UNIVERSIDAD DE SAN ANDRES

Seminario del Departamento de Economía

"Indicators of Monthly National Accounts"

Eduardo L. Salazar

(Instituto Torcuato Di Tella)

Martes 8 de Septiembre
11.00 hs.
Aula Chica de Planta Baja

Indicators of Monthly National Accounts

Eduardo L. Salazar[1,2], Richard J. Smith[3], Martin R. Weale[1] and Stephen Wright[4]
(1) National Institute of Economic and Social Research
2, Dean Trench Street
Smith Square
London SW1P 3HE
(2) Instituto Torcuato Di Tella
Minones 2159
(1428) Buenos Aires
Argentina
(3) Department of Economics
University of Bristol
8, Woodland Road
Bristol BS8 1TN
(4) Faculty of Economics and Politics
University of Cambridge
Sidgwick Ave.,
Cambridge CB3 9DD

August 17, 1998

Abstract

In this paper we look at the derivation of a monthly indicator of GDP, using information from available monthly series which provide an indication of short-term movements in output. Those series, such as industrial production or retail sales, are routinely examined to draw inferences about the state of the economy as a whole. They are, however, used as broad proxies for GDP. We propose instead a formal method based on a dynamic regression model, linking an underlying high-frequency relationship between GDP components and a number of indicator variables. An application to the components of the output measure of GDP for the United Kingdom is presented.

This paper summarises the results of a project initially done for the Office for National Statistics (ONS) of the United Kingdom, and jointly supported by H.M. Treasury. It has been brought to the state described here at which estimates of UK real GDP are produced on a monthly basis, with support from ESRC research grant L116251012. A steering committee has provided very useful feedback during the early part of the work. Richard Clare, Geoff Reed, Keith Vernon and Colin Yeend have been particularly helpful in providing us with both data and suggestions. Any errors remain our own.

Keywords: Interpolation, dynamic models, economic indicators.

JEL Classification: C22, E3.

1 The Motivation for Monthly National Accounts

A range of monthly series are currently available giving indications of short-term movements in output. In the United Kingdom, as in most other countries, these indicators provide only an incomplete picture of the output measure of GDP. However, as the only available information, they are nonetheless already exploited in various ways: financial commentators routinely examine monthly data on retail sales, the trade figures, and the output of the production industries; academic researchers exploiting high frequency econometric techniques make use of one or other of these series as the best available proxy for a broader measure of demand or output.

If these monthly data are to be used to draw inferences about the state of the economy as a whole, then it is desirable that there should be some formal procedure for grossing them up to represent the whole of GDP. Such a procedure is likely to produce estimates of GDP which are less satisfactory than those which might be produced by direct measurement. On the other hand, it is almost certainly more satisfactory than simply making a rough inference from whatever happen to be the latest numbers available.

The Office for National Statistics in the United Kingdom used to publish lagging, co-incident and leading cyclical indicators for the UK. They were calculated by O'Dea (1984) using a variant of the methods developed by the United States' Bureau of Economic Analysis and now maintained there by the US Conference Board. In their earlier stages it was unclear what represented the 'business cycle' but by the end of their life the indicators were meant to represent cyclical movements in real GDP. They were abandoned mainly because the forecasting power of the leading indicator was negligible. It was also the case, however, that the co-incident indicator was constructed by interpolating quarterly GDP; the interpolation was done using industrial production and retail sales as indicators but in a manner which was uninformed by statistical theory; it seemed logical that, if this indicator was to be replaced it should be replaced by an indicator of monthly GDP[1] in constant prices estimated using clearly specified methods. This paper describes a component by component approach to the construction of such an indicator.

[1] Stock & Watson (1989) argue for the use of a latent variable representing 'the state of the cycle' as a monthly indicator of economic activity. They suggest that this has the merit of not being affected by, for example, fluctuations of agricultural output; the latter do influence GDP. We have found, at least in the UK, that data users are familiar with the concept of GDP and rather less so with the idea that the state of the business cycle might be represented by a latent variable. By working with output components it is possible for interested users to look at movements in the output of industry and market services as a means of representing the output of the business sector.

Component	Availability of data/indicators	% of GDP (1990)
Agriculture and Fisheries	Few	1.9
Industry	100 % data availaible	27.7
Construction	Indicators available	7.2
Market Services	Strong correlations with other indicators	42.0
Non-market Services and Ownership of Dwellings	None	21.2

Table 1: Data coverage for monthly constant price accounts (output components)

Regarding the availability of monthly data, the components of the output[2] measure of real GDP fall into four broad categories: series for which data are already available; series for which there are obvious indicators; series which may bear systematic relation to other (possibly not directly related) monthly series; and series for which no monthly information is available. It is questionable whether interpolands of this last group of variables include any extra information. On the other hand if a monthly indicator of the whole of GDP is to be produced, it is necessary to interpolate these variables as well as those for which monthly indicators do exist. And, as we show in table 2, the sectors facing this difficulty (Agriculture and Public Services) contribute very little to the overall variability of GDP.

The published quarterly real GDP data have the property that the components of output do not add up to the total. This is because undisclosed 'other information' is used in addition to the information provided by the output series. This means that, in order to provide monthly series consistent with published GDP, which we denote as GDP(A), or average GDP, we have to allocate the residual between GDP(A) and the estimate of GDP calculated from the output indices alone, which we denote as GDP(O). This is discussed in section 6.

Table 1 shows the breakdown of the main components of output on this basis. However, the importance of each sector to GDP, shown in the last column may be a misleading indicator of the importance of each sector as a contribution to the variance of GDP. Table 2 shows the covariance matrix of the quarter on quarter changes to each component of GDP, with the variables themselves weighted by their share of GDP.

[2]There are two important reasons for using the output measure as our reference point. First of all the ONS regard this as the single best indicator of short-term movements in economic activity. Secondly, the expenditure measure does not offer a satisfactory alternative because there is no obvious means of interpolating changes to inventories which are very volatile.

	Industry	Construction	Agriculture	Non-market Services	Market Services
Industry	0.2825	0.00323	-0.0011	0.0001	0.1190
Construction		0.0322	0.0013	0.0031	0.0415
Agriculture			0.0027	-0.0011	-0.0026
Non-market Services				0.0089	0.0042
Market Services					0.1937

Table 2: The Covariance of Components of Output. This table shows the covariances of quarterly percentage changes weighted by 1990 output weights. Estimation period 1973Q1-1998Q1.

Taking the two tables together, the chance of finding a good monthly indicator of changes to output is higher than the first table suggests on its own. There is no monthly indicator of fluctuations of output by public sector services; despite the fact that they amount to over 21% of GDP, their contribution to the overall variance of GDP is very small, so that the absence of indicators does not matter much. The largest source of variance is industrial production, for which definitive monthly data exist. The second-largest source is market services, for which there are some monthly indicators. Moreover, the covariance between movements in market service output and movements in industry output suggests that monthly industrial production data, in some form or other, are going to be a useful guide to what happens to the market service sector.

The main purpose of this paper is to describe how an indicator of monthly GDP can be constructed; we also present a summary of the interpolation method using indicator variables and an account of the technique we use to complete the picture for those series for which there are no indicator variables present. A full account of the interpolation method using indicator variables, with the results of Monte Carlo tests, is presented by Salazar, Smith & Weale (1996)).

2 Estimation

In practice, there are two reasonably distinctive approaches to the problem of interpolation. The first method relies on the estimation of a regression equation linking low-frequency to high-frequency data. This approach was developed from the early work of Friedman (1962) by Chow & Lin (1971), Ginsburgh (1973) and Fernandez (1981) with the question of the estimation process being considered by Palm & Nijman (1984). It has the attraction that regression equations similar to those used in conventional econometric research and macroeconomic models are estimated. An underlying regression equation is produced explaining the low-frequency data by means of suitable aggregates of high-frequency data. Interpolands of low-frequency data can then be produced by means of the regression equation using coincident information about the high-frequency data. In producing coincident data, of course, extrapolation is required until low frequency data becomes available. Corrado (1986) describes an application of Fernandez's method to the United States' national accounts.

The second approach, suggested by Harvey & Pierse (1984) and Harvey (1989) relies on a state-space model, estimated by means of a Kalman filter. This method is often difficult to apply when there is a large number of possible interpolators, and a regression-based method has the advantage of clarity provided that it can deal adequately with dynamic issues. Regression methods are used by some statistical offices (e.g. in France and Italy) in the construction of quarterly data.

Chow & Lin (1971) suggested that the quarterly estimates of the interpoland should be regressed on the quarterly aggregates of monthly data which would then be used to interpolate the quarterly variable. The regression equation can be used to 'forecast' the interpoland on a monthly basis and least-squares adjustment of the type suggested by Stone, Champernowne & J.E.Meade (1942) is then used to make the monthly forecasts of the interpolands consistent with the known quarterly data. The extension of this method by Fernandez (1981) did not change this basic approach.

There are two shortcomings of Chow and Lin's method. Firstly, their approach relies on the quarterly regression equation being expressed in the levels of the variables of interest and they show how the monthly interpolands can be adjusted so that they add up to the known quarterly totals. Most regression equations, however, are usually expressed in logarithms so as to avoid problems of heteroscedasticity. And obviously, the logarithms of three monthly estimates do not add up to the logarithm of the quarterly estimate.

Secondly, the method, because it pre-dates much of the work which has been done on dynamic modelling, does not accomodate the possibility that there may be some dynamic structure linking the indicator variables to the interpoland. While the technique does not require the assumption that the regression errors are white noise, the specification of patterns of serial correlation does not offer a satisfactory alternative to the specification of a general dynamic structure (Hendry & Mizon 1978). Our procedure deals with both of these shortcomings and is therefore an important generalization of Chow and Lin's method.

Consider the following dynamic monthly regression equation linking the $j = 1, \dots, k$ observed indicator variables $\{x^j_{t,u}\}$ to the unobserved monthly interpoland $y_{t,u}$

$$\alpha(L)f(y_{t,u}) = \beta_0 + \sum_{j=1}^{k} \beta_j(L)x^j_{t,u} + \epsilon_{t,u}, \tag{1}$$

for $u = 1,2,3$ and $t = 0,1,...,T$. In (1), the subscript t indicates the particular quarter and the subscript u denotes the month within the t-th quarter, L is the monthly lag operator, so $\alpha(L) = 1 - \sum_{i=1}^{p} \alpha_i L^i$ and $\beta_j(L) = \sum_{k=0}^{q_j} \beta_{j,k} L^j$ are scalar lag polynomials of orders p and q_j, respectively, operating on the unobserved monthly dependent variable $f(y_{t,u})$ and the observed monthly indicator variable $\{x^j_{t,u}\}$. The functional form $f(.)$ used in constructing the interpoland in (1) is assumed known. The possibility that the dependent variable $f(y_{t,u})$ in (1) is a non-linear function of the interpoland $y_{t,u}$ reflects a frequent occurence in applied macro-econometric research; for example, a logarithmic transformation is often employed. Of course, the exogenous indicator variables $\{x^j_{t,u}\}$ may themselves also be transformations of other underlying variables. It is assumed that the lag lengths p and q_j are chosen sufficiently large so that the error terms $\{\epsilon_{t,u}\}$ may be assumed to possess zero mean, constant variance, and to be serially uncorrelated and uncorrelated with lagged values of $f(y_{t,u})$ and current and lagged values of $\{x^j_{t,u}\}$.

The regression equation (1) is quite general. For example,

a) if $\alpha_i = 0$ for all $i > 0$, then the model is essentially static in the level of $f(y_{t,u})$;

b) if $\alpha_1 = -1$ and $\alpha_i = 0$ for all $i > 1$, then the model involves the monthly first difference of $f(y_{t,u})$.

Other values for the parameters $\{\alpha_i\}$ allow a general specification of the dynamics in (1). In the special case in which the sum of the coefficients on the dependent variable is unity, that is $\sum_i \alpha_i = 1$, the left hand side of (1) may be re-expressed as a scalar lag polynomial of order $p-1$ operating on the first difference of the dependent variable $f(y_{t,u})$. When $\sum_i \alpha_i \neq 1$, there is a long-run relationship linking $f(y_{t,u})$ and $\{x^j_{t,u}\}$; in particular, if $f(y_{t,u})$ and $\{x^j_{t,u}\}$ are difference stationary, there exists a co-integrating relationship between $f(y_{t,u})$ and $\{x^j_{t,u}\}$. Furthermore, in this case a test of the restriction $\sum_i \alpha_i = 1$ corresponds to a test of the null hypothesis that there is no co-integrating relationship; see Engle & Granger (1987).

Estimation of the unknown parameters in (1) is not completely straightforward. The monthly variables $f(y_{t,u})$ are not observed whereas we do observe the quarterly aggregates of the interpolands $\{y_{t,u}\}$

$$y_t = \sum_{u=1}^{3} y_{t,u}.$$

Firstly, we need to aggregate (1) appropriately to yield a regression equation involving only the observable quarterly aggregates $y_t = \sum_{u=1}^{3} y_{t,u}$, which is then feasible for estimation, and secondly, we will need to deal with the implications of aggregation for the error structure of the resultant regression equation.

We may transform (1) into a regression equation involving only third-order lags of $f(y_{t,u})$ by pre-multiplying by a suitable polynomial function of the monthly lag operator L, whose coefficients will depend on $\{\alpha_i\}$. For simplicity, we deal only with the case in which the maximum lag length $p = 1$, hence $\alpha(L) = 1 - \alpha_1 L$.[3] Multiplying (1) through by $1 + \alpha_1 L + \alpha_1^2 L^2$ yields

$$f(y_{t,u}) = \alpha_1^3 f(y_{t-1,u}) + (1 + \alpha_1 + \alpha_1^2)\beta_0 + (1 + \alpha_1 L + \alpha_1^2 L^2)\left(\sum_{j=1}^{k} \beta_j(L) x^j_{t,u} + \epsilon_{t,u}\right) \tag{2}$$

[3] In the application, we limit ourselves to this case. More generally, the lag polynomial $\alpha(L)$ is factored in terms of its roots and each factor may then be treated using the method in this paper. See Salazar et al. (1996)).

Consequently, in the transformed regression equation (2) $f(y_{t,u})$ depends on its value three months previously as well as on the exogenous indicator variables. Aggregating (2) across the t-th quarter we obtain

$$\sum_{u=1}^{3} f(y_{t,u}) = \alpha_1^3 \sum_{u=1}^{3} f(y_{t-1,u}) + 3(1+\alpha_1+\alpha_1^2)\beta_0 + \sum_{u=1}^{3} \left\{ (1+\alpha_1 L + \alpha_1^2 L^2) \left(\sum_{j=1}^{k} \beta_j(L) x_{t,u}^j + \epsilon_{t,u} \right) \right\} \quad (3)$$

If (3) had been expressed in terms of $y_{t,u}$ rather than $f(y_{t,u})$ it would now involve the quarterly endogenous variable y_t and, thus, be feasible for estimation. To obtain an operational formulation of (3), we exploit the mean value theorem and express $f(y_{t,u}) = f(\bar{y}_t) + f'(y^*_{t,u})(y_{t,u} - \bar{y}_t)$, where $\bar{y}_t = y_t/3$ is the monthly average in quarter t and $y^*_{t,u}$ lies between $y_{t,u}$ and $\bar{y}_t$. If the error of approximation $y_{t,u} - \bar{y}_t$ is relatively small, we have

$$\sum_{u=1}^{3} f(y_{t,u}) \doteq 3f(\bar{y}_t) \quad (4)$$

note that the errors of approximation sum to zero; *viz.* $\sum_{u=1}^{3}(y_{t,u} - \bar{y}_t) = 0$. For a logarithmic transformation, the approximation becomes

$$\sum_{u=1}^{3} \ln y_{t,u} \doteq 3 \ln y_t - 3 \ln 3 \quad (5)$$

which can be seen to be equivalent to replacing the quarterly value $\sum_{u=1}^{3} \ln y_{t,u}$ by three times the geometric mean of the monthly values $y_{t,u}$, $u = 1, 2, 3$. The geometric mean is never larger than the arithmetic mean, but if monthly movements are small compared with the monthly average, the approximation error introduced should be of little importance.

As a result, the substitution of the approximation (4) into (3) provides a regression equation feasible for estimation. Note that the covariance structure of the error terms in (3) is a function of the parameter α_1. In the results that follow, we use maximum-likelihood estimation.[4] We show, in Salazar et al. (1996)) how to estimate the variance matrix of the resulting interpolated data.

3 Interpolation without Indicator Variables

Table 1 indicates that there are some sectors, agriculture and mainly public sector services, for which there are no obvious indicator variables available. One could nevertheless estimate (3) as a pure autoregression. However, the application of this approach to the public sector raised an interesting practical problem. The estimated coefficient α_1^3 in (3) was of the order of -0.2 when estimated on quarterly data. Extracting

[4]If the lag coefficient is known, perhaps because the monthly model reduces to one in first differences, the maximum-likelihood solution is that offered by Generalized least squares. For the more general case Salazar et al. (1996)) study the properties of the GLS and ML estimators of the parameters β_0, $\{\alpha_i\}$ and $\{\beta_{j,k}\}$ *via* Monte-Carlo experiments when the error terms $\{\epsilon_{t,u}\}$ are independently and identically distributed normal variates. The technique peformed well on samples of the size which we have available for practical estimation.

the cube root to estimate α_1 in (1) yielded figure of about -0.6. This implies an implausible amount of month-on-month movement for a variable which is generally believed to be smooth. We felt it would be better to look for a method which preserved the generally-accepted smoothness of the series.

Therefore, we constructed preliminary estimates of the monthly data $\{\tilde{y}_{t,u}\}$ from a simple two-sided moving average filter employing equal weights in terms of the monthly averages; that is

$$\tilde{y}_{t,1} = 2\bar{y}_t/3 + \bar{y}_{t-1}/3, \tilde{y}_{t,2} = \bar{y}_t, \tilde{y}_{t,3} = 2\bar{y}_t/3 + \bar{y}_{t+1}/3$$

where $\bar{y}_t = y_t/3$. We then assume the unobserved monthly data are linked to these preliminary estimates by the approximate model

$$\Delta_1 f(y_{t,u}) = \Delta_1 f(\tilde{y}_{t,u}) + \epsilon_{t,u} \tag{1}$$

where $\Delta_1 = 1 - L$ is the monthly difference operator and $\epsilon_{t,u}$ is as in section 2. Again, the functional transformation $f(.)$ used in the application discussed in section 5 is logarithmic.

4 Reconciliation of the Interpolands

The estimators of the parameters of the monthly regression equation (1) may then be used to produce fitted values of the interpolands $\{y_{t,u}\}$. These fitted values, however, need to be reconciled with the observed quarterly data $\{y_t\}$. Our estimate of $\{y_{t,u}\}$ minimises the sums of squares of the residuals in the regression equation (1) subject to the constraint that the interpolated monthly values in each quarter sum to the known quarterly totals, that is, $\sum_{u=1}^{3} y_{t,u} = y_t$.

For simplicity, we again confine attention to the first order case, by setting $p = 1$. There are observations available on the quarterly totals y_t for quarters $t = 1, ..., T$. Firstly, recall (1)

$$f(y_{t,u}) = \alpha_1 f(y_{t,u-1}) + \beta_0 + \sum_{j=1}^{k} \beta_j(L) x_{t,u}^{j} + \epsilon_{t,u} \tag{1}$$

where, for $t = 1$ (the first quarter in the sample), $u = 2, 3$ and for $t = 2, ..., T$ (the remainder of the sample), $u = 1, 2, 3$. At this stage the problem reduces to optimising the Lagrangean

$$\sum_{u=2}^{3} \epsilon_{1,u}^2 + \sum_{t=2}^{T} \sum_{u=1}^{3} \epsilon_{t,u}^2 + \sum_{t=1}^{T} \lambda_t \left(\sum_{u=1}^{3} y_{t,u} - y_t \right) \tag{2}$$

where λ_t is the Lagrange multiplier associated with the constraint $\sum_{u=1}^{3} y_{t,u} = y_t$ over $t = 1, ..., T$. The first-order conditions are given by

$$\nabla f(y_{t,u})(\epsilon_{t,u} - \alpha_1 \epsilon_{t,u+1}) + \lambda_t = 0 \tag{3}$$

where $\epsilon_{1,1} = 0$, $\epsilon_{T+1,1} = 0$, $u = 1, 2, 3$, $t = 1, ..., T$, and ∇ is the derivative operator.

Equation (3) can be solved jointly with the adding-up constraints, $\sum_{u=1}^{3} y_{t,u} = y_t$ over $t = 1, ..., T$ to produce estimates of the interpolands $\{\hat{y}_{t,u}\}$, $u = 1, 2, 3$ and the Lagrange multipliers $\hat{\lambda}_t$. The solution is inherently non linear, because the derivatives $\nabla f(.)$ in (3) are a function of the estimated interpolated data $\{\hat{y}_{t,u}\}$ which, in principle, necessitates the use of iterative methods. However, when the transformation $f(.)$ is logarithmic our experience indicates that the derivatives $\nabla f(.)$ in (3) may be satisfactorily evaluated at the monthly average $\bar{y}_t$ of the corresponding quarterly total y_t, hence avoiding further iteration. Further details concerning the solution of (3) are presented in Salazar et al. (1996)).

At the same time as interpolating the data, we are able to produce estimates of approximate expressions for the variances and covariances of the estimated interpolands $\{\hat{y}_{t,u}\}$. Including only terms of order $O_P(1)$, the source of error due to the estimation of the regression parameters is irrelevant, at least asymptotically. Hence, only the random component represented by the error terms $\{\epsilon_{t,u}\}$ is pertinent. Details of the requisite calculations are provided in Salazar et al. (1996)).

In the case when there are no indicator variables available, the approach to interpolation is essentially similar. It is necessary merely to substitute the expression for $\epsilon_{t,u}$ given in (1) into the Lagrangean (2). Details for the calculations of the interpolands and their approximate variances are set out in Appendix A.

5 Monthly Estimates of Constant Price GDP

5.1 Indicators for the Output Measure of GDP

The calculation of monthly estimates of output is done in three components. For the industries covered by the index of production, the index values simply indicate monthly output. Output is broken down into four further groups. For agriculture and public sector services we have to use the mechanical method of section 3, while for construction and private sector services we can use the indicator-variable approach of section 2. Our interpolated series begin in 1984, because the data on the output of materials used by the construction industry are not available before 1983.

Agriculture and Public Sector Services

For these two categories, there is no relevant monthly indicator variable, and there is no obvious reason to suppose that the output of agriculture (CKAP) or of other services (CKJC), the latter mainly ownership of dwellings and services produced by the public sector, should be closely linked to the various monthly data which do exist.

Table 3 suggests that these components of the output index are stable in logarithmic first differences. Accordingly, we apply the method set out in section 3 and Appendix A, interpolating the data on the basis of the quarter-on-quarter growth rate (assumed to relate from mid-month to mid-month of each quarter), but minimising the sum of the squared month-on-month changes which arise subject to the requirement that the monthly data add to the quarterly estimates.

Variable	Code	x	Order	$\Delta\ x$	Order	$\log(x)$	Order	$\Delta\ \log(x)$	Order
Agriculture	CKAP	-0.794	11	-3.154	10	-0.783	11	-3.326	10
Non-marketed Services		-2.441	3	-3.261	7	-2.820	3	-3.350	7

Estimation Period: 1973Q1-1998Q1

Table 3: ADF tests for Agriculture and Non-market Services

The standard errors of the error terms in the model of section 3 are estimated using the procedure described in Appendix A. We find that, for agriculture the standard error is 0.41%, while for public sector services it is 0.10%. However, these standand errors apply to the percentage growth rate from one quarter to the next before the adding up-constraints are taken into account. After allowance is made for these, we find that the average standard error in the level of the monthly data, measured as a proportion of its interpolated value, is 0.22% for agriculture and 0.06% for public sector services. The average monthly errors in the rates of change have to be calculated from the variance-covariance matrix of the interpolands, as shown in Appendix A. These are 0.33% for agriculture and 0.08% for public sector services.

Private Sector Services

The output of private sector services was interpolated by means of indicator variables. A preliminary search on quarterly data (making no adjustment for the moving-average error process) suggested that the growth in private-sector service output was related to growth in retail sales, to growth in manufacturing output (but not to the movements in the other components of the index of production) and to growth in imports of goods (using the Overseas Trade Statistics data).

Variable	Code	x	Order	$\Delta\ x$	Order	$\log(x)$	Order	$\Delta\ \log(x)$	Order
Quarterly variable									
Market Services	PSS1	-1.924	8	-2.931	8	-2.226	9	-3.543	8
Monthly variables									
Retail Sales	FAAM	-2.371	7	-4.686	8	-2.483	4	-5.221	8
Manufacturing	MANU	-2.574	15	-3.872	16	-2.531	15	-3.911	16
Imports	MOTS	-1.875	13	-4.977	12	-2.615	13	-5.287	12

Estimation Period: Quarterly variable 1973Q1-1998Q1, Monthly variables 1973M1-1998M3.

Table 4: ADF tests for Market Services and related Monthly Indicators

All the explanatory variables were I(1), as table 4 shows. No cointegrating vector could be identified linking the variables.In consequence, the underlying monthly equation links first differences in the indicator variables to first differences in the interpoland. The regression exercise led to the equation shown in table 3. This suggests that, while our indicators, manufacturing output, retail sales and imports may be reasonably good indicators of movements of private service output there is no long-run relationship between output of private services and these indicators. That should come as no great surprise; it indicates that the components of the sector which are not reflected in our indicators may follow stochastic trends of their own.

This equation is possibly the most important tool in the interpolation of constant price GDP. The diagnostic tests are all satisfactory, and the within-sample fit is good, with an R^2 in terms of the change in services output of 0.7 and a standard error of around 0.4%.

Table 5: Regression for Private Sector Services output

Dependent Variable is Δln Market Services

Variable	Code	coef.	t-val.	s.e.
Constant		0.004	5.233	0.001
$\Delta ln\ RetailSales_{-0}$	faam	0.368	6.831	0.054
$\Delta ln\ Manufacturing_{-0}$	manu	0.116	3.283	0.035
$\Delta ln\ RetailSales_{-1}$	faam	0.128	1.985	0.064
$\Delta ln\ Imports_{-1}$	mots	0.057	3.417	0.017

DW	R^2	s.e.
2.114	0.7269	0.00396

Sample Period: 73Q3 to 98Q1
Extrapolation Period: 98M4 to 98M9

Chow Test (forecast adequacy) F(4,90) = 0.4575 [0.7667]

Theil test based on forecast mean = 1.176

Theil test based on lagged value = 3.376

Bera-Jarque normality test = 3.305 [0.1915]

Serial correlation: F(1,93) = 0.4113 [0.5229] F(4,90) = 0.3183 [0.8651]

ARCH test: Chi(1) = 0.6416 [0.4231] Chi(4) = 2.155 [0.7073]

Chow test (parameter stability), F(5,89) = 1.126 [0.3526]

MSE of estimate of level data : 0.22%

MSE of month-on-month growth rate : 0.05%

MSE of rolling quarter-on-quarter growth rate : 0.03%

Construction

The estimate of a regression for construction output is complicated by the fact that some of the potentially-important indicators do not become available until the 1980s. For the period from 1983Q4 until 1993Q4 we have estimated an equation explaining movements in construction output from movements in output of constructional steelwork, orders received in volume terms and output of non-plastic building materials. This is less satisfactory than the market services' equation in a number of respects. First of all, the ADF statistics in table 6 are consistent with the view that, with the exception of orders received, the series even in first differences include unit roots. We have taken the view that the output data in log first differences are in fact stationary. Secondly, the Chow test suggests that parameter instability may be present; we were unable to find an equation using plausible indicators which avoided this problem unless it was estimated over a rather shorter period.

Variable	Code	x	Order	$\Delta\ x$	Order	$\log(x)$	Order	$\Delta \log(x)$	Order
Quarterly Variable									
Construction Output	BAEM	-2.475	5	-3.183	1	-2.566	9	-2.351	6
Monthly Variables									
Orders Received	FEAZ	-2.244	5	-11.062	4	-2.171	16	-10.787	4
Steelwork Output	FFAA	-2.147	17	-2.831	16	-2.138	17	-2.906	16
Other Bldg. Mat. Out.	FFBB	-2.355	15	-2.126	18	-2.350	15	-2.175	18

Estimation Period: Quarterly variable 1983Q1-1998Q2, Monthly variables 1983M1-1998M6.

Table 6: ADF tests for Construction output and related Monthly Indicators

With these shortcomings in mind, the error correction mechanism suggests that construction output is related to orders received in levels, but the relationships with the two components of manufacturing output relate only to differences. This seems very plausible; provided the two series have similar coverage one would expect construction output and orders received to have a common stochastic trend. The regression results for the interpolation of this component are shown in table 8.

Table 7: Regression results for Construction output

Dependent Variable is ln Construction Output

Variable	Code	coef.	t-val.	s.e.
$ln\ ConstructionOutput_{-1}$	baem	0.982	138.242	0.007
$ln\ OrdersReceived_{-1}$	feaz	0.021	3.362	0.006
Constant		-0.014	-0.399	0.036
$\Delta ln\ SteelworkOutput_{-1}$	ffaa	0.140	3.169	0.044
$\Delta ln\ OthBuildMatsOutput_{-1}$	ffbb	0.307	4.048	0.076
$\Delta ln\ SteelworkOutput_{-2}$	ffaa	-0.107	-2.491	0.043
$\Delta ln\ OthBuildMatsOutput_{-2}$	ffbb	-0.175	-2.238	0.079

DW	R^2	s.e.
2.323	0.9834	0.009612

Sample Period: 83Q3 to 98Q2
Extrapolation Period: 98M7 to 98M9

Chow Test (forecast adequacy) F(5,48) = 0.996 [0.4303]

Theil test based on forecast mean = 0.9903

Theil test based on lagged value = 0.7554

Bera-Jarque normality test = 1.559 [0.4586]

Serial correlation: F(1,52) = 1.679 [0.2007] F(4,49) = 0.6398 [0.6367]

ARCH test: Chi(1) = 1.284 [0.2572] Chi(4) = 2.232 [0.6933]

Chow test (parameter stability), F(7,46) = 3.043 [0.01024]

MSE of estimate of level data : 0.60%

MSE of month-on-month growth rate : 0.90%

MSE of rolling quarter-on-quarter growth rate : 0.39%

6 The Residual Error

Until 1990 there were three estimates published estimates of GDP calculated from output, income and expenditure data. Since then there has been a single measure of GDP which is output-based. Nevertheless, there are discrepancies between the published measure of GDP and the figure that can be calculated from the output indices[5].

We deal with this problem by aggregating the four monthly series described above together with industrial production to give an estimate of the output measure of GDP. We then use this as an indicator variable to interpolate the published measure of GDP, using the method described in section 3, but imposing

[5]See, for example, footnote 2 on p. 28 of *Economic Trends*, July 1996

a unit coefficient on the output measure of GDP instead of estimating it by regression. We did not use a regression because the discrepancy is always greater with more recent data and is subsequently revised away.

7 The Reliability of Estimates of Monthly GDP

The actual standard errors of the interpolated data depend on the period in question. table 9 presents the standard errors for the output measure of GDP, calculated as described above, for each of the months in the quarter, together with the mean standard errors associated with estimates of the month-on-month and rolling quarter-on-quarter changes in the interpoland.

Month in quarter	Levels (%)	Month on month (%)	Quarter on quarter (%)
1	0.1115	0.1302	0.1039
2	0.0856	0.1646	0.1037
3	0.1112	0.1645	0.0000

Table 9: Standard errors in GDP(O)

Table 10, in turn, shows the mean of the standard errors of the estimates of the published measure of GDP. These figures reflect the extra component arising from the discrepancy between GDP(O) and the published measure

Month in quarter	Levels (%)	Month on month (%)	Quarter on quarter (%)
1	0.1174	0.1383	0.1090
2	0.0891	0.1775	0.1088
3	0.1171	0.1774	0.0000

Table 10: Standard errors in GDP(A)

In general, the interpoland standard errors appear to be quite stable, and the results are consistent with the mechanics of the interpolation procedure. The standard errors associated with GDP(A) are slightly larger than those of the output measure, GDP(O). This is because we do not have the monthly information needed to interpolate expenditure and income components.

8 Conclusions: Monthly Estimates of GDP

Table 11 shows our interpolated estimates of the components of GDP, together with an aggregate output estimate calculated by adding up the components using 1990 weights (the GDP(O) column) and the interpolated published figure of GDP (the GDP(A) column.) In both cases the monthly data are fully consistent with published quarterly data. Figure 1 shows the month on month and rolling quarter on quarter percentage changes in GDP(A), from January 1985 to July 1998.

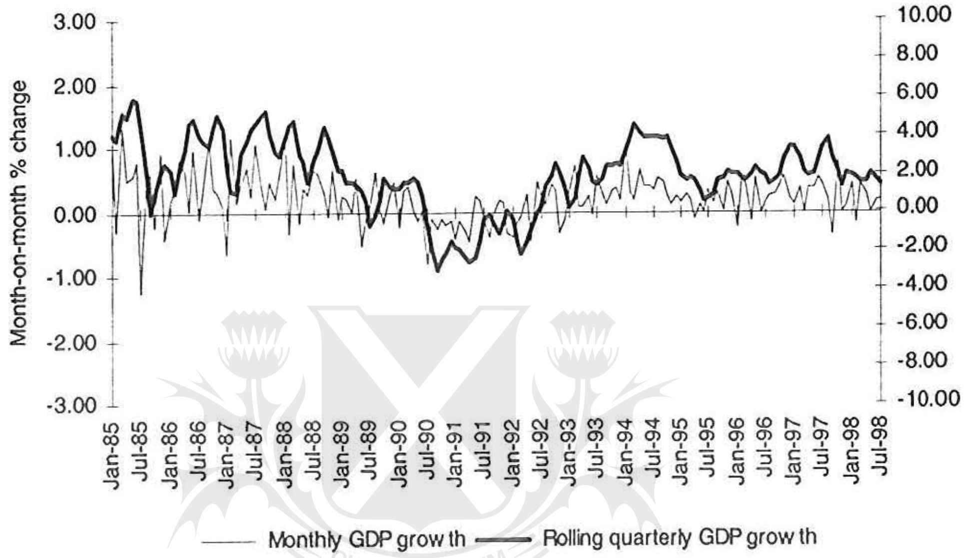

Figure 1: Month-on-month and rolling quarter-on-quarter growth rates of monthly GDP estimates, 1985M1-1998M7.

Our approach provides a robust procedure for the estimation of monthly national accounts. Quantification of measurement errors is an essential part of any data estimation. In this case, until more data are collected or improvements to the estimation methodology are made, the estimates of standard errors provide an indication of the precision with which inferences can be made about the state of the economy in any particular month.

A rather separate question concerns the utility of the data in a statistical sense. One would not expect monthly data to add very greatly to our understanding of the long-term performance of the economy. However, at a bare minimum our approach produces monthly indicators of the main components of the national accounts which are, over historical periods (due to the imposition of adding-up constraints) unambiguously superior to the existing available monthly indicators, which must at present, perforce, be used as proxies for the desired series. Since, as we note above, the underlying equations of key importance

Period	Industry	Agriculture	Construction	Market Services	Non-market Services	GDP(O)	GDP
Jan 1997	109.0	93.4	91.7	118.7	108.3	111.4	111.2
Feb 1997	108.7	93.7	93.0	119.6	108.3	111.8	111.6
Mar 1997	108.3	94.0	92.5	120.0	108.3	111.8	111.6
Apr 1997	109.3	94.4	93.2	120.2	108.3	112.2	112.0
May 1997	108.3	94.4	93.5	121.7	108.3	112.6	112.4
Jun 1997	110.0	94.0	93.2	122.2	108.4	113.2	113.0
Jul 1997	110.8	93.2	93.0	122.8	108.5	113.7	113.5
Aug 1997	110.2	92.6	93.7	123.4	108.6	113.9	113.7
Sep 1997	109.9	92.3	92.6	122.8	108.7	113.5	113.3
Oct 1997	109.6	92.3	95.0	124.7	108.9	114.4	114.1
Nov 1997	109.1	92.4	93.2	125.4	108.9	114.4	114.2
Dec 1997	109.2	92.5	95.0	125.3	108.9	114.5	114.3
Jan 1998	108.9	92.8	95.4	126.5	108.9	115.0	114.7
Feb 1998	108.6	92.9	96.9	126.0	108.8	114.8	114.6
Mar 1998	109.5	93.0	96.7	126.6	108.8	115.3	115.1
Apr 1998	110.8	93.0	96.4	126.6	108.8	115.6	115.4
May 1998	109.4	93.0	96.6	127.5	108.7	115.6	115.3
Jun 1998	110.2	93.0	95.9	127.5	108.7	115.8	115.5
Jul 1998	110.7	93.0	94.8	127.8	108.8	116.0	115.7

Table 11: Monthly Output Estimates, 1997M1 to 1998M7.

in interpolating the major aggregates also have good explanatory power, appear stable on the standard diagnostic tests, and perform respectably out of sample, it seems likely that there will also be significant additional useful information on movements within a given calendar quarter, and when the approach is used for pure extrapolation. This is an important topic for further investigation; however it lies outside the scope of the present study.

References

Chow, G. C. & Lin, A. L. (1971), 'Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series', *The Review of Economics and Statistics* **53**, 372–75.

Corrado, C. (1986), Reducing Uncertainty in Current Analysis and Projections: the Estimation of Monthly GNP, Federal Researve Board Special Study No. 209.

Engle, R. F. & Granger, C. W. J. (1987), 'Co-integration and Error Correction: Representation, Estimation and Testing', *Econometrica* **55**, 1–87.

Fernandez, R. (1981), 'A Methodological Note on the Estimation of Time Series', *Review of Economics and Statistics* **63**, 471–478.

Friedman, M. (1962), 'The Interpolation of Time Series by Related Series', *Journal of the American Statistical Association* **57**, 729–757.

Ginsburgh, A. (1973), 'A Further Note on the Derivation of Quarterly Figures from Annual Data', *Applied Statistics* **5**, 388–394.

Harvey, A. (1989), *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A. C. & Pierse, R. G. (1984), 'Estimating Missing Observations in Economic Time Series', *Journal of the American Statistical Association* **79**, 125–31.

O'Dea, D. (1975), *Cyclical Indicators for the Postwar British Economy.* NIESR Occasional Paper No. 28, Cambridge University Press.

Hendry, D. & Mizon, G. (1978), 'Serial Correlation as a Convenient Simplificiation, not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England', *Economic Journal* **88**, 549–563.

Palm, F. C. & Nijman, T. E. (1984), 'Missing Observations in the Dynamic Regression Model', *Econometrica* **52**, 1415–35.

Stock, J. H. & Watson, M. W. (1989), 'New Indices of Coincident and Leading Economic Indicators', *NBER Macroeconomics Annual.*

Salazar, E., Smith, R. & Weale, M. (1997), Interpolation by Means of a Regression Model: Estimation and Monte-Carlo Analysis, National Institute of Economic and Social Research, Discussion Paper No. 126.

Stone, J., Champernowne, D. & J.E.Meade (1942), 'On the Precision of National Income Estimates', *Review of Economic Studies* **10**, 111–135.

Appendix A: Interpolation without Indicator Variables

A.1 Data Estimation

When there are no indicator variables available, we adopt the following model as in (1):

$$\Delta_1 f(y_{t,u}) = \Delta_1 f(\tilde{y}_{t,u}) + \epsilon_{t,u}, \tag{A.1}$$

$u = 2,3,\ t = 1,\ u = 1,2,3,\ t = 2,...,T$, where Δ_1 is the first difference operator and $\tilde{\mathbf{y}}^h$ denotes the vector of monthly data constructed by the crude interpolation method described in section 3. Effectively, (A.1) corresponds to equation (1) section 2 with $\alpha_1 = 1$ and $\beta_0 + \sum_{j=1}^{k} \beta_j(L)x_{t,u}^j$ set equal to $(\Delta_1 f(\tilde{y}_{1,2}), \Delta_1 f(\tilde{y}_{1,3}), ..., \Delta_1 f(\tilde{y}_{T,3}))'$. The solution for the interpolands, $y_{t,u}$, $u = 1,2,3,\ t = 1,...,T$, may be obtained from equation (2) of section 4.

A.2 An Approximate Variance for the Interpolands

In order to estimate the error variance σ_ϵ^2, we proceed as in section 2. Firstly, multiply (A.1) by the lag polynomial $1 + L + L^2$. Hence,

$$\Delta_3 f(y_{t,u}) = \Delta_3 f(\tilde{y}_{t,u}) + u_{t,u}, \tag{A.2}$$

where $\Delta_3 = 1 - L^3$ and $u_{t,u} = (1 + L + L^2)\epsilon_{t,u}$. Secondly, aggregating (A.2) across quarter t gives

$$\sum_{u=1}^{3} f(y_{t,u}) - \sum_{u=1}^{3} f(y_{t-1,u}) = \sum_{u=1}^{3} f(\tilde{y}_{t,u}) - \sum_{u=1}^{3} f(\tilde{y}_{t-1,u}) + \sum_{u=1}^{3} u_{t,u}. \tag{A.3}$$

Applying the approximation (4) of section 2 to (A.3) results in

$$3[f(\bar{y}_t) - f(\bar{y}_{t-1})] \doteq \sum_{u=1}^{3} f(\tilde{y}_{t,u}) - \sum_{u=1}^{3} f(\tilde{y}_{t-1,u}) + u_t, \tag{A.4}$$

where $u_t = \sum_{u=1}^{3} u_{t,u}$. Now, u_t in (A.4) is a moving average process of order 1 with $var\{u_t\} = 19\sigma_\epsilon^2$ and $cov\{u_t, u_{t-1}\} = 4\sigma_\epsilon^2$. Hence, neglecting the approximation error in (A.4), an unbiased estimator for the error variance σ_ϵ^2 is given by $[\sum_{t=2}^{T} u_t^2 + 2\sum_{t=3}^{T} u_t u_{t-1}]/[27(T-1) - 8]$.