NEURAL NETWORKS AND GENETIC
ALGORITHMS AS FORECASTING TOOLS:
A CASE STUDY ON GERMAN REGIONS

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

* Department of Spatial Economics, Faculty of Economics and Business
Administration, VU University Amsterdam, The Netherlands; email:
patuelli@feweb.vu.nl; pnijkamp@feweb.vu.nl

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

*** Department of Economics, Faculty of Statistics, University of Bologna,
Italy; email: aura.reggiani@unibo.it
ABSTRACT

This paper develops and applies neural network (NN) models to forecast regional employment patterns in Germany. Computer-aided optimization tools that imitate natural biological evolution to find the solution that best fits the given case (namely, genetic algorithms, GAs) are also used to detect the best NN structure. GA techniques are compared with more ‘traditional’ techniques which require the supervision of experienced analysts. We test the performance of these techniques on a panel of 439 districts in West and East Germany. Since the West and East data sets have different time spans, the models are estimated separately for West and East Germany. The results show that the West and East NN models perform with different degrees of precision, mainly because of the different time spans of the two data sets. Automatic techniques for the choice of the NN architecture do not seem to outperform selection procedures based on the supervision of expert analysts.

1. Introduction

Key economic variables such as (un)employment have always been considered important indicators of the performance of labour markets, at both the local and the national level. Shocks to labour demand, which eventually lead to permanent changes in employment, are likely to be region- rather than country-specific (see, for instance, the theoretical models by Krugman, 1998, and the empirical evidence by Blanchard and Katz, 1992 and Decressin and Fatás, 1995). To allow policy makers to allocate public expenditures efficiently among regions, labour market forecasts at the regional level are a necessary
complement to forecasts at the national level. The aim of this paper is to forecast the short-term development of the volume of employment for a large number of regions, using only short time series.

Most methods commonly used to compute forecasts require the availability of long time series of (national) aggregates. However, when forecasts at the regional level are needed, the data available for the analysis are likely to include – as in the case of the German labour market – a high number of cross-sections and a small number of time periods. We propose several different techniques to compute employment forecasts at the regional level for 439 German regions: 326 regions are located in West Germany and 113 in East Germany.

Employment data at the regional level are available only at intervals of one year. For the West German regions, data from 1987 to 2001 are available, but for the Eastern regions, the number of observations over time decreases to nine: from 1993 to 2001. Since the number of regions for which the forecasts have to be computed is much higher than the number of time periods for which regional data are available, statistical techniques that are commonly used in time-series analysis cannot be applied. Instead, we propose techniques that exploit the panel nature of the data.

Regional data are often characterized by spatial heterogeneity. After more than a decade since German unification in 1990, Germany is still characterized by significant disparities among its regional labour markets. Regional disparities are visible not only between regions located in the former West and those in the former East Germany, but also within each of the two parts of the country. For example, the southern part of West Germany is developing faster than the rest of the country (see, for instance, Bade, 2006; Bayer and Juessen,
2007). The availability of panel data allows us to correctly identify similarities and differences across regions and obtain more reliable regional employment forecasts.

The adoption of a suitable functional form is critical. The choice between models that impose linear behaviour and models that allow for non-linear behaviour of the relevant variables is extensively discussed in the forecasting literature, though mostly in the context of time series analysis. Linear methods have been extensively used over the years because of their easy implementation and interpretation, although many empirical problems involve non-linear behaviour (Granger and Terasvirta, 1993), in particular when longer forecasting periods are concerned (Zhang, 2001). A number of authors (for instance, Stock and Watson, 1998; Swanson and White, 1997a; 1997b) have compared the performance of linear and non-linear methods – time-series regression versus neural networks (NNs), genetic algorithms, or fuzzy logic – in forecasting variables such as national employment, industrial production or corporate profits. However, the conclusions reached are sometimes contradictory. Stock and Watson (1998) find that non-linear methods, such as NNs, do not perform better than the other proposed linear time-series techniques proposed. Swanson and White (1997b, p. 459) find instead that it might be possible to improve macroeconomic forecasts by ‘using flexible specification econometric models’, whose specification ‘is allowed to vary over time, as new information becomes available’.

Attempts to compute labour market forecasts for German regions using linear techniques have been made by several authors, among others, Blien and Tassinopolous (2001) and Bade (2006). Blien and Tassinopolous (2001)
compute short-term employment forecasts for West German regions by combining a top-down and a bottom-up approach. 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-optimizing procedure. Bade (2006) forecasts the long-term development of regional shares in national employment by means of an extended ARIMA approach. Both methodologies require a number of constraints and economic, as well as econometric, assumptions. A non-linear non-conventional approach may help to overcome such constraints.

In the present paper, we compute short-term forecasts of employment at the regional level for East and West Germany by means of NNs. We assume an autoregressive relationship, in which future developments of employment are the result of its past developments. We also exploit the panel nature of the data by modelling region-specific characteristics.

Modelling panel data in the context of NNs is not straightforward. Nevertheless, NNs have some advantages over conventional techniques. For example, the asynchronic nature of the business cycles among regions may make conventional models rather complicated, imposing constraints that would limit the scope of the analysis. The advantage of artificial NNs is their flexibility and the absence of strong underlying modelling hypotheses; this makes them suitable for our purpose of computing employment forecasts at the regional level.

We find that in terms of forecasting performance NN models generally win over naïve extrapolation models, such as random walk, and ordinary least squares (OLS) panel estimations.
The paper is organized as follows. Section 2 illustrates the methodology adopted in our empirical analysis, that is, traditional NNs and NN models embedding genetic algorithm procedures (NNGA), and concludes with a brief discussion of issues related to neural forecasting. Section 3 describes the data employed in our experiments and the empirical application, which aims at estimating regional employment in West and East Germany by means of the NN and the NNGA approaches. Section 4 makes some concluding remarks and suggestions for future research.

2. Neural Network Models for the Estimation of Employment Variations

2.1. Neural Network Models

NNs are computation algorithms which were initially developed to imitate the functioning of the human brain. The main characteristic of NNs is their ability to find numerical solutions when the relationships between the variables are not fully known. Thus, they are particularly useful when one has a limited knowledge of the phenomenon examined.

Similarly to what happens in the human brain, calculation in NNs is distributed over a number of processing units (neurons) which work in parallel. These units are distributed in ‘layers’ and are internally connected through a set of weights. The layers are made up of units which represent either the input variables, the output variables, or intermediate (hidden) computation units. When no hidden layers are used, input and output units are directly linked, and the NN can be referred to as a ‘linear NN’ or as a 1-layer NN (see, for example,
Chandrasekaran and Manry, 1999). In fact, no computation is carried out at input layer level. In feedforward NNs, the most popular family of NN methods, the units of each layer are unidirectionally connected and transfer information only to units of the succeeding layer.

Following Fischer (2001b, p. 23), we define the generic processing unit \( u_{i,n} \) as:

\[
 u_{i,n} = \varphi(u_{n-1}) = \mathcal{S}(f(u_{n-1})) ,
\]

where \( u_{n-1} = \{u_{1,n-1}, ..., u_{k,n-1}\} \) is the preceding layer of units, and the transfer function \( \varphi \) can be decomposed into two separate functions: the activation function \( \mathcal{S} \), and the integrator function \( f \). The integrator function is used to aggregate the data entering the processing unit \( u_{i,n} \), thus providing a single input, which will be subsequently transferred to the processing units of the successive layer \( u_{n+1} \). The integrator function is a weighted sum:

\[
 v_{i,n} = f(u_{n-1}) = \sum_j w_{ij,n-1} u_{j,n-1} ,
\]

where \( u_{j,n-1} \) is the \( j \)th unit connected to unit \( u_{i,n} \), and \( w_{ij,n-1} \) is the weight connecting the two units (Fischer, 2001a). The activation function – most often a sigmoid/logistic function – computes the unit’s output and can be represented, as (Fischer, 2001b, p. 24):

\[
 \mathcal{S}(v_{i,n}) = \frac{1}{1 + \exp(-\beta v_{i,n})} ,
\]

where \( v_{i,n} = f(u_{n-1}) \), and \( \beta \) defines the slope of the curve. The value of \( \beta \) can be selected a priori or, for example, by means of sensitivity analysis.

---

1 An \( N \)-layer NN implies the computation of \( N \) sets of weights between the layers. Consequently, an NN with no hidden layers is a 1-layer NN (also called a linear NN). Similarly, an NN with one hidden layer is called a 2-layer NN.
A recursive modification of the weights employed in Equation (2) guides the ‘learning’ process (see, for instance, Rumelhart et al, 1986) of an NN. This recursive weight computation is often carried out by means of a Back-Propagation Algorithm (BPA). The BPA – as every other ‘supervised’ NN algorithm – uses input examples and their corresponding outputs, which are provided by the analyst, in order to map out and replicate the data-underlying behaviour. At each iteration of the algorithm the error generated by the NN is propagated backwards through the layers of units, with a consequent modification of the connection weights. Two parameters – learning rate and momentum – define, respectively, the extent and the duration (in terms of iterations) of the corrections (for more details, see Rumelhart et al, 1986).

In this paper we employ conventional feedforward NNs, in which the weights are computed by means of the BPA. The transfer function used is a sigmoid with $\beta = 1$, while the learning rate and momentum parameters are set at 0.9 and 1, respectively and, for reasons of comparison, are kept fixed during the iterative process. A discussion and empirical testing of adaptive learning rates and multiple learning parameter values can be found in Patuelli et al. (2006).

A noteworthy aspect of the NN models concerns the balance between network simplicity and complexity (in terms of the number of layers and computation units). An overly simple NN will not learn the relationship between the input and output variables, and therefore it will generate large errors (Fischer, 2001a). On the other hand, an NN that is too complex will lead to generalization problems (overfitting), causing high variance and unreliable
forecasts.\textsuperscript{2} Many techniques have been proposed to tackle the problem of overfitting. Here we use one of the most common methods: namely, ‘early stopping’, which consists of stopping the learning process (iterations) when the performance indices (the error computed) start to worsen. The NN model concerned will then run for the number of iterations selected by means of the early stopping method.

\textbf{2.2. The Implementation of Genetic Algorithms in Neural Networks}

The previous discussion highlights the difficulty of finding the best NN structure. The high number of choices that have to be made in order to obtain the final forecast generally requires the supervision of an expert analyst. In this paper, we test whether automatic procedures, such as genetic algorithms (GAs), can be a suitable substitute for ‘manual’ – and therefore subjective – techniques used to identify the best NN structure. GAs are used here as optimization procedures to choose the best NN structure and parameters; we should then expect GAs to provide better generalization properties and to reduce the time and work needed in the fine-tuning of NN models.

Genetic algorithms are optimization tools belonging to the class of evolutionary algorithms. They mimic natural biological evolution dynamics and are nowadays widely adopted in the scientific literature for various purposes (see, for example, Fischer and Leung, 1998; Reggiani et al, 2000; 2001).

Formally, GAs are stochastic search methods, which aim to solve an optimization problem that can be expressed as follows (Fischer and Leung, 1998; Nag and Mitra, 2002):

\textsuperscript{2} For a discussion of the model selection problem, see, for instance, Fischer (2000).
\[
\max \{ g(s) | s \in \Omega \},
\] (4)

where \( g \) is a fitness function,\(^3\) and \( s \) is a single ‘individual’ (candidate solution to the optimization problem) belonging to a ‘population’ of \( d \)-dimensional binary vectors called ‘strings’. These strings are used as to represent nature’s genotypes, which contain the genetic information (referred to as the ‘structure’) of an individual.

In our empirical application, the strings include two types of information: a) the NN learning parameters (learning rate, momentum and input noise – a small, randomly distributed disturbance effect); and b) the NN configuration. The NN configuration string contains the total number of layers and the number of computation units in each hidden layer.\(^4\)

Figure 1 shows the functioning of a GA (Fischer and Leung, 1998; Riechmann, 2001). The GA starts from an initial – randomly chosen – set of NN structures, \( m_0 \). Each structure is evaluated by means of the fitness function; here we use the mean square error computed over the input/output examples set. Genetic operators (namely, ‘selection’ and ‘recombination/crossover’; for more details, see Riechmann, 2001; Rumelhart et al, 1986) subsequently generate a new structure, leading to the successive ‘generation’ of NN structures. Lastly, a final operator (‘mutation’, see Fischer and Leung, 1998) introduces an exogenous, stochastic change in the structures.

---

\(^3\) In our case study, the fitness function is an objective function to be minimized on the training set. Fischer and Leung (1998) show how an objective function can be recoded into a fitness function.

\(^4\) In our case, the algorithm tests NN structures with up to two hidden layers, comprising a default maximum of 30 and 10 units in the first and second hidden layer, respectively. A larger number of units per layer can be considered to be superfluous – if not harmful – for NN generalization. As an implicit rule, the second hidden layer will always contain a smaller or equal number of units than the first hidden layer.
<table>
<thead>
<tr>
<th>t := 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creation of First Population $\overline{m}_0$</td>
</tr>
<tr>
<td>Evaluation of $\overline{m}_0$</td>
</tr>
<tr>
<td><strong>while</strong> Stopping Condition not Met</td>
</tr>
<tr>
<td>t := $t + 1$</td>
</tr>
<tr>
<td>Selection from $\overline{m}_{t-1}$ and Reproduction into $\overline{m}_t$</td>
</tr>
<tr>
<td>Recombination on $\overline{m}_t$</td>
</tr>
<tr>
<td>Mutation on $\overline{m}_t$</td>
</tr>
<tr>
<td>Evaluation of $\overline{m}_t$</td>
</tr>
<tr>
<td><strong>End</strong></td>
</tr>
</tbody>
</table>


Figure 1 – Structure of a standard GA

In our experiments, all the structures tested are selected for reproduction/recombination, so as to generate new sets of parameters. Here the mutation operator is limited to 10 per cent of the structures for each generation. Once the newly generated structures $\overline{m}_t$ have been computed, they are substituted for the old ones ($\overline{m}_0$), and their fitness is computed. The process continues until a stopping condition is met. In the present paper the stopping condition is set at ten iterations. At the end of this process, the best-fitting structure obtained in the last iteration is adopted as the NN architecture. At each iteration three structures were generated and evaluated, resulting in three final ‘optimized’ NN configurations to choose from. Although three structures and ten iterations might at first seem insufficient, they nevertheless seem to be enough for our empirical analysis. A combination of 100 iterations and a population size of 100 structures did not improve our results, while greatly increasing computation time.
Further issues regarding the application of NNs to forecasting should be discussed, such as the inclusion of time or the nature of the data-generating process. These are discussed in the next section, together with recent developments in neural forecasting.

2.3. Neural Forecasting

Forecasting is one of the main functions of NNs (see, for example, Lapedes and Farber, 1987; Weigend et al, 1990; Werbos, 1974). With regard to economics, several reviews of the use of NNs in business/financial applications can be found (we refer, for example, to Vellido et al, 1999), while a wider look of neural forecasting is provided by, among others, Zhang et al. (1998).

Neural forecasting is attractive to researchers and practitioners for a number of reasons, one being the weaknesses of both linear methods (which are meant to forecast future values which are linearly related to previous observations) and non-linear methods (which can indeed incorporate richer data information, but were developed for specific problems: for example, logit models are used for discrete choice problems). Additional issues pushing towards the utilization of NNs in forecasting relate to data quality. Problems like multicollinearity (in a panel data framework) and noise in the data invalidate the results of conventional regression analyses. The inclusion of time (correlation) is also a relevant issue in neural forecasting. The above-mentioned topics are briefly discussed in the remaining part of this section.

A critical issue for NNs is the nature of the data-generating process, in particular whether this has linear or non-linear characteristics. Zhang (2001) analyses the suitability of NNs for approximating linear data-generating
processes and finds that NNs ‘do have the competitive ability for linear time-series modeling and forecasting’ (Zhang, 2001, p. 1199). The reason is that (non-)linearity tests are only developed for specific functional forms, and it is therefore difficult to test multiple possible non-linear relations. NNs allow us to bypass the process of choosing the functional form of the model. This is particularly true when sufficiently long series of data are employed. In this case, as shown by Balkin and Ord (2000), NNs are able to detect possible non-linearity in the data and to outperform linear methods.

On the other hand, for the case of linear processes, NNs could be thought to overcomplicate forecasting. However, if we consider – in a panel data forecasting problem – the number of specific characteristics of single regions, we can expect these diverse characteristics to show up in economic data as outliers, which deviate from the average national trend. Furthermore, NNs have been found to provide a comparatively better performance than linear time series models when the data show more statistical noise or when specification/multicollinearity problems occur (Markham and Rakes, 1998, in Zhang, 2001). Though this finding was obtained in a comparison based on time series data, we should consider it to be particularly valuable with regard to the objectives of the present paper, which is concerned with highly disaggregated regional forecasts.

In addition to the nature and distributional properties of the data, a further aspect should be considered when discussing neural forecasting, that is, the extension of the data sets employed and the forecasting horizon. Balkin and Ord (2000) suggest that a long enough data series must be available in order for NNs to outperform simpler methods, though the authors do not provide
additional information supporting this claim. However, Tkacz (2001) also suggests that NNs are more useful for larger data sets. Furthermore, the author stresses, on the basis of multiple time-series-based experiments, that NNs provide a forecasting accuracy advantage when forecasts are carried out for longer time horizons (the author tests a single-quarter and a four-quarter forecast horizon).

This discussion has shown that NNs are particularly helpful, as a forecasting technique, when complex and large data sets are employed. A further issue to be addressed here is the inclusion of time (serial correlation) in NNs. Van Veelen et al. (2000) review the different solutions applicable to NNs dealing with time series data. The authors present two main approaches to the inclusion of time information in NNs: 1) explicit representation of time (which is used in the present paper); and 2) dynamic NN models. The latter NN paradigms – we refer, for example, to time delay neural networks (TDNNs) – have been extensively applied to time series forecasting, though, according to Van Veelen et al. (2000, p. 4), ‘they lost some attention in the last few years’. The authors also stress that such models are not free of problems. TDNNs, for example, make it challenging to employ a BPA, and are found to have poor heuristic properties. Other dynamic NN methods (for a review, see Hagan et al, 1996) resort to the recurrent NN paradigm, starting with the introduction of the Hopfield neuron (Hopfield, 1982).

The alternative approach of explicitly including time information is also criticized (see Van Veelen et al, 2000), since it does not include a dynamic framework. This shortcoming may possibly result in the incapacity of NNs to locate hidden time trends. But it should be pointed out that in most cases trends
or temporary shocks can be accounted for by including one or more counters or a periodic variable in an NN. This type of approach is adopted in our experiments, in which multiple observations per year are processed by the NN, and recognizing specific temporal shocks appears to be critical for maximizing the generalization of the NNs. However, this approach does not account for serial correlation between single region observations, which is instead included in the models by means of lagged variables. The operationalization of the approach discussed above is described in detail in Section 3.1.

Before detailing the actual procedure followed in the development of our NN models, we first describe next the data employed in our experiments.

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

3.1. The Data Set Available

The data set used in our experiments has been provided by the German Institute for Employment Research (Institut für Arbeitsmarkt und Berufsforschung – IAB), and includes information on the number of full-time workers employed every year on 30 June, subdivided into nine economic sectors. In addition to these variables, average regional daily wages earned by full-time workers are also available. All the data are collected for social security purposes, at single firm level.

The data refer to districts in West and East Germany at the NUTS 3 disaggregation level. The data for West Germany cover 15 years (1987 to

---

5 The 9 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.
2001), while the data for 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.

In an effort to identify labour market patterns in regions with a similar development, a 9-point variable identifying the ‘type of district’ is adopted. This variable (Böltgen and Irmen, 1997) was computed on the basis of the urbanization and agglomeration levels of each district. The districts are classified as:

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.

These data will be the basis for the forecasting experiments described in the next section. The aim of our experiments is to forecast 2-years-ahead employment variations in each German district.

3.2. The Neural Networks Models Developed

This section proposes a number of NN models that can be used for our forecasting purposes. The main independent variable in all our NN models is
the biannual (between \( t-2 \) and \( t \)) growth rate of regional full-time employment observed in the nine economic sectors in each district.

To exploit the panel structure of our data, we also use what we indicate as a ‘time’ variable. This can be done in two different ways. The first consists of using a periodic variable identifying the year to which the data refer. The variable is rescaled to the interval \((0, 1)\) and might resemble a trend variable in a time-series model. The second way to include time consists of adding a set of dummy variables (one dummy per year). The use of dummy variables to identify time periods might be compared to a ‘time fixed effects’ approach in a conventional panel modelling framework (Longhi et al., 2005).

A second group of variables can be added to capture the correlation across the observations belonging to the same – or a similar – district. First, a counter ranging from 1 to 326 in West Germany and from 1 to 113 in East Germany is added to (roughly) model district-specific characteristics. The variable is substituted for the more commonly used – when working with panel data – regional dummies (fixed effects), which would require the computation of an overly large number of weights. In panel data modelling, such an ‘incidental parameter problem’ can be avoided by using the ‘within transformation’ (see, for example, Hsiao, 2003). However, this solution does not seem appropriate in our case, since the most important input – sectoral employment – is added as growth rates. We call this variable the ‘district identifier’.

Alternatively, we might assume that regions with a similar degree of urbanization or agglomeration behave in similar ways, ceteris paribus. Consequently, the variable ‘type of district’ (described in Section 3.1) can be
added to the independent variables either as a counter, ranging from 1 to 9, or as a set of nine dummy variables.

Finally, selected models are enhanced with a further input variable: the lagged biannual growth rate of average daily wages earned by full-time workers. The rationale for the inclusion of this variable is the possible relationship between wages and employment.

In total, we computed nine different NN models, whose equation can be generically represented by the following relationship:

$$\Delta e_{i,t+2} = f[T, \text{district}, \Delta e_{i,1,t}, \ldots, \Delta e_{i,9,t}, \Delta w_{i,t}]$$

where $\Delta e_{i,t+2}$, the percentage variation of employment in region $i$ in the period $(t, t+2)$, is a function of: 1) the time variable $T$; 2) district characteristics (either the district identifier or urbanization/agglomeration types); 3) lagged employment variations in the nine economic sectors ($\Delta e_{i,1,t}, \ldots, \Delta e_{i,9,t}$); and 4) lagged variation in average daily wages, $\Delta w_{i,t}$.

The models can then be grouped according to the input variables used:

- Model A and all subsequent models starting with the letter A include sectoral employment and time as dummy variables (time fixed effects).
- Model B and all subsequent models starting with the letter B include sectoral employment and time as a periodic ordinal (trend) variable.
- The following models were developed, on the basis of Model A and Model B, as follows:
  - Model AC also includes the variable ‘district identifier’, to capture region-specific characteristics, while
With Model AD and Model AE also include the variable ‘type of district’ as a counter (Model AD) or as dummy variables (Model AE), in order to capture differences across districts with different urbanization/agglomeration characteristics.

- Model BD also includes the variable ‘type of district’, similarly to Model AD.

Finally, Models AW, ADW and BW (an ending with the letter W) use average daily wages as a further input. These models may be seen as extensions of Models A, AD and B, respectively.

The next section describes the validation process followed for all models, as well as the introduction of genetic algorithm-enhanced NN models (NNGA).

### 3.3. The Validating and Testing Procedure

#### 3.3.1. The Validation Phase

As already mentioned, our NN models use the employment growth rates for the time period \((t-2, t)\) in order to forecast the growth rates for the period \((t, t+2)\). Since the data for West and East Germany start from 1987 and 1993, respectively, the first available forecasting periods are 1989–91 and 1995–97. In the remainder of the paper we will refer to the generic \((t, t+2)\) interval, using the end year of the period (for example, 1989–91 will be referred to as 1991).

The first test phase of our NN experiments – referred to as the model validation phase – is summarized in Table 1. This phase is concerned with the evaluation of a set of alternative NN configurations (see, for instance, Fischer, 1998) and the selection – for each NN model – of the most suitable architecture and training threshold. For this phase we employed data for up to and including
the year 2000. The NN models concerning West Germany were validated on the basis of their performance for the years 1999 and 2000. The NN models for East Germany were instead validated using the year 2000 only, because of the shorter time span of the data set used. The use of a double validation set for the West Germany NN models – and the consequent computation of average statistical results – is expected to provide a more reliable validation of the NN models, as their performance tends not to be uniform across test sets, and to reduce the effect of time-specific shocks on the model validation.

Table 1 – Data utilization for validating the network configuration

<table>
<thead>
<tr>
<th>Models</th>
<th>Training</th>
<th>Validating</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Germany</td>
<td>1991–98</td>
<td>1999–2000</td>
</tr>
<tr>
<td>East Germany</td>
<td>1997–99</td>
<td>2000</td>
</tr>
</tbody>
</table>

In the validation phase, for every NN model we tested five configurations. First, a 1-layer NN was tested, as well as three 2-layer models containing 5, 10 and 15 hidden units, respectively, in one hidden layer. Finally, a 3-layer model was tested, using 5 units in each of the two hidden layers.

The NN models, trained as described above, were evaluated by means of two statistical indicators: MSE and MAE. Given the panel structure of the data,

---

6 The rationale for proceeding in ‘jumps’ of a few computation units in validating NN structures is in the lengthy testing process, and is supported in the empirical literature on NNs. Future experiments may address the behaviour of NNs for intermediate structures (for instance, using 4 or 7 hidden computation units), and will focus on 2-layer NN structures, since empirical evidence has proven that an NN with one hidden layer can approximate nearly any type of function (Cheng and Titterington 1994; Kuan and White 1994).

7 The statistical indicators employed in the validation of the NN models – and more in general in the forecasting literature – are commonly used for the evaluation of both time series and NN forecasts (see, for example, Zhang et al. 1998) and are computed as follows:

- Mean Square Error: \[ \text{MSE} = \frac{1}{N} \left[ \sum (y_i - y_i')^2 \right] \]
- Mean Absolute Error: \[ \text{MAE} = \frac{1}{N} \left[ \sum |y_i - y_i'| \right] \]
these indicators have been computed on the $ex \ post$ forecasts computed by district. Hence, contrary to their usual time-series interpretation, these indicators summarize the error of the forecasts across districts, rather than over time. On the basis of the above indicators, the best-performing structure of the validation phase was selected for each model, and was then employed in the subsequent test phase. The statistical performance of each NN structure was tested for an increasing number of iterations (training epochs), in order to find the optimal training period (after which the performance of the algorithm tends to deteriorate or to reach a plateau).

Subsequently, an additional NN structure, obtained by means of a genetic algorithm (GAs) NN optimizer (see Section 2.2), was selected for each model previously developed. In addition to varying architectures, each NNGA model also employs an alternative set of learning parameters (learning rate, momentum, and an additional input noise component), which are instead fixed in the manually-developed NN models (see Section 2.1). Tables A.1 and A.2 in Annex A summarize the input and network structure of the models developed for West and East Germany, respectively.

The GA-enhanced NN models (NNGA) are identified, in the remainder of the paper, by the GA suffix.

- Mean Absolute Percentage Error: $MAPE = 1/N \times \left[ \sum |y_i - y'_i|/y_i \right] \times 100$, where $y_i$ is the observed value (target); $y'_i$ is the forecast of the model adopted (NN); and $N$ is the number of observations. The MAPE is used in place of the MAE in the statistical evaluation of the $ex \ post$ forecasts, since the forecasting error is computed, in this case, on the employment levels resulting from the estimated growth rates.
3.3.2. The Testing Phase

In the test phase, the evaluation of the structures selected in the validation phase was provided by *ex post* forecasts carried out for the year 2001, in order to finally assess the statistical performance of the NN models developed above. Table 2 summarizes the data sets which were used at this stage. In this phase, the weights were reset to random initial values (between 0 and 0.1), and the models were retrained until the year 2000.

Table 2 – Data utilization for the test phase

<table>
<thead>
<tr>
<th>Models</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Germany</td>
<td>1991–2000</td>
<td>2001</td>
</tr>
<tr>
<td>East Germany</td>
<td>1997–2000</td>
<td>2001</td>
</tr>
</tbody>
</table>

The comparison of the 2001 *ex post* forecasts with the actual data allows us to statistically evaluate the models’ generalization properties. The statistical performance of the models is summarized and compared by means of the MSE and MAPE indicators. Further, the models can be compared – pairwise – by using forecast equality tests, such as the Morgan-Granger-Newbold (MGN) test (Granger and Newbold, 1977). Following Diebold and Mariano (1995) we compute the MGN test as:

\[
MGN = \frac{\hat{\rho}}{\sqrt{1 - \hat{\rho}^2}} \sqrt{N - 1}
\]

It may be argued that an assessment of the models’ performance should ideally be carried out by means of resampling techniques, or Monte Carlo simulations. However, in our case, the focus is on forecasting German employment variations for the most recent year available by exploiting the full extent of the data sets available. Resampling experiments would instead imply the repeated random selection of data subsamples for testing. The MGN test allows us to infer the statistical results found, and we can therefore consider it an acceptable compromise in this regard.
where $\hat{\rho}$ is the estimated correlation between the sum $S$ and the difference $D$ of the forecast error vectors ($N \times 1$) of the two models compared; and $N$ is the number of districts for which forecasts are carried out. The null hypothesis is that of equally accurate forecasts (no correlation between $S$ and $D$) and follows a $t$-student distribution with $(N - 1)$ degrees of freedom. The MGN test relies on two assumptions: the absence of serial correlation and of deviations from normality in the forecasting errors, which have been shown to significantly influence the reliability of the test (Tkacz, 2001). At this stage, we consider the assumption of no serial correlation to be feasible. In fact, in our panel forecast experiments, forecasts are not carried out over time, with a series of time-progressive forecasts, but over regions instead. The assumption of the MGN test would imply, in our case, horizontal correlation, that is, pairwise correlation between (the forecasts for) region 1 and region 2, regions 2 and 3, 3 and 4, and so on. On the contrary, what we attempt to capture with the NNs developed in the present paper is the specificity of each single region.

We also employ a non-parametric test: the sign test (ST) (Lehmann, 1998). The ST is based on the following idea: if two models, 1 and 2, are equally accurate, the number of forecasts of Model 2 which have a bigger error than that of Model 1 will be expected to be 50 per cent of the total number of forecasts obtained. Consequently, Model 1 will be considered superior to Model 2 if Model 2 has higher forecasting errors in more than 50 per cent of the cases. The test statistic is then computed as:

$$\text{ST} = \frac{C - \frac{N}{2}}{\frac{1}{2} \sqrt{N}},$$

(7)
where $C$ is the number of times in which Model 2 shows higher errors than Model 1, and $N$ is, again, the number of forecasts carried out (in our case, the number of districts concerned). The ST statistic follows a normal distribution $N(0, 1)$.

Finally, as benchmarks to our NN models, we will present two random walk (RW) models and an OLS model in Section 3.4.

### 3.4. Benchmark Forecasting Models

For comparison purposes, we propose random walk (RW) models and OLS panel regression. These techniques were chosen for their easy and fast implementation. Also, a comparison with RW models is the first step in the evaluation of any proposed econometric technique (Collopy et al, 1994), that is, a proposed methodology should be at least as accurate as a naïve extrapolation. On the other hand, RW models have shortcomings: for example, they do not exploit the potential explanatory power of explanatory variables. In our work, we employed two types of RW models, which are defined as follows:

a) **Random Walk Nat.:** this model assumes that the number of employees in each district in year $(t+2)$ is equal to the number of employees in year $t$. For example, the forecast for 2001 equals the number of employees in 1999, and the regional growth rates are equal to zero.

b) **Random Walk G.R.:** this model assumes, for the period $(t, t+2)$, the same regional growth rates recorded for the period $(t-2, t)$. Consequently, the regional growth rates of employment between 1999 and 2001 will be equal to those recorded between 1997 and 1999.

Both RW models were calculated separately for each district.
In addition to the RW models, we propose OLS regression models as additional benchmarks for the NNs. As for the NN models presented in Sections 3.2 and 3.3, we developed separate regression models for West and East Germany, employing the same basic variables as those in the NN models, that is, the lagged \((t-2, t)\) biannual growth rates of regional full-time employment observed in nine economic sectors (see Section 3.2). Additionally, time fixed effects (dummies) are added to account for year-specific shocks that affect all districts. No intercept was estimated, for comparability purposes. Because of the above settings and specification, the OLS models are particularly comparable to Model A. Following from Equation (5), the model estimated can be written as follows:

\[
\Delta e_{i,t+2} = b_1 \Delta e_{i,1,t} + \ldots + b_9 \Delta e_{i,9,t} + v_t,
\]

where \(v_t\) represents the dummy variable for year \(t\). Similarly, additional regression models could be carried out, for direct comparability with each NN model. We limit our comparative analysis to the above model because of space limitations.

The next two sections will present our empirical findings with regard to both the NN and the NNGA models. First, Sections 3.5 and 3.6 will show the results obtained for the former West and East Germany, respectively. Subsequently, Section 3.7 will conclude the discussion of our empirical experiments, focussing on the differences in the statistical performance of NN and NNGA models.
3.5. Estimation of West German Employment

As indicated in the previous section, nine NN models were developed and tested for each data set (West and East Germany). On the basis of the NN structures selected (as described in Section 3.3.1), we obtained *ex post* forecasts for the year 2001. The statistical indicators emerging from these experiments (see Section 3.3.2), and computed on the forecasts of full-time employment for each West German district, 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 actual future forecasts.

Table 3 shows that no model wins over the others for both statistical indicators, MSE and MAPE. Moreover, low MSE values tend not to occur along with low MAPE values. The ‘B-type’ models – primarily Model BW – seem to minimize the MSE, while the A-based models seem to perform better when considering the MAPE: Model AC has the best MAPE score. These dichotomous results for quadratic and non-quadratic errors require further investigation. Therefore, we carried out forecast equivalence tests to analyse whether the two models that minimize MSE and MAPE are significantly more accurate than the competing models. Generally, the MGN tests appear to provide different results from the sign test, which may be caused by differences in the indices used and by possible violations of the test assumptions. Size distortions would be found in the case of correlation between the single forecasts (see Section 3.3.2). The tests on Model BW – the winning model for the MSE indicator – did not confirm the dominance of the model. The MGN tests were mostly significant, while the sign tests were all insignificant. On the other hand, the tests carried out for Model AC – the winning model for the
MAPE indicator – show that the model outperformed all ‘B-type’ models and the benchmark models. Model AC also outperforms most of the five ‘A-type’ models. In general, while the results concerning Model BW are not consistent, the statistical evidence for Model AC seems to support its choice as a forecasting model for our case. The inclusion in the model of the district identifier variable, which models single district-specific characteristics, appears to be critical for forecasting West German employment. Overall, the ‘A-type’ models appear to outperform both the ‘B-type’ models and the benchmark models.

At the aggregate level, the NN 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 per cent, while the real growth rate recorded was about 2.87 per cent. The average aggregate occupational levels suggested by the models approximate the total number of employees, with an error as small as 0.34 per cent. Although this figure does not inform us about the district-level variance, it might be considered to be a generally acceptable error margin. The ‘B-type’ models tend to produce similar results, which are slightly higher than the average forecast, with a 1.5 per cent error. A graphical representation of the aggregate forecasts is given in Annex B (Figure B.1).
Table 3 – Statistical performance of the *ex post* forecasts for the year 2001; the case of West Germany

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE</th>
<th>MAPE</th>
<th>MGN</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21702742</td>
<td>2.0619</td>
<td>***</td>
<td>BW</td>
</tr>
<tr>
<td>B</td>
<td>8326739</td>
<td>2.6599</td>
<td>**</td>
<td>ST BW</td>
</tr>
<tr>
<td>AC</td>
<td>20259245</td>
<td>1.8809</td>
<td>***</td>
<td>n.a.</td>
</tr>
<tr>
<td>AD</td>
<td>25233824</td>
<td>2.0101</td>
<td>** ***</td>
<td>*** ***</td>
</tr>
<tr>
<td>AE</td>
<td>19857019</td>
<td>1.8979</td>
<td>***</td>
<td>*** ***</td>
</tr>
<tr>
<td>AW</td>
<td>12909349</td>
<td>1.9433</td>
<td>***</td>
<td>*** ***</td>
</tr>
<tr>
<td>ADW</td>
<td>8806876</td>
<td>2.1272</td>
<td>**</td>
<td>*** ***</td>
</tr>
<tr>
<td>BD</td>
<td>8057188</td>
<td>2.5651</td>
<td>**</td>
<td>*** ***</td>
</tr>
<tr>
<td>BW</td>
<td>7851692</td>
<td>2.7247</td>
<td>**</td>
<td>*** ***</td>
</tr>
<tr>
<td>RW Nat.</td>
<td>22748959</td>
<td>2.6999</td>
<td>2.1124</td>
<td>2.2095</td>
</tr>
<tr>
<td>RW G.R.</td>
<td>158622682</td>
<td>2.1124</td>
<td>2.2095</td>
<td>2.6999</td>
</tr>
<tr>
<td>OLS</td>
<td>24243766</td>
<td>2.5651</td>
<td>2.1124</td>
<td>2.2095</td>
</tr>
</tbody>
</table>

*** Rejection of forecast equivalence at the 99 per cent level
** Rejection of forecast equivalence at the 95 per cent level
* Rejection of forecast equivalence at the 90 per cent level
Because of the statistical variability of the results of the NN models shown in Table 3, we also considered, as a main performance indicator, the error generated by the pooled (averaged) forecast of our NN models, as suggested in Granger and Newbold (1986). A map visualization of these results is presented in Figures C.1 and C.2, respectively (in Annex C). The pooled forecasts appear to underestimate in particular the employment losses in East Germany, regardless of the models’ nature (NN or NNGA).

3.6. Estimation of East German Employment

The data set for East German employment contains information on the number of employees in 113 districts, for the period between 1993 and 2001. The data set has a smaller number of districts and is six years shorter than the one for West Germany. Consequently, only four years could be used for training and validating the models. One year was used for the test (see Table 1).

The structure of our NN models for the forecast of East German regional (full-time) employment was selected by using data for up to and including the year 2000. Annex A (Table A.2) provides the details on the configuration parameters of each NN model. The NN models were subsequently trained up to and including the year 2000, employing the year 2001 as an ex post testing period (see Table 2).

Table 4, which contains the results of the 2001 ex post forecasts, does not show homogeneity in the NN models’ results. As in the case of West Germany, the ‘B-type’ models show, on average, lower quadratic errors (MSE). However, these are not matched by comparable results for MAPE. This statistical evidence matches that observed for the West German models. Although this
aspect deserves further analysis, the results of the forecast equivalence tests provide a first model comparison.

We computed the MGN and sign tests for Models BW and AW, which show the lowest errors for MSE and MAPE, respectively. The sign tests show that Model AW outperforms all alternative NN and benchmark models, while Model BW outperforms the benchmarks only (with the exception of Model AD). The MGN tests are generally insignificant at this stage. Consequently, Model AW appears to be preferable to all alternative models in the case of the East German forecasts. The inclusion of the wage variable in the model appears to be critical. The RW models and the OLS regression generally appear to be outperformed by the NN models, though they appear to be competitive when the quadratic error (MSE) is considered.

The calculation of aggregated forecasts shows that all NN models predict a decrease in the occupational levels for the 1999–2001 period, clustering – in particular with regard to the ‘B-type’ models – around an estimated loss in employment of about 2.85 per cent. The observed decrease was 4.91 per cent (see Figure B.2 in Annex B for a graphical representation).

The NN models presented in this and the preceding section were subsequently compared with GA-enhanced otherwise-identical NN models. The next section summarizes the results obtained.

---

9 This issue deserves further investigation, which will be carried out in future research.
Table 4 - Statistical performance of the *ex post* forecasts for the year 2001; the case of East Germany

<table>
<thead>
<tr>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>Model</th>
<th>RW</th>
<th>RW G.R.</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>AC</td>
<td>AD</td>
<td>AE</td>
<td>AW</td>
<td>ADW</td>
<td>BD</td>
<td>BW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>13855323</td>
<td>10504722</td>
<td>36849400</td>
<td>10057159</td>
<td>18214516</td>
<td>50357944</td>
<td>17719380</td>
<td>9627065</td>
<td>8076538</td>
<td>9055105</td>
<td>10024545</td>
</tr>
<tr>
<td>MAPE</td>
<td>4.7083</td>
<td>5.1026</td>
<td>4.1479</td>
<td>5.9226</td>
<td>4.5186</td>
<td><strong>2.9360</strong></td>
<td>3.7931</td>
<td>5.1698</td>
<td>5.1244</td>
<td>7.0492</td>
<td>6.1832</td>
</tr>
<tr>
<td>MGN</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BW</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST BW</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGN</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST AW</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

*** Rejection of forecast equivalence at the 99 per cent level
** Rejection of forecast equivalence at the 95 per cent level
* Rejection of forecast equivalence at the 90 per cent level
3.7. NN Models vs. NNGA Models

In this section we present a summary of the results obtained, for the year 2001, by enhancing our NN models with a GA-based structure and parameter optimization (NNGA models, see Section 2.2). Table 5 summarizes our findings.

To compare these additional models with the previously-developed NN models, Table 5 first identifies which NNGA models show lower MSE or MAPE than their corresponding NN model. For each NNGA model, we test the equivalence of the ‘new’ (NNGA) and ‘old’ (NN) forecasts.

The comparison of NN and NNGA models shows that, in the ex post forecasts for the year 2001, the NNGA models perform ambiguously compared with the conventional NN models. Only one model (model AGA, for West Germany) improves the statistical reliability of its baseline model for both indicators, while the average error levels of the NNGA models are greater than those of conventional NN models. The MGN and sign tests – comparing each NNGA model with its baseline NN model – confirmed the above result, since only a limited number of tests were statistically significant. These tendencies are also visible at an aggregate level (for a graphical visualization of the average NN and NNGA aggregate forecasts, see Figures B.1 and B.2 in Annex B), suggesting that the inclusion of GA in the NN model-setting process does not lead – in our case study – to conclusive statistical results.

On the other hand, in spite of generally higher errors, the NNGA models still seem to perform slightly better than conventional NN models in terms of MAPE when pooled regional forecasts (average NN and NNGA MAPEs) are considered (see also Figures C.1 and C.2 in Annex C).
Table 5 – Comparative statistical performances of NN and NNGA models: *ex post* forecasts for the year 2001

<table>
<thead>
<tr>
<th></th>
<th>Model AGA</th>
<th>Model BGA</th>
<th>Model ACGA</th>
<th>Model ADGA</th>
<th>Model AEGA</th>
<th>Model AWGA</th>
<th>Model ADWGA</th>
<th>Model BDGA</th>
<th>Model BWGA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>West Germany</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>MAPE</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>MGN</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>ST</td>
<td>***</td>
<td></td>
<td>*</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>East Germany</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSE</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>MAPE</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>MGN</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>ST</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Y = NNGA model has lower error than corresponding NN model
N = NNGA model has higher error than corresponding NN model
*** Rejection of forecast equivalence at the 99 per cent level
** Rejection of forecast equivalence at the 95 per cent level
* Rejection of forecast equivalence at the 90 per cent level
Trying to explain the above differences between NN and NNGA models is indeed the most difficult part. Nevertheless, we can make a number of hypotheses on the basis of our results. First, the stochastic nature of the choice of the NNGA structures might play a role in determining a higher variance in the models’ performance. The GA-enhanced models develop more heterogeneous structures than the simpler NN models, which were set by means of a manual procedure (see Section 3.3.1). The settings chosen for the NNGA models may be expected to provide improved performance, because of the optimization procedure that generates them. This automated process also relieves the analyst from the lengthy process of manual choice of the network parameters and configuration.

Secondly, it might be seen as beneficial to allow the GA to go through a greater number of iterations than are used in the present paper, in order to have a wider set of alternatives examined by the software (this issue is briefly discussed in Section 2.2). On the other hand, a shortcoming of the NNGA approach is that computing time increases significantly, in particular for wide data sets such as ours.

A balance between the optimization and the computational time aspects of GAs should then be sought, in particular when database size becomes relevant.

4. Conclusions

The aim of this paper was to produce 2-years-ahead forecasts of the number of individuals employed in the 439 (NUTS 3) districts in Germany. Several models based on neural network (NN) and genetic algorithm (GA) techniques have been developed. Because of data availability, East and West German
districts have been analysed separately, and comparable sets of neural models were applied to Eastern and Western districts. The models were developed and configured both manually (NN models) and by means of GAs (NNGA models). The results of ex post forecasts for the year 2001 have been evaluated by means of MSE, MAPE, and of forecast equivalence tests: namely, the Morgan-Granger-Newbold (MGN) test and the sign test.

From the empirical point of view, concerning the models’ statistical results, it can be seen that the ‘B-type’ models, which utilize the time variable as a periodic ordinal variable (see Section 3.2), perform differently from the ‘A-type’ models, which use sets of time dummys (time fixed effects). In fact, the ‘A-type’ and ‘B-type’ models – for both West and East Germany – tend to minimize different indicators, namely MAPE and MSE, respectively. Though the reasons for this result require further investigation, one factor could be the different nature of the MSE and MAPE indicators. While the desirable conclusion of such analyses would be to identify a consistent and reliable NN model, this is made more difficult by the dichotomy in our results. Hence, there is the need for a statistical (inferential) assessment of forecasting quality. The results of the forecast equivalence tests carried out for the MSE and MAPE winning models suggest – for both the West and East Germany models – that the ‘A-type’ models, that is, an NN approach based on time dummys, appear to be preferable to the ‘B-type’ models, which employ a periodic time variable. Additionally, the modelling of district-specific characteristics and the inclusion of wages appear to be critical for the West and East German NN models, respectively. This result would imply that, for the homogeneous East German districts, wages are an important driving force of labour demand, while in the
more heterogeneous West Germany – where, for example, Bavaria has a largely
dominating economic position – single-district specificities influence the
distribution of economic results and the derived labour demand.

Still, a generalization of the results – at this stage of the analysis – can be
problematic. We then carried out an alternative approach, by considering the
use of pooled forecasts. Combined forecasts provide the average performance
of the analysed models or of a subset of them.\textsuperscript{10} The forecasts emerging by
separately pooling the results of the NN and NNGA models – are mapped in
Figure C.3 (in Annex C).

From a methodological viewpoint, the enhancement of the NN models with
GAs did not seem to significantly improve the NNs’ performance, showing
mixed results in our case study of labour market forecasting. Increased
iterations subsequently carried out did not bring any better results, suggesting
that the learning parameters and NN structure (that is, the settings which are
modified by the GA) have a limited weight in defining the performance of our
models.

In conclusion, our experiments show that NN and NNGA models for
forecasting regional German employment have different levels of reliability,
depending on the data sets used and the socioeconomic background. This is
certainly caused by the different time spans of the data sets for West and East
Germany. It should also be remarked that our empirical analysis has been based
on two main explanatory variables (employment and wages), and thus it cannot

\textsuperscript{10} It should be noted that Granger and Newbold (1986), who suggest the use of pooled forecasts,
originally referred to experiments based on time series data.
be comprehensive with regard to the many variables that may come into play when employment and social conditions are at stake.

Finally, concerning the policy issue of generalization, since the West German data set is much wider, both spatially and temporally, than the East German, the results obtained for the former might be considered to be more reliable and suitable for benchmarking considerations.

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 included in Annex C. Useful comments by three anonymous referees are gratefully acknowledged.

References


Chandrasekaran H, Manry M T, 1999, “Convergent design of a piecewise linear neural network” in *International Joint Conference on Neural Networks (IJCNN '99)* volume 2, 1339 – 1344


Kuan C-M, White H, 1994, “ANNs: an econometric perspective” Econometric Reviews 13 1 – 91


Annex A – Details of Model Experiments

The NN models presented in this 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. All the NN parameters are used in the default values suggested by the software employed (Neuralyst™, see Cheshire Engineering Corporation 1994).

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

<table>
<thead>
<tr>
<th>NN Models</th>
<th>Inputs</th>
<th>IU</th>
<th>HU</th>
<th>Epoch</th>
<th>LR</th>
<th>M</th>
<th>IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>Employment (GR), time (dummies)</td>
<td>22</td>
<td>10</td>
<td>900</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model B</td>
<td>Employment (GR), time (trend)</td>
<td>10</td>
<td>650</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Model AC</td>
<td>Employment (GR), time (dummies), district identifier</td>
<td>23</td>
<td>5</td>
<td>600</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AD</td>
<td>Employment (GR), time (dummies), type (ordinal)</td>
<td>23</td>
<td>10</td>
<td>600</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AE</td>
<td>Employment (GR), time (dummies), type (dummies)</td>
<td>31</td>
<td>10</td>
<td>200</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AW</td>
<td>Employment (GR), time (dummies), wage (GR)</td>
<td>23</td>
<td>5</td>
<td>750</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model ADW</td>
<td>Employment (GR), time (dummies), type (ordinal), wage (GR)</td>
<td>24</td>
<td>15</td>
<td>900</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model BD</td>
<td>Employment (GR), time (trend), type (ordinal)</td>
<td>11</td>
<td>10</td>
<td>300</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model BW</td>
<td>Employment (GR), time (trend), wage (GR)</td>
<td>11</td>
<td>1600</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
## NN Models enhanced with GAs

<table>
<thead>
<tr>
<th>Inputs</th>
<th>IU</th>
<th>HU</th>
<th>Epoch</th>
<th>LR</th>
<th>M</th>
<th>IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model AGA: Employment (GR), time (dummies)</td>
<td>22</td>
<td>24(1\textsuperscript{st}L), 5(2\textsuperscript{nd}L)</td>
<td>250</td>
<td>0.8279</td>
<td>0.2252</td>
<td>0.0071</td>
</tr>
<tr>
<td>Model BGA: Employment (GR), time (trend)</td>
<td>10</td>
<td>0</td>
<td>400</td>
<td>0.9013</td>
<td>0.3330</td>
<td>0.0118</td>
</tr>
<tr>
<td>Model ACGA: Employment (GR), time (dummies), district identifier</td>
<td>23</td>
<td>29</td>
<td>350</td>
<td>0.9492</td>
<td>0.1246</td>
<td>0.0101</td>
</tr>
<tr>
<td>Model ADGA: Employment (GR), time (dummies), type (ordinal)</td>
<td>23</td>
<td>27</td>
<td>600</td>
<td>0.9575</td>
<td>0.5977</td>
<td>0.0175</td>
</tr>
<tr>
<td>Model AEGA: Employment (GR), time (dummies), type (dummies)</td>
<td>31</td>
<td>24(1\textsuperscript{st}L), 8(2\textsuperscript{nd}L)</td>
<td>200</td>
<td>0.6892</td>
<td>0.0515</td>
<td>0.0198</td>
</tr>
<tr>
<td>Model AWGA: Employment (GR), time (dummies), wage (GR)</td>
<td>23</td>
<td>29(1\textsuperscript{st}L), 9(2\textsuperscript{nd}L)</td>
<td>350</td>
<td>0.6002</td>
<td>0.4409</td>
<td>0.0028</td>
</tr>
<tr>
<td>Model ADWGA: Employment (GR), time (dummies), type (ordinal), wage (GR)</td>
<td>24</td>
<td>24(1\textsuperscript{st}L), 10(2\textsuperscript{nd}L)</td>
<td>300</td>
<td>0.8294</td>
<td>0.1348</td>
<td>0.0076</td>
</tr>
<tr>
<td>Model BDGA: Employment (GR), time (trend), type (ordinal)</td>
<td>11</td>
<td>0</td>
<td>500</td>
<td>0.7982</td>
<td>0.2698</td>
<td>0.0164</td>
</tr>
<tr>
<td>Model BWGA: Employment (GR), time (trend), wage (GR)</td>
<td>11</td>
<td>0</td>
<td>1800</td>
<td>0.8416</td>
<td>0.2774</td>
<td>0.0187</td>
</tr>
</tbody>
</table>

Notes: IU = input units; HU = hidden units; LR = learning rate; M = momentum; IN = input noise; GR = growth rates; 1\textsuperscript{st}L = first hidden layer; 2\textsuperscript{nd}L = second hidden layer. All models have only 1 output unit; the activation function is a sigmoid.
Table A.2 – Parameter values of the NN models adopted; the case of East Germany

<table>
<thead>
<tr>
<th>NN Models</th>
<th>Inputs</th>
<th>IU</th>
<th>HU</th>
<th>Epochs</th>
<th>LR</th>
<th>M</th>
<th>IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>Employment (GR), time (dummies)</td>
<td>16</td>
<td>10</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model B</td>
<td>Employment (GR), time (trend)</td>
<td>10</td>
<td>5(1stL), 5(2ndL)</td>
<td>900</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AC</td>
<td>Employment (GR), time (dummies), district identifier</td>
<td>17</td>
<td>10</td>
<td>300</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AD</td>
<td>Employment (GR), time (dummies), type (ordinal)</td>
<td>17</td>
<td>5</td>
<td>300</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AE</td>
<td>Employment (GR), time (dummies), type (dummies)</td>
<td>25</td>
<td>15</td>
<td>300</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model AW</td>
<td>Employment (GR), time (dummies), wage (GR)</td>
<td>17</td>
<td>5(1stL), 5(2ndL)</td>
<td>200</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model ADW</td>
<td>Employment (GR), time (dummies), type (ordinal), wage (GR)</td>
<td>18</td>
<td>5(1stL), 5(2ndL)</td>
<td>200</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model BD</td>
<td>Employment (GR), time (trend), type (ordinal)</td>
<td>11</td>
<td>15</td>
<td>1100</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Model BW</td>
<td>Employment (GR), time (trend), wage (GR)</td>
<td>11</td>
<td>5</td>
<td>1000</td>
<td>0.9</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NN Models enhanced with GAs</th>
<th>Inputs</th>
<th>IU</th>
<th>HU</th>
<th>Epochs</th>
<th>LR</th>
<th>M</th>
<th>IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model AGA</td>
<td>Employment (GR), time (dummies)</td>
<td>16</td>
<td>26(1stL), 8(2ndL)</td>
<td>300</td>
<td>0.5685</td>
<td>0.799</td>
<td>0.0022</td>
</tr>
<tr>
<td>Model BGA</td>
<td>Employment (GR), time (trend)</td>
<td>10</td>
<td>19</td>
<td>1700</td>
<td>0.6878</td>
<td>0.3651</td>
<td>0.0230</td>
</tr>
<tr>
<td>Model ACGA</td>
<td>Employment (GR), time (dummies), district identifier</td>
<td>17</td>
<td>14</td>
<td>200</td>
<td>0.6385</td>
<td>0.0994</td>
<td>0.0019</td>
</tr>
<tr>
<td>Model ADGA</td>
<td>Employment (GR), time (dummies), type (ordinal)</td>
<td>17</td>
<td>16</td>
<td>200</td>
<td>0.9573</td>
<td>0.1433</td>
<td>0.0129</td>
</tr>
<tr>
<td>Model AEGA</td>
<td>Employment (GR), time (dummies), type (dummies)</td>
<td>25</td>
<td>16</td>
<td>100</td>
<td>0.9443</td>
<td>0.0666</td>
<td>0.0061</td>
</tr>
<tr>
<td>Model AWGA</td>
<td>Employment (GR), time (dummies), wage (GR)</td>
<td>17</td>
<td>8</td>
<td>200</td>
<td>0.5705</td>
<td>0.0272</td>
<td>0.0170</td>
</tr>
<tr>
<td>Model ADWGA</td>
<td>Employment (GR), time (dummies), type (ordinal), wage (GR)</td>
<td>18</td>
<td>6</td>
<td>100</td>
<td>0.8544</td>
<td>0.0764</td>
<td>0.0034</td>
</tr>
<tr>
<td>Model BDGA</td>
<td>Employment (GR), time (trend), type (ordinal)</td>
<td>11</td>
<td>29</td>
<td>1000</td>
<td>0.7201</td>
<td>0.4295</td>
<td>0.0196</td>
</tr>
<tr>
<td>Model BWGA</td>
<td>Employment (GR), time (trend), wage (GR)</td>
<td>11</td>
<td>13</td>
<td>200</td>
<td>0.6973</td>
<td>0.4033</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Notes: 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 a sigmoid.
Annex B – Aggregate Ex-Post Forecasts for the Year 2001

Note: The graph does not suggest employment levels for the year 2000 to be a linear interpolation of the years 1999 and 2001.

Figure B.1 – West Germany’s *ex post* forecasts for the year 2001

Note: The graph does not suggest employment levels for the year 2000 to be a linear interpolation of the years 1999 and 2001.

Figure B.2 – East Germany’s *ex post* forecasts for the year 2001
Annex C – Maps of Error Levels (Year 2001) in Germany

Figure C.1 – Map of error levels – in both West and East Germany – for the average of the nine NN adopted models, with reference to the employed by district (ex post forecasts for the year 2001)
Figure C.2 – Map of error levels – in both West and East Germany – for the average of the nine NNGA adopted models, with reference to the employed by district (ex post forecasts for the year 2001)
Figure C.3 – Map of estimated growth rates – in both West and East Germany – for the average of the nine NN adopted models, with reference to the employed by district (forecasts for the year 2003)