FORECASTING REGIONAL EMPLOYMENT IN GERMANY BY MEANS OF NEURAL NETWORKS AND GENETIC ALGORITHMS

Roberto Patuelli[*], Simonetta Longhi[**], Aura Reggiani[***] and Peter Nijkamp[**]

[*] School of Public Policy, George Mason University, Fairfax, Virginia; email: rpatuel1@gmu.edu
[**] Department of Spatial Economics, Free University of Amsterdam, The Netherlands; email: slonghi@feweb.vu.nl; pnijkamp@feweb.vu.nl
[***] Department of Economics, University of Bologna, Italy; email: reggiani@economia.unibo.it;

ABSTRACT

The aim of this paper is to develop and apply Neural Network (NN) models in order to forecast regional employment patterns in Germany. NNs are statistical tools based on learning algorithms with a distribution over a large amount of quantitative data. NNs are increasingly deployed in the social sciences as a useful technique for interpolating data when a clear specification of the functional relationship between dependent and independent variables is not available.

In addition to traditional NN models, a further set of NN models will be developed in this paper, incorporating Genetic Algorithm (GA) techniques in order to detect the networks' structure. GAs are computer-aided optimization tools that imitate natural biological evolution in order to find the solution that best fits the given case.

Our experiments employ a data set consisting of a panel of 439 districts distributed over the former West and East Germany,. The West and East data sets have different time horizons, as employment information by district is available from 1987 and 1993 for West and East Germany, respectively. Separate West and East models are tested, before carrying out a unified experiment on the full data set for Germany. The above models are then evaluated by means of several statistical indicators, in order to test their ability to provide out-of-sample forecasts. A comparison between traditional and GA-enhanced models is ultimately proposed.

The results show that the West and East NN models perform with different degrees of precision, because of the different data sets' time horizons.

1. Introduction

Variables such as employment and unemployment are always used as indicators of the performance of labor markets, both at local and national level. However, such data is usually available with a certain lag and is therefore only able to give information about the past development of labor markets. Therefore, in order to take decisions on the allocation of public expenditures among regions, policymakers are always in need of reliable regional labor market forecasts.

Among the regional labor market data needed by policymakers, employment is certainly one of the most important.

Several different methods to compute such forecasts have been proposed in the literature. One of the main issues concerns the choice between models imposing linear behavior and models allowing for nonlinear behavior of the relevant variables over time.

Some authors (see, e.g., Swanson and White, 1997b, 1997a; Stock and Watson, 1998) compare linear (mainly regression analysis) and nonlinear (neural networks, genetic algorithms, fuzzy logic) methods to make forecasts of variables such as employment, industrial production or corporate profits. They come to various conclusions. Stock and Watson (1998) conclude that, in the main, the nonlinear methods adopted in their analysis perform worse than linear methods. On the other hand, Swanson and White (1997a, 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*".

In this paper we aim to compute regional forecasts of employment in Germany using artificial neural network (NN) models.[1] We follow a kind of auto-regressive approach, in which future developments of employment are the result of past developments. However, our data are somewhat different than the time-series approaches proposed by the above-mentioned literature, since these data consist of a panel characterized by a high number of cross sections and a small number of time periods. We therefore try to exploit the panel nature of the data by estimating all regions in the same NN model.

Because of the asynchronic nature of business cycles among regions, conventional models may become very complicated, and may therefore impose many constraints that could limit the scope of the analysis. Artificial NNs are more suitable for our purpose of computing regional employment forecasts because of their flexibility and the absence of strong underlying modeling hypotheses.

The paper is organized as follows. Section 2 will illustrate the methods adopted in our empirical analysis, in particular: a) the NN technique; and b) the NNs embedding genetic algorithms (GA) procedures. Next, Section 3 will describe the empirical application, which aims to estimate – by means of the NN and the NNGA approaches – employment variations in West and East Germany. The paper ends with some concluding remarks and suggestions for future research.

2. Neural Network Models for the Estimation of Employment Variations

2.1. Neural Networks Models: a Brief Introduction

Neural Networks (NNs) are calculation algorithms, which resemble the functioning of the human brain. The main characteristic of NNs is their ability to find optimal solutions when the relationships between the variables are not fully known. This is particularly useful if only a limited knowledge of the phenomenon examined is available. Despite their no-modeling hypothesis, NNs are often compared with conventional statistical tools as generalized linear models or regressions (see, among others, Cheng and Titterington, 1994; Swanson and White, 1997b, 1997a; Baker and Richards, 1999). NNs

[1] For a brief historical presentation of NNs, see Ripley (1993).

have also been shown to be equivalent, in the case of binary choice, to a logit model (Schintler and Olurotimi, 1998).

In the human brain, calculation is distributed over a high number of simple units working in parallel and strictly related to each other. In artificial NNs, these units (or neurons) are distributed in layers and are internally connected through "weights". Layers can be comprised of units, referring to the input or output variables or to the NN hidden units. In feedforward NNs, every unit from each layer is connected – and transfers information – to every unit of the next layer. Since connections between units are in only one direction and there are no cycles, the input units are only connected to the first hidden layer's neurons, while the output units are only connected to the neurons belonging to the last hidden layer. 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 deployed in the NN, input and output units are directly linked. Fischer (2001b, p. 23) defines the generic processing unit u_i, belonging to $\mathbf{u} = \{u_1, ..., u_k\}$, as:

$$u_i = \varphi_i(\mathbf{u}) = \Im_i(f_i(\mathbf{u})), \qquad (1)$$

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 the unit's output and is usually constant over the same NN.[2] The integrator function is used for aggregating the units entering the processing units, thus providing a single input, by combining the inputs through the use of the weights vector $\mathbf{w_i}$. The function commonly used for this task is a weighed sum:

$$f_i(\mathbf{u}) = \sum_j w_{ij} u_j \ , \qquad (2)$$

where u_j is the jth unit connected to unit u_i, and w_{ij} is the connection weight associated to the two units (Fischer, 2001a).

The "learning process" of an NN is guaranteed by the recursive modification of the above weights, through which the NN can identify significant rules in data occurrence (see, for example, Rumelhart and McClelland, 1986). In order to find the optimal configuration of the network weights, a learning algorithm is required, which involves several computations. The Back-Propagation Algorithm (BPA) is frequently used for this scope of computation. The BPA requires the analyst to provide input examples and their correct – and known – outputs. These data allow the network to map out their underlying behaviour and 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. The obtained error[4] is then propagated backwards through the network, adjusting the NN weights. This process is repeated for each sample record and the complete cycle is carried out as many times as requested by the operator or until the error reaches a pre-defined low value. It should be noted that the algorithm "*will never exactly learn the ideal function, but rather it will asymptotically approach the ideal function*" (McCollum,

[2] Sigmoid or logistic functions are commonly used for activation. In our case study, a sigmoid activation function was used.

[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 we do not know where the optimal value is, the initial random definition is our best guess.

[4] The error term is often computed as the mean of the single units' squared errors. In our case study, the error is given by the following: $E_j = Y_j(1 - Y_j)(D_j - Y_j)$,

where the error term E_j is a function of the actual output Y_j and the difference between desired and actual output Dj.

1998). 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).[5] Also, because of the difficulties in reaching the desired error level, we chose to stop the learning process when the NN performance started to deteriorate, in order to avoid overtraining.

A further aspect of the NN models deals with the network complexity.[6] It is necessary to seek a balance between network simplicity and complexity. In fact, an overly simple NN is not able to learn complex relationships between the variables, and therefore smoothes out the underlying data structure generating a large bias (Fischer, 2001a). Alternatively, an NN that has a too complex structure would lead to generalization problems, created by data overfitting, by causing unreliable forecasts and high variance. In order to avoid overfitting, many techniques have been proposed, which are primarily based on the partial elimination of inputs and weights (pruning methods), or on an early stopping of the learning process once the performance indices start deteriorating (early stopping techniques). In our case study, the latter method was chosen for its simplicity.

Next Section 2.2 will illustrate the implementation of a computer-aided algorithm, namely, the Genetic Algorithm, within the NN structure choice process. Subsequently, Section 2.3 will illustrate the data set available, while Section 3 will describe the empirical analysis, which aims to estimate the variations in employment in West and East Germany.

2.2. The Implementation of Genetic Algorithms in Neural Networks

Genetic Algorithms (GAs) belong to the class of Evolutionary Algorithms (EAs). These are optimization tools that nowadays have a wide following in the scientific literature (see, e.g., Fischer and Leung, 1998; Reggiani et al., 2000, 2001). Their aim is to mimic natural biological evolution dynamics. In the social sciences, this is reflected in computational models, which outline the design and structure of evolutionary processes.

The implementation of an additional algorithm in NNs responds to the need for an optimization of the choice process regarding the network structure and parameters. The final objective is to obtain better generalization properties from the NN, as well as to reduce the time/work needed in the fine-tuning of the network. In this framework, GAs were used, being one of the most commonly employed classes of EAs. Referring to the "survival of the fittest" Darwinian law (Holland, 1975), GAs are stochastically-based search methods, which aim to tackle an optimization problem expressed as follows (Fischer and Leung, 1998; Nag and Mitra, 2002):

$$\max\{\mathrm{f}(\mathbf{s})|\mathbf{s}\in\Omega\}, \tag{3}$$

where f is called the *fitness function*,[7] and $\mathbf{s}$ is an *individual* (candidate solution) belonging to the *population* $\Omega=\{0,1\}^d$, made of d-dimensional binary vectors called *strings*.[8] These strings correspond, in GAs, to nature's genotype, which contains the genetic information of an individual (referred to as the *structure*). In our model, a genotype includes three types of information: input columns; network configuration; and network parameters. These are the three strings that uniquely define an NN

---

[5] In his work, Ripley (1993) reports positive opinions expressed by Fahlmann (1992) on the occurrence of local minima. Fahlmann reportedly stresses that, although NNs do fall within local minima, these are often the ones the analyst wants. He also points out how, in some cases, local minima are blamed for problems that are instead generated by a different cause.

[6] For a discussion of the model selection problem see, e.g., Fischer (2000).

[7] In our case study, the fitness function is an objective function to be minimized on the training set. Fisher and Leung (1998) show how an objective function can be recoded into a fitness function.

[8] For details of the encoding process, see Fischer and Leung (1998).

configuration. More in detail, the *network configuration* is represented by five features: the first one refers to the total number of layers, while the remaining four features specify the number of neurons in each hidden neuron (by imposing the value zero for each hidden neuron that is not employed in the model). The *network parameters* are learning rate; momentum; and input noise. For more details on network parameters, see Cooper (1999).

Figure 1 shows the steps of a standard GA (Fischer and Leung, 1998; Riechmann, 2001). The elaboration of new NN configurations starts from an initial – randomly chosen – array of individuals (population).[9] In this case, individuals are the NN structures that are candidate solutions to the given problem. These structures are first evaluated by means of the fitness function. Subsequently, an intermediate population is generated though the *selection* operator. The probability Pr of the k^{th} individual $\mathbf{s}_k$ being duplicated – i.e. being able to "reproduce" itself – is given by the value of its fitness function divided by the overall population fitness (Fischer and Leung, 1998):

$$\text{Pr}\ (\mathbf{s}_k \text{ is selected}) = \mathrm{f}(\mathbf{s}_k) \Big/ \sum_{k=1}^{P} \mathrm{f}(\mathbf{s}_k), \qquad (4)$$

In the next step, the selected individuals undergo the *recombination/crossover* operator, which crossbreeds two individuals, randomly chosen from the population (Riechmann, 2001). The procedure resembles the functioning of sexual reproduction and the "offspring" generated replaces the "parent" individuals (network configuration and network parameters) in the population, even when its fitness is worse than that of the parents. In our experiments, all the population individuals are crossbred, although only their last two features are modified, leaving the input columns list unchanged. The following step is represented by the *mutation* operator. This operator is applied with a uniform probability distribution, and operates by switching a bit in the strings to its opposite value, i.e. from 0 to 1 (Fischer and Leung, 1998). The mutation share in our experiments is set to 10%. As in the previous step, the new generated individuals substitute the old ones in the population. At this point, the fitness function of the individuals belonging to the new generation is evaluated and the complete process restarts from the selection phase until the stopping condition is met. In our experiments, we used – as the fitness criterion – the RMS Error on the training set. As a consequence, ten generations were generated before the stopping condition was met. After this process, the best-fitting NN configuration of the last generation was adopted in our NN application (see Section 3).[10]

The definition of an optimal NN configuration can, however, be a long process, which also depends on the setting of the genetic parameters and on the problem complexity. The advantages of the GA are in its automatic functioning, which relieves the analyst from the lengthy process of manual choice of the network parameters and configuration. Nevertheless, GAs do also have shortcomings. The main limitation of GAs concerns their search space. In particular, in the framework of their use in NNs, difficulties emerge when encoding/decoding "*the original search space* Σ *into some GA-space* Ω" (Fischer and Leung, 1998, p. 447), and when finding a global minimum. In order to solve this last problem, it is suggested that the BP algorithm should be used for the global search, and then use the GA for isolating the minimum (Shapiro, 2002). This particular method will not be employed in our case study.

---

[9] For our experiments, three individuals form a population, which is iterated through the algorithm for ten generations. Future research will aim at increasing both the population size and the number of generations.

[10] In other literature examples, the best-fitting configuration is selected, from all those generated, during the given number of generations (see, e.g., Fischer and Leung, 1998).

$t := 0$
Creation of First Population $\overline{m}_0$
Evaluation of $\overline{m}_0$
while Stopping Condition not Met
$t := t + 1$
Selection from $\overline{m}_{t-1}$ and Reproduction into $\overline{m}_t$
Recombination on $\overline{m}_t$
Mutation on $\overline{m}_t$
Evaluation of $\overline{m}_t$
End

Figure 1 – Structure of a standard GA.
Source: Riechmann (2001).

2.3. The Data Set Available

The data available for our experiments concern district units in West Germany and East Germany. The data on West Germany cover 15 years (1987 to 2001), while the data on East Germany are only available for 9 years (from 1993 to 2001). 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 on 30 June. The above data are also classified according to 9 economic sectors.[11] In addition to these variables, average regional daily wages earned by full-time workers are also available. Furthermore, in an effort to identify labor 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 9 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 described below.

[11] 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.

3. Empirical Analysis: Forecasting Regional Employment in West and East Germany

3.1. 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 9 economic sectors. To exploit the panel structure of our data and – more specifically – the correlation across observations of the same region over time, in our models we introduced what we indicate as the "time" variable. This variable was identified in two different ways in the models. First, a time variable that can be interpreted as a 'time fixed effect' in panel models (Longhi et al., 2002b). Alternatively, the time factor was also introduced as an array of dummy variables. On the basis of these considerations, 9 NN models in total have been adopted, which are the following. Model A employs time by means of dummy variables, while Model B employs a fixed effects time variable. In addition to the introduction of the time variable, further variables were employed in the NN models, in order to enrich their level of information. As a consequence, 7 additional NN models emerged (see Tables A.1 and A.2 in Annex A). Model C has the same inputs as Model A, plus a qualitative variable able to distinguish among 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., 2002a). Model D and Model E have the same inputs as Model A, plus the variable 'type of economic region'. The main difference between the two models is that the new variable was introduced as a qualitative variable in Model D, and as a dummy in Model E. 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 D, obtaining Model DW; and c) in Model B, obtaining Model BW. The characteristics of the various models are summarized in Annex A. All the models adopted use, as input variables, the growth rate of the sectoral employment.

As a second step, all the above models were estimated employing GAs as a method for automatically choosing the structure of the NNs (see Annex A for details on the NN architectures chosen for each model). The GA-enhanced models (NNGA) are therefore identified by the GA suffix. The structure of the NN models, in terms of number of layers and weights, was chosen by means of comparing of the results obtained from different settings of the networks on ex-post forecasts made on a test set. In the case of the NNs employing GA, only the number of training epochs has been 'freely' chosen, since we always accepted the structure automatically proposed by the genetic algorithm. Since, for each year, the NNs were trained on the basis of the 2 years-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.[12] 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

[12] 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.

The first test phase (referred to as the validation phase), which is summarized by Table 1, concerned the validation of an array of network configurations (see, e.g., Fischer, 1998). For both types of NN models, traditional and GA-enhanced, 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, a 2-year test set was used in the case of West Germany (1999-2000), while a 1-year test set was chosen for East Germany (2000). The use of a 2-year test set 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 2-year basis should 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 a 1-year test period, since, because of the limited length of the period covered by the data, only few years would have been available for the learning process of the NN.[13] For every NN model, five structures were experimented in the initial stage. First, a two-layer structure was tried, followed by three models employing three layers and containing 5, 10 and 15 neurons respectively in one hidden layer. Finally, a four-layer model was attempted, using 5 neurons for each of the two hidden layers.[14] The models trained as described above were subsequently evaluated by means of several statistical indicators.[15] The best-performing settings were then chosen for further development of the NNs.

Table 2 – Data utilization for the test phase

Models	*Training*	*Testing*
West Germany	1991-2000	2001
East Germany	1997-2000	2001

In the subsequent test phase, the evaluation of the chosen structures was provided by ex-post tests carried out on the year 2001 – 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 2000. The objective of this procedure was to obtain ex-post forecasts for the year 2001 that could be compared with the actual data, in order to evaluate the models' generalization properties.[16]

The next sections will explain and discuss the empirical findings from our experiments. These finding were obtained by both traditional NNs and NNs utilizing GAs. We will try to verify whether, in our case, GAs can provide us with better precision in forecasting employment variations, through a

---

[13] For the development of NN models comprising both West and East Germany – which are not reported at this stage of the work – we used the same number of years as for East Germany, its data set length being the common denominator in terms of data availability. The full set of data could indeed have been used, by providing – in the training phase – a major set of information on the districts belonging to the former West Germany. This possibility will be tested in further developments of this work.

[14] Future experiments will address various behaviors for the intermediate structures (e.g., 4 or 7 neurons). However, in the future we will focus on two- and three-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).

[15] 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: $MAPE = 1/N * [\Sigma_i |y_i - y_i^f| *100/ y_i]$,

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.

[16] For the final step – and ultimate aim of the experiments – of making forecasts at district level for the year 2003, all of the available data were employed, training the NNs until the year 2001. The results for this part of the experiment are not reported here, since no real data for 2003 are available at the moment for comparison.

selection process of the NN structure that is not based on human selection. First, the results obtained for the former West Germany will be shown and examined (Section 3.2), followed by those found for East Germany (Section 3.3). Section 3.4 will conclude the illustration of our empirical experiments, by focusing attention on the performance differences between NN and NNGA models.

3.2. Estimation of West German Employment

As indicated in the previous section, 18 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 1997-1999 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 the percentage employment variations and, for the NN structures subsequently chosen, provided the results shown in Annex B (Table B.1). Further details on the NN models structure that were finally chosen can be found in Annex A (Table A.1). The models were then retrained until the year 2000, while the year 2001 acted as a test set (see Table 2). The statistical indicators emerging from these experiments are presented in Table 3. These results will be the basis for the choice of a reduced array of NN models to be adopted for the employment forecasts regarding the year 2003.

Table 3 – Statistical performances of the ex-post forecasts for the year 2001; the case of West Germany

	Model A	*Model B*	*Model C*	*Model D*	*Model E*	*Model AW*	*Model DW*	*Model BD*	*Model BW*	*RW Nat.*
MSE	21702742	8326739	20259245	25233824	19857019	12909349	8806876	**8057188**	**7851692**	22748959
MAE	1753.51	1587.58	1636.79	1783.79	1603.13	**1367.95**	1424.77	1568.83	1612.12	2124.82
MAPE	2.0619	2.6599	1.8809	2.0101	1.8979	**1.9433**	2.1272	2.5651	2.7247	2.6999
	Model AGA	*Model BGA*	*Model CGA*	*Model DGA*	*Model EGA*	*Model AWGA*	*Model DWGA*	*Model BDGA*	*Model BWGA*	*RW G.R.*
MSE	13961510	8780426	29658114	26081503	16547757	17224990	16213865	8670100	9100656	158622682
MAE	**1381.53**	1726.75	2174.53	1919.93	1457.64	1501.44	1496.21	1767.46	1744.25	1599.18
MAPE	**1.8004**	2.9324	2.4809	2.1700	1.8681	**1.8339**	1.8531	2.9006	2.9419	2.1124

Note: The abbreviations are explained in Footnote 15.

It is clear from Table 3 that no model wins over the others for all the statistical indicators. However, Model AW has the best MAE value and one of the lowest MAPE. Model AGA also has a good value for MAE and the lowest MAPE. Furthermore, we can see that low values of MSE are never combined with good performance on the other indicators. Also, the ratio between the worst and best values of MSE is much wider than it is for MAE and MAPE.[17] This might be due to the fact that MSE is based on squared errors. One more significant fact is that the models enhanced by means of GAs provide good results. In fact, they show some of the best values for the statistical indicators.

At aggregate level, the models seem to suggest an increase in the number of employees from 1999 to 2001. The models forecast an average employment increase of 2.52%, while the real growth rate recorded was about 2.87%. The average aggregate occupational level obtained by the models therefore approximates the real growth rate with an error as small as 0.34%. Although this figure does not refer to the district level variance, it might be considered to be a widely acceptable error margin. Models

[17] Ratios calculated between the worst and best values of each of the indicators provided the following results: MSE = 3.50; MAE = 1.44; MAPE = 1.62.

AW and AGA also seem to provide accurate aggregate forecasts, as they show errors of 0.08% and 0.87%, respectively. In addition, models employing time as fixed effects (the "B" models) tend to offer analogous results, slightly higher than the average performance, around a 1.5% error rate. A graphical representation of the aggregate forecasts is given in Annex C (Figure C.1).

In addition, we considered as a main performance indicator – given the variability of the values of the statistical indicators shown in Table 3 – the error of the average of the 9 NN models, as well as the error of the average of the further 9 NNGA models (Granger and Newbold, 1986). Their graphical representation, by district, is mapped in Figures D.1 and D.2, respectively (in Annex D). Concerning the forecasts for the year 2003, we trained all the models till the year 2001. In the first stage of our analysis, we considered, as a 'synthesis' model, the growth rates of these Average NN models. Maps of Germany showing the degree of the estimated employment variations for 2003 can be found in Annex D (Figures D.3 and D.4).

As an alternative method for predicting and for comparison purposes, we also adopted the Random Walk (RW) technique. RW models were chosen for their easy and fast implementation and because they do not require specific software. However, a shortcoming of RW models is the fact that they do not exploit the potential of other variables possibly correlated to the one to be estimated. In our work, we utilized two types of random walk (RW) models, which are defined as follows:

a) Random Walk Nat.: this model hypothesizes that the regional number of employees for year $t+n$ is equal to the number recorded in year t. Since, in our case, forecasts are made for year t+2, the forecast for 2001 is therefore represented by the recorded number for 1999 and has growth rates equal to zero.
b) Random Walk G.R.: this model assigns to period (t; $t+n$) the same district growth rates recorded for the period (t-n; t). For example, the regional/district growth rates of employment between 1999 and 2001 will be equal to those recorded between 1997 and 1999.

Both types of random walk models were calculated on yearly total amounts of employees, separately for each district. We observe that the Random Walk G.R. model has better values in MSE and MAE than Random Walk Nat., but shows higher MAPE values. At aggregate level, the Random Walk G.R. model seems to better estimate the employment variation, while Random Walk Nat. has the highest error (see Figures C.1 and C.2 in Annex C).

3.3. 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 2001. The data set is therefore smaller than that for West Germany – which comprises 326 districts from 1987 to 2001 – and 6 years shorter. Consequently, only 5 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) (see again Table 1). Table B.2 in Annex B shows the results obtained in this phase of the process for the chosen configurations, while Annex A (Table A.2) provides the details on the structure and parameters of each NN. The above-mentioned models were subsequently trained until the year 2000, employing the year 2001 as a test period (see Table 2).

Table 4, which contains the results of the ex-post forecasts for the year 2001, does not show a strong homogeneity in the models' results. Once more, as for West Germany, the models based on time as a qualitative variable (Models B, BW and BD) have most of the lowest values for the MSE indicator, but do not show satisfying values for estimators based on absolute error. Still, as was the case for West

Germany, the best MSE values are never found together with by the best MAE and MAPE values. These phenomena seem to suggest that our models do not enable the consistent performance of MSE and the remaining indicators.[18]

Table 4 - Statistical performances of the ex-post forecasts for the year 2001; the case of East Germany

	Model A	*Model B*	*Model C*	*Model D*	*Model E*	*Model AW*	*Model DW*	*Model BD*	*Model BW*	*RW Nat.*
MSE	13855323	10504722	36849400	10057159	18214516	50357944	17719380	9627065	**8076538**	9055105
MAE	1938.55	2007.30	1980.93	2268.41	1961.44	**1825.19**	**1726.13**	2026.60	1981.71	2558.70
MAPE	4.7083	5.1026	4.1479	5.9226	4.5186	**2.9360**	**3.7931**	5.1698	5.1244	7.0492
	Model AGA	*Model BGA*	*Model CGA*	*Model DGA*	*Model EGA*	*Model AWGA*	*Model DWGA*	*Model BDGA*	*Model BWGA*	*RW G.R.*
MSE	**8848823**	13189747	43609565	11878101	**8520633**	9007534	9630108	12362021	11950440	10024545
MAE	2074.04	1978.65	**1914.06**	2000.42	2073.56	2199.91	1950.69	1935.92	1996.29	2257.91
MAPE	5.4320	4.8676	**3.6817**	4.9793	5.3941	5.8310	4.9871	4.7495	4.9881	6.1832

Note: The abbreviations are explained in Footnote 15.

Aggregated forecasts show all the models indicating a decrease in occupational levels between 1999 and 2001. Once again, the models, and in particular the B-models, seem to cluster around a numeric zone, where the average aggregated performance of the models can also be found. This result estimates losses in employment of about 2.85%, while the actual data show decreases around 4.91% of the total labor forces. The percentage difference between the two aggregate figures is thus in the order of 2.16% (see Figure C.2 in Annex C for a graphical representation of the aggregate results).

The results from the above experiments can now be compared with those calculated for the two random walk models, as introduced in Section 3.2. While the MSE indicator has values that are lower than the average performance[19] of the NNs, worse values were found for MAE and MAPE. On the other hand, at aggregate level, the RW models overestimated employment for 2001.

As in the case of West Germany, with reference to the forecasts for the year 2003, we then trained all the models till the year 2001. In this first stage, we considered the growth rates for both the average of the NN models and the average of the NNGA models. Figures D.3 and D.4 in Annex D show how the estimated growth rates (for the year 2003) are distributed for each district on the map of Germany. A general remark here concerns the evidence – for both the averaged NN and NNGA models – of positive growth rates for West Germany and negative growth rates for East Germany, for the year 2003. This first result seems to confirm the forecasting trends emerging from other studies (see, e.g., Bade, 2003).

3.4. Concluding Remarks: NN Models vs. NNGA Models

We now summarize the performances of the NNs with a particular focus on the application of GAs in the NN choice process. All the models developed so far were also carried out in their GA-enhanced version. Therefore, in any phase of our experiments, we have 9 traditional NN models and 9 similar NNGA models – with different layers and/or parameter settings.

Through a quick comparison of the registered performance of NN and NNGA models, we observed that:

[18] The same questions arose early in Section 3.2, while examining the results for West Germany's ex-post forecasts.

[19] Calculating an average of the growth rates proposed by all of the NNs, we obtain the following values for the statistical indicators: MSE: 14166882; MAE: 1942.24; MAPE: 4.6862.

a) GA-enhanced models seem to perform ambiguously with reference to the NNs. NNGA models show, for our statistical indicators, an average error level on the ex-post forecasts for 2001 that is 16% bigger than for traditional NN models. These results are, however, not crystal-clear. For West and East Germany models, on average, 9% and 3% differences, respectively, were found between NN and NNGA models. We also made an attempt to identify the statistical indicators for which NNGA models' performance is worse. This would help us to understand what type of error is induced by the GA structure selection. Unfortunately, the results are not clear. NNGA models show 16% higher errors in MSE for West Germany, while, in the case of East Germany, they perform *better* than NN models for the same indicator. Clearer results were obtained for the MAE and MAPE indicators. For all the data sets, NNGA models show higher error levels, ranging from 3% to 21% additional error. These tendencies are also visible at an aggregate level (see Figures C.1 and C.2 in Annex C). On the other hand, in spite of average higher errors, NNGA models can still provide some good results for our statistical indicators. Also, NNGA models seem to be slightly better than traditional NN models for the MAPE indicator when we compare the error levels of both the average NN and the average NNGA model (see Sections 3.2 and 3.3, as well as Annex D).
b) NN models organized on a two-layer basis[20] seem to show significant differences if compared with four-layer models. While two-layered models seem to minimize the MAPE indicator, the more complex models provide good results for the squared error indicator (MSE). This aspect is particularly evident for the NN models related to West Germany, while the situation is more ambiguous for East Germany.

Trying to explain the above differences between traditional NN models and those that use GA-chosen settings is indeed the most difficult part. A first aspect to be considered is that the software commonly available for NN applications is not suited for panel data use, but is designed to work on time series data.[21] Nevertheless, we can make some hypotheses on the basis of our results.

The stochastic nature[22] of the NN structures that embed GA might play a role in determining a higher variance in the networks' performance. The NNGA models seem to develop more heterogeneous structures than the 'traditional' NNs, which were chosen through a fixed procedure (see Section 3.1). Although the settings chosen by the GA models are supposed to be the best-performing (among the ones developed during its running time), a longer running time for the GA might be desirable, in order to have a wider set of alternatives examined by the software. A setback to this procedure is that computing time significantly increases, especially for wider data sets.

4. Conclusions

The aim of this paper was to make forecasts – at the time n+2 – on the number of people employed in the 339 districts in Germany. For this purpose, several models – based on NN and GA techniques – have been 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 NNs were developed in configurations chosen either manually or by means of GA supervision.

The results of ex-post forecasts on the year 2001, obtained by the NN and NNGA models, were evaluated by means of several statistical indicators. In addition, we compared the results of NN and

[20] Models for which the chosen architecture is based on a two-layer structure are: Models BGA, BDGA and BWGA for West Germany; Model BDGA for East Germany; Model BDGA for Germany. The four-layer models are: Models B, BW, AGA, EGA, AWGA, DWGA for West Germany; Models B, AW, DW and AGA for East Germany; Models B, AW, DW and AGA for Germany.
[21] The software used for our NN models is Neuralyst, version 1.4.
[22] Every operator in the GA, as described in Section 2.2, involves probability distributions.

NNGA models with the results of RW models, either hypothesizing stable employment (null variations) or replicating the same employment trends as in the previous 2-year period.

Our results lead to the following considerations:

a) The models' performances show different degrees of homogeneity with respect to the West and East data sets. While the NN and NNGA models built for West Germany show more or less homogeneous results, this is not the case for East Germany.[23] From a preliminary observation of Tables 3 and 4, the models utilizing the variable 'time fixed effect' seem to behave differently from the remaining models. In fact, while these NN models often show – for both West and East Germany – the best values for the MSE indicator, they never generate such acceptable values for the MAE and MAPE statistical indicators. This dichotomy could be due to the different nature of the MSE and MAE/MAPE indicators.[24]
b) Through all our experiments, we searched for an NN model that could be considered as the most consistent and reliable. We identified it in Model AW. This is a model employing time as a dummy variable, and wage information as an additional variable. Model AW consistently shows, through each data set, that values for the MAE and MAPE indicators are among the best results. Only the MSE values did not follow this tendency over our experiments, and presented high error rates. In fact, we previously saw that MSE behaves in an opposite way to MAE/MAPE in most of our experiments. However, given the high number of NN models involved (we adopted 18 models in total), more reflection is necessary in this context. In addition to the choice of an NN model, we considered the use of pooled forecasts, as defined in Granger and Newbold (1986). These combined forecasts would provide the average performance of the analyzed NNs or of the chosen sub-set of them.[25] The emerging forecasting values are mapped in Annex D.
c) The Random Walk models described above show different levels of precision if compared with the NN models. While, for West Germany, RW models perform worse than NNs, they tend to have similar performances for East Germany.[26] This difference might be due to the different time span of the data sets.
d) As we outlined in Section 3.3, the enhancement of GA in NN models did not seem to improve the networks' performance in a significant way. In some cases, NNGA models do perform well (we refer, in particular, to the West Germany models), while, in several other cases, they show a higher error level (see again Section 3.3).

In conclusion, our aim was to experiment and test NN and NNGA 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 socioeconomic background. Consequently, by means of our NN models, we offered 'bands' of forecast values instead of 'unique' values. It has to be remarked that our empirical analysis has been based only on two main variables (employment and wages), thus it cannot be comprehensive with regard to the many variables that come into play when employment and social conditions are at stake.

---

[23] The variance of the MAPE indicator (which is directly comparable between the data sets since it is based on relative errors) is equal to 0.18 for West Germany, while it is 0.57 for East Germany.

[24] Minimization of absolute error could in fact not coincide with minimization of squared error. We privilege percentage (absolute) error, consequently choosing MAPE over MSE and MAE.

[25] It has to be noted that the work from Granger and Newbold (1986) originally refers to time series data.

[26] It is interesting to underline that, for East Germany, the RW models show MSE values quite similar to the models using the 'time fixed effects' variable. For Germany as a whole, the RW G.R. model shows MAE and MAPE values better than those of the NN models, which strengthens our hypothesis that the West and East data sets bring, if employed in the same model, a lack of clarity, compared with previous empirical evidence. RW models did in fact perform worse on average than our NN models in the course of the experiments on the West and East data sets.

Further directions for research are therefore concerned with addressing the need for a longer data span enriched with more variables (e.g. unemployment) and, ultimately, the possibility of 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.

Acknowledgements

The authors wish to thank the Institute for Employment Research (IAB), Nuremberg (Germany), and particularly Uwe Blien and Erich Maierhofer, for providing the data. Special thanks are given to Professor Franz-Josef Bade (University of Dortmund) for kindly providing the maps illustrated in Annex D.

References

Bade, F.-J. (2003). Die Regionale Entwicklung der Erwerbstätigkeit Bis 2010. Dortmund, University of Dortmund.

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: 405-415.

Bellmann, L. and U. Blien (2001). Wage Curve Analyses of Establishment Data from Western Germany. *Industrial and Labor Relations Review* 54: 851-863.

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

Cooper, J.C.B. (1999). Artificial Neural Networks versus Multivariate Statistics: An Application from Economics. *Journal of Applied Statistics* 26: 909-921.

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

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

Fischer, M.M. (2000). Methodological Challenges in Neural Spatial Interaction Modelling: The Issue of Model Selection. In: *Spatial Economic Science. New Frontiers in Theory and Methodology*, edited by A. Reggiani. Berlin, Springer-Verlag.

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*, edited by M. Gastaldi and A. Reggiani. Aldershot, England, Ashgate**:** 3-19.

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

Fischer, M.M. and Y. Leung (1998). A Genetic-Algorithms Based Evolutionary Computational Neural Network for Modelling Spatial Interaction Data. *The Annals of Regional Science* 32: 437-458.

Granger, C.W.J. and P. Newbold (1986). *Forecasting Economic Time Series*. Orlando, Florida, Academic Press Inc.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbour, University of Michigan Press.

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

Longhi, S., P. Nijkamp, A. Reggiani and U. Blien (2002a). 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., P. Nijkamp, A. Reggiani and E. Maierhofer (2002b). Neural Network as a Tool for Forecasting Regional Employment Patterns in West Germany.

McCollum, P. (1998). An introduction to back-propagation neural networks, Encoder. URL: http://www.seattlerobotics.org/encoder/nov98/neural.html.

Nag, A.K. and A. Mitra (2002). Forecasting Daily Foreign Exchange Rates Using Genetically Optimized Neural Networks. *Journal of Forecasting* 21(7): 501-511.

Reggiani, A., P. Nijkamp and E. Sabella (2000). A Comparative Analysis of the Performance of Evolutionary Algorithms. In: *Spatial Economic Science. New Frontiers in Theory and Methodology*, edited by A. Reggiani. Berlin, Springer-Verlag**:** 332-354.

Reggiani, A., P. Nijkamp and E. Sabella (2001). New Advances in Spatial Network Modelling: Towards Evolutionary Algorithms. *European Journal of Operational Research* 128(2): 385-401.

Riechmann, T. (2001). *Learning in Economics: Analysis and Application of Genetic Algorithms*, Contributions to Economics. Heidelberg and New York: Physica.

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

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

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

Shapiro, A.F. (2002). The Merging of Neural Networks, Fuzzy Logic, and Genetic Algorithms. *Insurance: Mathematics and Economics* 31(1): 115-131.

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.

Swanson, N.R. and H. White (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. and H. White (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.

Annex A – 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: training tolerance: 0.1; testing tolerance: 0.3. The genetic algorithm's parameters are as follows: inclusion rate: 1; population size: 3; population mode: immigrate; crossovers: 1; mutation rate: 0.1; fitness criteria: training error.

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

	Inputs	*IU*	*HU*	*Epochs*	*LR*	*M*	*IN*
Model A	Employment (GR), time (dummies)	22	10	900	0.9	1	0
Model B	Employment (GR), time (qualitative)	10	5(1stL), 5(2ndL)	650	0.9	1	0
Model C	Employment (GR), time (dummies), district (fixed effects)	23	5	600	0.9	1	0
Model D	Employment (GR), time (dummies), district (qualitative)	23	10	600	0.9	1	0
Model E	Employment (GR), time (dummies), district (dummies)	31	10	200	0.9	1	0
Model AW	Employment (GR), time (dummies), wage (GR)	23	5	750	0.9	1	0
Model DW	Employment (GR), time (dummies), district (fixed effects), wage (GR)	24	15	900	0.9	1	0
Model BD	Employment (GR), time (qualitative), district (fixed effects)	11	10	300	0.9	1	0
Model BW	Employment (GR), time (qualitative), wage (GR)	11	5(1stL), 5(2ndL)	1600	0.9	1	0
Model AGA	Employment (GR), time (dummies)	22	24(1stL), 5(2ndL)	250	0.8279	0.2252	0.0071
Model BGA	Employment (GR), time (qualitative)	10	0	400	0.9013	0.3330	0.0118
Model CGA	Employment (GR), time (dummies), district (fixed effects)	23	29	350	0.9492	0.1246	0.0101
Model DGA	Employment (GR), time (dummies), district (qualitative)	23	27	600	0.9575	0.5977	0.0175
Model EGA	Employment (GR), time (dummies), district (dummies)	31	24(1stL), 8(2ndL)	200	0.6892	0.0515	0.0198
Model AWGA	Employment (GR), time (dummies), wage (GR)	23	29(1stL), 9(2ndL)	350	0.6002	0.4409	0.0028
Model DWGA	Employment (GR), time (dummies), district (GR), wage (GR)	24	24(1stL), 10(2ndL)	300	0.8294	0.1348	0.0076
Model BDGA	Employment (GR), time (qualitative), district (GR)	11	0	500	0.7982	0.2698	0.0164
Model BWGA	Employment (GR), time (qualitative), wage (GR)	11	0	1800	0.8416	0.2774	0.0187

IU = Input Units; HU = Hidden Units; LR = Learning Rate; M = Momentum; IN = Input Noise;
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.

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

	Inputs	*IU*	*HU*	*Epochs*	*LR*	*M*	*IN*
Model A	Employment (GR), time (dummies)	16	10	100	0.9	1	0
Model B	Employment (GR), time (qualitative)	10	5(1stL), 5(2ndL)	900	0.9	1	0
Model C	Employment (GR), time (dummies), district (fixed effects)	17	10	300	0.9	1	0
Model D	Employment (GR), time (dummies), district (qualitative)	17	5	300	0.9	1	0
Model E	Employment (GR), time (dummies), district (dummies)	25	15	300	0.9	1	0
Model AW	Employment (GR), time (dummies), wage (GR)	17	5(1stL), 5(2ndL)	200	0.9	1	0
Model DW	Employment (GR), time (dummies), district (fixed effects), wage (GR)	18	5(1stL), 5(2ndL)	200	0.9	1	0
Model BD	Employment (GR), time (qualitative), district (fixed effects)	11	15	1100	0.9	1	0
Model BW	Employment (GR), time (qualitative), wage (GR)	11	5	1000	0.9	1	0
Model AGA	Employment (GR), time (dummies)	16	26(1stL), 8(2ndL)	300	0.5685	0.799	0.0022
Model BGA	Employment (GR), time (qualitative)	10	19	1700	0.6878	0.3651	0.0230
Model CGA	Employment (GR), time (dummies), district (fixed effects)	17	14	200	0.6385	0.0994	0.0019
Model DGA	Employment (GR), time (dummies), district (qualitative)	17	16	200	0.9573	0.1433	0.0129
Model EGA	Employment (GR), time (dummies), district (dummies)	25	16	100	0.9443	0.0666	0.0061
Model AWGA	Employment (GR), time (dummies), wage (GR)	17	8	200	0.5705	0.0272	0.0170
Model DWGA	Employment (GR), time (dummies), district (GR), wage (GR)	18	6	100	0.8544	0.0764	0.0034
Model BDGA	Employment (GR), time (qualitative), district (GR)	11	29	1000	0.7201	0.4295	0.0196
Model BWGA	Employment (GR), time (qualitative), wage (GR)	11	13	200	0.6973	0.4033	0.0004

IU = Input Units; HU = Hidden Units; LR = Learning Rate; M = Momentum; IN = Input Noise;
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 B – Statistical Performances from the Validation Phase that Identifies the NN Structures

Table B.1 – Identification of the NN structures; the case of West Germany

	Model A	*Model B*	*Model C*	*Model D*	*Model E*	*Model AW*	*Model DW*	*Model BD*	*Model BW*
MSE	8.7872	19.8530	10.2190	7.1454	14.8275	9.3283	8.2178	19.5390	19.2870
MAE	2.1383	3.7784	2.3971	1.9263	3.0673	2.2227	2.1865	3.7460	3.7372
	Model AGA	*Model BGA*	*Model CGA*	*Model DGA*	*Model EGA*	*Model AWGA*	*Model DWGA*	*Model BDGA*	*Model BWGA*
MSE	9.9385	18.6356	28.8005	19.2710	12.8172	13.1477	12.8878	19.8720	21.1513
MAE	2.3783	3.6342	4.7588	3.6788	2.7682	2.8180	2.7912	3.7903	3.8918

Note: For an explanation of the abbreviations, see Footnote 15.

Table B.2 - Identification of the NN structures; the case of East Germany

	Model A	*Model B*	*Model C*	*Model D*	*Model E*	*Model AW*	*Model DW*	*Model BD*	*Model BW*
MSE	18.9200	43.4300	19.6600	19.6022	19.3368	18.6839	18.9898	41.4465	52.1626
MAE	3.3300	5.4736	3.4900	3.3665	3.4456	3.3549	3.3208	5.3772	6.1449
	Model AGA	*Model BGA*	*Model CGA*	*Model DGA*	*Model EGA*	*Model AWGA*	*Model DWGA*	*Model BDGA*	*Model BWGA*
MSE	23.2200	42.2797	20.6900	19.6679	22.4591	23.5909	20.4065	41.4455	62.0669
MAE	3.6800	5.3565	3.6400	3.4030	3.6707	3.7573	3.4287	5.3938	6.8123

Note: For an explanation of the abbreviations, see Footnote 15.

Annex C – Aggregate Ex-Post Forecasts for the Year 2001

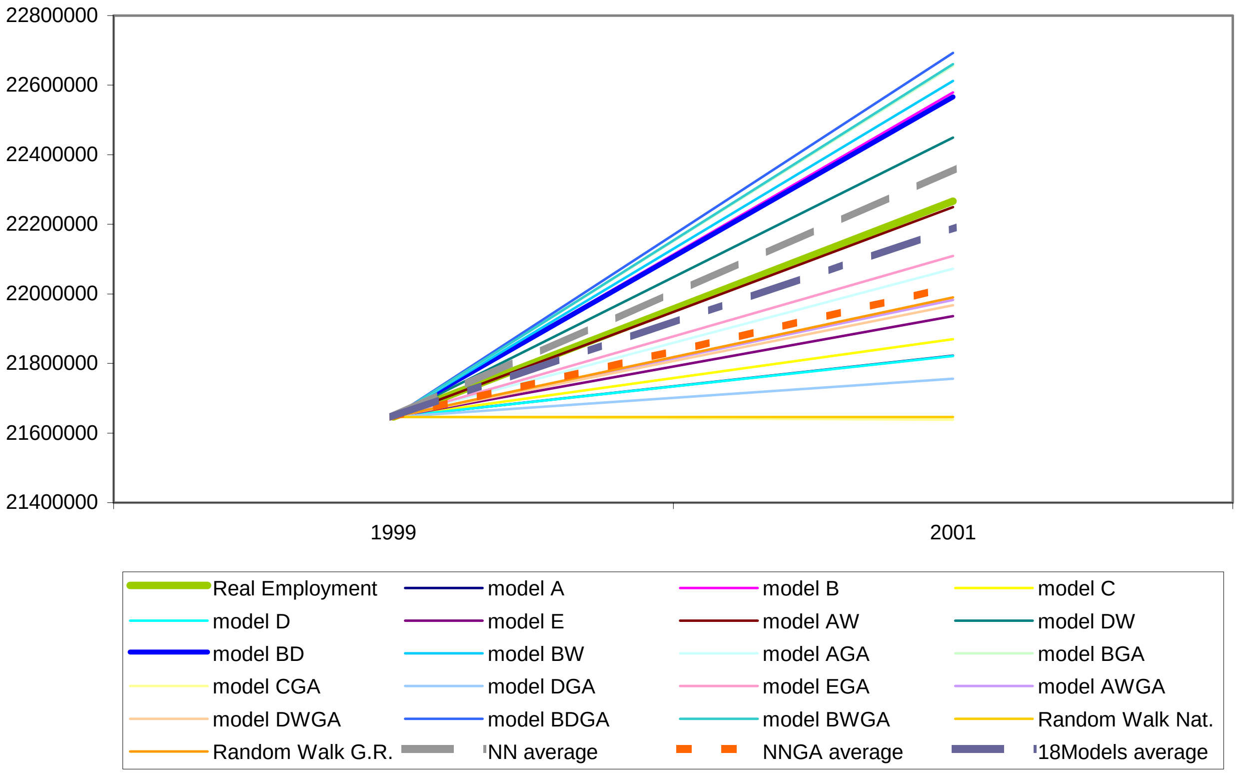

Figure C.1 – West Germany's ex-post forecasts for the year 2001

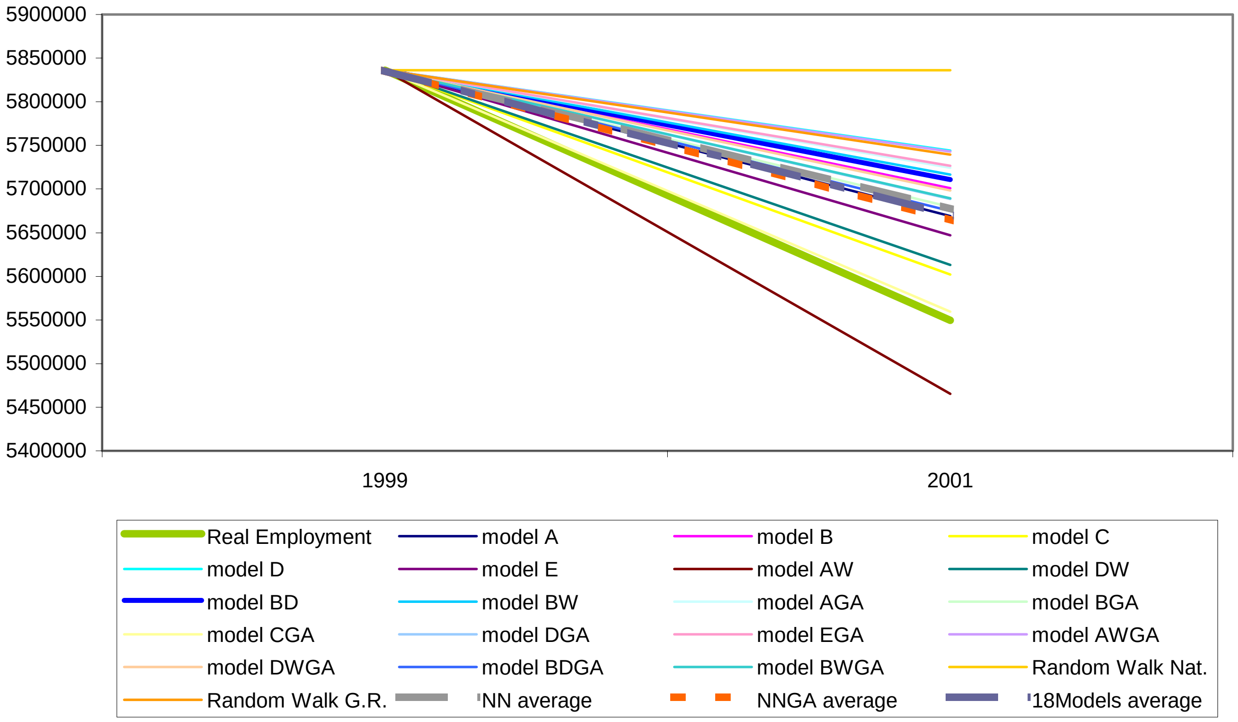

Figure C.2 – East Germany's ex-post forecasts for the year 2001

Annex D – Maps of Error Levels (Year 2001) and Estimated Growth Rates (Year 2003) in Germany

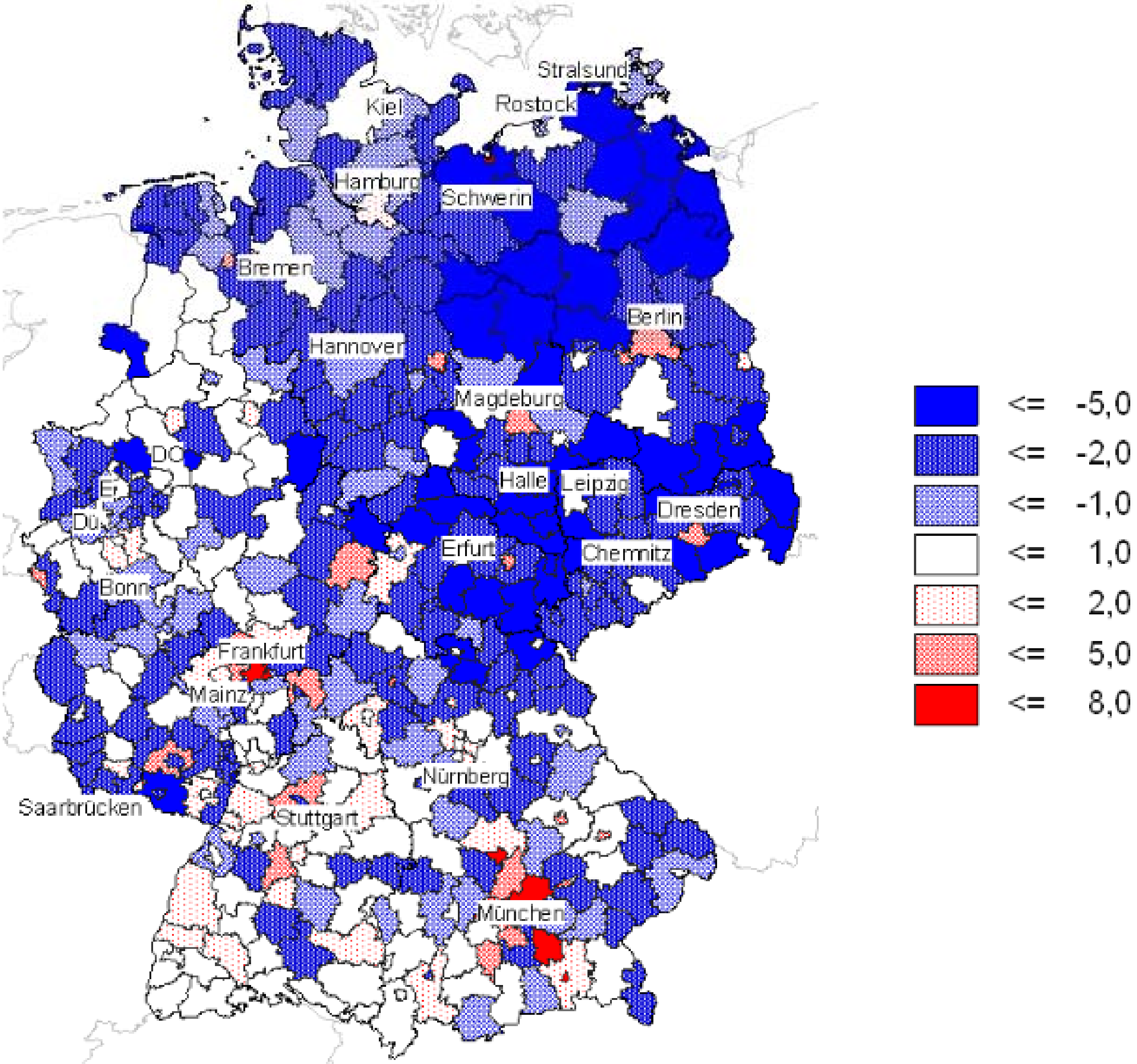

Figure D.1 – Map of error levels – in both West and East Germany – for the average of the nine NN adopted models (ex-post forecasts for the year 2001)

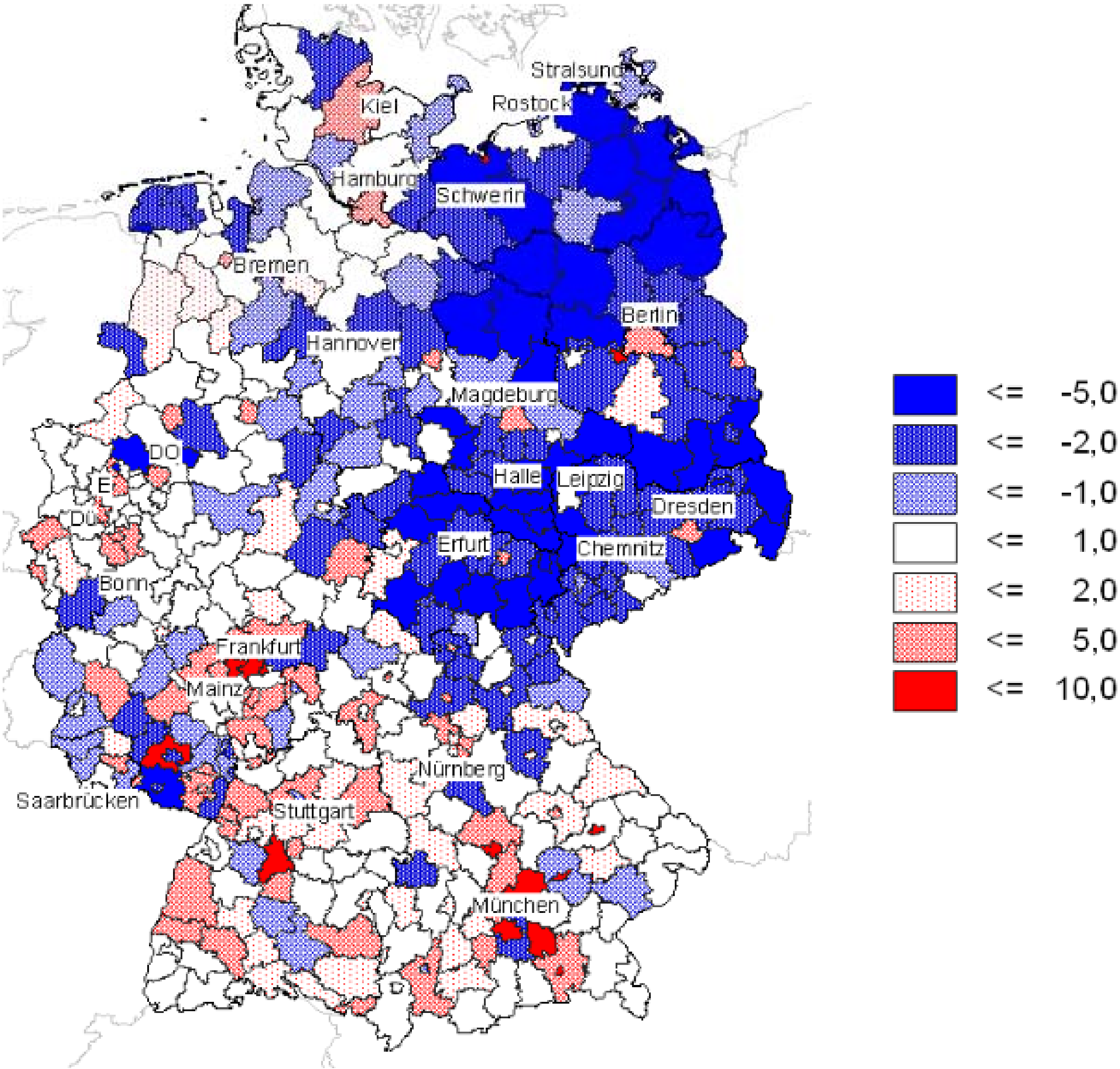

Figure D.2 – Map of error levels – in both West and East Germany – for the average of the nine NNGA adopted models (ex-post forecasts for the year 2001)

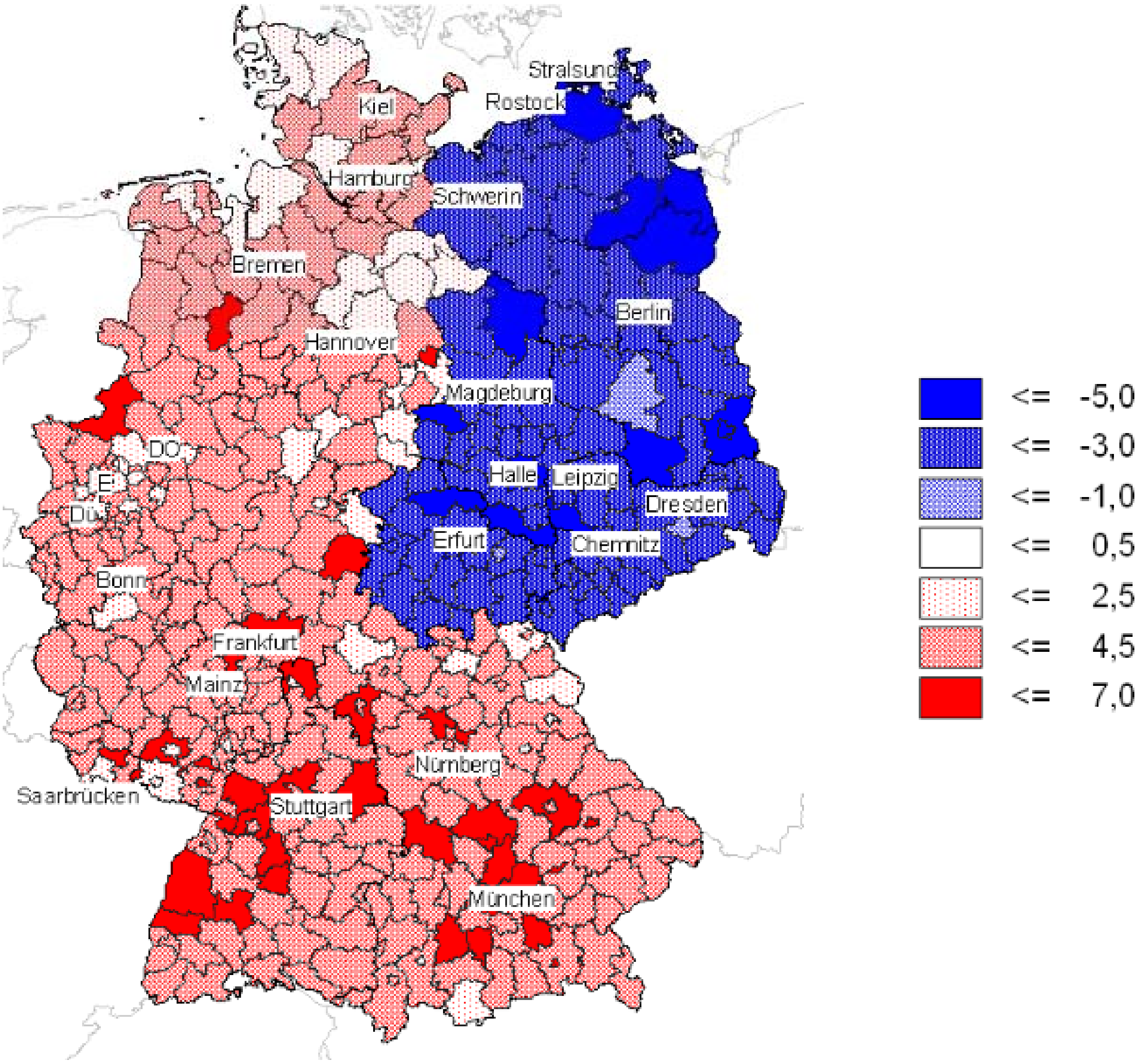

Figure D.3 – Map of estimated growth rates – in both West and East Germany – for the average of the nine NN adopted models (forecasts for the year 2003)

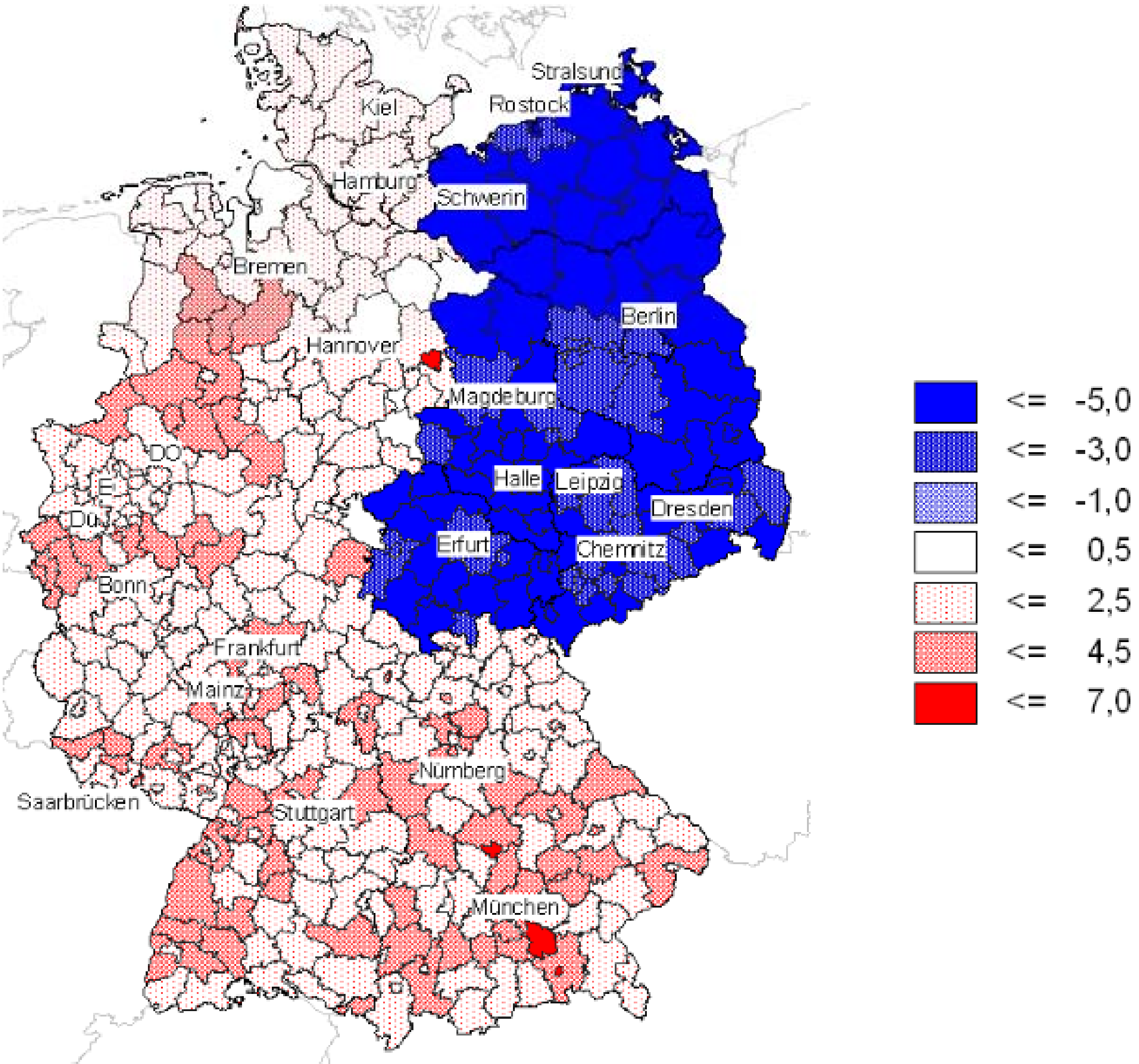

Figure D.4 – Map of estimated growth rates – in both West and East Germany – for the average of the nine NNGA adopted models (forecasts for the year 2003)