TI 2006-020/3
Tinbergen Institute Discussion Paper

New Neural Network Methods for Forecasting Regional Employment

*Roberto Patuelli**
*Aura Reggiani***
*Peter Nijkamp**
*Uwe Blien****

** Dept. of Spatial Economics, Free University of Amsterdam, The Netherlands;*
***Dept. of Economics, University of Bologna, Italy;*
**** Institut für Arbeitsmarkt und Berufsforschung (IAB), Nuremberg, Germany.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Paper submitted to *Spatial Economic Analysis*

NEW NEURAL NETWORK METHODS FOR FORECASTING REGIONAL EMPLOYMENT: AN ANALYSIS OF GERMAN LABOUR MARKETS

Roberto Patuelli[*], Aura Reggiani[**], Peter Nijkamp[*] and Uwe Blien[***]

[*] Department of Spatial Economics, Free University of Amsterdam, The Netherlands
[**] Department of Economics, University of Bologna, Italy
[***] Institut fuer Arbeitsmarkt und Berufsforschung (IAB), Nuremberg, Germany

ABSTRACT

In this paper, a set of neural network (NN) models is developed to compute short-term forecasts of regional employment patterns in Germany. NNs are modern statistical tools based on learning algorithms that are able to process large amounts of data. NNs are enjoying increasing interest in several fields, because of their effectiveness in handling complex data sets when the functional relationship between dependent and independent variables is not explicitly specified.
The present paper compares two NN methodologies. First, it uses NNs to forecast regional employment in both the former West and East Germany. Each model implemented computes single estimates of employment growth rates for each German district, with a 2-year forecasting range. Next, additional forecasts are computed, by combining the NN methodology with Shift-Share Analysis (SSA). Since SSA aims to identify variations observed among the labour districts, its results are used as further explanatory variables in the NN models.
The data set used in our experiments consists of a panel of 439 German districts. Because of differences in the size and time horizons of the data, the forecasts for West and East Germany are computed separately. The out-of-sample forecasting ability of the models is evaluated by means of several appropriate statistical indicators.

Keywords: neural networks, forecasts, regional employment, shift-share analysis, shift-share regression

JEL codes: C23, E27, R12

1. Introduction

The need for accurate forecasts of modern socio-economic (regional and national) systems has been growing in recent years. Most economic interventions, such as federal or EU funds distribution, require adequate policy preparation and analysis, usually made well in advance, and, often, at a disaggregated level. In this context, an emerging problem is the increasing level of disaggregation for which economic data are collected, and, hence, the imbalance between the number of disaggregated (regional) figures to be forecasted, and the quantity of observations (usually years) available. Although traditional econometric techniques can be useful in this respect, it is well known that, in addition to the many constraints and hypotheses that these econometric models have to cope with, such as the use of fixed regressors, the choice of the model specification – and, most important, of the explanatory variables to use – is crucial. An alternative approach, able to overcome some of these limitations, such as the choice of model and functional variables – especially in the framework of short-term forecasts – is provided by neural networks (NNs), a family of non-linear statistical optimization methods, which can provide a means to override some of these limitations (see, for example, Cheng and Titterington 1994). The NNs' capacity of learning from the data, and of finding functional relationships among variables, makes it possible to forgo strict statistical assumptions and specification problems, and to process data by means of a flexible statistical tool.

The present paper is concerned with the use of NNs in order to forecast regional employment change. Employment data are necessary in economic and regional policy analysis. Pension systems, social security reforms and annual policy-making tasks, such as the establishment of budget allocations, require detailed employment forecasts. The case study under analysis is the evolution of labour markets in Germany. In particular, our NN experiments focus on short-term employment forecasts, that is, forecasts for 2 years ahead. The paper describes a set of NN models developed with this aim in mind, and reviews the validation process and the statistical results of the NN models, which are evaluated for various test years.

The aim of our experiments is not the use of NNs in itself, since nowadays NNs are widely used in different research fields, but the exploration of the NNs' ability to forecast changes in economic variables in a panel data framework. While applications of NNs to time series – or to other pattern recognizance settings – are rather frequent, contributions on NNs dealing with panel data are limited (see, for example, Lin 1992), since panel data are usually analysed by means of various appropriate and specific panel estimation techniques (see, for example, Blien and Tassinopoulos 2001; Bade 2005). The high number of cross-sections in the data under analysis and the limited number of years for which information is available are a problematic issue for conventional econometric techniques. Here lies the rationale underlying our methodological choice of NN techniques.

A novel part of this paper is the incorporation of shift-share analysis (SSA). We will introduce several variants of SSA, including some modern specifications, known as spatial shift-share and shift-share regression (SSR). This class of methods will be integrated with the NN methodology employed in our paper. This may provide an interesting balance between a data-driven technique and a solid well-known research method.

The paper is organized as follows. Section 2 briefly illustrates NN theory, as well as the criteria to be used in the validation of their results. Then, Section 3 will introduce various classes of shift-share techniques. Section 4 will describe the data used in our experiments. Section 5 first explains the practical steps in the implementation of the NN models, and, subsequently, reviews the statistical results of the empirical application, which aims to estimate employment variations in the former West and East Germany for the year 2003. A new contribution to NN analysis is offered by embedding SSA components. The results of the NN models – comprising the NN models embedding SSA components –

are evaluated by means of appropriate statistical indicators and map visualizations. Finally, Section 6 offers some conclusions and sets future research directions.

2. Neural Network Models for Analysis and Forecasting

2.1. Neural Networks as a Statistical Optimization Tool

Neural networks (NNs), sometimes also called ‘artificial neural networks’, in order to differentiate them from actual biological networks, are optimization algorithms, whose main characteristic is the ability to find optimal goodness-of-fit solutions when the relationships between the variables are not fully or explicitly known, or when only a limited knowledge of the phenomenon examined is available. While traditional statistical models require an identification process for the set of regressors employed, as well as a specification of the relationship between dependent and independent variables, these steps are not necessary in NNs. Their no-modelling hypothesis could be considered a drawback in this regard because of the lack of theoretical economic (or behavioural) interpretation, which forces the analyst to accept the data-driven results of the NN models ‘as they are’. On the other hand, the limited possibilities of interpretation of the results is less relevant when the aim is, as in our case, to produce forecasts, rather than to explain the relationships between the driving factors. In addition, NNs are also more robust against statistical noise, since they store redundant information. In contrast with conventional statistical techniques, NNs do not efficiently process categorical variables when these have many ‘values’, while there is no set of unifying and optimal NN models. As a consequence, the performance of NNs is dependent on the implementation carried out by the analyst.

Because of their relatively simple application, NNs are attractive in various fields of socio-economic application. Reviews of NNs used in several fields can easily be found in the literature. Many examples could be listed, as well as academic journals entirely dedicated to the NN-related studies. A very concise and non-exhaustive selection of these is shown in Table 1. Generally, it should be underlined that NNs enjoy great scalability properties, as they can be applied to problem-solving related to practically any application area.

Table 1 – Some illustrative reviews of NN applications in different fields; NN journals

Field	*Authors*
Atmospheric sciences	Gardner and Dorling 1998
Business and finance	Wong et al. 1997; Wong and Selvi 1998; Chatterjee et al. 2000
Classification of medical data	Dreiseitl and Ohno-Machado 2002
Environmental modelling	Maier and Dandy 2000; Shiva Nagendra and Khare 2002
Medical imaging and signal processing	Miller et al. 1992
Transportation	Himanen et al. 1998
Neural Computing & Applications	(journal)
Neural Computing Surveys	(journal)
Neural Networks	(journal)
Neural Processing Letters	(journal)

Although NNs have sometimes been referred to as a ‘black box’ approach, they are definitely not such an obscure tool. The internal functions that process the different information inputs are, of course, selected by the analyst, as well as the algorithms that determine the direction and the degree of interaction of the factors during the computation process. As a matter of fact, NNs are often compared with conventional statistical methods, such as generalized linear models or simple regressions, in the light of an integrated utilization of all these methodologies. This kind of literature is now wide and diverse (see, among others, Cheng and Titterington 1994; Swanson and White 1997a, 1997b; Baker and Richards 1999; Sargent 2001), covering different fields. For example, Nijkamp et al. (2004) compared NNs with logit and probit models in an analysis of multimodal freight transport choice. In the labour market field, previous works by Longhi et al. (Longhi 2005; Longhi et al. 2005a; Longhi et al. 2005b) should be cited, particularly for their use of panel and cross-sectional data, instead of time series. NNs have also been shown to be equivalent, in the case of binary choice, to a logit model (Schintler and Olurotimi 1998).

Generally, we can define a set of rules for the evaluation and comparison of NNs, which we derive from Collopy et al. (1994):

- Comparison with widely-accepted 'conventional' models. Forecasts from the NN models should be at least as accurate as those generated by a naïve extrapolation, such as random walk.
- Test of the models' out-of-sample performance. The results of out-of-sample forecasts should be used in comparing different methodologies.
- Use of a satisfying sample size. The size of the sample has to allow for statistical inference.

As can be seen later on in the presentation of an empirical application, these three rules are respected in our experiments. In addition to these general validation guidelines, additional rules may also apply with regard to the actual implementation of NN models. These rules are important, in that they define the correct execution of NN modelling experiments, and the presentation of their results. We refer here to Adya and Collopy (1998):

- Provision of the in-sample performance of the models. Sample data provide the basis for the learning process (see next subsection), and are a benchmark for the evaluation of the generalization properties of the NN models.
- Generalization. The level of similarity between in- and out-of-sample performance provides an indication of the generalization potential of the models. In this regard, a generalization estimator was computed by the authors (see Patuelli et al. 2003).
- Stability. A similar performance over different data sets allows the stability of the forecasting tool, and its reliability, to be assessed.

Several attempts have been made to assess the usefulness or effectiveness of NNs. Some authors (see, for example, Swanson and White 1997a, 1997b; Stock and Watson 1998) compared NNs with linear and nonlinear methods as forecasting tools for variables such as employment, industrial production, or corporate profits. They came to various conclusions. Stock and Watson (1998) concluded that NNs, and nonlinear methods in general, mainly perform worse than linear methods. On the other hand, Swanson and White (1997b, p. 459) suggest that it could be possible to improve macroeconomic forecasts 'using flexible specification econometric models', whose specification 'is allowed to vary over time, as new information becomes available'. Finally, Adya and Collopy (1998), found that, most of the time, NNs seem to provide better forecasts than the models with which they are compared. Examining a string of studies which developed NNs for business forecasting, they find that, of the studies correctly validating and implementing the NN models, 88 per cent show that the NNs have a superior performance.

In order to fully understand the implications of the aforementioned rules and methodological comparisons, we first need to describe the functioning of NNs. The next subsection will give a very brief discussion of the main components and interactions of an NN.

2.2. Backgrounds of Neural Networks

Scientists have long been interested in the use of artificial NNs that could replicate the type of simultaneous information processing and data-driven learning seen in biological networks. Since Rosenblatt's first introduction of an artificial NN (Rosenblatt 1958) and the works of Werbos (1974), who provided a proper mathematical framework, and those of Rumelhart and McClelland (1986), who developed the most commonly used error-correction algorithm (backpropagation), many developments have been made in the NN framework.

NNs can be defined as systems of units (or *neurons*) that are distributed in layers and are internally connected. The layers comprise units, which can either refer to input variables (first layer), or to output variables (last layer). Intermediate layers composed of hidden units can also be used. When counting the number of layers of a NN, the input layer is usually not considered, since it does not take part in the data computation. Therefore, a NN with no hidden units has a 1-layer structure, while, accordingly, a NN with one layer of hidden units has a 2-layer structure.

In *feedforward* NNs, every unit from each layer is connected – and transfers information – to every unit of the next layer, while connections between pairs of units only go in one direction (there are no cycles, as in other types of NNs, such as *recurrent* NNs). Consequently, the input units are only connected to the units of the first hidden layer (if employed), while the output units are only connected to the neurons belonging to the preceding (hidden) layer. It follows that, in the case of a single hidden layer, this is the only intermediate level between input and output units, while, when a hidden layer is not employed in the NN, input and output units are directly linked. Figure 1 provides a graphic illustration of the structure of an NN.

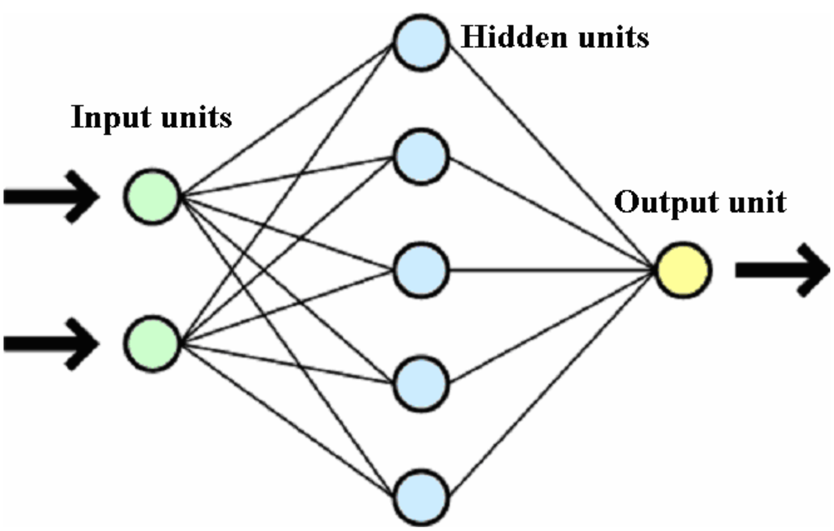

Figure 1 – A graphical illustration of a feedforward NN

Source: Image licence is with Creative Commons.[1]

Fischer (2001b, p. 23) defines the generic processing unit u_i, belonging to $\boldsymbol{u} = \{u_1, ..., u_k\}$, as:

$$u_i = \varphi_i(\boldsymbol{u}) = \Im_i(f_i(\boldsymbol{u})), \tag{1}$$

where the function φ_i can be decomposed into two separate functions: $\Im_i$ is the activation function, and f_i is the integrator function. The activation function computes each unit's output, and is usually constant over the same NN.[2] The integrator function is used for aggregating the information processed by the units of the preceding layer. This is done by combining the inputs by means of a set of weights, contained in vector $\boldsymbol{w_i}$. The function commonly used for this task is a weighed sum:

$$f_i(\boldsymbol{u}) = \sum_j w_{ij} u_j , \tag{2}$$

where u_j is the j^{th} unit connected to unit u_i, and w_{ij} is the connection weight associated with the two units (Fischer 2001a). The 'learning process' of an NN is guaranteed by the recursive modification of the aforementioned weights, through which the NN can identify significant rules in data occurrence (see, for example, Rumelhart and McClelland 1986). The 'knowledge' generated by the NN is therefore contained in the set of weights that are computed.

In order to find the optimal values for the NN weights, a learning algorithm is needed, which normally involves iterative computations. The Back-Propagation Algorithm (BPA) is the one most commonly used for this task. The BPA requires the analyst to provide input examples and their correct – and known – outputs. NN models that follow this kind of process are called *supervised* NNs. The sample data used allow the models to identify the behaviour underlying the data and to replicate it. The actual learning process is given by the comparison of the output generated from the current weight configuration[3] with the correct output, by means of a backward propagation of the obtained error[4] through the network. This process is repeated for each record of the sample, with a consequent re-adjustment of the weights. The cycle's stopping condition can be decided by the analyst on the basis of, for example, computing time, error level, or the number of iterations. It should be noted that the algorithm 'will never exactly learn the ideal function, but rather it will asymptotically approach the ideal function' (McCollum 1998), in addition to the local-minima problems that can arise.[5]

After this brief description of NN methods, we now offer a brief overview of shift-share methods as a complement to NN approaches, which can be integrated in a meaningful way to improve the statistical results of our experiments.

3. Shift-Share Analysis for Regional Growth Analysis

3.1. The Conventional Shift-Share Analysis Identities

Shift-Share Analysis (SSA) has, since its inception in the 1960s, been a popular analytical tool, among regional scientists and not only, for improving the understanding of changes in economic variables, such as employment or GDP, at the regional level. SSA can be usually employed in three ways: a) in forecasting; b) in strategic planning (that is, observing the weight of the effects); c) in policy evaluation (before-and-after analysis); and d) in decision making (Dinc et al. 1998; Loveridge and Selting 1998).

SSA was first introduced by Dunn (1960), and subsequently formalized by Fuschs (1962) and Ashby (1964). In SSA, the growth showed by economic variables is decomposed into several components. Using employment as an example, the conventional shift-share decomposition can be written as:

$$\Delta e_{ir} = [G + (G_i - G) + (g_{ir} - G_i)]e_{ir} \,, \tag{3}$$

where e_{ir} is the employment observed in region r for sector i; G is the overall national employment growth rate; G_i is the national growth rate of sector i; and g_i is the growth rate of region r in sector i. The employment growth rate Δe_{ir} is therefore decomposed into three components:

i) the *national effect* G;
ii) the *sectoral effect*, given by the difference between the sectoral and overall national growth rates, G_i and G;
iii) the *competitive effect*, given by the difference between the local and nationwide sectoral growth rates, g_i and G_i.

Each of the three components can be calculated for each region, over all the sectors, and nationwide. In particular, when summed nationwide, the sectoral and competitive effects sum to zero. This property is usually referred to as the 'zero national deviation' (ZND) property. The above identity has been studied and modified by several authors over the years. Alternative formulations of SSA also include an industry-structure approach where, in place of growth rates, industrial structures are compared (Ray 1990).

However, perhaps he most popular SSA extension is the one developed by Esteban-Marquillas (E-M) (1972):

$$\Delta e_{ir} = Ge_{ir} + (G_i - G)e_{ir} + (g_{ir} - G_i)e_{ir}^h + (g_{ir} - G_i)(e_{ir} - e_{ir}^h). \tag{4}$$

In this SSA formulation, e_{ir}^h is the homothetic employment of sector i in region r. Homothetic employment is calculated as $e_{ir}^h = eE_{ir}/E$, that is, region r's employment in sector i, as it would be if the sector had the same structure as the nation. The homothetic competitive effect (third component) measures 'a region's comparative advantage/disadvantage in [sector] i relative to the nation' (Esteban-Marquillas 1972, p. 43). The fourth and last component is called the *allocation effect*, as it is the product of the expected employment and the differential, which measures a region's competitive advantage in sector i. The claim of this model is that it isolates the competitive effect from its relationship with the sectoral effect. Critiques of the E-M model can be found in Stokes (1974) and in Haynes and Machunda (1987). The E-M extension is not considered in our experiments, since the competitive effect is computed in the same way as in conventional SSA, the only difference being that it is multiplied by the homothetic employment.

More generally, the main criticisms of SSA, according to Loveridge and Selting (1998), concern:

- its lack of theoretical content. In order to fill this gap, there have been attempts to link SSA to neoclassical microeconomics and factor demand for labour;
- aggregation problems. Finer categories increase the weight of the sectoral effect and shrink the competitive effect. But, it has to be remembered that other techniques are also sensitive to aggregation issues;
- weighting bias. It is not clear whether it is more convenient to use the base or the terminal year. Alternatively, the average of the two or a middle year could be used, or a 'dynamic shift-share' formulation (see Wilson 2000);
- instability of the competitive effect. This instability makes employment projections by means of SSA somewhat precarious. On the other hand, this issue does not exclude the use of SSA in forecasting, particularly, in the framework of NNs;

- interdependence of the sectoral and competitive effects.

A number of new SSA specifications have been developed over the years,[6] on the basis of the first technical advances described above, often focusing on the elimination of dependence among shift-share components, or trying to solve other deficiencies of SSA. However, the application of newer methodologies has often deprived the models of their contribution to understanding local phenomena (Loveridge and Selting 1998). While all types of decomposition can be obtained by adding and subtracting variables, all of them can be shown to be rooted in the simple SSA decomposition (Nazara and Hewings 2004). Consequently, the basic models and a few other modifications, widely accepted as standards, are still preferred by most analysts, because of their intuitive and simple specifications.

Despite the above considerations, the development of new SSA extensions still goes on. One of the most recent developments in this matter is the extension proposed by Nazara and Hewings (2004), also called 'spatial shift-share' by the authors, and described in the next subsection.

3.2. Spatial Shift-Share

The development of the recent shift-share extension termed 'spatial shift-share' is justified by the fact that spatial issues, such as spillovers, spatial competition, and so on, have not been considered in the application of SSA. There is therefore a need for the introduction of an element that accounts for the spatial structure which comprises a particular region. If we consider that regions are – as seems logical – interdependent and they influence each other, we note, in fact, that horizontal-influence relationships (region to region) are not enclosed in the traditional SSA formulation, while only hierarchical ones are accounted for (that is, nation to region).

Starting from this consideration, Nazara and Hewings modified the conventional shift-share identity in:

$$\Delta e_{ir} = [G + (\tilde{g}_{ir} - G) + (g_{ir} - \tilde{g}_{ir})]e_{ir}, \tag{5}$$

where $\tilde{g}_i$ is sector i's growth rate in the regions that are neighbours to region r. The neighbours' growth rate $\tilde{g}_i$ is formulated, for a generic (t, $t + n$) period, as:

$$\tilde{g}_i = \frac{\sum_{S=1}^{r} \tilde{w}_{rs} e_{is}^{t+n} - \sum_{S=1}^{r} \tilde{w}_{rs} e_{is}^{t}}{\sum_{S=1}^{r} \tilde{w}_{rs} e_{is}^{t}}, \tag{6}$$

where the employment levels of neighbouring regions are weighted according to a row-standardized weight matrix $\tilde{W}$, which defines the intensity of the neighbours' interaction with region r. This interaction can be defined in many ways: for instance, on the basis of geographical contiguity or economic flows. A simplified version of the weight matrix is employed in this paper, where the neighbours of a given region are empirically defined as the three regions that provide the highest number of individuals commuting toward the region considered.[7] In practical terms, the weight matrix employed here is an asymmetrical matrix with only three identical values differing from 0 for each region. The overall employment growth rate of the neighbours is subsequently computed.

As a consequence of the new variable presented in Equation (5), the sectoral and the competitive components change in meaning. In detail:

- the sectoral component now identifies the difference between the growth rate of region r's neighbours in sector i, and the national all-sector growth;
- the competitive component is the difference between sector i's growth rate in region r and in its neighbouring regions.

This recent decomposition is already the subject of further study and expansion. Fernández and López Menéndez (2005) developed a mixed Nazara-Hewings/E-M model, which employs both homothetic employment and the spatial connotation given by a geographical connectivity matrix. The interest in the SSA framework also goes beyond its deterministic nature. The next subsection describes a stochastic shift-share approach termed 'shift-share regression'.

3.3. Shift-Share Regression

One of the main critiques of SSA is the lack of hypothesis testing, which is due to shift-share's deterministic nature. A stochastic approach, based on regression techniques equivalent to shift-share, has been developed by Patterson (1991), and subsequently used by, among others, Möller and Tassinopoulos (2000), and by Blien and Wolf (2002) in the analysis of employment patterns in Eastern Germany. The model proposed by Patterson is rather simple, and strictly related to the conventional SSA approach:

$$\Delta e_{irt} = \alpha_i + \lambda_t + k_r + \varepsilon_{irt}, \tag{7}$$

where Δe_{irt} is the regional employment growth rate in sector i during period (t, t+1); α_i is the effect of sector i; λ_t incorporates time period t (period effect); k_r is a locational effect specific to region r; and ε_{irt} is stochastic noise. Möller and Tassinopoulos, as well as Blien and Wolf, propose extensions of this specification, incorporating additional variables, such as structural adjustment or region-type indicators, and qualification level of employees. Equation (7) suffers from perfect multicollinearity, and is therefore estimated by introducing a set of constraints (see Blien and Wolf 2002). A Weighted Least Squares (WLS) estimation procedure is suggested in order to reduce the impact of outliers.

This Shift-Share Regression (SSR) approach has been replicated, in this paper, in a simplified version. We are interested in introducing shift-share components in NNs in order to forecast overall regional employment. Therefore, we only employ the locational effects regressors, which are region-specific, as explanatory variables in NN models. In our case, the dependent variable is Δe_r, that is, the overall employment growth rate of region r. Equation (7) is therefore simplified as follows:

$$\Delta e_r = a + k_r + \varepsilon_{rt}. \tag{8}$$

In Equation (8), a is the intercept, while ε_{rt} is the stochastic noise for region r at time t. In this case, the locational effects variable is computed as the competitive effects used in conventional SSA. Consequently, there is a set of locational effects regressors: one for each sector. The model was estimated, by means of WLS,[8] for each 2-year period. We found most of the locational effect variables to be statistically significant (for details, see Tables A.1 and A.2 in Annex A). The multiple per-year estimations seem logical in the NN forecasting framework. The estimation of a single regression coefficient per sector would only change the scale of the independent variables introduced in an NN model, as they are multiplied by the corresponding regression coefficients. Computing a regression for each 2-year period enables what could be seen as a 'fine tuning' of the locational/competitive effect variables, the regression coefficient being different for each year. Certainly, the correctness of this procedure – from a methodological viewpoint – will have to be looked into more in-depth.

On the basis of the considerations of this and of the preceding sections, several NN-SS models were developed, using conventional and 'spatial' SSA formulations, as well as SSR. The next section illustrates the data employed for our analyses, and then Section 5 provides details of the NN models developed and their results.

4. The Data Set on German Regional Labour Markets

The data available for our experiments concern district units in the former West Germany and East Germany. The data on West Germany cover 17 years (1987 to 2003), while the data on East Germany are only available for 11 years (from 1993 to 2003). The number of districts is 326 for West Germany and 113 for East Germany, amounting to a total of 439 districts.

The data sets have been provided by the German Institute for Employment Research (Institut für Arbeitsmarkt und Berufsforschung – IAB), and include information on the number of full-time workers employed every year at 30 June. A graphical visualization of recent regional trends in the data (for the period 2001–03) is displayed in Figure C.1, in Annex C. The above-mentioned regional data are also classified according to nine economic sectors.[9] In addition to these variables, average regional daily wages earned by full-time workers are also available. Furthermore, in an effort to identify labour market patterns in similar regions, the 'type of economic region' variable was adopted. This variable, which is an index ranging from 1 to 9, follows the classification adopted by BfLR/BBR (Bundesforschungsanstalt für Raumordnung und Landeskunde / Bundesanstalt für Bauwesen und Raumordnung, Bonn). In fact, our West and East German districts may be grouped into the following nine economic regions (Bellmann and Blien 2001):

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;
9. Rural districts in regions with rural features.

The data set illustrated above will be the basis for our forecasting experiments, which are described below.

5. Forecasting Regional Employment in West and East Germany

5.1. Forecasting Employment by Means of Neural Networks

This section will illustrate the series of NN models that we developed for our forecasting purposes. The main inputs of our models are the growth rates of the number of workers regionally employed in the nine economic sectors. To exploit the panel structure of our data and – more specifically – the correlation across observations of the same regions over time, we introduced in our models what we describe as the 'time' variable. This variable was identified in the models in two different ways: 1) as a ‘time fixed effect’ in panel models (Longhi et al. 2005b); and 2) as a set of dummy variables. On the basis of these considerations, 12 NN models in total have been adopted, which start from two basic models: a) Model A, which employs time by means of dummy variables; and b) Model B, which employs a fixed effects time variable.

In addition to the time variable, further variables were employed in the NN models. Seven additional NN models have been applied (see Tables B.1 and B.2 in Annex B). Model AC has the same inputs as Model A, plus a qualitative variable able to distinguish between the districts. As in the case of the time fixed effects variable, this can be seen as the correspondent of cross-sectional fixed effects in a panel model (Longhi et al. 2002). Model AD and Model AE have the same inputs as Model A, plus the variable ‘type of economic region’, which was introduced in the two NN models as a qualitative variable (Model AD) and as a set of dummies (Model AE). Also, Model B was enhanced with the qualitative variable ‘type of economic region’, thereby obtaining Model BD. Finally, information about daily wages was introduced as a new input variable: a) in Model A, obtaining Model AW; b) in Model AD, obtaining Model ADW; and c) in Model B, obtaining Model BW.

Additional models were developed, by employing SSA-computed variables. We refer to these models as NN-SS models. As in some of the models presented above, the NN-SS models use Model B as a basis:

- *Model BSS* presents nine additional variables, which are the competitive effect coefficients calculated, for each sector, in the framework of conventional SSA. As a result, for each German district and each year, we have nine coefficients expressing regional competitiveness.
- Similarly, *Model BSSN* employs the competitive effect coefficients deriving from the Nazara and Hewings SSA extension.
- Finally, *Model BSSR* embeds variables computed in the SSR framework. The variables employed in this model are the product of the multiplication of the competitive effect variables used in Model BSS, and their regression coefficients, found in the analysis explained in Section 3 (for details on the coefficient values, see Tables A.1 and A.2 in Annex A).

The characteristics of the various models presented are summarized in Annex B. All the models adopted use, as input variables, the growth rate of the sectoral employment. Since, for each year, the NNs were trained on the basis of the 2-year lagged employment variations, the data used in our NN models started from 1991 (1989-1991) for West Germany and from 1997 (1995-1997) for East Germany.[10] The data set available for West Germany is six years longer and allows for larger training and testing periods.

Table 1 – Data utilization for validating the network configuration

Models	*Training*	*Validating*
West Germany	1991-1998	1999-2000
East Germany	1997-1999	2000

The first test phase (referred to as the validation phase), which is summarized by Table 1, concerned the validation of a number of network configurations (see, for example, Fischer 1998). For all NN models, we employed data until the year 2000. NN models related to the case study of West Germany were trained from 1991 to 1998, while NN models for East Germany were trained from 1997 until 1999. For validating the models, two 2-year test sets have been used in the case of West Germany (1999–2000), while one 2-year test set has been chosen for East Germany (2000). The use of two test sets in the choice of the NN structure is justified by the fact that the performance of the NNs is not uniform for different test sets. The use of statistical indicators calculated on a two-period basis may lead to choices that are less influenced by shocks that could have affected a particular year. However, experiments on East Germany had to be carried out on just one test period, since, because of the limited coverage of the data, only a few years would have been available for the learning process of the NN. For every NN model, we experimented with five structures in the initial stage. First, a 1-layer structure (see Section 2.2) was tried out, followed by three 2-layer models containing 5, 10 and 15 neurons, respectively, in one hidden layer. Finally, a 3-layer model was attempted, using 5 neurons for each of the two hidden layers.[11] The models trained as described above were subsequently evaluated by means of several statistical indicators.[12] The best-performing settings were then chosen for further development of the NNs.

In the subsequent test phase, the evaluation of the chosen structures was provided by *ex post* tests carried out for the year 2003 – for which actual data were available. Table 2 summarizes which data were used at this stage. In this phase, the weights were reset and the models were retrained from their respective initial year until the year 2002. The objective of this procedure was to obtain *ex post*, out-of-sample forecasts for the year 2003 that could be compared with the actual data, in order to evaluate the models' generalization properties.[13]

Table 2 – Data utilization for the test phase

Models	*Training*	*Testing*
West Germany	1991-2002	2003
East Germany	1997-2002	2003

The next sections will explain and discuss the empirical findings from our experiments. First, the results obtained for West Germany will be shown and examined (Section 5.2), followed by those found for East Germany (Section 5.3).

5.2. Estimation of West German Employment

As indicated in the previous section, 12 different models were developed and tested for each data set. The first step was the choice of the NN structure (in terms of number of layers and hidden neurons). The models were compared with respect to several configurations, using the years from 1991 to 1998 as the training period, and the years 1999 and 2000 (growth rates for 1997–99 and 1998–2000) as a validation period (see Table 1). The indicators computed on the basis of the years 1999 and 2000 were calculated on the basis of percentage employment variations. Further details on the structures of the NN models that were finally chosen can be found in Annex B (Table B.1). The models were then retrained until the year 2002, while the year 2003 acted as a test set (see Table 2). The statistical indicators emerging from these experiments are presented in Table 3. These results assess the statistical performance of the NN models, and will be the basis for the choice of a reduced set of models that will be adopted for actual employment forecasts (in this case, for the year 2005).

It is clear from Table 3 that the models which use Model B as a base (we will call them B-type models) and, in particular, Model BW, perform better than the others (which we call A-type models). Specifically, Models BSS and BSSN, embedding SSA, seem to provide promising results, improving on the performance of the simpler Model B. Also, the B-type models mostly outperform a naïve no-change random walk (see Theil's U statistic), while the A-type do not. Finally, it is important, in the evaluation of the NN and NN-SS models, to note that the B-type models exceed, in the *ex post* forecasts, their own statistical performance in the training set, while, again, the A-type models do not.

Table 3 – Statistical performances of the *ex post* forecasts for the year 2003: the case of West Germany

	A-type models						*B-type models*			*NN-SS models*		
	Model A	Model AC	Model AD	Model ADW	Model AE	Model AW	Model B	Model BD	Model BW	Model BSS	Model BSSN	Model BSSR
							Training					
MSE	6272983	9978277	5738632	6070688	5620431	7191179	19924559	19701038	25194368	22340774	22810078	21735874
MAE	1329.03	1620.98	1284.71	1364.94	1361.20	1410.64	2292.79	2344.16	2586.07	2446.76	2447.85	2373.33
MAPE	2.1312	2.4899	2.0652	2.1650	2.1963	2.2195	3.4124	3.6046	3.8323	3.6286	3.6272	3.5273
							Testing					
MSE	19924131	20653281	48389433	38130097	30658822	45534811	8464111	9769103	**6887958**	7190785	7902621	22116964
MAE	2612.97	2708.09	4283.15	3924.97	3484.53	4114.05	1661.20	1717.71	**1415.10**	1520.36	1584.99	2428.30
MAPE	5.0166	5.2581	8.1696	7.7009	6.7716	7.8912	3.3038	3.1697	**2.8078**	3.0592	3.1406	3.6179
Theil's U	1.3622	1.4120	3.3083	2.6069	2.0961	3.1132	0.5787	0.6679	**0.4709**	0.4916	0.5403	1.5121

Similarly to what has been presented above, the next section will illustrate the statistical results for the NN models forecasting employment in East Germany.

5.3. Estimation of East German Employment

The data set for East German employment contains information on the number of employees for 113 districts. Data are available for the period between 1993 and 2003. The data set is therefore smaller than that for West Germany (which comprises 326 districts from 1987 to 2003) and 6 years shorter. Consequently, only five years could be used for training, validating, and testing the models (see Table 1). The NN models were selected, structure-wise, by training the models from the year 1997 to the year 1999, and tested on the year 2000 (growth rate for 1998–2000). Annex B (Table B.2) provides the details on the structure and parameters of each NN model. The aforementioned models were subsequently trained until the year 2002, employing the year 2003 as a test period (see Table 2). The statistical results of the East German NN models for the 2003 *ex post* forecasts are presented in Table 4.

Table 4 - Statistical performances of the *ex post* forecasts for the year 2003: the case of East Germany

	A-type models						*B-type models*			*NN-SS models*		
	Model A	Model AC	Model AD	Model ADW	Model AE	Model AW	Model B	Model BD	Model BW	Model BSS	Model BSSN	Model BSSR
							Training					
MSE	22158313	19952364	8596534	21268095	9034762	21858611	37252966	33940993	33799007	32600242	38854476	31626312
MAE	1679.79	1544.25	1527.65	1727.03	1492.37	1697.19	2011.37	1902.13	1825.46	1901.07	2036.55	1868.62
MAPE	3.4297	3.1888	3.5745	3.5303	3.4504	3.4643	3.8528	3.6353	3.4996	3.7358	3.8583	3.6847
							Testing					
MSE	9614821	11553786	34344579	24497503	18994772	14620784	1016209	1381412	5556952	1194348	1400856	**916426**
MAE	1130.16	1536.84	2026.88	1697.88	1387.79	1371.51	618.41	714.47	1023.88	633.85	645.65	**595.45**
MAPE	3.1412	4.5443	5.1718	4.4177	3.4920	3.7587	2.1493	2.4442	2.8824	2.1511	2.1044	**2.0957**
Theil's U	0.2459	0.2955	0.8784	0.6266	0.4858	0.3740	0.0260	0.0353	0.1421	0.0305	0.0358	**0.0234**

Table 4 shows results that seem to be consistent with the ones obtained for the West German NN models, presented in Table 3. As in the West German case, the B-type models – based on time as a fixed effect – display most of the lowest errors for all the indicators. The NN-SS models, and, in particular, Model BSSR, employing SSA/SSR components, suggest an enhanced generalization power compared with the base model (Model B). The NN-SS models provide most of the best estimates, ranking among the top models in every statistical indicator.

The consistent results between the West and East German NN models make for interesting considerations, which will be illustrated in the next, concluding, section.

6. Conclusions

The aim of this paper was to make forecasts – at the time (t+2) – of the number of individuals employed in 339 NUTS 3 districts in Germany. For this purpose, several models – based on NN techniques – were developed. In particular, the districts were divided into West German and East German district data sets. Separate NN models were subsequently developed for the two zones.

The results of *ex post* forecasts for the year 2003 were evaluated by means of several statistical indicators (see Tables 3 and 4). In particular, we were interested in observing the results of NN models employing SSA/SSR variables. Our results lead to the following considerations:

a) The models' performance shows different error levels, for both the West and East data sets. From a preliminary observation of Tables 3 and 4, the models utilizing the variable 'time fixed effect' (B-type models and NN-SS models) seem to forecast better than the remaining models (A-type models). In fact, they provide the lowest error levels for both the West and East Germany models.
b) Through all our experiments, we searched for an NN model that could be considered as the most consistent and reliable. While previous work by Patuelli et al. (2004) found Model B to respond to these criteria (shift-share NN models were not included), the NN-SS models (SSA/SSR-enhanced) presented here seem to improve the performance of Model B. They displayed, for both West and East Germany, errors that were among the lowest found, only competing with the other B-type models. The A-type models, as said before, do not seem to be competitive.

In conclusion, our aim was to experiment and test NN models that could provide reliable forecasts for German employment at a district level. In doing so, we experienced different levels of result reliability, depending on different data sets and socio-economic background. It has to be remarked that most of our empirical analysis has been based on only a few main variables (such as employment, type of district, and wages). Thus, it can not be comprehensive with regard to the many variables that come into play when employment and social conditions are at stake. A step in this direction was the introduction of the SSA/SSR-enhanced NN models. By embedding shift-share components in the NNs, we move in the direction of integrating linear and non-linear methods. In addition, as in the case of Model BSSN, we also incorporate spatial information. The incorporation in the NNs of information on the performance of the 'neighbours' allows us to fill one of the gaps of conventional SSA, and maybe of NNs, that is, they do not include the spatial characteristics of the data.

Further directions for research, from the empirical viewpoint, are concerned with addressing the need for a longer data span enriched with more variables (for example unemployment or migration). Also, a comparison of the accuracy of forecasts for the (t+1) and (t+2) periods might help in evaluating the usefulness of neural computing for labour markets.

On the methodological side, it might be desirable to carry out a multicriteria analysis that could, if it were based on several appropriate criteria, objectively evaluate the models in terms of the basis of the final user's information needs. In addition, an actual integration of linear methods with NNs should be a main objective. Fulfilling such a task would make it possible to combine the benefits of both families of methods in a more complete approach to labour market analysis. This could, therefore, be exploited in the NN forecasting.

Also, a more in-depth analysis of the spatial linkages among districts in terms of (un)employment growth might help to achieve a better understanding of the regional phenomena. In this framework, the utilization of methods such as spatial filtering (Griffith 2003), possibly in a joint NN approach, seem to be desirable, in particular, from a policy perspective.

Acknowledgements

The authors wish to thank Prof. Günter Haag for kindly providing data on commuting flows. The first author also thanks Prof. Kingsley Haynes for a useful discussion of SSA.

References

Adya, M. and F. Collopy (1998) How Effective Are Neural Networks at Forecasting and Prediction? A Review and Evaluation. *Journal of Forecasting* 17 (5-6), pp. 481-95

Ashby, L.D. (1964) The Geographical Redistribution of Employment: An Examination of the Elements of Change. *Survey of Current Business* 44 (10), pp. 13-20

Bade, F.-J. (2005) Evolution of Regional Employment in Germany: Forecast 2001 to 2010. In: *Spatial Dynamics, Networks and Modelling*, A. Reggiani and P. Nijkamp (eds). Cheltenham, Edward Elgar, pp. 293-318 (forthcoming)

Baker, B.D. and C.E. Richards (1999) A Comparison of Conventional Linear Regression Methods and Neural Networks for Forecasting Educational Spending. *Economics of Education* 18, pp. 405-15

Bellmann, L. and U. Blien (2001) Wage Curve Analyses of Establishment Data from Western Germany. *Industrial and Labor Relations Review* 54, pp. 851-63

Blien, U. and A. Tassinopoulos (2001) Forecasting Regional Employment with the ENTROP Method. *Regional Studies* 35 (2), pp. 113-24

Blien, U. and K. Wolf (2002) Regional Development of Employment in Eastern Germany: An Analysis with an Econometric Analogue to Shift-Share Techniques. *Papers in Regional Science* 81 (3), pp. 391-414

Chatterjee, A., O.F. Ayadi and B.E. Boone (2000) Artificial Neural Network and the Financial Markets: A Survey. *Managerial Finance* 26 (12), pp. 32-45

Cheng, B. and D.M. Titterington (1994) Neural Networks: A Review from a Statistical Perspective. *Statistical Science* 9 (1), pp. 2-30

Collopy, F., M. Adya and J.S. Armstrong (1994) Principles for Examining Predictive Validity: The Case of Information Systems Spending Forecasts. *Information Systems Research* 5 (2), pp. 170-79

Cooper, J.C.B. (1999) Artificial Neural Networks versus Multivariate Statistics: An Application from Economics. *Journal of Applied Statistics* 26, pp. 909-21

Dinc, M., K.E. Haynes and L. Qiangsheng (1998) A Comparative Evaluation of Shift-Share Models and their Extensions. *Australasian Journal of Regional Studies* 4 (2), pp. 275-302

Dreiseitl, S. and L. Ohno-Machado (2002) Logistic Regression and Artificial Neural Network Classification Models: A Methodology Review. *Journal of Biomedical Informatics* 35 (5/6), pp. 352-59

Dunn, E.S. (1960) A Statistical and Analytical Technique for Regional Analysis. *Papers and Proceedings of the Regional Science Association* 6, pp. 97-112

Esteban-Marquillas, J.M. (1972) A Reinterpretation of Shift-Share Analysis. *Regional and Urban Economics* 2 (3), pp. 249-55

Fahlmann, S.E. (1992) Comments on comp.ai.neural.nets, item 2198

Fernández, M.M. and A.J. López Menéndez (2005) *Spatial Shift-Share Analysis: New Developments and Some Findings for the Spanish Case*. Paper presented at the 45th Congress of the European Regional Science Association, August, Amsterdam, The Netherlands

Fischer, M.M. (1998) Computational Neural Networks: An Attractive Class of Mathematical Models for Transportation Research. In: *Neural Networks in Transport Applications*, V. Himanen, P. Nijkamp and A. Reggiani (eds). Aldershot, England, Ashgate Publishing Ltd, pp. 3-20

Fischer, M.M. (2001a) Central Issues in Neural Spatial Interaction Modeling: the Model Selection and the Parameter Estimation Problem. In: *New Analytical Advances in Transportation and Spatial Dynamics*, M. Gastaldi and A. Reggiani (eds). Aldershot, England, Ashgate, pp. 3-19

Fischer, M.M. (2001b) Computational Neural Networks - Tools for Spatial Data Analysis. In: *GeoComputational Modelling. Techniques and Applications*, M.M. Fischer and Y. Leung (eds). Berlin, Springer-Verlag, pp. 15-34

Fuschs, V.R. (1962) Statistical Explanations of the Relative Shift of Manufacturing among Regions of the United States. *Papers of the Regional Science Association* 8, pp. 1-5

Gardner, M.W. and S.R. Dorling (1998) Artificial Neural Networks (The Multilayer Perceptron): A Review of Applications in the Atmospheric Sciences. *Atmospheric Environment* 32 (14-15), pp. 2627-36

Griffith, D.A. (2003) *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin, New York, Springer

Haynes, K.E. and Z.B. Machunda (1987) Considerations in Extending Shift-Share Analysis: Note. *Growth and Change* 18 (Spring), pp. 69-78

Himanen, V., P. Nijkamp and A. Reggiani (eds) (1998) *Neural Networks in Transport Applications*, Ashgate, Aldershot

Kuan, C.-M. and H. White (1994) ANNs: an Econometric Perspective. *Econometric Reviews* 13, pp. 1-91

Lin, C.-F.J. (1992) The Econometrics of Structural Change, Neural Network and Panel Data Analysis, PhD thesis, University of California, San Diego

Longhi, S., P. Nijkamp, A. Reggiani and U. Blien (2002) Forecasting Regional Labour Markets in Germany; an Evaluation of the Performance of Neural Network Analysis. Paper presented at the 42nd Congress of the European Regional Science Association, Dortmund, Germany

Longhi, S. (2005) Open Regional Labour Markets and Socio-Economic Developments, PhD thesis, Vrije Universiteit, Amsterdam

Longhi, S., P. Nijkamp, A. Reggiani and U. Blien (2005a) Developments in Regional Labour Markets in Germany: A Comparative Analysis of the Forecasting Performance of Competing Statistical Models. *Australasian Journal of Regional Studies* 11 (2), pp. 175-96

Longhi, S., P. Nijkamp, A. Reggiani and E. Maierhofer (2005b) Neural Network Modeling as a Tool for Forecasting Regional Employment Patterns. *International Regional Science Review* 28 (3), pp. 330-46

Loveridge, S. and A.C. Selting (1998) A Review and Comparison of Shift-Share Identities. *International Regional Science Review* 21 (1), pp. 37-58

Maier, H.R. and G.C. Dandy (2000) Neural Networks for the Prediction and Forecasting of Water Resources Variables: A Review of Modelling Issues and Applications. *Environmental Modelling & Software* 15, pp. 101-24

McCollum, P. (1998) An Introduction to Back-Propagation Neural Networks, Encoder. URL: http://www.seattlerobotics.org/encoder/nov98/neural.html

Miller, A.S., B.H. Blott and T.K. Hames (1992) Review of Neural Network Applications in Medical Imaging and Signal Processing. *Med Biol Eng Comput* 30 (5), pp. 449-64

Möller, J. and A. Tassinopoulos (2000) Zunehmende Spezialisierung oder Strukturkonvergenz? Eine Analyse der Sektoralen Beschäftigungsentwicklung auf Regionaler Ebene. *Jahrbuch für Regionalwissenschaft* 20 (1), pp. 1-38

Nazara, S. and G.J.D. Hewings (2004) Spatial Structure and Taxonomy of Decomposition in Shift-Share Analysis. *Growth and Change* 35 (4), pp. 476-90

Nijkamp, P., A. Reggiani and W.F. Tsang (2004) Comparative Modelling of Interregional Transport Flows: Applications to Multimodal European Freight Transport. *European Journal of Operational Research* 155 (3), pp. 584-602

Patterson, M.G. (1991) A Note on the Formulation of the Full-Analogue Regression Model of the Shift-Share Method. *Journal of Regional Science* 31 (2), pp. 211-16

Patuelli, R., S. Longhi, A. Reggiani and P. Nijkamp (2003) A Comparative Assessment of Neural Network Performance by Means of Multicriteria Analysis: An Application to German Regional Labour Markets. *Studies in Regional Science* 33 (3), pp. 205-29

Patuelli, R., S. Longhi, A. Reggiani, P. Nijkamp and U. Blien (2004) New Experiments with Learning Models for Regional Labour Market Forecasting. Paper presented at the 51st Annual North American Meetings of the Regional Science Association International, November, Seattle, WA

Ray, D.M. (1990) Standardizing Employment Growth Rates of Foreign Multinationals and Domes Firms in Canada from Shift-Share to Multifactor Partitioning. International Labour Organisation, International Labour Office, Geneva, Working paper

Ripley, B.D. (1993) Statistical Aspects of Neural Networks. In: *Networks and Chaos: Statistical and Probabilistic Aspects*, O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (eds). London, Chapman & Hall, pp. 40-123

Rosenblatt, F. (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review* 65, pp. 386-408

Rumelhart, D.E. and J.L. McClelland (1986) *Parallel Distribute Processing: Explorations in the Microstructure of Cognition*. Cambridge, Massachusetts, MIT Press

Sargent, D.J. (2001) Comparison of Artificial Neural Networks with Other Statistical Approaches. *Cancer* 91 (S8), pp. 1636-42

Schintler, L.A. and O. Olurotimi (1998) Neural Networks as Adaptive Logit Models. In: *Neural Networks in Transport Applications*, V. Himanen, P. Nijkamp and A. Reggiani (eds), Ashgate

Shiva Nagendra, S.M. and M. Khare (2002) *Artificial Neural Network Based Line Source Emission Modelling: A Review*. Paper presented at ACE 2002: International Conference on Advances in Civil Engineering, January, Kharagpur; India

Stock, J.H. and M.W. Watson (1998) A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. NBER Working Paper 6607

Stokes, H.K. (1974) Shift-Share Once Again. *Regional and Urban Economics* 4 (1), pp. 57-60

Swanson, N.R. and H. White (1997a) A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *The Review of Economic and Statistics* 79, pp. 540-50

Swanson, N.R. and H. White (1997b) Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models. *International Journal of Forecasting* 13, pp. 439-61

Werbos, P. (1974) Beyond Regression: New Tools for Predicting and Analysis in the Behavioral Sciences, PhD thesis, reprinted by Wiley & Sons, 1995, Harvard University

Wilson, P. (2000) The Export Competitiveness of Dynamic Asian Economies 1983-1995. *Journal of Economic Studies* 27 (6), pp. 541-65

Wong, B.K., T.A. Bodnovich and Y. Selvi (1997) Neural Network Applications in Business: A Review and Analysis of the Literature (1988-1995). *Decision Support Systems* 19 (4), pp. 301-20

Wong, B.K. and Y. Selvi (1998) Neural Network Applications in Finance: A Review and Analysis of Literature (1990-1996). *Information & Management* 34 (3), pp. 129-39

Annex A – Details of Shift-Share Regression Parameter Estimates

Tables A.1 and A.2 present the regression coefficients found when regressing the districts' overall growth rates on the competitive effect variable seen in Equation (3), for West and East Germany, respectively. A competitive effect variable was used for each of the nine industry sectors. WLS regressions were carried out for each year (that is, for each 2-year period).

Table A.1 – Shift-share regression parameters for the competitive effect variables: the case of West Germany

Sector	*87–89*	*88–90*	*89–91*	*90–92*	*91–93*	*92–94*	*93–95*	*94–96*	*95–97*	*96–98*	*97–99*	*98–00*	*99–01*	*00–02*	*01–03*
Primary sector	0.060***	0.109***	0.087***	0.051***	0.042***	0.061***	0.022***	0.012***	0.012*	0.015***	0.028***	0.028***	0.018***	0.021***	0.035***
Industry goods	0.246***	0.195***	0.195***	0.269***	0.295***	0.244***	0.231***	0.211***	0.221***	0.197***	0.242***	0.256***	0.265***	0.195***	0.183***
Consumer goods	0.038***	0.049***	0.053***	0.074***	0.085***	0.072***	0.058***	0.053***	0.032***	0.054***	0.053***	0.057***	0.038***	0.036***	0.044***
Food manufacturing	0.030**	0.019	0.061***	0.033***	0.031***	–0.021	0.025***	0.024***	0.015**	0.018***	0.000	0.020*	0.017	0.001	0.015*
Construction	0.044**	0.073***	0.039**	0.043**	0.038*	0.099***	0.096***	0.067***	0.046***	0.004	0.002	–0.022	–0.001	0.058***	0.062***
Distributive services	0.156***	0.146***	0.109***	0.090***	0.135***	0.140***	0.107***	0.137***	0.115***	0.152***	0.167***	0.093***	0.197***	0.186***	0.158***
Financial services	0.060***	0.075***	0.056***	0.066***	0.052***	0.033*	0.068***	0.099***	0.100***	0.105***	0.097***	0.075***	0.117***	0.112***	0.118***
Household services	0.029***	0.058***	0.116***	0.057***	0.052***	0.042**	0.045***	0.043***	0.060***	0.084***	0.074***	0.090***	0.058***	0.077***	0.053***
Services for society	0.161***	0.106***	0.139***	0.188***	0.080***	0.110***	0.127***	0.092***	0.164***	0.181***	0.093***	0.097***	0.155***	0.209***	0.201***

Notes:
*** Significant at the 99 per cent level.
** Significant at the 95 per cent level.
* Significant at the 90 per cent level.

Table A.2 – Shift-share regression parameters for the competitive effect variables: the case of East Germany

Sector	*93–95*	*94–96*	*95–97*	*96–98*	*97–99*	*98–00*	*99–01*	*00–02*	*01–03*
Primary sector	0.077***	0.097***	0.073***	0.056***	0.056***	0.054***	0.011	0.035***	0.040***
Industry goods	0.150***	0.103***	0.096***	0.135***	0.135***	0.114***	0.157***	0.139***	0.104***
Consumer goods	0.008	0.002	0.011	–0.011	–0.011	0.035**	0.040***	0.035**	0.026**
Food manufacturing	0.035**	0.017	0.009	0.009	0.009	0.013	0.035***	0.015	–0.001
Construction	0.151***	0.144***	0.187***	0.210***	0.210***	0.158***	0.172***	0.076***	0.102***
Distributive services	0.181***	0.211***	0.123***	0.139***	0.139***	0.115***	0.191***	0.195***	0.141***
Financial services	0.043***	0.089***	0.091***	0.101***	0.101***	0.126***	0.176***	0.166***	0.097***
Household services	0.004	–0.027	0.055**	–0.002	–0.002	0.140***	0.098***	0.086***	0.031
Services for society	0.208***	0.175***	0.306***	0.288***	0.288***	0.267***	0.302***	0.275***	0.252***

Notes:
*** Significant at the 99 per cent level.
** Significant at the 95 per cent level.

Annex B – Details of Model Experiments

The NN models used in the present paper were computed using the network parameters shown in the table below. In addition, the following parameters were used: learning rate: 0.9; momentum: 1; input noise: 0; training tolerance: 0.1; testing tolerance: 0.3.

Table B.1 – Parameter values of the NN models adopted; the case of West Germany

	Inputs	*IU*	*HU*	*Epochs*
Model A	Employment (GR), time (dummies)	22	10	900
Model AC	Employment (GR), time (dummies), district (fixed effects)	23	5	600
Model AD	Employment (GR), time (dummies), district (qualitative)	23	10	600
Model ADW	Employment (GR), time (dummies), district (fixed effects), wage (GR)	24	15	900
Model AE	Employment (GR), time (dummies), district (dummies)	31	10	200
Model AW	Employment (GR), time (dummies), wage (GR)	23	5	750
Model B	Employment (GR), time (qualitative)	10	5(1^{st}L), 5(2^{nd}L)	650
Model BD	Employment (GR), time (qualitative), district (fixed effects)	11	10	300
Model BW	Employment (GR), time (qualitative), wage (GR)	11	5(1^{st}L), 5(2^{nd}L)	1600
Model BSS	Employment (GR), time (qualitative), SSA regional component	19	15	100
Model BSSN	Employment (GR), time (qualitative), SSA spatial regional component	19	5	400
Model BSSR	Employment (GR), time (qualitative), SSA modified competitive effect	19	5	900

Notes:
IU = Input Units; HU = Hidden Units; GR = Growth Rates; 1^{st}L = First Hidden Layer; 2^{nd}L = Second Hidden Layer
All models have only 1 Output Unit; the Activation Function is always a Sigmoid.

Table B.2 – Parameter values of the NN models adopted; the case of East Germany

	Inputs	*IU*	*HU*	*Epochs*
Model A	Employment (GR), time (dummies)	16	10	100
Model AC	Employment (GR), time (dummies), district (fixed effects)	17	10	300
Model AD	Employment (GR), time (dummies), district (qualitative)	17	5	300
Model ADW	Employment (GR), time (dummies), district (fixed effects), wage (GR)	18	5(1stL), 5(2ndL)	200
Model AE	Employment (GR), time (dummies), district (dummies)	25	15	300
Model AW	Employment (GR), time (dummies), wage (GR)	17	5(1stL), 5(2ndL)	200
Model B	Employment (GR), time (qualitative)	10	5(1stL), 5(2ndL)	900
Model BD	Employment (GR), time (qualitative), district (fixed effects)	11	15	1100
Model BW	Employment (GR), time (qualitative), wage (GR)	11	5	1000
Model BSS	Employment (GR), time (qualitative), SSA regional component	19	5(1stL), 5(2ndL)	200
Model BSSN	Employment (GR), time (qualitative), SSA spatial regional component	19	5(1stL), 5(2ndL)	300
Model BSSR	Employment (GR), time (qualitative), SSA modified competitive effect	19	5(1stL), 5(2ndL)	300

Notes:
IU = Input Units; HU = Hidden Units; GR = Growth Rates; 1stL = First Hidden Layer; 2ndL = Second Hidden Layer
All models have only 1 Output Unit; the Activation Function is always a Sigmoid.

Annex C – Map of Observed Growth Rates (Years 2001–03) in Germany

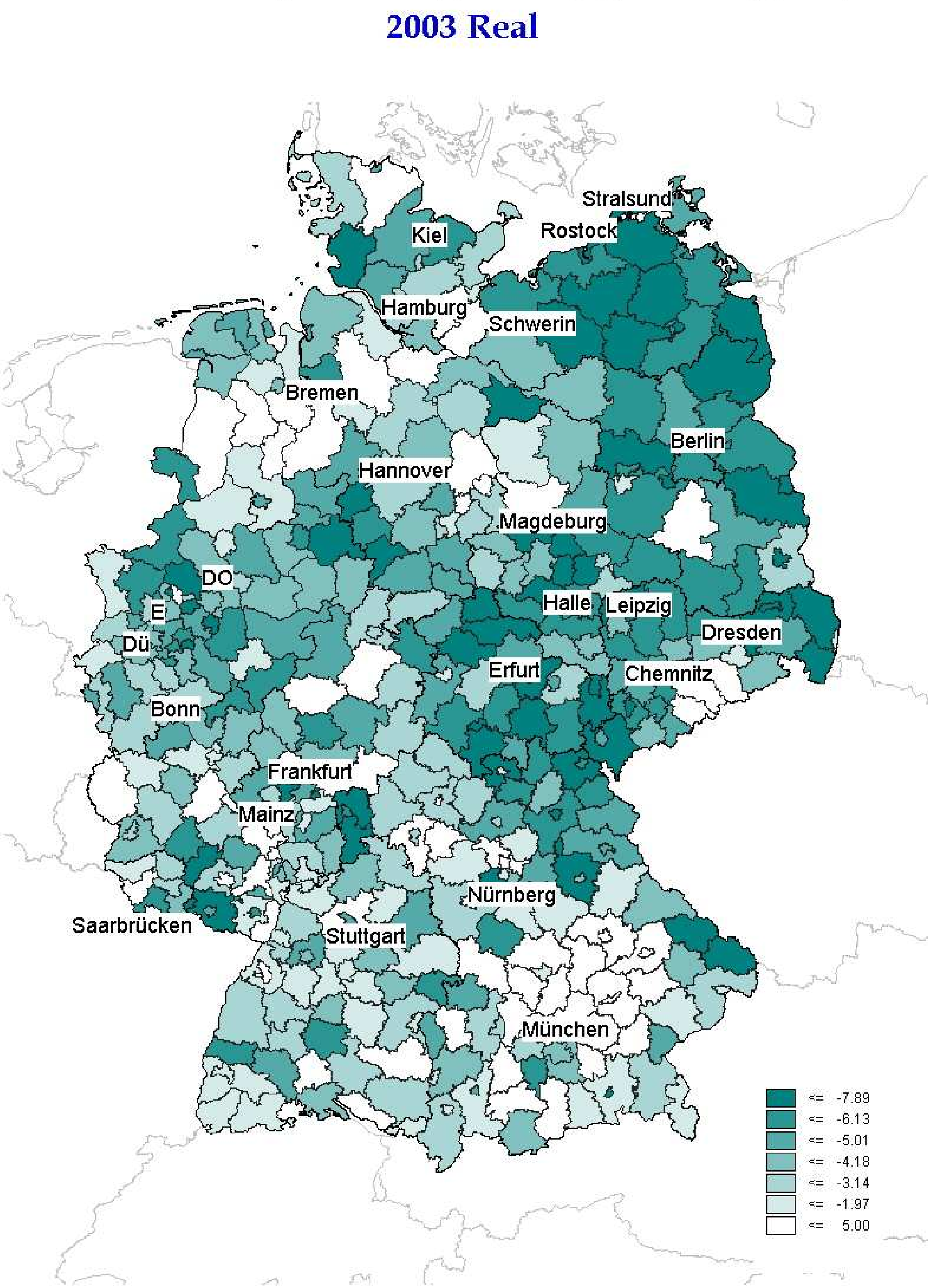

Figure C.1 – Observed full-time employment growth rates in Germany, years 2001–03

Notes

[1] Creative Commons, http://creativecommons.org/licenses/by/1.0/.

[2] Sigmoid functions are most commonly used as activation functions. For example, Adya and Collopy (1998) find that, for all the studies they collected on the business application of NNs, the activation function, when specified, was always a sigmoid. The sigmoid function is often used because it introduces non-linearity, by reducing the activation level of computing units to the [0, 1] interval. Another advantage of the sigmoid function is its simple derivative function.

[3] The starting set of weights is usually randomly defined, so that a large error is generated at first (Cooper 1999). On the other hand, Ripley (1993, p. 50) points out that the initial values 'should be chosen close to the optimal values, so as to seek the correct values are used'. Since, in our case, the optimal value of the weights is unknown, a set of random weights is used.

[4] The error term is often computed as the mean of the single units' squared errors. In our experiments, the error is computed as $E_j = Y_j(1 - Y_j)(D_j - Y_j)$, where the error term E_j is a function of the actual output Y_j, and of the difference between the expected and the actual output of the model, *Dj*.

[5] A shortcoming of the BPA is that the algorithm is only expected to reach a stationary error, which can indeed be a non-global minimum (Ripley 1993). Fahlmann (1992, as reported by Ripley) stresses that, although NNs do fall into local minima, these are often the ones the analyst wants to reach. He also points out how, in some cases, local minima are blamed for problems which are in fact the result of other causes.

[6] For a review of SSA identities, see Loveridge and Selting (1998), and Dinc et al. (1998).

[7] Data on the commuting flows were kindly provided by Gunther Haag (STASA, Stuttgart, Germany), and refer to the year 2002. Future research would ideally also look at changes in the commuting patterns, so to also have a 'dynamic' definition of 'neighbours'.

[8] The weights are computed, in our case, as the ratio between the regional and the national overall employment levels, in a base year.

[9] The 9 economic sectors are the following: 1) primary sector; 2) industry goods; 3) consumer goods; 4) food manufacturing; 5) construction; 6) distributive services; 7) financial services; 8) household services; 9) services for society.

[10] Our models employ the employment variation between years (t-2; t) in order to forecast the variation for the period (*t*, *t*+2). Consequently, if the data start from 1987, the first forecasted interval is 1989-1991. We refer to this forecast as a forecast for 1991.

[11] Future research should address various behaviours for the intermediate structures (for example, 4 or 7 neurons). However, in the future, we will focus on 2- and 3-layer NN configurations, as empirical evidence has proved that an NN with one hidden layer can approximate nearly every type of function (Cheng and Titterington 1994; Kuan and White 1994).

[12] The models are compared using the following statistical indicators:

- Mean Absolute Error: MAE = $1/N * [\Sigma_i |y_i - y_i^f|]$;
- Mean Square Error: MSE = $1/N * [\Sigma_i (y_i - y_i^f)^2]$;
- Mean Absolute Percentage Error: MAPE = $1/N * [\Sigma_i |y_i - y_i^f| *100/ y_i]$;
- Theil's U: MSE (Model) / MSE (random walk),

where y_i is the observed value (target); y_i^f is the forecast of the model adopted (NN); and N is the number of observations/examples. The common interpretation of these indicators is that the estimation is better, the closer the value is to zero. The MAPE indicator was not used in the testing phase of the NN models, but only for *ex post* forecasts evaluation.

[13] For the final step – and ultimate aim of the experiments – of making forecasts at district level for the year 2005, all of the available data will be employed, training the NNs until the year 2003. The results for this part of the experiment are not reported here, since at present no real data for 2005 are available for comparison.