INTERNATIONAL REGIONAL SCIENCE REVIEW 28, 3: 330–346 (July 2005)

NEURAL NETWORK MODELING AS A TOOL FOR FORECASTING REGIONAL EMPLOYMENT PATTERNS

SIMONETTA LONGHI
Department of Spatial Economics, Free University, Amsterdam, the Netherlands, slonghi@feweb.vu.nl

PETER NIJKAMP
Department of Spatial Economics, Free University, Amsterdam, the Netherlands, pnijkamp@feweb.vu.nl

AURA REGGIANI
Department of Economics, Faculty of Statistics, University of Bologna, Bologna, Italy, reggiani@economia.unibo.it

ERICH MAIERHOFER
Institut fuer Arbeitsmarkt und Berufsforschung (IAB), Nuremberg, Germany, erich.maierhofer@iab.de

This article analyzes artificial neural networks (ANNs) as a method to compute employment forecasts at a regional level. The empirical application is based on employment data collected for 327 West German regions over a period of fourteen years. First, the authors compare ANNs to models commonly used in panel data analysis. Second, they verify, in the case of panel data, whether the common practice of combining forecasts of the computed models is able to produce more reliable forecasts. The technique currently employed by the German authorities to compute such regional employment forecasts is comparable to a simple naïve no-change model. For this reason, ANNs are also compared to this undemanding technique.

Keywords: *regional forecasts; employment; panel data; neural networks*

To be able to allocate public expenditures efficiently among regions, policy makers are always in need of reliable regional labor market forecasts. In this connection, the regional data required to make such forecasts are becoming increasingly available. Yet in many cases, the number of regions for which the forecasts are needed is generally much higher than the number of time periods for which data for each region are available. As a result, econometric techniques that are commonly used in

DOI: 10.1177/0160017605276187

time-series analysis—generally characterized by a high number of observations over time—are not easily generalized and applied to panel data. However, nowadays, econometric techniques especially designed to deal with panel data are becoming easier to implement in standard statistical packages. But because of the constraints imposed for the parameter estimations (see, for example, Baltagi 2001; and Hsiao 2003), such techniques might be not flexible enough to produce reliable short-term forecasts.

Regions tend to specialize in specific industries and economic sectors (Krugman 1991). Each region produces specific ranges of goods and is therefore likely to be affected by regional-specific fluctuations. In other words, different shocks to labor demand, which eventually lead to permanent changes in employment growth, are likely to be region- rather than country-specific (see also Krugman 1998). For example, Blanchard and Katz (1992) for the United States and Decressin and Fatás (1995) for Europe both found that labor market disturbances are asymmetrically distributed across regions. As a result, a panel estimator imposing equal slopes of the coefficients across regions might lead to incorrect forecasts. To take into account such regional heterogeneity, and to obtain reliable regional forecasts, more sophisticated models are needed. The aim of this article is, therefore, to propose artificial neural networks (ANNs) as an alternative technique to produce such regional labor market forecasts.

The success of ANN techniques may be due to their ability to approximate arbitrary functions of the data rather well (Kuan and White 1994). ANNs have often been applied to highly uncertain financial forecasting problems, interesting examples of which are provided by Donaldson and Kamstra (1996), Kuan and Liu (1995), and Franses and van Griensven (1998). In spatial and regional economics, several recent contributions can be found in Nijkamp, Reggiani, and Sabella (2001); Haag (2001), Reggiani, Nijkamp, and Sabella (2000); and Himanen, Nijkamp, and Reggiani (1998). An overview of ANN applications in the broader field of economics can be found in Herbrich et al. (1999).

One of the most interesting properties of ANNs is their ability to work and forecast even on the basis of incomplete, noisy, and fuzzy data. Furthermore, they do not require—a priori—any hypothesis and do not impose any functional form between inputs and output (Reggiani, Nijkamp, and Sabella 2000). For this reason, such forecasting techniques might be very useful in those cases where knowledge of the functional form relating inputs and output is lacking, or when a prior assumption about such a relationship should be avoided.

Of course, ANNs also have drawbacks, one of the most important being their interpretation. Because of the way they are designed, some authors refer to ANNs as “black boxes,” which might approximate the relationship between dependent and explanatory variables rather well but, at the same time, might make such a relationship difficult to interpret from a behavioral/economic point of view. However, when the main focus is on forecasting a certain variable rather than explaining the relationship between that variable and the explanatory ones, such a drawback might be considered a minor side effect.

Further advantages and disadvantages of ANNs compared with other techniques can be fully appreciated only after an empirical application of such techniques. To investigate the characteristics of this method, in this article we apply ANNs to forecast employment developments in 327 regions belonging to the former West Germany. The German government allocates its resources across German regions on the basis of a certain number of labor market indicators, such as the number of persons employed in each region and the regional unemployment rate. However, at present, German institutions do not compute any forecast on regional employment or unemployment. As a result, decisions concerning both actual and future public expenditures are based on present or past—rather than future—regional labor market developments. Econometrically, the model used by the German policy maker is similar to a naïve no-change forecasting model, in which the number of people employed in year t is assumed to be equal to the number of people employed in the previous year ($t - 1$).

We compare ANNs to panel data techniques[1] with respect to many aspects. Such comparison involves an assessment of the forecasts' reliability, as well as a comparison of the flexibility and computational effort required by each technique to produce their final forecasts.

The rest of the article is organized in four sections. We start by introducing the specific forecasting problem and the statistical indicators used to compare the models' performance. We then continue with a brief illustration of the specific ANN method used and of the calibration process needed to obtain the final forecasts. In the empirical part of the article, we introduce the data set and illustrate and compare the results of ANNs with the results of panel models. Furthermore, we compute forecasts obtained by combining the previous models. The last section summarizes and concludes.

Forecasting Regional Employment

The Forecasting Problem

The data available contain employment figures and regional average wages in the West German regions and is structured as a panel of 327 cross-sections and fourteen time periods from 1987 to 2000. More details on the data set used can be found in later sections. The aim of our exercise is to forecast the volume of employment in each region r (with r ranging from 1 to 327) in year t, given the number of people employed in each economic sector s (with s ranging from 1 to 9) at time $t - 1$. Our forecasting problem may therefore be formalized in the following way:

$$E_{rt} = f(E_{1r(t-1)}; E_{2r(t-1)}; \ldots E_{9r(t-1)}; W_{r(t-1)}; \text{other terms}) + e_{rt}, \tag{1}$$

where the dependent variable E_{rt} is the total number of people employed in region r at time t. The explanatory variables are the number of people employed in region r in the 9 economic sectors at time $t-1$ (i.e., $E_{1r(t-1)}$; $E_{2r(t-1)}$; . . . $E_{9r(t-1)}$), and the average wage earned in year $t-1$ by a full-time worker employed in region r ($W_{r(t-1)}$). The other terms are region- or time-specific characteristics that we add to some of the models to have more reliable forecasts. Such explanatory variables are described in more detail in the following sections. e_{rt} is the remaining—white noise—disturbance.

Finally, f represents the functional form through which the set of inputs is combined to approximate the output. The main difference between the models compared in the sections below consists of the way the function f is modeled. In the naïve no-change-forecasting model, for example, (1) might be rewritten as

$$E_{rt} = \Sigma_s E_{sr(t-1)} + \varepsilon_{rt} = E_{r(t-1)} + \varepsilon_{rt}, \tag{2}$$

where the coefficients of $E_{1r(t-1)}$; $E_{2r(t-1)}$; . . . $E_{9r(t-1)}$ are all equal to 1 and the coefficient of $W_{r(t-1)}$ is zero.

Evaluation of the Models' Performance

The models' performance is analyzed on ex-post forecasts for the year 2000 by means of statistical indicators common in the time-series literature (see, e.g., Swanson and White 1997a, 1997b; and Fauvel, Paquet, and Zimmerman 1999). Given the panel structure of our data, however, for each time period t we have R regional forecasts, with R being the total number of regions (327). Therefore, such indicators are not computed on a series of subsequent ex-post forecasts over time, but on one-year ex-post forecasts over regions. As a result, our indicators summarize the forecasts' variability across regions, rather than across time. Thus, the forecasting error is computed as the difference between the total number of employees in region r in the year 2000 (E_{r2000}) and the total number of employees in region r in the year 2000 predicted by the model (E^f_{r2000}). The global error is therefore computed as the sum across regions of (a function of) the forecasting errors.

The statistical indicators we use to compare our models are listed below.

The mean absolute error (MAE) represents a loss function that equally weights small and large errors. We compute the MAE as the arithmetic mean of the absolute error of the forecast: $1/R * [\Sigma_r \mid E_{r2000} - E^f_{r2000} \mid]$. The absolute forecasting error is averaged across regions.

The same forecasting error ($E_{r2000} - E^f_{r2000}$) might have a different impact on regions of different sizes. By rescaling each error on the basis of the size of the region, we obtain the mean absolute percentage error (MAPE) as the second indicator to assess the models' performance. We compute the MAPE as $1/R * [\Sigma_r \mid (E_{r2000} - E^f_{r2000})/E_{r2000} \mid]$.

The mean square error (MSE) represents a loss function that gives bigger weight to large than to small errors. In the calculation of the MSE, the square of the forecasting error is averaged across regions: $1/R * [\Sigma_r \mid E_{r2000} - E^f_{r2000} \mid^2]$.

Finally, we decompose the MSE into its three components: the bias proportion ($BP = [E_{2000} - E^f_{2000}]^2/MSE$); the variance proportion ($VP = [\sigma^f - \sigma]^2/MSE$); and the covariance proportion ($CP = 2\sigma^f\sigma[1 - \rho(E_{2000}E^f_{2000})]/MSE$). In these three formulas, E_{2000} and E^f_{2000} are, respectively, the average across regions of the total number of people employed and of its forecast. σ^f and σ are the standard deviations—computed across regions—of the forecasted and observed values. Finally, ρ is the correlation coefficient between the real and the observed series of values. Clearly, ρ too is computed on cross-sectional—rather than on time-series—data.

While BP measures the distance between the observed and forecast average value, VP indicates how the variability of the forecast diverges from the variability of the sample. CP, instead, measures the unsystematic error of the forecast (Fauvel, Paquet, and Zimmerman 1999).

To be suitable for real empirical applications, a forecasting model needs to beat the no-change-forecasting model. Such model's characteristic can be easily analyzed by means of the U-Theil inequality coefficient (the Theil statistic), which is computed as the ratio between the MSE of each model and the MSE of the no-change model (Granger and Newbold 1986). The proposed model outperforms the no-change model when the U-Theil inequality coefficient is lower than 1 (see, e.g., Fauvel, Paquet, and Zimmerman 1999; and Swanson and White 1997b).

Although there are other methods for comparing the models' performance, the indicators used here, being among the most common, are likely to be appropriate to assess the differences between the models proposed and compared in the following sections.[2]

Proposed Methodology: ANNs

ANN Architecture

ANN models were originally designed to understand and imitate the functioning of the human brain (see Cheng and Titterington 1994). For this reason, their name and terminology are strictly connected to terms commonly used in neuroscience. For a historical review of ANNs, we may refer to Taylor (1997). For a review of ANNs from a statistical perspective, we refer to, among others, Cheng and Titterington (1994) and Kuan and White (1994).

An ANN is made of units—called neurons—representing the inputs, and connections between units—called weights—which may be seen as the parameters to be estimated. In Figure 1, the units are represented as circles and the weights are represented by arrows. For our forecasts, we use a single-hidden-layer[3] feedforward ANN, in which the neurons are organized in three layers: input, output, and one hidden layer only. While the number of units of the input and output layer is

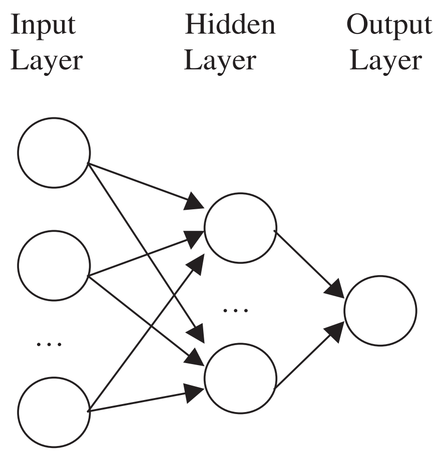

FIGURE 1. Example of Artificial Neural Network (ANN) Structure

defined by the problem investigated, the number of units belonging to the hidden layer (as explained in the following sections) is chosen by the researcher.

Clearly, in this case the number of parameters to be estimated is usually higher than the number of explanatory variables included in the model. For this reason, ANNs might be more powerful forecasting tools than panel data regressions. However, at the same time, it might also be more cumbersome to estimate them.

Labor market forecasts using ANN models have been made by Swanson and White (1997a, 1997b) and Stock and Watson (1998). In these studies, ANNs are found to perform at most slightly better than the other time-series techniques proposed by the authors. We differ from such papers in the data we use, which has a panel rather than a time-series structure.

In more detail, our ANN model may be considered as a special case of (1) (see also Fischer 2001a):

$$E_{rt} = \Psi[\Sigma_j w_j \phi_j(\Sigma_n w_{jn} x_n)] + \varepsilon_{rt}\,, \tag{3}$$

where the dependent variable E_{rt} is the total number of workers employed in region r at time t. The N explanatory variables—input units of our models—are identified by the term x_n, where n ranges from 1 to N. From equation 1, $x_n = (E_{1r(t-1)}; E_{2r(t-1)}; \ldots E_{9r(t-1)}; W_{r(t-1)};$ other terms). The term w_{jn} represents the weights connecting the jth hidden unit with the nth input unit, while w_j are the weights connecting the jth hidden unit with the output unit. The expression $\Sigma_n w_{jn} x_n$ is, therefore, the internal value of the jth hidden unit. Finally, ϕ_j and Ψ are the "transfer functions," which transform the internal value of each unit into the input signal for the connected units. We implement both transfer functions as sigmoid ($\Psi[z] = \phi_j[z] = [1 + e^{-z}]^{-1}$).

We might interpret (3) in the following way: each input unit x_n sends signals ($w_{jn} x_n$) to each unit j belonging to the succeeding layer (the hidden layer). Each unit j of the hidden layer computes the sum of all signals coming from each input unit x_n, and then amplifies or attenuates them via the transfer function (ϕ_j): $y_j = \phi_j(\Sigma_n w_{jn} x_n)$. This value ($y_j$) is then sent as an input signal to the output unit. The output unit computes the sum of all signals originating from the hidden units, and amplifies or attenuates them via the transfer function, thus obtaining the output $E_{rt} = \Psi(\Sigma_j w_j y_j)$. Since, in a feedforward ANN, each layer is only connected with the preceding and succeeding layer, in our simple model both input and output units are only connected with the hidden units, these being connected with both input and output units, but not with other hidden units.

Some important problems have to be addressed here. They concern the choice of the number of hidden units, and the computation of the weights. We start with the second of these problems.

Computation of the Weights

To compute the ANN weights, we start from an initial—randomly chosen—set of weights, calculate the preliminary output, and then modify the weights by means of a recursive algorithm called backpropagation.[4] The backpropagation algorithm consists of two steps: in the first step, the error is back propagated through the network; while in the second step, the weights are modified in the following way.

The error term of the output unit is computed as a function of the difference between the desired and the actual output ($e = g[E^f_{rt} - E_{rt}]$). On the basis of this error, the new weights connecting the jth hidden unit with the output unit (w_j) are computed as a function of the previous weights (w'_j), the weights just before the previous weights (w''_j), and the error (e): $w_j = g'(w'_j; LR{*}e; M{*}[w'_j - w''_j])$, where LR (learning rate) is a parameter belonging to the interval (0; 1] that determines the proportion of the error that is used to compute the new weights. In our exercise, LR is initially set to a high value and then reduced accordingly, when the root of the MSE[5] decreases. M (momentum) is another parameter belonging to the interval (0; 1] that allows for a change in weights to persist for a certain number of adjustment cycles. For simplicity, we set the value of the parameter M to zero, since our results seem to appear quite robust to changes in the value of this parameter. A more thorough analysis of this issue is, however, left for future more specific researches.

The weights connecting the nth input unit with the jth hidden unit are modified in a similar way. The error term of the jth hidden unit is computed as a function of two terms: the first is the desired output, and the second is the difference between the error of the output unit and the old weights connecting the output with the jth hidden unit ($e_j = g[e - w'_j; E_{rt}]$). On the basis of this error, the new weight connecting the jth hidden unit with the nth input unit is then computed as a function of the previous weights (w'_{jn}), the weights just before the previous weights (w''_{jn}), and the error (e_j): $w_{jn} = g''(w'_{jn}; LR{*}e_j; M{*}[w'_{jn} - w''_{jn}])$.

We call one iteration of the backpropagation algorithm an "epoch." The process automatically stops when, for each example belonging to the training set, the error is smaller than a certain percentage τ, or when the algorithm has reached a certain number of iterations (early stopping, see below).

Choice of the Hidden Units

The number of hidden units is a fundamental characteristic of the ANN architecture. An inappropriate choice of the ANN architecture, as well as an inadequate computation of the weights (learning procedure), can cause the failure of the ANN in generalizing the pattern of examples presented. A network that is too complicated might overfit the data, leading to a poor performance. On the other hand, a network that is too simple might not "learn" at all, again causing poor network performance (Fischer 2001b).

We try to avoid overfitting of the network using the technique of early stopping.[6] This technique aims at avoiding too many iterations of the backpropagation algorithm and therefore reduces the possibility that the model will have a poor forecasting performance. The generalizing abilities of the network should gradually improve when the root MSE of the validation data decreases (Church and Curram 1996). However, in our models this minimum is reached after a number of iterations, which is different depending on the number of hidden units of the network.

In our exercise, we choose the number of hidden units via a process of trial and error, after comparing a large number of architectures with numbers of hidden units ranging from 0 to 20. We compute the root MSE of each network every 100 epochs and keep details of the characteristics of the network reaching the overall minimum root MSE[7] of the validation set.

As we have already mentioned, the process of weight estimation starts with a randomly chosen initial set of weights. This introduces uncertainty and makes the comparison among different ANN architectures and models more difficult. To avoid such uncertainty, we start each model with the same set of weights. This choice might reduce the flexibility of our models and, under certain circumstances, might lead to suboptimal results. However, always using the same initial set of weights makes the choice of the architecture, as well as the replication of our results, much easier. Sensitivity analysis on the effects of changes in the initial set of weights is left for future research.

Empirical Application

The Data Set

The data used in this article is part of a bigger database gathered by the German Institute for Employment Research. The information for this database is collected from individual firms and contains micro-data about all people employed in West

TABLE 1. Aggregation of West German Regions in Nine Types of Regions

Group	*Type*	*Number of Districts*
A. Regions with urban agglomeration	1. Central cities	39
	2. Highly urbanized districts	42
	3. Urbanized district	23
	4. Rural districts	14
B. Regions with tendencies toward agglomeration	5. Central cities	21
	6. Highly urbanized districts	61
	7. Rural districts	37
C. Regions with rural features	8. Urbanized districts	43
	9. Rural districts	47

Germany who are covered by the social insurance system. Since such information was originally collected for the administrative purposes of the social security system, the measurement errors affecting our data are probably rather low and not systematic. For more information on the Institut für Arbeitsmarkt und Berufsforschung (IAB) database, we refer to Blien and Tassinopoulos (2001).

The data set available for our forecasting exercise does not contain individual data, but only information about labor market aggregates at the regional level, and is structured as a panel of 327 West German regions covering a period of fourteen years, from 1987 to 2000. Because of its location in the east, the region of Berlin is excluded from the data set. The variables available are the number of full-time workers employed each year on June 30, classified in nine economic sectors.[8] Mean regional daily wages earned by such full-time workers are available as well.

To group regions that might have similar labor market behavior, we also adopted the BfLR/BBR (Bundesforschungsanstalt für Raumordnung und Landeskunde/ Bundesanstalt für Bauwesen und Raumordnung, Bonn) definition of "type of economic region." This classification, which is represented by an index ranging from 1 to 9 (see Table 1), is computed according to the size of population and to the centrality of the location of each region (see Bellmann and Blien 2001).

ANN FORECASTS

In this section, we present and compare a number of ANN models aiming at forecasting the growth rate[9] of total regional employment at time $t(E_{rt})$. Each model differs from the others in terms of the inputs used.

Regional growth rates of employment in the nine economic sectors ($E_{1r(t-1)}$; $E_{2r(t-1)}$; . . . $E_{9r(t-1)}$), and regional growth rates of average daily wages ($W_{r(t-1)}$) are inputs used in all the models proposed below. We now introduce the inputs that we called other terms in equation 1.

When working with panel data, we are used to modeling regional-specific unobserved characteristics by means of regional dummies and time-specific unobserved characteristics by means of time dummies. We try to take into account the effect of specific regional and time characteristics in our ANNs by using the same modeling strategy.

Modeling regional-specific characteristics as regional dummies would require a number of additional explanatory variables equal to the number of regions in our data set—327—and a number of additional parameters, which is too high when compared with the number of observations in the data set. In panel data modeling, this kind of problem is solved by means of the "within transformation" (Hsiao 2003). However, since our ANNs are computed on growth rates, this solution does not seem appropriate for our estimations. We therefore try to model regional effects by means of a discrete variable in the interval (0; 1), which assumes a different value for each region *r*. This variable, which we call "regional effects," is computed as $(1/R)*r$, where r ranges from 1 to 327, and $R = 327$. Since our exercise aims at verifying the forecasting characteristics of ANNs in the case of panel data, we believe it might also be interesting to analyze the effects of this (unorthodox) way of modeling regional-specific characteristics.

As an alternative to this way of modeling regional-specific characteristics, we might assume that regions with a similar degree of urbanization behave in similar ways. In that case, we might only add dummies corresponding to the above-mentioned IAB "type of region" classification. As an alternative to the use of dummy variables, we might model such regional characteristics—that distinguish urban from rural regions—in the same way as we previously computed the "regional effects." The main difference is that r now ranges from 1 to 9, and $R = 9$. Since this kind of input variables ("type of region") might be overlapping with the previous one ("regional effects"), we do not use both of them in the same model.

Similarly, we try to model the effect of time-specific characteristics—"time effects"—either by means of time dummies or by means of a discrete variable computed in the same way as we previously computed the "regional effects." In this case, r ranges from 1 to 13—since the last year (2000) was completely set aside—and $R = 13$. Since this last choice produced better results than the time dummies, we only present the models which include the variable "time effects."

Since we used the technique of "early stopping" to avoid overtraining, no noise has been added to the data. For the same reason, and to keep our exercise simple, we also did not use any pruning technique, such as the ones suggested by Kaashoek and van Dijk (1997) or by Morgan, Curry, and Beynon (2000).

As previously mentioned, we divide the data set into three groups: the training set includes data from 1987/88 to 1996/97, while the validation set includes data for 1997/98, which leads to forecasts for the growth rate of employment between 1998 and 1999. We therefore have 3,270 examples to train the network and 327 examples to be used as the validation set. After having decided the best ANN architecture by means of the training and validation set, we once more reinitialize the set of

TABLE 2. Comparison of the Forecasts for the Year 2000 in the 327 West German Regions

	(1) ANN-RFE	*(2) ANN-TFE*	*(3) ANN-TD*	*(4) ML*	*(5) ML-1-9*	*(6) C-EW*	*(7) C-IM*
MAE	912	866	934	1,403	1,590	1,030	961
MAPE	.01142	.01158	.01209	.02052	.02180	.01436	.01325
RMSE	2,425	2,125	2,354	2,686	3,063	2,300	2,238
MSE	5,882,280	4,517,417	5,542,897	7,214,122	9,380,761	5,288,860	5,009,401
BP	.05780	.02070	.03001	.15535	.18010	.08800	.07083
VP	.57595	.45017	.53443	.46485	.50079	.50314	.50122
CP	.36914	.53214	.43854	.38239	.32163	.41167	.43080
U-Theil	.68278	.59835	.66279	.75613	.86224	.64742	.63009

Note: MAE = mean absolute error; MAPE = mean absolute percentage error; RMSE = root mean squared error; MSE = mean square error; BP = bias proportion; VP = variance proportion; CP = covariance proportion; U-Theil = U-Theil inequality coefficient (the Theil statistic).

weights, shift the data set one year on, and train the ANNs on the data from 1988/89 to 1997/98 to find a new set of weights. This new set of weights is then used on data for 1998/99 (the test set, made up of the remaining 327 examples) to compute forecasts for employment growth between 1999 and 2000. The final employment forecast is obtained by reconverting the growth rate in the employment levels for the year 2000. All model comparisons are therefore based on absolute values of the employment variable rather than on its growth rates.

Since we have continuous data, and since in the validation and test set the variables might assume values that are outside the range of values of the training set, we further rescaled all variables to the interval 0.05-0.95.

In summary, the models we present and compare in this section are[10]

ANN-RFE: The twelve inputs of this model are (1) the growth rates of employment in the nine economic sectors, (2) the growth rate of average regional daily wages, (3) the variable "time effects," and (4) the variable "regional effects." The model has ten hidden units, and the learning algorithm is stopped after six hundred epochs.

ANN-TFE: The twelve inputs of this model are (1) the growth rates of sectoral employment, (2) the growth rate of average daily wages, (3) the variable "time effects," and (4) the variable "type of region." The model has thirteen hidden units, and the learning algorithm is also stopped after six hundred epochs.

ANN-TD: The twenty inputs of this model are (1) the growth rates of sectoral employment, (2) the growth rate of average daily wages, (3) the variable "time effects," and (4) nine dummies—instead of one single variable—identifying the nine types of regions. The model has seventeen hidden units, and the learning algorithm is stopped after twenty-seven hundred epochs.

The results of these three ANN models are shown in the first three columns of Table 2. Each of these models' MAPE is close to 1 percent, the poorest being model

ANN-TD, with an error of 1.2 percent. The systematic error of the forecast, measured by the BP, seems to be acceptably low (Pindyck and Rubinfeld [1998], as cited by Fauvel, Paquet, and Zimmerman [1999], suggested that a BP that is not higher than 0.1 or 0.2 is considered to be “good”). In all three cases, the U-Theil indicator, with values all around 0.60-0.70, confirms that the ANN models we estimated are clearly able to outperform the naïve no-change model. We can therefore conclude that the ANN approach may be worthwhile for the computation of such forecasts. The comparison among the ANN models demonstrates that model ANN-TFE outperforms the other two ANN models in almost all aspects considered.[11]

In the next section, we estimate maximum likelihood models for panel data and compare them with the results obtained from the ANN models presented in this section.

Panel Models' Forecasts

In this section, we forecast regional employment in West Germany using models derived from panel data econometrics. We compute our estimations using the fixed effects, the generalized method of moment, and the maximum likelihood estimators. Since the maximum likelihood method is the one that offers the best regional employment forecasts for the year 2000, we only present the results of this last model.[12]

The model to be estimated may be formalized in the following way (for details on the maximum likelihood random effect estimator [ML], we refer to, among others, Baltagi [2001] and Hsiao [2003]):

$$y_{rt} = \alpha_r + \beta' X_{rt-1} + u_{rt}, \tag{4}$$

where y_{rt} is employment in region r at time t (the term E_{rt} of equation 1) and X_{rt} corresponds to the inputs of the ANNs presented in the previous sections, namely, employment in region r, time $t-1$, and sector s, and the average regional wages in region r at time $t-1$. The components of X_{rt-1} are $E_{1r(t-1)}$; $E_{2r(t-1)}$; . . . $E_{9r(t-1)}$; $W_{r(t-1)}$ of equation 1. Finally, β are the parameters to be estimated. Both regional effects α_r and error term u_{rt} are assumed to be random and normally distributed.

Similar to the ANN models, the parameters of the ML models have been estimated using data from 1987 to 1999. Ex-post forecasts were then computed for the year 2000, and the results compared with those of ANNs.

In a recent paper, Diebold and Kilian (2000) found that in time-series models, pretesting for unit roots is needed for a better selection of the forecasting model. Though our employment data seems to be nonstationary, the model in equation 4 has been estimated using data on levels, first differences, and growth rates. In our case, the model estimated in levels seems to outperform the other two kinds of models in forecasting regional employment in 2000. For this reason, in Table 2 we only show the results of the models computed on levels.

The results of the model computed using data for all 327 regions (ML) are shown in column 4. The results of column 5 are based on the estimation of nine separate ML models (one for each type of region). After the estimation, the results for the nine groups were combined to compute the indicators on all 327 regions. We call this model ML-1-9. According to Table 2, the model that is estimated on the whole data set seems to offer better forecasts for the year 2000 than the one estimated separately for the nine types of regions. One possible explanation for this result is that each of the nine models that are combined to obtain ML-1-9 is computed on a relatively small number of observations. Of the two maximum likelihood models, only the ML model of column 4 seems to clearly outperform the naïve model. Comparing the indicators of the best ANN model—column 2—with the ones of the best ML model—column 4—we can conclude that ANNs seem to outperform the conventional models proposed.

As one referee correctly suggested, the labor market might be evolving over time. In that case, not allowing the parameters to change over time might cause the forecasts to be unreliable. If the effect of these changes in labor market relationships are also significant over a relatively short time period of less than fifteen years, then forecasting techniques such as the Kalman filter or state space methods ought to be chosen when making comparisons with the forecasting performance of ANN models. However, due to the complexity and difficulty of implementing such models, we limited the current analysis to panel data estimators, leaving these more complicated estimations for future research.

The next step of the analysis consists in verifying whether combining forecasts of models with different characteristics will improve our regional forecasts for the year 2000.

Combined Forecasts

In a time-series setting, the combination of forecasts is usually a good alternative to the choice of only one among many competing models. The advantages of combining forecasts in time series are well known: Granger and Newbold (1986) demonstrated that combined forecasts generally tend to outperform the best individual forecasts. Combining forecasts is particularly interesting when the competing models have very different characteristics. However, since in our empirical exercise we compute forecasts over a time period of only one year, these results might not apply to our models. Despite this caveat, in this section we compute pooled models combining forecasts of the ANN and the ML models proposed in the previous sections.

We combine forecasts in a linear way: $E_{r2000} = \Sigma_j k_j E_{jr2000}$. The forecast of total employment in region r in the year 2000 (E_{r2000}) is computed as a weighted average of the forecasts computed by the J models that we combine. E_{jr2000} is therefore total employment in region r in the year 2000 forecast by the jth model (where j ranges from 1 to J).

We compute the weights (k_j) in two different ways. In the first case, we give equal weights to each model, so that $k_j = 1/J$. In the second case, we give to each model a weight which is a function of the inverse of the MSE of the combined models. In this case, $k_j = \mathrm{MSE}_j^{-1}/\Sigma_i \mathrm{MSE}_i^{-1}$.

We present the results of such combined models in columns (6) and (7) of Table 2. While model C-EW combines forecasts using equal weights, model C-IM computes the weighted combination of the single models. In both cases, the models combine only the best ANN—ANN-TFE of column 2—and the best ML—of column 4—model, since we found that combining all seven models of Table 2 gives worse results than only combining the two best ones. Both combined models seem to perform quite well compared with all other models. The model combining weighted forecasts seems to outperform the model combining forecasts using equal weights. Despite the good performance of the combined forecasts, the best model among the ones presented in Table 2 seems to be the ANN model of column 2. This result, which might seem inconsistent with what is commonly observed in empirical time-series analysis (see, e.g., Stock and Watson 1998), might be due to the panel structure of our data and therefore to the way in which our experiment is designed and our indicators are computed. Future research directions should aim at collecting more regional employment data over time to be able to make forecasts over longer time periods and to compare the different models in a more complete way.

Summary and Conclusions

In this article, we estimate and compare a number of different models designed to compute ex-post forecasts of regional employment in 327 West German regions for the year 2000. Currently, labor market forecasts for Germany at such a disaggregated level are computed using a naïve no-change model. The main purpose of our analysis is therefore to assess whether ANNs represent a convenient way to compute forecasts able to outperform the naïve model. To analyze the performance of ANN models in a more complete way, we also compare ANN forecasts to those computed by means of maximum likelihood methods. We also assess the convenience of combining forecasts from different methods.

As expected, almost all models proposed are able to easily outperform the naïve model. Among such models, the ones based on ANN methods seem to offer more accurate forecasts for the year on which we are testing the models. At this stage of the analysis, the conventional models we proposed rely on the assumption of constant parameters over time. Since this assumption might be too strong, future research should also analyze the performance of more complex models which involve time-varying parameters.

In our empirical application, ANN models have proved to be flexible forecasting tools. However, the process of calibration of the network's architecture is also somewhat time-consuming and accompanied by a certain amount of uncertainty as to whether the network has reached a global or local minimum.

Finally, since our data only comprise information for fourteen time periods, it is still quite difficult to find the right balance between the number of time periods to be used for estimation purposes and the number of time periods to be used to test our models. At this stage of the analysis, we test our models on only one year. Future research should aim at using longer time series and testing the models on subsequent forecasts over a period of more than one year, thus rendering the model comparison clearer and more complete.

NOTES

1. The panel structure of our data, comprising a high number of cross-sections and a small number of time periods, poses problems—and requires forecasting methods—that might be quite different from what is common in time-series analysis.

2. Many statistical tests that have been proposed to compare models' performance (such as the test proposed by Diebold and Mariano 1995) in time-series analysis are not straightforward to generalize to a panel data setting. In time-series analysis, the correlation runs only in one direction, from past to current and future observations, but not vice versa. When cross-sections are involved (as in the case of panel data), since each region may affect all other regions involved in the estimation, the correlation usually runs in more directions. This might eventually have an effect on the reference distribution of the tests, with the consequence that the naïve application of such tests to our forecasts would probably lead to misleading results.

3. Since—among others—Kuan and White (1994) demonstrated that a network with only one hidden layer is able to approximate almost any kind of functional form, we only analyze such kinds of artificial neural networks (ANNs).

4. Though the backpropagation is not the most efficient algorithm, it is one of the most simple and easy to implement.

5. We split the data set into three (mutually exclusive) subsets: the first, called the "training set," is used for parameter estimation; the second, called the "validation set," is used in the process of fine tuning of the parameters. The third, called the "test set," is used to assess the performance of the model. The root mean squared error to which we refer here is computed on the training set.

6. Since our backpropagation algorithm does not converge (unless we set the threshold τ at a very high value, around 80 percent), the use of the early stopping technique is necessary.

7. Of course, by comparing only this (relatively) small set of networks we might obtain suboptimal results, since we cannot be sure that the overall best ANN architecture was among those evaluated.

8. These are primary sector, industry goods, consumer goods, food manufacturing, construction, distributive services, financial services, household services, and services for society.

9. Our data might be affected by problems such as measurement errors. When measurement errors are serially correlated over time, the computation of first differences (or growth rates) might reduce the problem. Furthermore, some previous experiments that we made on this data showed that ANN forecasts based on growth rates are more precise than ANN forecasts based on levels. This result might be due to the persistency of regional employment levels.

10. The NN software used in this exercise is Neuralyst, version 1.4, while the conventional models have been estimated using Stata 7.

11. From the previous discussion, it is clear that the results of ANN models depend upon a certain number of parameters (such as the learning rate and the momentum) and choices (e.g., the number of

hidden units and the number of epochs). For this reason, it becomes quite cumbersome to identify the performance of the ANN models to the combined change of such parameters and choices. To statistically test the significance of the difference between the proposed ANN models is a hard task. First of all, to make statistical comparison, we should compare the "average performance" of our models. Pizarro, Guerrero, and Galindo (2000) suggested that such average performance should be computed over the training-test-validation sets that can be randomly drawn from the data. However, since we have panel data, a random choice of the training-test-validation sets seems unwise. Furthermore, statistical tests rely on the assumption of gaussian residuals. The models we estimated do not meet this requirement.

12. Nonlinearity tests in panel data are not a straightforward generalization of the corresponding tests used in time-series analysis. Though we did not compute any formal nonlinearity test, previous analyses based on our data set (see Reggiani, de Graaff, and Nijkamp 2002; and Andergassen, Nijkamp, and Reggiani 2003) indicate the existence of a critical state for the system concerned and hence the nonlinear behavior of our variables.

References

Andergassen, R., P. Nijkamp, and A. Reggiani. 2003. Analysis of regional labour market dynamics: In search of indications for self-organised criticality. *Journal of Geographical Systems* 5: 275-90.

Baltagi, B. H. 2001. *Econometric analysis of panel data.* London: Wiley.

Bellmann, L., and U. Blien. 2001. Wage curve analyses of establishment data from western Germany. *Industrial and Labor Relations Review* 54: 851-63.

Blanchard, O. J., and L. F. Katz. 1992. Regional evolutions. *Brookings Papers on Economic Activity* 1: 1-75.

Blien, U., and A. Tassinopoulos. 2001. Forecasting regional employment with the ENTROP method. *Regional Studies* 35: 113-24.

Cheng, B., and D. M. Titterington. 1994. Neural networks: A review from a statistical perspective. *Statistical Science* 9: 2-30.

Church, K. B., and S. P. Curram. 1996. Forecasting consumers' expenditure: A comparison between econometric and neural network models. *International Journal of Forecasting* 12: 255-67.

Decressin, J., and A. Fatás. 1995. Regional labour market dynamics in Europe. *European Economic Review* 39: 1627-55.

Diebold, F. X., and L. Kilian. 2000. Unit-root tests are useful for selecting forecasting models. *Journal of Business and Economic Statistics* 18: 265-73.

Diebold, F. X., and R. S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253-63.

Donaldson, R. G., and M. Kamstra. 1996. Forecast combining with neural networks. *Journal of Forecasting* 15: 49-61.

Fauvel, Y., A. Paquet, and C. Zimmerman. 1999. Short-term forecasting of national and provincial employment in Canada. Working Paper R-99-6E. Applied Research Branch, Strategic Policy, Human Resources Development Canada, Hull, Quebec.

Fischer, M. M. 2001a. Central issues in neural spatial interaction modeling: The model selection and the parameter estimation problem. In *New analytical advances in transportation and spatial dynamics*, ed. M. Gastaldi and A. Reggiani, 3-19. Aldershot, UK: Ashgate.

———. 2001b. Neural spatial interaction models. In *Geocomputational modelling*, ed. M. M. Fischer and Y. Leung, 195-219. Heidelberg, Germany: Springer.

Franses, P. H., and K. van Griensven. 1998. Forecasting exchange rates using neural networks for technical trading rules. *Studies in Nonlinear Dynamics & Econometrics* 2: 109-16.

Granger, C. W. J., and P. Newbold. 1986. *Forecasting economic time series*. Orlando, FL: Academic Press.

Haag, G. 2001. A neural network approach for mobility panel analysis. In *Geocomputational modelling*, ed. M. M. Fischer and Y. Leung, 220-35. Heidelberg, Germany: Springer.

Herbrich, R., M. Keilbach, T. Graepel, P. Bollmann-Sdorra, and K. Obermayer. 1999. Neural networks in economics: Background, applications and new developments. In *Computational techniques for modelling learning in economics*, ed. T. Brenner, 169-93. Dordrecht, the Netherlands: Kluwer Academic.

Himanen, V., P. Nijkamp, and A. Reggiani. 1998. *Neural networks in transport application*. Aldershot, UK: Ashgate.

Hsiao, C. 2003. *Analysis of panel data*. Cambridge: Cambridge University Press.

Kaashoek, J. F., and H. K. van Dijk. 1997. A simple strategy to prune neural networks with an application to economic time series. Discussion Paper 97-123/4. Rotterdam, the Netherlands: Tinbergen Institute.

Krugman, P. 1991. *Geography and trade*. Cambridge, MA: MIT Press.

———. 1998. Space, the final frontier. *Journal of Economic Perspectives* 12: 161-74.

Kuan, C.-M., and T. Liu. 1995. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics* 10: 347-64.

Kuan, C.-M., and H. White. 1994. ANNs: An econometric perspective. *Econometric Reviews* 13: 1-91.

Morgan, P., B. Curry, and M. Beynon. 2000. Pruning neural networks by minimization of the estimated variance. *European Journal of Economic and Social Systems* 14: 1-16.

Nijkamp, P., A. Reggiani, and E. Sabella. 2001. Neural network and spatial interaction analysis of european commodity flows. In *New analytical advances in transportation and spatial dynamics*, ed. M. Gastaldi and A. Reggiani, 89-115. Aldershot, UK: Ashgate.

Pindyck, R. S., and D. L. Rubinfeld. 1998. *Econometric models and economic forecasts*. Boston: McGraw-Hill.

Pizarro, J., E. Guerrero, and P. L. Galindo. 2000. A statistical model selection strategy applied to neural networks, proceedings of the European Symposium on Artificial Neural Networks, Bruges, Belgium, 55-60.

Reggiani, A., T. de Graaff, and P. Nijkamp. 2002. Resilience: An evolutionary approach to spatial economic systems. *Networks and Spatial Economics* 2: 211-29.

Reggiani, A., P. Nijkamp, and E. Sabella. 2000. A comparative analysis of the performance of evolutionary algorithms. In *Spatial economic science: New frontiers in theory and methodology*, ed. A. Reggiani, 332-54. Berlin: Springer-Verlag.

Stock, J. H., and M. W. Watson. 1998. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. NBER Working Paper 6607. Cambridge, MA: National Bureau of Economic Research.

Swanson, N. R., and H. White. 1997a. Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting* 13: 439-61.

———. 1997b. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economic and Statistics* 79: 540-50.

Taylor, J. C. 1997. The historical background, handbook of neural computation 1. Pp. A1.1:1-A1.1:7. http://library.thinkquest.org/18242/data/resources/nnhistory.pdf?tqskip1=1&tqtime=0429.