

Paper to be presented at the 42nd European Regional Science Association (ERSA)
congress in Dortmund August 27th – 31st, 2002

Forecasting Regional Labour Markets in Germany An Evaluation of the Performance of Neural Network Analysis

S. Longhi^a, P. Nijkamp^b, A. Reggiani^c, U. Blien^d

Abstract

The aim of the present paper is to forecast regional employment developments in the 327 West German districts. Using a Neural Networks (NNs) methodology, we try to identify the existence of underlying structural relationships between the input variables – data on regional and sectoral employment and wages – and the future development of employment at a district level. In order to offer reliable forecasts for the years 2000 and 2001, a variety of NN models have been developed and compared. The emerging results confirm the ability of NNs to capture complex data structures in the training and test phases and hence to ‘extrapolate’ useful information in a multi-regional context. With regard to the forecasting phases, our analysis highlights the necessity of carrying out further research experiments – introducing additional economic background variables – in order to get more insight into the mechanisms and structures of spatio-temporal employment data.

^a Tinbergen Institute Amsterdam, Keizersgracht 482, 1017 EG Amsterdam (The Netherlands); e-mail: longhi@tinbergen.nl

^b Department of Spatial Economics, Free University, De Boelelaan 1105, 1085 HV Amsterdam (The Netherlands); e-mail: pnijkamp@feweb.vu.nl;

^c Department of Economics, Faculty of Statistics, University of Bologna, Piazza Scaravilli 2, 40126 Bologna (Italy); e-mail: reggiani@economia.unibo.it;

^d Institut fuer Arbeitsmarkt und Berufsforschung (Institute for Employment Research - IAB), Regensburger Str. 104, D-90478 Nuernberg (Germany); e-mail: uwe.blien@iab.de

1. Issues of Modern Labour Market Analysis

Wage and (un)employment are variables usually considered main indicators for the functioning of the labour market (see, for example, Rouwendal et al., 1987). Based on such indicators, many studies try to explain, from a theoretical perspective, important labour market characteristics, such as the level of equilibrium unemployment or the presence of wage rigidities, often using micro data. We may refer here, for example, to Pissarides (1985, 2000), who has performed studies of search frictions and job matching based on the hypothesis of heterogeneity of jobs and workers, or to the analysis of Shapiro and Stiglitz (1984), who explain the existence of wages that are higher than the market clearing level as based on the firm's decisions to increase the level of work effort in its plants. Examples of an empirical application of such theories can, *inter alia*, be found in Bunzel et al. (2000). Recent studies try to combine the theories referred to above in a common analysis framework (see, for example, Rocheteau, 2001). Clearly, detailed information on labour market characteristics is necessary for an understanding of the empirical aspects of malfunctioning labour markets.

The existence of such labour market inefficiencies is usually regarded as connected to the occurrence of (persistent) unemployment, one of the most important social problems in modern economies. The labour market is therefore an important concern of governments. As can be seen from Figure 1, governments spend a large percentage of their budgets on labour market programmes with the goal of increasing the number of jobs and reducing unemployment.

Furthermore, in the last decades, several western economies have faced high rates of unemployment characterised by strong regional differences either between core and peripheral regions or between urban centres and their hinterlands (Gleave, 1987). According to OECD statistics, regional disparities in employment have increased considerably in many countries during the 1970s and the 1980s, and have either stabilised or continued to increase between 1985 and 1997 (OECD, 2000). As a consequence, increasing attention is nowadays given to regional development issues in order to gain better insight into the specific labour market conditions that lead to regional disparities, as well as to understand the forces that drive or create barriers to the development of regions that structurally lag behind the national, European or world average¹.

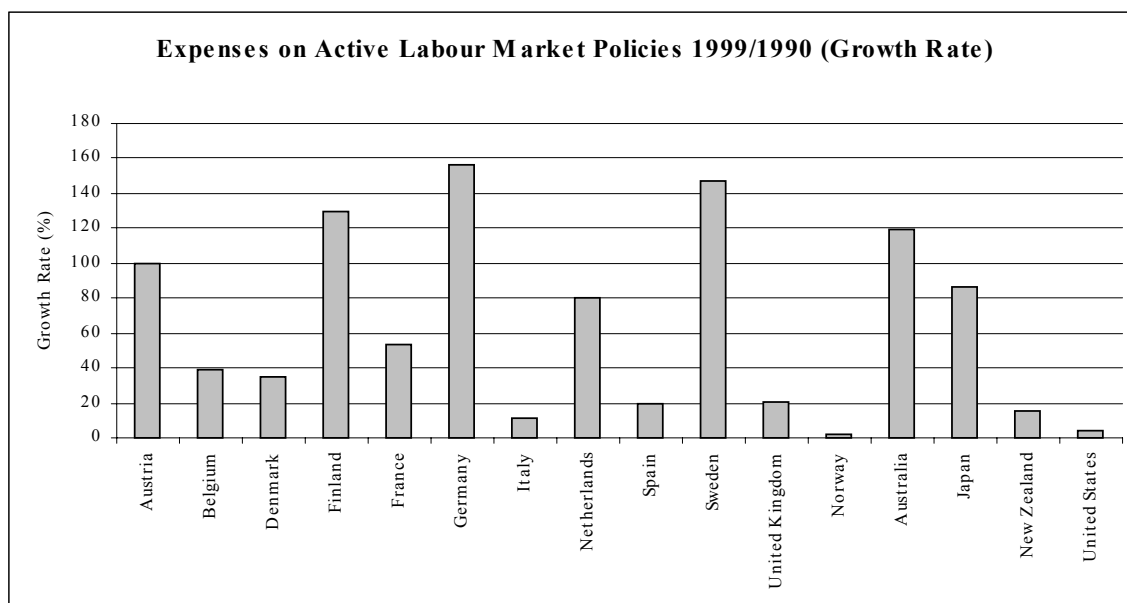


Figure 1: Expenses on active labour market policies
Source: Our elaboration of data from OECD/DICE² (2001)

Within this framework, labour market studies able to detect specific regional development patterns and to identify characteristics of winning and losing regions can be extremely useful in helping governments to develop more effective labour market policies for reducing disparities and favouring consistent development among regions.

As pointed out by Fischer et al. (1987), regions are, in principle, non-homogeneous, and labour markets are segmented by industry, job type, occupational structure, as well as by their geographic location. On the labour supply side, for example, the evolution of a regional labour force is strictly related with changes in labour force participation as well as with changes in net migration (see Fischer et al., 1991). On the labour demand side, many hypotheses have been proposed to explain regional differences in unemployment and in the timing of spatial business cycles. Some of these hypotheses explain regional performance in terms of regional industry structure; others emphasise the importance of the position of the regions concerned in the spatial hierarchy, and in the level of integration of these regions in the national economy (Fischer et al., 1991). In recent years Krugman (1991, 1998a, 1998b), in particular, describes the economy as a self-organising system in spatial clusters – regions – often characterised by an uneven distribution of activity among locations. This tendency of regions to specialise themselves in specific industries and economic sectors will ultimately influence related labour market behaviour by causing a subdivision of regions into urban agglomerations and rural areas, as well as into core and peripheral regions. In addition, regional labour

markets may be affected by shocks on the demand side that can influence employment in a more or less transitory way, depending on both the degree of regional specialisation and the willingness of workers and firms to migrate to other regions (Decressin et al., 1994).

Regional performance is also connected with changes in technology. In particular, Fischer et al. (1991) have pointed out the complexity of the relationships linking technological change and labour market development, emphasising the necessity of focussing the analysis at a meso- (spatial labour markets and sectors) or micro- (firms) level. As a consequence, given the existence of regional disparities, it seems straightforward and necessary to compare region-specific developments. In their paper, for example, Blanchard et al. (1992) explore whether or not labour market disturbances are symmetrically distributed across US regions; a similar analysis is performed for the EU by Decressin et al. (1994). Since they produce specific ranges of goods, regions are affected by different specific fluctuations. This implies that different shocks to labour demand, which eventually lead to permanent changes in employment growth, are region- instead of country-specific.

The study of labour markets from a regional perspective presents various intriguing problems: one of them concerns the level of data (dis)aggregation. Many authors (Bauman et al., 1983; Nijkamp et al., 1984) have shown that different conclusions may be drawn for the same problem, simply by using different levels of data aggregation. Non-linearity and regional differences, for example, may introduce aggregation bias in the macro relation, and therefore lead to incorrect results (Rouwendal et al., 1987). There is the additional problem of the ecological fallacy if the conclusions drawn from aggregated data are transferred to individual economic actors or units.

One interesting study on regional wages is the Blanchflower and Oswald ‘*wage curve*’ (1994), a relationship linking the level of wages to the unemployment rate. In contrast to neo-classical models, which state that there is a positive relationship between wages and unemployment, Blanchflower and Oswald (1990, 1994, 1995) claim that the level of local unemployment and wages are linked in a negative way. Although other time-series studies (see, for example, Chiarini et al., 1997; Johansen, 1997) have also proved the existence of a long-run trade-off between wages and unemployment, the Blanchflower-Oswald empirical findings, based on data from 12 countries, have led to an impressive number of subsequent related studies³ in which attempts have been made to replicate their experiments using other data sets. In a recent study Nijkamp and Poot

(2001) have applied modern meta-analysis techniques to these kinds of studies in order to test the robustness of the wage curve elasticity.

A second problem concerning the regional approach to labour market analysis lies in the characterisation of the observational unit: the “labour market area”. The generally accepted definition states that a spatial labour market is a “*region within which there is a clear labour market pattern defined by the spatial range of employment opportunities open to a worker without changing his place of residence*” (Fischer et al., 1987, p. 3). Consequently, the regional labour market pattern depends not only on market employment opportunities, but also on transport accessibility to regions, and on knowledge of alternative employment opportunities. This definition of the regional labour market implies that the size of the unit of analysis is continuously subject to change. The development of the modern Information and Communication Technology (ICT), for example, as well as improvements in transportation systems have significantly increased the “regional” labour market’s area during the last decade. Unfortunately, the economic definition of a spatial labour market is often rather rigid and is usually based on administrative demarcations that are strictly related to the level at which data are collected.

Nowadays the increasing accessibility of statistical information at both the macroeconomic and the microeconomic level, along with the availability of more powerful computers, facilitates the analysis of employment problems from a more disaggregated perspective. The availability of detailed information clearly demonstrates the need for more advanced, complicated models and analytical frameworks that are able to properly describe the behaviour of labour market variables, in order to offer more reliable regional forecasts, which may enable policy-makers to develop effective labour market policies. From this perspective, it may be difficult to design a simple integrated and operational spatial model that links the available information in a theoretically and methodologically satisfactory way. Conventional models may become very complicated, and may impose many constraints that could reduce the scope of the analysis.

Some new tools emerging from the field of Artificial Intelligence (AI), particularly Neural Network (NN) models, are recently coming to the fore and gaining popularity due to their flexibility and their ability to solve complex problems. A NN is a computational paradigm that tries to imitate the functioning of the human brain in order to solve a wide range of problems. Like the human brain, NN models are able to “learn”

from a set of examples and generalise from them to find the right solution to “new” problems from the same category as the examples presented. Unlike many conventional statistical methods, NNs do not require any kind of a priori hypothesis about the underlying model structure. A brief overview on NNs is given in Section 2; for a more complete introduction on NNs we refer, among others, to Sarle, 1997; Cheng et al., 1994; Kuan et al., 1994. The NN approach has been applied and tested against conventional methods on economic problems – mainly in the transportation field – by Reibnegger et al., 1991; Altman et al., 1994; Dougherty et al., 1994; Kuan et al., 1995; Camastra et al., 1997; Himanen et al., 1998; Reggiani et al., 2000.

The purpose of this paper is now to propose a NN approach in order to forecast regional employment developments, with a particular focus on the West German labour market. The paper is therefore organised as follows: after a brief overview of the German labour market and the NN methodology (Sections 2 and 3), Section 4.1 will introduce our case study, addressing West German employment forecasts. Results of the experiments carried out will be given in Sections 4.2 and 4.3. Finally, Section 5 will conclude with some specific remarks and directions for future research.

2. Basic Information about the German Labour Market

To understand the significance of forecasts for the labour markets of Germany and of other countries, it is necessary to understand the situation in the country we are using as an example. Compared with other European countries, Germany has a relatively high rate of people out of work (for an overview of the German labour market see Blien et al., 2002). In 2000 the unemployment rate was just below the average for EU countries, according to data from the Labour Force Survey. The labour market situation in Germany was marked by high and, until 1997, increasing levels of unemployment. Since then a slight improvement has taken place. In the year 2000 the annual average number of people registered as unemployed was approximately 3.9 million (1997: 4.4 million). This corresponds to an unemployment rate of 9.6% (national definition: registered unemployed as a proportion of the civilian labour force). The rate was 11.4% in 1997.

The development of unemployment over the course of time is especially worrying, since even after an economic upswing, unemployment has remained high. During the economic recovery in the late 1980s and early 1990s, unemployment decreased less than employment increased. As a consequence of the persistent employment crisis, a process of selection among the unemployed has taken place and a hard-core group of unemployed people has developed. A large proportion of them are long-term unemployed individuals. In addition to the long-term unemployed, hard-core unemployment is also used to refer to people whose employment is repeatedly interrupted by periods of unemployment.

It must also be pointed out that the German labour market situation is characterised by considerable regional discrepancies, especially between the western and the eastern parts. Long after unification (1990) unemployment remained high in the east: in 2000 the unemployment rate in western Germany was 7.8 % whereas the figure for eastern Germany was 17.4%. Table 1 illustrates the amount of employment covered by social security in western versus eastern Germany.

Table 1: Employment Covered by Social Security in Germany (end of year)
Source: Employment statistics of the Federal Employment Services

	1996	1997	1998	1999	2000
Employment (West)	22126949	22043258	22223461	22694019	23075334
Change in %		-0,38	0,82	2,12	1,68
Employment (East)	5298337	5097548	5143506	5061546	4904773
Change in %		-3,79	0,90	-1,59	-3,10

The unemployment problems are at least partly due to the fact that in Germany conditions are not as favourable to the development of employment as they are in other European countries, for instance, the Netherlands. The table above shows small gains in the western part of Germany during recent years, but even losses in the east. Since the economy began to recover in 1998, employment performance in western Germany has improved. With the onset of the recession this favourable development came to a halt in 2001/2002.

In recent years the German recipe for success – a reliance on technological innovation with a well-trained workforce – has ceased to produce the positive results for the labour market that it produced in the past. According to a new assessment, this might at least partly be due to the fact that German industry has specialised on segments of the

world's product markets that are characterised by a low rate of technical progress, by inelastic demand or by both factors (Appelbaum and Schettkat, 1993; Möller, 2001).

In its regional aspects Germany is characterised by a situation that is unique in the world. The transformation process that a regional economy has undergone – from state socialism to a market economy or to capitalism – has not been done previously in any independent state. The process was facilitated through unification with one of the largest western economies. The territory of the former German Democratic Republic is now part of the Federal Republic of Germany. In the east the process of transformation was pushed through very quickly: western institutions were established there during the early nineties. Wages doubled within a few years, so that their level is now about 75 % of that of the west. Since productivity is lower in the east, at about 65% of that of the west, there are still severe labour market problems in eastern Germany.

Within both parts of the country, marked regional disparities in the labour market are visible. In former West Germany, the southern parts of the country are developing more rapidly than the north. In addition, there are great differences with respect to important indicators that can be measured on a small scale. It is necessary to look at relatively small regional units to see great differences in employment and unemployment.

To counteract the disparities in the regional labour markets, a relatively large amount of spending is budgeted by the political institutions responsible for labour market policy. Every year about 22 billion Euros (approximately 19 billion US\$) is spent on active labour market policy measures, i.e. training and job creation measures, wage subsidies and other schemes. To allocate these funds efficiently, the people who are responsible for the budget continuously demand regional forecasts about the labour market. In many cases, however, these are unavailable, and they are restricted to using statistical information, which is only available after a time lag. The allocation of funds to the various regions and programmes is therefore based on information about the past. To counteract imbalances in the labour market, however, information about the future is needed, which is only available from forecasts. This is a vital political interest, since the amount of money to be distributed, though large, is restricted. It is necessary to create forecasts for small regions since labour market disparities can be found at this level.

The use of forecasting techniques is intended to improve the allocation of the funds of labour market policy to the regions of Germany. The Institute for Employment Research (IAB) is involved in the process of calculating indicators that reflect the current problems in local labour markets. The funds for active labour market policy are

distributed according to a formula developed at the IAB (cf. Blien, 1998). Four indicators concerning the situation and development of the labour market yield the results upon which the monetary distribution is based. One of these indicators is the development of employment in the past. Since this indicator fluctuates considerably over time and between regions, finding an indicator that could represent current and future developments would be a great improvement. If such an indicator were used, the distribution of funds could be oriented towards resolving prospective problems and it would be possible to counteract labour market imbalances more effectively. Therefore, the intention is to calculate one indicator by using the forecast of regional employment.

There is also a scientific motivation for the development of forecasts. A major purpose of research is to identify the structure of causality in a given field of interest. Forecasts are appropriate for testing analyses of this kind. The structure of causality identified in research about the past can be extrapolated to a future situation. If these forecasts are reliable, the researcher can be confident of being able to identify the main influences that are being sought. It is possible to evaluate the results of a forecast, which serves as an additional test of whether the model used adequately represents the causal structure inferred from a theory.

Compared with this situation, analyses of past developments are always in danger of overfitting. Random fluctuations might affect the data and a model might represent these perfectly but show poor results when applied to future situations. The construction of forecasts, however, can help to discriminate between random and systematic variation. The labour market is a particularly interesting field for forecasts, since it is known that there is a lag between developments in product markets and developments in the labour market. This lag can be exploited for forecasts.

Both motivations for inquiry, one related to labour market policy and the other concerned with scientific explanations, are relevant in the present context, which deals with regional forecasts for employment in western Germany for a time period of two years. The purpose of our project was to forecast not global employment, but the development of regional employment relative to the national average.

We intend to use new and innovative methods of forecasting to see whether their results are better than those obtained with a method based on an entropy optimization approach that has recently been developed (cf. Blien and Tassinopoulos, 2001).

3. Neural Network Techniques: A Brief Overview

As mentioned briefly in the first section, NN functioning is inspired by human brain organization, in which calculation is based on the principle of distribution of activity to a high number of simple calculation units, strictly related and working in parallel. The artificial NN is therefore made up of neuron-cells, internal connections between neurons (the so-called weights), and input/output connections with the external world. By recursively modifying a set of weights linking input, output and ‘hidden’ neurons, the NN can ‘learn’ from a set of empirical examples by finding meaningful structures that exist between the variables presented (see, for example, Rumelhart et al., 1986). As a result, since they do not require *a priori* any explicit hypothesis about or tie between the variables under analysis, NNs are very useful when knowledge of a precisely specified statistical model able to explain the phenomenon being examined is lacking.

The NN learning process, which consists of finding the best set of weights, can be formulated in terms of the minimisation of the square error function between actual and desired output such that the underlying relationship/model represented by the training set of examples can be approximated/learned. One of the main problems that could occur in this phase is to get stuck at a local minimum; therefore different techniques have been developed in order to reach global minima and to escape from local minima (see, for example, Fischer, 2001).

A second kind of problem usually encountered when working with NNs is model selection, consisting of the choice of the number of hidden units, as well as of the tuning of the NN parameters. As in statistics, where there is a sort of bias-variance trade-off, it is considered good practice to keep the NN complexity low. However, a NN that is too simple or too inflexible will have a large bias and will not be able to reach a good approximation of the structure underlying the data; no learning will result from this situation. On the other hand, a NN that is too complex with respect to the data to be analysed will probably overfit the data, again causing generalisation problems. In the literature, many different methods of avoiding overfitting have been proposed. The most popular ones are pruning techniques consisting of removing inputs and/or weights, as well as early stopped training methods consisting of training the NN until the error on a further validation data set deteriorates. The last method, in particular, does not require the convergence of the training process.

Despite the fact that these are models in which no modelling hypothesis is needed and no exact function underlying the variables and the data is imposed, NNs can often be related to more conventional statistical tools (cf. Blien and Lindner, 1994). For example, feedforward NNs without any hidden units are basically generalised linear models, while feedforward NNs with one hidden layer are closely related to projection pursuit regression (Sarle, 1997). Furthermore, Schintler et al. (1998) have demonstrated that NNs, in the case of binary choice, are equivalent to a logit model. In this paper a feedforward network with hidden layers is used in the forecasts. As a training procedure, backpropagation is applied.

4. The Case Study: Forecasting Regional Employment in West Germany

4.1 Introduction: the Data Set and the Experiments Carried Out

The NN technique mentioned in the previous section is applied in a case study related to the western German labour market in order to provide short-term forecasts of regional employment. Employment in different target years is modelled. Since the database was available until 1999 and the target years are 2000 and 2001, the horizon of the forecast spans one or two years⁴. The aim of this research project is to forecast employment on the basis of available information from the past about the number of employees, as subdivided into nine economic sectors. The usefulness for forecasting purposes of some additional external variables will also be explored.

The database from the German employment statistics available at the Institute for Employment Research (IAB) is organised as a panel containing data about all people employed in the 327 western German districts, subdivided into 9 economic sectors⁵: primary sector, industry goods, consumer goods, food manufacture, construction, distributive services, household services and services for society. The related time series covers 13 years, from 1987 to 1999. In addition to the data on the number of employees, information about daily wages is available. One additional independent non-geographic characteristic of the districts – the district type – can also be used to improve the results of the models. This qualitative variable is represented by a number ranging from 1 to 9,

and indicates the ‘type of economic region’ the district belongs to (details can be found in Table 2).

Table 2: The economic regions

Source: IAB - Institute for Employment Research, Nuremberg, Germany

Type of district (BfLR/BBR-typology):	Type of region
A. Regions with urban agglomeration	1. Central cities 2. Highly urbanised districts 3. Urbanised districts 4. Rural districts
B. Regions with tendencies towards agglomeration	5. Central cities 6. Highly urbanised districts 7. Rural districts
C. Regions with rural features	8. Urbanised districts 9. Rural districts

Note: BBR is the Bundesanstalt für Bauwesen und Raumordnung, Bonn, whose former name was Bundesforschungsanstalt für Raumordnung und Landeskunde (BfLR).

Since the adoption of the variable ‘number of employees’ expressed in absolute terms might lead to unclear results (the so-called spurious regression results), the experiments carried out in this study have been based on growth rates.

A second problem encountered during the analysis consisted of the choice of the best NN architecture. Usually, in order to find the best architecture for a NN, the data set has to be split into three randomly chosen sub-sets: the training set, validation set and test set. While the training set is used to find the best set of weights, the validation set is used to tune the NN parameters and to find the best architecture. The purpose of the test set is to evaluate the performance of the models proposed. Since the information, in our case, is organised as pooled cross-section and time series data, we do not have much freedom in choosing the three sets; rather than choosing them randomly⁶, we have selected the three sub-sets according to the time periods. In addition, the NN structure of the network is not given *a priori*; many free parameters have to be tuned via a process of trial and error in order to find architecture that is able to offer good results. Because it is not possible to know whether a global minimum of the error function has been reached in the forecasting context, many phases are necessary to reach stable results. As a consequence, we have carried out the experiments in three different phases. In the first phase, we trained the models on the set ranging from 1987/88 to 1996/97 and tested them on the remaining two years’ data set (1997/98 and 1998/99), in order to decide on the best architecture for the NN. Once good architecture was found, the models were re-trained on the set ranging from 1987/88 to 1997/98 and then tested on the remaining

year's data (1998/99). In this second phase of the experiments, the resulting ex-post forecast was used to compare the models proposed, as well as their generalisation features. Finally, in the third phase, the whole data set was used as a training set in order to assess employment forecasts for the year 2000.

The major problem that emerged in the first phase of the analysis, regarding the choice of the NN structure, concerned the choice of the number of training epochs. Our NN results appeared quite sensitive to these parameter changes: when the number of epochs is (relatively) low, the indicators for both test sets (1997/98 and 1998/99) appear to improve. In a second stage, the statistical indicators for the first test set (1997/98) declined in quality, but the statistical indicators for the second test set (1998/99) continued to improve.

We may choose the number of training epochs in three ways. The first option consists of choosing a number of epochs that minimises the first test set (1996/97), while the second option consists of choosing a number of epochs that minimises the second test set (1997/98). We found that the number of epochs necessary to reach the minimum for the second test set is always higher than the number of epochs necessary to reach the minimum for the first test set. The third option corresponds to the model in which the number of epochs chosen falls somewhere between those in the previous two cases. Although no exact rules for making a choice exist, this third method seems to offer the best results in the ex-post forecast. Consequently, in all the models presented here, the number of training epochs was found according to this method.

The adopted models were ultimately investigated and compared with the aim of forecasting future employment on the basis of available information from the past about the number of employees, as subdivided into nine economic sectors. This will be illustrated in the next section.

4.2 Forecasts for the Year 2000: the Models Adopted and their Results

In a previous stage of our research (Longhi et al., 2001) we analysed and evaluated NN models based on the available information about the number of employees, subdivided into eleven economic sectors, in order to predict future employment. In the current stage of research, the analysis focuses on the same variables classified into a different number of sectors (nine instead of eleven).

Because our data set contains a large number of cross-sections, it is not possible to consider the organisation of the information as pooled cross-section and time series

data: this would require too many weights. Information about the time period must therefore be introduced in the models as a new variable. Table 3 compares⁷ the two different ways in which the variable ‘time’ can be introduced in the model: as dummy variables (Model A) or as qualitative⁸ variables (Model B). As we can see by comparing the two test years for each model in Table 3, the model in which time is introduced as dummy variables seems to have a more stable behaviour than the model in which time is defined as qualitative variable. It should be noted that the results of Table 3 still refer to the training phase and should not be used to compare the different models. The comparison among the models is done on the base of the test set’s results of Table 4. We may read a small difference between the two test sets (1997/98 and 1998/99) of Table 3 as an indicator of a more stable NN’s performance.

Because the year 1991 represents a structural break in our series, we have also carried out a set of further experiments in which the information available for the years before 1991 is eliminated. We found that this model did not show an improvement over the previous results.

A second step in our research consisted of verifying whether or not the introduction of information about the economic district (the correspondence of cross-sectional fixed effects in a panel model) was able to improve the performance of Model A. The information about the districts was introduced in the model as a number ranging from 1 (set as minimum value) to 327 (set as maximum value). It was not possible to introduce this kind of information as a qualitative variable, since this would have required too many different symbols. The resulting Model C is also shown in Table 3.

Table 3: Identification of the NN structure for the models forecasting 2000
The indicators refer to the growth rates

	Time: dummy Model A		Time: qualit. Model B		Time: dummy, “Fixed effects” Model C		Time: dummy, Type: qualit. Model D		Time: dummy, Type: dummies Model E	
Test:	‘97- ‘98	‘98- ‘99	‘97- ‘98	‘98- ‘99	‘97- ‘98	‘98- ‘99	‘97- ‘98	‘98- ‘99	‘97- ‘98	‘98- ‘99
ARV	0.921	0.941	1.6155	2.441	0.971	0.957	0.9065	0.914	0.928	0.956
MSE	2.522	2.925	4.413	7.570	2.658	2.973	2.480	2.841	2.541	2.971
	Model AW		Model DW							
Test:	‘97- ‘98	‘98- ‘99	‘97- ‘98	‘98- ‘99						
ARV	0.890	0.959	0.852	0.995						
MSE	2.436	2.979	2.333	3.093						

Note: Table A.1 of the Annex offers a summary of all models proposed

As a third step, we aimed to verify whether using the variable ‘type of economic region’ could improve the results of Model A, which had seemed the best up to that point in the process. Once again, we were able to introduce this latter variable either as a single qualitative variable (Model D), or as a number of dummies (Model E), both shown in Table 3. We also tried introducing this variable in Model B, but were unable to obtain good results.

As mentioned above, an additional external variable that could improve the performance of our models viz. daily wages is available. Since all the previously proposed models seem to produce quite similar results, we introduced this new variable in all of them. The introduction of the variable ‘daily wages’ in Model B and Model E yielded disappointing results, hence we removed it. We will now show the results from the combined Model AW (in which W refers to wage) in comparison with those from Model A, as well as the results from Model DW in comparison with those from Model D. It should be noted that the two ‘original’ Models A and D show homogeneous values in the statistical indicators under analysis, while the two models utilising information about wage (Models AW and DW) display rather different values among the statistical indicators of the first and the second test sets.

In the second phase of our analysis (see the previous sections) we re-trained the models on the set ranging from 1987/88 to 1997/98 and tested them on the remaining year’s data (1998/99). In this way we could use the resulting ex-post forecast to compare the proposed models, as well as to analyse their generalisation characteristics. After calculating the growth rates predicted by the models, we re-converted⁹ them into the original scale measure (total number of employees). In Table 4 the statistical indicators inherent to each model are shown.

Table 4: Comparative analysis among the adopted models (test year: 1998/99)
The indicators refer to the total number of employees.

TRAIN SET	Model A	Model B	Model C	Model D	Model E	Model AW	Model DW
MSE	1636633.3	3921973	1740527	1507558	1627749	1707834	1503297
MAE	732.57958	1166.9	736.6439	696.9285	722.95	746.5229	696.8544
MAPE	1.22584909	1.872259	1.225942	1.177418	1.20813	1.253605	1.182383
TEST SET	Model A	Model B	Model C	Model D	Model E	Model AW	Model DW
MSE	1890058.3	1743597	2728751	2269495	15802724	2807580	3398917
MAE	709.146915	704.5804	829.5734	772.5858	2496.202	808.3172	950.1482
MAPE	1.16066761	1.155527	1.278273	1.227717	5.359818	1.210859	1.43503

From Table 4 we can see that Model D and Model DW are the ones that best fit the training data. When the results are “generalised” to the test set, both Models D and DW appear to produce poorer results than Models A and B. Model B, and to a lesser extent Model A, seem to perform better with the test set than with the training set. This could mean that these two models are quite robust in generalising the data set presented during the training phase. The poor performance of Model D and Model DW in generalising the training set might result from several causes. Firstly, this could simply be a problem inherent in the training phase: the results of training phases are very unstable, and can vary drastically if the number of training epochs is changed. Secondly, the additional variables introduced in Models D and DW might not be able to explain the behaviour of the dependent variable (employment). From Table 4, therefore, we can conclude that Models A and B seem to be the best ones.

Figure 2 shows¹⁰, in a graphical form, that at a national level Model E is the one that generates the highest forecasting error. Model A and Model B, the two models that do not use any kind of information apart from the sectoral composition of employment, seem to offer the best ex-post forecasts for 1999 (the test set). Furthermore, Model A tends to mirror the path of the observed values (the training set) better than Model B, which fails to identify the point of change (1993 instead of 1992). Models D and DW offer very similar forecasts, while Models A and AW produce quite different values. In particular, Model A seems to perform better than Model AW. The subsequent Figure 3 compares the forecasts for the year 2000 offered by all models proposed so far. If we neglect the errors generated in the ex-post forecast of 1999, we see that all models predict similar growth rates from 1999 to 2000.

We can, therefore, conclude that the basic Model A seems to be the best one. A second conclusion is that the information about wages seems not to be so relevant in improving the performance of the adapted models. This result may be due to the NN set of weights linking inputs and output. Since none of the weights connecting the groups of inputs¹¹ with the output can be dropped, the NN may not be able to distinguish between the natures of the inputs themselves. To verify whether or not this is the case, we have tried to introduce wages into the NN models in which employment is not subdivided into economic sectors by obtaining a model that may look like an autoregressive model of order one (AR(1) model). The results attained were quite poor compared to those given by the models presented above.

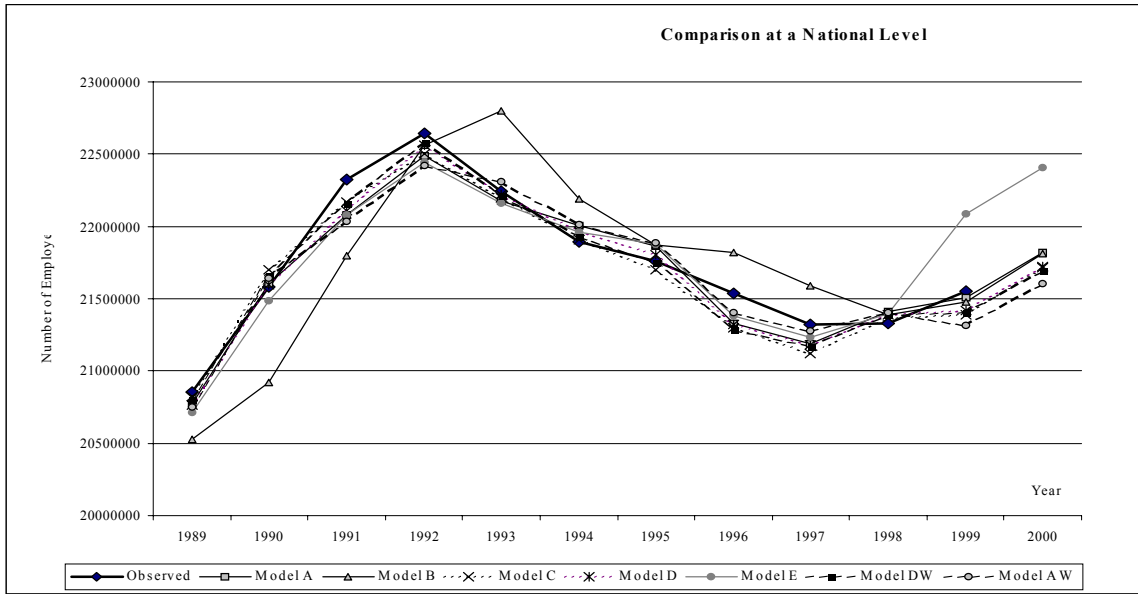


Figure 2: Training and test performance of all models at a national level

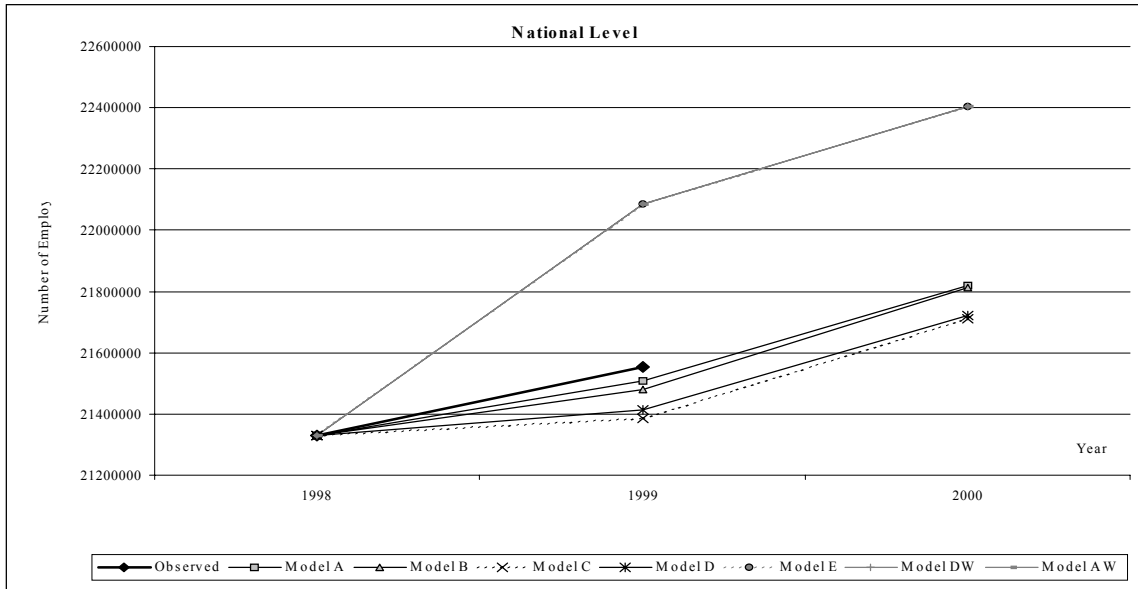


Figure 3: Test performance of all models trained until 1998 and tested ex-post on 1999, at a national level

Further graphical analyses of the models' behaviour in each district gave us additional evidence that Model E tends to systematically overestimate the change from 1998 to 1999. The forecast for the number of employees in the year 2000 is, of course, affected by this error.

Finally, to offer predictions for the year 2000, we trained the NNs on the whole data set available (years from 1987 until 1999). Figure 4 shows the national forecasts offered by all the models, and how this allows them to be grouped into two clusters. The highest forecasts are offered by Models A, B and DW, while the lowest are offered by Models

C, E, D and AW. The distance between the highest and the lowest forecasts at a national level is of the order of magnitude of about 130,000 employees.

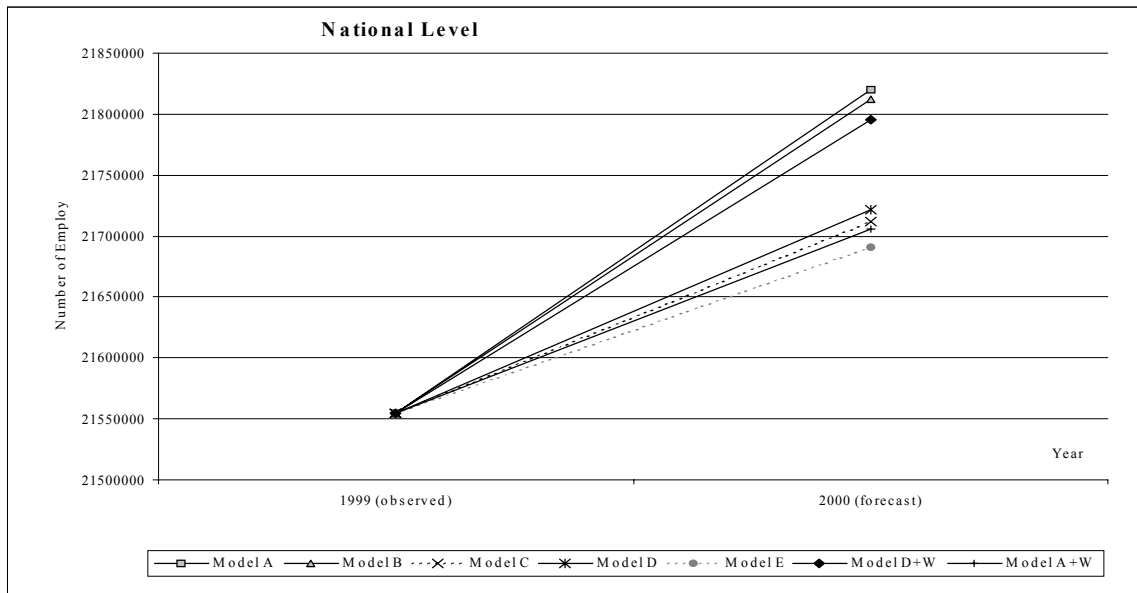


Figure 4: National forecasts for the year 2000 emerging from the adopted models trained on the whole data set

4.3 Forecasts for the Year 2001: the Models Adopted and Their Results

The models proposed in the previous section were not able to create forecasts for the year $(t + 2)$, because in that case they need inputs for the year $(t + 1)$. Consequently, to predict the employment situation for the year 2001, we had to change the previous approach slightly: we based our analysis on the growth rates between t and $(t + 2)$ instead of on the growth rates between t and $(t + 1)$; all other aspects of the analysis remained unvaried¹². Table 5 shows the statistical indicators we estimated for some of the models in order to predict the employment values for the year 2001 (the test set refers to the year 1999).

Table 5: Identification of the NN structure for the models forecasting 2001
The indicators refer to growth rates. (Test set: 1997/99)

	Model A-2001	Model B-2001	Model AW-2001	Model D-2001	Model DW-2001
ARV	0.962621	1.491785	0.91135	0.85706	0.840135
MSE	7.551256	11.68939	7.1514	6.722227	6.591309

By comparing Models A-2001 and D-2001 with Models AW-2001 and DW-2001, we can see that in this case the introduction of the wage variable is useful for improving the performance of all models adopted (only for the Model B-2001 was this not the case).

Furthermore, the somewhat more ‘complex’ model, in which the variable of the type of economic region is also included (Models D-2001 and DW-2001), seems to perform better than all the other models. This result is reflected in Table 6, which shows that Models D-2001 and DW-2001 perform well. These models fit the training data well and appear to be the best ones for generalising the examples presented.

Finally, Figure 5 shows that the forecasts offered by all models produce similar values, with the exception of Model B-2001, which displays the worst performance on both the training and test set. Therefore, we can be fairly confident on the robustness of the other relative forecasts.

Table 6: Comparative analysis of the models proposed to forecast the year 2001 (test set: 1997/99)

The indicators refer to the total number of employees.

TRAIN SET	Model A-2001	Model B-2001	Model AW-2001	Model D-2001	Model DW-2001
MSE	5562937	16535292	6466852	5870054	5507190
MAE	1323.555	2178.409	1436.935	1340.144	1336.435
MAPE	2.18174	3.393514	2.359092	2.179108	2.183237
TEST SET	Model A-2001	Model B-2001	Model AW-2001	Model D-2001	Model DW-2001
MSE	4151746	8398732	3679887	3322544	3356230
MAE	1112.172	1663.389	1054.36	1011.711	1003.07
MAPE	1.8977	2.598186	1.832554	1.782615	1.757255

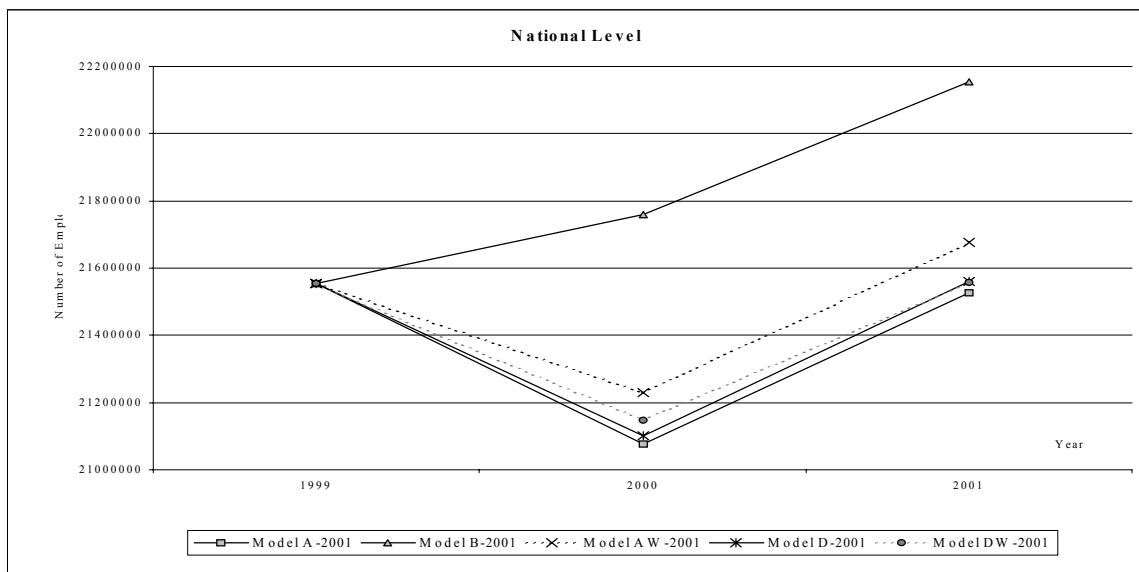


Figure 5: National forecasts for the year 2001 offered by the five models proposed

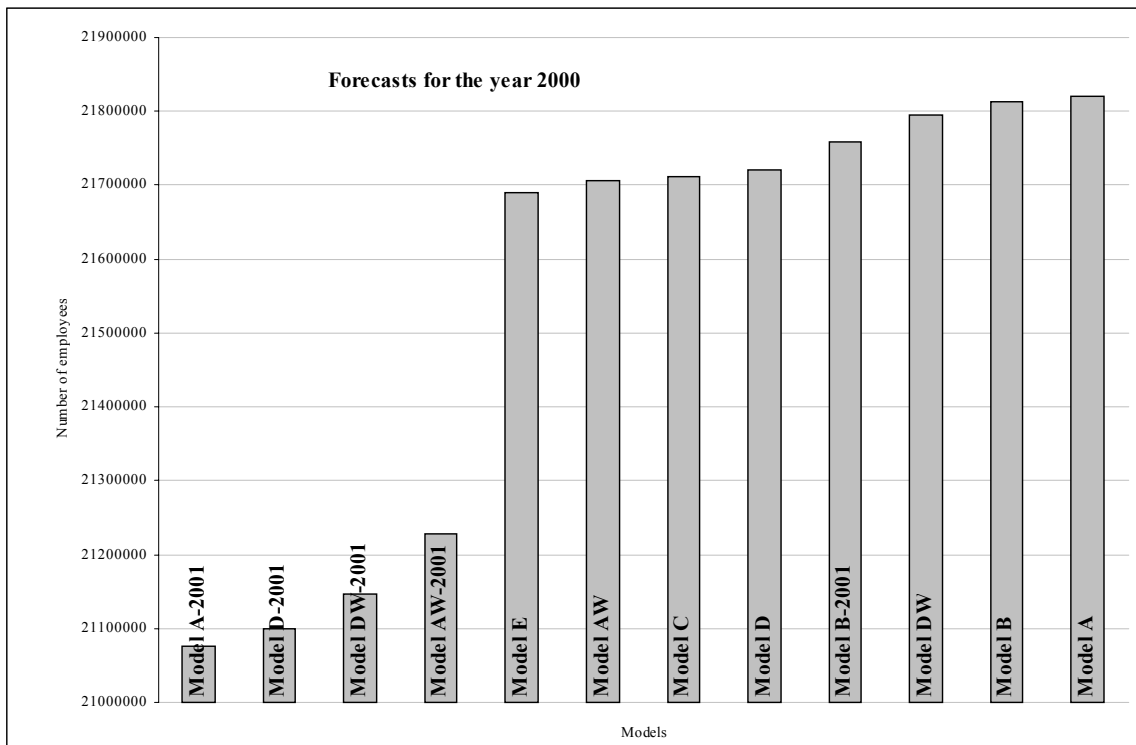


Figure 6: Comparison of the national forecasts for the year 2000 offered by each of the proposed models

Since the models above are able to offer forecasts for both the years 2000 and 2001, the last step in our analysis consists of comparing the forecasts for the year 2000 offered by the two classes of models presented (viz. the models specifically designed to forecast employment in the year 2000 – shown in the previous section – and the models designed to offer a forecast for the year 2001). Figure 6 orders the forecasts of the number of employees produced by each of the proposed models, clustering them into two groups. In the first group we find all the models tuned to make forecasts with a lag of two years. The only exception is Model B-2001, which shows value forecasts that are more homogeneous with the second cluster. In the second group we find all the remaining models. The difference between the result from Model A and that from Model A-2001 is 745,000 employees.

We can therefore conclude that Model B-2001 seems rather satisfactory, although the indicators in Table 8 pointed out that different models might also be good ones. In this context, further experiments, based on multidimensional approaches like Multicriteria Analysis (MCA), could be useful in choosing the ‘most suitable’ model or models.

5. Concluding Remarks

In the present paper we have illustrated the various results of an investigation into the dynamics of regional employment in West Germany at a district level, using a neural network (NN) methodology. The analysis emphasised the ability of NNs to capture the data structures – at least in the training and test phases – and hence to ‘extrapolate’ useful information within a spatial context. However, further remarks and observations are in order here.

- i. The assumption that the past evolution of the sectoral composition of employment might explain an important part of the future development is, undoubtedly, an issue which needs to be reflected upon more thoroughly, especially with regard to the socio-economic political factors underlying employment dynamics and their trajectory.
- ii. As a consequence, more information concerning the economic environment of the districts is useful, since the set of available data does not allow us to model any kind of ‘explicative’ or ‘causal’ relationship in the dynamic path of the employment variable. The reliability of our estimations would certainly benefit from the inclusion of additional economic information.
- iii. The problem of moderately informative data is particularly important for the NN model, which does not impose any hypothesis upon the ‘functional’ relationships under analysis.
- iv. Finally, as already mentioned, recent technological advances in the last decade, such as the diffusion and adoption of the modern Information and Communication Technology (ICT), as well as logistical improvements in transportation systems, have significantly widened the size of the “regional” labour market areas. Therefore, the hypothesis that the district level data is too disaggregated to allow patterns to be inferred from it has to be considered as well.

In conclusion, an integration of the data and variables under analysis with additional information concerning several economic variables may offer a better possibility for modelling behavioural and dynamic relationships concerning employment. This is especially true in light of recent globalisation phenomena and socio-economic-political dynamics. In this context, recently created tools for analysis, like the Self-Organising

Criticality (SOC) approach, which attempts to investigate the critical conditions leading to significant transformations of the dynamic system's behaviour, could offer interesting insights into spatio-temporal employment patterns and development.

Acknowledgements

The authors wish to thank Erich Maierhofer for his help with the data and for his valuable suggestions and comments on the research reported in this paper.

¹ According to its website, the European Union (EU) spends almost the 35% of its budget on regional policy initiatives. Various structural funds have been established with the aim of reducing gaps in development between the regions of the EU, as well as disparities in the standards of living for their inhabitants. The fund's contributions amounted to 8 billion Euros per year in 1989, and rose up to 32 billion per year in 1999. In the period 2000-2006, 28 billion per year (at the prices of 1999) should be spent for the same purpose (Source: EU web site. URL: http://europa.eu.int/comm/regional_policy/).

² The OECD source is: www.oecd.org (database on labour market programmes); the DICE source is: www.cesifo.de (database on labour markets).

³ See, for example, Janssens et al. (1998) for Belgium; Baltagi et al. (1998), Buettner (1999) and Wagner (1994) for Germany; Groot et al. (1992) for the Netherlands; Kennedy et al. (2000) for Australia; and Morrison et al. (2000) for New Zealand. Sato (2000), instead, tries to place Blanchflower and Oswald's empirical findings into a more theoretical framework.

⁴ In particular, we have adapted two kinds of models. One tuned to offer forecasts with a time span of one year in order to make predictions for the year 2000; the other tuned to offer forecasts with a time span of two years in order to make predictions for the year 2001.

⁵ The original information was subdivided into almost 30 economic sectors: for our analysis, however, we have aggregated these sectors into the smaller number of 9 groups.

⁶ Fischer et al. (2000) and Fischer (2001b), for example, have demonstrated that the common practice splitting the data set into training, validation and test sets might produce inaccurate conclusions.

⁷ The models are compared using the following statistical indicators:

Average Relative Variance: $ARV = [\sum_i (y_i - y_i^f)^2] / [\sum_i (y_i - y^a)^2]$

Mean Absolute Error: $MAE = 1/N * [\sum_i |y_i - y_i^f|] * 100$

Mean Square Error: $MSE = 1/N * [\sum_i (y_i - y_i^f)^2]$

Mean Absolute Percentage Error: $MAPE = 1/N * [\sum_i |y_i - y_i^f| * 100 / y_i]$

Where: y_i is the observed value (target); y_i^f is the forecast of the model adopted (NN); y^a is the average of the observed values; N is the number of observations/examples.

The common interpretation of these indicators is that the estimation is better the closer the value to zero.

⁸ The NN software we adopted allows qualitative variables to be introduced in the data set as inputs, which can also be defined in a non-numeric format.

⁹ These values were calculated at a district level by applying the growth rate given as output by the models to the observed total number of employees in that district in the previous year.

¹⁰ Note that these figures compare the performance of the models at a national level and could therefore hide the models' real performance at a district level.

¹¹ The information identifying the year to which each example belongs (the dummies) is different in nature from the information about the sectoral growth rate of employment and the growth rate of average wages.

¹² This is only the fastest way of making such forecasts for 2001. More sophisticated approaches are left for future researches.

References

- Altman E. I., Marco G., Varetto F. (1994) Corporate Distress Diagnosis: Comparison using Linear Discriminant Analysis and Neural Network (The Italian Experience), *Journal of Banking and Finance*, 18, pp. 505-529.
- Appelbaum, E.; Schettkat, R. (1993) Employment Developments in Industrialized Economies: Explaining Common and Diverging Trends, Discussion Paper FS I 93- 313, Wissenschaftszentrum Berlin für Sozialforschung, Berlin.
- Baltagi B. H., Blien U. (1998) The German Wage Curve: Evidence from the IAB Employment Sample, *Economics Letters*, 61, pp. 135-142.
- Baumann J. H., Fischer M. M., Schubert U. (1983) A Multiregional Labour Supply Model for Austria: the Effects of Different Regionalisations in Multiregional Labour Market Modelling, *Papers of the Regional Science Association*, 52-1983, pp. 53-83, re-printed in Fischer M. M., Button K., Nijkamp P. (eds.), *Regional Housing and Labour Markets*, Edward Elgar Publishing, Cheltenham, UK, 1996, pp. 366-396.
- Blanchard O. J., Katz L. F. (1992) Regional Evolutions, *Brookings Papers on Economic Activity*, 1, pp. 1-75.
- Blanchflower D. G., Oswald A. J. (1990) The Wage Curve, *Scandinavian Journal of Economics*, 92, pp. 215-235.
- Blanchflower D. G., Oswald A. J. (1994) *The Wage Curve*, Cambridge, MIT Press.
- Blanchflower D. G., Oswald A. J. (1995) An Introduction to the Wage Curve, *Journal of Economic Perspectives*, 9 - 3, pp. 153-167.
- Blien, U. (1998) Die regionale Mittelverteilung für die aktive Arbeitsmarktpolitik: Politikberatung des IAB für Zwecke der Maßnahmesteuerung, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 31/4, pp 674-689.
- Blien, U., Lindner, H.-G. (1994) Applications of Artificial Neural Networks, in Faulbaum, F. (1994) (ed.) *SoftStat '93. Advances in Statistical Software 4*, Stuttgart etc.: Gustav Fischer, pp. 463-470.
- Blien, U., Tassinopoulos, A. (2001) Forecasting Regional Employment with the ENTROP Method, *Regional Studies* 35/2, pp. 113-124.
- Blien, U., Walwei, U., Werner, H. (2002) Labour Market Policy in Germany. Job Placement, Unemployment Insurance and Active Labour Market Policy in Germany, IAB Topics (forthcoming).
- Buettner T. (1999) The Effect of Unemployment, Aggregate Wages, and Spatial Contiguity on Local Wages: an Investigation with German District Level Data, *Papers on Regional Science*, 78, issue 1, pp. 47-67.
- Bunzel H., Christensen B. J., Jensen P., Kiefer N. M., Mortensen D. T. (eds) (2000) *Panel Data and Structural Labour Market Models*, North Holland.
- Camagra F., A. M. Colla (1997) Short Term Load Forecasting based on Correlation Dimension Estimation and Neural Nets, *Lecture Notes in Computer Science, 1327 – Artificial Neural Networks ICANN 1997*, pp. 1035-1040.
- Cheng B., D. M. Titterton (1994) Neural Networks: A Review from a Statistical Perspective, *Statistical Science*, 9 issue 1, pp. 2-30.
- Chiarini B., Piselli P. (1997) Wage Settings, Wage Curve and Phillips Curve: the Italian Evidence, *Scottish Journal of Political Economy*, 44, 5, pp. 545-565.
- Decressin J., Fatás A. (1994) Regional Labour Market Dynamics in Europe, Discussion Paper no. 1085, Centre for Economic Policy Research, London.
- Dougherty M. S., M. R. Cobbett (1994) *Design and Evaluation of the Neural Network based Short Term Traffic Forecasting System*, Central European Road Data Information Exchange Network, drive II, PN: V2044.
- Fischer M. M. (2001) Neural Spatial Interaction Models, in Fischer M. M., Leung Y. (eds.), *GeoComputational Modelling*, Springer, 2001, pp. 195-219.
- Fischer M. M., Nijkamp P. (1987) Spatial Labour Markets Analysis: Relevance and Scope, in Fischer M. M., Nijkamp P. (eds.) *Regional Labour Markets*, North Holland, 1987, pp. 1-33.
- Fischer M.M., Nijkamp, P. (1991) *Major issues in regional labour market analysis*, in Boyce D., Nijkamp P. and Shefer D. (eds.), *Regional Science-Retrospect and Prospect of Regional Science*, pp. 43-78. Springer, Heidelberg.
- Fischer M. M., Reismann M. (2000) Evaluating Neural Spatial Interaction Modelling by Bootstrapping, Paper Presented at the 6th World Congress of the Regional Science Association International, Lugano, Switzerland.
- Gleave D. (1987) Dynamics in Spatial Variations in Unemployment, in Fischer M. M., Nijkamp P. (eds.), *Regional Labour Markets*, North Holland, 1987, pp. 269-288.

- Groot W., Mekkelholt E., Oosterbeek H. (1992) Further Evidence on the Wage Curve, *Economics Letters*, 38, pp. 355-359.
- Himanen V., P. Nijkamp, A. Reggiani (ed) (1998) *Neural Networks in Transport Application*, Ashgate, Aldershot.
- Janssens S., Konings J. (1998) One more Wage Curve: the case of Belgium, *Economics Letters*, 60, pp. 223-227.
- Johansen K. (1997) The Wage Curve: Convexity, Kinks and Composition Effects, *Applied Economics*, 29, pp. 71-78.
- Kennedy S., Borland J. (2000) A Wage Curve for Australia?, *Oxford Economic Papers*, 52, pp. 774-803.
- Krugman P. (1991) *Geography and Trade*, MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Krugman P. (1998a) Space, the Final Frontier, *Journal of Economic Perspectives*, 12, pp. 161-174.
- Krugman P. (1998b) What's New About the New Economic Geography?, *Oxford Review of Economic Policy*, 14 - 2, pp. 7-17.
- Kuan C.-M., T. Liu (1995) Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks, *Journal of Applied Econometrics*, 10 - 4, pp. 347-364.
- Kuan C.-M., H. White (1994) ANNs: an Econometric Perspective, *Econometric Reviews*, 13, pp. 1-91.
- Longhi S., Nijkamp P., Reggiani A., Maierhofer E. (2001) A Neural Network Approach to Forecast the Development of Regional Labour Market in West Germany, Paper Presented to the 41st Conference of the European Regional Science Association, Zagreb, Croatia.
- Morrison P. S., Papps K. L., Poot J. (2000) Local Labour Markets, Migration and Wage Determination: Theory and Evidence for the Wage Curve in New Zealand, Paper Presented at the NZESG Workshop, Victoria University of Wellington, New Zealand.
- Möller, J. (2001): "Income and Price Elasticities in Different Sectors of the Economy. An Analysis of Structural Change for Germany, the UK and the US", in Ten Raa, T., Schettkat, R. (2001) (eds.) *The Growth of Service Industries. The Paradox of Exploding Costs and Persistent Demand*, Cheltenham, Northampton: Edward Elgar.
- Nijkamp P., Poot J. (2001) The Last Word on the Wage Curve? A Meta-analytical Assessment, Discussion Paper, Tinbergen Institute, Amsterdam.
- Nijkamp P., Rietveld P., Rima A. (1984) Information Content of Data from Different Spatial Aggregation Levels, in Nijkamp P., Rietveld P. (eds.), *Information Systems for Integrated Regional Planning*, North Holland, 1984, pp. 215-228.
- OECD (2000) *Employment Outlook*, Organization for Economic Co-operation and Development, June 2000.
- Pissarides C. A. (1985) Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages, *The American Economic Review*, 75, issue 4, pp. 676-690.
- Pissarides C. A. (2000) *Equilibrium Unemployment Theory*, MIT Press.
- Reggiani A., P. Nijkamp, E. Sabella (2000) A Comparative Analysis of the Performance of Evolutionary Algorithms, in Reggiani A. (eds.), *Spatial Economic Science. New Frontiers in Theory and Methodology*, Springer-Verlag, Berlin, 2000, pp. 332-354.
- Reibnegger G., G. Weiss, G. Werner-Felmayer, G. Judmaier, H. Wachter (1991) Neural Networks as a Tool for Utilizing Laboratory Information: Comparison with Linear Discriminant Analysis and with Classification and Regression Trees, *Proceedings of the National Academy of Sciences of the United States of America*, 88 - 24, pp. 11426-11430.
- Rocheteau G. (2001) Equilibrium Unemployment and Wage Formation with Matching Frictions and Worker Moral Hazard, *Labour Economics*, 8, pp. 75-102.
- Rouwendal J., Nijkamp P. (1987) Regional Economic Research on Labour Markets, in Fischer M. M., Nijkamp P. (eds.), *Regional Labour Markets*, North Holland, 1987, pp. 95-115.
- Rumelhart, McClelland (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, Massachusetts.
- Sarle W. S., (ed.) (1997), Neural Network FAQ, Periodic Posting to the Usenet Newsgroup comp.ai.neural-nets, URL: <ftp://ftp.sas.com/pub/neural/FAQ.html>.
- Sato Y. (2000) Search Theory and the Wage Curve, *Economics Letters*, 66, pp. 93-98.
- Schintler L. A., Olurotimi O. (1998) Neural Networks as Adaptive Logit Models, in Himanen V., Nijkamp P., Reggiani A. (eds.), *Neural Networks in Transport Applications*, Ashgate, 1998.
- Shapiro C., Stiglitz J. E. (1984) Equilibrium Unemployment as a Worker Discipline Device, *The American Economic Review*, 74 - 3, pp. 433-444.
- Wagner J. (1994) German Wage Curves, 1979-1990, *Economics Letters*, 44, pp. 307-311.

Annex

Table A.1: Summary of the models proposed

Model A	Basic NN with time defined as dummies
Model B	Basic NN with time defined as qualitative
Model C	NN time as dummies + fixed effects
Model D	NN time as dummies + type as qualitative
Model E	NN time as dummies + type as dummies
Model AW	Basic NN with time defined as dummies + wage
Model DW	NN time as dummies + type as qualitative + wage
Model A-2001	Basic NN with time defined as dummies for the year 2001
Model B-2001	Basic NN with time defined as qualitative for the year 2001
Model AW-2001	Basic NN with time defined as dummies + wage for the year 2001
Model D-2001	NN time as dummies + type as qualitative for the year 2001
Model DW-2001	NN time as dummies + wage + type as qualitative for the year 2001

Model A is a three-layer NN with 21 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 500 epochs to avoid overfitting.

Model B is a three-layer NN with 10 inputs, 5 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 800 epochs to avoid overfitting. The learning rate is set at 0.5.

Model C is a three-layer NN with 22 inputs, 9 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 150 epochs to avoid overfitting.

Model D is a three-layer NN with 22 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 350 epochs to avoid overfitting.

Model E is a three-layer NN with 30 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 350 epochs to avoid overfitting. The main difference between model D and model E is the way in which the qualitative variable “type” is introduced in the models. While model D treats the variable as qualitative information, model E treats it as a number of dummies variables.

Model AW is a three-layer NN with 22 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 200 epochs to avoid overfitting.

Model DW is a three-layer NN with 23 inputs, 9 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 200 epochs to avoid overfitting.

Model A-2001 is a three-layer NN with 20 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 400 epochs to avoid overfitting.

Model B-2001 is a three-layer NN with 10 inputs, 5 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 350 epochs to avoid overfitting.

Model AW-2001 is a three-layer NN with 21 inputs, 9 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 850 epochs to avoid overfitting.

Model D-2001 is a three-layer NN with 21 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 550 epochs to avoid overfitting.

Model DW-2001 is a three-layer NN with 22 inputs, 10 hidden neurones and 1 output. The activation function is a sigmoid, and the learning process was forced to stop after 450 epochs to avoid overfitting.