

A Rank-order Analysis of Learning Models for Regional Labor Market Forecasting

ROBERTO PATUELLI¹, SIMONETTA LONGHI², AURA REGGIANI³, PETER NIJKAMP¹ and UWE BLIEN⁴

¹ Department of Spatial Economics, Free University of Amsterdam, The Netherlands;
email: rpatuell@gmu.edu; pnijkamp@feweb.vu.nl

² ISER, University of Essex, UK; email: slonghi@essex.ac.uk

³ Department of Economics, University of Bologna, Italy; email: reggiani@economia.unibo.it

⁴ Institut fuer Arbeitsmarkt und Berufsforschung (IAB), Nuremberg, Germany;
email: uwe.blien@iab.de

ABSTRACT

Using a panel of 439 German regions we evaluate and compare the performance of various Neural Network (NN) models as forecasting tools for regional employment growth. Because of relevant differences in data availability between the former East and West Germany, NN models are computed separately for the two parts of the country.

The comparisons of the models and their ex-post forecasts have been carried out by means of a non-parametric test: viz. the Friedman statistic. The Friedman statistic tests the consistency of model results obtained in terms of their rank order. Since there is no normal distribution assumption, this methodology is an interesting substitute for a standard analysis of variance. Furthermore, the Friedman statistic is indifferent to the scale on which the data are measured.

The evaluation of the ex-post forecasts suggests that NN models are generally able to correctly identify the fastest-growing and the slowest-growing regions, and hence predict rather well the correct ranking of regions in terms of their employment growth. The comparison among NN models – on the basis of several criteria – suggests that the choice of the variables used in the model may influence the model's performance and the reliability of its forecasts.

Keywords: forecasts, regional employment, learning algorithms, rank order test

JEL Classification: C23, E27, R12

1. INTRODUCTION

It is critical in modern economies to be able to predict regional economic growth as well as trends that might affect regional socio-economic development. To better guide both national and local policies, forecasts of socio-economic variables at a disaggregated (e.g., sectoral or regional) level are becoming more and more relevant. Employment is without any doubt a key variable in this regard: employment directly influences private expenditures, savings and many other parts of the economy. It is therefore important to be able to produce reliable forecasts of its variation. In this paper we evaluate different models to compute forecasts of the regional employment evolution in Germany.

Blien and Tassinopoulos (2001) and Bade (2005) propose new methodologies to compute labor market forecasts for German regions. More specifically, Blien and Tassinopoulos (2001) suggest a combination of top-down and bottom-up techniques to compute short-term forecasts for West German regions. Their forecasts take into account regional autonomous trends that are then combined with expectations about the development of single industrial sectors, by means of an entropy-optimising procedure. Bade (2005) uses an extension of the ARIMA approach to forecast the long-term development of regional shares in the national employment. However, both methodologies require a number of constraints and economic, as well as econometric, assumptions.

Similarly to Blien and Tassinopoulos (2001), Longhi et al. (2005) and Patuelli et al. (2003b) proposed a set of models to compute short-term forecasts of employment at the regional level, using data on West German regions. These models are based on a learning/forecasting algorithm called Neural Network (NN), which has been applied to several fields ranging from financial forecasts to transport economics (see, e.g., Himanen et al., 1998; Reggiani et al., 2000). Over the years, the NN models proposed in Longhi et al. (2005) and Patuelli et al. (2003b) have been constantly updated with newly acquired data, and evaluated on their ability to forecast regional employment variations using statistical indicators.

The choice of a proper forecasting model is not an easy task. For example, the stochastic nature of NN models tends to produce unstable forecasts over time, data sets and statistical indicators, thus requiring a methodology able to rank the models on the basis of their performance. Patuelli et al. (2003a) apply Multicriteria Analysis (MCA) for choosing among NN models with the aim of defining a set of models which would overall perform better than the others when several aspects of the statistical performance are taken into account. They consider several dimensions of NN performance: indicators

assessing the error levels of the forecasts, their computational complexity, the stability of their in-sample forecasts, and their generalization properties (out-of-sample forecasts).

However, further aspects should be taken into account when evaluating a forecasting technique. In addition to assessing the “quality” of forecasts in terms of errors from the observed values, it is also necessary to examine their consistency by analyzing, e.g., whether the NN models show a stable rank order in their performance over time, statistical indicators, or other performance measures. In this paper we extend the analysis by Patuelli et al. (2003a) by using the Friedman (1937) statistic to assess the presence of regularities in the rank orders of the NN models’ aggregate and disaggregate performance.

The paper unfolds as follows. Section 2 briefly illustrates the NN models of Longhi et al. (2005) and Patuelli et al. (2003b), as well as their results. Section 3 introduces the Friedman statistic, while Section 4 shows the empirical application of the test. Section 5 summarizes the results and suggests future research directions.

2. NEURAL NETWORKS AS FORECASTING MODELS

NNs are calculation algorithms originally developed to mimic the operations of the human brain (see, e.g., Rumelhart and McClelland, 1986). The attractiveness of NNs is mainly based on their ability to approximate mathematical relationships between variables, even when these relationships are not known *a priori*. Given their ability to obtain good fit in data sets of any complexity, NNs are nowadays applied to a wide range of problems, in particular in those situations where the simulation of complex relationships between independent variables is necessary, as in applied finance or environmental analysis. For a review of NN applications in economics see, among others, Herbrich et al. (1999). Comprehensive descriptions of the NN technique can be found, among others, in Bossomaier (2000). NN models have often been compared to more traditional forecasting methodologies with contrasting results (see, among others, Cheng and Titterton, 1994; Swanson and White, 1997a,b; and Fischer, 1998, 2001a).

In a NN model, calculation is distributed over a high number of simple units working in parallel, the neurons. Neurons are organized in layers and are internally connected through a set of weights (w_j ; w_{jn}). In the models evaluated in this paper (called feedforward NNs) the neurons are organized in only three groups of layers: one input layer, some hidden layers, and one output layer. Neurons from each layer are connected to every neuron of the next layer. Every input neuron (the explanatory variables in a regression framework: x_n) is connected only to the neurons of the first hidden layer, while only the

neurons of the last hidden layer are connected with the neurons in the output layer (the dependent variable in a regression framework: E_{rt}). Only in case of no hidden layers input and output neurons are directly connected.

The data entering each neuron are aggregated by an activation function ($\Psi; \phi_j$) which computes the output of the unit before passing the result to the units of the next layer. In case of only one hidden layer (see Fischer, 2001a,b):

$$E_{rt} = \Psi[\sum_j w_j \phi_j(\sum_n w_{jn} x_n)] + \varepsilon_{rt} \quad (1)$$

The “learning process” of an NN consists in the computation of the network weights via a recursive algorithm, called Back-Propagation Algorithm (BPA). The BPA modifies the network weights step by step by minimizing the difference between the observed data and the output obtained using the current set of weights. To identify the best NN architecture and to generate the final forecasts, the data set is divided into three (mutually exclusive) sub-sets: the first, called the ‘training set’, is used for parameter estimation; the second, called the ‘validation set’, is used in the process of fine-tuning of the parameters. The third, called the ‘test set’, is used to assess the performance of the model.

The models developed by Longhi et al. (2005) and Patuelli et al. (2003b) employed two similar data sets: one for West Germany, one for East Germany. The data set, provided by the German Institute for Employment Research (Institut für Arbeitsmarkt und Berufsforschung – IAB), is a panel of 439 regions containing data on the number of individuals employed full-time each year on 30 June, divided in nine economic sectors,¹ as well as the average of their daily wages at regional level. The data for the 326 West German regions is available for the period 1987–2003, while the data for the 113 East German regions is available only for the period 1993–2003. Information on the year of collection of the data, as well as on the level of urbanization of each region (the ‘type of region’, see Blien and Tassinopoulos, 2001) are also available. Such data are used to compute forecasts of the growth rate of East and West German employment over a two-year period.

Inputs of the models are the growth rate of sectoral employment; the growth rate of daily wages; an indicator of the year of data collection; and an indicator of the regional degree of urbanization.

¹ The nine economic sectors are: 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.

Information identifying the year was used in two different ways: as a set of dummy variables (one dummy for each year); or as a string variable.² The degree of urbanization was used a string variable.

The five NN models compared in this paper are summarized in Table 1.

TABLE 1 ABOUT HERE

All models use the growth rate of sectoral employment as well as information about the year of collection of the data. This last variable is represented by dummy variables in all models, except Model B, where it is used as string. Model D uses the same inputs as Model A, plus the information on the urbanization level. Model DW and Model AW add the growth rate of daily wages to the inputs of Model D and Model A respectively.

All these models have been used to compute ex-post forecasts for the years 2001, 2002 and 2003. The results of such models are used in our empirical application in Section 4.

The NN models were evaluated by Longhi et al. (2005) using only statistical indicators such as the Mean Absolute Error (MAE); the Mean Square Error (MSE) and the Mean Absolute Percentage Error (MAPE):

$$MAE = 1/N * \left(\sum_i |y_i - y_i^f| \right) \quad (2)$$

$$MSE = 1/N * \left[\sum_i (y_i - y_i^f)^2 \right] \quad (3)$$

$$MAPE = 1/N * \left(\sum_i |y_i - y_i^f| * 100 / y_i \right) \quad (4)$$

where y_i is the observed value in region i (target); y_i^f is the forecast of the model adopted; and N is the number of regions for which the forecast is computed. The common interpretation of these indicators is that the better estimation, the closer the value to zero.

Patuelli et al. (2003b) add new criteria to evaluate further aspects of the NN performance. The generalization properties of the NN models are evaluated using the ‘generalization’ criterion (*Gen*), which combines the performance of the models – measured by the three abovementioned indicators – over different parts of the data sets: the part used to compute the network weights (*train*), and the part used to compute the ex-post forecasts (*test*):

² The commercial software used for our experiments, Neuralyst, allows the use of non-numeric (string) input and output variables. The software associates numeric values between 0 and 1 to the values of each variable. The interpretation and mapping of the values are automatically taken care of by the algorithm.

$$Gen = \frac{MSE_{train} - MSE_{test}}{MSE} + \frac{MAE_{train} - MAE_{test}}{MAE} + \frac{MAPE_{train} - MAPE_{test}}{MAPE} \quad (5)$$

where \overline{MSE} , \overline{MAE} and \overline{MAPE} are the averages of the indicators computed over the two parts of the data set. Gen can be positive or negative, depending on the performance of the model over the two parts of the data set. A negative value of Gen might suggest that the model is overfitting the data, thus being unable to compute reliable ex-post forecasts. The indicator increases when the difference between training and forecasting error increases. Consequently, a value close to 0 indicates that similar error levels are found in the two parts of the data set. Finally, a higher (positive) value of the indicator indicates lower errors in the ex-post forecasts.

Patuelli et al. (2003a) find that Model AW is the most consistent model, showing the best performance for two out of three statistical indicators. However, in more recent forecasts such model shows a weaker performance, being outperformed by Model B. Furthermore, none of the proposed models had been able to correctly predict the decrease in full employment in Germany in the beginning of the 2000s (on this topic, see Statistisches Bundesamt, 2002).

The MCA approach neglects two important aspects of the models' performance. The first relates to whether such models are able to correctly reproduce/forecast the ranking of regions in terms of growth rates of employment. The second relates to whether the models' rankings are consistent over time (after updating the data set used for the analysis) and statistical indicators. The Friedman statistic test proposed in the next section should enable a better evaluation of the models' performance.

3. RANK ORDER TEST: THE FRIEDMAN STATISTIC

The main objective of the test based on the Friedman (1937) statistic is the "*isolation of factors which account for variation in the variable studied*" (p. 675). The test based on the Friedman statistic is commonly implemented in data mining software and extensively used in the statistical literature. Applications can be found in biology (e.g. Edgar, 2004), medicine (e.g. Efficace et al., 2004) and economics (e.g. Frees, 1995). Profit and Tschemig (1998) use an iterative Friedman test to rank results from a survey on possible solutions to German unemployment, while Scarelli and Venzi (1997) proposed the Friedman test for the hierarchization of the alternatives in MCA.

We use the Friedman test to assess the correlation among measurements – e.g. statistical indicators – of a set of alternatives (the NN models). The Friedman test is a distribution-free measure of the variability observed for one variable over two or more factors and is computed over intra-sample rankings as:

$$S^* = \frac{12S}{Kn(n+1)} \quad (6)$$

with

$$S = \sum_{j=1}^n (S_j - S_e)^2 \quad (7)$$

where n is the number of observations of the variable and K is the number of factors according to which the variable is evaluated. In our case n is equal to the number of NN models compared, while the K factors are the statistical indicators. S is the deviance of the sums of the rankings (S_j) of each j^{th} observation (in our case the NN models' ranks are summed through the factors) from the average rank sum S_e . Therefore, if the j^{th} observation is ranked first under two hypothetical factors, the sum (S_j) of its rankings is two. For a sufficient number of cases the S^* statistic has a χ^2 distribution, with $(n - 1)$ degrees of freedom. When $K = 3$ and $n \leq 15$, or $K = 4$ and $n \leq 8$, the critical values table prepared by Friedman should be consulted.

In the computation of a Friedman test the alternatives can be organized in a two-way table as in Table 2, where each row represents a factor according to which the alternative is measured, while each column contains the rank positions of the alternatives according to the factors. The null hypothesis H_0 is that the rankings by row are uncorrelated. If H_0 is true, then the sums (by column) of the rankings S_j will differ only because of sampling error. When the value of the Friedman statistic exceeds the critical values for the chosen significance level, H_0 is rejected, thus indicating that the rankings by row are unlikely to be uncorrelated.

TABLE 2 ABOUT HERE

A non-significant value of the test might imply two different scenarios: when the differences among the rank sums S_j are too small to be statistically significant, the rankings could be either uncorrelated or, in the opposite case, inversely correlated. In case of two factors and four alternatives, for example,

rankings like (1, 2, 3, 4) and (4, 3, 2, 1) generate equal rank sums. Although the null hypothesis is not rejected, the rankings are not uncorrelated, but inversely correlated.

The Friedman test can be considered as a generalization of the Spearman's rank correlation test to the case when more than two factors are compared. Since the Friedman test is exclusively based on rank order positions, the observations can be multiplied by any arbitrary number without affecting the test results. This property is useful when working on linear or logarithmic transformation of the data.

The Friedman test can be employed as a substitute for the analysis of variance when the observations are not normally distributed, since it does not require normally distributed data. This is particularly advantageous in economics, where the normality condition is rarely met. Furthermore, in case of large data sets, the Friedman test has been found to be four times faster to compute than the analysis of variance, with an information loss ranging from 9 to 36 per cent, which decreases with the number of factors considered (Friedman, 1937). In light of this consideration it would be desirable to employ additional factors in our analysis in the course of future research.

4. COMPARING THE PERFORMANCE OF NEURAL NETWORK MODELS

Rank Order Tests for Regional Performance

The first aspect to be investigated regards the ability of the NN models to correctly predict the hierarchy on the changes in employment at district level. As a preliminary test to our analysis, a Friedman test carried out on observed employment variations in 2001, 2002 and 2003, suggested that there are hierarchies among the German districts that are consistent over the years. The test computed over the years 2001, 2002 and 2003 rejected the null hypothesis H_0 with a 99 per cent confidence level, suggesting that (fairly) stable patterns of growth among the districts are present. The result is not trivial, since, in the presence of regional change, the rank position of the districts in terms of (employment) growth might change greatly. Similar results were obtained also when the rankings were split in broad groups – of 30 and ten districts, for West and East Germany, respectively.³

The above analysis suggests that there is in fact a ranking of the districts to be forecasted. For the purpose of testing the NN models' ability to reproduce this hierarchy, we compare, separately for 2001, 2002 and 2003, the rankings computed on real employment variations to the rankings generated ex-post

³ The ranking of the districts belonging to each layer was structured so that every district within a certain layer would have the same position. Consequently, the 30 West German districts with the higher growth rates were ranked 1st, followed by 30 more districts ranked 31st, and so on.

by each forecasting model. A test was carried out for each NN model, for each year of forecast (2001, 2002 and 2003), separately for East and West Germany. The results are shown in Table 3.

TABLE 3 ABOUT HERE

The Friedman test computed on the observed growth rate of employment in West German regions suggests that the rankings of regions are consistent over the three ex-post forecasts. The test computed for the forecasts for the years 2001, 2002 and 2003 rejects the null hypothesis of random correlation at 99 per cent confidence level, thus suggesting that the models are able to correctly predict the pattern of growth among the regions. Contrasting regional evolutions that might significantly modify the rank position of each region might be correctly identified by the NN models evaluated.

The Friedman test for East German regions rejects the hypothesis of uncorrelated rankings for 7 out of 15 models. This failure to correctly predict the ranking of regions in terms of growth rate of employment is probably due to the high errors on the models' ex-post forecasts for East German regions, which might be a consequence of the shorter time span available for East German data. In 2001 all models fail to reproduce the rankings of regions, while their performances seem to improve in 2002 and 2003, where the hypothesis of uncorrelated rankings is rejected at a 95 per cent confidence level. The disappointing performance of the models in the year 2001 might be caused by an exogenous shock that changed regional employment evolutions somehow, therefore making forecasts more difficult.⁴ The above results, however, do not tell about the quality of the forecasts in terms of relative error; the error probabilities in Table 3 will be used as evaluation criterion in the rank order analysis of the NN models in the next section.

A visual representation of the observed and forecasted rankings for the three years is shown in Figures 1, 2 and 3, where Model B has been selected as an example. The left map plots the observed rankings, while the right map plots the rankings generated by the ex-post forecasts of Model B. While for West Germany the rank of each region seems to be predicted with a small error, the errors for East German regions seem bigger. In fact, while West Germany shows a rather heterogeneous performance over the districts, East German districts tend to show a more homogeneous (negative) performance, which is more difficult to reproduce in terms of rank order.

FIGURES 1, 2 AND 3 ABOUT HERE

⁴ This possibility is not explored here, since it goes beyond the scope of the present analysis.

Rank Order Tests on the Model Results

The second aspect of the model's performance we are interested in is the rank order of NN models. We test whether the ranking of the NN models are consistent over the three ex-post forecasts and five statistical indicators. The NN models are evaluated and compared separately for East and West Germany, and the three ex-post forecasts, using the four statistical indicators introduced in Section 2, in addition to the results of the S-Tests computed in the previous subsection and shown in Table 3. The NN model ranked first for a given year is the one that shows the lowest error or, in the case of the S-Test, the lowest error probability when the null hypothesis is rejected.

Table 4 show the rank orders of each NN model for the five criteria. The first panel of Table 4 shows the rankings of the models computed on data for West Germany, while the second panel shows the rankings of the models computed on data for East Germany. The last row of each panel – West-Sum and East-Sum – shows the sums of the models' rankings that are used to compute the Friedman test. When the sums differ between the NN models, the test is significant and suggests that the rankings analyzed are correlated. The last row of Table 4 shows the results for the test carried out on all models: those computed for West Germany and those computed for East Germany. Columns from 1 to 5 show the ranking of each model for each criterion. Column 6 shows the critical value of the Friedman tests, while column 7 shows the associated probability of uncorrelated rankings.

TABLE 4 ABOUT HERE

The Friedman test rejects the null hypothesis of uncorrelated rankings for both West and East Germany. The test computed for West Germany has a value of 10.5067 and is statistically significant at 5 per cent; the test for East Germany has a value of 23.0933 and is statistically significant at the 1 per cent. Similarly to the results for East and West, the test computed on the whole country, shown in the last row of Table 4, suggests the existence of a correlation among the rankings observed. Note that the statistical significance of the separate tests for West and East Germany does not imply statistically significant results for the combined data set, since the NN models might rank differently for West and East Germany. This might bring non-significant results for the analysis on the whole country. The significance of the test for the combined data set further supports the previous results.

Because of the way the Friedman test is computed, the results are sensitive to the number of criteria employed. In a small data set, rank order variations tend to have a bigger weight in the determination of the rank order sums, therefore making the identification of clear-cut hierarchies more difficult. Further, the Friedman test follows a χ^2 distribution only for a sufficient number of alternatives and criteria.

The implications of the above considerations were evident once we carried out further analyses that employed only a portion of the factors examined earlier. The objective of these analyses was to observe the correlation of the rank orders for smaller set of criteria. More in detail, we carried out additional analyses to verify the existence of consistent hierarchies between the NN models' performances over time and for different statistical indicators. The selected NN models were tested separately by year – carrying out a comparison of the rank orders obtained in terms of MSE, MAE and MAPE for each year – and by statistical indicator – comparing each indicator's rank orders over the 2001–2003 interval. Both analyses were carried out separately for the former West and East Germany, as well as for Germany as a whole. The analyses did not provide mixed results. These additional analyses are an example of how a limited number of factors may influence the analysis and may complicate the identification of significant results, also because of a higher loss of information (see Section 2).

In summary, our analyses showed two different aspects of the reliability of NN models. The district-level estimates proved to be consistent with the hierarchies incorporated in the (observed) data. This suggests that NN models are able to correctly identify the contrasting performances of districts, in terms of ranking.

The 'aggregate' analyses showed that the rankings of the NN models are not uncorrelated. It follows that the NN models tested here show a pattern of results, where some models perform better than others. The order of "preference" of these models is proven to be consistent for different data sets and, to a certain degree, over time and different evaluation indices.

CONCLUSIONS

This paper has statistically analyzed the consistency of five Neural Network (NN) models to forecast employment in Germany. We used the Friedman statistic as a test to verify the ability of the NN models to reproduce the observed rankings of regions in terms of employment growth. The results suggest that the models are able to effectively reproduce the rankings of German regions in terms of employment growth rates. This suggests that the NN models do not level out the variability among the

districts into a blend average performance: to a certain extent they are able to correctly predict the rankings observed among the regions' employment growth rates.

The second set of analyses of our study centered around the aggregate performance of the NN models. We analyzed whether the NN models showed a consistent performance (in terms of rank order), over time and statistical indicators. The tests on the models computed for West Germany rejects the hypothesis of uncorrelated rankings. The test on the models computed for East Germany is insignificant in the first year (the year 2001) and significant for the subsequent two years (2002 and 2003). The test on all NN models – those computed for West Germany, and those computed for East Germany – rejects the hypothesis of uncorrelated rankings. This suggests a correlation between the rank orders found for the West and East German NN models, which were developed separately. As a consequence, we might imply that, although carried out separately, the two sets of NN models do not bring to a blurred out ranking of the models. The resulting rankings of the NN models still seem to be correlated, according to the evaluation criteria used.

When a wide set of criteria is used, the models show fairly stable rankings. The tests carried out for East and West Germany separately, and on Germany as a whole rejected the null hypothesis of uncorrelated rankings. Additional analyses should use more evaluation criteria (such as spatio-temporal autocorrelation, or the computation of indicators for both East and West Germany) and different NN models. Also, a comparison of the Friedman test to other rank order estimators might provide a wider look on the implications of using distribution-based or non-parametric tests (such as Friedman's) or the possibility of employing pairwise comparisons instead of summarizing measures like the sum of the rank positions, which was employed in computing our tests.

Finally, more reflection and research is needed on how to generate more reliable and consistent NN models; and how to evaluate them. Future research should address these issues, by expanding the set of tools used in evaluating the NN models, e.g., shift-share analysis or different classification approaches, such as the one recently proposed by Berardi et al. (2004).

REFERENCES

- Bade, F.-J. (2005). Evolution of Regional Employment in Germany: Forecast 2001 to 2010. In A. Reggiani and P. Nijkamp (eds), *Spatial Dynamics, Networks and Modelling* (pp. 293-318). Cheltenham (UK): Edward Elgar (forthcoming)
- Berardi, V. L., Patuwo, B. E., Hu, M. Y. (2004). A Principled Approach for Building and Evaluating Neural Network Classification Models. *Decision Support Systems*, 38(2), 233-246
- Blien, U., Tassinopoulos, A. (2001). Forecasting Regional Employment with the ENTROP Method. *Regional Studies*, 35(2), 113-124

- Bossomaier, T. R. J. (2000). Complexity and Neural Networks. In T. R. J. Bossomaier, D. G. Green (eds), *Complex Systems* (pp. 368-406). Cambridge (UK): Cambridge University Press
- Cheng, B., Titterton, D. M. (1994). Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 9(1), 2-30
- Edgar, R. C. (2004). MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Research*, 32(5), 1792-1797
- Efficace, F., Biganzoli, L., Piccart, M., Coens, C., Van Steen, K., Cufer, T., Coleman, R. E., Calvert, H. A., Gamucci, T., Twelves, C., Fargeot, P., Bottomley, A. (2004). Baseline Health-related Quality-of-life Data as Prognostic Factors in a Phase III Multicentre Study of Women with Metastatic Breast Cancer. *European Journal of Cancer*, 40(7), 1021-1030
- Fischer, M. M. (1998). Computational Neural Networks: An Attractive Class of Mathematical Models for Transportation Research. In V. Himanen, P. Nijkamp, A. Reggiani (eds), *Neural Networks in Transport Applications* (pp. 3-20). Aldershot, England: Ashgate Publishing Ltd
- Fischer, M. M. (2001a). Central Issues in Neural Spatial Interaction Modeling: the Model Selection and the Parameter Estimation Problem. In M. Gastaldi, A. Reggiani (eds), *New Analytical Advances in Transportation and Spatial Dynamics* (pp. 3-19). Aldershot, England: Ashgate
- Fischer, M. M. (2001b). Computational Neural Networks – Tools for Spatial Data Analysis. In M. M. Fischer, Y. Leung (eds), *GeoComputational Modelling. Techniques and Applications* (pp. 15-34). Berlin, Germany: Springer-Verlag
- Frees, E. W. (1995). Assessing Cross-Sectional Correlation in Panel-Data. *Journal of Econometrics*, 69(2), 393-414
- Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32, 675-701
- Herbrich, R., Keilbach, M., Graepel, T., Bollmann-Sdorra, P., Obermayer, K. (1999). Neural Networks in Economics. Background, Applications and New Developments. In T. Brenner (ed.), *Computational Techniques for Modelling Learning in Economis* (pp. 169-193). Dordrecht: Kluwer Academic Publisher
- Himanen, V., Nijkamp, P., Reggiani, A. (eds) (1998). *Neural Networks in Transport Application*: Ashgate, Aldershot
- Longhi, S., Nijkamp, P., Reggiani, A., Maierhofer, E. (2005). Neural Network as a Tool for Forecasting Regional Employment Patterns in West Germany. *International Regional Science Review*, 28(3), 330-346
- Patuelli, R., Longhi, S., Reggiani, A., Nijkamp, P. (2003a). 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), 205-229
- Patuelli, R., Longhi, S., Reggiani, A., Nijkamp, P. (2003b). Forecasting Regional Employment in Germany by means of Neural Networks and Genetic Algorithms. Paper presented at the 50th Annual North American Meetings of the Regional Science Association International. Philadelphia, PA
- Profit, S., Tschmig, R. (1998). Germany's Labor Market Problems: What to Do and What Not to Do? *Ifo Studien*, 44(3), 307-325
- Reggiani, A., Nijkamp, P., Sabella, E. (2000). A Comparative Analysis of the Performance of Evolutionary Algorithms. In A. Reggiani (ed.), *Spatial Economic Science. New Frontiers in Theory and Methodology* (pp. 332-354). Berlin: Springer-Verlag
- Rumelhart, D. E., McClelland, J. L. (1986). *Parallel Distribute Processing: Explorations in the Microstructure of Cognition*. Cambridge, Massachusetts: MIT Press
- Scarelli, A., Venzi, L. (1997). Nonparametric Statistics in Multicriteria Analysis. *Theory and Decision*, 43, 89-105
- Statistisches Bundesamt (2002). *6,8 Mill. Menschen in Deutschland arbeiten Teilzeit*. Retrieved November 3rd, 2004, from <http://www.destatis.de/presse/deutsch/pm2002/p3770031.htm>
- Swanson, N. R., White, H. (1997a). Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models. *International Journal of Forecasting*, 13, 439-461
- Swanson, N. R., White, H. (1997b). A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks. *The Review of Economic and Statistics*, 79, 540-550

TABLES

Table 1. NN models Compared

<i>Model's Name</i>	<i>Input Variables</i>
Model A	Growth rate of sectoral employment; year: dummies
Model B	Growth rate of sectoral employment; year: string
Model D	Growth rate of sectoral employment; year: dummies; urbanization level: string
Model AW	Growth rate of sectoral employment; growth rate of daily wages; year: dummies
Model DW	Growth rate of sectoral employment; growth rate of daily wages; year: dummies; urbanization level: string

Table 2. An example of the data used for the Friedman test

	X_1	X_2	X_3
<i>Factor 1</i>	x_{11}	x_{21}	x_{31}
<i>Factor 2</i>	x_{12}	x_{22}	x_{32}
...
<i>Factor K</i>	x_{1K}	x_{2K}	x_{3K}
S_j	S_1	S_2	S_3

Table 3. Districts' rankings by year and the Friedman statistic

West Germany				East Germany			
Years	Models	S^*	Prob.	Years	Models	S^*	Prob.
2001	Model A	440.026***	$2.13 * 10^{-5}$	2001	Model A	117.7792	0.3358
	Model AW	446.5338***	$8.48 * 10^{-6}$		Model AW	108.4741	0.5767
	Model B	444.2226***	$1.18 * 10^{-5}$		Model B	116.0121	0.3785
	Model D	436.9179***	$3.26 * 10^{-5}$		Model D	116.6204	0.3636
	Model DW	434.5176***	$4.51 * 10^{-5}$		Model DW	88.2692	0.9524
2002	Model A	434.8807***	$4.29 * 10^{-5}$	2002	Model A	139.1886**	0.0417
	Model AW	437.5326***	$3.00 * 10^{-5}$		Model AW	139.8547**	0.0384
	Model B	430.5571***	$7.62 * 10^{-5}$		Model B	138.4807**	0.0455
	Model D	437.2602***	$3.11 * 10^{-5}$		Model D	127.2986	0.1531
	Model DW	443.2864***	$1.38 * 10^{-5}$		Model DW	129.6665	0.1215
2003	Model A	391.4998***	0.0067	2003	Model A	139.9823**	0.0378
	Model AW	389.2529***	0.0083		Model AW	139.5668**	0.0398
	Model B	382.9143**	0.0148		Model B	137.9870**	0.0483
	Model D	394.4091*	0.0050		Model D	137.0545	0.0540
	Model DW	405.5177***	0.0016		Model DW	143.2781**	0.0247

*** significant at 1%; ** significant at 5%

*** significant at 1%; ** significant at 5%

Table 4. NN models' rankings by East and West Germany, by year and indicator

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Model	Model	Model	Model	Model		
		A	AW	B	D	DW		
		Rankings					S*	Prob.
West Germany	MSE 2001	4	3	1	5	2		
	MSE 2002	4	2	1	3	5		
	MSE 2003	2	4	1	5	3		
	MAE 2001	4	1	3	5	2		
	MAE 2002	4	2	1	3	5		
	MAE 2003	2	4	1	5	3		
	MAPE 2001	3	1	5	2	4		
	MAPE 2002	4	2	1	3	5		
	MAPE 2003	2	4	1	5	3		
	Gen 2001	4	5	1	3	2		
	Gen 2002	4	2	1	3	5		
	Gen 2003	2	4	1	5	3		
	S-Test 2001	3	1	2	4	5		
	S-Test 2002	4	2	5	3	1		
	S-Test 2003	3	4	5	2	1		
		West-Sum	49	41	30	56	49	10.5067**
East Germany	MSE 2001	3	5	2	1	4		
	MSE 2002	1	2	4	5	3		
	MSE 2003	2	3	1	5	4		
	MAE 2001	3	2	4	5	1		
	MAE 2002	2	3	1	5	4		
	MAE 2003	2	3	1	5	4		
	MAPE 2001	3	1	4	5	2		
	MAPE 2002	2	3	1	5	4		
	MAPE 2003	2	3	1	5	4		
	Gen 2001	2	4	3	5	1		
	Gen 2002	2	3	1	5	4		
	Gen 2003	2	3	1	5	4		
	S-Test 2001	1	4	3	2	5		
	S-Test 2002	2	1	3	5	4		
	S-Test 2003	2	3	4	5	1		
		East-Sum	31	43	34	68	49	23.0933***
Germany	EW-Sum	80	84	64	124	98	27.0933***	0.00002

*** significant at 1%; ** significant at 5%

FIGURES

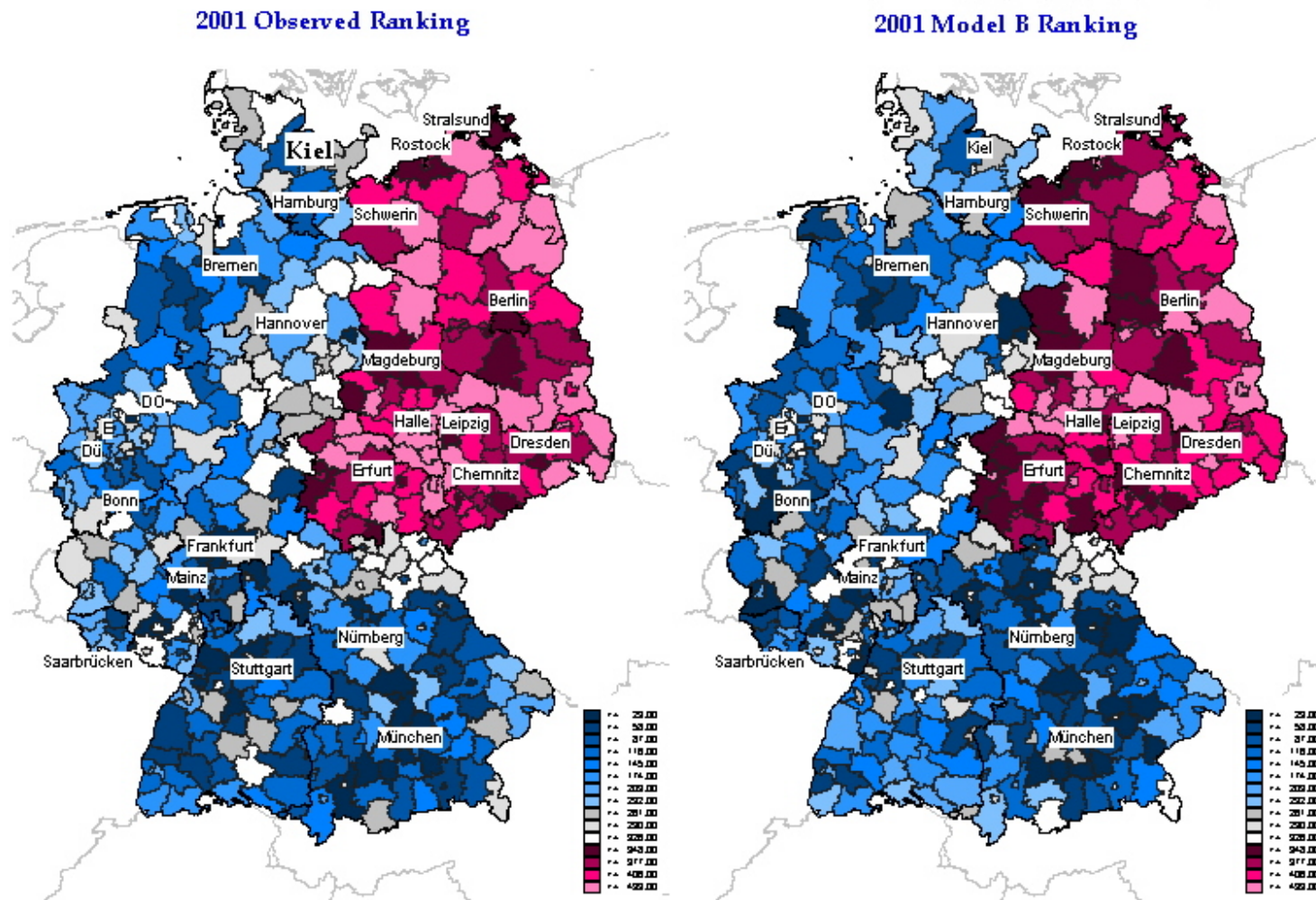


Figure 1 – Graphical representation of observed and forecasted growth rate district rankings for the year 2001 (Model B)

Note: The districts' rankings are here split in layers containing about 30 districts each. The colors for West and East Germany are blue and red, respectively; the shading of the layers goes from dark to light, from the first-ranked to last-ranked districts. The legend shows the cumulative rank positions of the districts. The layers for East Germany are listed subsequently to the West German ones.

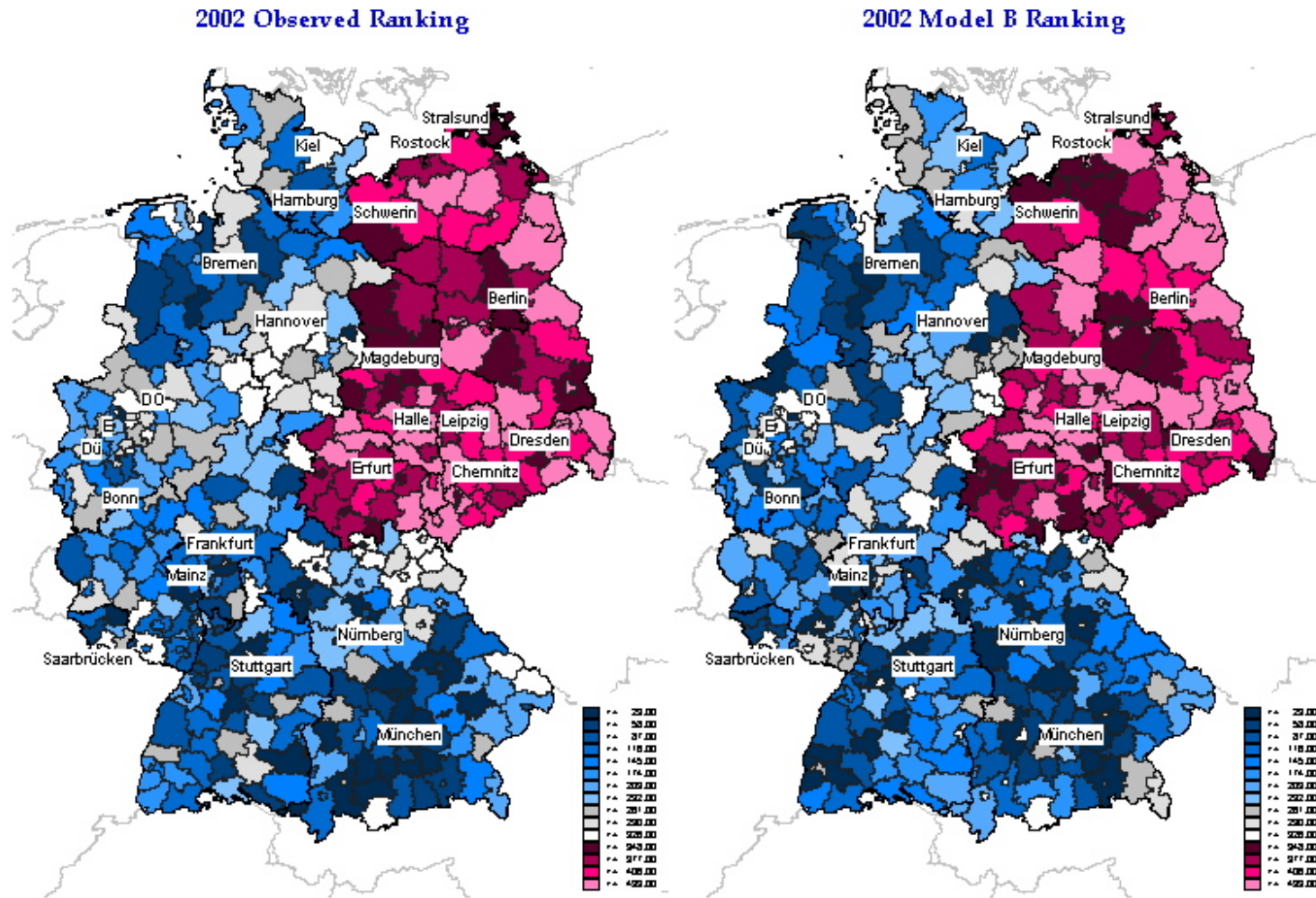


Figure 2 – Graphical representation of observed and forecasted growth rate district rankings for the year 2002 (Model B)

Note: The districts' rankings are here split in layers containing about 30 districts each. The colors for West and East Germany are blue and red, respectively; the shading of the layers goes from dark to light, from the first-ranked to last-ranked districts. The legend shows the cumulative rank positions of the districts. The layers for East Germany are listed subsequently to the West German ones.

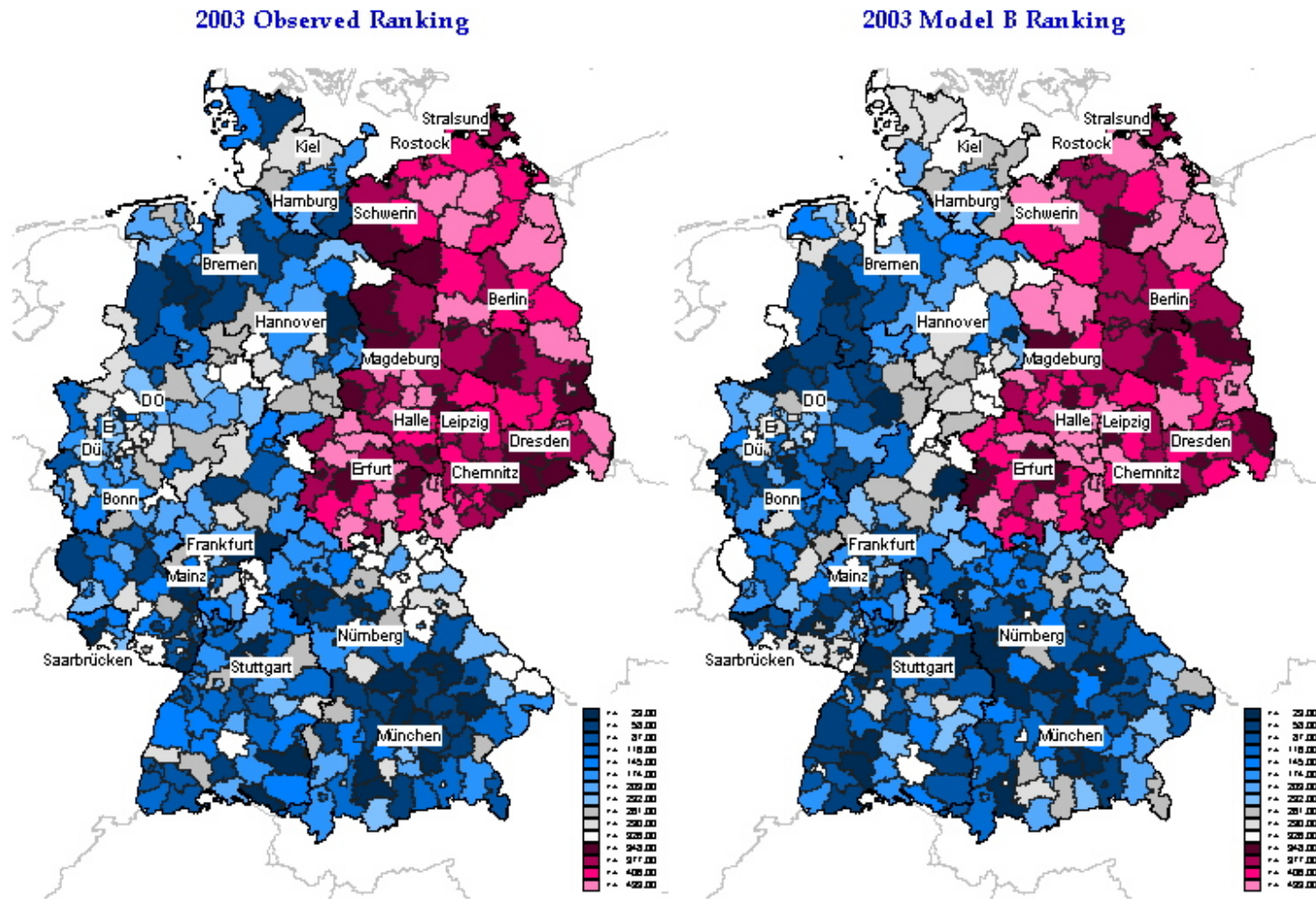


Figure 3 – Graphical representation of observed and forecasted growth rate district rankings for the year 2003 (Model B)

Note: The districts' rankings are here split in layers containing about 30 districts each. The colors for West and East Germany are blue and red, respectively; the shading of the layers goes from dark to light, from the first-ranked to last-ranked districts. The legend shows the cumulative rank positions of the districts. The layers for East Germany are listed subsequently to the West German ones.