

Oberhofer, Walter; Blien, Uwe; Tassinopoulos, Alexandros

**Conference Paper**

## Forecasting Regional Employment With A Generalised Extrapolation Method

40th Congress of the European Regional Science Association: "European Monetary Union and Regional Policy", August 29 - September 1, 2000, Barcelona, Spain

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Oberhofer, Walter; Blien, Uwe; Tassinopoulos, Alexandros (2000) : Forecasting Regional Employment With A Generalised Extrapolation Method, 40th Congress of the European Regional Science Association: "European Monetary Union and Regional Policy", August 29 - September 1, 2000, Barcelona, Spain, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/114825>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Forecasting Regional Employment with a Generalised Extrapolation Method

Paper prepared for presentation at the 40<sup>th</sup> European Congress 'European Monetary Union and Regional Policy' in Barcelona 2000

16.05.2000

Walter Oberhofer\*, Uwe Blien\*\*, Alexandros TASSINOPOULOS\*\*,

**Abstract:** Forecasts of regional employment serve a double purpose. On the one hand they include a strict test of the identification of causal processes, which generate differing rates of employment in regions. On the other hand they are useful for the planning of the measures of active labour market policy.

In the paper an innovative method is applied to generate this forecast. In common extrapolation problems the entries of a base matrix are used to estimate the elements of an unknown matrix. In the present problem of forecasting the base matrix is given for a time period in the past and the matrix to be estimated is related to a future period. In this paper a generalisation of this approach is used since the data about the past is given for a number of points in time and not only for a single period.

The technique developed to take this information into account is related to well-known methods of panel analysis based on models with fixed effects. Additionally, it is similar to the techniques of entropy optimisation. The objective function to compare the past and the future is a modified chi-squared statistic, which is approximately equal to the relative entropy function often used in gravity models and in input-output analysis.

The technique is demonstrated with respect to the example of employment in all 327 (western) German districts for a time span of two years. The method chosen uses any available information extensively. Therefore, the estimates are reliable, as is shown in an ex-post forecast.

\*University of Regensburg, Universitätsstraße 31, 93040 Regensburg, Germany

\*\*Institute for Employment Research (IAB), Regensburger Str. 104, 90327 Nuremberg, Germany

## **1. Two purposes of employment forecasts**

The problem the paper is related to is the forecast of employment in the 327 districts of Western Germany. This is not only a technical problem as it might seem at first sight. A major purpose of scientific research is the identification of the structure of causes in a given field of interest. To test analyses of this kind forecasts are appropriate. The structure of causes identified in an analysis of the past can be extrapolated to a future situation. If these forecasts are reliable, the researcher can be confident that he is able to identify the main influences he is looking for. It is possible to evaluate the results of a forecast which serves as an additional test whether the model used adequately represents the causal structure inferred from a theory. The evaluation can be done either with an ex-post forecast (a simulated extrapolation for a time period which has already passed) or - after waiting for some time - with a 'real' forecast (related to the future).

Compared to this situation, analyses of past developments are always in the danger of overfitting. Random fluctuations might affect the data and a model might represent these perfectly but show poor results when transferred to the future. The construction of forecasts, however, can help to discriminate between random and systematic variation.

Apart from this scientific purpose, there is the policy aspect. Any institution that carries out measures of active labour market policy has an urgent need of prospective information. To counteract imbalances on the labour market at an early point in time, information about the future is necessary. Reliable forecasts can help to concentrate measures on disadvantaged segments of the labour market before severe damages occur. The use of pure statistical information is not sufficient, since it is available only with a lag. If the allocation of money to the various programmes and regions is based on statistical information, it is based on information about the past.

Both motivations, that related to labour market policy and that concerning scientific explanations, are relevant in the present context, which deals with regional forecasts of employment in western Germany for a time horizon of two years. The data basis for the forecast is a pseudo panel for the employment in 11 industries and 327 districts (Stadt- und Landkreise) of western Germany for 11 years (1987-1997). The special research question is the employment in the single districts in the year 1999.

The main concern of the paper is the methodology necessary for an extrapolation, only a brief discussion of the theoretical aspects is given. The econometric techniques used in the

forecast are completely new developed. They connect two large fields of statistical method, since they are related to regression analysis and to entropy optimisation procedures.

In the next section of the paper the approach chosen is presented. Then, in section 3, the data basis is described. The details of the method used can be found in the section 4. Then, the results obtained are described and an ex-post forecast is evaluated. Finally a conclusion is given.

## **2. Theoretical and empirical approach**

Market economies are based on decentralised decisions of the economic subjects. In principle it can be seen only ex post whether these decisions are compatible. This is one of the reasons why forecasts of economic developments are very difficult. There is hope, however, that it is possible to gain forecasts of the labour market which are reliable up to a certain degree. The labour market reacts to changes of basic parameters of the economy with a lag, especially compared with the development of goods markets. Firms react to changes in the demand for their products first by changing their capacity utilisation. Hires and layoffs come later. This lag can be used for forecasts of the labour market. It is subject to the discussion about leading indicators for economic development (Lahiri, Moore 1991). The lag is especially large in the case of Germany, since in this country the firms have to consider legal restrictions of layoffs which are stronger than in the Anglo-Saxon countries. Therefore, German firms can be expected to follow at least partly a policy of labour-hoarding.

The forecast can be prepared by using information about employment expectation of the firms. Because of the lag between the demand situation on goods markets and the personnel policy of the firms, it can be expected that the employment expectations are a relatively solid basis for future decisions of the firms. Additionally, theoretical arguments about regional specialisation and location are important to include causal factors in the preparation of the forecast:

- From Krugman's (1991, 1998) analyses and from classical location theory (cf. Puu 1997 for an overview) it is known that regions tend to specialise in single industries. These industries show differing patterns of business cycles. Therefore, the exploitation of the composition of employment with respect to industries is important.

There is a new discussion about the impacts of structural change on employment, partly following Appelbaum, Schettkatt (1993, 1994, Schettkat 1997). In this discussion it is

shown that changes in the elasticity of demand on products markets may have important employment effects (cf. Möller 1998 for an empirical analysis and Blien 1999 for a regional application).

- The location of a region is important for its fate. There are other arguments from location theory showing this. The quality of a location changes slowly according to changes in transportation technology, institutional restrictions and political changes. Since the advantages or disadvantages of the location of a region are relatively stable, this property can be used for causal analyses of regional developments and for forecasts.
- The type of a region is important. There are strong trends, which affect all regions of the same type. E. g. in the last twenty years, there is a loss of the importance of large city centres compared to the periphery of agglomerations. Therefore, there is a tendency to suburbanisation.

The econometric technique used has been specially developed for the purpose of the forecast. To analyse the employment development an auto-regressive panel model with fixed effects is applied. Since the estimation of trends is included, the results could be used directly for an extrapolation of future developments.

The forecast uses basically the employment statistics of the Federal Employment Services. A pseudo-panel is constructed which includes repeated measurements on the aggregate level, for districts. For a further improvement of the forecast, additional external information is included, which is of heterogeneous form. Normally, it is available only referring to aggregates, i. e. with respect not to single districts, but with respect to larger regions. Other examples are assessments of the future developments of special industries which are based on surveys about firms' business expectations. Usually the results of these surveys are available only for the whole country, not for small regional units.

In principle it would be possible to integrate both sources of information by estimating first a base matrix X of employment for 1999 with a panel model. X contains the data about the districts in the rows and about the industries in the columns. In a second step a matrix Y could be estimated which refers to the future and which is most similar to the base matrix X, but satisfies the requirements specified by the additional external information.

This task could be done by minimising a distance function between X and Y under the constraints given by additional external information. A distance function with 'nice' properties is a modified relative entropy:

$$d = E_R(Y, X) = \sum_{i,j=1} y_{ij} \left[ \ln \left( \frac{y_{ij}}{x_{ij}} \right) - 1 \right] + \sum_{i,j=1} x_{ij}$$

Here,  $i$  is the index of regions, i. e. districts,  $j$  the index of industries.  $x_{ij}$  and  $y_{ij}$  represent the single entries of the matrices  $X$  and  $Y$ . If used in constrained optimisation, a matrix  $Y$  is generated that is similar to the base matrix  $X$ , since small elements of  $Y$  are small in  $X$ . This effect is called the property of structure conservation. Additionally, under the assumption of a multinomial distribution, the estimated matrix is the most probable one, given the specified constraints. The entropy function was optimised in an employment forecast (Blien, Tassinopoulos 1998) with the same data base as in the present paper. For the estimations the ENTROP algorithm (Blien, Graef 1998) was developed. Here, instead of the relative entropy the  $\chi^2$ -criterion is used:

$$\chi^2(Y, X) = \sum_{\substack{i, j \\ x_{ij} \neq 0}} \frac{(y_{ij} - x_{ij})^2}{x_{ij}}$$

But there is only a slight difference between the two objective functions, since it can be shown that they are identical up to a Taylor approximation. But the approach chosen here differs in more than this respect from previous work. A special advantage of the new method is the use of ‘soft’ constraints. It is possible to specify constraints which do not have to be exactly binding, but could be violated to a certain degree. The extent of this violation can be specified ex-ante. An additional advantage of the method used here is the fusion of this extrapolation techniques with regression analysis as can be seen in detail from section 4.

The approach chosen here implies a combination of top down and bottom up methods. On the one hand, in the forecast of the employment for 1999 for the 327 districts (Stadt- und Landkreise) of Western Germany, autonomous developments of these regional units are relevant. On the other hand, a projection of global employment is broken down into small regional units. There are global influences affecting employment in all regions in nearly the same way, for example the interaction of the business cycle with the institutions of the Federal Republic of Germany. It can be argued that the global component of regional development is of special relevance in Germany (with the exception that it is always necessary to discriminate between the western and the eastern part of the country), since important economic processes affect the whole country. Wage bargaining for example is rather centralised, since it takes place at the level of industries, not at the level of regions or of firms. Therefore, it would be misleading to realise a forecast for the country on the basis of regional information only.

The contrary would also be misleading, since it is not possible to see regional developments as entirely determined by general influences. Large variation in the level of local unemployment rates indicates large disparities on regional labour markets. The unemployment rate (average for the country for 1997) in western Germany was 10.8 % with

the extreme values of 4.4 % (district of Erding) and of 20.1 % (district of Bremerhaven). It is therefore necessary to identify the sources of this variation and to supplement the top down method by bottom up strategies.

For the top down part of the forecast especially the proportion of the industries present in the regions is used. The importance of the industrial structure is known from the bulk of literature on shift-share analyses (Brown 1969, Holden et al. 1989, Patterson 1991, Selting, Loveridge 1994). The business cycle affects different industries in differing ways. Especially manufacturing reflects slumps and peaks in the demand on the world market to a greater extent than the service sector of the economy.

The business cycle of an industry is synchronised because of spill-over effects. An especially interesting example is Germany's automobile industry. Phases of relatively good competitiveness compared to car production in other major countries (e. g. in Japan) are followed by phases of relatively low productivity and falling shares in sales on the world market. Though the performance of the individual companies in a country differs, there is a common trend affecting all car producers. The spill-over effects between the companies are described in general terms by the new growth theory (Lucas 1988, Romer 1986) and in more detail by evolutionary economics (Nelson, Winter 1982, Dodi, Nelson 1994, Nelson 1995). The growth perspectives of industries also differ as can easily be seen from the growing share of employment which is absorbed by the service sector and is described as tertiarisation of the economy.

Information that can be exploited for the forecast ends mainly in 1997, so the time horizon is two years. The methods used have the advantage of extensively taking into account any available information. They are therefore preferable to standard shift-share or regression techniques.

### **3. Data**

The data used in the forecast is from the employment statistics of western Germany. The data basis is not a sample; it comprises a sequence of complete cross-sections of all people employed on any 30 June of the period between 1987 and 1997 (statistics based on location of the workplace).

The employment statistics include information about the entire population of people in gainful employment and covered by the social insurance system, i. e. about 80 % of all

employment in Germany. Two main groups of people whose data is not included are civil servants and workers with an income lower than DM 620 a month.

The employment statistics give continuous information on employment spells, earnings, job and personal characteristics. It is based on microdata delivered by individual firms about their individual employees. For every employee a new record is generated every year. The same is done if he or she changes establishment. From 1975 to 1995, over 600 million records about the employment of individual workers are included in the files of the Federal Employment Services (Bundesanstalt für Arbeit, see Bender et al. 1996).

Originally, the data was collected for administrative purposes of the social security system. Since they are used to calculate the pensions of retired people, the income and duration information is very reliable. No problem of recalling from memory and reporting is encountered, as is the case in population surveys.

Since the employment statistics contain detailed regional information, an assessment of regional employment is possible. Every employed person is registered at his or her place of work. Several additional variables are available. Here the information about the respective industry is used. This variable has 11 categories (cf. Table 1) The classification of industries follows the work on structural change (Dietz 1988) and has the advantage of separating in particular the various components of the service sector. The evaluation of the data of the tertiary sector showed that the overall gains are not equally distributed among all types of services. Especially the financial services and the services for the society (e.g. education, health service, research) show - in contrast to other types - a relative increase of people employed in their area. This implies a crucial differentiation between the types of service in order to forecast the growing importance of this sector with more accuracy. To control for industry-specific business cycles the forecast is based on a calculation of the prospects of each industry.

Table 1: Industries (IAB typology):

1. Primary sector	7. Construction
2. Energy/mining	8. Distributive services
3. Goods-producing industry	9. Financial services
4. Capital goods	10. Household services
5. Consumer goods	11. Services for society
6. Manufacture of food products and beverages	



For every year  $t$  a data matrix  $X(t)$  is obtained. Its rows  $i$  are the districts of western Germany ( $i = 1 \dots 327$ ), and its columns  $j$  the industries ( $j = 1 \dots 11$ ). The data refer to eleven years ( $t = 1 \dots 11$ ), since the time span under observation is 1987 to 1997.

#### 4. Econometric method

The application of the generalised extrapolation method includes three steps. In the first step, which is a preparatory one, the external information, which is available about the matrix to be estimated, has to be made consistent. In the second step an appropriate regression model has to be specified and in the third step the estimations are carried out. We start the outline of the method with the preparatory step.

##### 4.1 Consistent constraints

External constraints are important in solving the problem of extrapolation. An obstacle often encountered in empirical work is that these constraints may be inconsistent, since the data basis can include errors. In this case the set of solutions of the extrapolation problem is empty.

The errors in the data can be caused simply by "noise" or because of sampling errors. An additional possibility is that the constraints may be based on expert judgements which are accurate only up to an assessment error. The constraints may be inconsistent because the underlying definitions of facts (e. g. the specifications of industries) may differ or they may refer to differing time periods. A typical source of all these inconsistencies is that the constraints may be taken over from various external sources and the design of the surveys, which is the basis of the information used, may vary.

In the following a procedure is discussed that has the purpose of transforming the constraints to make them consistent. The correction is carried out in a way that minimises changes. Only linear constraints in the form of equations are taken into account.

$$Ay = b \tag{1}$$

The vector  $y$  is obtained by setting  $\text{vec}(Y) = y$ , with  $\text{vec}()$  being the operator which transforms a matrix to a vector by stacking the columns of the matrix. The matrix  $Y$  represents the dependent variables i. e. it contains the distribution of employment (the number of employees in defined regions and industries) at the year  $T+t_0$ .  $t_0$  is the time horizon of the forecast. In the

following, it is set  $t_0 = 1$  for reasons of simplicity.  $Y$  has  $n$  rows and  $m$  columns.  $A$  is an  $r \times k$  matrix of coefficients with  $r < k = nm$ .  $r$  is the number of constraints and the vector  $b = (b_f)$  the right hand side of the equations. If the rank of  $A$  is smaller than  $r$ , the system (1) might be inconsistent.

In many extrapolation problems only the row and the column sums are given. Then, the matrix  $A$  contains only ones and zeroes and the vector  $b$  represents the row and column sums. Because the row sums and column sums are not independent, the matrix  $A$  is singular.

To correct the constraints it is assumed that the right hand side of (1) is stochastic:

$$b = c + u$$

The vector  $u$  represents an unobserved error term with a diagonal and nonsingular variance-covariance matrix  $D = (d_f)$ . The variance  $d_f$  is a measure of the accuracy of  $b_f$ . To obtain a consistent system of constraints, the right hand side  $b$  has to be substituted by an estimated  $\hat{b}$ . The system (1) is consistent if there is a vector  $z$  with

$$Az = b$$

Therefore,  $\hat{b}$  is calculated by using an appropriate vector  $z$  so that:

$$\hat{b} = Az$$

with  $\hat{b}$  as close as possible to  $b$ . To find a vector  $z$  the following minimisation problem has to be solved:

$$(b - Az)'D^{-1}(b - Az) \rightarrow \text{Min}$$

The first order conditions are:

$$(A'D^{-1}A)z = A'D^{-1}b \tag{2}$$

The system of  $k$  unknown variables  $z_i$  has always at least one solution though the matrix of coefficients

$$A'D^{-1}A$$

is singular. It has the same rank as  $A$ , i. e. a rank smaller than  $k$ . Since the right hand side

$$\hat{b} = Az$$

is independent of the special solution of (2), a special pseudo inverse can be used to find a solution:

$$z = (A'D^{-1}A)^+ A'D^{-1}b \tag{3}$$

Thus, the problem of finding consistent constraints is reduced to the problem of solving a system of linear equations.

#### 4.2 Relative entropy and $\chi^2$ -criterion

In the following it is assumed that the system of constraints is consistent. Often the extrapolation of a basis matrix  $X = (x_{ij})$  of the size  $n \times m$  to a matrix  $Y = (y_{ij})$  with the same size  $n \times m$  is done by using the relative entropy as an objective function:

$$E'(Y) = \sum_{\substack{i,j \\ x_{ij} \neq 0}} y_{ij} \ln \left( \frac{y_{ij}}{x_{ij}} \right) \quad (4)'$$

A minimisation of  $E'$  without any constraints gives  $y_{ij}$  not identical to the  $x_{ij}$ . They differ by the factor  $e^{-1}$ . To avoid this effect, a slightly modified definition of the entropy can be used:

$$E(Y) = \sum_{\substack{i,j \\ x_{ij} \neq 0}} y_{ij} \left[ \ln \left( \frac{y_{ij}}{x_{ij}} \right) - 1 \right] + \sum_{i,j} x_{ij} \quad (4)$$

This function<sup>1</sup> is minimised subject to the following constraints:

$$\text{Avec}(Y) = \mathbf{b} \quad (5)$$

If  $\mathbf{b}$  contains the row and column sums, optimising (4)' produces the same results as optimising (4). For all  $x_{ij} = 0$ , the corresponding  $y_{ij}$  are set  $y_{ij} = 0$ . If only row and column sums are specified the matrix  $A$  contains  $n + m$  rows. This is the number of constraints. The rank of  $A$  is  $n + m - 1$ . Minimising the relative entropy leads to the well-known equation:

$$y_{ij} = x_{ij} \alpha_i \beta_j, \quad 1 \leq i \leq n, 1 \leq j \leq m \quad (6)$$

with  $\alpha_i$  and  $\beta_j$  as dual variables<sup>2</sup> (Lagrange multipliers). These can be obtained iteratively by using the RAS-method (Deming, Stephan 1940, Bacharach 1970), which is identical with the Iterative Proportional Fitting Algorithm (IPF) used in log-linear models (cf. Bishop et al. 1975).

$$x_{ij}^{s+1} = \frac{x_{ij}^s b_i^r}{\sum_p x_{ip}^s} \quad \text{all } i$$

$$x_{ij}^{s+1} = \frac{x_{ij}^s b_j^c}{\sum_p x_{pj}^s} \quad \text{all } j$$

---

<sup>1</sup> If  $x_{ij}$  and  $y_{ij}$  do not represent absolute but relative frequencies, which could be interpreted as probabilities, (4) is obviously reduced to (4)'. In this case  $E(Y)$  is called the "information gain" (following Kullback 1968).

<sup>2</sup> If it is set  $\alpha_i = R_i$ ,  $\beta_j = S_j$ , and  $x_{ij} = A_{ij}$ , it can easily be seen why the iterative procedure normally applied to find the solution (6) is called the RAS method. The property of the solution, that all resulting  $y_{ij}$  can be calculated by using constant factors specific for the respective row  $i$  and column  $j$  is called "double proportionality".

Here, the  $b_i^r$  represent the row and the  $b_j^c$  the column sums,  $s$  is the step in the iteration process. In one step the RAS operations are carried out for all rows and then for all columns. This is a very simple procedure, including only linear operations, which has contributed much to the widespread use of the relative entropy as an objective function in solving extrapolation problems, especially in input-output analysis. Other applications are gravity models (cf. Wilson 1970, Nijkamp 1975, Batten 1983).

Instead of the relative entropy a different function can be used. (4) can be written as

$$\begin{aligned}
 E(Y) &= \sum_{\substack{i,j \\ x_{ij} \neq 0}} y_{ij} \left[ \ln \left( \frac{y_{ij} - x_{ij}}{x_{ij}} + 1 \right) - 1 \right] + \sum_{i,j} x_{ij} \\
 &\approx \sum_{\substack{i,j \\ x_{ij} \neq 0}} y_{ij} \frac{y_{ij} - x_{ij}}{x_{ij}} - \sum_{i,j} y_{ij} + \sum_{i,j} x_{ij} \\
 &= \sum_{\substack{i,j \\ x_{ij} \neq 0}} (y_{ij} - x_{ij})^2 / x_{ij} = \chi^2
 \end{aligned} \tag{7}$$

In step two the famous Taylor approximation  $\ln(u+1) \cong u$  is used, which is especially accurate if  $u \ll 1$ . Therefore, the weighted sum of squares, i. e. the  $\chi^2$ -criterion defined by (7), is close to the relative entropy, if  $(y_{ij} - x_{ij}) / x_{ij} \ll 1$  (the proof is slightly modified after Kádas, Klafszky 1976).

It is possible to use the  $\chi^2$  criterion instead of the relative entropy to estimate a matrix  $Y$  on the basis of a matrix  $X$  subject to linear constraints. Since the relative entropy and the  $\chi^2$  criterion are only approximately equivalent, the size of the deviations in practical applications of the two functions is important. Additionally, it is important by which function better results are generated. This can be assessed by using known matrices  $X$  and  $Y'$  and checking the deviation between  $Y'$  and the extrapolated  $Y$ . This experiment was carried out with input-output tables. The results differed only slightly and the comparison with the known exact table showed that there is no clear ranking of the objective functions. In some cases the results of an extrapolation were better if the weighted sum of squares was used, in others the application of the relative entropy was superior (Oberhofer 1998).

Therefore, in many cases it is better to use the  $\chi^2$  criterion, since its optimisation leads to simple linear expressions and there is a close relationship to standard techniques of regression analysis. Both aspects will be made clearer in the following.

An application of the  $\chi^2$  criterion implies the minimisation of this Lagrange function:

$$\sum_{\substack{i,j \\ x_{ij} \neq 0}} (y_{ij} - x_{ij})^2 / x_{ij} - 2\lambda'(\text{Avec}(Y) - b) \rightarrow \text{Min} \quad (7)'$$

To simplify the following exposition a special case is regarded, where only the row and the column sums of the matrix Y are specified. The first order conditions can be written as:

$$y_{ij} = x_{ij}(1 + \alpha_i + \beta_j), \quad 1 \leq i \leq n, \quad 1 \leq j \leq m, \quad (8)$$

Again  $\alpha_i$  and  $\beta_j$  are elements of the vector  $\lambda$  with the Lagrange multipliers. Comparing (8) and (6) it is obvious that (8) has a simpler structure, since it is a linear system, which can easily be solved. Formally (8) is equivalent to a two-way panel model with fixed effects. The effects  $\alpha_i$  and  $\beta_j$  can be calculated exactly from the specification of the constraints and are not gained by an estimation.

#### 4.3 Extrapolation with panel data and regression analysis

It is now assumed that data is available which has a panel structure. Therefore, the data can be interpreted as consisting of a sequence of matrices:

$$X(1), X(2), \dots, X(T) \quad T \geq 1$$

The aim of the extrapolation implies the calculation of a matrix  $Y = X(T+1)$ . Again the resulting matrix is subject to r linear constraints

$$\text{Avec}(Y) = b \quad (9)$$

Additionally, it is assumed that the sequence of  $X(t)$  can be modelled in a regression approach:

$$x_{ij}(t) = \mu_{ij} + u_{ij}(t), \quad 1 \leq i \leq n, 1 \leq j \leq m \quad (10)$$

As usual, the  $u_{ij}(t)$  are assumed as being mutually independent:

$$\text{Var}(u_{ij}(t)) = v_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq m$$

According to the regression approach (10) the  $\mu_{ij}$  are estimated by  $\hat{\mu}_{ij}$ . A simple extrapolation is possible by setting:

$$y_{ij} = \hat{\mu}_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq m$$

To include the constraints (9) in the estimations, the results must satisfy the following condition (with  $M = (\mu_{ij})$ ):

$$\text{Avec}(M) = b$$

Then, the Lagrange equation is

$$\sum_{i,j,t} \frac{(x_{ij}(t) - \mu_{ij})^2}{v_{ij}} - 2\lambda'(Avec(M) - b) \rightarrow \text{Min} \quad (11)$$

If  $T = 1$  and  $v_{ij} = \sigma^2 \mu_{ij}$  the results are equivalent to the extrapolation (7)', as can be easily shown. In the frame work of a feasible generalized least squares procedure the  $v_{ij}$  are estimated by

$$\hat{v}_{ij} = x_{ij}(1), \quad \text{with} \quad x_{ij}(1) \neq 0,$$

(11) is simplified to:

$$\sum_{\substack{i,j \\ x_{ij}(1) \neq 0}} (x_{ij}(1) - \mu_{ij})^2 / x_{ij}(1) - 2\lambda'(Avec(M) - b) \rightarrow \text{Min} \quad (12)$$

with the extrapolation

$$Y = M$$

#### 4.4 Discussion

The relative entropy is widely used to solve problems of extrapolating a matrix  $X$  to a matrix  $Y$ . Since this objective function is approximately equivalent to the  $\chi^2$  criterion, the approach chosen here is a fusion of common extrapolation techniques with regression analysis. It has been shown that the standard extrapolation technique can be described as a simple problem of regression analysis. Two separately developed branches of econometric methods and statistics turn out to be closely related.

This interpretation is the basis of several generalisations:

1. The regression equation can be modified. Instead of using the simple equation (10) various kinds of panel models with time dependent co-variables can be estimated.
2. A sequence of matrices can be used instead of only one basis matrix. Auto-regressive models are also possible.
3. The variances  $v_{ij}$  can be modified to express special weightings of the data.

Advantages of the new technique are its flexibility and the extensive use of the available information. It is possible to use the full information content of the data in one estimation process together with external information. To include theoretical considerations the models can be adapted in many ways. The extrapolation can be based on various influences represented by exogenous variables, which are regarded as important. The trend of the development can be directly taken from panel  $(X(1)...X(T))$ , which is the basis of the forecast.

Additionally, all kinds of external information, expert judgements etc. can be included in the form of external constraints.

Therefore, forecasts generated with the new method can be expected to be more reliable than those based on traditional techniques.

The treatment in section 4.1 shows that it is possible to include information which is affected by errors and therefore inconsistent. According to an a-priori assessment, the size of these errors can be included in a step of calculations carried out before starting the genuine estimations. This first step facilitates the inclusion of "soft" information, which is valuable and important to be observed, but which is not very accurate.

In previous work with entropy optimising procedures, especially with the ENTROP method (cf. Blien, Graef 1998, Blien, Tassinopoulos 1998), soft information is observed by including inequalities in the estimation process. The ENTROP procedure is a generalisation of the traditional RAS method, since not only the row and the column sums can be included, but all kind of linear equations and inequalities.

To specify soft information with ENTROP, it is assumed that the exact value  $b_k$  of the right hand side of a constraint  $k$  lies between two bounds  $b_k^l \leq b_k \leq b_k^u$ . Two cases are possible for an individual constraint. If a constraint is binding, it could be substituted by an equation, e. g. by  $b_k^l = b_k$ . If it is not, it could be dropped without changing the result of the extrapolation. However, the researcher is often not interested to specify exact bounds, but to obtain a result which is close to  $b_k$ . Then  $b_k$  is measured with an error  $b_k = b_k + u_k$ , as it is assumed in section 4.1. In this case the method introduced here is more appropriate.

## **5. Preparation of the forecast and results**

The data basis for the forecast were the matrices  $X(t)$  of the employment statistics for the years  $t$  from 1987 to 1995. The aim of the analysis was the extrapolation to 1997. Since the actual employment matrix of 1997 is known, this is an ex post forecast and in a comparison the error of the calculated matrix could be obtained.

Additionally to the data of the employment statistics 'external' information of the year 1995 about employment in 1997 was available. These sources were mainly based on the Establishment Panel of the Institute for Employment Research (IAB). The IAB has built up this survey to gain information about the structures and developments of the demand side of

the labour market. Major determinants of corporate employment and personnel policies can be analyzed. This survey is the only existing establishment panel representative for employment of a large economy (Bellmann et al. 1996).

For the present forecast a question of the 1995 wave of the panel was used that concerned the prospects of employment in the respective establishments. Since the IAB panel is not representative for small regional units this information had to be aggregated according to industries. The results could then be included as external information related to column sums of the matrix  $\hat{X}$  (1997) to be estimated.

In addition forecasts on industry specific business developments were included, which were published 1995 by the German research institutes IFO and IW. Since the available external data were contradictory, they were made consistent in the described way (cf. Section 4.1).

Then, the following simple regression model was used:

$$x_{ij}(t) = a_{ij}x_{ij}(t-1) + u_{ij}(t) \quad 1987 \leq t \leq 1995, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m \quad (13)$$

The estimates were obtained by minimising the weighted mean square i. e. the  $\chi^2$ -criterion.

$$\sum_{i,j,t} (x_{ij}(t) - a_{ij}x_{ij}(t-1))^2 / v_{ij} \quad (14)$$

with

$$v_{ij} = \sum_t x_{ij}(t)$$

This is equivalent to a feasible generalized least squares procedure.

The model used is a very simple one. Since individual coefficients  $a_{ij}$  are estimated, region specific effects are included and so are effects of variables which do not change in time e. g. the type and the location of a region. More sophisticated models are possible, the one used here serves primarily as a reference to test whether the method is useful.

All  $n \cdot m = 327 \cdot 11$  elements of the matrix were based on a uniform regression model. The matrix elements follow individual trends which differ much. Therefore, the extrapolation of the status quo situation

$$x_{ij}(1997) = x_{ij}(1995)$$



is robust and – in many cases – better than uniform regression models. Consequently, a status-quo extrapolation was used as a base line for a comparison with the regression extrapolation. To assess the value of the external information, a pure regression extrapolation was carried out without any external constraints.

As a measure for the error the mean absolute deviation of the ‘real’ from the extrapolated employment in per cent  $F_1$  was used:

$$F_1 = \frac{100}{nm} \sum_{i,j} |x_{ij}(1997) - \hat{x}_{ij}(1997)| / x_{ij}(1997) \quad (15a)$$

This error measure implies the problem that the deviation of small  $x_{ij}$  is weighted relatively high. To avoid this disadvantage, the  $\chi^2$ -criterion can be used as an error measure  $F_2$ :

$$F_2 = \frac{100}{nm} \sum_{i,j} (x_{ij}(1997) - \hat{x}_{ij}(1997))^2 / x_{ij}(1997) \quad (15b)$$

Table 1 shows a comparison of the reliability of the differing forecast methods based on the two error measures. Row one with the naive status-quo extrapolation serves as a reference.

Table 1: Error of differing forecast methods

	Mean absolute error (percent)	$\chi^2$ -criterion
Status-quo extrapolation	7.99	40.88
Regression extrapolation without external information	9.05	47.99
<b>Generalized extrapolation method</b> (regression extrapolation with external information)	7.14	30.33

For the time span of two years, which is relevant here, the status quo extrapolation generates a result, which is better than a pure regression extrapolation without including any external information. The forecast obtained by the generalised extrapolation method, which improves the regression by using external information, is better than both the simple regression and the status-quo forecast.

## 6. Conclusion

In this paper a new method is presented. It is appropriate for forecasting and for other extrapolation problems. Here, the results of an application in forecasting regional employment for a two-year time span is shown.

Even by using a simple version of the regression model, which is one genuine part of the new generalised extrapolation method, results are obtained which are better than the status-quo forecast. It can be expected, that the results could be improved significantly, if more exogenous variables from the employment statistics would be included. The discussion about shift-share techniques in regional economics shows that only a smaller part of regional developments can be explained by the structure of industries, which is the most important set of variables here.

The generalised extrapolation method can be used for many problems of empirical economics. Besides forecasting, the disaggregation of aggregate units is one possible application. Others are the estimation of transition or flow matrices, e. g. in Markov processes. The matrices could be from input-output analysis, from migration analysis or from demographic accounting.

Application in forecasting can help to satisfy the demand for prospective information in regional and active labour market policy. Strict tests of economic theories are also possible.

## References:

- Appelbaum, Eileen; Schettkat, Ronald (1993): "Employment Developments in Industrialized Economies: Explaining Common and Diverging Trends" (Discussion Paper FS I 93- 313, Wissenschaftszentrum Berlin für Sozialforschung)
- Appelbaum, Eileen; Schettkat, Ronald (1994): "The End of Full Employment? On Economic Development in Industrialized Countries", in: *Intereconomics* 29: 122-130
- Bacharach, M. (1970): "Biproportional Matrices and Input-Output-Change", Cambridge
- Balderjahn, Ingo; Mathar, Rudolf; Schader, Martin (1998) (Hrsg.): "Classification, Data Analysis, and Data Highways" (Proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation), Berlin etc.: Springer
- Batten, David F. (1983): "Spatial Analysis of Interacting Economics", Boston usw.: Kluwer-Nijhoff
- Bellmann, Lutz; Düll, Herbert; Kühl, Jürgen; Lahner, M.; Lehmann, U. (1996): "Patterns of Enterprise Flexibility in Germany. Results of the IAB Establishment Panel 1993 to 1996" (Project report of the Institute for Employment Research), Nürnberg
- Bender, Stefan; Hilzendegen, Jürgen; Rohwer, Götz; Rudolph, Helmut (1996): "Die IAB-Beschäftigtenstichprobe 1975-1990. Eine praktische Einführung", Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg: Bundesanstalt für Arbeit
- Bishop, Y.; Fienberg, S.; Holland, P. (1975): "Discrete Multivariate Analysis: Theory and practice", Cambridge, London
- Blien, Uwe (1999): "Arbeitslosigkeit und Entlohnung auf regionalen Arbeitsmärkten. Theoretische Analyse, ökonometrische Methode, empirische Evidenz und wirtschaftspolitische Schlußfolgerungen für die Bundesrepublik", Habilitationsschrift, Veröffentlichung in der Reihe Beiträge zur Arbeitsmarkt- und Berufsforschung in Vorbereitung, Nürnberg: Bundesanstalt für Arbeit
- Blien, Uwe; Graef, Friedrich (1998): "Entropy Optimizing Methods for the Estimation of Tables", in: Balderjahn et al. (1998): 3-15
- Blien, Uwe; Tassinopoulos, Alexandros (1998): "Forecasting Regional Employment with the ENTROP Method", paper prepared for presentation at the EALE Conference, Blankenberge
- Brown, H. J. (1969): "Shift and Share Projections of Regional Economic Growth, an Empirical Test", in: *Journal of Regional Science* 9: 1-18
- Deming, W. Edwards; Stephan, Frederick F. (1940): "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", in: *The Annals of Mathematical Statistics* 11/1, March, S. 427 ff.
- Dietz, Frido (1988): "Strukturwandel auf dem Arbeitsmarkt. Entwicklung bei den sozialversicherungspflichtig beschäftigten Arbeitnehmern nach Wirtschaftszweigen, Berufen und Qualifikationen zwischen 1974 und 1986", in: *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 21/1
- Holden, D.; Nairn, A.; Swales, J. K. (1989): "Shift-share Analysis of Regional Growth and Policy: a Critique", in: *Oxford bulletin of economics and statistics* 51 (1): 5-34

- Kádas, Sándor A.; Klafszky, Emil (1976): "Estimation of the Parameters in the Gravity Model for Trip Distribution: a New Method and Solution Algorithm", in: *Regional Science and Urban Economics* Vol. 6: 439-457
- Krugman, Paul (1991): "Geography and Trade", Cambridge (Mass.) etc.: MIT Press
- Krugman, Paul (1998): "Space: The Final Frontier", in: *Journal of Economic Perspectives* 12/2: 161-174
- Kullback, Salomon (1968): *Information Theory and Statistics*", zweite Auflage, New York: Dover
- Lahiri, K.; Moore, G. H., eds. (1991): "Leading Economic Indicators", Cambridge et al.: Cambridge University Press
- Lucas, Robert E. (1988): "On the Mechanics of Economic Development", in: *Journal of Monetary Economics* 22: 3-42
- Möller, Joachim (1998): "Income and Price Elasticities in Different Sectors of the Economy. An Analysis of Structural Change for Germany, the UK and the USA", unpublished paper of the University of Regensburg
- Möller, Joachim; Tassinopoulos, Alexandros (1998): "Zunehmende Spezialisierung oder Strukturkonvergenz? Eine Analyse der sektoralen Beschäftigungsentwicklung auf regionaler Ebene", in: *Jahrbuch für Regionalwissenschaft* (im Erscheinen)
- Nelson, Richard R. (1995): "Recent Evolutionary Theorizing About Economic Change", in: *Journal of Economic Literature* 33/1: 48-90
- Nelson, Richard R.; Winter, Sidney G. (1982): "An Evolutionary Theory of Economic Change", Cambridge (Mass.), London: Belknap
- Nijkamp, Peter (1975): "Reflections On Gravity And Entropy Models", in: *Regional Science and Urban Economics* 5: 203-225
- Oberhofer, Walter (1998): "Schätzung von Matrizen bei unvollständiger Information", *Regensburger Diskussionsbeiträge zur Wirtschaftswissenschaft*, Nr. 314
- Patterson, M. G. (1991): "A note on the formulation of a full-analogue regression model of the shift-share-method", in: *Journal of Regional Science*, 31 (2): 211-216
- Puu, Tönu (1997): "Mathematical Location and Land Use Theory", Berlin etc.: Springer
- Romer, Paul M. (1986): "Increasing returns and long-run growth", in: *Journal of Political Economy* 94/5: 1002-1038
- Schettkat, Ronald (1997): "Die Interdependenz von Produkt- und Arbeitsmärkten. Die Wirtschafts- und Beschäftigungsentwicklung der Industrieländer aus der Produktmarktperspektive", in: *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 30/4: 721-731
- Selting, A.; Loveridge, S. (1994): "Testing Dynamic Shift-share", in: *Regional Science Perspectives* 24: 23-41
- Wilson, Alan G. (1970): "Entropy in urban and regional modelling", London: Pion Ltd.