38th Congress of the European Regional Science Association
28 August - 1 September 1998, Vienna

Miquel Clar, Raúl Ramos, Jordi Suriñach
Grup d'Anàlisi Quantitativa Regional. Universitat de Barcelona
Avda. Diagonal, 690 - 08034 Barcelona (Spain)
mclar@eco.ub.es, rrlobo@eco.ub.es, surinach@eco.ub.es

MEASURING REGIONAL MANUFACTURING PRODUCTION. AN ANALYSIS FOR THE SPANISH REGIONS.

ABSTRACT: In a big amount of economies (NUTS-I) the evolution of manufacturing production is analysed using Gross Domestic Product (GDP) and Gross Added Value (GAV) data from National Accounts. In Spain, the problem of using these data is that they are not available as soon as it would be desirable. In consequence, it is not possible to analyse the short term evolution of the industrial output through them. To solve these problems the Institute of Statistics of Spain (*Instituto Nacional de Estadística* -INE-) constructs a monthly Industrial Production Index (IPI) from data belonging to a survey addressed to firms. At a regional level (NUTS-II), the difficulties to monitor the evolution of manufacturing production are even bigger due to the nearly absence of official data. During the last years, different public and private institutions have started to construct indices for some Spanish regions, but they do not use an homogeneus methodology and the indices are not directly comparable.

In this paper, we summarize and extent the main results of previous studies about the possibility of using different indirect methods to analyse the short term evolution of regional industrial production. In concrete, two statistic and an econometric method are considered. First, we study the possibility of extending the methodology proposed by the Regional Institute of Statistics of Catalonia (*Institut d'Estadística de Catalunya* -IEC-) to other Spanish Regions. Second, we analyse the relationships between electric energy consumption for industrial purposes and industrial production. Third, following Israilevich and Kuttner (1993), we apply a *state-space* model to obtain estimates of the industrial production indices using the Kalman Filter and the method of maximum likelihood. Next, to validate the indices obtained through these three methods we compare them with regional indices obtained by direct methods for the regions where they exist. Finally, we expose the main conclusions remarking the implications for public policy in relation with elaboration of regional statistics.

1. Introduction

Even though service industries have become more important in developed economies during the last few decades, the evolution of manufacturing activities is still crucial to determine the current economic conditions. The most commonly used data to monitor manufacturing evolution are Industrial Gross Domestic Product (IGDP) and/or Industrial Gross Added Value (IGAV) obtained from the National Accounts. But these data are not available as soon as it would be desirable and that is the reason because manufacturing indices have gained popularity in recent years, not only in the United States but also in Europe. In fact, manufacturing indices have become valuable tools to check regularly how national and regional economies change over short periods of time and, more generally, as leading economic indicators.

Basically, manufacturing indices can be classified in two big categories: *qualitative* and *quantitative*. On one hand, *qualitative* indices are based in businessmen appreciations of their firms economic evolution obtained as the difference between positive and negative answers to specific surveys. The main disadvantage of these indices is that they only offer an approximation to the direction of the economic evolution and they do not provide information about the quantity or the value of produced output in the period considered. Moreover, they are very sensitive to the current phase of the economic cycle as a result of the excessive instability of the *proxy* variable used.

On the other hand, *quantitative* indices are estimates of the real added value of the industries considered. They can be obtained through the use of direct methods, mainly surveys addressed to firms, or using pre-existent information. In both cases, these indices reflect the joint evolution of the quantity and quality of the industrial production excluding the effect of prices. Although they can be used as *proxy* variables of the value of industrial production in a several variety of models (for example, models of regional growth or structural change) and base information for constructing other indices (as activity synthetic indices or quarterly accounts), it is important to remark that they are though to be used as key indicators of the state of the considered manufacturing sector.

The main characteristic of quantitative indices is that they are constructed from numeric data related to the amount of industrial goods produced in a territory during a concrete period of time. Depending on the source of these data, one can distinguish two different methods for elaborating quantitative indices: *direct methods* (if the data used to construct the indices come from a survey addressed to a sample of firms previously defined as representative of the economic structure of the considered territory) and *indirect methods* (if the indices are not based in specific survey data but in pre-existent information related to the variable of interest).

In the case of *direct methods*, the process of collecting survey data necessarily implies to design an appropriate questionnaire and define a sample of products and firms which represents properly the sectorial composition and geographical distribution of the region's industrial output. It is very important to choose properly products and firms that represent better the relevant sectors of the economy and, consequently, the method used to define the sample should not be completely random. Under this assumption, big and medium firms are preferred to be included in the sample to small ones because the first are thought to be more precise in providing the required information and also they do it more quickly. Moreover, it is desirable that the sample remains stable for a long period of time and small firms have a shorter life than big ones. In this context, the selection of the appropriate sample implies to define, in a sequential way, small sectorial and geographic populations as a guide to choose the most important *J* products and *N* firms in the economy. Without any doubt, this method provides the best quantitative indices but it has several disadvantages: *a)* the costs of using this method are very high as a result of the process of designing the survey, selecting the sample and collecting and treating of the data; *b)* the selected sample and the chosen base year loose validity as the industrial and the geographical structure of the economy changes over time introducing a systematic bias in the considered index; and, *c)* when the base year or the sample is changed, the new index is not comparable with the previous one and it is required an additional effort to connect the two series.

On the other hand, the main advantages of *indirect methods* are their low cost in relation to direct ones and their good performance in terms of forecasting accuracy if the researcher is able to find *proxy* variables closely related to manufacturing production.

In this paper, we summarise and extent the main results of previous studies about the possibility of using different indirect methods to analyse the short term evolution of regional industrial production in the Spanish regions (Clar *et al.*, 1997a, 1997b and 1998). First, a brief review is made on different statistics sources available in Spain including National Accounts (annual and quaterly), Regional Accounts, the different sinthetic indicators available at national and regional level -IPI, ICI and ISAEC- and other simple indicators. Second, we study the Spanish experiences on the elaboration of indirect quantitative indicators to monitor the evolution of regional industrial activity. In particular, two methodologies are considered: one based on electric energy consumption and the other based on pre-existent information, studying the possibility of extending it to other Spanish regions. And third, following Israilevich and Kuttner (1993), we apply a *state-space* model to obtain estimates of the growth of rate of industrial production indices using the Kalman Filter for the regions with indices obtained by direct methods. Finally, we expose the main conclusions remarking the implications for public policy in relation with elaboration of regional statistics.

2. Available indicators for the Spanish regions

The Spanish National Institute of Statistics (INE) offers two different National Accounts: at annual and quarterly frequencies. However, both kind of data are published with a considerable delay. For example, the INE presents four types of estimates before publishing the definitive data for quarter *t* (Muñoz *et al*., 1996): first estimate (*t*+80 days), preliminary estimate (*t*+8 months), provisional estimate (*t*+20 months), second provisional estimate (*t*+32 months) and definitive estimate (*t*+44 months). Althoug this delay is considerable, the situation is similar in the rest of OECD countries[1]. Spanish Regional Accounts also suffer from similar delays: the last definitive data belongs to 1991, there are provisional data for 1992 and 1993 and a preliminary estimate for 1994.

To solve this problem, to monitor the short term evolution of economic activity at a national level, two sinthetics indexes are elaborated: the Industrial Production Index (IPI) and the Industrial Climate Index (ICI). The IPI, elaborated by the INE, is a monthly quantitative index, that it is obtained from a specific survey addressed to nearly 9000 industrial establishments. Nowadays the base year of the index is 1990 and so the weights used to combine the different sectorial indexes reflect the productive structure of this year. The historical series with base

1990 start in 1975. The ICI, elaborated by MINER (*Ministerio de Industria y Energía*), is obtained from the information contained in a monthly survey of businessmen appreciations. So, it is a qualitative direct index that can be used to analyse the evolution of the economic activity at national and regional level. In particular, the applied methodology consists in combining information about businessmen expectatives (+, - or =) and the evolution of commands and finished products stock in the following way (EC Comission, 1991): *a)* first, the series of businessmen opinions is deseasonalised obtaining the trend; *b)* next, a sinthetic indicator is built as the arithmetic average of the three series; and, *c)* the obtained serie is filtered using an autoregressive of order two and power 0.5 in twenty months, *AR*(2)20. At a regional level, the methodology for obtaining ICIs is similar, although the first step is usually omitted[2].

Apart from these indicators, there are other sinthetic indexes for some Spanish regions. As an example, the *Cambra de Comerç, Indústria i Navegació de Barcelona* (COCINB) publishes every two months a sinthetic index for *Catalunya* (ISAEC), which is a quantitative index, also available for industry, construction, and services, elaborated following four steps[3]: *a)* partial indicators for the Spanish and Catalan economy are chosen (those indicators available at a regional level that reflect better the evolution of the economic activity; *b)* estimation of the trend-cycle signal of theses partial indicators applying the LAM filter [INE (1983); Cristóbal and Quilis (1995)]; *c)* trimestralization of sectorial GAV using the same method as INE (Chow-Lin) [Suriñach *et al.* (1996); Artís *et al.* (1997c)]; and, *d)*multivariate techniques are applied to combine all the information.

However, to analyse industrial activity at a regional level is more difficult as a consequence of the nearly absence of survey data (due mainly to the associated high costs) that will permit this kind of analysis. As an alternative, during the last years, there have been some attempts to elaborate regional industrial production indices from indirect data. In this sense, during the last years the efforts of regional governments, regional institutes of statistics and private institutions have been important but still insufficient: only nine of the seventeen Spanish regions at NUTS-II level have their own quantitative manufacturing index: *Andalucía*, *Asturias*, *Baleares*, *Canarias*, *Catalunya*, *Euskadi*, *Madrid*, *Navarra,* and *La Rioja* and they do not use the same methodology. *Andalucía*, *Asturias* and *Euskadi*'s indices are constructed using direct methods while the rest of these regions use indirect methods. *Baleares, Canarias,* *Madrid*, *Navarra* and *La Rioja* use the electric energy consumption for industrial purposes as base information to elaborate their quantitative manufacturing index. In *Catalunya*, the IEC uses the national IPI as the starting point to elaborate their regional quantitative manufacturing index. As a result, there are difficulties to compare the different Spanish regional quantitative manufacturing indices: the methodology, the number of branches considered, the starting date, the base year, ..., are different.

3. Experiences in elaborating indirect indicators to analyse the evolution of regional industrial activity in Spain

In Spain nowadays, basically two methodologies are applied to obtain indirect indicators of the regional industrial activity. One methodology is based on taking electric energy consumption for industrial purposes (provided by electric companies for 29 industrial branches) as a *proxy* of industrial production. The other, applied by the IEC to the catalan region, uses the same information that INE for the whole nation: the national IPIs at the maximum industrial aggregation level. These information is censored to make it representative of the region and weighted in function of the catalan productive structure.

3.1. The relationships between electric energy consumption and industrial production

In Spain during the seventies and eighties, some studies [Sanz (1979), Molina and Sanz (1985)] proposed to use the electric energy consumption for industrial purposes as a *proxy* variable to predict the short term evolution of the industrial production. Nowadays, *Andalucía*, *Baleares, Canarias*, *Madrid*, *Navarra* and *La Rioja* use this variable as an indicator of the regional industrial production. This methodology has the advantage of its low costs. However, the adequacy of this strategy depends on the validity of the following hypothesis: *a)* the industrial production is directly related with electric energy consumption in every industry; *b)* the production function is stable, at least at short term (this fact implies that the ratio electric energy by unit of output is constant over time and so, the technologic level does not change); and, *c)* as in other methods based on Laspeyres indices, it is supposed that relative weights of every sector are the same as in the base year. The main disadvantages of this methodology are due to the process of collecting data from electric companies: they use a sectorial classification different to the usual, so an additional effort is required to make them compatible as there are problems to calculate the proper relative weights of every considered branch. In fact, if they are inadequate, a systematic bias can be introduced as energy-intensive branches would be overweighted.

At a regional level, and in spite of its low cost, the obtained indices do not reflect the real evolution of industrial production as a consequence of the lack of validity of some hypotheses. In this sense, the liability of the indicators could be improved incorporating information about consumption of other energetic sources and some measure of efficiency, for example the elasticity output/energy.

3.2. The IEC methodology

In a previous work (Clar *et al.*, 1997a, 1997b and 1998), we have extended the indirect method proposed by the IEC to other Spanish regions with the aim that the obtained indices will be directly comparable providing information about the evolution of the regional economies. The starting point of the IEC methodology is that, as a result of its elaboration process, the national general IPI can be expressed in the following terms:

$$IPI = \sum_{i=1}^{N} IPI_i \alpha_i \ , \qquad (1)$$

where $\alpha_i = X_i / X$ 0is the relative weight of the branch i in the total national output and IPI_i represents the IPI of the branch i for every one of the N considered branches. Applying the same process, a similar expression to (1) can be obtained for the industrial production index of the region j:

$$IR_j = \sum_{i=1}^{N} IR_{ji} \alpha_{ji} \ , \qquad (2)$$

where IR_j is the general industrial production index of the region j, IR_{ji} is the industrial production index of the branch i in region j and $\alpha_{ji} = X_{ji} / X_j$ is the relative weight of the branch i for every branch of the N considered branches.

The idea behind IEC's methodology is that national sectorial industrial production indices depend on sectorial regional indicators IR_{ji} as a result of the elaboration process of the national indices. In fact, the relationship between sectorial IPIs and sectorial regional indicators IR_{ji} for the 17 Spanish regions can be approximated, under certain restrictive assumptions, as follows:

$$IPI_i = \sum_{j=1}^{17} IR_{ji}\mu_{ji} , \qquad (3)$$

where μ_{ji} is the relative weight of branch i in region j in the total production of branch i at a national level $\mu_{ji} = X_{ji} / X_j$.

The results of comparing the original data with the indices constructed applying (3) for *Andalucía*, *Asturias* and *Euskadi* (regions that have their own survey to elaborate industrial production indices by direct methods) show that this methodology provide a good approximation, at least at yearly frequencies, to the real evolution of the regional production as it can be seen in table 1, where the values of *MAPE*[4] (Mean Average Percentual Error) are shown.

Table 1. Values of *MAPE* for the indicators elaborated using the IEC method

IEC IPI	***MAPE***			***Sample***
	Month	Quarter	Year	
Andalucía	5.67%	4.63%	3.36%	1/86-12/96
Asturias	4.32%	2.33%	1.29%	1/90-12/96
Euskadi	6.83%	2.54%	0.67%	1/86-12/96

However, the results are not completely satisfactory. In this sense, it is important to remark that the liability of the proposed indices for a concrete region depends on: *a)* the degree of geographical concentration of the manufacturing production; *b)* the aggregation level of the base information, the weight of the regional manufacturing production in the national one; *c)* the similarity of the regional productive structure to the national one; and, *d)* the stability of the productive structure. The proposed methodology is, then, justified for the case of Catalunya (see Clar *et al.*, 1998), but the liability of the obtained indices for other regions can only be guaranteed at yearly (and quarterly) frequencies.

4. A latent variable model to measure regional manufacturing production

4.1. The Israilevich and Kuttner's model[5]

Israilevich and Kuttner (1993) propose an alternative method for estimating a monthly regional production model. Their model involves treating the regional industrial production index as a latent variable[6], which depends on capital (*proxied* by electric energy consumption for industrial purposes[7]) and labour. With their model they try to solve one problem which is common to some indirect methods (for example, the IEC's method): the estimates of the regional production index rely entirely on the validity of the national index of industrial production as a perfect indicator of regional output. Another important feature of the model is its imposition of consistency between the estimated monthly index and the observed (or predicted) annual GDP or GAV series.

The starting point of the model is a regional monthly production function. Following much of the existing literature, they propose to use a first-differenced logarithmic version of a Cobb-Douglas specification:

$$\Delta X_{t,s} = \gamma + \phi \cdot \Delta e_{t,s} + \theta \cdot \Delta l_{t,s} + \eta_{t,s}, \tag{4}$$

where $X_{t,s}$ represents the industrial production of region j in month s of year t, $e_{t,s}$ is the electric energy consumption for industrial purposes and $l_{t,s}$ is the number of worked hours in the region during the considered period. γ, ϕ, and θ denote the parameters of the production function.

As regional fluctuations are correlated with national fluctuations, the national industrial production index can serve as a useful indirect measure of the regional industrial activity. In fact, the national index can be seen as a "noisy" indicator of regional activity:

$$\Delta X_{t,s}^{nac} = \mu + \delta \cdot \Delta X_{t,s} + v_t, \tag{5}$$

where a bigger value of δ implies a bigger correlation between national and regional fluctuations.

Finally, the imposition of consistency with the annual estimates implies that:

$$\Delta X_t^{anual} = \frac{1}{12} \cdot \sum_{s=1}^{12} \sum_{j=0}^{11} \Delta X_{t,s-j} \ . \qquad (6)$$

The relationships expressed in equations (4) to (6) can be put together (facilitating its joint estimation) using a *state-space* form.

Many conventional dynamic models can be easily written in a *state-space* form. The *state-space* form offers a more flexible way of treating the identification and estimation of dynamic models and this is the reason why they have been widely used by economists in the last years.

The general *state space* form applies to a univariate or multivariate time series denoted by (Y_1, Y_2, ..., Y_{t-1}, Y_t)[8]. The idea behind *state-space* models is that these observable variables Y_t, subjected to noise denoted by ε_t, capture the dynamics of an unobservable variable or variables α_t, known as state vector. The main objective of expressing a model in its *state-space* form is to obtain information about the behaviour of α_t taking into account its relationship with Y_t.

The relationship between Y_t, an $nx1$ vector, and α_t, an $mx1$ vector, is supposed to be linear and it is expressed by the following equation, known as measurement equation:

$$Y_t = Z_t \cdot \alpha_t + d_t + \varepsilon_t \ , \qquad (7)$$

where Z_t is an nxm matrix, d_t is an exogenous variables $nx1$ vector and ε_t is an $nx1$ vector of serially uncorrelated disturbances with mean zero and known covariance matrix H_t: $\varepsilon_t \sim Niid(0_{nx1}, H_{nxn})$.

Although, in general, the elements of α_t are not observable, they are known to be generated by a first-order Markov process:

$$\alpha_t = T_t \cdot \alpha_{t-1} + c_t + R_t \cdot \theta_t \ , \qquad (8)$$

where T_t is an mxm matrix, c_t is an $mx1$ vector of exogenous variables which influence α_t, R_t is an mxg matrix and θ_t is a $gx1$ vector of serially uncorrelated disturbances with mean zero and covariance matrix Q_t: $\theta_t \sim Niid(0_{mx1}, Q_{mxm})$.

Equation (8) is known as transition equation and together with the measurement equation, they form the *state-space* model. Z_t, d_t, H_t, T_t, c_t, R_t and Q_t are known as system matrix.

The specification of the *state-space* model is completed by two further assumptions concerning the initial state vector values and the covariance matrix of the disturbances:

$$\begin{gathered} E(\alpha_0) = a_0\,; \\ var(\alpha_0) = P_0\,; \end{gathered} \tag{9}$$

$$E(\varepsilon_t \cdot \theta'_s) = 0 \quad s,t = 1,\ldots,T\,; \tag{10}$$

$$\begin{gathered} E(\varepsilon_t \cdot \alpha_0) = 0 \quad t = 1,\ldots,T; \\ E(\theta_t \cdot \alpha_0) = 0 \quad t = 1,\ldots,T. \end{gathered} \tag{11}$$

The definition of α_t depends on the characteristics of the system considered, but it is important to remark that usually there is more than one possible *state-space* form for every system, being preferred models with a lower number of parameters.

The main difficulty of applying *state-space* models in a economic framework is due to the dependence of the system parameters Z_t, H_t, T_t, R_t and Q_t on unknown parameters. These parameters are known as hyperparameters and they determine the stochastic properties of the model. In Engineering, hyperparameters are usually known but in Economics an additional effort is required to obtain estimates of their values.

A possible expression of equations (4), (5) and (6) in the *state-space* form is shown in equations (12) and (13).

$$\begin{pmatrix} \Delta X_t^{anual} \\ \Delta X_{t,12}^{nac} \\ \Delta X_{t,11}^{nac} \\ \dots \\ \Delta X_{t,1}^{nac} \end{pmatrix} = \begin{pmatrix} 1/12 & 2/12 & \dots & 12/12 & 11/12 & \dots & 0 \\ \delta & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \delta & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & \delta & 0 & \dots & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta X_{t,12} \\ \dots \\ \Delta X_{t,1} \\ \Delta X_{t-1,12} \\ \dots \\ \Delta X_{t-1,1} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \dots & 0 \\ \mu & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \mu & 0 & \dots & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \Delta e_{t,12} \\ \dots \\ \Delta e_{t,1} \\ \Delta l_{t,12} \\ \dots \\ \Delta l_{t,1} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} v_{t,12} \\ \dots \\ v_{t,1} \end{pmatrix}.$$

Measurement equation (12)

$$\begin{pmatrix} \Delta X_{t,12} \\ \dots \\ \Delta X_{t,1} \\ \Delta X_{t-1,12} \\ \dots \\ \Delta X_{t-1,1} \end{pmatrix} = \begin{pmatrix} 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 \\ 1 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \Delta X_{t-1,12} \\ \dots \\ \Delta X_{t-1,1} \\ \Delta X_{t-2,12} \\ \dots \\ \Delta X_{t-2,1} \end{pmatrix} + \begin{pmatrix} \gamma & \phi & \dots & 0 & \theta & \dots & 0 \\ \gamma & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \gamma & 0 & \dots & \phi & 0 & \dots & \theta \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \Delta e_{t,12} \\ \Delta e_{t,11} \\ \dots \\ \Delta e_{t,1} \\ \Delta l_{t,12} \\ \Delta l_{t,11} \\ \dots \\ \Delta l_{t,1} \end{pmatrix} + \begin{pmatrix} 1 & \dots & 0 \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 1 \\ 0 & \dots & 0 \\ 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \end{pmatrix} \cdot \begin{pmatrix} \eta_{t,12} \\ \eta_{t,11} \\ \dots \\ \eta_{t,1} \end{pmatrix}.$$

Transition equation (13)

4.2. Estimation and validation of the results

4.2.1. The Kalman Filter

The estimation of this model can be dealt easily using the Kalman Filter. The Kalman filter is a recursive procedure for computing the optimal estimates of the state vector at time t, using information available at time t-1 and updating these estimates as additional information becomes available. To apply the Kalman filter, the considered model must be expressed in the *state-space* form. In fact, like in most economic applications, here the Kalman filter is seen as a tool to obtain estimates of latent variables using information from related observable variables.

The Kalman filter, proposed originally by Kalman (1960) and Kalman and Bucy (1961), is formed by two sets of equations which are applied sequentially. The first stage of the estimation procedure consists in obtaining the optimal predictor of the next observation of the state vector (time t) using all the available information (until t-1). The second step updates this predictor incorporating the additional information available at time t.

The first step for the *state-space* model proposed in equations (7) to (11) can be summarised as follows. Let a_{t-1} denote the optimal estimator of α_{t-1} based on the observations up to and including Y_{t-1}. The associated *mxm* error estimation covariance matrix P_{t-1} is given by:

$$P_{t-1} = E\left[(\alpha_{t-1} - a_{t-1}) \cdot (\alpha_{t-1} - a_{t-1})'\right]. \tag{14}$$

Once a_{t-1} and P_{t-1} are known, the optimal estimator of α_t restricted to these values is given by:

$$a_{t/t-1} = \alpha_t \; / \; a_{t-1} = E(\alpha_t \; / \; a_{t-1}) = E(T_t \cdot a_{t-1} + c_t + R_t \cdot \theta_t) = T_t \cdot a_{t-1} + c_t \, , \tag{15}$$

with a covariance matrix of the estimation errors equal to:

$$P_{t/t\text{-}1} = E\left[(\alpha_t - a_t \; / \; a_{t-1}) \cdot (\alpha_t - a_t \; / \; a_{t-1})'\right] = T_t \cdot P_{t\text{-}1} \cdot T_t' + R_t \cdot Q_t \cdot R_t' \, . \tag{16}$$

Equations (15) and (16) form the first step of the estimation procedure and they are known as prediction equations.

The second step starts once the observation Y_t is available and the estimator of α_t, $a_{t/t-1}$ can be updated. The equations that summarise this second stage are known as updating equations and are given by the following expressions:

$$a_t = a_{t/t} = a_{t/t-1} + P_{t/t-1} \cdot Z_t' \cdot F_t^{-1} \cdot (Y_t - Z_t \cdot a_{t/t-1} - d_t); \tag{17}$$

$$P_t = P_{t/t-1} - P_{t/t-1} \cdot Z_t' \cdot F_t^{-1} \cdot Z_t \cdot P_{t/t-1} \, ; \tag{18}$$

$$F_t = Z_t \cdot P_{t/t-1} \cdot Z_t' + H_t \, . \tag{19}$$

Taken together equations (16) to (19) form the Kalman filter. But before applying these equations, in most cases, it is necessary to obtain estimates of the hyperparameters (in equations

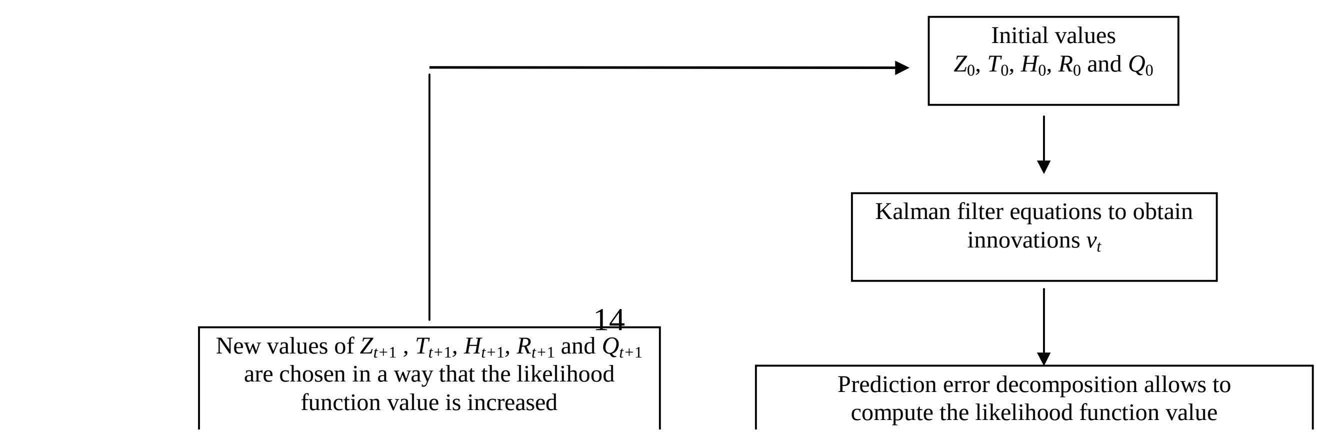
(12) and (13): γ, ϕ, θ, μ, δ and the variance of their error terms) and the initial values of a_0 or $a_{1/0}$ and the associated covariance matrix P_0 or $P_{1/0}$.

4.2.2. Hyperparameters and initial values

The classical theory of maximum likelihood estimation can be adapted to obtain estimates of the hyperparameters, although the expression of the likelihood function is usually too complicated to obtain the desired values from its analytical expression. These difficulty can be easily dealt using the numerical expression of the likelihood function and numerical optimisation procedures to find its maximum value. The maximum likelihood estimation of the hyperparameters can be summarised in four steps (see figure 1): *a)* the first step involves analysing whether the model is identified and determining the unknown hyperparameters that must be estimated: Z_t, T_t, H_t, R_t and Q_t. Initial values for these parameters are chosen; *b)* using these initial values, the Kalman filter equations are applied to obtain the innovations values v_t (one-period forecast errors) conditioned to the initial values of the hyperparameters; *c)* following the prediction error decomposition (Harvey, 1989), the value of the likelihood function can be obtained from the innovation values and using a numerical optimisation procedure, we can determine if its value is a maximum or not; and, *d)* in case the value is a maximum, the estimation procedure of hyperparameters will be finished. If the value is not a maximum, the chosen optimisation procedure will proportionate new initial values for the hyperparameters coming back again to step *b)*.

Another procedure for estimating the values of hyperparameters involves the application of the *EM* algorithm, first developed by Dempster *et al.* (1977) and introduced in this framework by Shumway and Stoffer (1982) and Watson and Engle (1983).

Figure 1. Maximum likelihood estimation of the hyperparameters



Adapted from Cuthberson *et al.* (1992), p. 214.

One of the main critiques against the application of *state-space* models in Economics is related with the instability associated to both procedures. In this sense, Hackl and Westlund (1996) found that the outcome of Kalman filtering is very sensitive to the specification of the *state-space* model. Their conclusion is that extreme care must be exercised when applying Kalman filtering and a reasonable protection in many cases consists in specifying the model as simply as possible and to choose a priori specified hyperparameters instead of estimating them simultaneously with the state vector.

For the case of the considered model for the Spanish regions (*Andalucía*, *Asturias* and *Euskadi*), data for the number of worked hours in manufacturing is not available at a monthly basis[9], so we have used as a *proxy* of labour input the number of industrial workers in the region which form part of the General Social Security System. The regional electric energy consumption for industrial purposes data was only available from January 1993 to December 1996. As a result of the short number of available observations (36 after differenciating), the maximum likelihood approach was not appropriate to obtain estimates of the hyperparameters. This is the reason why, following the approximation proposed by Hackl and

Westlund (1996), we have used a priori information (panel data) to estimate hyperparameters for the considered regions. Results are summarised on table 2.

Table 2. Estimates of the hyperparametres using additional information (annual data)

Andalucía		***Asturias***		***Euskadi***		
γ	-0.82	γ	-0.49	γ	-0.77	$X_{jt}=\gamma_j+\phi\cdot K_{jt}+\theta\cdot L_{jt}$
ϕ	0.38	ϕ	0.38	ϕ	0.38	R^2=0.92
θ	0.69	θ	0.69	θ	0.69	
μ	1.58	μ	4.85	μ	4.01	$GAV_{Esp}=\mu_j+\delta_j\cdot GAV_j$
δ	0.82	δ	0.68	δ	0.65	R^2_{And}=0.81; R^2_{Ast}=0.72; R^2_{Eus}=0.72

To solve the problem of the initialisation of the Kalman filter, there are two kind of procedures depending if the state vector is stationary or not. A *state-space* model is stationary if the eigenvalues of the matrix T_t in equation (2) are inside the unit circle and there are enough observations of the considered system to affirm that the model has reached stationarity. In this situation, it is possible to approximate the initial values of the state vector from the unconditional mean of the considered process. Following Harvey (1981), these values can be obtained using the first m available observations to estimate the equation (1) by *OLS* and starting the Kalman filter at time m+1. Another alternative consists in considering the initial values as unknown hyperparameters and estimate them by any of the procedures described in the section before. However, when the model is not stationary, the initial conditions are not well defined and the previous solutions can not be applied. The most usual solution in this case consists in treating the initial conditions as diffuse introducing complementary equations to the usual Kalman filter. In the literature, different ways of introducing these equations have been proposed [Harvey and Phillips (1979); Anderson and Moore (1979)-Kitagawa and Gersch (1984); Harvey (1989); Ansley and Kohn (1989)-de Jong (1991)-Kohn and Ansley (1986)], although none of them has found to be completely satisfactory. Most research on Kalman filtering and *state-space* models focuses on solving this problem.

In this application, the problem of unknown initial values was solved using the approximation proposed by Harvey (1984), which forms part of the first approach. Once the hyperparameters have been estimated and the problem of initial values has been solved, it is straight forward to obtain estimates of the growth rates of the regional industrial production indices and their values. Results are shown in figures 2 to 4.

Figures 2a and 2b.
Evolution of the Industrial Production-*Andalucía* (quarterly index and growth rates)

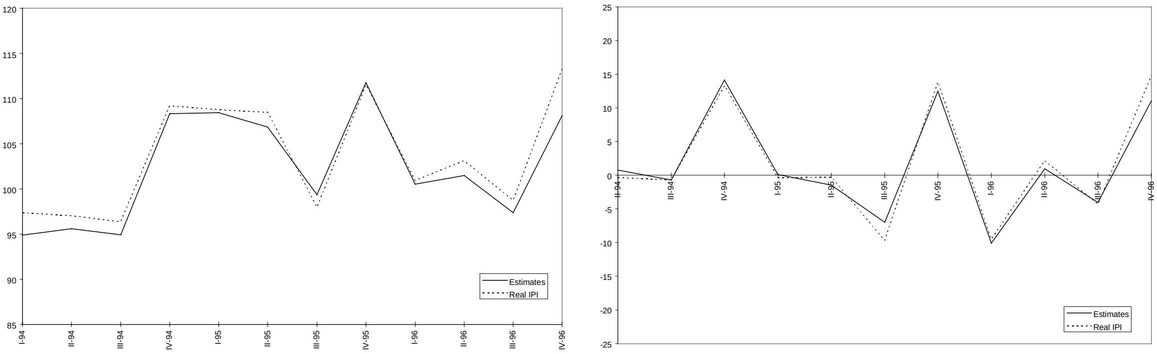
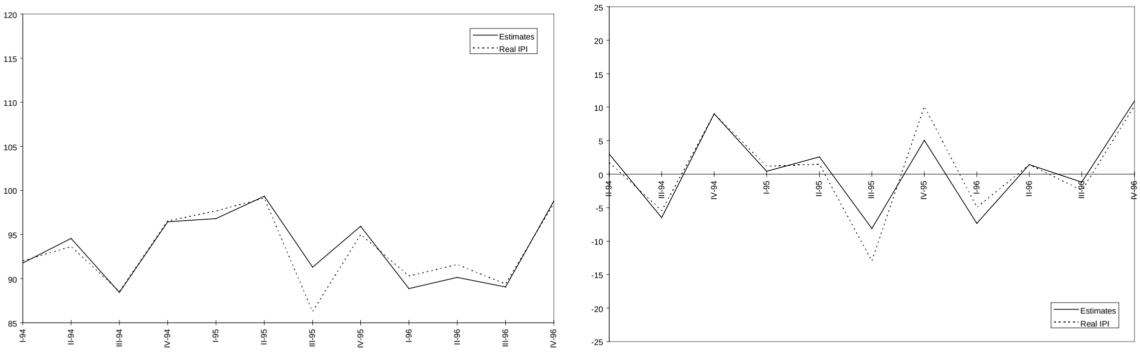
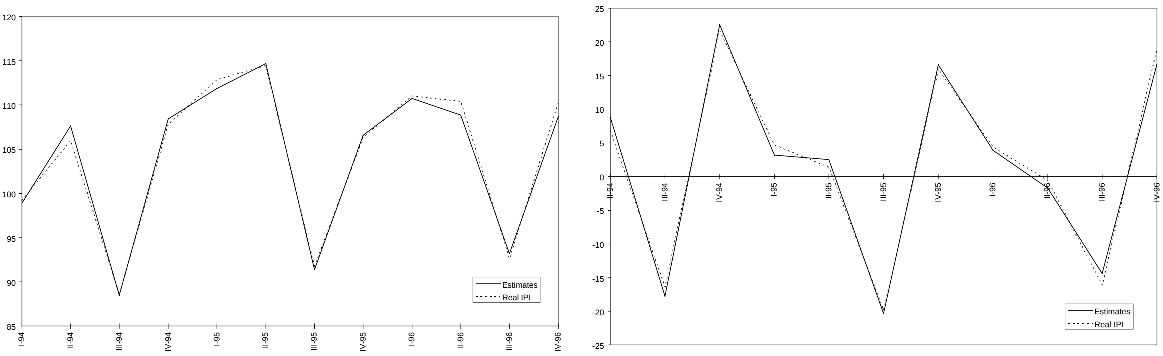


Figure 3a and 3b
Evolution of the Industrial Production-*Asturias* (quarterly index and growth rates)



Figures 4a and 4b
Evolution of the Industrial Production-*Euskadi* (quarterly index and growth rates)



As it is shown in the figures, the obtained results provide a good approximation to the evolution of the indices elaborated by direct methods, at least at a quarterly frequency. As an additional way to validate their results, Israilevich and Kuttner also propose to calculate a measure of how informative are national fluctuations about regional ones. They called this measure *pseudo*-R^2 (due to its similarity with the usual R^2 in linear regression models) and it can be computed using the following expression:

$$pseudo - R^2 = 1 - \frac{\sigma_v}{var(\Delta X_{t,s}^{nac})}. \qquad (20)$$

The values of the *pseudo*-R^2 for the considered regions together with the values of *MAPE*, which show that the adjustment is satisfactory, can be found in table 3. In table 3, *MAPE* values are shown obtained when applying the IEC's methodology for the sample considered here. Comparing the values obtained with both methodologies, it seems clear that the methodology based on *state-space* models works better than the IEC's one for these regions.

Table 3. Values of the *pseudo-R*^{2} and *MAPE*

S-S IPI	***pseudo-R*^{2}**	***MAPE***		
1994-1996		Month	Quarter	Year
Andalucía	0.66	4.89%	1.48%	1.24%
Asturias	0.48	3.10%	1.09%	0.77%
Euskadi	0.50	6.34%	0.66%	0.45%

IEC IPI	***MAPE***		
1994-1996	Month	Quarter	Year
Andalucía	8.52%	7.16%	4.94%
Asturias	5.04%	3.43%	1.84%
Euskadi	5.74%	1.48%	1.14%

However, a comment must be done. The values of *MAPE* in monthly terms are still too high. Three reasons can be accounted to explain this fact. First, it should be remembered here that as the number of worked hours is not available at a regional level, we have used the number of affiliated workers which does not reflect seasonal fluctuations in worked hours due to holidays. A possible way to obtain better results, could be to correct the number of affiliated workers in function of their different preferences for holidays. Second, the short number of observations available affects negatively the performance of the applied methodology. Better results would be obtained incorporating new observations when available. Third, in this period the Spanish Economy has suffered a short, but deep, crisis and the beginning of an expansive phase which makes more difficult the estimation of these indices (*MAPE* values applying the IEC's methodology are considerably higher in this period than in the full sample, compare table 3 and table 1).

5. Conclusions

During the last years, some public and private regional institutions have published different indices to monitor the short term evolution of the manufacturing production in some Spanish regions. The main problem with these indices is that they are not strictly comparable as they are elaborated using different methodologies.

The main objective of this paper has been to compare the relative performance of three indirect methods to obtain homogeneus indices of the evolution of regional manufacturing in the Spanish regions. The main advantage of using indirect methods is their relative low cost as they used pre-existent information. The analysis has shown that indirect methods provide a good approximation to the evolution of regional manufacturing production at quarterly and yearly frequencies, but there are problems in their application at a monthly frequency due to defficiencies on available data at a regional level. The implication of these results in relation with the elaboration of regional statistics is that an efficient allocation of the available resources to elaborate statistics would better imply an improvement of the pre-existent data instead of developing new high-cost statistics.

6. Final notes

[1] The most common pattern in this countries is to release their initial accounts a little more than two months after the reference quarter, to revise them during that year and continue revising them on an annual basis for two to four years (Smith, 1993).

[2] For more details, see Cordero *et al.* (1996), Aranda, D. *et al.* (1994) and EC Comission (1991).

[3] For more details, see Artís *et al.* (1994, 1997a and 1997b).

[4] $MAPE = {}^{1}\!/_{T} * \sum_{t=1}^{T} \frac{|I_t - \hat{I}_t|}{I_t} * 100$ where I_t is the value of real IPI and $\hat{I}_t$ is its estimate.

[5] The results in this section have been obtained using SAS 6.12 software. We would like to thank Philip Israilevich for facilitating us part of the GAUSS code used in their computations.

[6] The econometric technique used here to deal with latent variables is based in models in *state-space* form and its estimation by means of the Kalman filter. See Harvey (1982, 1987), Engle and Watson (1987) or Aoki (1990) for extensive reviews.

[7] See Moody (1974) for more details.

[8] The notation used here strongly follows Harvey (1989).

[9] At regional level these data are only available at quarter frequency.

7. References

- Anderson, B.D. and J.B. Moore (1979): *Optimal Filtering*, Prentice-Hall, Englewoods Cliffs.
- Ansley, C.F. and R. Kohn (1989): ‘Filtering and Smoothing in *State-Space* Models with Partially Diffuse Initial Conditions’, *Journal of Time Series Analysis*, 11, 275-293.
- Aoki, M. (1990): *State Space Modelling of Time Series*, Springer-Verlag, Berlin.
- Aranda, D., A. González and A. Petitibó (1994): ‘Las Encuestas de Opinión Empresariales: un Instrumento para Conocer la Coyuntura Industrial’, *Economía Industrial*, 229, Ministerio de Industria y Energía.
- Artís, M., J. Pons, M.A. Sierra and J. Suriñach (1994): ‘Elaboració d’un Sistema d’Indicadors d’Activitat per a l’Economia Catalana’, *Perspectiva Econòmica de Catalunya*, 176, 83-102.
- Artís, M., J. Pons, M.A. Sierra and J. Suriñach (1997a): ‘Estimación de la Actividad Económica a Corto Plazo Mediante Indicadores de Coyuntura’, *Revista de Economía Aplicada*, 13, 129-147.
- Artís, M., E. Pons, J. Pons and J. Suriñach (1997b): *Evolución Cíclica de las Comunidades Autónomas y Análisis Cíclico*, Escuela de Economía Regional, Universidad Internacional Menéndez Pelayo, Santander.

- Artís, M., E. Pons, J. Pons and J. Suriñach (1997c): *Comptabilitat Econòmica de Catalunya i Mètodes de Trimestralització. Components de la Demanda*, Document de Treball DOCT97R02, Grup d'Anàlisi Quantitativa Regional, Universitat de Barcelona.
- Clar, M., R. Ramos and J. Suriñach (1997a): *A Methodological Proposal for Elaborating Regional Manufacturing Production Indices*, Communication presented at the 37th European Regional Science Association, Rome, 26-29 August.
- Clar, M., Ramos, R. y Suriñach, J. (1997b): *Análisis Comparativo de Métodos Indirectos para la Elaboración de un Indicador de la Actividad Industrial Regional*, Communication presented at the XXIII Reunión de Estudios Regionales, Valencia, 18-21 de November.
- Clar, M., R. Ramos and J. Suriñach (1998): *Anàlisi de la Informació Estadística pel Seguiment de l'Activitat Industrial. Una Proposta per a les Economies Regionals.* Working Paper 98R15, Grup d'Anàlisi Quantitativa Regional, Universitat de Barcelona.
- Cordero, G., A. Galloso, A. Pavón and E. Rodríguez (1996): *Los Indicadores de Clima Industrial Regionales como Instrumento para el Análisis Espacial del Ciclo en la Industria: Metodología y Resultados",* Working Paper SGPR-96002, Dirección General de Planificación, Secretaría de Estado de Hacienda, Ministerio de Economía y Hacienda.
- Costa, A. and J. Galter (1994): 'L'IPPI, un Indicador molt Valuós per Mesurar l'Activitat Industrial Catalana', *Revista d'Indústria*, 2ª etapa, 3, 6-15.
- Cristóbal, A. and E. Quilis (1995): 'Señal de Ciclo-Tendencia frente al Ajuste Estacional en la Contabilidad Nacional Trimestral', *Boletín Trimestral de Coyuntura*, 55, pp. 69-102.
- Cuthberson, K., S. Hall and M. Taylor (1992): *Applied Econometric Techniques*. New York: Phillip Allan.
- de Jong, P. (1991): 'The Diffuse Kalman Filter', *The Annals of Statistics*, 19, 1073-1083.
- Dempster, A., N. Laird and D. Rubin (1977): 'Maximum Likelihood from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society*, Series B, 39, 1-38.
- EC Comission (1991): 'Business and Consumer Survey', *European Economy*, Supplement B, Special Edition.
- Engle, R.F. and M.W. Watson (1987): 'The Kalman Filter: Applications to Forecasting and Rational-Expectations Models', in Bewley, T.F. (ed.), *Advances in Econometrics: Fifth World Congress, Econometric Society Monograph*, vol. 13, 245-283, Cambridge University Press, Cambridge.
- EUSTAT: *Índice de Producción Industrial*, different volumes.
- Gardner, G., A.C. Harvey and G.D.A. Phillips (1980): 'An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering', *Applied Statistics*, 29, 311-312.
- González, M., P. Revilla and P. Rey (1992): 'Los Nuevos Índices de Producción y Precios Industriales', *Situación* 3-4, 109-117.
- Gobierno de Navarra (1992): 'El Índice de Actividad Industrial de la Comunidad de Navarra', *Boletín de Economía de Navarra, 2ª época* 4, 101-113.
- Hackl, P. and A.H. Westlund (1996): 'Demand for International Telecommunication: Time-Varying Price Elasticity', *Journal of Econometrics*, 70, 243-260.
- Hamilton, J.D. (1994a): *Time Series Analysis*, Princeton University Press, Princeton.
- Hamilton, J.D. (1994b): 'State-Space Models', in Engle, R. F. and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, 3039-3080, Elsevier Science B.V, Amsterdam.
- Harvey, A.C. (1981): *The Econometric Analysis of Time Series*, Deddington, Oxford.
- Harvey, A.C. (1984): 'Dynamic Models. The Prediction Error Decomposition and State-Space Models' in Hendry, D.F. and K.F. Wallis (eds.), *Econometrics and Quantitative Econometrics*, Oxford, Basil Blackwell, 37-59.
- Harvey, A.C. (1987): 'Applications of the Kalman Filter in Econometrics', in Bewley, T.F. (ed.), *Advances in Econometrics: Fifth World Congress*, Econometric Society Monograph, vol. 13, 285-313, Cambridge University Press, Cambridge.
- Harvey, A.C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Harvey, A.C. and G.D.A. Phillips (1979): 'Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances', *Biometrika*, 66, 69-58.
- Instituto de Estadística de Andalucía (1997): *Índice de Producción Industrial. Metodología del Cambio de Base y Presentación de Resultados*.
- Institut d'Estadística de Catalunya: *L'Índex de Producció de Productes Industrials*, different volumes.

- Instituto de Estadística de la Comunidad de Madrid (1991): *Índice de Actividad Industrial de la Comunidad de Madrid: Metodología y Resultados,* Working Paper, Comunidad de Madrid, Consejería de Economía.
- Instituto de Estadística de la Comunidad de Madrid: *Boletín de Coyuntura*, different volumes.
- Instituto Nacional de Estadística (1982): *Números Índice de la Producción Industrial Base 1972. Monografía Técnica*.
- Instituto Nacional de Estadística (1983): *Contabilidad Nacional Trimestral de España. Metodología y Series,* Madrid.
- Instituto Nacional de Estadística: *Números Índice de la Producción Industrial*, different volumes.
- Israilevich, P. and K. Kuttner (1993): “A Mixed Frequency Model of Regional Output”, *Journal of Regional Science*, 33, 321-342.
- Junta de Andalucía (1988): *Memoria Técnica y Metodológica del Índice de Producción Industrial de Andalucía*, Consejería de Fomento y Trabajo.
- Kalman, R.E. (1960): ‘A New Approach to Linear Filtering and Prediction Problems’, *Transactions ASME, Journal of Basic Engineering*, 82, 35-45.
- Kalman, R.E. and R.S. Bucy (1961): ‘New Results in Linear Filtering and Prediction Theory’, *Transactions ASME, Journal of Basic Engineering*, 83, 95-108.
- Kitagawa, G. and W. Gersch (1984): ‘A Smoothness Priors-State Space Modelling of Time Series with Trend and Seasonality’, *Journal of the American Statistical Association,* 82, 1032-1063.
- Kohn, R. and C.F. Ansley (1986): ‘Estimation, Prediction and Interpolation for ARIMA Models with Missing Data’, *Journal of the American Statistical Association,* 81, 751-761.
- Molina, A. and R. Sanz (1985): *Un Indicador Mensual del Consumo de Energía Eléctrica para Usos Industriales*, Documento de Trabajo 8510, Banco de España, Servicio de Estudios.
- Moody, C.E. (1974): ‘The Measurement of Capital Services by Electrical Energy’, *Oxford Bulletin of Economics and Statistics*, 36, 42-52.
- Muñoz, J., E. Pons and J. Pons (1996): ‘Les Revisions de les Estimacions de la Comptabilitat Nacional’, *Questiió*, 20.
- Prado, C. (1988): ‘Elaboración de un Índice de Producción y Precios Industriales‘, *Ekonomiaz*, 11, 297-313.
- SADEI (1993): *Índice de Producción Industrial de Asturias. Año 1991,* Principado de Asturias, Consejería de Hacienda, Economía y Planificación.
- Sanz, R. (1979): ’Modelización del Índice de Producción Industrial y su Relación con el Consumo de Energía Eléctrica’, *Cuadernos Económicos del ICE*, 11-12, 227-259.
- Shumway, R.H. and D.S. Stoffer (1982): ‘An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm’, *Journal of Time Series Analysis,* 3, 253-264.
- Smith, P. (1993): ‘The Timeliness of Quaterly Income and Expenditure Accounts: An International Comparison‘, *Australian Economic Indicators,* September.
- Snyder, R.D. and G.R. Saligari (1996): ‘Initialisation of the Kalman Filter with Partially Diffuse Initial Conditions’, *Journal of Time Series Analysis*, 17, 409-424.
- Suriñach, J., E. Pons and J. Pons (1996): *Comptabilitat Econòmica de Catalunya i Mètodes de Trimestralització*, IEC, Barcelona. Also *Revista Econòmica de Catalunya*, 1996, 30, 38-56.
- Watson, M.W. and R.F. Engle (1983): ‘Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression’, *Journal of Econometrics,* 23, 385-400.