Cat. No. KS-AN-01-016-EN-I

A Monthly Indicator of GDP for the Euro-zone

EUROPEAN COMMISSION

eurostat

THEME 1
General statistics

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2001

Printed in Belgium

PRINTED ON WHITE CHLORINE-FREE PAPER

INDEX

INDEX

A Monthly Indicator of GDP for the Euro-zone

Roberto ASTOLFI
2SDA Statistical Studies and Data Analysis
17, Rue des Bains, L-1212 Luxembourg
e-mail:roberto.astolfi@planistat.lu

Dominique LADIRAY
Eurostat / A6
Jean Monnet Building, L-2920 Luxembourg
e-mail:dominique.ladiray@cec.eu.int

Gian Luigi MAZZI
Eurostat / A6
e-mail:gianluigi.mazzi@cec.eu.int

Fabio SARTORI
Eurostat / A6
e-mail:fabio.sartori@cec.eu.int

Regina SOARES
Eurostat / A6
e-mail:regina.soares@cec.eu.int

1. Introduction

In the last few years the interest in short-term statistics has considerably increased. The availability of statistics suitable to give an image of the economy with a short delay and in a reliable way has become one of the main challenges for Eurostat. This attitude justifies, for example, the increasing role played by quarterly national accounts, which give a complete and coherent picture of the economic situation.

In addition to quarterly accounts several monthly statistics are currently available supplying indications of short-term movements in demand, output, prices and wages. Despite of their early availability and higher frequency, these statistics provide only a partial and thus incomplete picture of the economy.

It is well known that the capability of identifying correctly the short-term pattern of an economic phenomenon increases with the frequency of the observations. On the other hand, a higher frequency is often associated with a higher volatility in the series; this is considered the most important limit to the use of high frequency data.

The project of deriving a monthly indicator of GDP gives the opportunity to combine the demand for a short-term indicator of the whole economy and the requirements of an accounting framework. The final objective would be the compilation of a complete set of monthly national accounts, even if the currently available set of monthly statistics seems to be too weak to ensure the necessary basic information for such a compilation.

Eurostat is working in research projects concerning this field. In Eurostat [3] some synthetic guidelines are given to compute the estimation of a monthly indicator of GDP. This paper presents the methodology and the first results of a project for the estimation of a monthly indicator of GDP.

The approach proposed in this paper is based on separate estimates of each output component of national accounts, which are then used to obtain a monthly estimation of GDP. In fact most of the available monthly indicators directly refer to output components. Results are then compared with a direct estimation of monthly GDP by using all information available.

Section 2 is devoted to the presentation of the estimation methodology. Section 3 shows the first results for the Euro-zone monthly GDP. Section 4 presents conclusions and future developments.

2. The methodology

Chow and Lin [2] suggested that the basic approach for the estimation of monthly series out of quarterly ones should rely on a regression analysis between the quarterly dependent variable and the quarterly aggregates of some monthly explanatory variables. The estimated regression coefficients are then used to interpolate the quarterly figures and thus obtain the monthly series. Since monthly estimates need to be consistent with the quarterly known figures, an appropriate least-squares adjustment is then used to ensure this consistency. This approach has been extended by Bournay and Laroque [1], Fernández [4] and Litterman [6] in order to adapt the model to different structures of the error term.

There are two shortcomings of Chow and Lin's method:

- in their approach the quarterly regression equation are expressed in levels; however, most economic regression equations are usually expressed in logarithms to avoid problems of heteroscedasticity and volatility in the data, see Pinheiro and Coimbra [7];
- the method pre-dates much of the work on dynamic modelling and so does not allow any dynamic structure linking the indicator variables to the interpoland; by contrast a dynamic specification is required whenever dependent variables and interpolands are co-integrated, see Gregoir [5].

The proposed approach, based on Salazar, Smith, Weale and Wright [8], deals with both these shortcomings.

Let $y_{t,s}$ denote an unobserved high frequency monthly scalar series of interest, which is to be estimated using the observed low frequency quarterly aggregates y_t. Here the index $t = 1,\ldots,T$ enumerates the different quarters, and $s = 1, 2, 3$ enumerates the different months in the same quarter.

The relation between the unobserved series and the observed aggregates may be expressed in a compact form:

$$y_t = \sum_{s=1}^{3} c_s y_{t,s} \quad . \tag{2.1}$$

The coefficients c_s, $s = 1, 2, 3$, are invariant and known: they only depend upon the nature of the problem and of the variables involved. If $y_{t,s}$ is a flow variable then $c_s = 1$, $s = 1, 2, 3$, (and the problem is in general referred to as "distribution"); if $y_{t,s}$ is a stock variable then $c_s = 0$, $s = 1, 2$, and $c_3 = 1$ (in this case, the problem is called "interpolation"); then if $y_{t,s}$ is an index, $c_s = \frac{1}{3}$, $s = 1, 2, 3$.

By introducing the high frequency lag operator L, that is $Ly_{t,s} = y_{t,s-1}$ and the aggregator polynomial $c(L)$ defined by:

$$c(L) = \sum_{h=1}^{3} c_h L^{3-h} = c_1 L^2 + c_2 L + c_3 \quad ,$$

then (2.1) may be simply written as:

$$y_t = c(L) y_{t,3} \quad .$$

In the high frequency domain a simple dynamic regression model can be written:

$$\alpha(L) f(y_{t,s}) = \mathbf{x}'_{t,s}\beta + \varepsilon_{t,s} \tag{2.2}$$

where $f(y_{t,s})$ denote a general non-linear transformation of the dependent variable (typically the logarithmic one), $\{x^j_{t,s}\}_{j=1}^{k}$ denote k observable monthly explanatory variables and $\alpha(L) = 1 - \Sigma_{i=1}^{p} \alpha_i L^i$ is a lag polynomial of order p.

The high frequency error terms $\varepsilon_{t,s}$ are supposed to be i.i.d. $N(0, \sigma^2_\varepsilon)$. Moreover the term $\mathbf{x}'_{t,s}\beta$ expresses compactly the constant and the distributed lag effect of the $x^j_{t,s}$, that is:

$$\mathbf{x}'_{t,s}\beta = \beta_0 + \sum_{j=1}^{k} \beta_j(L) x^j_{t,s}$$

where:

$$\beta_j(L) = \sum_{i=1}^{q_j} \beta_{ij} L^i$$

are lag polynomials of order q_j, $j = 1,...,k$. The Chow and Lin method can be obtained from (2.2) by substituting $\alpha(L) = 1$ and $f(y_{t,s}) = y_{t,s}$.

Transforming the high frequency regression model (2.2) into a low frequency relation between the variables requires dealing with two problems, namely transforming the high frequency dynamic induced by $\alpha(L)$ into a low frequency dynamic and then dealing with the non-linear transformation $f(y_{t,s})$.

To accomplish the first task it is necessary to factorise $\alpha(L)$ at its inverse roots $\alpha(L) = \Pi_{i=1}^{p}(1-\rho_i L)$ and then to pre-multiply both sides of (2.2) by $\Pi_{i=1}^{p}(1+\rho_i L+\rho_i^2 L^2)$ since:

$$\prod_{i=1}^{p}(1+\rho_i L+\rho_i^2 L^2)\times\prod_{i=1}^{p}(1-\rho_i L) = \prod_{i=1}^{p}(1-\rho_i^3 L^3) \; .$$

We obtain:

$$\prod_{i=1}^{p}(1-\rho_i^3 L^3) f(y_{t,s}) = \prod_{i=1}^{p}(1+\rho_i L+\rho_i^2 L^2)(\boldsymbol{x}'_{t,s}\beta+\varepsilon_{t,s})$$

which may be compactly expressed as:

$$\theta(L^3) f(y_{t,s}) = \gamma(L)(\boldsymbol{x}'_{t,s}\beta+\varepsilon_{t,s}) \tag{2.3}$$

where $\theta(L^3)$ is the polynomial whose inverse roots are $\{\rho_i^3\}_{i=1}^{p}$ and $\gamma(L) = 1+\Sigma_{i=1}^{2p}\gamma_i L^i = \Pi_{i=1}^{p}(1+\rho_i L+\rho_i^2 L^2)$. Pre-multiplying both sides of (2.3) by the aggregator polynomial $c(L)$ leads to a model expressed in the low frequency domain, that is:

$$\theta(L^3)\big(c(L) f(y_{t,3})\big) = \gamma(L)\big(c(L)\boldsymbol{x}'_{t,3}\big)\beta + u_t \tag{2.4}$$

where the error terms u_t are given by:

$$u_t = \gamma(L)\big(c(L)\varepsilon_{t,3}\big)$$

and so are no longer white noise.

The use of a non-linear transformation does not allow, in the general case, to derive the term $c(L)f(y_{t,3})$ in (2.4) from the quarterly aggregates y_t, so we introduce the following approximation:

$$c(L) f(y_{t,3}) \cong c(1) f(\bar{y}_t) \tag{2.5}$$

where $\bar{y}_t$ expresses the monthly average of the quarterly aggregate y_t, that is $\bar{y}_t = y_t / c(1)$, which means $\bar{y}_t = y_t/3$ for a flow variable and $\bar{y}_t = y_t$ for a stock variable or an index. It can be shown that (2.5) is equivalent to a first order approximation of $f(y_{t,s})$ through the mean value theorem, see Salazar, Smith, Weale and Wright [8].

Introducing the approximation (2.5) inside (2.4) we obtain a dynamic regression model in the low frequency domain which is feasible for estimation, namely:

$$\theta(L^3)\big(c(1)f(\bar{y}_t)\big) = \gamma(L)\big(c(L)\boldsymbol{x}'_{t,3}\big)\boldsymbol{\beta} + u_t \quad . \tag{2.6}$$

So the final estimation of the monthly figures $y_{t,s}$ require a two-step procedure:

- a dynamic regression involving (an approximation) of the quarterly aggregates of the dependent variable $c(1)f(\bar{y}_t)$ and the quarterly aggregates of the monthly explanatory variables $c(L)\boldsymbol{x}_{t,3}$ is estimated. The estimation can be obtained *via* a full maximum likelihood approach or *via* an iterative procedure;
- the estimated regression parameters are then used to obtain a first estimate of the monthly interpolands $\hat{y}_{t,s}$; since these $\hat{y}_{t,s}$ are not consistent with the known quarterly figures, that is (2.1) does not hold, this consistency is achieved through a Lagrange multipliers constrained optimisation, for details see Salazar, Smith, Weale and Wright [8].

The general model (2.6) gives rise in practice to three different specifications according to the available monthly information:

- a multiple ECM regression in case of co-integration between the dependent and the explanatory variables;
- a multiple regression in case of no co-integration;
- a general autoregressive model where no monthly indicator is available.

3. Estimation of monthly GDP

In order to achieve our objective we need to define an estimation strategy. As stated in Section 1, our first approach will involve separate estimates of the different output components of national accounts which are then used to obtain a monthly estimation of GDP. An alternative approach will estimate directly the monthly GDP by using all the available information. The first approach should produce more reliable figures, whereas the second one can be computationally simpler. These two approaches will be compared below in this Section.

3.1 Data availability

In order to obtain the best estimate for each economic sector and for the GDP as whole it is important to verify which monthly series are available as significant indicators in the interpolation process. A synthetic analysis of data availability at the Euro-zone level has shown the following results:

- for industry and construction we have good and reliable monthly statistics: respectively the industrial production index and the construction output index. Since the Euro-zone construction output index series was too short for our purposes, a back-recalculation of the series *via* an ECM regression model has been performed;
- for trade and transport services, the industrial production index is still a reliable

indicator since these services are mainly addressed to enterprises. In addition other useful indicators can be found, in principle, in the deflated turn over of the retail trade and new car registration;

– for agriculture, forestry and fishery, financial services and other services including public administration, we were not able to find any relevant monthly indicator.

After this analysis our data set has been completely defined by including, for the period 1991 Q1 to 2000 Q4, quarterly seasonally adjusted values for the GDP, the six main economic sectors and monthly figures concerning industrial production, construction output, deflated turnover of retail trade and new car registration.

3.2 Estimation

For the economic sectors where co-integrated indicators are available, at a first stage an ECM multiple regression model has been estimated including the maximum number of explanatory variables. Then non-significant indicators have been marginalised in order to obtain an optimal specification.

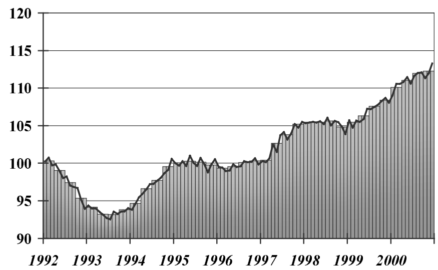
For example, the model for industry has been an ECM regression of the form:

$$\Delta \ln IND_t = \eta_0 + \eta_1 \ln IND_{t-1} + \eta_2 \ln IPI_{t-1} + \eta_3 \Delta \ln IPI_t + \nu_t \tag{3.1}$$

where IND_t is the gross added value of industry in quarter t and IPI_t is the quarterly aggregation of the industrial production index. The specification (3.1) has been used since there is statistical evidence that the available indicators are co-integrated with the corresponding quarterly aggregates. Figure 3.1 shows the output of the estimation of (3.1) and the resulting monthly interpolation of IND_t.

Figure 3.1 Regression output and monthly interpolation of gross added value of industry

Dependent variable is $\Delta \ln IND_t$			
Variable	*coeff.*	*t-val*	*s.e.*
Constant	2.777	3.175	0.875
$\ln IND_{t-1}$	0.705	7.566	0.093
$\ln IPI_{t-1}$	0.203	3.091	0.066
$\Delta \ln IPI_t$	0.724	5.952	0.122
Durbin-Watson	2.099		
R^2	0.8053		
s.e.	0.00453		



As we can see the regression tests diagnostics are rather satisfactory, as it was expected. Interpolated monthly data have a quite regular pattern and they are less volatile than the indicator, which is completely in line with the philosophy of the adopted model.

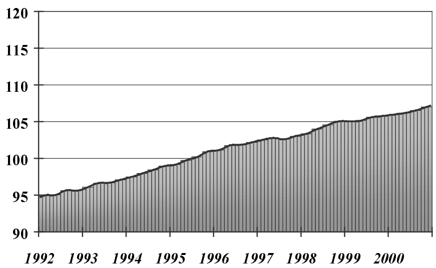
In the case where no indicators were available, a simple auto-regressive model has been fitted to the original series. As an example, the model for other services including public administration has been:

$$\Delta \ln OSER_t = \eta_0 + \eta_1 \Delta \ln OSER_{t-1} + \nu_t \ . \qquad (3.2)$$

Figure 3.2 shows the regression output of (3.2) and the corresponding interpolated series.

Figure 3.2 Regression output and monthly interpolation of gross added value of other services including public administration

Dependent variable is $\Delta \ln OSER_t$			
Variable	*coeff.*	*t-val*	*s.e.*
Constant	0.003	3.941	0.001
$\Delta \ln OSER_{t-1}$	0.275	1.933	0.142
Durbin-Watson	2.366		
R^2	0.09646		
s.e.	0.002315		



Since the quarterly series is quite regular and is not characterised by any significant fluctuation, the lack of additional information on monthly basis can not be considered as a real problem. In addition the share of this aggregate on the GDP is quite small and mainly constant over the time. The same considerations apply to the case of agriculture, forestry and fishery even if in this case the series is less regular. Financial services are the only sector that seems quite problematic for the future. In fact the importance of this aggregate is continuously growing and unexpected shocks can not be captured without any external monthly information.

For the final estimation of monthly GDP the six interpolated series have been used as monthly indicators in an ECM multiple regression of the form:

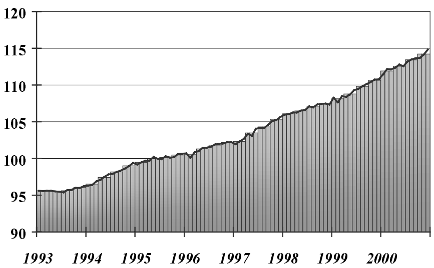
$$\begin{aligned}\Delta \ln GDP_t &= \eta_0 + \eta_1 \ln GDP_{t-1} + \eta_2 \ln AGR_{t-1} + \eta_3 \ln IND_{t-1} + \eta_4 \ln COS_{t-1} + \\ &+ \eta_5 \ln TSER_{t-1} + \eta_6 \ln FSER_{t-1} + \eta_7 \ln OSER_{t-1} + \eta_8 \Delta \ln AGR_t + \\ &+ \eta_9 \Delta \ln IND_t + \eta_{10} \Delta \ln COS_t + \eta_{11} \Delta \ln TSER_t + \eta_{12} \Delta \ln FSER_t + \\ &+ \eta_{13} \Delta \ln OSER_t + \eta_{14} \Delta \ln IND_{t-1} + \eta_{15} \Delta \ln COS_{t-1} + \eta_{16} \Delta \ln TSER_{t-1} + \nu_t\end{aligned} \qquad (3.3)$$

where the dependent variable is the GDP at quarter *t* and the explanatory variables are the quarterly aggregates of the six monthly output components, respectively agriculture, forestry and fishery (AGR_t), industry (IND_t), construction (COS_t), trade and transport services ($TSER_t$) financial services ($FSER_t$) and other services including public administration ($OSER_t$). Figure 3.3 shows the output of the regression (3.3) and the interpolated values for GDP.

Figure 3.3 Regression output and monthly interpolation of gross domestic product estimated trough the six output components

Dependent variable is $\Delta \ln GDP_t$							
Variable	*coeff.*	*t-val*	*s.e.*	*Variable*	*coeff.*	*t-val*	*s.e.*
Constant	2.980	228.230	0.013	$\Delta \ln IND_t$	0.241	74.208	0.003
$\ln GDP_{t-1}$	–0.092	–2.484	0.000	$\Delta \ln COS_t$	0.076	28.267	0.003
$\ln AGR_{t-1}$	0.027	18.023	0.002	$\Delta \ln TSER_t$	0.282	12.871	0.022
$\ln IND_{t-1}$	0.257	332.760	0.001	$\Delta \ln FSER_t$	0.249	20.692	0.012
$\ln COS_{t-1}$	0.070	84.728	0.001	$\Delta \ln OSER_t$	0.209	13.527	0.016
$\ln TSER_{t-1}$	0.228	62.821	0.004	$\Delta \ln IND_{t-1}$	–0.014	–3.671	0.004
$\ln FSER_{t-1}$	0.289	81.844	0.004	$\Delta \ln COS_{t-1}$	–0.005	–5.543	0.001
$\ln OSER_{t-1}$	0.221	129.919	0.002	$\Delta \ln TSER_{t-1}$	–0.079	–3.936	0.020
$\Delta \ln AGR_t$	0.027	8.193	0.003				

Durbin-Watson	1.718
R^2	0.9998
s.e.	0.00007704



The regression output reveals very significant coefficients and the Durbin-Watson test shows no evidence of residual serial correlation. Interpolated values are characterised by a quite regular pattern without any significant abnormal movement.

This is a very satisfactory result since, in producing a monthly estimate of such a crucial variable as the GDP, it is important to avoid any excessive volatility of estimates. Too volatile estimates could create wrong expectations among users, so that the advantage of producing a high frequency indicator will be largely compensated by the negative effect of wrong decisions.

3.3 An alternative approach

As mentioned in Section 3.1, an alternative indicator of monthly GDP could be compiled on the basis of the following ECM multiple regression model:

$$\Delta \ln GDP_t = \eta_0 + \eta_1 \ln GDP_{t-1} + \eta_2 \ln IPI_{t-1} + \eta_3 \ln DTRT_{t-1} + \\ + \eta_4 \ln IR_{t-1} + \eta_5 \Delta \ln IPI_t + \eta_6 \Delta \ln DTRT_t + \eta_7 \Delta \ln IR_t + \eta_8 \Delta \ln IPI_{t-1} + \\ + \eta_9 \Delta \ln DTRT_{t-1} + \eta_{10} \Delta \ln IR_{t-1} + v_t$$

where the explanatory variables are the quarterly aggregates of the industrial production index (IPI_t), the deflated turnover of retail trade ($DTRT_t$) and the interest rate (IR_t).

One of the disadvantages of this approach is that the estimates are, in a certain sense, “outside” the framework of national accounts. Moreover the specificity of the different output component of GDP will be less evident when working at an aggregated level.

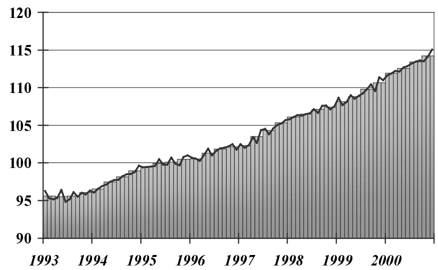
These estimates of the monthly GDP, summarised in Figure 3.4 have been produced in order to compare them with the results of Section 3.2. As in Section 3.2, a general ECM regression model including all available variables has been estimated and then non-significant indicators have been marginalised to obtain the best fitting model.

Figure 3.4 Regression output and monthly interpolation of gross domestic product estimated trough the available monthly indicators

Dependent variable is $\Delta \ln GDP_t$

Variable	*Coeff.*	*t-val*	*s.e.*	*Variable*	*coeff.*	*t-val*	*s.e.*
Constant	1.199	4.970	0.241	$\Delta \ln DTRT_t$	–0.167	–2.608	0.065
$\ln GDP_{t-1}$	0.888	39.801	0.022	$\Delta \ln IR_t$	–0.040	–3.203	0.013
$\ln IPI_{t-1}$	0.053	3.796	0.014	$\Delta \ln IPI_{t-1}$	–0.199	–3.319	0.060
$\ln DTRT_{t-1}$	0.030	2.355	0.013	$\Delta \ln DTRT_{t-1}$	0.187	1.914	0.098
$\ln IR_{t-1}$	–0.003	–3.761	0.001	$\Delta \ln IR_{t-1}$	0.043	2.972	0.015
$\Delta \ln IPI_t$	0.513	7.336	0.070				

Durbin-Watson	2.529
R^2	0.9401
s.e.	0.001471



A simple comparison of the results of the two approaches suggests that the interpolation obtained in Section 3.2 is more smooth and presents a smaller variance. Both series fit quite well the quarterly aggregates even if the second one displays more erratic movements. At this stage we prefer the first approach for the following reasons:

- its lower volatility;
- a more encompassing information for users since the main output components of GDP are also estimated;
- the possibility of extending this approach by taking into account also the demand side.

4. Conclusions

In this paper the Eurostat strategy for the estimation of a monthly indicator of GDP has been synthetically presented. Results seem to be quite encouraging even if more in-depth analysis is still needed.

Further work should deal with the following aspects:

- monthly extrapolation when a complete quarter is still not available;
- dynamic simulation over a given time period to assess the ability of monthly estimates to produce accurate estimations of quarterly figures;
- identification of additional information sources to improve the quality of the estimates in particular in the services domain;
- analysis of the utility of short-term qualitative business and consumer surveys to extrapolate quantitative variables such as the industrial production index or the construction output index, currently used in the estimation process of monthly GDP.

In addition it will be useful to test the impact of seasonal adjustment practices on the interpolation of quarterly national accounts. For example whether the adoption of a direct seasonal adjustment methodology for the Euro-zone quarterly national accounts could improve the quality of the monthly estimates, as it is expected from a theoretical point of view.

Finally it is important to note that, as for all short-term statistics, the usefulness of a monthly GDP is strongly related to the delay of its publication. A detailed analysis has to be made to investigate the impact of using an incomplete or forecasted set of information on the quality and the reliability of the GDP estimates. On the basis of such an analysis, Eurostat will decide if a monthly GDP will be calculated and from when onwards such an indicator will be published.

References

[1] Bournay J., Laroque G., (1979), *Réflexions sur la méthode d'elaboration des comptes trimestriels*, Annales de l'INSEE 36, 3-30.

[2] Chow G., Lin A.L., (1971), *Best linear unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series*, The Review of Economics and Statistics 53(4), 372-375.

[3] Eurostat, (1999), *Handbook on Quarterly National Accounts*.

[4] Fernández R.B., (1981), *A Methodologic Note on the Estimation of Time Series*, The Review of Economics and Statistics 63(3), 471-478.

[5] Gregoir S., (1994), *Note on temporal Disaggregation with simple dynamic Models*, Workshop on Quarterly National Accounts, Paris 5-6 December 1994.

[6] Litterman R.B., (1983), *A random Walk, Markov Model for the Distribution of Time Series*, Journal of Business and Economic Statistics 1(2), 169-173.

[7] Pinheiro M., Coimbra C., (1993), *Distribution and Extrapolation of Time Series by Related Series Using Logarithms and Smoothing Penalities*, Economia v.XVII, 359-374.

[8] Salazar E.,Smith R., Weale M., Wright S., (1994), *Indicators of monthly National Accounts*, Workshop on Quarterly National Accounts, Paris 5-6 December 1994.