i^*\right)\right\} > 0$, $rk\left\{\eta_i^0\right\} > 0$, or $rk\left\{\eta_i^*\right\} > 0$. Two (or more) series are considered to share a common trend if there exists a linear combination of the series that is stationary (integrated

39 As a consequence of the series being $I(1)$, matrix $\prod$ in eq. (5.5) has less than full rank; its determinant is 0.

40 This structure is equivalent to the frequently used $\alpha\beta'$ decomposition in cointegration analysis; see e.g. Lütkepohl (2005: Ch. 6.3).

of order 0). In a small equation system, it suffices to determine the rank of Π (or H, respectively) since $\mathrm{rk}(\Pi)$ equals the number of cointegrating relations. However, we have to pursue another strategy since we are not able to estimate Π directly.

We have seen before that the entire structure of a GVAR model can be understood as a factor model. In these, joint stochastic trends can be generated by nonstationary factors. However, nonstationarity can even be inherent in the idiosyncratic components. This can be crucial for the identification of the factor space and the number of factors (in particular for the stationary factors). In addition, nonstationary idiosyncratic components may forestall the existence of stationary combinations of series although the series have a common stochastic trend, thus eliminating cointegration: $y_{i,m,t} - \beta y_{i',m,t} = \left(\theta_{i,m} - \beta\theta_{i',m}\right) f_t + \upsilon_{i,m,t} - \beta\upsilon_{i',m,t}$ and $(\beta = \left(\theta_{i,m}\right)^{-1}\theta_{i',m})$ can be stationary only if both $\upsilon_{i,m,t}$ and $\upsilon_{i',m,t}$ are stationary. Bai/Ng (2004) argue that a factor model in first differences allows to track nonstationarity in the components:

$$\Delta y_{i,m,t} = \theta_{i,m}\Delta f_t + \Delta\upsilon_{i,m,t} \tag{5.15}$$

with $\Delta\chi_{i,m,t} = \theta_{i,m}\Delta f_t$ the difference of the common component, f_t the vector of factors and $\theta_{i,m}$ the loadings of the factors corresponding to variable m in region i. For $I(1)$ variables $y_{i,m,t}$, the model in differences is stationary, and results from standard factor analysis become applicable. For any number of factors r, $\Delta f_{(r)t}$ and $\Delta\upsilon_{(r)i,m,t}$ can be retained from principal component analysis and used further in the *Panel Analysis of Nonstationarity in Idiosyncratic and Common Components* (PANIC) proposed by Bai/Ng (2004). In this analysis, the first (strongest) factor $f_{(1)t} = f_{(1)0} + \sum_{\tau=1}^{t}\Delta f_{(1)\tau}$ and all $\upsilon_{(r)i,m,t}$ are tested for unit roots by ADF-test, and the number of nonstationary common components can be determined from the entire system of factors by the MQ_c statistics. We augment their procedure by testing the first factor for HEGY-type seasonal unit roots and unit roots with structural breaks in addition to the ADF test. Results for the *PANIC* tests (without linear trend) and for the determined number of factors are provided in Table 5.3.

The presented results suggest that, given the existence of at least one common factor, nonstationary factors affect both log unemployment and log employment. Nonstationarity of the first factor can not be rejected at reasonable significance levels not only in the reported Augmented Dickey-Fuller tests but also if we account for a structural break in January 2005 or in February 2007, the latter suggested as a break date of the unemployment factor by the Zivot/Andrews (1992) unit-root test. According to the MQ_c statistic, we can reject nonstationarity of more factors in the entire system than in the employment subsystem. All in all, the occurrence of more than one nonstationary common factor can be rejected in most specifications.

Table 5.3: Nonstationarity in idiosyncratic and common components

	Full system (Y_t)		Unemployment (U_t)		Employment (L_t)	
	2005	2011	2005	2011	2005	2011
Unfiltered series						
# factors, PCA	3	3	2	2	2	2
First factor $I(1)$ (ADF)	Yes	Yes	Yes	Yes	Yes	Yes
# $I(1)$ factors (MQ_c)	1	2	1	1	2	2
# $I(1)$ idios. comp.	13	13	8	4	5	6
Series filtered for deterministic seasonal figure (season dummies)						
# factors, PCA	3	3	2	1	2	2
First factor $I(1)$ (ADF)	Yes	Yes	Yes	Yes	Yes	Yes
# $I(1)$ factors (MQ_c)	1	1	1	1	2	2
# $I(1)$ idios. comp.	17	17	9	7	6	6
Cross-sec. dimension (N)	20	20	10	10	10	10

All numbers presented refer to unit roots rejected at the 90% confidence level. Tests are carried out without linear trend.

When it comes to the idiosyncratic components, we can reject nonstationarity at the 90% confidence level of ADF tests only in seven of the twenty investigated untransformed series; their majority seems to have $I(1)$ idiosyncratic components. If we additionally control for regular seasonal patterns (before decomposing the series in the common and the idiosyncratic components), evidence for nonstationary idiosyncratic components becomes even stronger as shown in the lower panel of Table 5.3. Hence, because most often one or two (but up to three) nonstationary components are inherent in each combination, it is unlikely that we find a stationary linear combination of the series.

To ascertain this, we investigate the existence of cointegration in small subsystems, that is in pairwise relations between two series, and in the relation between regional series and the corresponding local average. We split system (5.5) in small subsystems and determine the rank of the subsystem's elements of Π; the number of cointegrating relations equals the matrix rank (see Johansen, 1991, 1995). Results for some tests, using Johansen's trace testing procedure and critical values at the 95% level, are reported in Table 5.4.

The second and third column in Table 5.4 report the cointegration rank of a bivariate intra-region VEC, entailing only unemployment and employment. Columns four to seven show, across the regions, the number of pairwise cointegration relations of the same variable in two regions; the maximum of possible relations would be nine (if the series in a region cointegrates with the series in every other region). Columns eight to eleven refer to cointegration between a series in region i

and the corresponding local average. Interestingly, we find hardly any evidence for cointegration between unemployment and employment inside a region. Cointegration between pairs of regional unemployment, or between unemployment in region i and unemployment averaged over the surrounding regions is not supported in general. In contrast, employment seems to cointegrate across some regions, that is to form stationary linear combinations driven by the same trend. For the sampling period ending in December 2005, we have joint trends between Rhineland-Palatine/Saarland and both Baden-Wurttemberg and Hesse (two of its neighbours), and between Lower Saxony/Bremen and North Rhine-Westphalia. The cointegrating relations change with the sampling period: with data ending in 2011, we find Berlin/Brandenburg (in the East) cointegrating with Bavaria (in the South-East), Rhineland-Palatinate/ Saarland (in the South-West) and North, but no evidence for cointegration with the Eastern German regions Saxony and Saxony-Anhalt/Thuringia. When analysing further periods, evidence for some of these relations vanishes, whereas cointegration relations between other pairs of regions become significant.

Table 5.4: Number of cointegrating relations per regional subsystem

RD	U_i, L_i		U_i, U_j		L_i, L_j		U_i, U_i^*		L_i, L_i^*	
	2005	2011	2005	2011	2005	2011	2005	2011	2005	2011
Nord	0	.	0	0	0	1	0	0	0	.
BB	0	.	0	0	0	3	0	0	0	0
SAT	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0
BY	0	0	0	1	0	1	1	0	0	0
BW	0	0	0	0	1	0	0	0	1	0
RPS	.	0	1	0	2	2	1	0	.	1
H	.	0	1	1	1	1	.	0	0	0
NRW	0	0	0	0	1	0	0	0	0	0
NSB	0	0	0	0	1	0	0	0	1	0

Reported results refer to the period from January 1996 to December 2005 and 2011, respectively. Tables using the 99% critical values and the Hannan-Quinn Information Criterium, results over other sampling periods (till Dec. 2007 and 2009) and for cointegration in three-region systems (instead of the two-region systems in columns 4-7) are available from the corresponding author.

A dot represents a full rank of the subsystems matrix Π. This implies stationarity of all series in the subsystem which is neither consistent with the other results on cointegration (presented in the same line) nor with the unit-root tests in Table 5.1.

To put the previous findings in a nutshell: we have shown first that the series are unlikely to form stationary linear combinations across the regions and that we cannot employ a cointegration relation. Then, a full model in VEC form, i.e.

including the series in lagged log-levels, will be misspecified with a (supposedly insignificant) $I(1)$ term on the right hand. Probably, a system in first differences (i.e. a VEC under the explicit restriction that all elements in Π are zero) produces more accurate forecasts.[41] Second, we have found only evidence for semi-strong cross-sectional dependence.

5.5 Selection and inclusion of indicators

This section focuses on the appropriate determination of the component $\zeta_{i,t}$ in equation system (5.11). It entails seasonal dummies, a dummy for the pre-2005 period (to account for the important labour-market reforms), and additionally the information provided by leading indicators. Often it is argued that the inclusion of a small number of indicators with a high information content performs better in forecasting than a larger number of indicators with less information (see e.g. Stock, 2001 or, especially for Germany, Gaggermeier, 2006).

To approximate the business cycle expectations, we use a set of publicly available national indicators: series of the Stock Market Index (DAX, at the end of a month), and the Wholesale Index provided by the German Federal Bank[42]; in particular the value of the major German enterprises and the sales within the economy can be considered as easily observable metrics regarding economic prospects. Alternatively, we use judgemental indicators regarding business situation and expectations (two indicators gained from a survey amongst financial experts, provided by the ZEW Centre for European Economic Research Mannheim[43]; two from a management survey, provided by the ifo institute Munich[44]).

In addition, we test the information content of some labour market series on vacancies and participants in a number of active labour-market policy (ALMP) programmes; the series are register data collected by the German Federal Employment Agency (FEA) at the same regional level as the employment and unemployment series. The metric of vacancies covers all job offers that are reported to the FEA. It underestimates the real number of job offers, though it may still serve as a prospective indicator. The measure for ALMP participation

41 Regarding the problem of uncertain cointegration when forecasting, Stock (2001: p. 578) argues: "However, even if cointegration is correctly imposed, it remains to estimate the parameters of the cointegrating vector, which are, to first-order, estimated consistently (and at the same rate) if cointegration is not imposed. If cointegration is imposed incorrectly, however, asymptotically biased forecasts with large risks can be produced."

42 See http://www.bundesbank.de/statistik/statistik_zeitreihen.php, series BBK01.WU3140 (DAX) and M.DE.N.I.IT2.ACM01.V.I (Wholesale Index); accessed last 03.12.2012.

43 See ftp://ftp.zew.de/pub/zew-docs/div/konjunktur.xls (accessed last 30.11.2012).

44 The series Geschäftsbeurteilungen (R5) and Geschäftserwartungen (R6) are available at http://www.cesifo-group.de/ifoHome/facts/Time-series-and-Diagrams/Zeitreihen/Reihen-Geschaeftsklima-Deutschland.html, last accessed 30.11.2012.

includes participants in job-training schemes and other programmes for which participants are counted neither as employed nor as unemployed. Thus, persons benefitting from subsidies employment are not included in this metric. ALMP programmes reduce the reported number of unemployed – during the programme, participants have reduced search activity and often are not able to take on a job instantaneously. Thus, approximately contemporaneous numbers on ALMP may reduce unexplained fluctuations in unemployment and help fitting the model. As well, it may be that unemployed persons benefit more from wage subsidies during a cycles upturn (or an expanding labour market), and relatively more from training programmes and their long-run effects during a economic contraction. Shifts in the number of training schemes may be related with the business cycle.

To measure the effect of climate (which seems interesting in the light of unemployment's development during the warm winters 2006 and 2007), we use a set of publicly available metrics on temperature, sun-shine, wind force and precipitation collected by the German Climate Service (Deutscher Wetterdienst) at 40 stations all over Germany[45]. We use minimum Temperature within a month (tnn), cloud amount (nmm), total monthly precipitation (rss) and average windforce (fmm). The climate indicators are averaged over those stations located within the territory of each RD to receive the region-specific value.

Most series are available over the whole sampling period starting in January 1996 (or even before). An exception are the metrics on vacancies and ALMP participants firstly reported in the FEA data in January 2000. Each indicator becomes available at the same (or with less) delay as the target series. In a number of unit-root tests analogously to those provided for the target variables in Table 5.1, the climate variables all show up to be stationary, whereas business-cycle indicators as well as ALMP participation and vacancies supposedly contain unit roots. We account for this by analysing and including the non-stationary leading indicators in first differences.

Uncertainty in the h-step ahead predictor $\hat{\xi}_{i,T+h} = E\left(x_{i,T+h|T}\right)\hat{\beta}$ stems from two sources, uncertainty about the value of the indicator $x_{i,t+h}$ and uncertainty about the relationship between $x_{i,t+h}$ and $y_{i,t+h}$ which is described by the parameter vector β. A good indicator is, on the one hand, significantly correlated with the variable of interest. Here, the temporal lead of an indicator is crucial: the same indicator variable may show high correlation with the target variable at a certain lead, and weak correlation at other leads. On the other hand, it should have a certain temporal lead to the variables of interest, such that the relevant

45 See http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=_dwdwww_klima_umwelt_klimadaten_deutschland, last accessed 30.11.2012.

observations of the indicator have realized already or, to be more precise, are observed with sufficient accuracy (with little measurement error) in the period when the forecast is made. Inaccurate observation of the indicator and data underlying major revisions increase the uncertainty in the model (for the discussion on data revisions and real-time forecasting see e.g. Jacobs/van Norden, 2011). Hence, the forecast variance will be smaller if the indicator's values are known. Nevertheless, the time delay between indicator and target variable should not be too large to be economically reasonable.

To restrict the number of relations tested out, we determine for each indicator the correlation with log employment and, respectively, log unemployment shown at any lead between zero and thirty-six months (or, due to the shorter observation period, twenty-four months for ALMP and Vacancies). Ordering the correlation according to their absolute values allows us to determine the optimum lead, i.e. the time delay for which the mostly significant relation with regional employment or unemployment can be expected. In general, there is no clear timing for the peak in the degree of correlation between two variables. Thus, we determine for each indicator-(un)employment-region combination the three leads with the highest correlation, which are shown in Table 5.5.

Table 5.5: Lead of the indicator variables, in months

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Leads related with Δ unemployment										
Δ dax	29, 17, 5	28, 5, 29	4, 5, 16	29, 5, 4	28, 4, 5	5, 21, 11	5, 29, 28	5, 29, 17	28, 29, 16	28, 5, 29
Δ wholesale	36, 12, 24	36, 12, 24	36, 12, 24	36, 12, 0	36, 12, 24	36, 12, 24	36, 24, 12	36, 24, 12	36, 24, 12	36, 12, 24
Δ ifo-sit.	5, 7, 4	30, 28, 11	29, 30, 1	29, 30, 1	7, 28, 4	7, 4, 5	5, 29, 7	28, 5, 7	29, 7, 4	30, 28, 29
Δ ifo-exp.	36, 8, 0	0, 24, 36	0, 36, 24	24, 36, 0	0, 2, 4	0, 9, 8	36, 24, 0	36, 24, 0	36, 26, 0	26, 29, 2
Δ zew-sit.	10, 11, 3	10, 22, 11	34, 11, 23	34, 11, 10	10, 11, 22	5, 10, 11	10, 29, 11	10, 29, 11	10, 11, 29	10, 23, 22
Δ zew-exp.	27, 0, 15	27, 26, 36	27, 15, 3	3, 15, 27	14, 15, 27	0, 14, 15	27, 26, 15	27, 15, 0	26, 27, 2	26, 27, 2
Δ almp	0, 12, 1	0, 12, 16	4, 0, 16	1, 4, 16	0, 12, 24	12, 24, 0	0, 12, 24	0, 12, 6	0, 6, 12	7, 0, 6
Δ vacancies	3, 15, 2	11, 10, 23	2, 14, 10	2, 14, 1	2, 14, 1	2, 3, 15	2, 3, 11	11, 23, 2	11, 23, 2	2, 11, 10
fmm	7, 4, 6	31, 36, 35	31, 30, 29	36, 35, 25	5, 4, 7	4, 3, 5	4, 3, 5	25, 23, 24	3, 4, 2	23, 33, 25
nmm	24, 25, 26	3, 0, 2	32, 35, 31	0, 2, 3	2, 3, 4	26, 25, 27	25, 24, 23	25, 24, 26	25, 24, 26	3, 2, 4
rss	11, 10, 9	36, 35, 25	0, 8, 7	18, 19, 20	10, 11, 36	19, 36, 20	8, 10, 5	8, 19, 18	11, 10, 19	23, 28, 25
tnn	12, 11, 17	27, 28, 26	8, 9, 20	31, 20, 32	0, 10, 9	10, 9, 8	10, 11, 9	18, 11, 10	10, 11, 22	27, 28, 26
Leads related with Δ employment										
Δ dax	36, 16, 4	29, 5, 36	16, 29, 4	16, 5, 29	29, 5, 16	36, 29, 17	16, 36, 4	36, 16, 4	16, 28, 36	16, 28, 4
Δ wholesale	7, 19, 31	23, 11, 35	35, 23, 11	23, 35, 11	23, 35, 11	32, 20, 8	19, 7, 31	19, 7, 31	19, 7, 31	23, 11, 35
Δ ifo-sit.	34, 22, 36	34, 29, 11	29, 25, 34	29, 4, 35	28, 11, 34	10, 4, 16	36, 34, 22	34, 22, 29	29, 36, 2	28, 22, 29
Δ ifo-exp.	32, 24, 19	24, 0, 36	0, 14, 36	0, 24, 32	32, 0, 8	0, 8, 32	32, 11, 15	32, 24, 0	32, 17, 14	32, 0, 14
Δ zew-sit.	14, 11, 15	34, 10, 11	34, 30, 10	34, 10, 11	10, 9, 34	12, 10, 31	30, 11, 9	34, 11, 15	34, 23, 11	34, 23, 11
Δ zew-exp.	10, 23, 22	26, 22, 27	26, 27, 15	26, 15, 14	26, 14, 15	23, 11, 0	22, 10, 26	10, 22, 23	26, 14, 22	26, 14, 15
Δ almp	12, 0, 24	0, 12, 24	0, 12, 24	0, 12, 24	21, 24, 12	24, 12, 0	12, 0, 24	0, 12, 24	6, 18, 12	6, 18, 0
Δ vacancies	22, 23, 10	22, 10, 23	2, 10, 22	2, 14, 1	22, 10, 21	7, 19, 22	10, 22, 23	23, 22, 10	23, 11, 10	22, 10, 23
fmm	11, 13, 10	8, 9, 7	8, 9, 7	27, 3, 15	2, 13, 3	2, 3, 1	3, 2, 4	25, 26, 28	25, 26, 28	25, 21, 24
nmm	21, 17, 16	21, 22, 9	9, 8, 10	9, 5, 33	15, 14, 26	26, 27, 15	25, 26, 24	26, 25, 14	25, 21, 26	21, 22, 14
rss	11, 12, 13	8, 7, 6	32, 33, 31	7, 8, 6	8, 10, 9	8, 7, 9	8, 11, 10	8, 11, 9	8, 25, 26	36, 21, 25
tnn	10, 8, 11	33, 32, 34	33, 32, 34	33, 32, 34	9, 8, 10	9, 8, 10	10, 9, 8	10, 9, 19	19, 28, 25	33, 34, 32

Lead ordered by correlation (in absolute value, declining) of the indicator with the target series.

5.6 Forecast evaluation

5.6.1 Forecast construction and evaluation method

We evaluate out-of-sample forecasts at the three- and twelve-months horizon, with the end of the sampling window rolling from July 2005 to December 2010. I.e., the first three-month forecast is made for October 2005, the last twelve-month forecast predicts the labour market quantities in December 2011. In each estimation we include just a single indicator at one lead. The coefficients are estimated region by region with a GVAR model in differences (we rejected cointegration, thus the coefficients in H are restricted to zero) without leading region (we found only evidence for semi-strong cross-sectional dependence); the coefficients are then inserted in the full model (containing all regions) for forecasting. h-step ahead forecasts are received recursively by accumulating one-step ahead forecasts for the monthly differences starting at the last episode in the estimation sample, τ,

$$E\left\{\Delta\mathbf{Y}_{\tau+h}\right\} = \hat{G}^h \Delta\mathbf{Y}_\tau + \sum_{\eta=1}^{h} \hat{G}^{(h-\eta)} \hat{\boldsymbol{\Xi}}_{\tau+\eta} \tag{5.16}$$

which for log unemployment and employment levels results in forecasts

$$\begin{aligned} E\left\{\mathbf{Y}_{\tau+H}\right\} &= Y_\tau + \sum_{h=1}^{H} E\left\{\Delta\mathbf{Y}_{\tau+h}\right\} \\ &= Y_\tau + \sum_{h=1}^{H} \hat{G}^h \Delta\mathbf{Y}_\tau + \sum_{h=1}^{H}\sum_{\eta=1}^{h} \hat{G}^{(h-\eta)} \hat{\boldsymbol{\Xi}}_{\tau+\eta} \end{aligned} \tag{5.17}$$

Thus, with 66 different sampling periods (with last 'observed' period τ) and 6 leads (indexed with $\ell \in \{1,\ldots,l\}$) per indicator j, we carry out 396 estimations (and predictions) per indicator. Error measures averaged over these 398 different GVAR predictions are reported separately per regional FEA division i, target variable m, indicator j and horizon $h = \{3,12\}$ in Table 5.6; identical measures for comparison models without spatial interdependence are displayed in Tables 5.7 and 5.8. Whereas we have used logs in the estimation, we let the reported values refer to the forecast error of the variables in levels: i.e., $e_{m,i,\tau+h,j,\ell} = \exp(E\{y_{m,i,\tau+h} \mid \mathcal{I}(j,\ell)_{i,\tau}\}) - \exp(y_{m,i,\tau+h})$ denotes the h-periods ahead forecast error for variable m in region i with sample end τ which is yielded with indicator j at lead ℓ [that is, with information $\mathcal{I}(j,\ell)$]. Using this, we define the Mean Average Percentage Forecast Error (MAPFE) as

$$MAPFE_{m,i,j,h} = \frac{1}{lT}\sum_{\ell=1}^{l}\sum_{\tau=1}^{T}\frac{|e_{m,i,\tau+h,j,\ell})|}{\exp\left(y_{m,i,\tau+h}\right)} \tag{5.18}$$

We estimate and predict also a GVAR model without leading indicator, i.e. a model where the component Ξ_t entails only seasonal dummies and a break dummy to control for the pre-2005 period. In addition to the forecast errors referring to the GVAR forecasts, we report the MAPFEs resulting from the corresponding single-region VAR models without local average or other forms of spatial linkages. This enables us to identify the prospective information provided by the indicators and the forecast contribution achievable through considering spatial interdependence. To test the significance of differences in the performance between two indicators j_1 and j_2, we employ a panel version of the Diebold-Mariano (DM) test similar to the one presented by Bernoth/Pick (2011). However, the presented results rely on the Mean Absolute Error rather than the Root Mean Squared Error used in their analysis. With $d_\tau = \left(\left|e_{m,i,\tau+h,j_2,\ell_2}\right| - \left|e_{m,i,\tau+h,j_1,\ell_1}\right|\right)$ as the difference between the absolute errors, the individual DM test is defined as

$$DM(\mathcal{I}[j_1,\ell_1],\mathcal{I}[j_2,\ell_2])_{i,m,h} = \frac{\frac{1}{T}\sum_{\tau=1}^{T} d_\tau}{\sqrt{\frac{2\pi}{T}\sum_{\theta=-\infty}^{\infty}(d_\tau - \bar{d})(d_{\tau+\theta} - \bar{d})}} \tag{5.19}$$

which will take on negative values if indicator j_2 at lead ℓ_2 yields on average smaller forecast errors than indicator j_1 at lead ℓ_1. The individual DM statistic has a standard normal limiting distribution, and so does the panel statistics as well:

$$\overline{DM}(\mathcal{I}[j_1],\mathcal{I}[j_2])_{m,h} = \frac{1}{\sqrt{Nl}}\sum_{i=1}^{N}\sum_{\ell=1}^{l} DM(\mathcal{I}[j_1,\ell_1],\mathcal{I}[j_2,\ell_2])_{i,m,h} \sim \mathcal{N}(0,1) \tag{5.20}$$

5.6.2 General forecast performance and the contribution of spatial information

We can see from Tables 5.6, 5.7 and 5.8 that the twelve-month forecast error is between 2.5 and 4.5 times the forecast error at three-month horizon; the average/median multiplier is between 3.5 and 4. The almost linear increase of the error's size with the horizon can be expected because of the unit root (which implies persistence of shocks), see e.g. Dickey/Bell/Miller (1986) or Lütkepohl (2005): shocks are persistent, and thus the probabilistic uncertainty accumulates. Employment forecasts have, with exception of Berlin/Brandenburg (BB), approximately equal performance across the three Tables. In BB, the VAR estimation produces significantly smaller forecast errors than the models using differences on the left hand. For unemployment, in contrast, the VAR in levels predicts in general worse than the VAR in differences and the GVAR; on the twelve months horizon, forecast errors are approximately 1-1.5 percentage points (0.01-0.015) higher throughout most regions and indicators if integration of the series is not accounted for.

Table 5.6: Mean Absolute Percentage Forecast Error, GVAR (in differences)

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Unemployment; 3-month horizon										
none	0.0330	0.0252	0.0373	0.0355	0.0541	0.0378	0.0331	0.0308	0.0240	0.0307
dax	0.0330	0.0254	0.0371	0.0354	0.0540	0.0378	0.0332	0.0310	0.0242	0.0306
wholesale	0.0330	0.0252	0.0373	0.0355	0.0540	0.0381	0.0330	0.0309	0.0241	0.0309
ifo sit.	0.0330	0.0256	0.0377	0.0358	0.0533	0.0374	0.0332	0.0308	0.0238	0.0311
ifo exp.	0.0335	0.0254	0.0378	0.0358	0.0543	0.0377	0.0333	0.0312	0.0244	0.0316
zew sit.	0.0326	0.0251	0.0379	0.0356	0.0527	0.0366	0.0329	0.0304	0.0239	0.0312
zew exp.	0.0341	0.0259	0.0380	0.0355	0.0541	0.0380	0.0335	0.0310	0.0246	0.0316
almp	0.0290	0.0226	0.0354	0.0311	0.0471	0.0397	0.0329	0.0310	0.0256	0.0320
vacancies	0.0278	0.0223	0.0347	0.0308	0.0456	0.0368	0.0333	0.0291	0.0229	0.0319
fmm	0.0334	0.0257	0.0372	0.0351	0.0545	0.0384	0.0334	0.0311	0.0242	0.0306
nmm	0.0335	0.0250	0.0369	0.0353	0.0540	0.0379	0.0332	0.0308	0.0240	0.0310
rss	0.0329	0.0251	0.0376	0.0350	0.0542	0.0369	0.0325	0.0306	0.0241	0.0311
tnn	0.0332	0.0254	0.0378	0.0358	0.0540	0.0371	0.0332	0.0310	0.0241	0.0311
Unemployment; 12-month horizon										
none	0.0847	0.0734	0.0956	0.1007	0.1688	0.1622	0.1067	0.1045	0.0927	0.0929
dax	0.0846	0.0738	0.0955	0.1015	0.1687	0.1618	0.1065	0.1047	0.0933	0.0941
wholesale	0.0851	0.0741	0.0960	0.1013	0.1690	0.1620	0.1069	0.1046	0.0927	0.0937
ifo sit.	0.0828	0.0753	0.0948	0.0987	0.1593	0.1543	0.1032	0.1012	0.0873	0.0881
ifo exp.	0.0882	0.0760	0.0968	0.1024	0.1694	0.1579	0.1058	0.1044	0.0921	0.0954
zew sit.	0.0807	0.0770	0.0923	0.0938	0.1436	0.1399	0.0961	0.0942	0.0842	0.0850
zew exp.	0.0867	0.0765	0.0969	0.1021	0.1702	0.1620	0.1066	0.1044	0.0925	0.0938
almp	0.0908	0.0643	0.0794	0.0817	0.1863	0.1907	0.1264	0.1282	0.1127	0.1059
vacancies	0.0852	0.0625	0.0758	0.0770	0.1763	0.1805	0.1258	0.1170	0.0989	0.0866
fmm	0.0868	0.0752	0.0973	0.1018	0.1712	0.1658	0.1090	0.1078	0.0947	0.0933
nmm	0.0864	0.0700	0.0928	0.1007	0.1683	0.1613	0.1061	0.1044	0.0923	0.0918
rss	0.0825	0.0730	0.0958	0.1010	0.1659	0.1516	0.1007	0.1004	0.0908	0.0926
tnn	0.0859	0.0745	0.0952	0.0997	0.1668	0.1562	0.1055	0.1035	0.0929	0.0936

Table 5.6 continued

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Employment; 3-month horizon										
none	0.0044	0.0061	0.0067	0.0067	0.0036	0.0031	0.0032	0.0030	0.0038	0.0042
dax	0.0044	0.0061	0.0067	0.0067	0.0036	0.0031	0.0032	0.0031	0.0038	0.0042
wholesale	0.0044	0.0060	0.0067	0.0066	0.0036	0.0031	0.0032	0.0031	0.0038	0.0042
ifo sit.	0.0044	0.0063	0.0070	0.0068	0.0036	0.0030	0.0032	0.0030	0.0036	0.0043
ifo exp.	0.0045	0.0062	0.0068	0.0068	0.0036	0.0031	0.0032	0.0031	0.0038	0.0042
zew sit.	0.0042	0.0062	0.0070	0.0069	0.0036	0.0030	0.0031	0.0029	0.0035	0.0044
zew exp.	0.0045	0.0063	0.0070	0.0068	0.0037	0.0031	0.0033	0.0031	0.0038	0.0044
almp	0.0041	0.0055	0.0068	0.0066	0.0045	0.0034	0.0039	0.0035	0.0049	0.0044
vacancies	0.0039	0.0054	0.0068	0.0063	0.0040	0.0034	0.0037	0.0037	0.0044	0.0042
fmm	0.0044	0.0061	0.0068	0.0067	0.0037	0.0031	0.0034	0.0032	0.0038	0.0043
nmm	0.0043	0.0060	0.0067	0.0067	0.0036	0.0031	0.0032	0.0031	0.0038	0.0043
rss	0.0043	0.0061	0.0068	0.0067	0.0036	0.0030	0.0031	0.0030	0.0038	0.0043
tnn	0.0044	0.0061	0.0068	0.0067	0.0036	0.0029	0.0032	0.0030	0.0038	0.0043
Employment; 12-month horizon										
none	0.0178	0.0299	0.0254	0.0266	0.0140	0.0131	0.0137	0.0121	0.0155	0.0157
dax	0.0179	0.0301	0.0254	0.0267	0.0140	0.0132	0.0138	0.0121	0.0155	0.0159
wholesale	0.0177	0.0299	0.0251	0.0264	0.0139	0.0131	0.0136	0.0120	0.0155	0.0156
ifo sit.	0.0180	0.0308	0.0256	0.0270	0.0132	0.0125	0.0128	0.0113	0.0145	0.0154
ifo exp.	0.0181	0.0305	0.0258	0.0271	0.0139	0.0130	0.0137	0.0123	0.0155	0.0161
zew sit.	0.0175	0.0303	0.0252	0.0268	0.0113	0.0111	0.0118	0.0099	0.0134	0.0152
zew exp.	0.0180	0.0304	0.0258	0.0269	0.0138	0.0131	0.0137	0.0123	0.0156	0.0159
almp	0.0189	0.0279	0.0242	0.0255	0.0188	0.0175	0.0182	0.0167	0.0344	0.0195
vacancies	0.0184	0.0271	0.0232	0.0249	0.0171	0.0169	0.0170	0.0168	0.0196	0.0190
fmm	0.0187	0.0301	0.0257	0.0268	0.0142	0.0136	0.0143	0.0125	0.0156	0.0158
nmm	0.0180	0.0293	0.0251	0.0265	0.0140	0.0131	0.0137	0.0123	0.0155	0.0155
rss	0.0179	0.0297	0.0255	0.0267	0.0136	0.0119	0.0129	0.0116	0.0154	0.0157
tnn	0.0179	0.0302	0.0257	0.0268	0.0135	0.0121	0.0135	0.0119	0.0155	0.0155

Scale: 1.0 = 100%.

Table 5.7: Mean Absolute Percentage Forecast Error, VAR (not differenced)

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Unemployment; 3-month horizon										
none	0.0381	0.0274	0.0390	0.0367	0.0605	0.0391	0.0377	0.0336	0.0253	0.0359
dax	0.0379	0.0280	0.0387	0.0378	0.0604	0.0393	0.0376	0.0341	0.0276	0.0363
wholesale	0.0381	0.0277	0.0393	0.0371	0.0597	0.0399	0.0377	0.0342	0.0263	0.0362
ifo sit.	0.0385	0.0291	0.0386	0.0376	0.0609	0.0388	0.0383	0.0328	0.0245	0.0358
ifo exp.	0.0387	0.0276	0.0408	0.0379	0.0618	0.0398	0.0387	0.0308	0.0242	0.0368
zew sit.	0.0364	0.0278	0.0397	0.0365	0.0615	0.0387	0.0392	0.0350	0.0245	0.0371
zew exp.	0.0404	0.0290	0.0401	0.0368	0.0614	0.0400	0.0374	0.0340	0.0263	0.0360
almp	0.0288	0.0222	0.0363	0.0337	0.0475	0.0390	0.0274	0.0326	0.0259	0.0380
vacancies	0.0330	0.0244	0.0373	0.0356	0.0520	0.0420	0.0278	0.0373	0.0286	0.0403
fmm	0.0382	0.0274	0.0387	0.0364	0.0608	0.0392	0.0381	0.0341	0.0258	0.0361
nmm	0.0381	0.0273	0.0392	0.0369	0.0608	0.0394	0.0379	0.0338	0.0253	0.0360
rss	0.0383	0.0275	0.0392	0.0368	0.0607	0.0392	0.0375	0.0339	0.0254	0.0360
tnn	0.0380	0.0276	0.0392	0.0370	0.0600	0.0391	0.0377	0.0334	0.0257	0.0364
Unemployment; 12-month horizon										
none	0.1108	0.0795	0.1114	0.1146	0.1965	0.1800	0.1226	0.1047	0.0949	0.1070
dax	0.1120	0.0833	0.1094	0.1206	0.1844	0.1728	0.1234	0.1050	0.1049	0.1083
wholesale	0.1110	0.0819	0.1137	0.1175	0.1923	0.1871	0.1220	0.1082	0.0968	0.1069
ifo sit.	0.1135	0.0920	0.1130	0.1216	0.1891	0.1661	0.1201	0.0984	0.0897	0.1034
ifo exp.	0.1131	0.0816	0.1184	0.1194	0.2018	0.1852	0.1246	0.0964	0.0854	0.1114
zew sit.	0.1110	0.0823	0.1129	0.1144	0.1755	0.1609	0.1226	0.1023	0.0904	0.1104
zew exp.	0.1171	0.0885	0.1176	0.1162	0.2004	0.1851	0.1258	0.1106	0.0996	0.1014
almp	0.0695	0.0580	0.0879	0.0947	0.1804	0.1922	0.0921	0.1136	0.0943	0.1205
vacancies	0.0964	0.0871	0.1052	0.1078	0.1813	0.2115	0.0785	0.1076	0.1065	0.1381
fmm	0.1125	0.0793	0.1089	0.1114	0.1979	0.1823	0.1247	0.1071	0.0960	0.1075
nmm	0.1117	0.0785	0.1112	0.1145	0.1985	0.1804	0.1228	0.1048	0.0943	0.1070
rss	0.1107	0.0799	0.1123	0.1141	0.1956	0.1780	0.1217	0.1067	0.0954	0.1074
tnn	0.1088	0.0795	0.1117	0.1161	0.1921	0.1798	0.1228	0.1055	0.0960	0.1086

Table 5.7 continued

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Employment; 3-month horizon										
none	0.0046	0.0047	0.0079	0.0077	0.0043	0.0035	0.0040	0.0035	0.0039	0.0050
dax	0.0046	0.0048	0.0074	0.0075	0.0049	0.0039	0.0045	0.0042	0.0038	0.0056
wholesale	0.0045	0.0048	0.0078	0.0076	0.0043	0.0035	0.0040	0.0035	0.0038	0.0049
ifo sit.	0.0047	0.0049	0.0077	0.0075	0.0048	0.0039	0.0041	0.0036	0.0042	0.0052
ifo exp.	0.0046	0.0049	0.0080	0.0077	0.0045	0.0036	0.0040	0.0030	0.0033	0.0052
zew sit.	0.0048	0.0048	0.0078	0.0078	0.0048	0.0036	0.0045	0.0042	0.0042	0.0055
zew exp.	0.0047	0.0050	0.0083	0.0078	0.0045	0.0036	0.0041	0.0035	0.0039	0.0054
almp	0.0038	0.0043	0.0071	0.0071	0.0048	0.0038	0.0044	0.0039	0.0038	0.0045
vacancies	0.0038	0.0045	0.0072	0.0069	0.0045	0.0040	0.0039	0.0041	0.0042	0.0048
fmm	0.0045	0.0047	0.0077	0.0077	0.0044	0.0035	0.0040	0.0036	0.0039	0.0051
nmm	0.0046	0.0048	0.0080	0.0077	0.0043	0.0035	0.0040	0.0036	0.0039	0.0051
rss	0.0046	0.0047	0.0079	0.0077	0.0043	0.0035	0.0040	0.0035	0.0039	0.0050
tnn	0.0044	0.0047	0.0078	0.0076	0.0043	0.0035	0.0040	0.0036	0.0040	0.0050
Employment; 12-month horizon										
none	0.0173	0.0169	0.0228	0.0224	0.0194	0.0185	0.0157	0.0137	0.0165	0.0171
dax	0.0177	0.0172	0.0218	0.0229	0.0190	0.0173	0.0155	0.0145	0.0150	0.0171
wholesale	0.0169	0.0173	0.0231	0.0227	0.0186	0.0185	0.0152	0.0134	0.0160	0.0164
ifo sit.	0.0189	0.0186	0.0229	0.0223	0.0202	0.0188	0.0161	0.0140	0.0180	0.0178
ifo exp.	0.0170	0.0171	0.0238	0.0222	0.0189	0.0181	0.0149	0.0111	0.0136	0.0170
zew sit.	0.0172	0.0179	0.0228	0.0230	0.0167	0.0162	0.0147	0.0139	0.0149	0.0177
zew exp.	0.0170	0.0184	0.0241	0.0226	0.0200	0.0185	0.0160	0.0142	0.0171	0.0180
almp	0.0123	0.0173	0.0205	0.0231	0.0207	0.0203	0.0167	0.0166	0.0151	0.0175
vacancies	0.0166	0.0176	0.0196	0.0207	0.0179	0.0194	0.0146	0.0125	0.0141	0.0196
fmm	0.0173	0.0171	0.0220	0.0220	0.0197	0.0186	0.0158	0.0141	0.0164	0.0173
nmm	0.0173	0.0171	0.0229	0.0224	0.0194	0.0186	0.0158	0.0138	0.0165	0.0172
rss	0.0173	0.0172	0.0230	0.0225	0.0193	0.0184	0.0155	0.0137	0.0165	0.0173
tnn	0.0166	0.0168	0.0228	0.0227	0.0187	0.0186	0.0155	0.0139	0.0165	0.0174

Estimated equation: $y_{i,t} = A_1 y_{i,t-1} + A_2 y_{i,t-2} + BX_{i,t} + u_{i,t}$
Scale: 1.0 = 100%.

Table 5.8: Mean Absolute Percentage Forecast Error, VAR in differences (DVAR)

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Unemployment; 3-month horizon										
none	0.0329	0.0257	0.0364	0.0345	0.0563	0.0375	0.0347	0.0325	0.0246	0.0335
dax	0.0328	0.0259	0.0363	0.0345	0.0563	0.0374	0.0348	0.0325	0.0247	0.0332
wholesale	0.0329	0.0257	0.0364	0.0346	0.0563	0.0378	0.0347	0.0326	0.0246	0.0336
ifo sit.	0.0329	0.0261	0.0368	0.0348	0.0560	0.0372	0.0346	0.0323	0.0244	0.0337
ifo exp.	0.0334	0.0261	0.0370	0.0349	0.0568	0.0375	0.0351	0.0330	0.0251	0.0344
zew sit.	0.0327	0.0257	0.0371	0.0344	0.0553	0.0363	0.0341	0.0318	0.0244	0.0337
zew exp.	0.0340	0.0264	0.0372	0.0346	0.0564	0.0377	0.0352	0.0327	0.0249	0.0344
almp	0.0291	0.0227	0.0335	0.0306	0.0520	0.0390	0.0329	0.0349	0.0284	0.0360
vacancies	0.0276	0.0220	0.0331	0.0305	0.0503	0.0364	0.0322	0.0331	0.0252	0.0360
fmm	0.0333	0.0260	0.0363	0.0343	0.0567	0.0381	0.0349	0.0328	0.0248	0.0329
nmm	0.0333	0.0254	0.0358	0.0343	0.0562	0.0376	0.0348	0.0324	0.0248	0.0336
rss	0.0328	0.0256	0.0368	0.0344	0.0565	0.0365	0.0338	0.0324	0.0247	0.0339
tnn	0.0331	0.0259	0.0370	0.0347	0.0563	0.0368	0.0349	0.0326	0.0247	0.0336
Unemployment; 12-month horizon										
none	0.0842	0.0734	0.0964	0.1005	0.1718	0.1618	0.1089	0.1063	0.0948	0.0974
dax	0.0843	0.0741	0.0961	0.1011	0.1712	0.1616	0.1085	0.1063	0.0954	0.0984
wholesale	0.0846	0.0740	0.0969	0.1013	0.1720	0.1616	0.1092	0.1067	0.0948	0.0983
ifo sit.	0.0822	0.0751	0.0953	0.0979	0.1639	0.1558	0.1039	0.1026	0.0891	0.0932
ifo exp.	0.0874	0.0762	0.0977	0.1021	0.1735	0.1569	0.1085	0.1066	0.0938	0.1001
zew sit.	0.0802	0.0763	0.0926	0.0924	0.1488	0.1405	0.0983	0.0956	0.0873	0.0894
zew exp.	0.0859	0.0762	0.0976	0.1019	0.1738	0.1615	0.1091	0.1059	0.0946	0.0985
almp	0.0845	0.0655	0.0818	0.0829	0.1946	0.1914	0.1253	0.1298	0.1163	0.1173
vacancies	0.0805	0.0618	0.0783	0.0788	0.1828	0.1822	0.1181	0.1229	0.1015	0.0999
fmm	0.0861	0.0750	0.0979	0.1022	0.1740	0.1655	0.1108	0.1092	0.0976	0.0968
nmm	0.0846	0.0698	0.0932	0.1000	0.1714	0.1607	0.1087	0.1061	0.0953	0.0962
rss	0.0820	0.0729	0.0970	0.1004	0.1694	0.1504	0.1019	0.1027	0.0949	0.0977
tnn	0.0852	0.0744	0.0954	0.0997	0.1702	0.1549	0.1088	0.1055	0.0957	0.0980

Table 5.8 continued

	Nord	BB	SAT	S	BY	BW	RPS	H	NRW	NSB
Employment; 3-month horizon										
none	0.0043	0.0061	0.0071	0.0073	0.0038	0.0031	0.0041	0.0034	0.0039	0.0048
dax	0.0044	0.0062	0.0070	0.0073	0.0038	0.0031	0.0041	0.0035	0.0039	0.0048
wholesale	0.0044	0.0061	0.0070	0.0073	0.0038	0.0031	0.0041	0.0034	0.0039	0.0048
ifo sit.	0.0044	0.0064	0.0072	0.0074	0.0038	0.0030	0.0038	0.0033	0.0037	0.0048
ifo exp.	0.0045	0.0063	0.0072	0.0074	0.0038	0.0031	0.0041	0.0035	0.0039	0.0049
zew sit.	0.0043	0.0063	0.0073	0.0075	0.0037	0.0029	0.0038	0.0032	0.0036	0.0050
zew exp.	0.0045	0.0063	0.0073	0.0074	0.0039	0.0031	0.0042	0.0035	0.0039	0.0050
almp	0.0041	0.0054	0.0072	0.0072	0.0045	0.0033	0.0049	0.0045	0.0049	0.0051
vacancies	0.0041	0.0053	0.0071	0.0068	0.0041	0.0033	0.0047	0.0045	0.0045	0.0049
fmm	0.0044	0.0062	0.0071	0.0074	0.0039	0.0031	0.0042	0.0035	0.0039	0.0049
nmm	0.0043	0.0060	0.0070	0.0073	0.0038	0.0030	0.0041	0.0035	0.0039	0.0049
rss	0.0043	0.0062	0.0071	0.0074	0.0038	0.0030	0.0039	0.0034	0.0039	0.0049
tnn	0.0044	0.0062	0.0071	0.0074	0.0038	0.0029	0.0041	0.0034	0.0039	0.0049
Employment; 12-month horizon										
none	0.0178	0.0300	0.0257	0.0274	0.0143	0.0133	0.0147	0.0127	0.0158	0.0165
dax	0.0179	0.0302	0.0257	0.0276	0.0143	0.0135	0.0146	0.0126	0.0157	0.0167
wholesale	0.0178	0.0299	0.0254	0.0272	0.0143	0.0133	0.0147	0.0126	0.0158	0.0164
ifo sit.	0.0182	0.0308	0.0259	0.0278	0.0137	0.0128	0.0136	0.0117	0.0147	0.0164
ifo exp.	0.0181	0.0305	0.0262	0.0279	0.0144	0.0132	0.0149	0.0130	0.0157	0.0171
zew sit.	0.0177	0.0304	0.0255	0.0275	0.0120	0.0113	0.0129	0.0102	0.0137	0.0160
zew exp.	0.0181	0.0305	0.0261	0.0277	0.0143	0.0133	0.0147	0.0128	0.0159	0.0168
almp	0.0183	0.0279	0.0250	0.0261	0.0190	0.0171	0.0188	0.0175	0.0213	0.0210
vacancies	0.0181	0.0265	0.0242	0.0260	0.0174	0.0166	0.0184	0.0180	0.0199	0.0202
fmm	0.0187	0.0303	0.0260	0.0275	0.0145	0.0138	0.0152	0.0129	0.0159	0.0166
nmm	0.0178	0.0293	0.0254	0.0275	0.0143	0.0133	0.0148	0.0128	0.0158	0.0163
rss	0.0179	0.0299	0.0258	0.0276	0.0140	0.0120	0.0139	0.0122	0.0158	0.0166
tnn	0.0179	0.0303	0.0259	0.0277	0.0140	0.0123	0.0147	0.0125	0.0158	0.0162

Estimated equation: $\Delta y_{i,t} = A\Delta y_{i,t-1} + BX_{i,t} + u_{i,t}$

This equation is theoretically equivalent to that used for Table 5.7 if $-(I-A_1-A_2)\sim 0$ and $A=-A_2$.

Scale: 1.0 = 100%.

In each Regional Division, the forecast errors for employment and unemployment reflect roughly the same number of persons. The ratio of the relative forecast errors (for unemployment over employment) in a Regional Division is more or less proportional to its unemployment rate. Thus, it doesn't come as a surprise that we predict unemployment with higher relative forecast errors in those regions where the unemployment rate is small: Bavaria (BY) and Baden-Wurttemberg (BW) show at the moment unemployment rates between four and five percentage points. That is, in these two regions a MAPFE of 17% regarding unemployment at the 12-months horizon corresponds to a forecast error of roughly 0.75 percentage points in the unemployment rate. However, this should be sufficient to provide a short illumination of the absolute size of the forecast errors and to allow for a proper judgement of the results; in the following, we focus on discussing the forecast accuracy across the various specifications.

In general, we find an improvement of the GVAR models compared to the corresponding VARs (estimated in both first differences and in levels with an additional lag) with regard to forecast accuracy. For unemployment forecasts at the 3-month horizon, the VAR models in levels (Table 5.7) show smaller MAPFEs than the corresponding GVARs only in six indicator-region combinations (out of 130 tabulated). Here, the average of ratio of the GVAR-related over the VAR-related MAPFE, $\frac{1}{NJ}\sum_{i=1}^{N}\sum_{j\in J}\frac{MAPFE_{GVAR,i,j}}{MAPFE_{VAR,i,j}}$, mounts to 0.914. At the one-year horizon, we find an improvement in the forecast accuracy in 108 of 130 GVAR vs. VAR comparisons. The ratio is on average 0.891; this number means that the forecast error in the unemployment forecasts is reduced on average by roughly ten percent. We find a similar pattern in the short-run employment forecasts where the average GVAR-VAR MAPFE ratio is 0.911, and where we find smaller relative forecast errors for the GVAR in 102 of 130 region-indicator combinations. At the twelve-month horizon we find more accurate forecasts for the GVAR only in 71 of the 130 comparisons, and the GVAR-VAR forecast error ratio exceeds one.

Amongst the employment forecasts at the three-month horizon, the RD of Berlin/Brandenburg (BB) is somewhat outstanding insofar that for each indicator the GVAR is outperformed by the VAR in levels. Here, spatial interdependence seems to bear not only irrelevant but misleading information with regard to employment. This misinformation amplifies over time and affects also the other (partially) Eastern German regions: in BB, the relative loss in accuracy (mounting to 1.708) is at the 12-month horizon two to three times as high as the relative loss at the 3-month horizon (an average ratio of 1.279), and in the RDs Nord, Saxony (S) and Saxony-Anhalt/Thuringia (SAT) the ratios mount at the 12-months horizon to 1.028, 1.079 and 1.026, respectively. That is, we find the GVAR to perform on average neither better nor worse than the level VAR when forecasting employment one year ahead; in the other three scenarios, the improvement of the GVAR is (even statistically) significant.

Forecast accuracy is more similar between the GVAR and the DVAR in Table 5.8; the amount of the improvement in forecast accuracy is obviously smaller. The average MAPFE-ratios mount to 0.979 for unemployment at the three-month horizon, to 0.987 in the predictions for unemployment one-year ahead, to 0.935 for three-months ahead employment forecasts and to 0.978 for employment at the twelve-month horizon. When we count the pairwise comparisons where the GVAR performs better than the DVAR, we find with regard to unemployment forecasts, that the (non-spatial) DVAR yield smaller forecast errors at the three-month horizon in 52 pairwise comparisons and larger errors in 77 comparisons. At the twelve-month horizon, DVARs have a better performance in 41 and a worse in 86 comparisons. The spatial GVAR models tend to predict unemployment more accurate than the DVAR in Western Germany with the exception of Baden-Wurttemberg), and less accurate in Eastern Germany. In forecasting employment, however, the GVAR outperforms the DVAR (again with the exception of Baden-Wurttemberg at the three-month horizon). All in all, accounting for spatial co-development seems to provide some information which can be used for prediction.

5.6.3 Forecast comparison across the indicators

Since we have found evidence for slightly more accurate forecasts of the GVAR model so far, we will focus on these when comparing the forecast content of the indicators. Across all forecasts for which DM panel tests are reported in Table 5.9, a number of trends may be recognized. First, the variation of the forecast errors between the different indicators is rather small. It is hardly possible to identify a clearly superior leading indicator, albeit we have to admit that our analysis might be somewhat 'conservative' against them: the DM test does not account for the nested structure and is thus in favor of the more parsimonious model without indicator (see Clark/West, 2007). At the very short run, no indicator seems significantly superior to the model without indicators with regard to forecasts for the two target variables at the same time. Precipitation yields on average smaller forecast errors in predicting both employment and unemployment; the differences to the forecasts without indicator are however not significant. Vacancies and ALMP can contribute significantly to the forecast accuracy of unemployment, albeit at the cost of (for ALMP even significantly) less accurate employment forecasts. Surprisingly, business cycle indicators improve neither unemployment nor employment forecasts at the three-months horizon.

Table 5.9: Cross-indicator comparison: Panel Diebold-Mariano tests

Tested Indicator	Smaller MAE than Model ... without Indicator U	without Indicator L	with Precipitation (rss) U	with Precipitation (rss) L	with ZEW situation U	with ZEW situation L
3-month horizon						
dax	-4.152 ***	-3.904 ***	-2.260 **	-3.147 ***	0.903	0.060
wholesale	-1.293	0.039	-1.267	-1.702 *	1.671 *	1.099
ifo sit.	-1.956 *	-0.573	-2.736 ***	-2.255 **	-0.831	0.435
ifo exp.	-4.509 ***	-5.685 ***	-5.274 ***	-5.694 ***	-1.169	-1.460
zew sit.	-2.068 **	-1.323	-2.372 **	-2.094 **	–	–
zew exp.	-3.802 ***	-6.892 ***	-5.266 ***	-7.465 ***	-1.462	-2.498 **
almp	5.127 ***	-4.667 ***	4.422 ***	-4.863 ***	4.566 ***	-2.746 ***
vacancies	7.559 ***	-1.144	6.396 ***	-1.443	7.432 **	-0.347
fmm	-0.391	-6.629 ***	-1.709 *	-5.653 ***	0.890	-1.206
nmm	-1.237	-1.789 *	-2.527 **	-2.445 **	0.475	0.273
rss	1.061	1.227	–	–	2.372 **	2.094 **
tnn	-1.095	-0.604	-1.872 *	-1.147	0.837	1.741 *
12-month horizon						
dax	-2.186 **	-2.961 ***	-8.051 ***	-7.883 ***	-7.226 ***	-6.992 ***
wholesale	-4.614 ***	5.124 ***	-10.708 ***	-5.853 ***	-8.204 ***	-5.884 ***
ifo sit.	4.212 ***	2.301 **	-1.235	-1.719 *	-5.865 ***	-6.159 ***
ifo exp.	-1.593	-4.006 ***	-7.762 ***	-9.963 ***	-7.632 ***	-7.346 ***
zew sit.	7.332 ***	6.250 ***	3.384 ***	3.481 ***	–	–
zew exp.	1.222	-0.607	-6.737 ***	-7.213 ***	-6.196 ***	-6.697 ***
almp	-3.979 ***	-11.084 ***	-5.037 ***	-11.627 ***	-8.369 ***	-12.185 ***
vacancies	-1.110	-8.562 ***	-2.245 **	-9.478 ***	-5.646 ***	-10.818 ***
fmm	-5.212 ***	-7.033 ***	-9.761 ***	-9.425 ***	-8.399 ***	-6.984 ***
nmm	2.155 **	-1.276	-4.981 ***	-6.140 ***	-5.976 ***	-5.895 ***
rss	9.460 ***	8.915 ***	–	–	-3.384 ***	-3.481 ***
tnn	1.262	2.740 ***	-5.314 ***	-3.730 ***	-6.284 ***	-5.214 ***

*/**/***: Significant at the 10%/5%/1% confidence level

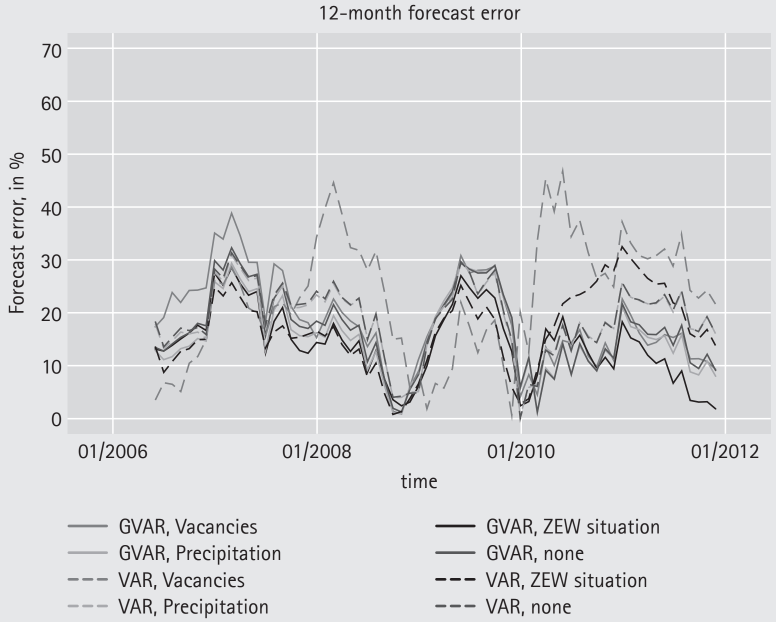
On the one-year horizon the picture changes. Labour market related indicators now lead in general to worse forecasts. Judgements on the current economic situation, in contrast, improve the forecast accuracy relative to forecasts without indicator significantly for both target variables; other business cycle indicators (even the expectations collected together with the judgements on the current situation) don't show a similar potential with regard to labour market forecasts. However, the highest score of a DM panel test against the model without indicator is again achieved by those forecasts using precipitation as an indicator. To strengthen the evidence of the two best-performing indicators, precipitation and the ZEW economic situation index, we test forecasts of all other indicators against these. As the third to sixth column in Table 5.9 show, forecasts employing either precipitation or ZEW situation are more accurate than any other. However, when comparing the two against each other, we find precipitation outperformed by the business-cycle indicator.

5.6.4 Development of forecast errors over time

To complete the analysis of the forecast errors, we present and discuss their development over time. To demonstrate a pattern that seems to us generalizable, we show only a figure for forecasts regarding unemployment at the twelve-month horizon in a single region for a small selection of indicators. We choose the regional division Baden-Wurttemberg (BW) since here the forecast errors are relatively high and differences between the models are more pronounced. Similar graphs for unemployment and employment forecasts at the three- and twelve-month horizon across all regions are provided in Appendix 5.B. The figures include forecasts without indicator, with vacancies (one candidate for the best short-run indicator), ZEW situation (best long-run indicator) and precipitation (relatively good performance in both long and short run). For these, we include additionally to the GVAR forecasts even the VAR predictions.

Figure 5.1 shows (alike most Graphs in Appendix 2) that the Absolute Percentage Forecast Errors at the same point in time are highly correlated across the various forecasting models. Forecast errors in the 12-month ahead unemployment predictions are in general high at the beginning of our evaluation period; they decline to a first minimum of forecast's inaccuracy in the predictions made for late 2008 and early 2009, increased again to a peak for the second half-year of 2009, declined afterwards etc. This pattern holds for most forecast regardless of the indicator employed, and for most regions. That is, the accuracy of forecasts seems to be affected to a large extent by shocks or innovations which are not accounted for in the models.

Figure 5.1: Development of forecast accuracy: Unemployment, $h = 12$, Baden-Wurttemberg



We can see as well that the order of the different forecasting models changes over time. According to Figure 5.1, the GVAR forecast relying on ZEW situation achieved the smallest error of the displayed forecasts approximately from Fall 2007 to Summer 2008, and from late 2010 till the end of 2011. In the time between, we find periods where the (non-spatial) VAR with ZEW situation as indicator, the VAR including vacancies or the GVARs with precipitation and without indicator performed best. This suggests that findings on the comparative performance of indicators are quite sensitive with regard to the period used for evaluation. Furthermore, the figure present the forecast improvement of precipitation in the proper light. An eye-glass is needed to distinguish between the precipitation forecasts and those without indicator; the further line is most often extremely close to the latter. That is, despite its statistical significance, the forecast improvement of precipitation (and supposedly the other climate indicators) is quantitatively marginal. Finally, we can observe that VAR forecasts employing indicators often show a good performance over a certain time span (e.g. vacancies in 2009) but extremely inaccurate forecasts in other periods (vacancies in early 2008 and in 2010). In contrast, the GVAR forecasts rarely reach this good performance but even less are far off. Thus, the main advantage of the GVAR forecasts seems to be in their reduced sensitivity with regard to the indicators.

5.7 Conclusion

The focus of this paper is on forecasting regional labour markets. It is broadly accepted that two aspects regarding the modeling strategy are essential for the accuracy of forecast: a parsimonious model focusing on the important structures, and the quality of prospective information. Thus, our aim is twofold. First, we establish a Global VAR framework, a technique that considers spatio-temporal dynamics in a multivariate setting, that allows for spatially heterogeneous slope coefficients, and that is nevertheless feasible for data without extremely long time dimension. Second, we use this framework to analyse the prospective information regarding the economy due to spatial co-development of regional labour markets in Germany. We employ the same model to examine the information content of a set of commonly used business-cycle indicators, and compare it with the predictive information provided by labour-market immanent indicators and climate variables.

The GVAR model has the advantage of allowing for both cross-sectionally strong (factor) and weak (spatial) dependence. Through estimation of the exponent of divergence we can distinguish between the two from inside the modeling framework. We find in the data support for semi-strong cross-sectional dependence which seems reasonable since Germany is a polycentric economy, in contrast to the UK or continental France where clearly dominant spatial units exist. Because of the less-than-strong interregional dependence, it is sufficient to account for the joint influence of the other regions by constructing spatially weighted aggregates. These local averages are considered to be weakly exogenous. As a second specification issue we investigate the existence of common nonstationary trends. However, nonstationary idiosyncratic components forestall cointegration. Thus, our basic GVAR specification is a model in first differences without imposing cointegration relations and without including a dominant region.

In our forecasting experiment we estimate this basic specification as well as GVAR models which are augmented by leading indicators and use them to predict unemployment and employment in the FEA Regional Divisions three-months and twelve-months ahead. For comparison we forecast for each indicator even a VAR in differences (the same model as the GVAR without local averages) and a VAR in levels (with lag order increased by one). The forecasting experiment starts with data until June 2005, the sample expands month-by-month until December 2010. The forecast accuracy is evaluated by comparing Mean Absolute Percentage Forecast Errors and panel Diebold-Mariano tests.

We can indeed assess a systematic improvement in the forecast accuracy due to accounting for spatial interdependence. The degree of improvement depends on the target variable and the horizon but is robust across all indicators. Albeit, there exist for any horizon regions where all GVAR forecasts for one of the two target variables are less accurate than the corresponding non-spatial forecasts, regardless of the indicator. The indicators themselves are evaluated only by comparison amongst the GVAR-based forecasts. At the three-month horizon, vacancies are the only indicator with significant forecast content regarding one target variable (unemployment) that does not come at a significant cost in forecasting the second target variable, compared to forecasts without indicator. At the twelve-month horizon, forecasts on both target variable can be improved significantly (relative to forecasts without indicators) by judgements on the current economic situation collected in the ZEW and the ifo business-tendency surveys. However, the simulated out-of-sample forecasts assess a similarly high forecast content to Precipitation. The latter produces on average significantly more accurate regional labour-market forecasts than the ifo business situation index, and is only outperformed by ZEW economic situation. Thus, even the best performing indicators seem to provide at the moment only limited prospective information.

5.A Exponent of divergence – simulation

```
% function factor_exponent()
% Program version: GNU Octave 3.2.4 -- filename: factor_exponent.m
function [epsilon s_epsilon Lambdaval] = factor_exponent(Y,reps)
   [T N] = size(Y);
   Llength = (N-2) * reps;
   Lambdaval = zeros(Llength,2);
   row = 0;
   z1 = [1:1:N]';
   for n = 2:N-1
      for rep = 1:reps
          row = row + 1;
          z2 = unifrnd(0,1,N,1); z = [z1 z2]; z = sortrows(z,2);
          idselect = z(1:n,1);
          X = Y(:,idselect);
          S = cor(X,X);
          if n < 7
             lambda = eig(full(S));
          elseif n > 6
             lambda = eigs(S,1);
          end;
          Lambdaval(row,1) = log(abs(lambda(1,1)));
          Lambdaval(row,2) = log(n);
      end;
   end;
   % use correlation rather than covariance
   S = cor(Y,Y);
   lambda = eigs(S,1);
   for rep = 1:reps
   row = row + 1;
      Lambdaval(row,1) = log(abs(lambda(1,1)));
      Lambdaval(row,2) = log(N);
   end;
   y = (Lambdaval(:,1)) - (Lambdaval(:,2));
   x1 = ones(length(Lambdaval),1); x2 = (Lambdaval(:,2)); x = [x1 x2];
   [b_ols, s_ols, r_ols] = ols(y,x);
   s_bols = (x'*x)\eye(length(b_ols)) * s_ols;
   epsilon = b_ols(2,1); s_epsilon = s_bols(2,2);
end;
```

Table 5.10: Simulated critical values for the exponent of divergence (–ε)

Model	T	N	mean	5%	10%	90%	95%
nofactor	10	100	–0.4671	–0.5582	–0.5416	–0.3853	–0.3596
nofactor	20	100	–0.4360	–0.4945	–0.4821	–0.3873	–0.3736
nofactor	50	100	–0.5085	–0.5389	–0.5326	–0.4838	–0.4758
nofactor	100	100	–0.5145	–0.5351	–0.5309	–0.4963	–0.4899
nofactor	10	200	–0.6176	–0.6898	–0.6741	–0.5622	–0.5423
nofactor	20	200	–0.5915	–0.6346	–0.6261	–0.5567	–0.5465
nofactor	50	200	–0.6344	–0.6561	–0.6528	–0.6159	–0.6092
nofactor	100	200	–0.6287	–0.6447	–0.6409	–0.6156	–0.6120
weakfactor	10	100	–0.1535	–0.4698	–0.4246	0.1223	0.1943
weakfactor	20	100	–0.0590	–0.3209	–0.2716	0.1508	0.2256
weakfactor	50	100	–0.1092	–0.2588	–0.2227	0.0007	0.0234
weakfactor	100	100	–0.1089	–0.1916	–0.1719	–0.0470	–0.0282
weakfactor	10	200	–0.2213	–0.5851	–0.5163	0.0503	0.1205
weakfactor	20	200	–0.1312	–0.4151	–0.3463	0.0705	0.1176
weakfactor	50	200	–0.1374	–0.2813	–0.2400	–0.0406	–0.0201
weakfactor	100	200	–0.1223	–0.1971	–0.1809	–0.0687	–0.0557
strongfactor	10	100	0.1601	–0.1563	–0.0883	0.3937	0.4648
strongfactor	20	100	0.2477	0.0235	0.0745	0.4203	0.4583
strongfactor	50	100	0.0921	–0.0099	0.0118	0.1663	0.1892
strongfactor	100	100	0.0289	–0.0345	–0.0181	0.0724	0.0868
strongfactor	10	200	0.1302	–0.2058	–0.1170	0.3690	0.4166
strongfactor	20	200	0.1958	–0.0089	0.0423	0.3425	0.3829
strongfactor	50	200	0.0639	–0.0253	–0.0042	0.1301	0.1486
strongfactor	100	200	0.0155	–0.0318	–0.0210	0.0524	0.0640

5.B Figures: Development of forecast accuracy

Figure 5.2: Relative forecast errors over time: Unemployment, $h = 3$

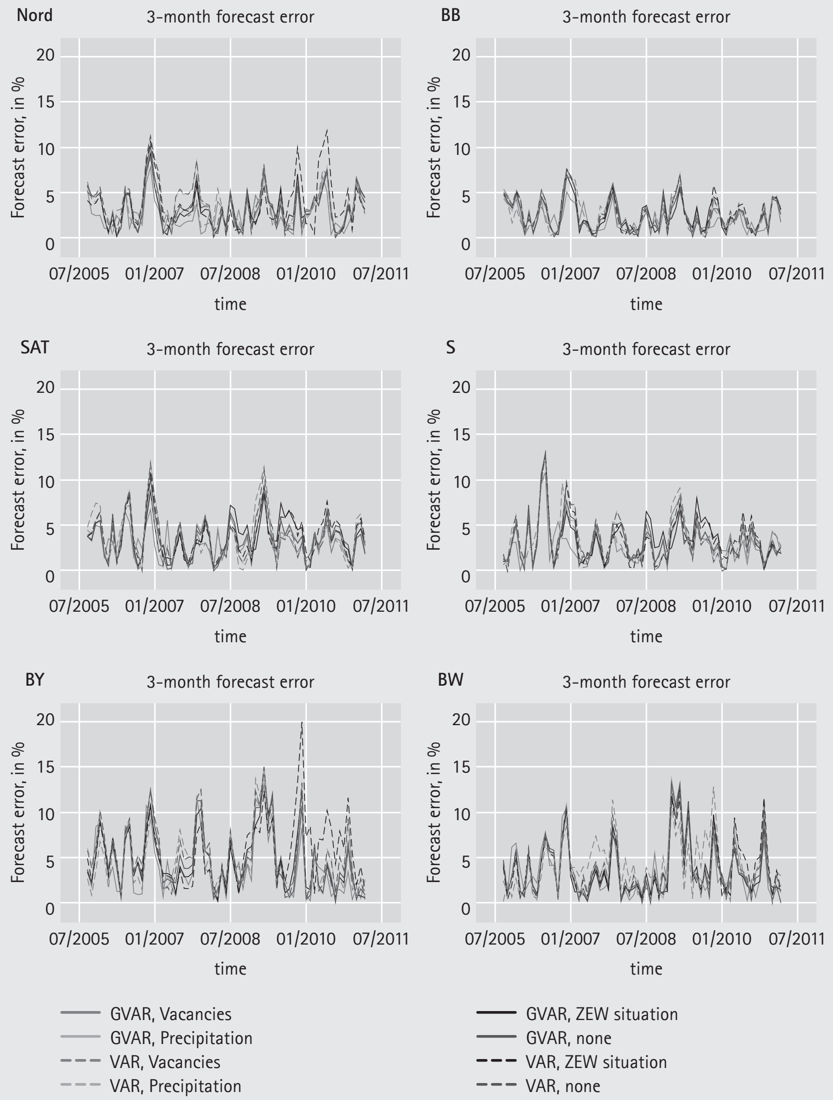


Figure 5.2 continued

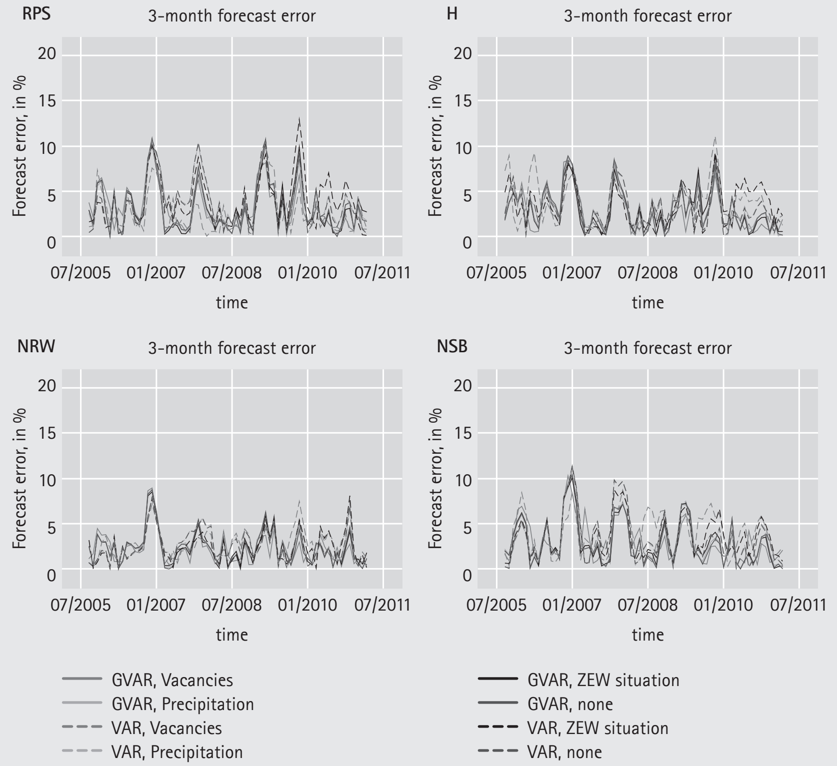


Figure 5.3: Relative forecast errors over time: Employment, *h* = 3

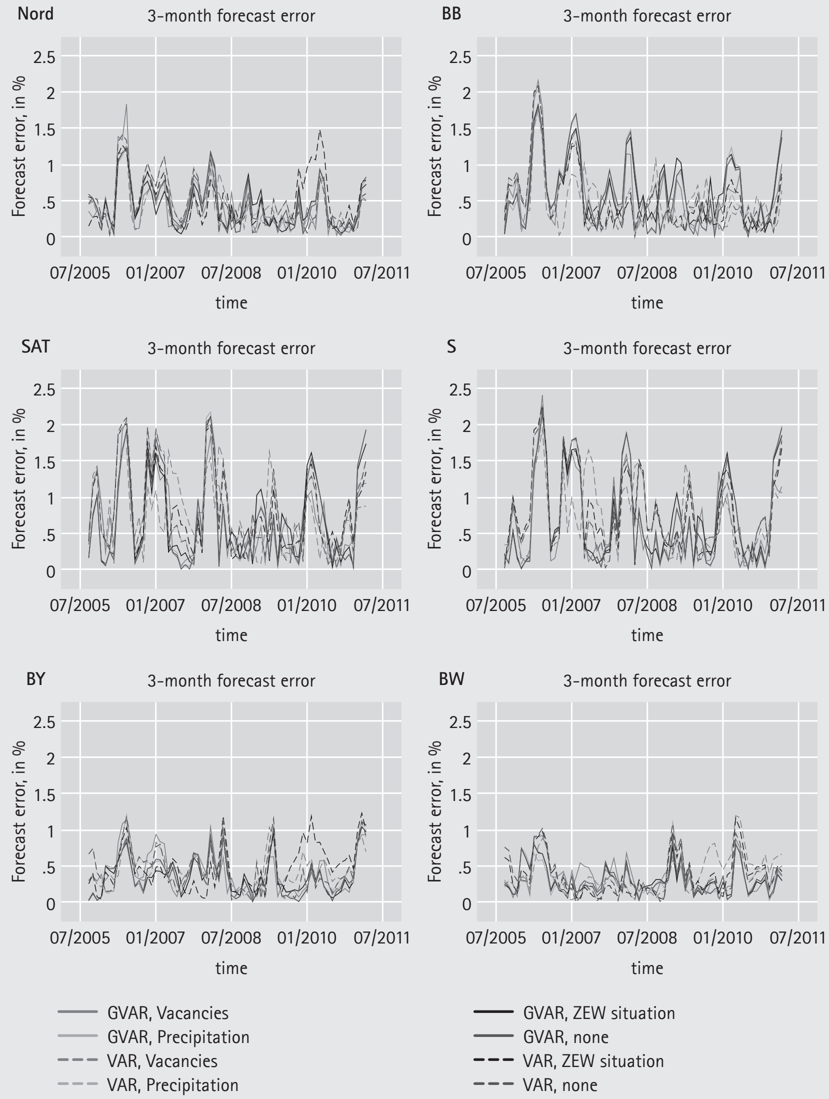


Figure 5.3 continued

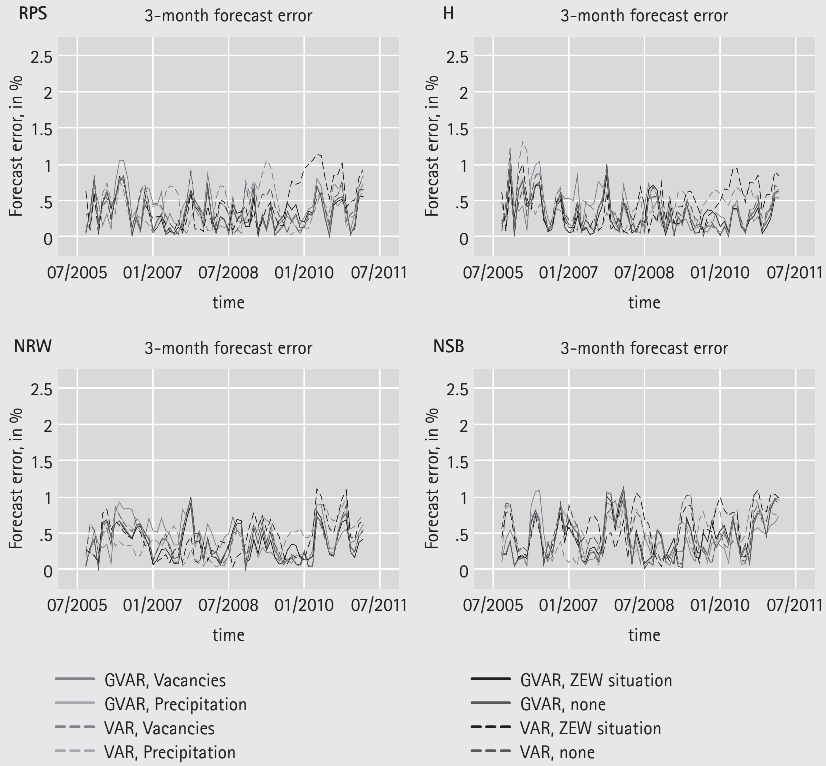


Figure 5.4: Relative forecast errors over time: Unemployment, *h* = 12

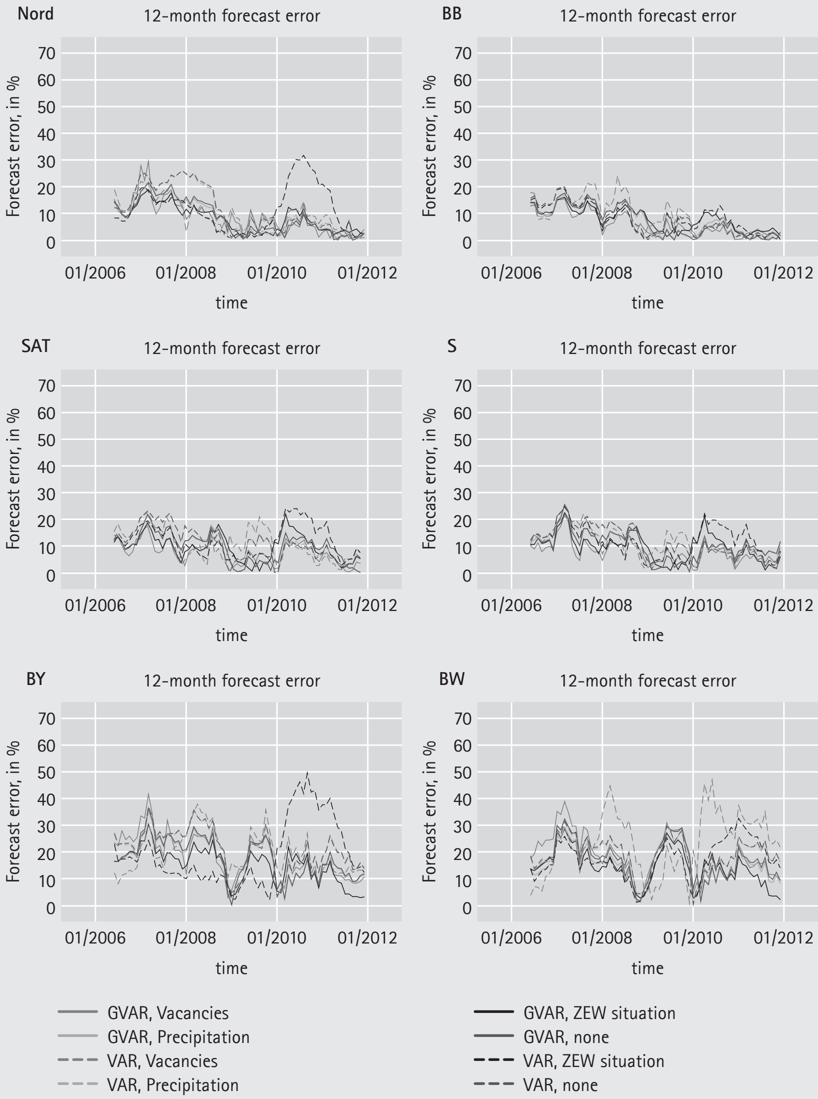


Figure 5.4 continued

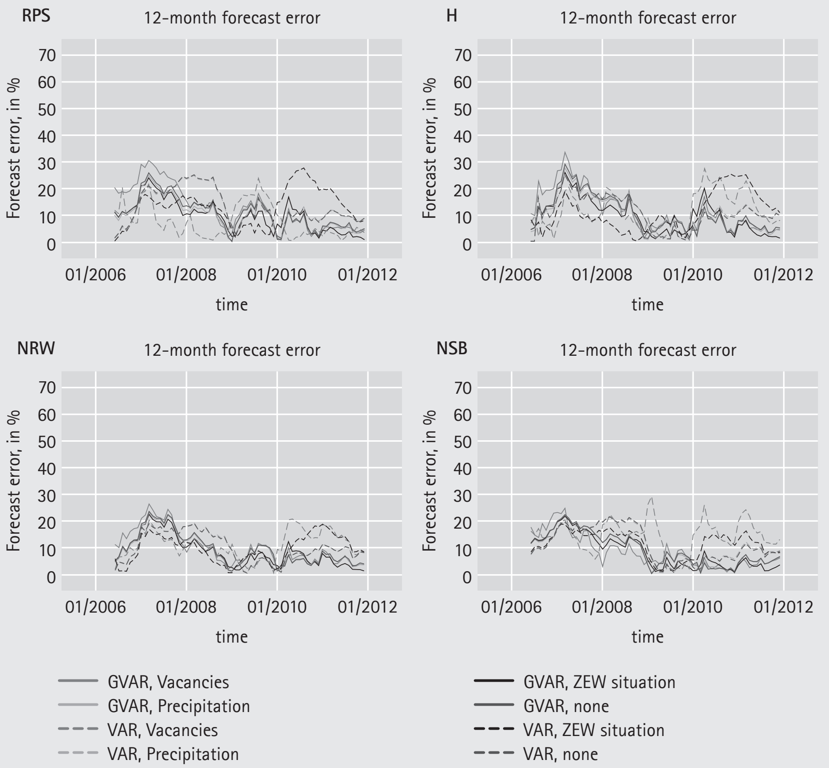


Figure 5.5: Relative forecast errors over time: Employment, $h = 12$

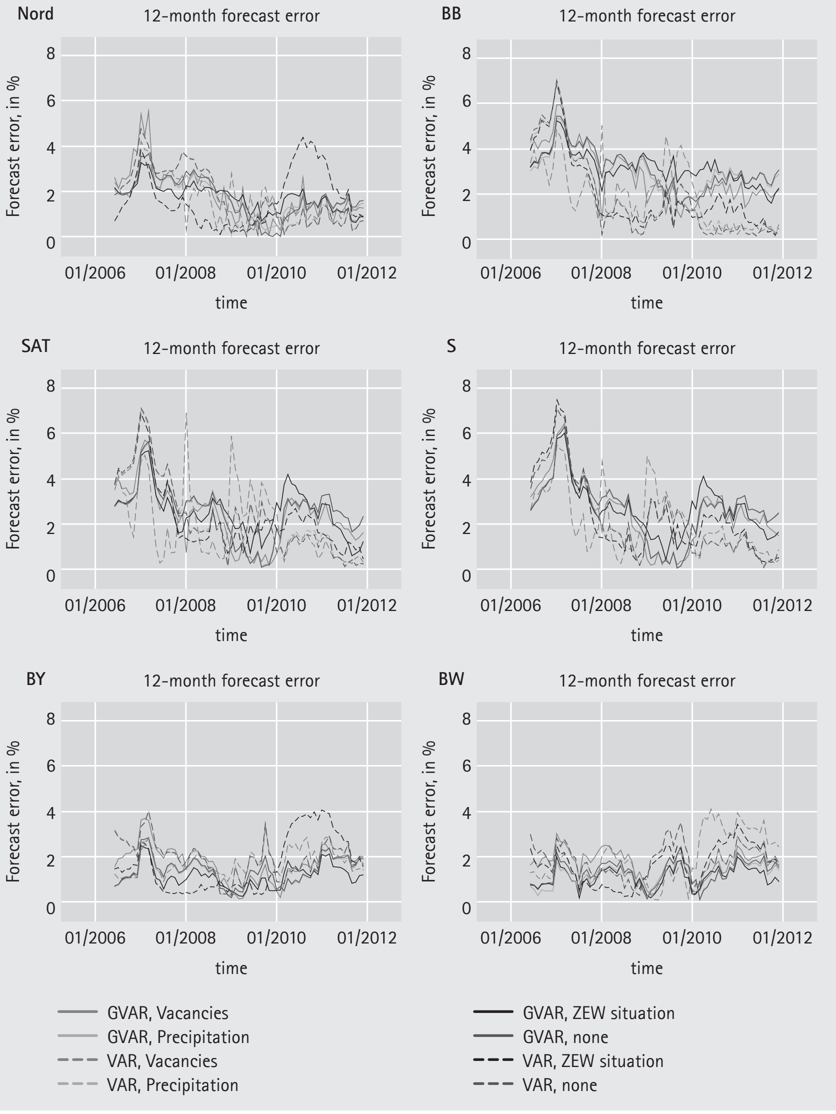
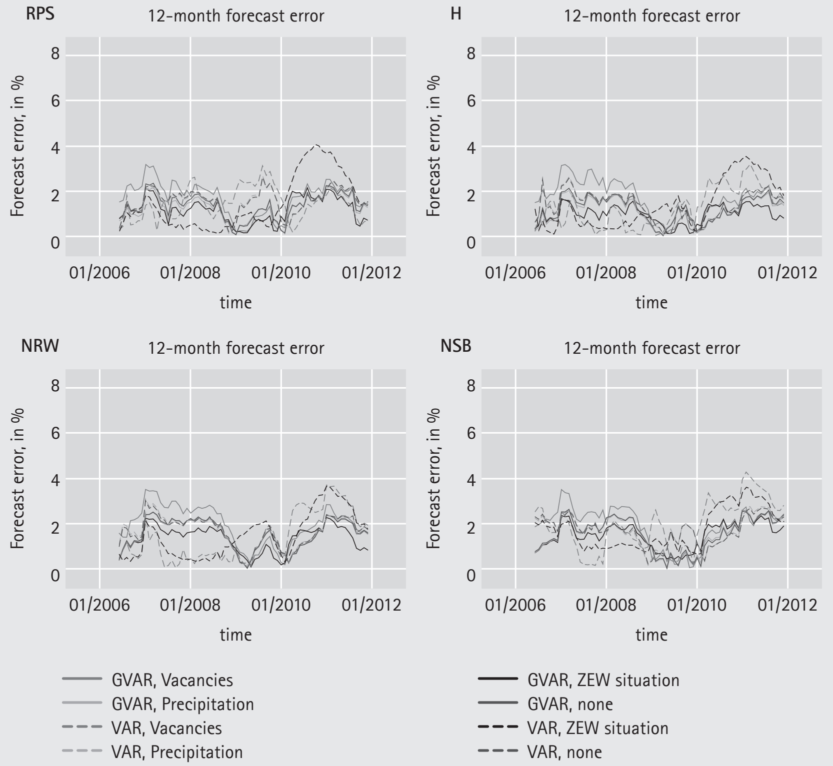


Figure 5.5 continued



Chapter 6

Do they Run with the Pack? The Formation of Experts' Expectations on Labour Markets

By Norbert Schanne[46]

46 Thanks to Silke Delfs and Reinhilde Zimmer for providing access to the survey used in this study. I gratefully acknowledge research assistance by Bastian Stockinger and Pirmin Schanne, and comments by numerous colleagues at IAB. I would like to thank seminar participants at University Regensburg, IAB Nuremberg, the 32nd International Symposium on Forecasting at Boston, the Econometric Society European Meeting (EEA/ESEM) 2012 at Malaga, and the CIRET 2012 at Vienna. Remaining errors are to my fault.

Abstract

Expectations regarding the economic development might be correlated due to various reasons: because individuals use the same public information and similar evaluation methods, and because of social learning or herding amongst peers. We analyse to what extent expectations are driven by herd behaviour, and if it contributes to make expectations more realistic.

In a novel survey the CEOs of the local departments of the German Federal Employment Agency report their expectations on unemployment in the short run. In this data we can discriminate between close and less-close peers to overcome the reflection problem and to quantitatively assess answers regarding the two questions.

We find strong evidence for herding in expectation formation. The size of effect is robust across various specifications and remains even when controlling for forecasts from experts external to the survey. The social multiplier approximately doubles the effect of observable information included in the model. Compared to counterfactual expectations without herding constructed from the estimates, herding seems to improve the accuracy of the expectations.

Keywords: Economic expectations; Expectation formation; Herding; Information cascades; Labour market forecasts; Peer effects; Social learning; Spatial dependence

Jel: C 31; D 83; E 24; J 64

6.1 Introduction

Expectations regarding forthcoming events and the future development of the economy are essential for the plans and decisions of economic and political actors. There's an ongoing debate on the formation and the (frequently missing) rationality of individuals' expectations, focusing on two questions: Which information (private or public, to-date or outdated etc.) do people evaluate in which way when formulating their expectations? Are individuals' expectations really looking forward, or do individuals when announcing their expectation pursue some other objective than predicting future outcomes (e.g. building up reputation)? In the present study on unemployment expectations we will address in particular the first of the two questions: our focus is to empirically ascertain the impact of social learning in expectation formation – that is, of observing and mimicking the expectations of other individuals – and to separate it from the effect of learning from public information.

Professional forecasters in the US[47], a small, intensely interactive group, seem to have fairly realistic expectations, using available information efficiently and adapting quickly to new trends (depending on the sample and the investigation design, see e.g. Chroushore, 1998, 2010 for an overview addressing to a large extent inflation expectations). Expectations of non-forecasters are, however, found to be less realistic. Private households seem to learn on average slowly about economic trends, with only a small part of the population receiving new information (e.g. from reading newspapers) and a large though declining fraction of the population relying on outdated information (Carroll, 2003; Curtin, 2003). Nevertheless, they even tend to extrapolate past trends into the future, causing overly pessimistic expectations at the end and too optimistic early in a recession (Tortorice, 2012).

Surprisingly, most studies on unemployment expectations treat individuals' responses as independent from each other although they, when asked about their sentiments, report similar expectations. Of course, it might be argued that – at least for professional forecasters – individuals expect a similar development because they utilize the same or at least related information; as well, they employ similar methods and models (see Zarnowitz/Lambros, 1987; Keane/Runkle, 1990; Hey, 1994). Thus, strongly diverging expectations generated using the same data and methods would indeed come at a surprise; diverging expectations would result from diverging (private) information. However, correlation of expectations

47 To our knowledge, forecasting studies using European data focused on GDP, inflation and stock-market expectations; corresponding studies analysing unemployment expectations in European countries are to date missing.

might even be due to social interaction if people either communicate their private information to other persons or conclude on the information received by other individuals from observing the announcements, actions and decisions amongst their peers. In the extremum of 'herding' (Banerjee, 1992) or an 'informational cascade' (Bikhchanandi/Hirshleifer/Welch, 1992) social learning may cause total disregard of the respective personal information.[48] As a consequence, expectations can follow blindly a wrong direction or, on the opposite, converge fast to the correct direction/value. Where the informational cascade ends up and when it is broken, is a question of the initial direction and of (public) information that becomes available throughout the cascade. Various studies, e.g. Bikhchanandi/Hirshleifer/Welch (1998), Smith/Sørensen (2000), Çelen/Kariv (2004), Chamley (2004), Manski (2004), Acemoglu/Ozdaglar (2011) and Acemoglu et al. (2011), provide deeper (theoretical, simulation and experimental) insight under which conditions – unboundedness of private believes, continuous rather than discrete signals, repeated decisions in a stationary environment, etc. – social learning warrants convergence to the correct expectation/decision. However, it might be interesting even to quantitatively assess the effect of social learning, and to test if herding contributes to more realistic expectations – despite the conceptual differences between herding theory (modeling a sequence of decisions in continuous time in a static world) and empirical herding analysis (with data observed in discrete time in a dynamic world) as it has been emphasized by Welch (2000).

Some papers – addressing expectations on unemployment and other macroeconomic figures such as GDP growth, inflation or stock market indices – use the serially lagged consensus, that is the expectations' average, as herding variable (e.g. Bewley/Fiebig, 2002; Rangvid/Schmeling/Schrimpf, 2012). This might be a misspecified measure since herding should arise amongst fairly contemporaneous observations and not with those announced a month or quarter

48 Both models, as well as a number of follow-up studies, have a similar structure: Individuals receive private signals; they observe the behaviour of other persons who had to decide earlier, and conclude from their observations on the social aggregate over previous signals; and the individuals make their decisions according to a combination of their private signal and the socially aggregated signal. In both models two (or three) subsequent individuals acting in the same fashion are sufficient to initiate a non-optimal social outcome (be it sitting in the worse restaurant, or having the wrong believe with regard to future development of unemployment), and only the first individual of these two is required to have the wrong information. The private information of all following individuals becomes in general irrelevant once the cascade has started and the herd began to move. Bikhchanandi/Hirshleifer/Welch (1998) and Chamley (2004) demonstrate that cascades occur with a high probability.
The models of Banerjee (1992), Bikhchanandi/Hirshleifer/Welch (1992), Smith/Sørensen (2000) and Acemoglu et al. (2011) describe games in which each individual has only to make a singular decision. Learning from mistakes, or gaining reputation as a person with good information, is not possible in this framework whereas both might be possible if decisions have to be made (expectations have to be formed) repeatedly (as in Manski, 2004). However, the effects of learning and, in particular, of reputation will be only minor if the correct outcome varies across time and if the private signals regarding this outcome are redistributed every period: then, a person knows about the relative reliability of the (observable) public information compared to the private signal and the socially aggregated information; but not the reliability of her signal's current realization or the reliability of the signal she presumes a single peer to have received.

ago. Even fewer studies test for herding in contemporaneous expectations. Pons-Novell (2003) provide evidence of herding amongst some groups of forecasters in the Livingston Survey. Rülke/Tillmann (2011), in contrast, reject herding of the FOMC members with regard to unemployment sentiments. Both studies rely on the contemporaneous consensus. As we will argue, this measure may be also problematic: empirical identification of social learning, herding and imitation among peers is crucial, but not warranted in a linear-in-means peer model.

Manski (1993) shows that, in a model entailing on the right hand an expectation about the peers' observations of the dependent variable and expectations about the distribution of explanatory variables amongst the peers, the parameters of this model are frequently not identified when estimating conditional expectations. If a variable's expectation is linear (that is, estimable by the equally weighted arithmetic mean) the expectation of the dependent variable on the left hand will be reflected by the expectation of the dependent variable (across the peers) on the right hand, and likewise will be the expectations of the explanatory variables. The reflection problem may be overcome by nonlinear exclusion restrictions e.g. due to a nonlinearity in the expectation or by imposing a network structure amongst the peers (for theoretical discussions and an overview across various applications see Soetevent, 2006 and Bramoullé/Djebbari/Fortin, 2009).

The survey employed in this paper – collected amongst the CEOs of the local departments of the German Federal Employment Agency (FEA) – allows to assess to each respondent her location. By assuming more intense communication amongst geographical neighbours and amongst CEOs forced to meet frequently, we are able to discriminate between close and less-close peers. Thus, we can impose a network structure amongst the survey participants which so far has not been possible in the literature on (macro-)economic expectations. The cross-sectionally dependent model is estimable with spatial econometric techniques since we observe a complete network and thus have hardly a problem of omitted peers, missing network nodes or 'edge effects' (omitted spatial neighbours). The dependence structure is similar to a spatial Durbin model (e.g. LeSage/Pace, 2009), i.e. a model including a spatially lagged dependent variable to account for the 'endogenous effect' amongst peers, spatially lagged exogenous variables for the 'contextual effect' as well as a spatially correlated error term for the 'correlated effect'.

We model unemployment-growth expectations to be affected by previous local unemployment and vacancies as the observable market fundamentals according to a matching function, private (for us and for other CEOs unobservable) signals about job destruction, job creation and plant closures, and additionally by endogenous and contextual peer effects. We find strong evidence for the interdependence of the expectations among peers which we interpret as social learning. Our estimate for the social multiplier – a measure for the endogenous peer effect – mounts approximately to two, robustly across various specifications; that is, social interaction roughly doubles the effect which we assess directly to observable characteristics. The size of this effect persists even if we include public professional unemployment forecasts as an alternative source of social information. Furthermore, we find evidence for the herd collapsing at the time when *information* on a trend reversal becomes available, and not at the time of the trend reversal itself; the herd re-establishes a few months later. This suggest that the CEOs are susceptible to information cascades. However, though expectations in our survey are overly pessimistic, we find that social learning brings them (to a small degree) closer to the realized development. Herding seems indeed to make expectations more rational.

However, our paper is limited with regard to some aspects. First, we abstract from strategic herding due to reputation effects (described e.g. by Scharfstein/Stein, 1990; Lamont, 2002; Ottaviani/Sørensen, 2006); admittedly, we wouldn't be able to distinguish reputation effects from social learning with the available information. The labour-market experts in the FEA survey gain reputation from effectively reducing unemployment, not from having expectations close to those of their principals; hence, reputation bias is likely more severe amongst professional forecasters.[49] Second, we do not directly test if expectations are rational although we touch this issue in Section 6.7. Supposedly, the costs of underestimating unemployment (e.g. in terms of lacking the resources necessary to get unemployed quickly into training or work) may exceed the costs of overestimating (e.g. in terms of participation in active labour market policy, ALMP, below the capacity frontier). Thus, rational CEOs would have an asymmetric loss function whereas we employ mean-square loss as it is custom. Third, we leave to further research whether the reported expectations correspond to subsequent action.

The paper is organized as follows. Section 6.2 describes the survey and the record data considered throughout the investigation. Section 6.3 discusses the basic model, Manski's Reflection Problem and a number of estimation issues arising in spatially autoregressive panel models. Section 6.4 entails the analysis of the expectation formation process in general. Sections 6.5, 6.6 and 6.7 focus on particular aspects of expectation formation: learning from public information, the detection of an informational cascade through its fragility and the question if expectations become more realistic by herding. Central results are recapitulated in the conclusion.

49 Note that Keane/Runkle (1990) argue that survey responses amongst professional forecasters are more realistic. However, various studies finding reputation bias stronger amongst professional forecaster than amongst non-professional may provide support for our perspective.

6.2 Data

6.2.1 The FEA Management Survey

The following section describes the Management Survey of the FEA which has been started in November 2008.[50] In addition to this survey we use record data generated inside the FEA in labour-market administration processes and provided by the FEA statistics. The responsibility for the survey is at the FEA's labour-market monitoring service. The survey is collected at a monthly frequency amongst the CEOs (Vorsitzende der Geschäftsführung, VG) of the local offices of the FEA. Probably, these are more expert than the participants in the Michigan Survey of Consumers but have less forecasting expertise than the respondents in the Survey of Professional Forecasters or the Livingston Survey (as the three surveys frequently employed in the literature on unemployment expectations). Reporting date is around the record day of FEA statistics; the realized numbers of unemployed persons, participants in labour-market programmes and employees in the current month are unknown at the date of response, the corresponding numbers of the previous months have already been published (except the number of employees according to the FEA register which is first released with a three-months delay). The set of questions and the possible items corresponding to the questions vary; a short summary is provided in Table 6.9 in the appendix. However, the question which is of central interest in our paper has been observed continuously since the beginning:

> *How do you expect unemployment in your district to develop within the next three months (besides the usual seasonality)?*

The answer allows five items [Decline strongly, Decline, Stay constant, Increase, Increase strongly] which are associated numerically to the values $\{-2,-1,\ldots,2\}$. Although the variable is ordinal and only defined on a 5-groups Likert scale, we treat it like an interval variable throughout the analysis; this allows us not only to use quantiles of this variable but even means and standard deviations. However, we try to be careful about interpreting numbers.

Respondents typically answer in consultation with their top-level staff; thus, the answers can be considered as an institutional expectation rather than an individual. A major advantage is that the answers are not anonymously. We know the agency district (that is the location) of each respondent. As a consequence, we

50 Results of the current and recent waves are available in the FEAs intranet. For further information and data access, contact the corresponding author.

can easily assign to each respondent her geographical neighbours, her Regional Division[51] (RD) and thus those regions reporting to the same principal, as well as those respondents sharing the same benchmarking group[52]. Responses are available for all 176 agency districts for every month since November 2008 (till April 2012, the last month considered here); as participation in the survey is not voluntary, it is not affected by non-response or panel mortality.

Throughout the entire observation period, 676 times a CEO expected unemployment to increase strongly within the next tree months, 2056 times to increase (moderately), 2249 times to remain at the same level, 2371 times to decline, and only 40 times to decline strongly. These (the 40 decline-strongly responses) are observed in only 12 districts, of which in three districts the CEOs expected unemployment to decline strongly five times and in one district ten times. The asymmetry in the answers may be partly due to the German job miracle: unemployment remained, against any economic intuition, rather low during the credit-crunch crisis – and thus had not too much potential to decline during the recovery.

Table 6.1: Unemployment sentiments – descriptive statistics

Statistic	Overall	I/2009	II/2009	I/2010	II/2010	I/2011	II/2011
Median	0.000	1.000	1.000	0.000	-1.000	-1.000	0.000
Mean	0.129	1.335	1.014	0.127	-0.609	-0.775	-0.390
Std. Dev.	0.985	0.509	0.577	0.824	0.529	0.447	0.607
Within Std. Dev.	0.981	0.326	0.406	0.719	0.372	0.329	0.434
Cross Std. Dev.	0.543	0.504	0.562	0.620	0.526	0.440	0.579

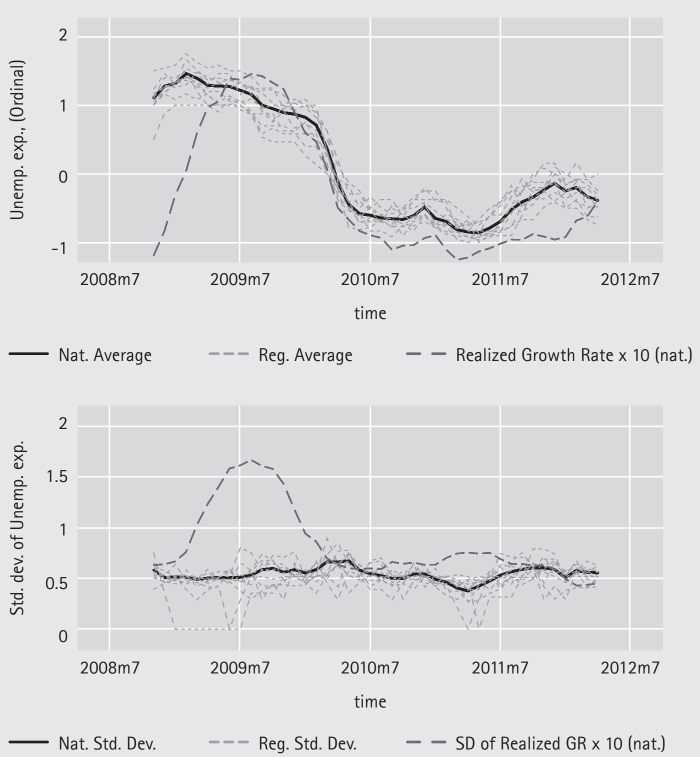
Descriptive figures allow a first illumination of the data. Common statistics are reported in Table 6.1; numbers in column 2 refer to the entire observation period whereas columns 3 to 8 show the corresponding statistics calculated for the respective half-year. According to the median over the entire observation period and the semi-annual medians the CEOs seem to expect a more or less cyclical

51 Regional Divisions form the intermediate organisational level between the local and the national. The ten RDs have approximately the size of the major German federal states and collect between 8 (Berlin/Brandenburg) and 33 (North Rhine-Westphalia) agency districts.

52 Each agency district belongs to a benchmarking group implemented according to comparable economic conditions (the procedure is described by Rüb/Werner, 2008 and Blien/Hirschenauer/Thi Hong Van, 2010); these groups do not coincide with the Regional Divisions. CEOs (and their top-level staff) have to participate in periodical meetings of their benchmarking group.

development of unemployment, such that increases balance with decreases and that, over a longer period, unemployment remains rather constant. However, they seem to be slightly pessimistic; the total mean and the period-specific averages are always higher than the corresponding median values. The dispersion of the sentiments is dominated by the variation over time. The within standard deviation[53] has the same size as the total standard deviation; the cross-sectional standard deviation (computed using the deviations from the time-specific mean) has significantly smaller size. However, a large amount of the temporal variation in the sentiments seems to be due to the variation in the first half-year in 2010. The standard deviation in other half-years mounts to a size similar to that of the cross-sectional standard deviation.

Figure 6.1: Development of unemployment sentiments



53 Within Std. Dev. = $\sqrt{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)^2}$; Cross Std. Dev. = $\sqrt{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}-\bar{y}_t)^2}$.

The early year 2010 is even more outstanding when we inspect the series of monthly national and regional mean expectations in Figure 6.1. Sentiments of local CEOs are averaged at the RD level to compute regional means. The figures indicate little variation in expectations across regions. They seem to develop in a similar fashion, more or less parallel to the national average; only in a single RD the expectations move towards 'declining unemployment' earlier than the average as the lowest dashed line (in the period from December 2009 to March 2010) in the upper diagram of Figure 6.1 indicates. Time-specific standard deviations with regard to the national and each regional data set (the lower diagram) are in general close to or below 0.5 (in one RD, all CEOs gave the same response in early 2009); only in Spring 2010 it exceeded 0.5. Hence, the distribution of responses seems to exhibit a shift of the first moment but stable second moments.

Figure 6.2: Sentiments and realized unemployment growth

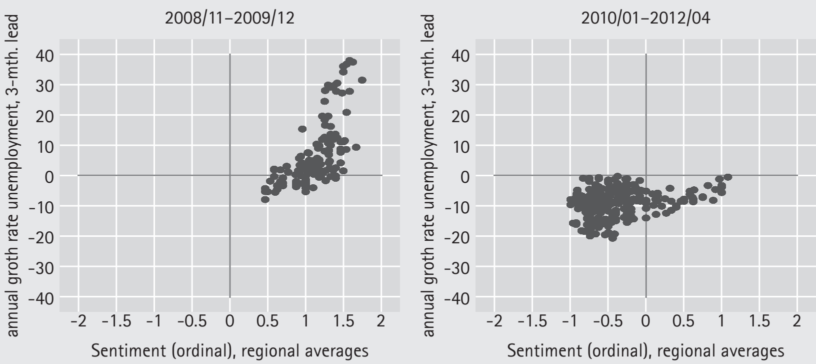


Figure 6.2 contrasts the expected and the realized development of unemployment, both monthly averaged over the local districts in each regional division. We divide the data in two samples because of the expectations' shift in early 2010. Throughout the entire period before January 2010 all regional averages across expectations take on values above 0.5; that is, the majority of CEOs in every regional division expected unemployment to increase. Averages higher than 1.5 – corresponding with a regional majority of CEOs expecting a strong increase – are however rare. This concentration of sentiments' averages does not correspond to the realized development of unemployment at the forecast horizon: we observe both declining unemployment (with a regional average growth rate of roughly -2.5 percent within three months, or -10 percent annually) as well as strong increases (up to 10 percent within a quarter, corresponding with an annual growth rate of roughly 40 percent).

In the period from January 2010 onward both expectations and realized unemployment growth shifted downward; all regional averages in the right sentiment-realization diagram are located in the second and third quadrant. That is, unemployment declined everywhere within the next quarter (with an estimated annual growth rate up to -20 percent). Nevertheless, the sentiments seem again too pessimistic: In a number of regional divisions, the majority of CEOs still expected unemployment to increase (in particular in early 2010), and the mass of the distribution is slightly right from the value corresponding to a moderate decline (in the direction of the 'no change' item).

6.2.2 Prospective public information

Public information plays a key role particularly in models for expectations of non-professional forecasters. In the model of Bikhchanandi/Hirshleifer/Welch (1992), external public information may cause the collapse of an informational cascade; thus, the effect associated with herding likely exhibits a response. In the model of Carroll (2003), only a fraction of private households receive news about professional forecasts and adapt their expectations accordingly; social learning might interact with learning from public information. We discuss both in detail later, hence it seems adequate to clarify our notion of public information before.

Though the local labour market data reported by the FEA statistics which is available to each CEO as well as outside the Federal Employment Agency is in general public data, we consider it here as private information with regard to trend reversals: cyclical shifts may be hard to detect besides the seasonal and irregular fluctuation. Another source of public information are the business-cycle forecasts published by the major German economic research institutes.[54] Their forecasts on GDP growth and unemployment in the years 2009 till 2011 are listed by month of release in Table 6.2.

54 Those forecast which attract the greatest deal of attention are, supposedly, published by the Ifo Institute for Economic Research (Ifo, Munich), the Institute for the World Economy (IfW, Kiel), the ZEW Centre for European Economic Research (ZEW, Mannheim), the Macroeconomic Policy Institute (IMK, Dusseldorf), the Halle Institute for Economic Research (IWH, Halle) and the Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI, Essen); these institutes are (or have been) moreover involved in the Joint Economic Diagnosis (Gemeinschaftsdiagnose). With regard to labour markets, forecasts of the Institute for Employment Research (IAB, Nuremberg) receive high attention as well.

Table 6.2: Availability of public information on trend reversals

Released in	Publishing institute	GDP growth rate forecasts for				unemployment forecasts (in '000 persons) for			
		2009	2010	2011	2012	2009	2010	2011	2012
Oct 08	IMK	0.2	–	–	–	3,263	–	–	–
Dec 08	ifo	-2.2	-0.2	–	–	3,471	3,971	–	–
Dec 08	IfW	-2.7	0.3	–	–	3,665	3,949	–	–
Dec 08	IMK	-1.8	–	–	–	3,882	–	–	–
Jan 09	RWI	-4.3	0.5	–	–	3,727	4,633	–	–
Mar 09	IfW	-3.7	-0.1	–	–	3,642	4,251	–	–
Apr 09	IMK	-6.0	-0.5	–	–	3,718	4,688	–	–
Jun 09	ifo	-6.3	-0.3	–	–	3,581	4,337	–	–
Jun 09	IfW	-6.0	0.4	–	–	3,576	4,365	–	–
Jun 09	RWI	-5.0	1.2	–	–	3,481	4,133		–
Jul 09	IMK	-6.5	-0.4	–	–	3,575	4,448	–	–
Sep 09	IfW	-4.9	1.0	–	–	3,444	3,881	–	–
Oct 09	IMK	-5.0	1.2	–	–	3,470	4,075	–	–
Dec 09	ifo	-4.9	1.7	1.2	–	3,426	3,607	3,617	–
Dec 09	IfW	–	1.2	2.0	–	–	3,827	3,935	–
Dec 09	IMK	-4.9	2.0	–	–	3,424	3,600	–	–
Jan 10	RWI	–	1.4	1.6	–	–	3,475	3,565	–
Mar 10	IfW	–	1.2	1.8	–	–	3,443	3,275	–
Apr 10	IMK	–	1.5	1.4	–	–	3,382	3,313	–
Jun 10	ifo	–	2.1	1.5	–	–	3,233	3,043	–
Jun 10	IfW	–	2.1	1.2	–	–	3,199	2,952	–
Jun 10	IMK	–	2.0	1.5	–	–	3,226	3,048	–
Jun 10	RWI	–	3.4	2.2	–	–	3,250	3,055	–
Sep 10	IfW	–	3.4	1.7	–	–	3,235	2,958	–
Oct 10	IMK	–	3.5	1.9	–	–	3,236	2,933	–
Dec 10	ifo	–	3.7	2.4	–	–	3,242	2,943	–
Dec 10	IfW	–	3.7	2.3	1,3	–	3,252	2,984	2,778
Dec 10	IMK	–	3.7	2.5	–	–	3,240	2,963	–
Jan 11	RWI	–	–	2.9	2,4	–	–	2,875	2,467
Mar 11	IfW	–	3.6	2.8	1,6	–	3,244	2,992	2,803
Apr 11	IMK	–	–	2.7	1,7	–	–	2,944	2,758
Jun 11	ifo	–	–	3.3	2,3	–	–	2,944	2,683
Jun 11	IfW	–	–	3.6	1,6	–	–	2,970	2,687

Released in	Publishing institute	GDP growth rate forecasts for				unemployment forecasts (in '000 persons) for			
		2009	2010	2011	2012	2009	2010	2011	2012
Jun 11	IMK	–	–	4.0	2,3	–	–	2,949	2,740
Jun 11	RWI	–	–	2.9	2,3	–	–	2,965	2,650
Sep 11	IfW	–	–	2.8	0,8	–	–	2,984	2,868
Sep 11	RWI	–	–	2.9	1.0	–	–	2,965	2,805
Oct 11	IMK	–	–	3.2	0.7	–	–	2,977	2,865
Dec 11	ifo	–	–	3.0	0.4	–	–	2,975	2,800
Dec 11	IfW	–	–	2.9	0.5	–	–	2,976	2,863
Dec 11	IMK	–	–	3.0	-0.1	–	–	2,976	2,900
Mar 12	IfW	–	–	–	0.7	–	–	–	2,775
Mar 12	IMK	–	–	–	0.3	–	–	–	2,876
Jun 12	ifo	–	–	–	0.7	–	–	–	2,866
Jun 12	IfW	–	–	–	0.9	–	–	–	2,866

Forecasts on unemployment (and GDP growth) published by the major German economic research institutes. ifo = ifo institute Munich; IfW = Institute for the World Economy Kiel; IMK = Institute for Macroeconomic Policy; RWI = RWI Essen. The ZEW centre for European economic research Mannheim (ZEW) published only GDP growth rate forecasts. In January 2012, the Halle Institute for Economic Research (IWH) didn't provide their forecasts retrospectively for the period before 2011; thus, their forecasts are not listed.

It might be easier to identify a trend reversal in these publicly available forecasts, as we will demonstrate with Table 6.2; for comparison, note that realized unemployment numbers had been on average 3.258 million in 2008, 3.415 million in 2009, 3.238 million in 2010 and 2.972 million in 2011. All institutes expected unemployment to rise in 2009; throughout most of the year 2009, by five to ten percent within the remaining year and by another ten to 25 percent in the following. In late 2009, the expected rise of unemployment became smaller, and in January 2010 the first institute expected unemployment to rise by less than five percent within the same and the subsequent year. In the next forecast published in the first quarter of 2010, a research institute already predicted a decline in unemployment.

6.3 Empirical design

6.3.1 The basic model of labour market expectations

In the following paragraphs, we will construct a fairly simple model which we hold to be valid in describing the CEOs expectation formation process. The core of the model consists of three variables related to two components: First, we suppose that the CEOs expect the short-run unemployment dynamics to follow a matching process: jobs are created by matching unemployed persons and vacant jobs. Thus, unemployment and vacancies (included in annual differences to eliminate seasonality and non-stationarity) are two central observable indicators describing the market fundamentals.[55] Second, CEOs typically will receive information on planned job destruction (firm closures and job separation) which is not collected by the FEA statistics; this, as well as information on vacancies not reported to the statistics, form private signals on labour market dynamics. We denote the sentiments with y and the vector entailing the market fundamentals (unemployment and vacancies in logs) with x.

Furthermore, we assume that the CEOs take also information from recent periods into consideration; this may result in belief persistence or slow/delayed adaption to new information. The entire information available up to period $t-1$ is already incorporated in the sentiments reported in the previous period; thus, we include the first lag of the dependent variable (denoted with $L.y$) as additional regressor on the right hand (following Carroll, 2003). The autoregressive component may also reflect that CEOs adapt to prediction errors over time – which in fact would correspond to a Moving-Average (MA) process.

It is unlikely that the CEOs consider only the reported market fundamentals referring to the district they are responsible for. Regional labour markets are interdependent, thus it would be irrational not to look at unemployment and vacancies in other regions. The development in other regions, or the aggregate development, may help to assess whether ups and downs in unemployment are national phenomena affecting all regions or a local phenomenon leveling across the regions; it has been shown that consideration of spatial co-developments provides prospective information (e.g. Schanne/Wapler/Weyh, 2010; Mayor/Patuelli, 2012). CEOs aggregate the information observable across other regions, that is across those districts under responsibility of their peers, to a contextual effect: a conditional expectation which we denote with $E(x|p)$ wherein p represents the characteristics defining the peer group or the social network. In addition to this, we expect a CEO and her top level staff to communicate with other CEOs/their top level staff (i.e. with the persons considered as their peers) in their daily business. Supposedly labour-market expectations are subject matter of communication in the time between two waves of the survey. Then, a (rational) CEO would draw information from the (preliminary) sentiments announced by her peers; the aggregate over the peers' sentiments is $E(y|p)$.

55 Current unemployment and vacancies are two of the numbers published monthly by the FEA at various regional levels; their development is typically discussed in the first chapter of the monthly German Labour Market Report (together with employment which is published with a delay of one/two months). Validity of vacancies as a leading indicator regarding unemployment at a three-months forecast horizon has been shown recently by Schanne (2010).

Thus, we need to consider four sources of information in our model of expectation formation: the own history of sentiments (which accounts for learning from own mistakes as well as past information), statistical (verifiable) information directly related to an observation, an aggregate of contextual statistical (verifiable) information related to the peers, and social learning. A regression model with these four sources of information can be written as follows:

$$y = \alpha + x'\beta + E(x|p)'\gamma + E(y|p)\eta + L.y\,\phi + e \tag{6.1}$$

When estimating, that is when taking expectations conditional on x and p, and under the assumption that the unobservable signals (the disturbances) are somewhat related amongst peers with $E(e|x,p) = g(p;\delta)$, eq. (6.1) becomes:

$$\begin{aligned} E(y|x,p) = {} & \alpha + E(x'\beta|x,p) + E(E(x|p)'\gamma|x,p) + E(E(y|p)\eta|x,p) \\ & + E(L.y|x,p)\,\phi + g(p;\delta) \end{aligned} \tag{6.2}$$

With linear social expectations on y (in the endogenous peer-related effects), the *Reflection Problem* (see Manski, 1993) becomes obvious when applying the Law of Iterated Expectations to eq. (6.2). When sequentially conditioning first on p and then on x, it is possible to receive reduced-form parameters in the scalars $\frac{\phi}{1-\eta}, \frac{\alpha}{1-\eta}$ and the vector $\frac{1}{1-\eta}(\beta\eta + \gamma)$ but, with exception of β not the structural parameters themselves from the following equation:

$$E(y|x,p) = \frac{\alpha}{1-\eta} + x'\beta + E(x'|p)\frac{\gamma + \beta\eta}{1-\eta} + E(L.y|x,p)\frac{\phi}{1-\eta} + \frac{g(p;\delta)}{1-\eta} \tag{6.3}$$

However, a nonlinearity in the model might be sufficient to overcome the identification problem in eq. (6.2) (see Manski, 1993: Sec. 3; Brock/Durlauf, 2001). Identification in models for an ordered categorial variable will be investigated in Subsection 6.3.3.

Spatially autoregressive models will be an alternative to a nonlinear functional form (see Manski, 1993: Sec. 2.6) if sampling issues are negligible (because the data covers the entire population or the sampling accounts for this particular structure): Since distance (as well as neighbourhood relations or communication intensity) to a third unit varies across observations, a weighted average with weights accounting for these distance relations shows variation itself. The intuition behind is that one needs to discriminate between peers and non-peers (or close and distant peers), that is to observe variation, in order to identify the effect of peers (an argument raised in a similar form by De Giorgi/Pellizari/Redaelli, 2010).

Likewise, identification of spatially autoregressive effects in a cross section or in a panel saturated with time-specific effects is impossible if a spatial system entails only neighbouring regions; without time-specific effects, the temporal variation identifies the endogenous peer effect (see, e.g., Kelejian/Prucha, 2002; Baltagi, 2006; Baltagi/Liu, 2009). Identification issues in spatial autoregressive panels are discussed in Subsection 6.3.2.

In our data we are able to investigate alternative peer-relation structures. First, we know the geographical location of each respondent. In daily business communication of (top level) staff members in different agencies is typically most intense with those in close agency districts, that is with those providing job-seekers and vacancies within commuting distance. Thus, using geographical distance (or contiguity) of regions as a measure for communication structures (and for peer relations) in expectation formation seems plausible. Political structures in which groups of CEOs meet periodically provide an alternative network structure which may serve as a robustness check. Here, utilization of benchmarking groups as social network seems advantageous over using Regional Divisions since it is not possible to separate social learning from reputation effects when analysing the latter.

6.3.2 Estimation of spatially autoregressive panels

In terms of spatial econometrics, eq. (6.1) is a dynamic panel regression with a spatially lagged dependent variable (endogenous effect), spatially lagged exogenous variables (contextual effect) and, likely, spatially autocorrelated disturbances (correlated effect). With different sets of weights for the endogenous, the contextual and the correlated effect, the empirical model becomes:

$$y_{i,t} = \phi y_{i,t-1} + \eta \sum_{j=1}^{n} w_{(1)ij} y_{j,t} + \alpha + x_{i,t}\beta + \sum_{j=1}^{n} w_{(2)ij} x_{j,t}\gamma + e_{i,t} \tag{6.4}$$

where $y_{i,t}$ denotes the sentiments of the CEO in region $i=1,\ldots,n$ at time $t=2,\ldots,T$ and $x_{i,t}$ the row vector with the information on log unemployment and log vacancies. The disturbances allow in principle for possible time-invariant unobserved heterogeneity, serial and network autocorrelation:

$$e_{i,t} = \mu_i + u_{i,t} \tag{6.5a}$$

$$u_{i,t} = \delta \sum_{j=1}^{n} w_{(3)ij} u_{j,t} + \nu_{i,t} \tag{6.5b}$$

$$\nu_{i,t} = \rho \nu_{i,t-1} + \varepsilon_{i,t} \tag{6.5c}$$

Eqs. (6.4) and (6.5) are in matrix notation with ι_T a T x 1 vector of ones, I_T the T-dimensional unit matrix (likewise for dimensions n and nT), L_T the lag operator matrix (a T x T matrix with the first sub-diagonal containing ones and all other elements set to zero), $W_{(1)}$, $W_{(2)}$, $W_{(3)}$ as n x n matrices containing the spatial weights, Y, e and ε as Tn-dimensional column vectors arranged such that $Y = (Y_1', \ldots, Y_T')'$ with $Y_t = (y_{1,t}, \ldots, y_{n,t})'$, μ a n-dimensional column vector of region-specific fixed effects and $X = (X_1', \ldots, X_T')'$ with $X_t = (x_{1,t}', \ldots, x_{n,t}')'$ a (Tn) x k matrix:

$$Y = \phi\left(L_T \otimes I_n\right)Y + \eta\left(I_T \otimes W_{(1)}\right)Y + \iota_{Tn}\alpha + X\beta + \left(I_T \otimes W_{(2)}\right)X\gamma + e \tag{6.6a}$$

$$e = \left(\iota_T \otimes I_n\right)\mu + \left(I_T \otimes \left(I_n - \delta W_{(3)}\right)^{-1}\right)\left(\left(I_T - \rho L_T\right)^{-1} \otimes I_n\right)\varepsilon \tag{6.6b}$$

The sequences of weights w_{ij} in a matrix $W \in \left\{W_{(1)}, W_{(2)}, W_{(3)}\right\}$ are assumed to be exogenous triangular arrays satisfying standard regularity conditions:

1. $w_{ii} = 0 \quad \forall i \in \{1, \ldots, n\}$, i.e. the main-diagonal elements of W are zero.
2. $w_{ij} \geq 0 \quad \forall i, j \in \{1, \ldots, n\}$, i.e. there are no negative spatial weights.
3. W has bounded row norm $\| W \|_\infty = \max_{j \in \{1,\ldots,n\}} \sum_{i=1}^{n} \left|w_{ij}\right| \leq c$ and bounded column norm $\| W \|_1 = \max_{i \in \{1,\ldots,n\}} \sum_{j=1}^{n} \left|w_{ij}\right| \leq c$ with $c < \infty$.[56]

Assume further $\delta \in \left(\frac{1}{\min\left\{\left|\lambda_{W_{(3)}}\right|\right\}}, \frac{1}{\max\left\{\left|\lambda_{W_{(3)}}\right|\right\}}\right)$ and $\eta \in \left(\frac{1}{\min\left\{\left|\lambda_{W_{(1)}}\right|\right\}}, \frac{1}{\max\left\{\left|\lambda_{W_{(1)}}\right|\right\}}\right)$ where $\left\{\left|\lambda_W\right|\right\}$ denotes the sequence of moduli corresponding to the combinations of real and complex eigenvalues extracted from W. That is, δ and η are, in absolute value, smaller than the inverse of the largest eigenvalue of the corresponding weights matrix. If these conditions hold, $\left(I_n - \delta W_{(3)}\right)$ and $\left(I_n - \eta W_{(1)}\right)$ will be finite and invertible, and the stochastic process will be cross-sectionally weakly dependent (see Chudik/Pesaran/Tosetti, 2011). Assume for serial dependence that $\left|\phi\right| < 1$ and $\left|\rho\right| < 1$, and for spatiotemporal dependence that $\left|\left(I_n - \delta W_{(3)}\right)\left(1 - \rho\right)\right| > 0$ and $\left|\left(I_{Tn} - \eta\left(I_T \otimes W_{(1)}\right) - \phi\left(L_T \otimes I_n\right)\right)\right| > 0$ and so that the process is covariance stationary.

In general, the matrices associated with the endogenous, the contextual and the correlated effect may but do not need to be different. Depending on the estimation technique, the parameters are still identified when using the same weights for only

56 Thus, W has bounded spectral norm $\| W \| \leq c$. The spectral norm of a matrix A is the square root of the largest eigenvalue of AA', i.e. $\| A \| = \left[\max\left\{\lambda_{(AA')}\right\}\right]^{\frac{1}{2}}$; it is the matrix equivalent to the euclidean vector norm.

two or all three components. Eq. (6.6) may be estimated consistently[57] either by (Quasi) Maximum Likelihood[58] (ML) or by Generalized Method of Moments (GMM) whereas the OLS estimator is biased due to the endogeneity of WY_t (see Anselin, 1988, 2001). Here, we rely on GMM[59] since various instrumentation tests allowing better judgement on identification are not available for the ML estimators (see the discussion on model identification in Gibbons/Overman, 2012). The spatially lagged dependent variable WY_t is instrumented by (first- and second-order) spatially lagged exogenous/predetermined explanatory variables WX_t, W^2X_t, WY_{t-1}. The error-component parameters are (with exception of fixed effects) estimated in separate steps: The correlated effect parameter δ by a first-order spatially autoregressive (SAR 1) panel estimator (see Mutl/Pfaffermayr, 2011 as well as Kelejian/Prucha, 1998, Kapoor/Kelejian/Prucha, 2007), the serial autocorrelation (AR 1) parameter ρ by pooled OLS. Utilization of the same weights for modeling the contextual effects and for constructing the instruments may affect the validity of GMM and 2SLS estimators. However, a sufficiently high partial R^2 statistic of the excluded instruments in the first-stage regression (explaining WY_t) and insignificance (low significance) when testing for overidentifying restrictions (Hansen/Sargan test) indicate an appropriate instrumentation strategy. Given valid instruments the parameters η, β and γ are identified. δ is identified if there exists a consistent estimator for α, β, γ, η and ϕ; however, δ is of minor interest for us.

We let the weights used for the endogenous effects and the contextual effect vary across the specifications in order to establish robustness of the results: as first set of weights we use an indicator variable where regions are considered as peers if the distance between their centroids is smaller than 88.66 km (the mean between percentile 90 and percentile 95 of all pairs of inverse distances; this value is chosen such that any region has at least two peers). The binary information is row-normalized such that $\sum_{j=1}^{n} w_{ij} = 1 \quad \forall i$. As alternatives to this truncated inverse-distance metric, we consider contiguity (indicating that regions share a border) and affiliation with a benchmarking group, both as row-normalized information. We always employ the same set of weights for the endogenous effect and the excluded instruments. In our preferred specification, the contextual effect is constructed with a fourth metric: we directly use inverse distance between regions as continuous weight, rather than row-normalized indicator variables.

57 Nickell-Bias (Nickell, 1981) in the estimate of ϕ is negligible when T is sufficiently large or when the variance of the time-constant error component converges to zero. Thus, we estimate the model in levels and treat the serially lagged dependent variable as weakly exogenous.

58 The ML estimator relies on the n-dimensional (or nT-dimensional) multivariate Gaussian distribution and thus accounts explicitly for the simultaneity of the observations.

59 Two-stage least squares, 2SLS, is considered as a version of GMM; we use 2SLS for estimation of eq. (6.4).

6.3.3 Identification in nonlinear regression models

Procedures for ordered categorial limited dependent variables (LDV) as an alternative estimation strategy come along with two advantages. First, the values of the dependent variable do not reflect equally sized intervals; a procedure like *ordered probit* treats that more adequately than a linear approach. Second, most models for LDVs employ the cumulative density function (CDF). The generated nonlinearity might allow identification under certain conditions (see e.g. Manski, 1993; Brock/Durlauf, 2001 and Brock/Durlauf, 2003 for the bivariate and multinomial-logit case).

However, Bajari/Krainer (2004) argue convincingly that continuous exclusion restrictions are still necessary for identification in the ordered-probit peer effects model. Appropriate exclusion restrictions can for example be derived from distinction between first and second-order neighbours in the peer-related effects while aggregating linearly over the CDFs at individual level.

Estimation of the spatially-autoregressive ordered probit with ML requires structural assumptions on the latent variable y_{ij} and its joint distribution over $i=1,\dots,n$ and $t=1,\dots,T$. Let the endogenous peer effect and the autoregressive component refer to the latent variable such that they can move to the left hand. Further, the parameters on the right hand are identified only up to a variance-scaling parameter σ such that the stochastic process can be represented by a standardized distribution (with unit variance). Eq. (6.6a) becomes, in reduced form with $H_{(\phi,\eta)}=\left[I_{Tn}-\phi\left(L_T\otimes I_n\right)-\eta\left(I_T\otimes W_{(1)}\right)\right]$

$$Y=H_{(\phi,\eta)}^{-1}\left[X\frac{1}{\sigma}\beta+\left(I_T\otimes W_{(2)}\right)X\frac{1}{\sigma}\gamma\right]+H_{(\phi,\eta)}^{-1}\frac{1}{\sigma}\left(\alpha+e\right) \tag{6.7}$$

Let $\upsilon=H_{(\phi,\eta)}^{-1}\left(\alpha+e\right)$. Then, the observable LDV $\tilde{y}_{it}$ takes on the value j if $\upsilon_{it}\in\left(a_{j-1},a_j\right]$. We assume, in contrast to the linear model, cross-sectional and serial independence of $e_{i,t}$ in order to reduce the complexity of the estimator. Though, we need to integrate over a nT-dimensional probability density function (PDF) $f\left(\upsilon_{1,1},\dots,\upsilon_{n,T}\right)$ assumed to be multivariate-Gaussian. Both the disturbances υ_{it} and the local scores $g_{it}\left(X,W;\theta\right)=\left\{H_{(\phi,\eta)}^{-1}\left[\mathrm{X}\frac{1}{\sigma}\beta+\left(I_T\otimes W_{(2)}\right)X\frac{1}{\sigma}\gamma\right]\right\}_{it}$ are cross-sectionally interdependent. Additionally, the local scores $g_{it}\left(X,W;\theta\right)$ entail data from all observations. When we abstract for notational simplicity from the time dimension and use only the n-dimensional PDF, and arrange the observations according to the values of the observed LDV $\tilde{y}_i$, the likelihood can be written as

$$\mathcal{L} = \underbrace{\int_{-\infty}^{a_1 - g_1(X,W;\theta)} \cdots \int_{-\infty}^{a_1 - g_{n_0}(X,W;\theta)}}_{i:\tilde{y}_i=1} \cdots \underbrace{\int_{a_{j-1} - g_{n_{j-1}+1}(X,W;\theta)}^{a_j - g_{n_{j-1}+1}(X,W;\theta)} \cdots \int_{a_{j-1} - g_{n_j}(X,W;\theta)}^{a_j - g_{n_j}(X,W;\theta)}}_{i:\tilde{y}_i=j}$$
$$\cdots \underbrace{\int_{a_{J-1} - g_{n_{J-1}+1}(X,W;\theta)}^{\infty} \cdots \int_{a_{J-1} - g_n(X,W;\theta)}^{\infty}}_{i:\tilde{y}_i=J} f(\upsilon_1,\ldots,\upsilon_n)\, d\upsilon_1 \cdots d\upsilon_n \tag{6.8}$$

In contrast to the case with independent observations, we are not able to split the likelihood into multiplicative-separable parts (or the log-likelihood into additive-separable). Hence, this likelihood is hardly solvable with conventional probabilistic (that is 'frequentistic') methods whereas Bayesian statistics may still provide a solution (see e.g. LeSage/Pace, 2009: Ch. 10; Franzese/Hays, 2009; Wang/Kockelman, 2009).[60]

A computationally simple approximation can be derived from a specification similar to eq. (6.4). Like in eq. (6.7), β and γ are identified only up to the scaling parameter σ. The serially and spatially lagged dependent (latent) variable is approximated by the serially lagged LDV and by the spatially weighted average[61] over the LDV, respectively. Since the re-scaled stochastic process is standard-normal by assumption, we divide these by the conditional standard deviation of $y_{i,t}$ estimated in the linear model (without fixed effects), such that both $\frac{1}{\sigma_{\tilde{y}}}\tilde{y}_{i,t-1}$ and $\sum_{j=1}^{n} w_{ij} \frac{1}{\sigma_{\tilde{y}}}\tilde{y}_{j,t}$ have unit variance. Then we estimate the latent process

$$y_{i,t} = \phi \frac{1}{\sigma_{\tilde{y}}}\tilde{y}_{i,t-1} + \eta \sum_{j=1}^{n} w_{(1)ij} \frac{1}{\sigma_{\tilde{y}}}\tilde{y}_{j,t} + x_{i,t}\frac{1}{\sigma}\beta + \sum_{j=1}^{n} w_{(2)ij}\, x_{j,t}\frac{1}{\sigma}\gamma + \frac{1}{\sigma}e_{i,t} \tag{6.9}$$

with ordered probit for independent observations while instrumenting for the spatially lagged LDV with the same internal instruments as in the linear case.

The condition for the identification of peer effects – existence of nonlinear exclusion restrictions – is the same as in the linear case. Furthermore, nonlinear estimation requires additional assumption (e.g. on standardization, or on independence of residuals) and computational challenges. Hence, we use the ordered probit estimations only as a robustness check with regard to the linear-form assumption.

60 We adapt the MATLAB code for a spatial probit model provided by Jim LeSage and employ a Gibbs sampler to draw from a Truncated Multivariate Normal (TMVN) distribution when generating the latent variable. To achieve a positive definite multiplier $H_{(\phi,\eta)}$, we restrict ϕ to the interval $\left[-\frac{1}{\varphi}, \frac{1}{\varphi}\right]$ with φ the largest eigenvalue of $I_n - \eta W_n$; to keep it finite, we replace the multiplier matrix $H_{(\phi,\eta)}$ by $\tilde{H}_{(\phi,\eta)} = \left[(I_T - \phi L_T) \otimes (I_n - \eta W_n)\right]^{-1}$ if $\left|H_{(\phi,\eta)}\right| < 10^{-6}$. However, since $(\hat{\eta}, \hat{\phi})$ frequently end up at the joint frontier of the parameter space and the estimates are not stable, we do not present results here.

61 We are aware that the (weighted) mean of the observed LDV is, in contrast to the median or any other quantile, not a well-defined statistic for categorial variables. Nevertheless, it might be a suitable approximation for the social expectation $E(y \mid p)$.

6.4 Expectation formation: Evidence for herding

6.4.1 Results

In the following we present results when estimating the linear model over the entire observation period. The focus of the analysis is the economic structure in expectation formation. Here we focus on the mere existence of herding and its distinction from other peer effects; enlightening the mechanism behind is postponed to a later section.

Table 6.3: Parameter estimates (distance)

Coefficient		Pure AR	X_t	X_t, MX_t	W_dY_t, X_t	W_dY_t, X_t, MX_t
α	cons	-0.013**	0.058***	0.120***	-0.008	0.005
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ϕ	AR(1)	0.838***	0.707***	0.539***	0.444***	0.445***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
β	U_{t-1}	–	-0.116*	0.595***	-0.070	0.311***
			(0.07)	(0.09)	(0.06)	(0.08)
	V_{t-1}	–	-0.639***	-0.151***	-0.087**	-0.091**
			(0.03)	(0.04)	(0.03)	(0.04)
γ	MU_{t-1}	–	–	-4.554***	–	-1.794***
				(0.23)		(0.26)
	MV_{t-1}	–	–	-2.617***	–	-0.647***
				(0.09)		(0.14)
η	SAR(1)	–	–	–	0.520***	0.459***
					(0.02)	(0.02)
μ	Fixed Eff.	yes [a]	yes [a]	yes [a]	yes [a]	yes [a]
	Time Eff.	no	no	no	no	no
δ	SAR(1)			0.328	-0.373	-0.336
ρ	AR(1)			-0.059	-0.032	-0.046
σ_u	(RMSE)	.517	.501	.473	.451	.449
Sargan test					45.461***	9.763*
Partial R^2					0.753	.669
Wu-Hausman					5.414**	1.288

W_d: Binary distance, row st. – M: inverse distance – WY_t instrumented with WX_t, W^2X_t, WY_{t-1}.

Standard errors in parenthesis – Asterisks mark standard significance levels (90%/ 95%/ 99%).

[a]: Region-specific fixed effects significant in ca. 10 regions.

The first four columns of Table 6.3 refer to estimations where the model components are included stepwise: we start with a pure autoregressive model, estimate then a model without peer effects (only conditioning on X_t and Y_{t-1}) and such where the peer-effect components (MX_t and WY_t) are included separately; we model the contextual effect by a continuous inverse-distance decay matrix (denoted with M) whereas we employ a row-standardized binary cut-off distance matrix W_d to describe the social-learning network. The last column in the table shows parameter estimates for eq. (6.4). This is our preferred specification because it ensures on the one hand identification of η, γ and δ in the light of Propositions 1 and 4 in Bramoullé/Djebbari/Fortin (2009) as $I_n, M, W_d, W_d M, W_d^2, W_d^3$ are linearly independent; on the other hand, W_d reflects that in densely populated regions more districts are within commuting distance and that a CEO is likely to communicate even with close CEOs without their regions being adjacent to each other. Estimations with alternative weights, modifications in the disturbance structure and nonlinear functional form accounting for the discreteness of Y (ordered probit) are reported in Table 6.4 in the robustness section.

A high AR term in general expresses that current disturbances or innovations have little effect on believes whereas there is long memory with regard to the history of innovations. The evidence for long memory with regard to errors can be interpreted twofold: as slow adaption to new information, or as strong effect of learning from personal mistakes, averaging over the entire history of errors. In the latter case, a moving-average process in the disturbance with a short lag length could be understood to describe CEOs repeating misbelief without personal learning. We find that the parameter estimate associated with the AR term declines only little when we include additional information on the market fundamentals: including local information results in a parameter shift of -0.13 (the difference between $\hat{\phi}$ in the first and second column of Table 6.3), consideration of contextual information lets the parameter shift by an additional -0.168. If we include the endogenous peer effect which accounts for social learning, the parameter estimate decreases again by an additional -0.095 down to 0.445. This AR parameter value is robust across most specifications which account for the endogenous effect (for social learning) and the local market fundamentals, regardless of the further information included in the estimation (only the estimates in the random-effects and in the ordered probit IV model, see Table 6.4, deviate significantly from that value). This remainder of the serially-autoregressive process seems to reflect the influence of signals received up to the previous period. Serial correlation in the residual (shown in the rows denoted with δ: AR(1) ε_t) which supposedly captures the short-run effect of personal learning in expectation formation seems negligible.

We would like to briefly discuss the signals which can be drawn from observable market fundamentals, i.e. the parameters in β and γ.[62] An increase in local unemployment seems throughout most models correspond to expectations of

62 In an earlier version of the paper with data until May 2011, we found parameter estimates that could be translated into a matching function with an elasticity with respect to labour-market tightness of approximately 0.3. The current estimates do not support such clear evidence of a matching function.

unemployment rising further on; with exception of the model including only local information and the own history, its effect is insignificant or significantly positive. With regard to the effect associated with local vacancies we can state that the sign is plausible insofar that unemployment is expected to rise when vacancies decline; furthermore the effect's height is stable across the models if we account at least for either the contextual or the endogenous peer effect. Vacancies in surrounding regions have an impact on unemployment sentiments which shows in the same direction. The size of the effect seems stronger at the first glance. However, the variation of this spatial average is much smaller than the variation of vacancies themselves, thus the larger size of the coefficient does not imply a relatively stronger impact. Contextual unemployment has, alike vacancies and in contrast to recent local unemployment, a negative sign. Declining unemployment in surrounding regions increases the probability that a CEO expects local unemployment to rise. This suggests a kind of crowding-out amongst the unemployed in close regions.

However, our central interest is on the endogenous peer effect, that is on the estimate for η; we interpret this as the effect of social learning (or herding). The estimate $\hat{\eta}$ in our preferred specification is 0.459. The corresponding estimates in the robustness checks are always significantly positive, mounting to approximately 0.4 to 0.5 throughout most specifications presented in Tables 6.3 and 6.4. A herding parameter of 0.459 corresponds to a social multiplier of $\frac{1}{1-\eta}=1.85$ (when abstracting from mis-specification due to negligence of the limited definition range of the dependent variable). The spectral norm of the corresponding spatial multiplier matrix $(I-\eta W)^{-1}$ has an almost identical value; it is 1.86. That is, the direct signal extractable from variation in each variable in our preferred model is approximately doubled due to social interaction as follows from eq. (6.3); without social learning, the impulse of a market fundamental would require twice its amount to cause the same effect on the expectations.

Note that the parameter can be considered identified in terms of IV estimation in our preferred specification since the partial R^2 of the excluded instruments is relatively high whereas the Sargan statistics is just weakly significant. The Sargan test statistics declines even further when we do not use the second-order lags of unemployment and vacancies which are only week instruments in our preferred specification and which we only include for better comparison with the models in the robustness section. In contrast, the Sargan test rejects validity of the instruments in the model without contextual effect. This emphasizes that the contextual variables should be included in the second stage equation explaining Y_t and not only in the first stage explaining WY_t.

6.4.2 Robustness

In the following, we present a number of robustness checks. Table 6.4 report estimations of the linear model according to eq. (6.6a) where we use alternative networks (spatial structures) in the endogenous and the contextual effect, modify the disturbance structure, or employ an ordered probit rather than the linear specification.

The regressions in the first five columns differ from the preferred model with regard to the spatial weights matrices. The parameter estimates do in general not deviate strongly from the corresponding estimates in the last column of Table 6.3. The parameters in the contextual effect have smaller size when we use a row-normalized binary matrix (W_c, W_d, W_p) instead of the continuous distance-decay matrix M. This result is rather intuitive since the discrete weighting schemes result in less smooth network averages, i.e. show higher variation than those generated with m_{ij} as weights (though still less variation than the corresponding variable itself). The autoregressive parameters in these five models are not significantly different from our preferred specification. In contrast to this, the herding parameters show some deviations, even though the values are not far apart, in the range between 0.381 and 0.512. The major difference across the first five models is in the validity of instrumentation. The Sargan tests tend to be rejected if we use discrete geographical relations for both the endogenous and the contextual effect.

Columns six and seven show estimates in which we used alternative disturbance structures. The two-way error-component model with individual and time-specific fixed effects controls for cross-sectional error correlation by ruling out the time-specific average disturbance or the average factor dependence (since $\hat{\mu}_t = \sum_{i=1}^{n} \lambda_i f_t$). Significance of fixed effects in eight to ten from 176 regions – that is region-specific effects not deviating significantly from the average in more than 90% of the regions – suggests that excluding region dummies from estimation, or estimation with random effects will not cause serious bias. Indeed, most parameters are not too different. The major difference can be observed in the herding parameter which is significantly smaller than 0.459 in both estimations; nevertheless, both estimates for η are still significantly positive. Even if we control for cross-sectional correlation in a very rigorous way by time-specific dummies the estimate for the herding parameter mounts to a significantly positive value of 0.287.

The ordered-probit estimation – which is identified under the assumption that the standard-deviation of the discrete dependent variable conditional on X is 0.449 (the estimate for $\sigma_{\tilde{y}|X} = \sigma_u$ in our preferred model) – deviates from the linear estimations with regard to two points. First, we find the highest estimates for both the serially and the cross-sectionally autoregressive parameters, ϕ and η, amongst all models accounting for the entire information about local and contextual market fundamentals. Second, the sign associated with vacancies in surrounding regions reverses; for us, it seems implausible that increasing vacancies should be associated with unemployment expected to rise.

Table 6.4: Parameter estimates under alternative weights matrices and disturbance structures

		geographical contiguity		Shared benchmarking class		Distance	Ind.+Time Eff.	Rand. Eff.	Ord. Probit (IV)[b]
Coefficient		W_cY_t, X_t, MX_t	W_cY_t, X_t, W_cX_t	W_pY_t, X_t, MX_t	W_pY_t, X_t, W_pX_t	W_dY_t, X_t, W_dX_t	W_dY_t, X_t, MX_t	W_dY_t, X_t, MX_t	(bootstrap s.e.)
ϕ	AR(1)	0.458 ***	0.463 ***	0.455 ***	0.461 ***	0.445 ***	0.437 ***	0.517 ***	0.599 ***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
β	U_{t-1}	0.336 ***	0.192 *	0.307 ***	0.154 *	0.272 ***	0.336 ***	0.376 ***	0.079 **
		(0.08)	(0.10)	(0.08)	(0.09)	(0.10)	(0.09)	(0.08)	(0.04)
	V_{t-1}	-0.099 ***	-0.116 ***	-0.086 **	-0.102 ***	-0.095 ***	-0.087 **	-0.088 ***	-0.109 ***
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
γ	WU_{t-1}	-2.317 ***	-0.648 ***	-2.155 ***	-0.572 ***	-0.632 ***	-1.335 ***	-2.028 ***	-0.440 ***
		(0.26)	(0.13)	(0.25)	(0.13)	(0.14)	(0.46)	(0.25)	(0.07)
	WV_{t-1}	-1.020 ***	-0.235 ***	-0.915 ***	-0.172 **	-0.109	-0.624 ***	-0.713 ***	0.569 ***
		(0.13)	(0.07)	(0.13)	(0.08)	(0.08)	(0.19)	(0.13)	(0.10)
η	SAR(1)	0.381 ***	0.453 ***	0.410 ***	0.480 ***	0.512 ***	0.287 ***	0.381 ***	0.542 ***
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.02)	(0.02)
δ	Fixed Eff.	yes [a]	yes [a]	yes [a]	yes [a]	yes [a]	yes [a]	no	–
	Time Eff.	no	no	no	no	no	yes	no	–
	SAR(1)	-.276	-.351	-.150	-.229	-.376	-.336	-.186	–
	AR(1)	-.039	-.039	-.039	-.04	-.035	-.046	-.107	–
σ_u	(RMSE)	0.454	0.456	0.447	0.449	0.451	0.441	0.461	–
N		7216	7216	7216	7216	7216	7216	7216	7216
Partial R^2		.480	.562	.610	.623	.613	.301	.599	.554
Sargan test		2.134	15.340 ***	5.041	5.959 *	22.418 ***	10.091 *	4.677	
Wu-Hausman		6.076 **	13.514 ***	11.976 ***	9.811 ***	3.434 *	21.878 ***	0.176	–

W_c: Binary contiguity, row st. – W_p: Binary, same SGB3-benchmark type, row st. – W_d: Binary distance, row st. – M: inverse distance (continuous) – WY_t instrumented with WX_t, W^2X_t, WY_{t-1}.

Standard errors in Parenthesis – Asterisks mark standard significance levels (90%/ 95%/ 99%).

[a]: Region-specific fixed effects significant in ca. 10 regions.

[b]: Threshold values between categories at -1.975, 1.249, 3.514 and 6.049.

6.5 Social learning or joint adaption to news?

Carroll (2003) presents a model where the average sentiment at time t on unemployment in $t + h$, $E(Y_t[u_{t+h}]) = \bar{Y}_t[u_{t+h}]$ in a survey of interest (in our case the sentiment Y_t refers always to 3-months ahead unemployment u_{t+3}) depends to a fraction λ on current-period public information $N_t[u_{t+h}]$ (e.g. published in newspapers) and to an amount $1-\lambda$ to prospective information which has already been available in the previous period. Recursion leads to the average expectation

$$\begin{aligned}\bar{Y}_t[u_{t+h}] &= \lambda N_t[u_{t+h}] + (1-\lambda)(\lambda N_{t-1}[u_{t+h}] + (1-\lambda)(\lambda N_{t-2}[u_{t+h}] + \ldots)) \\ &= \lambda N_t[u_{t+h}] + (1-\lambda)\bar{Y}_{t-1}[u_{t+h}]\end{aligned} \tag{6.10}$$

Carroll (2003) interprets λ as the fraction of a population that receives news. However, the model can be adapted to individual sentiments wherein λ reflects an individual's probability to get new information rather than the informed-population's share; Carroll's population model can be derived by averaging across independent individuals. Public information can be integrated into a simple model of expectation formation in addition to market fundamentals and other variables. This allows us to identify the partial contribution of public information to individuals' sentiments.

Hence, in order to compare the effect of public information and social learning, we estimate two equations (both two times, either with public forecasts released in the current or with those published in the previous month): In the first, denoted with PI, we explain current unemployment expectations by the local log unemployment stock and log vacancies observed in the previous period, lagged expectations and by the news regarding unemployment in the future.[63] In the second (PI+SL), we augment this model by the contextual and the endogenous peer effect, that is by spatially lagged market fundamentals and the spatially lagged expectations; with regard to public information, there won't be a contextual peer effect since it is in principle observable to all persons in the survey. The latter model with contemporaneous news can be written as

$$y_{i,t} = \phi y_{i,t-1} + \eta \sum_{j=1}^{n} w_{(1)ij} y_{j,t} + \alpha + x_{i,t}\beta + \sum_{j=1}^{n} w_{(2)ij} x_{j,t}\gamma + \lambda N_t + e_{i,t} \tag{6.11}$$

63 The institutes do not forecast unemployment at the same horizon and the same frequency that we have in the FEA management survey. Hence, we construct our 'public information variable' as follows: We first average forecasts published in the same month; in each month, we then consider only those forecasts published latest (if no new forecast is available, we extrapolate the average forecast from the previous month). Since most institutes publish forecasts for the current and the subsequent year, we use in the first six months of a year the forecast for the respective year and from July to September a weighted average between the forecasts regarding current (linearly declining weight) and next year (linearly increasing weight).

instead of eq. (6.4); for PI, restrict $\eta = 0$. Note that in this model the parameter associated with the public-information variable will only reflect the probability that a person will directly notice the new information; this effect may be amplified by social learning from informed persons. Results are reported in Table 6.5; for means of comparison we add the estimates of our preferred model from Table 6.3.

Table 6.5: Estimation: Public information vs. social learning

Param.	Variable	Herding (SL)	Public Info (PI)		Combined (PI+SL)	
α	cons	0.005	-0.724 ***	-0.412 ***	-0.107	0.091
		(0.01)	(0.12)	(0.12)	(0.12)	(0.11)
ϕ	AR(1)	0.445 ***	0.526 ***	0.529 ***	0.444 ***	0.445 ***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
β	U_{t-1}	0.311 ***	0.594 ***	0.617 ***	0.315 ***	0.306 ***
		(0.08)	(0.09)	(0.09)	(0.08)	(0.08)
	V_{t-1}	-0.091 **	-0.138 ***	-0.146 ***	-0.090 **	-0.091 **
		(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
γ	MU_{t-1}	-1.794 ***	-4.379 ***	-4.731 ***	-1.816 ***	-1.750 ***
		(0.26)	(0.23)	(0.23)	(0.26)	(0.27)
	MV_{t-1}	-0.647 ***	-2.274 ***	-2.497 ***	-0.631 ***	-0.643 ***
		(0.14)	(0.10)	(0.09)	(0.14)	(0.14)
η	SAR(1)	0.459 ***			0.451 ***	0.463 ***
		(0.02)			(0.03)	(0.03)
λ	N_t		0.251 ***		0.036	
			(0.04)		(0.04)	
	N_{t-1}			0.157 ***		-0.017
				(0.03)		(0.03)
δ	Fixed Eff.	yes	yes	yes	yes	yes
	SAR(1)	-.319	.324	.326	-.295	-.317
	AR(1)	-.037	-.049	-.054	-.036	-.037
RMSE		0.449	0.472	0.473	0.449	0.449
Partial R^2		.554	–	–	.525	.533
Sargan		9.763 **	–	–	9.486 *	9.891 **
Wu-Hausman		1.288	–	–	0.862	1.512

Standard errors in parenthesis – Asterisks mark standard significance levels (90%/ 95%/ 99%).

The spatially lagged dependent variable WY_t instrumented with $WX_t, W^2X_t,\ WY_{t-1}$.

We find that the effect of contemporaneous public information is in general larger than the effect of the one-month lagged information; however, both are significantly positive. The size of the estimate $\hat{\lambda} = 0.251$ in the second column of Table 6.5 – the estimation which is most similar to eq. (10) in Carroll (2003) – is even smaller than his estimate for the effect of public information (which mounts to 0.36), and far from the corresponding estimate (mounting to 0.851) in Curtin (2003) which refers as well to individual rather than averaged expectations.

The effect vanishes completely when we account for herding. Note that the two parameters η and λ reflect only the partial effects of social learning and public information. The estimates for the herding parameter have a similar size as in our preferred model from Section 6.4.2 (without public information); in contrast, $\hat{\lambda}$ is not significantly different from zero. If we use lagged news rather than contemporaneous, we get the same result. This indicates that most of the (direct) effect which had been attributed to public information before is in fact due to social learning. However, the spillover of information through social networks even works as an amplifying device for public information.

6.6 Is there an informational cascade?

So far we have found evidence for a significant amount of herding in the unemployment expectations reported in the FEA management survey. The exact nature is however unclear: Does social learning guide a herd towards the correct outcome as it could be expected because market fundamentals form continuous signals (Smith/Sørensen, 2000) or because repeated announcements reduce the information set (Manski, 2004)? Or does it behave like an informational cascade (Bikhchanandi/Hirshleifer/Welch, 1992, 1998; Chamley, 2004) in which a herd may converge not only to a correct but even to an incorrect herd belief (or at least remain at an incorrect expectation for a long time and converge just slowly)? The two are different insofar that private information becomes at most irrelevant within a cascade, and that a cascade is fragile:[64] It can be broken by a person with strong believe in her private information or by public information made available after the cascade has started. We will employ this relation – public announcement that the economy develops in a new direction which supposedly causes a drastic change of individual believes with regard to future unemployment (without sharp changes in real unemployment/vacancies data) – to investigate the existence of an informational cascade.

64 Bikhchanandi/Hirshleifer/Welch (1992) write that "conceptually, [their] paper differs from Welch's and Banerjee's in emphasizing the fragility ... cascades can explain not only uniform behavior but also drastic change such as fads."

The theoretical concept of an information cascade can be modeled empirically as a pair of structural breaks: Before the arrival of new external information, private information is of little importance whereas social learning should be dominant (that is, η should be high). During the period τ_1 in which, supposedly, public information becomes available and the subsequent adaption takes place, social learning pales in comparison to unobservable but even observable private information. The adjustment process presumably lasts for more than a month but not too long, somewhat between a quarter or half a year. In the final stage (starting in τ_2) a new informational cascade establishes. Social learning again dominates the direct influence of a CEO's own private information. However, the level of the social equilibrium, that is the average expectation, should have changed between the initial and the new cascade. In addition, the residual dispersion should be wider in the period when public information arrives since it reflects not only the innovations themselves but also the uncertainty about information. Thus, we model the two structural breaks explicitly and estimate the following equation without fixed effects[65]:

$$\begin{aligned} y_{i,t} = e_{i,t} &+ \mathbf{I}(t<\tau_1)_{i,t}\left(\alpha_1 + x_{i,t}\beta_1 + \sum_{j=1}^{n} m_{ij}x_{j,t}\gamma_1 + \sum_{j=1}^{n} w_{ij}y_{j,t}\eta_1 + \phi_1 y_{i,t-1}\right) \\ &+ \mathbf{I}(\tau_1 \le t < \tau_2)\left(\alpha_2 + x_{i,t}\beta_2 + \sum_{j=1}^{n} m_{ij}x_{j,t}\gamma_2 + \sum_{j=1}^{n} w_{ij}y_{j,t}\eta_2 + \phi_2 y_{i,t-1}\right) \\ &+ \mathbf{I}(\tau_2 \le t)\left(\alpha_3 + x_{i,t}\beta_3 + \sum_{j=1}^{n} m_{ij}x_{j,t}\gamma_3 + \sum_{j=1}^{n} w_{ij}y_{j,t}\eta_3 + \phi_3 y_{i,t-1}\right) \end{aligned} \tag{6.12}$$

We consider January 2010 as the month when new information is published, since at this date the first institute revised its unemployment forecast from *rising* to *remaining equal*, see Section 6.2. Results for eq. (6.12) estimated analogously to our preferred specification (albeit with a reduced set of instruments) are shown in Table 6.6.

We find that information/belief persistence does not change over time: we can not reject that $\hat{\phi}_1 = \hat{\phi}_2 = \hat{\phi}_3$ at reasonable significance levels. The influence of market fundamentals is at most equal between period one (before the first break) and period three (after the second break); only the influence of local unemployment changes substantially. In contrast, most parameters associated with unemployment or vacancies are significantly different between periods two and three. The parameter estimates determining the average sentiment ($\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$) are not significantly different from zero. However, equality of $\hat{\alpha}_2$, $\hat{\alpha}_3$ with 0.079, the

65 Because of OLS or random effects estimation, the dynamics parameters will be biased upwards with the size of bias depending on the ratio between σ_μ^2 and σ_u^2 (the variances of the time-invariant and time-varying error component) but not on T; see Sevestre/Trognon (1985).

point estimate of α_1, can be rejected at least weakly. In accordance with our theoretical considerations regarding an informational cascade, social learning is less important in the break period: $\hat{\eta}_1$ and $\hat{\eta}_3$ are significantly positive, in contrast to $\hat{\eta}_2$. Moreover, $\hat{\eta}_2$ is smaller than the further at the 90% confidence level. Thus, we observe social learning both before the collapse and a while after once a new cascade might have established, but find hardly evidence for herding at the time of the potential collapse and short after.

Table 6.6: Informational cascade – Structural break estimation

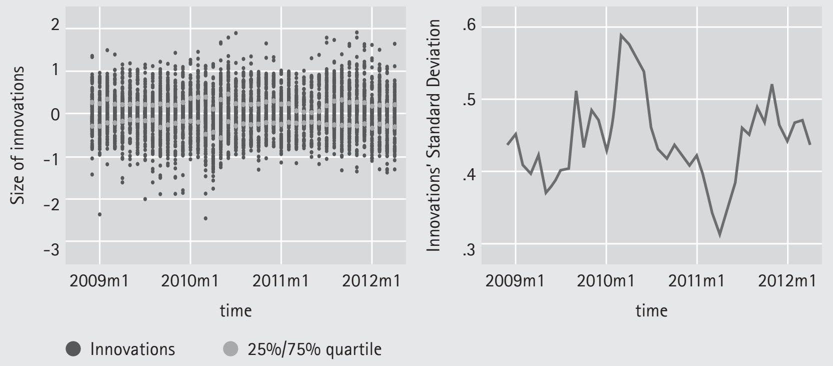
Parameter	Variable	$t < 01/2010$	01–06/2010	$07/2010 \leq t$
$\alpha_{I(t \in [t_1, t_2))}$		0.079	-0.087	-0.088
		(0.09)	(0.08)	(0.08)
$\phi_{I(t \in [t_1, t_2))}$	AR(1)	0.442 ***	0.412 ***	0.446 ***
		(0.02)	(0.02)	(0.02)
$\beta_{I(t \in [t_1, t_2))}$	U_{t-1}	0.575 ***	0.436 *	-0.040
		(0.14)	(0.23)	(0.15)
	V_{t-1}	-0.063	0.005	-0.139 **
		(0.06)	(0.09)	(0.06)
$\gamma_{I(t \in [t_1, t_2))}$	MU_{t-1}	-1.842 ***	2.042	-1.457 **
		(0.44)	(1.52)	(0.66)
	MV_{t-1}	-0.421	-1.426 ***	-0.509 **
		(0.26)	(0.64)	(0.22)
$\eta_{I(t \in [t_1, t_2))}$	SAR(1)	0.414 ***	0.110	0.393 ***
		(0.06)	(0.14)	(0.05)
IV statistics (for instrumenting WY_t with WX_t, WY_{t-1}):				
Partial R^2		0.561	0.161	0.446
Sargan test		J = 7.638 ~ χ^2_6		(p-val. = 0.266)
Wu-Hausman		H = 5.975 ~ $F_{3,7017}$		(p-val. = 0.000)

Standard errors in parenthesis. Asterisks mark standard significance levels (90%/95%/99%).

Estimation deviates from baseline model with regard to instruments and error components: carried out without fixed effects and without second-order spatial lags of exogenous variables.

To illuminate the behaviour of unobservable information, we look at the distribution of innovations (or disturbances) resulting from estimation of eq. (6.12). The left panel in Figure 6.3 plots the disturbances by month, with the first and third quartile highlighted. The right panel reports the innovations' time-specific standard deviation. Both figures show that the distribution becomes slightly wider dispersed in the first half-year of 2010: The standard deviation exceeds 0.5 in more than a single month only from February till May 2010. I.e., unexplained variation is stronger, private (unobserved) signals are more outstanding in this period than before January 2010 or after June 2010.

Figure 6.3: Dispersion of innovations over time



Our findings give in general support towards the fragility of herding amongst CEOs, that is, towards the existence of informational cascades. At the time when new information regarding oppositely directed future development of the labor market is published, social learning becomes insignificant. Then, the herding parameter is significantly smaller than in the periods both before that date and some months later. As well, the dispersion of (only ex-post measurable) private signals becomes wider.

However, it might be difficult to identify (or reject) the existence of a cascade in real-world data because of various reasons. On the one hand, achieved statistical significance is affected by a longer observation period since parameters converge over both n and T. Hence, the parameters corresponding to the short period between two cascades are necessarily estimated relatively imprecise.

On the other hand, shifts in the expectations might not only be due to the arrival of public *information* on a trend reversal and the subsequent shift in social beliefs but also due to the trend reversal itself, that is due to a shift of the world's *true state* (e.g. the shift from a recession to an upturn). Then, breaks in the parameters would result from the new correct value and not from the new information. Nonetheless, it is unlikely that a new correct value of unemployment has formed exactly in early 2010 because of three reasons: first, unemployment declined, seasonally adjusted, at a smooth rate in the time between Spring 2009 and January 2012 – that is, we do not observe a trend reversion in realized unemployment. Second, the quarterly GDP growth rate has been in a range between 0.5 and one from the second quarter 2009 till 2011, with exception of the second quarter 2010 where it mounted to approximately two.[66] And third, published GDP

66 See Federal Statistical Office of Germany, 2012 and the Federal Statistical Office's press release nr. 277, August 14th 2012.

growth rate forecasts were revised substantially in and after June 2010, from values between one and two percent to more than three percent. Research on Okun's Law in Germany from the 1990s (though admittedly outdated) assigns the unemployment threshold of the output gap to GDP growth rates between 1.5 and two percent. According to this theory, unemployment should have remained stable in the first half of 2010, a further decline in unemployment expectable only in late 2010. If the reversion of sentiments would be due to a change in the real world, the direction of herding would either have followed the correct value's shift with a delay of three quarters, or it would have anticipated this shift. Hence, a causation of the sentiments' reversion by a shift in the published unemployment forecasts, that is by information, is more plausible.

6.7 Are sentiments more rational due to herding?

Another topic of interest is whether the expectations are rational (or realistic), and whether social learning contributes to make sentiments more realistic: Despite public conviction, herding might contribute to the rationality of expectations through social aggregation of information. In order to analyse if the forecasts deviate systematically from the bisector in a prediction-realization diagram – as it has been proposed by Mincer/Zarnowitz (1969) besides several other criteria for the evaluation of expectations and forecasts – the realized values u_{t+h} are regressed on the forecasts $Y_t\left[u_{t+h|\mathcal{I}_t}\right]$ and a constant (maintaining our previous notation):

$$u_{t+h} = \vartheta_1 + \vartheta_2 \; Y_t\left[u_{t+h|\mathcal{I}_t}\right] + \upsilon_{t+h} \tag{6.13}$$

The estimated slope $\hat{\vartheta}_2$ is tested against the value 1, the intercept $\hat{\vartheta}_1$ against 0. Here, we rely on a similar concept. A first graphical inspection of the relation between sentiments as reported in the survey and realized unemployment development has been provided in Section 6.2. The estimates from a Mincer-Zarnowitz (MZ) regression will be shown subsequently.

However, for answering the question in this section, this test will have only limited use when applying it directly to the sentiments in the survey. The results in the previous sections have provided evidence for social learning in the formation of sentiments. Hence, in order to analyze whether social learning really contributes to more rational expectations we need to construct counterfactual sentiments in which we assume isolation of agents. In these hypothetical sentiments the effect of social learning is eliminated. To get a direct counterpart to this counterfactual, we construct as well hypothetical sentiments with social learning. For both, we use the estimates of eq. (6.4) in our preferred specification, and in both we include even the estimated innovations $\hat{e}_{i,t}$ as a measure for the private signals. Thus, we consider the following prediction as our hypothetical sentiments under social learning:

$$Y_{i,t}^{(s)}\left[u_{t+h|X_t,MX_t,WY_t}\right]=\left\{(H_{\hat{\eta},\hat{\phi}})^{-1}\left(\iota_{Tn}\hat{\alpha}+X\hat{\beta}+(M_n\otimes I_T)X\hat{\gamma}+\hat{e}\right)\right\}_{i,t} \tag{6.14}$$

The counterfactual under isolation uses the same parameter estimates. The social multiplier $(I-\hat{\eta}W)^{-1}$ which represents the simultaneous interaction between responses is however replaced by its spectral norm, a representation of the scale but not the interaction of the social multiplier. That is, the remaining score (which still accounts for a contextual peer effect) is multiplied with a constant such that the variance of the predictor under isolation is comparable to the predictor under social learning:

$$\begin{aligned}Y_{i,t}^{(i)}\left[u_{t+h|X_t,MX_t}\right]=\Big\{&\|(I_n-\hat{\eta}W_n)^{-1}\|\cdot\left(I_n\otimes(I_T-\hat{\phi}L_T)^{-1}\right)\\&\times\left(\iota_{Tn}\hat{\alpha}+X\hat{\beta}+(M_n\otimes I_T)X\hat{\gamma}+\hat{e}\right)\Big\}_{i,t}\end{aligned} \tag{6.15}$$

In the presented results, both hypothetical expectations $Y^{(i)}\left[u_{t+h|X,MX}\right]$ and $Y^{(s)}\left[u_{t+h|X,MX,WY}\right]$ are rounded to the next integer in [-2.2]; that is, they have the same ordinal scale as the expectations from the survey. The two ordinal hypothetical expectations are then used as regressors in MZ regressions. We will consider social learning as improving forecast accuracy if the slope in the regression with the 'hypothetical herding' is significantly closer to 1 than the slope coefficient under 'hypothetical isolation' (without the intercept significantly more distinct from 0), or vice versa.

The question in the survey refers to unemployment's development within the next three months (besides the typical seasonal development). Thus, we use the three-month difference between annual growth rates of unemployment three months ahead and today, $u_{t+3}=\Delta\left[\frac{(U_{t+3}-U_{t-12+3})}{U_{t-12+3}}-\frac{(U_t-U_{t-12})}{U_{t-12}}\right]\cdot 100$ to construct the realization variable for the MZ regression. However, since the expectation variables – both the survey responses and the hypothetical expectations – are ordinal, we use an ordinal scale (likewise coded from – 2 to + 2) where we define a growth-rate difference (in percentage points) in the interval (-2.5, +2.5) as no-change, in the interval from (–12.5, 2.5] and [2.5, 12.5), respectively, as moderate change (decline or growth), and growth-rate differences up to –12.5 (or above 12.5 percentage points) as strong change.

Table 6.7: Realized development vs. hypothetical sentiments

$\tilde{u}$	Social Learning $Y^{(s)}[u\|_{X, MX, WY}]$					Isolation $Y^{(i)}[u\|_{X, MX}]$					Sum
	-2	-1	0	1	2	-2	-1	0	1	2	
2	1	34	5	106	**143**	6	30	6	113	**134**	289
1	66	802	254	**435**	276	104	750	289	**411**	279	1,833
0	135	1,135	**580**	449	77	154	1,059	**682**	383	98	2,376
-1	50	**594**	766	397	46	73	**564**	793	352	71	1,853
-2	**0**	21	205	102	9	**2**	26	180	100	29	337
	252	2,586	1,810	1,489	551	339	2,429	1,950	1,359	611	

Table 6.7 crosses the two hypothetical expectation with the realized values; the values on the bisector are in bold letters. At the first glance, there is hardly any relationship detectable between hypothetical expectations and realized values – whereas we can replicate the survey responses with our predictions (under herding) pretty well. If at all, the cross-tables hint at a more or less horizontal pattern, that is sentiments which are uncorrelated with realized unemployment development.

The results of the corresponding MZ regression are shown in Table 6.8. The sentiments – both the survey responses and the two hypothetical sentiments – are far from perfect as can be seen from Table 6.8 and as we have already supposed for the survey responses in Sec. 6.2. The MZ hypothesis is rejected in each case. The slope in the two regressions with the discrete dependent variable is positive but closer to zero than to one. However, the slope under hypothetical social learning is significantly steeper than the slope under hypothetical isolation.

Table 6.8: Mincer-Zarnowitz regressions

Coefficient	Survey responses	Hypothetical social learning	Hypothetical isolated
	Realized growth (ordinal)		
Intercept $\hat{\vartheta}_1$	-.024	-.009	-.011
	(.01)	(.01)	(.01)
Slope $\hat{\vartheta}_2$	.045	.105	.085
	(.01)	(.01)	(.01)
Robust standard errors in parentheses.			

A horizontal line would express that there isn't, on average, any relation between realized values and expectations at all; a slope of one would indicate unbiased (that is rational) expectations. Since the slope of the hypothetical sentiments under social learning is closer to one than the hypothetical sentiments under isolation we conclude that social learning improves the rationality of sentiments.

6.8 Conclusion

In this paper we have analysed unemployment expectations with a particular focus on social learning in expectation formation. A novel survey amongst the CEOs of the local departments of the German Federal Employment Agency allows us to employ geographical structure and organizational networks in order to discriminate between close and less-close peers in communication and social learning. Thus, we have been able to deal with the reflection problem in the empirical analysis of social networks which allowed us to identify the effect of social learning and herding in the formation of unemployment expectations.

We have presented evidence for the influence of socially aggregated expectations in individual's expectation formation in Section 6.4; the results have been robust across various specifications. The estimate for the social multiplier in our preferred specification mounts to 1.86; that is, social interaction approximately doubles the impact on the average expectation which we have assessed directly to observable or private information. We have found that this effect does persist with a similar size when we account for public information on unemployment forecasts. The social multiplier still mounts to approximately 1.5 if we account for other contemporaneous effects in a rigorous way.

We have only to some extent been successful in detecting the nature of social information aggregation inherent in the survey's responses. First, we have rejected that the CEOs' sentiments only mimic unemployment forecasts published by professional forecasters in Germany; the estimated contribution of herding in sentiment formation is robust against controlling for the latter. Second, we have found a shift in believes accompanied by higher private uncertainty (or wider dispersion of unobservable private signals) and increased impact of observable information in the first half-year of 2010 – in succession to the economic research institutes' predictions that Germany has passed most of the crisis. This pattern hints at the existence of an informational cascade in the CEOs' announced unemployment expectations. However, our last result – that expectations do at least weakly become more realistic due to herding – is not systematically conformable with cascading of sentiments. The latter provides good news for economic tendency surveys amongst experts insofar that, despite the threat of a misleading cascade, they seem to aggregate information efficiently.

In future research – if the data cover more than one business-cycle turn and more than one expectation reversal – it might become possible to reject the cascading hypothesis, to provide more robust evidence for the informational efficiency of social learning, or to identify the origin of a herd: be it by pursuing a similar strategy to ours, or (if the data is much longer) by making a more detailed use of the timing of events. The evidence for social learning in unemployment expectations could be strengthened on the one hand if we can ascertain a similar endogenous peer effect in the further questions enclosed in the survey; supposedly, social learning amongst the same agents covers more than one issue. On the other hand, it would be interesting to verify that the responses are honest, i.e. that the expectations are followed by the adequate actions. However, both is beyond the scope of this paper. Finally, when setting up other surveys on expectations, in particular amongst experts or other small groups with a high probability of interaction, the information extracted from their responses may be improved by adequately accounting for social learning. For this, it is however necessary not only to ask for their expectations but even to collect (and provide) some information on their networks.

6.A Additional information regarding the survey

Table 6.9: Availability of questions in the FEA Management Survey (4/2012)

Question	Availability	Items/Scale
■ How do you expect unemployment in your district to develop within the next three months? (besides the usual seasonal pattern)	11/2008 – 4/2012	5-item Likert
■ Have more or less lay-offs been announced by the employers?[a]	11/2008 – 1/2011	5-item Likert
■ How do you expect employment in your district to develop within the next three months? (besides the usual seasonal pattern)	2/2011 – 4/2012	5-item Likert
■ Have more or less lay-offs of contingent workers been announced by the employers?[a]	11/2008 – 6/2010	5-item Likert
■ How do you expect employment in contingent work in your district to develop within the next three months (besides the usual seasonal pattern?)	2/2011 – 4/2012	5-item Likert
■ Do you observe an increase in the demand for contingent workers?	9/2009 – 6/2010	yes/no
■ Do you observe more lay-offs or more job creation in contingent work?[a]	7/2010 – 1/2011	5-item Likert
■ Do you observe lay-offs of workers subsequently to support by “reduced hours compensation”? If yes, how many?	3/2009 – 5/2010	yes/no (+ number)
■ Do you need to give more advise regarding “reduced hours compensation”?[a]	11/2008 – 6/2010 11/2011 – 4/2012	5-item Likert
■ Do you need to give more advise regarding “transitional companies”?[a]	11/2008 – 4/2012	5-item Likert
■ Do more or less employees in your district have contracts with “transitional companies”?[a]	11/2008 – 4/2012	5-item Likert
■ Do you observe excess demand for specialist workers? (If yes, in which occupations?)	7/2010 – 10/2011 01/2012	yes/no (+ text field in some waves)
■ Do you observe more intra-firm transitions from vocational training to regular work?[a]	7/2010 – 10/2010 2/2011 – 10/2011 2/2012 – 4/2012	5-item Likert

[a]: (compared to one year before)

The five item Likert scale is centered around zero; in general a value of −2 corresponds to the answer ‘declines strongly’, a value of +2 to the answer ‘increases strongly’. An exception is the question ‘Do you observe more lay-offs or more job creation in contingent work’ where we coded ‘much more job creation’ with a value of −2 and ‘much more lay-offs’ with +2.

Chapter 7

Resume

The aim of this thesis has been to contribute to closing the gap between economic theory and econometric practice in the analysis of regional labour market dynamics. Nowadays, economic theory on regional labour markets agrees that prices and quantities are cross-sectionally dependent between local labour markets, and that a local shock (e.g. on productivity) will result in spatially heterogeneous effects on labour supply, labour demand and the wage level (as well as land prices) if labour is neither perfectly mobile nor perfectly immobile, that is if transport/commuting/relocation costs are neither negligible nor infinite.

I have briefly sketched a variety of econometric models dealing with these spatial dependence and heterogeneity structures. As I have argued, the spatial econometric mainstream models are advantageous compared to other approaches if they are correctly specified. Moreover, it is necessary to employ a spatial lag or spatial Durbin model if the focus is on estimating the respective autocorrelation parameter or if research demands the calculation of spatial impact measures or a spatial multiplier. However, as I have discussed, identification of the parameters depends crucially on restrictions which should be made explicit. A second disadvantage of the mainstream models is in their sensitivity regarding the spatial weights employed in estimation. Furthermore, the heterogeneity in the impact measures follows in general the major patterns in the geography described by the spatial weights: Because the spatial multiplier smooths local variation, only a limited amount of effect heterogeneity can be represented by spatially autoregressive models.

The disadvantages may be overcome by a number of nonparametric methods relying either on geographic kernels or on eigendecomposition statistics. On the one hand, they allow for inferential procedures which are either robust across various alternative spatial weights or do not utilize these weights at all. On the other hand, the nonparametric approaches are more flexible with regard to the geographic pattern they are accounting for; they can deal with a higher degree of effect heterogeneity. Hence, the criticism on the spatial econometric mainstream (raised prominently by Gibbons/Overman, 2012) suggests that it might be a superior research strategy to employ some of these alternative approaches if the focus of enquiry is not on identifying a well defined spatial autocorrelation parameter or a specific impact measure.

When analysing the persistence of shocks in local unemployment rates, it is first and foremost important to find a less biased, more efficient (and hopefully more parsimonious in regard to the number of parameters) way to account for the heterogeneity in the adjustment parameters which shows up in region-specific estimations. Using a spatial lag to model the heterogeneity of effects would insufficiently account for the existing variation; it would overestimate the nation-wide heterogeneity and underestimate the local one. It is furthermore important to reduce the amount of cross-sectional dependence so as to get valid inference. For this, estimation of a spatial autocorrelation parameter is irrelevant; nonparametric methods will do the same (if not a better) job under less restrictive assumptions. Utilization of spatial regimes and of an eigenvector-based spatial filter interacted with the serial lagged dependent variable in a dynamic SFGWR approach allows the illumination of regional patterns in unemployment's adjustment to shocks. We find that a large extent of effect heterogeneity can be accounted for by differentiating regions according to their settlement structure; core regions show higher shock persistence than peripheral ones. Likewise, the spatial patterns in persistence are represented best by a selection of eigenvectors corresponding to regional patterns (showing 3-5 peaks and 3-5 troughs when plotted in a map); this again hints at persistence (or hysteresis) as a regional phenomenon.

In the forecasting study I analyse the predictive content of accounting for the spatial co-development of regional labour markets. From a model building perspective I would denote the employed Global VAR model as a model with spatially lagged exogenous (predetermined) variables for which the parameters are allowed to be heterogeneous. Temporally lagged local averages across the neighbouring regions assess the joint impact of other regions' recent development and its predictive information. Because of the heterogeneous parameters, these local averages can be considered to incorporate even the reduced form of a spatial lag model; explicitly accounting for the spatial lag would likely produce less accurate forecasts (Kelejian/Prucha, 2007b). The imposed structure results from specifications tests; in addition to the employed cross-sectionally weak dependence, the GVAR allows for strong dependence on a dominant (factor-like) region. The existence of a strong factor is however rejected from the data. Spatial autocorrelation in the disturbance is irrelevant for the analysis: for region-by-region model fitting, the local averages can be considered weakly exogenous; the construction of confidence intervals around the forecasts (for which the system's variance-covariance matrix would be required) is not necessary since I evaluate point forecasts; the evaluation criterion (improvement of out-of-sample forecasts) makes in-sample inference or parameter identification second-rate. Hence, estimation of a well defined spatial lag or spatial error process is not required by the GVAR method, albeit several elements of model building are related to the spatial econometric mainstream. In the empirical application, I find evidence that accounting for spatial co-development along this path improves the forecast accuracy compared to forecasts in which the regions are considered independent. Without providing evidence for my claim, I suppose that my proceeding produces more accurate forecasts than estimation and prediction with a spatially autocorrelated Panel VAR (as developed by Mutl, 2005) which is closer to conventional methods. A second issue not included in the study but discussed in the literature is that the GVAR estimations (and predictions) are to some extent sensitive with regard to the spatial weights employed in locally averaging: Schanne/Wapler/Weyh (2010) compare two different weighting schemes in univariate GVAR forecasting with automatized lag selection; the resulting models yield distinct predictions without one scheme being superior to the other in terms of accuracy. Also beyond the scope of the study is the issue of explicitly investigating the spatio-temporal diffusion of shocks. The construction of spatial Impulse-Response functions would be possible along the lines discussed in Pesaran/Schuermann/Weiner (2004) and Holly/Pesaran/Yamagata (2011); they constitute a kind of spatio-temporal impact measure. However, it would be difficult to provide an economic intuition behind tracing these impulse responses without a dominant region as the origin of shocks.

A conventional spatially autoregressive model is employed when investigating the sentiments of labour market experts. Herding can be understood as an endogenous peer effect, i.e. as a cross-sectionally autoregressive process which is identifiable only within a spatial or network structure. Since the spatial autoregression coefficient is the parameter of interest, there is no way around estimating it. I argue that information likely propagates along a spatial structure since the experts have to communicate more intensive with those whose area of responsibility is within commuting distance when carrying out their daily business. However, in order establish robustness, I consider alternative reasonable communication networks, I look at the instruments' validity, and I consider alternative disturbance structures; in short, I'm convinced that I indeed identify a herding parameter and do not estimate only some undefined (or possibly spurious) correlation.

To sum up, I have shown that spatial structures can not only be observed in regional labour market data and explained by economic theory on regional labour markets; they are even tractable in empirical research if the investigation strategy allows for both spatial dependence and spatial heterogeneity in a sufficiently flexible way. Applied research employs to a large extent spatial autocorrelation models which are well implemented in standard software packages. In my view, these parametric methods come along with a number of disadvantages that are hardly addressed: parameter identification only under (frequently implicit) restrictions; sensitivity with regard to the spatial weights employed; a (too) strong focus on correlation and not on effect heterogeneity; and, because of the infinite sum $\sum_{q=0}^{\infty}\rho^{q}W_{(n)}^{q}$ in the spatial multiplier, too smooth impact measures which can not account for heterogeneity at a less than nation scale. A lesson to be learned from this thesis would be that for many purposes in empirical regional research estimation methods are available which are better suited than the spatial econometric mainstream; hence it frequently might be worth taking more care on the question which results can really be achieved with the employed method, and how validity of the findings can be strengthened. Although this seems quite technical at the first glance, it has even practical implications: for example, if only the homogeneous parameter estimate in the study on unemployment persistence in Chapter 4 were taken into consideration, one would assess a lower degree of shock persistence (or faster convergence to a 'natural rate') than that found with heterogenous-parameter methods.

Bibliography

Acemoglu, Daron; Dahleh, Munther A.; Lobel, Ilan; Ozdaglar, Asuman (2011): Bayesian Learning in Social Networks. In: Review of Economic Studies, Vol. 78, p. 1201–1236.

Acemoglu, Daron; Ozdaglar, Asuman (2011): Opinion Dynamics and Learning in Social Networks. In: Dynamic Games and Applications, Vol. 1, p. 3–49.

Aldashev, Alisher (2012): Occupational and Locational Substitution: Measuring the Effect of Occupational and Regional Mobility. In: Labour, Vol. 26, No. 1, p. 108–123.

Anselin, Luc (2001): Spatial Econometrics. In: (Baltagi, 2001), chap. 14, p. 310–330.

Anselin, Luc (1991): Spatial Dependence and Spatial Heterogeneity: Model Specification Issues in the Spatial Expansion Paradigm. In: Casetti, Emilio; Jones, John P. (Eds.) Applications of the Expansion Method, chap. 15, London, New York (NY): Routledge, p. 264–279.

Anselin, Luc (1990): Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis. In: Journal of Regional Science, Vol. 30, No. 2, p. 185–207.

Anselin, Luc (1988): Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

Aragon, Yves; Haugthon, Dominique; Haughton, Jonathan; Leconte, Eve; Malin, Eric; Ruiz-Gazen, Anne; Thomas-Agnan, Christine (2003): Explaining the Pattern of Regional Unemployment: the Case of the Midi-Pyrénées Region. In: Papers in Regional Science, Vol. 82, No. 2, p. 155–174.

Armstrong, Harvey; Taylor, Jim (2000): Regional Economics and Policy. Oxford: Blackwell, 3rd ed.

Badinger, Harald; Url, Thomas (2002): Determinants of Regional Unemployment: Some Evidence from Austria. In: Regional Studies, Vol. 36, p. 977–988.

Bai, Jushan (2004): Estimating Cross-section Common Stochastic Trends in Nonstationary Panel Data. In: Journal of Econometrics, Vol. 122, p. 137–183.

Bai, Jushan; Ng, Serena (2011): Principal Components Estimation and Identification of the Factors, mimeo.

Bai, Jushan; Ng, Serena (2004): A PANIC Attack on Unit Roots and Cointegration. In: Econometrica, Vol. 72, No. 4, p. 1127–1177.

Bai, Jushan; Ng, Serena (2002): Determining the Number of Factors in Approximate Factor Models. In: Econometrica, Vol. 70, No. 1, p. 191–221.

Bailey, Natalia; Kapetanios, George; Pesaran, M. Hashem (2012): Exponent of Cross-sectional Dependence: Estimation and Inference. Discussion paper 6318, IZA.

Bajari, Patrick; Krainer, John (2004): An Empirical Model of Stock Analysts' Recommendations: Market Fundamentals, Conflict of Interest, and Peer Effects. Working Paper 10665, NBER.

Baltagi, Badi H. (2006): Random Effects and Spatial Autocorrelation with Equal Weights. In: Econometric Theory, Vol. 22, p. 973–984.

Baltagi, Badi H. (Ed.) (2001): A Companion to Theoretical Econometrics. Companions to Contemporary Economics, Malden (MA): Blackwell.

Baltagi, Badi H.; Liu, L. (2009): Spatial Lag Tests with Equal Weights. In: Economics Letters, Vol. 104, p. 81–82.

Baltagi, Badi H.; Song, S.H.; Jung, B.C.; Koh, W. (2007): Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. In: Journal of Econometrics, Vol. 140, p. 5–51.

Banerjee, Abhijit V. (1992): A Simple Model of Herd Behavior. In: Quarterly Journal of Economics, Vol. 107, No. 3, p. 797–817.

Banerjee, Anindya; Marcellino, Massimiliano (2009): Factor-augmented Error Correction Models. In: Castle, J.L; Shephard, N. (Eds.) The Methodology and Practice of Econometrics – A Festschrift for David Hendry, chap. 9, Oxford: Oxford University Press, p. 227–254.

Banerjee, Anindya; Marcellino, Massimiliano; Masten, Igor (2009): Forecasting with Factor-augmented Error Correction Models, mimeo.

Basile, Roberto; De Benedictis, Luca (2008): Regional Unemployment and Productivity in Europe. In: Papers in Regional Science, Vol. 87, No. 2, p. 173–192.

Bayer, Christian; Juessen, Falko (2007): Convergence in West German Regional Unemployment Rates. In: German Economic Review, Vol. 8, No. 4, p. 510–535.

Beaulieu, J. Joseph; Miron, Jeffrey A. (1993): Seasonal Unit Roots in Aggregate U.S. Data. In: Journal of Econometrics, Vol. 55, p. 305–328.

Bera, A.; Simlai, P. (2005): Testing Spatial Autoregressive Model and a Formulation of Spatial ARCH (SARCH) Model with Applications. Mimeo.

Bernanke, Ben S.; Boivin, Jean; Eliasz, Piotr (2005): Measuring the Effects of Monetary Policy: A Factor-augmented Vector Autoregressive (FAVAR) Approach. In: Quarterly Journal of Economics, Vol. 120, No. 1, p. 387–422.

Bernoth, Kerstin; Pick, Andreas (2011): Forecasting the Fragility of the Banking and Insurance Sectors. In: Journal of Banking & Finance, Vol. 35, No. 4, p. 807–818.

Bester, C. Alan; Conley, Timothy G.; Hansen, Christian B. (2011): Inference with Dependent Data using Cluster Covariance Estimators. In: Journal of Econometrics, Vol. 165, No. 2, p. 137–151.

Bewley, Ronald; Fiebig, Denzil G. (2002): On the Herding Instinct of Interest Rate Forecasters. In: Empirical Economics, Vol. 27, p. 403–425.

Bikhchanandi, Sushil; Hirshleifer, David; Welch, Ivo (1998): Learning from the Behavior of Others: Conformity, Fads, Informational Cascades. In: Journal of Economic Perspectives, Vol. 12, p. 151–170.

Bikhchanandi, Sushil; Hirshleifer, David; Welch, Ivo (1992): A Theory of Fads, Fashion, Custom, and Cultural Change as Information Cascades. In: Journal of Political Economy, Vol. 100, p. 992–1026.

Blanchard, Olivier J.; Summers, Lawrence H. (1986): Hysteresis and the European Unemployment Problem. In: NBER Macroeconomics Annual, Vol. 1, p. 15–78.

Blien, Uwe; Hirschenauer, Franziska; Thi Hong Van, Phan (2010): Classification of Regional Labour Markets for Purposes of Labour Market Policy. In: Papers in Regional Science, Vol. 89, p. 859–880.

Blundell, Richard; Bond, Stephen (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. In: Journal of Econometrics, Vol. 87, p. 114–143.

Bramoullé, Yann; Djebbari, Habiba; Fortin, Bernard (2009): Identifiction of Peer Effects through Social Networks. In: Journal of Econometrics, Vol. 150, p. 41–55.

Brock, William A.; Durlauf, Steven N. (2003): Multinomial Choice with Social Interactions. Technical Working Paper 288, NBER.

Brock, William A.; Durlauf, Steven N. (2001): Discrete Choice with Social Interactions. In: Review of Economic Studies, Vol. 68, p. 235–260.

Bronars, Stephen G.; Jansen, Dennis W. (1987): The Geographic Distribution of Unemployment Rates in the U.S. – A Spatial Time-Series Analysis. In: Journal of Econometrics, Vol. 36, p. 251–279.

Brueckner, Jan K. (1987): The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model. In: Mills, Edwin S. (Ed.) Handbook of Regional and Urban Economics, Vol. 2, chap. 20, Amsterdam: Elsevier, p. 821–845.

Brunsdon, Christopher; Fotheringham, A. Stewart; Charlton, Martin (2000): Geographically Weighted Regression as a Statistical Model. Mimeo.

Brunsdon, Christopher; Fotheringham, A. Stewart; Charlton, Martin (1998): Geographically Weighted Regression – Modelling Spatial Non-Stationarity. In: The Statistician (Journal of the Royal Statistical Society, Series D), Vol. 47, No. 3, p. 431–443.

Bun, Maurice J.G.; Carree, Martin A. (2005): Bias-Corrected Estimation in Dynamic Panel Data Models. In: Journal of Business and Economic Statistics, Vol. 23, p. 200–210.

Cahuc, Pierre; Zylberberg, André (2004): Labor Economics. Cambridge (MA), London: MIT Press.

Card, David (1990): The Impact of the Mariel Boatlift on the Miami Labor Market. In: Industrial and Labor Relations Review, Vol. 43, p. 245–257.

Carroll, Christopher D. (2003): Macroeconomic Expectations of Households and Professional Forecasters. In: Quarterly Journal of Economics, Vol. 118, p. 269–298.

Casetti, Emilio (1972): Generating Models by the Expansion Method: Application to Geographical Research. In: Geographical Analysis, Vol. 4, No. 1, p. 81–91.

Çelen, Boğaçhan; Kariv, Shachar (2004): Distinguishing Informational Cascades from Herd Behavior in the Laboratory. In: American Economic Review, Vol. 94, No. 3, p. 484– 498.

Chamley, Christophe P. (2004): Rational Herds: Economic Models of Social Learning. Cambridge (UK), New York (NY): Cambridge University Press.

Chroushore, Dean (2010): An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data. In: The B.E. Journal of Macroeconomics, Vol. 10, No. 1.

Chroushore, Dean (1998): Evaluating Inflation Forecasts. Working paper 98–14, Federal Reserve Bank of Philadelphia.

Chudik, Alexander; Pesaran, M. Hashem (2013): Large Panel Data Models with Cross-Sectional Dependence: A Survey. Working Paper 153, Federal Reserve Bank of Dallas.

Chudik, Alexander; Pesaran, M. Hashem (2011): Infinite-dimensional VARs and Factor Models. In: Journal of Econometrics, Vol. 163, No. 1, p. 4–22.

Chudik, Alexander; Pesaran, M. Hashem; Tosetti, Elisa (2011): Weak and Strong Cross Section Dependence and Estimation of Large Panels. In: Econometrics Journal, Vol. 14, No. 1, p. C45–C90.

Clark, Todd E.; West, Kenneth D. (2007): Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. In: Journal of Econometrics, Vol. 138, No. 1, p. 291–311.

Cliff, Andrew D.; Ord, J. Keith (1973): Spatial Autocorrelation. London: Pion Press.

Cliff, Andrew D.; Ord, J. Keith (1972): Testing for Spatial Autocorrelation Among Regression Residuals. In: Geographical Analysis, Vol. 4, No. 3, p. 267–284.

Conley, T.G.; Topa, G. (2002): Socio-Economic Distance and Spatial Patterns in Unemployment. In: Journal of Applied Econometrics, Vol. 17, p. 303–327.

Conley, Timothy G. (1999): GMM Estimation with Cross Sectional Dependence. In: Journal of Econometrics, Vol. 92, p. 1–45.

Corrado, Luisa; Fingleton, Bernard (2012): Where is the Economics in Spatial Econometrics? In: Journal of Regional Science, Vol. 52, No. 2, p. 210–239.

Cracolici, Maria F.; Cuffaro, Miranda; Nijkamp, Peter (2007): Geographical Distribution of Unemployment: An Analysis of Provincial Differences in Italy. In: Growth and Change, Vol. 38, p. 649–670.

Curtin, Richard T. (2003): Unemployment Expectations: The Impact of Private Information on Income Uncertainty. In: Review of Income and Wealth, Vol. 49, No. 4, p. 539–554.

Davidson, J. (1994): Stochastic Limit Theory. Advanced Texts in Econometrics, Oxford, New York (NY): Oxford University Press.

De Giorgi, Giacomo; Pellizari, Michele; Redaelli, Silvia (2010): Identification of Social Interactions through Partially Overlapping Peer Groups. In: American Economic Journal: Applied Economics, Vol. 2, p. 241–275.

Decressin, Jörg; Fatás, Antonio (1995): Regional Labor Market Dynamics in Europe. In: European Economic Review, Vol. 39, p. 1627–1655.

Dees, Stephane; Di Mauro, Filippo; Pesaran, M. Hashem; Smith, L. Vanessa (2007): Exploring the International Linkages of the Euro Area: A Global VAR Analysis. In: Journal of Applied Econometrics, Vol. 22, p. 1–38.

Dickey, David A.; Bell, William R.; Miller, Robert B. (1986): Unit Roots in Time Series Models: Tests and Implications. In: The American Statistician, Vol. 40, p. 12–26.

Eicker, Friedhelm (1967): Limit Theorems for Regressions with Unequal and Dependent Errors. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, p. 59–82.

Elhorst, J. Paul (2003): The Mystery of Regional Unemployment Differentials: Theoretical and Empirical Explanations. In: Journal of Economic Surveys, Vol. 17, No. 5, p. 709–748.

Elhorst, J. Paul (1995): Unemployment Disparities between Regions in the European Union. In: Armstrong, H.W.; Vickerman, R.W. (Eds.) Convergence and Divergence among European Unions, London: Pion, p. 209–221.

Engle, Robert F.; Granger, Clive W.J. (1987): Co-Integration and Error Correction: Representation, Estimation, and Testing. In: Econometrica, Vol. 55, p. 251–276.

Epifani, P.; Gancia, G.A. (2005): Trade, Migration, and Regional Unemployment. In: Regional Science and Urban Economics, Vol. 35, p. 625–644.

Ertur, Cem; Le Gallo, Julie (2003): An Exploratory Spatial Data Analysis of European Regional Disparities, 1980–1995. In: Fingleton, B. (Ed.) European Regional Growth, Berlin, Heidelberg: Springer, p. 55–98.

Ertur, Cem; Le Gallo, Julie; Baumont, Catherine (2006): The European Regional Convergence Process, 1980–1995: Do Spatial Regimes and Spatial Dependence Matter? In: International Regional Science Review, Vol. 29, No. 1, p. 3–34.

Federal Statistical Office of Germany (2012): National Accounts: Gross Domestic Product (Quarterly and Annually), Long Time Series from 1970. http://www.destatis.de/EN/FactsFigures/NationalEconomyEnvironment/NationalAccounts/DomesticProduct/Tables/Tables.html, accessed last November 14th, 2012.

Florax, Raymond J.G.M.; Folmer, Hendrik; Rey, Sergio J. (2003): Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology. In: Regional Science and Urban Economics, Vol. 33, p. 557–579.

Forni, Mario; Hallin, Marc; Lippi, Marco; Reichlin, Lucrezia (2005): The Generalized Dynamic-Factor Model: One-Sided Estimation and Forecasting. In: Journal of the American Statistical Association, Vol. 100, No. 471, p. 830–840.

Forni, Mario; Hallin, Marc; Lippi, Marco; Reichlin, Lucrezia (2004): The Generalized Dynamic Factor Model: Consistency and Rates. In: Journal of Econometrics, Vol. 119, p. 231–255.

Forni, Mario; Hallin, Marc; Lippi, Marco; Reichlin, Lucrezia (2000): The Generalized Dynamic-Factor Model: Identification and Estimation. In: The Review of Economics and Statistics, Vol. 82, No. 4, p. 540–554.

Fotheringham, A. Stewart; Charlton, Martin; Brunsdon, Christopher (1997): Two Techniques for Exploring Non-Stationarity in Geographical Data. In: Geographical Systems, Vol. 4, p. 59–82.

Franzese, Robert J. Jr.; Hays, Jude (2009): The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation, and Presentation, presented at the Annual Meeting of the Public Choice Society, 6 Mar 2009.

Fuchs, Johann; Hummel, Markus; Klinger, Sabine; Wanger, Susanne; Zika, Gerd (2010): Der Arbeitsmarkt schließt an den vorherigen Aufschwung an. IAB Kurzbericht 18, IAB.

Fujita, M; Krugman, P.; Venables, A.J. (1999): The Spatial Economy. Cities, Regions, and International Trade. Cambridge (MA), London: The MIT Press.

Gaggermeier, Christian (2006): Indikatoren-Modelle zur Kurzfristprognose der Beschäftigung in Deutschland. IAB Forschungsbericht 6, IAB, Nürnberg.

Getis, Arthur (1995): Spatial Filtering in a Regression Framework: Examples Using Data on Urban Crime, Regional Inequality and Government Expenditures. In: Anselin, L.; Florax, R.J.G.M. (Eds.) New Directions in Spatial Econometrics, Berlin, Heidelberg: Springer, p. 172–185.

Getis, Arthur (1990): Screening for Spatial Dependence in Regression Analysis. In: Papers of the Regional Science Association, Vol. 69, p. 69–81.

Getis, Arthur; Griffith, Daniel A. (2002): Comparative Spatial Filtering in Regression. In: Geographical Analysis. An International Journal of Theoretical Geography, Vol. 34, No. 2, p. 130–140.

Gibbons, Stephen; Overman, Henry G. (2012): Mostly Pointless Spatial Econometrics? In: Journal of Regional Science, Vol. 52, No. 2, p. 172–191.

Gracia-del Barrio, Pedro; Gil-Alana, Luis A. (2009): New Revelations About Unemployment Persistence in Spain: Time-Series and Panel Data Approaches Using Regional Data. In: Applied Economics, Vol. 41, No. 2, p. 219–236.

Granger, Clive W.J. (1969): Spatial Data and Time Series Analysis. In: London Papers in Regional Science, Vol. 1, p. 1–24.

Greene, William H. (2000): Econometric Analysis. Upper Saddle River (NJ): Prentice Hall, 4th ed.

Griffith, Daniel A. (2008): Spatial Filtering-Based Contributions to a Critique of Geographically Weighted Regression GWR. In: Environment & Planning A, Vol. 40, p. 2751– 2769.

Griffith, Daniel A. (2003): Spatial Autocorrelation and Spatial Filtering. Gaining Understanding through Theory and Scientific Visualization. Advances in spatial science, Berlin, Heidelberg: Springer.

Griffith, Daniel A. (2000): A Linear Regression Solution to the Spatial Autocorrelation Problem. In: Journal of Geographical Systems, Vol. 2, p. 141–156.

Griffith, Daniel A. (1996): Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-referenced Data. In: The Canadian Geographer, Vol. 40, No. 4, p. 351–367.

Griffith, Daniel A.; Peres-Neto, Pedro R. (2006): Spatial Modelling in Ecology: The Flexibility of Eigenfuntion Spatial Analyses. In: Ecology, Vol. 87, p. 2603–2613.

Haining, R. (2003): Spatial Data Analysis. Theory and Practice. Cambridge, New York (NY): Cambridge University Press.

Hampel, Katharina; Kunz, Marcus; Schanne, Norbert; Wapler, Rüdiger; Weyh, Antje (2008): Regional Employment Forecasts with Spatial Interdependencies. In: Knobel, C.; Kriechel, B.; Schmid, A. (Eds.) Regional Forecasting on Labour Markets, chap. 5, München, Mering: Rainer Hampp Verlag, p. 68–88.

Hastie, Trevor; Tibshirani, Robert (1993): Varying–Coefficient Models. In: Journal of the Royal Statistical Society, Series B (Methodological), Vol. 55, No. 4, p. 757–796.

Herwartz, Helmut; Niebuhr, Annekatrin (2011): Growth, Unemployment and Labour Market Institutions: Evidence from a Cross-section of EU Regions. In: Applied Economics, Vol. 43, No. 30, p. 4663–4676.

Hey, John D. (1994): Expectations formation: Rational or adaptive or …? In: Journal of Economic Behavior & Organization, Vol. 25, p. 329–349.

Holly, Sean; Pesaran, M. Hashem; Yamagata, Takashi (2011): Spatial and Temporal Diffusion of House Prices in the UK. In: Journal of Urban Economics, Vol. 69, p. 2–23.

Huber, Peter J. (1967): The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, p. 221–233.

Hujer, Reinhard; Rodrigues, Paulo J.M.; Wolf, Katja (2008): Dynamic Panel Data Models with Spatial Correlation. In: Jahrbücher für Nationalökonomie und Statistik, Vol. 228, No. 5+6, p. 612–629.

Im, Kyung S.; Pesaran, M. Hashem; Shin, Yongcheol (2003): Testing for Unit Roots in Heterogeneous Panels. In: Journal of Econometrics, Vol. 115, p. 53–74.

Jacobs, Jan P.A.M.; van Norden, Simon (2011): Modeling Data Revisions: Measurement Error and Dynamics of "True" Values. In: Journal of Econometrics, Vol. 161, p. 101–109.

Jimeno, Juan F.; Bentolila, Samuel (1998): Regional Unemployment Persistence. In: Labour Economics, Vol. 5, p. 25–51.

Johansen, Søren (1995): Likelihood-based Inference in Cointegrated Vector Autoregressive Models. Oxford: Oxford University Press.

Johansen, Søren (1991): Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. In: Econometrica, Vol. 59, No. 6, p. 1551–1580.

Kapoor, M.; Kelejian, H.H.; Prucha, I.R. (2007): Panel Data Models with Spatially Correlated Error Components. In: Journal of Econometrics, Vol. 140, No. 1, p. 97–130.

Keane, Michael P.; Runkle, David E. (1990): Testing the Rationality of Price Forecasts: New Evidence from Panel Data. In: American Economic Review, Vol. 80, No. 4, p. 714– 735.

Kelejian, H.H.; Prucha, I.R. (2007a): HAC Estimation in a Spatial Framework. In: Journal of Econometrics, Vol. 140, p. 131–154.

Kelejian, H.H.; Prucha, I.R. (2007b): The Relative Efficiencies of Various Predictors in Spatial Econometric Models Containing Spatial Lags. In: Regional Science and Urban Economics, Vol. 37, p. 363–374.

Kelejian, H.H.; Prucha, I.R. (2004): Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations. In: Journal of Econometrics, Vol. 118, p. 27–50.

Kelejian, H.H.; Prucha, I.R. (2002): 2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights. In: Regional Science and Urban Economics, Vol. 32, p. 691–707.

Kelejian, H.H.; Prucha, I.R. (1999): A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. In: International Economic Review, Vol. 40, p. 509–533.

Kelejian, H.H.; Prucha, I.R. (1998): A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. In: Journal of Real Estate Finance and Economics, Vol. 17, p. 99–121.

Kelejian, H.H.; Robinson, D.P. (1993): A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model. In: Papers in Regional Science, Vol. 72, p. 297–312.

Kochendörfer-Lucius, Gudrun; Pleskovic, Boris (Eds.) (2009): Spatial Disparities and Development Policies. Washington, D.C.: World Bank.

Kosfeld, Reinhold; Dreger, Christian (2006): Thresholds for Emplyoment and Unemployment. A Spatial Analysis of German Regional Labour Markets 1999–2000. In: Papers in Regional Science, Vol. 85, No. 4, p. 523–542.

Kunz, Marcus (2012): Regional Unemployment Disparities in Germany. No. 331 in IAB-Bibliothek, Nuremberg: IAB/Bertelsmann, diss.

Lamont, Owen (2002): Macroeconomic Forecasts and Microeconomic Forecasters. In: Journal of Economic Behavior & Organization, Vol. 48, No. 3, p. 265–280.

Lancaster, Tony (2000): The Incidental Parameter Problem since 1948. In: Journal of Econometrics, Vol. 95, p. 391–413.

Lee, Chien-C.; Chang, Chun-P. (2008): Unemployment Hysteresis in OECD Countries: Centurial Time Series Evidence with Structural Breaks. In: Economic Modelling, Vol. 25, p. 312–325.

Lee, L.-F. (2004): Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. In: Econometrica, Vol. 72, No. 6, p. 1899–1925.

Lee, Lung-F.; Yu, Jihai (2010): Spatial Panels: Random vs. Fixed Effects, mimeo.

LeSage, James; Pace, R. Kelley (2009): Introduction to Spatial Econometrics. Boca Raton (FL), New York (NY): CRC Press.

Levin, Andrew; Lin, Chien-F.; Chu, Chia-S.J. (2002): Unit Root Tests in Panel Data: Asymptotic and Finite–Sample Properties. In: Journal of Econometrics, Vol. 108, p. 1–24.

Longhi, Simonetta; Nijkamp, Peter (2007): Forecasting Regional Labour Market Developments under Spatial Autocorrelation. In: International Regional Science Review, Vol. 30, No. 2, p. 100–119.

López-Bazo, Enrique; del Barrio, Tomáas; Artis, Manuel (2002): The Regional Distribution of Spanish Unemployment: A Spatial Analysis. In: Papers in Regional Science, Vol. 81, No. 3, p. 365–389.

Lucas, Robert E.jr.; Prescott, E.C. (1974): Equilibrium Search and Unemployment. In: Journal of Economic Theory, Vol. 7, p. 188–209.

Lütkepohl, Helmut (2005): New Introduction to Multiple Time Series Analysis. Berlin, Heidelberg, New York: Springer.

Maddala, Gangadharrao S.; Wu, Shaowen (1999): A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test: Evidence From Simulations and the Bootstrap. In: Oxford Bulletin of Economics and Statistics, Vol. 61, No. S1, p. 631–652.

Manski, Charles F. (2004): Social Learning from Private Experience: The Dynamics of the Selection Problem. In: Review of Economic Studies, Vol. 71, p. 443–458.

Manski, Charles F. (1993): Identification of Endogenous Social Effects: The Reflection Problem. In: Review of Economic Studies, Vol. 60, No. 3, p. 531–542.

Mariano, Roberto S. (2001): Simultaneous Equation Model Estimators: Statistical Porperties and Practical Implications. In: Baltagi, Badi H. (Ed.) A Companion to Theoretical Econometrics, chap. 6, Malden (MA): Blackwell, p. 122–143.

Marston, Stephen T. (1985): Two Views of the Geographic Distribution of Unemployment. In: Quarterly Journal of Economics, Vol. 100, No. 1, p. 57–79.

Mayor, Matías; López, Ana Jesús (2008): Spatial Shift-Share Analysis Versus Spatial Filtering: An Application to Spanish Employment Data. In: Empirical Economics, Vol. 34, p. 123–142.

Mayor, Matias (Sr.); Patuelli, Roberto (2012): Short-Run Regional Forecasts: Spatial Models Through Varying Cross-Sectional and Temporal Dimensions. Quaderni DSE working paper 835, University of Bologna.

Miller, Alan (2002): Subset Selection in Regression. Boca Raton: Chapman & Hall/CRC.

Mincer, Jacob A.; Zarnowitz, Victor (1969): The Evaluation of Economic Forecasts. In: Mincer, Jacob A. (Ed.) Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance, chap. 1, Cambridge: NBER, p. 1–46.

Mitchell, William; Bill, Anthea (2004): Spatial Dependence in Regional Unemployment in Australia. In: Carlson, E. (Ed.) A Future that Works: Economics, Employment and the Environment. Proceedings of the 6th Path to Full Employment Conference and the 11th National Conference on Unemployment, Newcastle: The University of Newcastle, p. 312– 326.

Molho, Ian (1995): Spatial Autocorrelation in British Unemployment. In: Journal of Regional Science, Vol. 35, No. 4, p. 641–658.

Möller, Joachim (2010): The German Labor Market Response in the World Recession – Demystifying a Miracle. In: Zeitschrift für Arbeitsmarktforschung, Vol. 42, p. 325–336.

Moran, P.A.P. (1950): Notes on Continuous Stochastic Phenomena. In: Biometrika, Vol. 37, p. 17–23.

Moretti, Enrico (2011): Local Labor Markets. In: Ashenfelter, Orley; Card, David (Eds.) Handbook of Labor Economics, Vol. 4b, chap. 14, Amsterdam: Elsevier, p. 1237–1313.

Morrision, P. (2005): Unemployment and Urban Labour Markets. In: Urban Studies, Vol. 42, p. 2261–2288.

Mortensen, D. (2003): Wage Dispersion: Why Are Similar Workers Paid Differently? Cambridge (MA), London: MIT Press.

Mortensen, D.T.; Pissarides, C.A. (1999): New Developments in Models of Search in the Labour Market. In: Card, D.; Ashenfelter, O. (Eds.) Handbook of Labor Economics, Vol. 3B, chap. 39, Amsterdam, New York (NY): Elsevier, p. 2567–2627.

Mur, Jesús; Angulo, Ana (2009): Model Selection Strategies in a Spatial Setting: Some Additional Results. In: Regional Science and Urban Economics, Vol. 39, p. 200–213.

Mutl, J. (2006): Misspecification of Space: An Illustration Using Growth Convergence Regressions. Tech. Rep., Goethe University, presented at the International Workshop on Spatial Econometrics and Statistics, Rome.

Mutl, J. (2005): Dynamic Panel Data Models with Spatially Correlated Disturbances. Ph.D. thesis, University of Maryland.

Mutl, Jan; Pfaffermayr, Michael (2011): The Hausman Test in a Cliff and Ord Panel Model. In: Econometrics Journal, Vol. 14, p. 48–76.

Newey, Whitney K.; West, Kenneth D. (1987): A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. In: Econometrica, Vol. 55, No. 3, p. 703–708.

Neyman, Jerzy; Scott, Elizabeth L. (1948): Consistent Estimation from Partially Consistent Observations. In: Econometrica, Vol. 16, p. 1–32.

Nickell, Stephen J. (1981): Biases in Dynamic Models with Fixed Effects. In: Econometrica, Vol. 49, p. 1417–1426.

Niebuhr, Annekatrin (2003): Spatial Interaction and Regional Unemployment in Europe. In: European Journal of Spatial Development, Vol. 5, p. 1–26.

Nijkamp, Peter (2009): Regional Development as Self-Organized Converging Growth. In: Kochendörfer-Lucius, G.; Pleskovic, B. (Eds.) Spatial Disparities and Development Policy, Washington, D.C.: World Bank, p. 265–281.

Onatski, Alexei (2010): Determining the Number of Factors from Empirical Distribution of Eigenvalues. In: Review of Economics and Statistics, Vol. 92, No. 4, p. 1004–1016.

Onatski, Alexei (2009): Testing Hypotheses About the Number of Factors in Large Factor Models. In: Econometrica, Vol. 77, No. 5, p. 1447–1479.

Ottaviani, Marco; Sørensen, Peter N. (2006): The Strategy of Professional Forecasting. In: Journal of Financial Economics, Vol. 81, No. 2, p. 441–466.

Ottaviano, G.; Thisse, J.-F. (2004): Agglomeration and Economic Geography. In: Henderson, V.; Thisse, J.-F. (Eds.) Handbook of Regional and Urban Economics, Vol. 4, Amsterdam, New York: Elsevier, p. 2563–2608.

Oud, Johan H.L.; Folmer, Henk; Patuelli, Roberto; Nijkamp, Peter (2012): Continuous-Time Modeling with Spatial Dependence. In: Geographical Analysis, Vol. 44, p. 29–46.

Pagan, A.; Ullah, A. (1999): Nonparametric Econometrics. themes in modern econometrics, Cambridge: Cambridge University Press.

Partridge, Mark D.; Rickman, Dan S. (2010): Computable General Equilibrium (CGE) Modelling for Regional Economic Development Analysis. In: Regional Studies, Vol. 44, p. 1311–1328.

Patacchini, Eleonora; Zenou, Yves (2007): Spatial Dependence in Local Unemployment Rates. In: Journal of Economic Geography, Vol. 7, p. 169–191.

Patuelli, Roberto (2007): Regional Labour Markets in Germany: Statistical Analysis of Spatio-Temporal Disparities and Network Structures. Ph.d. thesis, VU Amsterdam.

Patuelli, Roberto; Griffith, Daniel A.; Tiefelsdorf, Michael; Nijkamp, Peter (2012): Spatial Filtering Methods for Tracing Space-Time Developments in an Open Regional System: Experiments with German Unemployment Data. In: Frenkel, A.; Nijkamp, P.; McCann, P. (Eds.) Societies in Motion: Regional Development, Industrial Innovation and Spatial Mobility, Cheltenham Northampton: Edward Elgar, p. 247–268.

Patuelli, Roberto; Griffith, Daniel A.; Tiefelsdorf, Michael; Nijkamp, Peter (2011): Spatial Filtering and Eigenvector Stability: Space-time Models for German Unemployment Data. In: International Regional Science Review, Vol. 34, No. 2, p. 253–280.

Peña, Daniel; Poncela, Pilar (2004): Forecasting with Nonstationary Dynamic Factor Models. In: Journal of Econometrics, Vol. 119, p. 291–321.

Pede, V.O.; Florax, R.J.G.M.; Holt, M.T. (2009): A Spatial Econometric Star Model with an Application to U.S. County Economic Growth, 1969–2003. Working Paper 09-03, Purdue University, Department of Agricultural Economics.

Peng, Roger (2008): A Method for Visualizing Multivariate Time Series Data. In: Journal of Statistical Software. Code Snippets, Vol. 25, p. 1–17.

Pesaran, M. Hashem; Schuermann, T.; Weiner, Scott M. (2004): Modeling Regional Interdependencies using a Global Error-Correcting Macroeconometric Model. In: Journal of Business and Economics Statistics, Vol. 22, p. 129–162.

Pesaran, M. Hashem; Schuermann, Til; Smith, L. Vanessa (2009): Forecasting Economic and Financial Variables with Global VARs. In: International Journal of Forecasting, Vol. 25, p. 642–675.

Pesaran, M. Hashem; Tosetti, Elisa (2007): Large Panels with Common Factors and Spatial Correlations. Discussion Paper 3032, IZA.

Pesaran, M.Hashem (2006): Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. In: Econometrica, Vol. 74, No. 4, p. 967–1012.

Pissarides, C.A.; Wadsworth, J. (1989): Unemployment and the Inter-regional Mobility of Labour. In: The Economic Journal, Vol. 99, p. 739–755.

Pons-Novell, Jordi (2003): Strategic Bias, Herding Behaviour and Economic Forecasts. In: Journal of Forecasting, Vol. 22, p. 67–77.

Pötscher, B.M.; Prucha, I.R. (2001): Basic Elements of Asymptotic Theory. In: (Baltagi, 2001), chap. 10, p. 201–229.

Rangvid, Jesper; Schmeling, Maik; Schrimpf, Andreas (2012): What do Professional Forecasters' Stock Market Expectations tell us about Herding, Information Extraction and Beauty Contests? http://staff.cbs.dk/JRangvid/HOE.pdf.

Rao, C. Radhakrishna (1967): Least Squares Theory using an Estimated Dispersion Matrix and its Application to Measurement of Signals. In: Le Cam, Lucien M.; Neyman, Jerzy (Eds.) Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Berkeley (CA): University of California Press, p. 355–372.

Robinson, P.M. (2011): Asymptotic Theory for Nonparametric Regression with Spatial Data. In: Journal of Econometrics, Vol. 165, p. 5–19.

Rüb, F.; Werner, D. (2008): "Den Arbeitsmarkt" gibt es nicht – Arbeitsmarktregionen des SGB II im Vergleich. In: Jahrbuch für Regionalwissenschaft, Vol. 28, p. 93–108.

Rülke, Jan-Christoph; Tillmann, Peter (2011): Do FOMC members herd? In: Economics Letters, Vol. 113, No. 2, p. 176–179.

Sarno, Lucio; Taylor, Mark P. (1998): Real Exchange Rate under the Recent Float: Unequivocal Evidence of Mean Reversion. In: Economics Letters, Vol. 60, p. 131–137.

Schanne, Norbert (2010): Forecasting Regional Labour Markets with GVAR Models and Indicators. http://www-sre.wu.ac.at/ersa/ersaconfs/ersa10/ERSA2010finalpaper1044.pdf.

Schanne, Norbert; Wapler, Rüdiger; Weyh, Antje (2010): Regional Unemployment Forecasts with Spatial Interdependencies. In: International Journal of Forecasting, Vol. 26, p. 908–926.

Schanne, Norbert; Weyh, Antje (2009): What Makes Start-ups out of Unemployment Different? Discussion Paper 4, IAB, Nuremberg.

Scharfstein, David S.; Stein, Jeremy C. (1990): Herd Behavior and Investment. In: American Economic Review, Vol. 80, p. 465–479.

Sevestre, Patrick; Trognon, Alain (1985): A Note on Autoregressive Error Component Models. In: Journal of Econometrics, Vol. 28, p. 231–245.

Smith, Lones; Sørensen, Peter (2000): Pathological Outcomes of Observational Learning. In: Econometrica, Vol. 68, No. 2, p. 371–398.

Soetevent, Adriaan R. (2006): Empirics of the Identification of Social Interactions; An Evaluation of the Approaches and their Results. In: Journal of Economic Surveys, Vol. 20, p. 193–228.

Song, Frank M.; Wu, Yangru (1997): Hysteresis in Unemployment. Evidence from 48 U.S. States. In: Economic Inquiry, Vol. 35, p. 235–243.

Stock, James H. (2001): Forecasting Economic Time Series. In: (Baltagi, 2001), chap. 27, p. 562–584.

Stock, James H.; Watson, Mark W. (2011): Dynamic Factor Models. In: Clements, Michael P.; Hendry, David F. (Eds.) Oxford Handbook of Economic Forecasting, Oxford: Oxford University Press, p. 35–60.

Stock, James H.; Watson, Mark W. (2002): Macroeconomic Forecasting Using Diffusion Indexes. In: Journal of Business and Economic Statistics, Vol. 20, No. 2, p. 147–162.

Suedekum, Jens (2005): Increasing Returns and Spatial Unemployment Disparities. In: Papers in Regional Science, Vol. 84, p. 159–181.

SVIMEZ (Ed.) (2009): Rapporto SVIMEZ 2009 sull'Economica del Mezzogiorno [SVIMEZ Report 2009 on the Economy of the Mezzogiorno]. Bologna: Il Mulino.

Swamy, P.A.V.B. (1970): Efficient Inference in a Random Coefficient Model. In: Econometrica, Vol. 38, No. 2, p. 311–323.

Taylor, Jim; Bradley, Steve (1997): Unemployment in Europe: A Comparative Analysis of Regional Disparities in Germany, Italy and the UK. In: Kyklos, Vol. 50, No. 2, p. 221–245.

Tiefelsdorf, Michael; Griffith, Daniel A. (2007): Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach. In: Environment and Planning A, Vol. 39, No. 5, p. 1193–1221.

Tortorice, Daniel L. (2012): Unemployment Expectations and the Business Cycle. In: The B.E. Journal of Macroeconomics, Vol. 12, No. 1: Topics.

Tracy, Craig A.; Widom, Harold (1994): Level-Spacing Distributions and the Airy Kernel. In: Communications in Mathematical Physics, Vol. 159, p. 151–174.

Tyrowicz, Joanna; Wójcik, Piotr (2011): Nonlinear Stochastic Convergence. Analysis of Regional Unemployment Rates in Poland. In: The Review of Economic Analysis, Vol. 3, p. 59–79.

Tyrowicz, Joanna; Wójcik, Piotr (2010a): Regional Dynamics of Unemployment in Poland. A Convergance Approach. In: Caroleo, F.E.; Pastore, F. (Eds.) The Labour Market Impact of the EU Enlargement: A New Regional Geography in Europe, Berlin, Heidelberg: Springer, p. 149–173.

Tyrowicz, Joanna; Wójcik, Piotr (2010b): Unemployment Convergence in Transition. In: Marelli, E.; Signorelli, M. (Eds.) Economic Growth and Structural Features of Transition, London: Palgrave Macmillan, p. 236–259.

van Ommeren, J.; Rietveld, P.; Nijkamp, P. (1997): Commuting: In Search of Jobs and Residences. In: Journal of Urban Economics, Vol. 42, p. 402–421.

vom Berge, Philipp (2012): Search Unemployment and New Economic Geography. In: The Annals of Regional Science, Vol. 50, No. 3, p. 731–751.

Wang, Xiaokun; Kockelman, Kara M. (2009): Bayesian Inference for Ordered Response Data with a Dynamic Spatial Ordered Probit Model. In: Journal of Regional Science, Vol. 49, No. 5, p. 877–913.

Wasmer, E.; Zenou, Y. (1997): Equilibrium Urban Unemployment. Discussion Paper 368, Centre for Economic Performance (CEP).

Welch, Ivo (2000): Herding among Security Analysts. In: Journal of Financial Economics, Vol. 58, p. 369–396.

White, Halbert (1980): A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. In: Econometrica, Vol. 48, p. 817–838.

Wooldridge, J.M. (2002): Econometric Analysis of Cross Section and Panel Data. Cambridge (MA): MIT Press.

Zarnowitz, Victor; Lambros, Louis A. (1987): Consensus and Uncertainty in Economic Prediction. In: Journal of Political Economy, Vol. 95, No. 3, p. 591–621.

Zenou, Yves (2011): Search, Wage Posting, and Urban Spatial Structure. In: Journal of Economic Geography, Vol. 11, No. 3, p. 387–416.

Zenou, Yves (2009): Urban Labor Economics. Cambridge (UK), New York: Cambridge University Press.

Zenou, Yves; Smith, Tony E. (1995): Efficiency Wages, Involuntary Unemployment and Urban Spatial Structure. In: Regional Science and Urban Economics, Vol. 25, p. 547–573.

Zierahn, Ulrich (2012): Regional Labour Market Disparities – A New Economic Geography Perspective. Ph.D. thesis, University Kassel.

Zivot, Eric; Andrews, Donald W.K. (1992): Further Evidence On The Great Crash, The Oil Price Shock, and The Unit Root Hypothesis. In: Journal of Business and Economic Statistics, Vol. 10, p. 251–270.

Abstract

In the dynamics of local labour markets there exist interdependencies between regions which generate spatial patterns with regard to both observable measures (e.g. unemployment rates or employment growth) and economic relations (e.g. the size of policy effects or the speed of adjustment to impulses). Conventional methods of the spatial econometric mainstream are in principle suitable to account for these patterns. However, results are sensitive with regard to the underlying assumptions and structures – which are rarely subject of critique.

This thesis argues for careful application of spatially autoregressive methods if such a process is necessary for answering the issue of research. If it is not, the thesis encourages using alternative, often non- or semi-parametric methods. These are more robust with regard to misspecification and allow – when chosen suitably – a more focused investigation. Three applications in the field of labour-market related regional studies serve as examples.

An analysis of the persistence of shocks with spatially heterogeneous convergence parameters shows a high degree of hysteresis of regions after local shocks. The results emphasize the importance of regional structures (with a size in between NUTS I and NUTS II) in the convergence of local labour markets. After a shock, rural areas approach a potential equilibrium faster than cities or agglomeration cores.

Specification tests for a multiregional forecasting model confirm a polycentric structure in the regional labour market development in Germany. Neither is there evidence of a dominating leader region nor are there stable common trends in the long run across the regions. Accounting for recent spatial co-development with the neighbouring regions improves the accuracy of short-run labour market forecasts. The performance gain has a similar size as that due to the information provided by standard business cycle or weather indicators.

To what extent expectations of local policy makers are due to common regional developments and to what extent they are influenced by the expectations of their neighbours or peers, respectively, is the issue of the third study. The focus is on the identification and isolation of this peer effect. Due to the strong peer effect, soft and private information on the future of the labour market which is available only to a few persons quickly circulates to other persons – that the expectations of persons in a survey are subject to dynamically learning within these social or regional networks and hince provide signals on turning points faster than they show up in hard data is exactly the effect prospective indicators such as the "IAB Konjunkturbarometer" exploit.

Kurzfassung

Hinsichtlich der Dynamik lokaler Arbeitsmärkte bestehen Wechselwirkungen zwischen Regionen, die räumliche Muster sowohl in beobachtbaren Größen (etwa der Arbeitslosigkeit oder dem Beschäftigungswachstum) wie auch in Wirkungszusammenhängen (etwa der Größe von Politikeffekten oder der Dauer, die Schocks nachwirken) erzeugen. Herkömmliche Methoden der räumlichen Ökonometrie sind zwar prinzipiell in der Lage, diesen Mustern Rechnung zu tragen. Allerdings sind ihre Ergebnisse sensitiv bezüglich der zugrunde gelegten Annahmen und Strukturen – die aber nur selten hinterfragt werden. Die vorliegende Dissertation argumentiert für eine selbstkritische Verwendung von räumlich-autoregressiven Verfahren, wenn ein derartiger Prozess zur Beantwortung einer Fragestellung notwendig ist, und andernfalls für eine stärkere Nutzung von alternativen, oftmals nicht- oder semiparametrischen Methoden: Letztere sind robuster hinsichtlich von Fehlspezifikationen und erlauben – bei entsprechender Auswahl – eine stärker fokussierte Analyse. Drei Beispiele aus der arbeitsmarktbezogenen Regionalforschung verdeutlichen dies.

Eine Untersuchung der Persistenz von Schocks mithilfe von räumlich heterogenen Konvergenzparametern zeigt erstens das insgesamt lange Verharrungsvermögen von Regionen nach lokalen Schocks. Zweitens wird die die Bedeutung einer mittleren regionalen Struktur (unwesentlich kleiner als die meisten Flächenländer) in der Konvergenz von lokalen Arbeitsmärkten bestätigt. Drittens kehren ländliche Räume schneller zu einem potenziellen Gleichgewicht zurück als Kernstädte bzw. Agglomerationszentren.

Spezifikationstests für ein Multiregionales Prognosemodell bestätigen eine polyzentrische Struktur der regionalen Arbeitsmarktentwicklung in Deutschland. Weder gibt es eine eindeutige Führungsregion noch zeigen sich langfristige gemeinsame Trends zwischen Regionen. Auf kurze Frist bestätigt sich der Prognosegehalt gemeinsamer regionaler Entwicklungen. Das Verbesserungspotenzial liegt in einer ähnlichen Größenordnung wie der Informationsgehalt von gängigen Konjunktur- und Wetterindikatoren.

Wie stark die Erwartungen von lokalen Entscheidungsträgern von gemeinsamen regionalen Entwicklungen und wie stark sie von den Erwartungen ihrer Nachbarn (bzw. ihrer Peers) getrieben sind, wird abschließend untersucht. Das Augenmerk liegt auf der Identifikation und Isolation des Peer-Effektes. Der starke Peer-Effekt führt dazu, dass weiche Informationen über künftige Arbeitsmarktentwicklungen, die nur wenigen zur Verfügung steht, schnell an andere weitergeben wird – prospektive Indikatoren wie das IAB-Konjunkturbarometer bauen genau darauf, dass die Erwartungen der befragten Akteure dieser Dynamik in sozialen bzw. regionalen Netzen folgen und Umschläge schneller anzeigen, als dies in den realen Entwicklungen sichtbar ist.

Empirical research in economics and social studies often treat regions within a country as independent islands. Moreover, they implicitly assume that effects or relations are homogenous across all regions. Both assumptions seem unrealistic: for example, the impact of a global shock is likely to vary from one region to another.

In this issue on regional labour markets, Norbert Schanne employs novel methods in spatial econometrics to describe and forecast their development. The analysis particularly focusses on the heterogeneity of regional dynamics and the spatially structured interdependency between locations.

ISBN 978-3-7639-4095-0

W. Bertelsmann Verlag

Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit
Weddigenstraße 20-22 ▪ 90478 Nürnberg ▪ www.iab.de