An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth*

Discussion Paper 127 (revised)

James Mitchell[1], Richard J. Smith[1,2], Martin R. Weale[1,4], Stephen Wright[3] and Eduardo L. Salazar[1]

[1]National Institute of Economic and Social Research
2, Dean Trench Street
Smith Square
London SW1P 3HE

[2]Department of Economics
University of Warwick
Coventry CV4 7AL

[3]Department of Economics
Birkbeck College
University of London
London W1T 1LL

[4]Statistics Commission,
10, Great George Street,
London SW1P 3AE

June 4, 2004

Abstract

A range of monthly series are currently available giving indications of short-term movements in output in the United Kingdom. As the only available information, these indicators are routinely exploited in various ways although they only provide an incomplete picture of gross domestic product (GDP). The main aim of this paper is to suggest a formal and coherent procedure for grossing these monthly data up to represent the whole of GDP. Although the resultant estimates of GDP would be worse than those obtained by direct measurement, they should be more satisfactory than simply making an informal inference from whatever monthly data are available. Our examination of the efficacy of the method for estimation of the state of economic activity indicates a rather satisfactory outcome.

*This paper summarises the results of a project initially undertaken for the Office of National Statistics (ONS) of the United Kingdom, and jointly supported by HM Treasury. A steering committee provided very useful feedback during the early part of the work. Richard Clare, Geoff Reed, Keith Vernon and Colin Yeend were particularly helpful in providing both data and suggestions. We also wish to thank Stephen Hall for helpful comments. The current state of the work described here at which estimates of UK real GDP are produced on a monthly basis was achieved with support from ESRC research grant L116251012. Richard Smith also gratefully acknowledges partial financial support for this and an earlier revision from the ESRC under the project R000237334 "Automatic Leading Indicators" and a 2002 Leverhulme Major Research Fellowship. Any errors remain our own.

1 The Motivation for Monthly Estimates of GDP

Macro-economic policy-making in real time faces the perennial problem of uncovering what is actually happening to the economy. Movements of seasonally adjusted real GDP (referred to subsequently simply as GDP) and related estimates of the output gap are widely regarded as important predictors of future inflation and thus are relevant to the problem of inflation targeting. Estimates of GDP are typically produced quarterly with the first estimates in the United Kingdom available about twenty five days after the end of the quarter to which they relate.[1] In many countries including the UK monetary policy is set more frequently than quarterly and in order to do so policy makers need not only to anticipate first estimates of GDP growth but also to estimate what is happening within each quarter. A range of monthly series are currently available giving indications of short-term movements in output. As the only available information, they are already exploited in various ways: financial commentators routinely examine monthly data on retail sales, the trade figures, and the output of the production industries in order to assess the state of the economy and likely developments in monetary policy; 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 worse than those which might be produced by direct measurement. On the other hand, it would certainly be more satisfactory than simply making an informal inference from whatever happen to be the latest numbers available.

The Office for National Statistics (ONS) in the United Kingdom used to publish lagging, co-incident and leading cyclical indicators for the UK. These indicators were constructed (O'Dea 1975) 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 intended to represent cyclical movements in real GDPand the co-incident indicator was constructed by interpolating quarterly GDP. The interpolation method used industrial production and retail sales as indicators but in a manner which was uninformed by statistical theory. The indicator system was abandoned mainly because the forecasting power of the leading indicator was negligible. However, a logical step was to see whether the co-incident indicator could be developed into an indicator of monthly seasonally adjusted GDP in constant prices estimated using clearly specified statistical methods.

[1]Longer lags are typical elsewhere although EUROSTAT is aiming for a reduction to forty-five days for members of the European Union.

Other researchers have approached the problem of constructing a co-incident indicator from a different perspective. Stock & Watson (1989) produced a co-incident indicator of economic activity in the United States. Their indicator was an estimated latent variable derived from a dynamic factor model based on a number of variables believed to coincide with the business cycle in the United States. The authors argued that it was valuable to distinguish indicators of the business cycle from monthly movements in GDP, because GDP can be affected by variables such as weather conditions and other acts of God unrelated to the state of the cycle. Two strands of research have emerged from this. One (Stock & Watson 2002) builds on Stone (1947) using static principal components to summarise a wide range of data, and defines the cycle as the first principal component. The second (Forni, Hallin, Lippi & Reichlin 2001) develops a dynamic factor model from spectral analysis, again aiming to represent the state of the economy by a latent variable.

Despite the desirability of purging the distortionary effects of non-economic events from data, the use of a latent variable as a key indicator of the state of the economy is open to two fundamental objections. First of all, there are bound to be difficulties in communicating its meaning to data users and policy makers; they will inevitably want to relate it to GDP. Secondly, there is no obvious means of verifying the performance of the models used to generate it.

In this paper, therefore, we focus on the construction of a monthly indicator of GDP[2] and the way it can be combined with short-term forecasting methods to produce an estimate of quarterly GDP growth about three weeks ahead of the first official data. The indicator has been published monthly by the National Institute of Economic and Social Research since April 1998. The early estimates of quarterly GDP growth produced are compared here with the first official estimates, allowing us to provide a real-time assessment of the approach.

2 Data Availability

Output indicators are generally regarded as being better guides than expenditure indicators to short-term movements in GDP (Office for National Statistics 2004, Background Note 2). The first estimate of GDP produced by the Office for National Statistics (ONS) is based on output rather than expenditure data (*National Accounts, Concepts, Sources and Methods* 1998, p. 212). It is therefore natural to attempt to build an estimate of monthly GDP from the output side. The components of the output measure of real GDP fall into three broad categories: series for which data are already available; series for which there are

[2] If desired, smoothing methods might also be used to produce estimates of "underlying" GDP from the monthly indicator. Their performance would presumably need to be assessed in the light of the capacity of underlying GDP to predict future GDP. Such underlying measures ought to provide reasonable forecasts because they should remove short-run $I(0)$ noise but not the long-run $I(1)$ signal in GDP.

Component	Code	Availability of data/indicators	% of GDP (1995)	(2002)
Production	CKYM	100 % data available	26.6	21.4
		Index of Production		
Agriculture	GDQA	No indicators	1.8	1.0
Construction	GDQB	Input and Output Indicators available	5.2	6.3
		available only with Delay		
		Qualitative DG-ECFIN Survey since 1985		
Private Services	GDQE, GDQH,	Qualitative DG-ECFIN Survey since 1997	43.9	52.8
	GDQN	Index of Retail Sales		
		Indices of Production Activity		
		Monthly Trade Data		
Public Services	GDQU	No indicators	22.5	18.6

Table 1: Data Coverage for Monthly Constant Price Accounts (Output Components)

indicators and series for which no monthly information is available.[3] It is debatable whether interpolands of this last group of variables include any extra information about the business cycle. However, 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.

We work with a decomposition of the output measure of GDP into five sectors. Table 1 shows these together with the codes describing the series in the ONS databank[4]. The terms Private and Public Services are a convenient shorthand. Public Services covers all education and health as well as public administration and defence. Some components of the first two items are marketed. Ownership of dwellings is included in Private Services despite the fact that part of this industry is run by the state. Nevertheless, as will become clear, the behaviour of the two service sectors is very different.

Table 1 also shows the availability of indicator variables [5] and each sector as a proportion of GDP in 1995 and in 2002; it is clear from this that the importance of the different sectors can change markedly over a relatively short time period. Nevertheless, much the most important sectors are production and public and private services. However, the size of each sector may be misleading when assessing its importance

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

[4]The data series for business services and government output consistent with current definitions go back only to 1983, while the data on output at current prices, needed to weight together the the components of private sector service output begin in 1986. For years earlier than 1986 we have constructed a series of government output in constant prices using earlier data series published in 1990 prices. We then use the output weights current used in the production index to calculate a series for private sector output by deducting the government component from total services output. This simple procedure gives a series with a correlation of 0.995 in one-period growth rates with the estimates of private sector output produced by chain-linking over the period 1986Q2-2003Q4 for which the comparison can be made.

[5]The ONS also publishes a monthly index of distribution output going back to 1995. This is relevant to private services output. However, it is published with a delay of eight weeks after the period in question and is therefore of little use in producing prompt estimates of economic activity.

	Production	Agriculture	Construction	Private Services	Public Services
Production	13.71	-0.02	1.72	8.62	-0.15
Agriculture		0.08	0.04	-0.16	-0.02
Construction			2.18	3.67	0.07
Private Services				27.65	-0.03
Public Services					0.74

Notes. Sample Period: $1986Q2 - 2003Q4$. The calculation of these covariances is explained in footnote 7.

Table 2: The Covariance Matrix of Weighted Components of Output Growth: 1973Q2-2003Q4

of its contribution to short-term movements in GDP.[6] Table 2 shows the covariance matrix of $10^3\times$ the quarter-on-quarter log changes of each component of GDP with the variables themselves scaled by their percentage shares of GDP in 2002.[7] Thus, (to first order), the scaled elements sum to $10^3\times$ the log change in GDP, and the variance measures shown are in effect contributions to the variance of the growth rate of GDP.

Together, Tables 1 and 2 indicate that the areas where the data are relatively weak are also those which do not contribute very greatly to the overall variability of GDP. For example, there is no monthly indicator of fluctuations of output by Public Services. However, despite amounting to around 20% of GDP, Public Services are a very small contributor to the overall variance of GDP; hence the absence of such indicators may not matter greatly. The largest source of variance is Private Services for which it is necessary to produce a good model from the available indicators if satisfactory interpolands of GDP are to be produced. The second largest source is Production for which monthly data exist. Moreover the covariance between movements in Private Services output and movements in Production suggests that monthly production data in some form or other will be a factor behind understanding what happens to the Private Services sector.

Published quarterly real GDP data have the property that the components of output do not sum to the total measured at basic prices. This arises because undisclosed 'other information' is used in addition to the information provided by the output series. To move from GDP at basic prices to GDP at market

[6]The movements in the relative sizes of the sectors over time do not complicate the overall modelling process. What is important in producing an aggregate from estimates for the individual components is that the weights used are the same as those used by the ONS in producing its data.

[7]Let z_{it} denote the output index for the ith sector in quarter t and $\hat{Z}$ the array whose (i,t)th element, $\hat{z}_{it} = 1000 \times \log(z_{it+1}/z_{it})$. The diagonal matrix Π has elements which are the weights shown in Table 1 for 2002. The table shows $\Pi(\hat{Z}P\hat{Z}'/T)\Pi$, where $P = I_T - T^{-1}\iota\iota'$, ι is a T-vector of units and T is the total length of the sample. Official service sector data are not published at the level of aggregation shown here; the component series for private services have been weighted together to produce a chain-linked index.

prices, some allowance must be made for movements in the factor cost adjustment at constant prices. We do this by calculating an estimate of GDP from the output data, GDP(O). We then use this as an indicator variable to interpolate both GDP(B), GDP at basic prices, and GDP(M), GDP at market prices.

The main purpose of this paper is to describe how an indicator of monthly GDP may be constructed. Section 3 briefly surveys earlier approaches to interpolation and also presents a summary of the interpolation method of this paper using indicator variables.[8] An account of the technique used for those series for which there are no available indicator variables is given in section 4. Section 5 describes the reconciliation of the fitted values of the interpolated monthly series with the available quarterly totals. The estimated equations which underpin the interpolation of the constant price disaggregated outputs of Table 1 are presented in section 6 whereas section 7 explains how we arrive at an estimate for monthly GDP at market prices. Monthly estimates of GDP are presented in section 8 and the method for the production of early estimates of quarterly GDP given in section 9. Section 10 concludes the paper. Appendix A presents full details of the interpolation method used when no indicator variables are available. Appendix B outlines how the model of section 3 may be cast in state-space form and estimated using the Kalman filter. Appendix B also offers a brief comparison of the interpolands of section 5 with alternatives based on the Kalman filter.

3 Estimation

The interpolation procedure of this paper relies on the specification of a regression equation linking the 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 parameter estimation addressed by Palm & Nijman (1984). An attractive feature of this method is the similarity between the regression equation specification and those encountered in conventional econometric research and macro-economic models. The interpolation procedure consists of two steps. First of all, the unobserved high-frequency interpolands are eliminated by aggregation of the underlying high-frequency regression model. The resultant estimable regression relationship explains observed low-frequency data on the interpolands in terms of suitable aggregates of observable high-frequency data on the indicators. Secondly, and subsequent to parameter estimation, estimates of the unobserved high-frequency interpolands may then be produced by means of the high-frequency regression equation using data on the observable high-frequency indicators.[9]

Chow & Lin (1971) suggested that the quarterly estimates of the interpoland should be regressed on the quarterly aggregates of the monthly indicators. The estimated regression equation can then be used to

[8] A full account of the method, together with an investigation of its properties *via* a set of Monte Carlo experiments, is provided in Salazar, Smith & Weale (1997).

[9] Corrado (1986) describes an application of Fernandez's method to the United States' National Accounts. Similar regression-based methods are used by some statistical offices (e.g. in France and Italy) in the construction of quarterly data.

‘forecast’ the interpoland on a monthly basis with least-squares adjustment of the type suggested by Stone, Meade & Champernowne (1942) used to ensure the monthly forecasts of the interpolands are consistent with the known quarterly totals. The basis for the extension of this method by Fernandez (1981) uses essentially the same approach.

There are a number of 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. Many macro-econometric regression equations, however, are often expressed in logarithms in an attempt to mitigate problems of heteroscedasticity. An immediate and obvious difficulty is that the logarithms of three monthly estimates do not add up to the logarithm of the quarterly estimate. Secondly, this method, because it pre-dates much of the work on dynamic modelling, does not accommodate the possibility of 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, an *a priori* 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 these shortcomings and is therefore an important generalization of Chow and Lin’s method. A third difficulty is the assumption that the explanatory variables are exogenous influences on the interpoland. A concern about simultaneity arises, of course, in any regression setting where the regressors are contemporaneous with the regressands. In principle it should be possible to devise an instrumental variable version to deal with the this issue, although we have not done so here as it is beyond the scope of the current paper. An alternative approach based on an extension of the bi-variate method described by Harvey & Chung (2000) which assumes that both low and high frequency variables are driven by underlying latent variables might also prove to be efficacious.

The problem addressed in this paper then is the estimation of the unobserved monthly interpoland $y_{t,m}$, where the subscript t indicates the particular quarter and m the month within that quarter, $m = 1, 2, 3$, $t = 1, ..., T$. The basis for interpolation is a monthly regression equation linking $y_{t,m}$ to the observed monthly indicator variables $x^j_{t,m}$, $j = 1, ..., k$. Firstly, to allow for the possibility that the dependent variable in the monthly regression equation is a non-linear function of the interpoland $y_{t,m}$, let

$$h_{t,m} \equiv h(y_{t,m}),$$

where the function $h(.)$ is assumed to be known, for example, the logarithmic transformation. Of course, the exogenous indicator variables $x^j_{t,m}$, $j = 1, ..., k$, in the regression may themselves also be transformations of other underlying variables. Secondly, consider a dynamic monthly regression equation defined in terms of $h_{t,m}$ linking the k observed indicator variables $x^j_{t,m}$, $j = 1, ..., k$, to the unobserved monthly interpoland

$y_{t,m}$:

$$\alpha(L)h_{t,m} = \beta_0 + \sum_{j=1}^{k} \beta_j(L)x^j_{t,m} + \epsilon_{t,m},\ m = 1, 2, 3, t = 1, ..., T. \tag{3.1}$$

where L is the monthly lag operator and $\alpha(L) = 1 - \sum_{i=1}^{p} \alpha_i L^i$ and $\beta_j(L) = \sum_{i=0}^{q_j} \beta_{j,i} L^i$ are scalar lag polynomials of orders p and q_j respectively operating on the unobserved transformed monthly dependent variable $h_{t,m} = h(y_{t,m})$ and the observed monthly indicator variable $x^j_{t,m}$, $j = 1, ..., k$. It is assumed that the lag lengths p and q_j, $j = 1, ..., k$, are chosen sufficiently large so that the error terms $\{\epsilon_{t,m}\}$ may be assumed to be white noise (zero mean, constant variance, and serially uncorrelated) and uncorrelated with lagged values of $h_{t,m}$ and current and lagged values of $\{x^j_{t,m}\}$.

The regression equation (3.1) is quite general. For example, if $\alpha_i = 0$, $i = 1, ..., p$, then the model is static in the level of $h_{t,m}$. Other values for the parameters $\{\alpha_i\}$ allow a general specification of the dynamics in (3.1). In the special case in which the sum of the coefficients on the dependent variable and its lags is zero, that is, $\sum_{i=1}^{p} \alpha_i = 1$, the lag polynomial $\alpha(L)$ on the left hand side of (3.1) may be re-expressed as a scalar lag polynomial of order $p - 1$ operating on the first difference of the dependent variable $h_{t,m}$. For example, if $\alpha_1 = 1$ and $\alpha_i = 0$, $i = 1, ..., p$, then the model involves the monthly first difference of $h_{t,m}$. When $\sum_{i=1}^{p} \alpha_i \neq 1$, there is a long-run relationship linking the levels $h_{t,m}$ and $x^j_{t,m}$, $j = 1, ..., k$; in particular, if $h_{t,m}$ and $x^j_{t,m}$, $j = 1, ..., k$, are difference stationary, there exists a co-integrating relationship between $h_{t,m}$ and $x^j_{t,m}$, $j = 1, ..., k$. Furthermore, in this case, a test of the restriction $\sum_{i=1}^{p} \alpha_i = 1$ corresponds to a test of the null hypothesis that there is no co-integrating relationship linking $h_{t,m}$ and $x^j_{t,m}$, $j = 1, ..., k$; see Engle & Granger (1987).

Estimation of the unknown parameters in (3.1) is not completely straightforward. The monthly variables $h_{t,m} = h(y_{t,m})$ are not observed whereas we do observe the quarterly aggregates of the interpolands $\{y_{t,m}\}$

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

Firstly, therefore, we need to aggregate (3.1) in such a manner as to yield a regression equation involving only the observable quarterly aggregates y_t, $t = 1, ..., T$. Secondly, we will need to deal with the implications of aggregation for the error structure of the resultant regression equation. For simplicity of exposition, we deal here only with the case in which the maximum lag length of $\alpha(L)$ is unity, that is, $p = 1$. Hence, we write $\alpha(L) = 1 - \alpha L$.[10]

To achieve the first objective, we note that $(1 + \alpha L + \alpha^2 L^2)(1 - \alpha L) = 1 - \alpha^3 L^3$. Thus, we may transform (3.1) into a regression equation involving only third-order lags of $h_{t,m}$ by pre-multiplying (3.1) by $(1 + \alpha L + \alpha^2 L^2)$ which yields

[10] In the application we also 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 described here. See Salazar et al. (1997) for further details.

$$h_{t,m} = \alpha^3 h_{t-1,m} + (1+\alpha+\alpha^2)\beta_0 + (1+\alpha L+\alpha^2 L^2)\left(\sum_{j=1}^{k}\beta_j(L)x_{t,m}^j + \epsilon_{t,m}\right). \quad (3.2)$$

Consequently, in the transformed regression equation (3.2), $h_{t,m}$ depends on its value three months previously as well as on the exogenous indicator variables. Aggregating (3.2) across quarter t:

$$\sum_{m=1}^{3} h_{t,m} = \alpha^3 \sum_{m=1}^{3} h_{t-1,m} + 3(1+\alpha+\alpha^2)\beta_0 + \sum_{m=1}^{3}\left\{(1+\alpha L+\alpha^2 L^2)\left(\sum_{j=1}^{k}\beta_j(L)x_{t,m}^j + \epsilon_{t,m}\right)\right\} \quad (3.3)$$

If (3.3) had been expressed in terms of $y_{t,m}$ rather than $h_{t,m} = h(y_{t,m})$ it would now involve the quarterly endogenous variable y_t and thus be feasible for estimation. To obtain an operational formulation of (3.3), we exploit Taylor's theorem which yields $h(y_{t,m}) = h(\bar{y}_t) + h'(\bar{y}_t)(y_{t,m} - \bar{y}_t) + O[(y_{t,m} - \bar{y}_t)^2]$, where $\bar{y}_t = y_t/3$ is the monthly average in quarter t and $h'(\cdot)$ is the first derivative of $h(\cdot)$. Hence, $h(y_{t,m}) \doteq h(\bar{y}_t) + h'(\bar{y}_t)(y_{t,m} - \bar{y}_t)$ with an error of approximation which is second order in the difference $y_{t,m} - \bar{y}_t$, $m = 1, 2, 3$. Aggregating across quarter t, we have:

$$\begin{aligned}\sum_{m=1}^{3} h_{t,m} &\doteq 3h(\bar{y}_t) + \sum_{m=1}^{3} h'(\bar{y}_t)(y_{t,m} - \bar{y}_t) \\ &= 3h(\bar{y}_t),\end{aligned} \quad (3.4)$$

as the errors of approximation sum to zero, *viz.* $\sum_{m=1}^{3}(y_{t,m} - \bar{y}_t) = 0$. This approximation will be particularly accurate if the errors of approximation $y_{t,m} - \bar{y}_t$, $m = 1, 2, 3$, are relatively small or if the function $h(\cdot)$ is slowly changing across the monthly values $y_{t,m}$, $m = 1, 2, 3$, in quarter t.[11] For a logarithmic transformation, (3.4) becomes:

$$\sum_{m=1}^{3} \log y_{t,m} \doteq 3\log y_t - 3\log 3, \quad (3.5)$$

which is equivalent to replacing the geometric mean of the monthly values $y_{t,m}$, $m = 1, 2, 3$, by the arithmetic mean $\bar{y}_t$. 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. The approximation is likely to be better than a conventional first-order approximation because the sum of the errors of the approximation is itself close to zero. Experimentation suggests that for a variable like constant price GDP which is unlikely to change by more than a percentage point in any month (3.5) is a very good approximation and that higher order terms raise no concerns.

Consequently, the substitution of the approximation (3.4) into (3.3) provides a regression equation feasible for estimation. Note that the covariance structure of the error terms in (3.3) is a function of the

[11] By the mean value theorem $h(y_{t,m}) = h(\bar{y}_t) + h'(y_{t,m}^*)(y_{t,m} - \bar{y}_t)$, where $y_{t,m}^*$ lies between $y_{t,m}$ and $\bar{y}_t$. If $h'(y_{t,m}^*)$ is approximately constant, $m = 1, 2, 3$, then $\sum_{m=1}^{3} h_{t,m} \doteq 3h(\bar{y}_t)$.

parameter α. In the results that follow, we use maximum-likelihood estimation.[12]

Appendix B sets up the interpolation problem described above in state-space form for comparison purposes with the approach suggested by Harvey & Pierse (1984) and Harvey (1989). The state-space reformulation may then be readily estimated using the Kalman filter, but, this reformulation suffers from being less transparent relative to the more familiar regression framework exploited above.

4 Interpolation without Indicator Variables

Table 1 indicates that there are some sectors, Agriculture and Public Services, for which there are no obvious indicator variables available. One could nevertheless estimate (3.3) as a pure autoregression. However, the application of this approach to the public sector raised an interesting practical problem. The estimated coefficient α^3 in (3.3) was of the order of -0.2 when estimated on quarterly data. Extracting the cube root to estimate α in (3.1) yielded a 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. Consequently, we felt it would be better to seek a method which preserved the generally-accepted smoothness of the series.

We therefore constructed preliminary estimates of the monthly data $\{\tilde{y}_{t,m}\}$ from a simple two-sided moving average filter employing equal weights in terms of the monthly averages; *viz.*

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

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

$$\Delta_1 h_{t,m} = \Delta_1 h(\tilde{y}_{t,m}) + \epsilon_{t,m}, \quad (4.1)$$

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

5 Reconciliation of the Interpolands

The estimators of the parameters of the monthly regression equation (3.1) may then be used to produce fitted values of the interpolands $\{y_{t,m}\}$. These fitted values, however, need to be reconciled with the observed quarterly data $\{y_t\}$. Our estimate of $\{y_{t,m}\}$ minimises the sums of squares of the residuals in the regression equation (3.1) subject to the constraint that the interpolated monthly values in each quarter sum to the known quarterly totals, that is, $\sum_{m=1}^{3} y_{t,m} = y_t$.

[12] Salazar et al. (1997) study the properties of the ML estimators of the parameters β_0, $\{\alpha_i\}$ and $\{\beta_{j,k}\}$ *via* Monte-Carlo experiments when the error terms $\{\epsilon_{t,m}\}$ are independently and identically distributed normal variates. The ML technique peformed well on samples of the size available in practice.

For simplicity, we again confine attention to the first order case; *viz.* $p = 1$. We assume that observations are available on the quarterly totals y_t for quarters $t = 1, ..., T$. Firstly, recall (3.1)

$$h_{t,m} = \alpha h_{t,m-1} + \beta_0 + \sum_{j=1}^{k} \beta_j(L) x_{t,m}^j + \epsilon_{t,m}, \tag{5.1}$$

where, for the first quarter, $t = 1$, $m = 2, 3$ and, for the remainder, $m = 1, 2, 3$, $t = 2, ..., T$. The problem then reduces to optimising the Lagrangean

$$\sum_{m=2}^{3} \epsilon_{1,m}^2 + \sum_{t=2}^{T} \sum_{m=1}^{3} \epsilon_{t,m}^2 + \sum_{t=1}^{T} \lambda_t \left(\sum_{m=1}^{3} y_{t,m} - y_t \right), \tag{5.2}$$

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

$$h'(y_{t,m})(\epsilon_{t,m} - \alpha\epsilon_{t,m+1}) + \lambda_t = 0, m = 1, 2, 3, t = 1, ..., T, \tag{5.3}$$

where $\epsilon_{1,1} = 0$ and $\epsilon_{T+1,1} = 0$.

Equation (5.3) can be solved jointly with the adding-up constraints, $\sum_{m=1}^{3} y_{t,m} = y_t$, $t = 1, ..., T$, to produce estimates of the interpolands $\hat{y}_{t,m}$, $m = 1, 2, 3$, $t = 1, ..., T$, and the Lagrange multipliers, $\hat{\lambda}_t$, $t = 1, ..., T$. The solution is inherently nonlinear because the derivatives $h'(.)$ in (5.3) are a function of the estimated interpolated data $\{\hat{y}_{t,m}\}$, which, in principle, necessitates the use of iterative methods. However, when the transformation $h(.)$ is logarithmic, our experience indicates that the derivatives $h'(.)$ in (5.3) 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 (5.3) are presented in Salazar et al. (1997).

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 $\{\hat{y}_{t,m}\}$. Including only terms of order $O_p(1)$, the source of error due to the estimation of the regression parameters is asymptotically irrelevant. Hence, only the random component represented by the error terms $\{\epsilon_{t,m}\}$ is pertinent. Details of the requisite calculations are provided in Salazar et al. (1997).

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 $\epsilon_{t,m}$ given in (4.1) into the Lagrangean (5.2). Details for the calculations of the interpolands and their approximate variances are set out in Appendix A.

6 Monthly Estimates of Constant Price GDP

The components of output fall into three categories in the calculation of monthly GDP. For industries covered by the index of production, the index values simply indicate monthly output. Of the remaining four industries identified, we are, for Agriculture, Construction and Public Services[13] forced to use the approach described in section 4, while for Private Services we use the indicator variable method of section 3.

Agriculture, Construction and Public Services

There are no relevant monthly indicator variables for Agriculture and Public Services. Moreover, there is no obvious reason to presume that the output of Agriculture or Public Services should be closely linked to the various monthly data which do exist. For Construction, although there are monthly indicators, the only timely monthly indicator is the survey run on behalf of the European Commission (DG-ECFIN) which asks construction firms whether the trend of output in recent months has been rising, stable or falling. We were unable to find a satisfactory model built round this; the reason may well be that the responses to these questions can lag output movements substantially (Weale 2004).

Variable	Code	x	Order	$\Delta_1 x$	Order	$\log(x)$	Order	$\Delta_1 \log(x)$	Order
Agriculture	gdqa	-3.165	1	-7.541	3	-3.471	1	-7.580	3
Construction	gdqb	-1.353	0	-5.484	1	-2.164	2	-5.656	1
Public Services	puse	-1.351	0	-8.394	0	-1.819	0	-8.891	0

Sample period 1976Q4-2003Q4.
The heading x indicates that the test is applied to the raw series. The headings $\Delta_1 x$, $\log(x)$ and $\Delta_1 \log(x)$ indicate that the tests are applied to these transformations of the raw data. When applied to the (log) first differences, $\Delta_1 x$ or $\Delta_1 \log(x)$, ADF test statistics are calculated using ADF regressions with an intercept and lagged first differences of the dependent variable. When applied to (log) levels, x or $\log(x)$, ADF regressions with an intercept, a linear time trend and lagged first differences of the dependent variable are used. The relevant lower 5% present and -3.45 when it is present. The lag order, "Order" is chosen by BIC with a maximum order of 24 for monthly variables and 12 for quarterly variables. All regressions are estimated over the same sample period.

Table 3: ADF Tests for Agriculture, Construction and Public Services

Table 3 presents augmented Dickey-Fuller (ADF) statistics for the output of Agriculture, Construction and Public Services. The sample period of the ADF tests depends on the lag order used. In this and subsequent tables of ADF statistics the sample periods shown relate to the tests with the highest lag order given in the tables. The lag order selected is that of the ADF regression equation which delivers the lowest BIC value with BIC calculated from the standard error of the regression equation.

This table suggests that these components of the output index are stationary in logarithmic first differences. Accordingly, we apply the method set out in section 4 and Appendix A, interpolating the data on the basis of the quarter-on-quarter growth rate (assumed to be measured between the middle months of

[13]For the precise meaning of the terms public and private services see section 2.

adjacent quarters), minimising the sum of the squared month-on-month changes subject to the requirement that the monthly data add to the quarterly estimates.

The equation standard error for the model of section 4 is estimated using the procedure described in Appendix A. We find that for Agriculture the estimated standard error is 0.41%, while for Public Services it is 0.10%. However, these standard error estimates apply to the percentage growth rate from one quarter to the next before the adding-up constraints are taken into account. After due allowance is made, we find that the average estimated 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 Services. The average monthly errors in the rates of change have to be calculated as shown in Appendix A. These are 0.33% for Agriculture and 0.08% for Public Services.

Private Services

The output of Private Services was interpolated by means of indicator variables. A preliminary search on quarterly data (making no adjustment for the moving-average error process present in equation (3.3)) suggested that the growth in Private Services output was related to growth in retail sales and to growth in manufacturing output (but not to the movements in the other components of the index of production). The qualitative data on recent output movements in the sector produced by the European Commission have been available only since 1997 and are therefore not used in our modelling.

Table 4 shows ADF statistics for the variables concerned. Once again we are able to accept the hypothesis that the variables of interest are $I(1)$. No co-integrating vector could be identified linking the variables, the hypothesis that the lagged level variable did not enter the equation in differences ($\alpha = 1$ in equation 3.2) was easily accepted, $\chi^2_1 = 0.06$. Such an outcome should not be too surprising; it merely indicates that the components of this sector which are not reflected in our indicators may follow their own individual stochastic trends. In consequence, the underlying monthly equation links first differences in the indicator variables to first differences in the interpoland.

Table 5 displays the resultant estimated regression equation. This equation is possibly the most important tool in the interpolation of constant price GDP. The diagnostic tests all appear satisfactory and the within-sample fit is good, with an R^2 of 0.60 and an estimated standard error of around 0.4%.[14] The estimated equation is expressed in monthly terms, linking observed monthly changes in the indicator variables to the unobserved monthly changes to Private Services output.

[14]If $\alpha = 1$, equation (3.3) can be estimated by generalised least squares since the structure of the variance matrix of the residuals is known. The diagnostic tests are calculated after transforming the independent and dependent variables which renders the transformed equation errors as independently and identically distributed.

Variable	Code	x	Order	$\Delta_1 x$	Order	$\log(x)$	Order	$\Delta_1 \log(x)$	Order
Private Services	pss1	-1.251	2	-4.718	1	-2.328	2	-5.215	1
Retail Sales	rett	-0.173	2	-17.949	1	-2.503	2	-18.784	1
Manufacturing	ckyy	-2.208	1	-25.285	0	-2.188	1	-25.206	0
Import Volume	imvo	-1.588	2	-21.262	1	-2.840	2	-19.670	1

Sample Period: Quarterly Variable 1976Q4-2003Q4 Monthly Variables 1975M4-2003M12.
See Table 3 for other notes.

Table 4: ADF Tests for Private Services and Related Monthly Indicators

Dependent Variable is $\Delta_1 \log$(Private Services)

Variable	Code	Coeff.	t-value	s.e.
Constant		0.005	6.401	0.001
$\Delta ln\ RetailSales_{-0}$	rett	0.370	7.959	0.047
$\Delta ln\ Manufacturing_{-0}$	ckyy	0.289	4.058	0.071
$\Delta ln\ Manufacturing_{-1}$	ckyy	-0.247	-2.220	0.112
$\Delta ln\ Manufacturing_{-2}$	ckyy	0.245	3.235	0.076
	DW	R^2	s.e.	
	2.306	0.6031	0.004534	

Sample Period: 1973Q3 to 2003Q4.

Chow Test (forecast adequacy) F(4,113) = 1.033 [0.3933]
Theil test based on forecast mean = 1.024
Theil test based on lagged value = 1.5
Bera-Jarque normality test = 2.635 [0.2678]
Serial correlation: F(1,116) = 2.89 [0.09182] F(4,113) = 0.7498 [0.5601]
ARCH test: Chi(1) = 0.2253 [0.635] Chi(4) = 4.262 [0.3717]
Chow test (parameter stability), F(5,112) = 1.448 [0.2126]
MSE of estimate of level data : 0.2474%
MSE of month-on-month growth rate [in % points] : 0.3691%
MSE of rolling quarter-on-quarter growth rate [in % points]: 0.1661%

Table 5: Regression Equation for Private Sector Services

7 The Residual Error and Adjustment to Market Prices

Until 1990 there were three published estimates of GDP calculated from output, income and expenditure data. Since then there has been a single measure of GDP at basic prices, 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 because the former takes into account information from income and expenditure data.[15]

We deal with this problem by aggregating as described in section 2. We use the published weights shown in Table 1 to aggregate the four monthly interpolated series described above together with industrial production which thereby gives an estimate of the output measure of GDP, GDP(O). We then use this measure as an indicator variable to interpolate the published measure of GDP at basic prices, GDP(B), using the method described in section 5, but imposing a unit coefficient on GDP(O) reflecting a very strong prior on the nature of the underlying relationship.[16] Thus we assume an underlying relationship between the monthly variables of the form

$$\Delta_1 \log(GDP(B)_{t,m}) = \Delta_1 \log(GDP(O)_{t,m}) + \varepsilon^B_{t,m}.$$

where $\varepsilon^b_{t,m}$ is a disturbance term initially assumed to be zero but then estimated through solving the Lagranean problem of the form given by equation (5.2).We also have to move from GDP at basic prices, GDP(B), to the headline figure, GDP at market prices, GDP(M). We deal with this in the same way. We assume that percentage changes are matched, one for one, but with $\varepsilon^M_{t.m}$ a disturbance term,

$$\Delta_1 \log(GDP(M)_{t,m}) = \Delta_1 \log(GDP(B)_{t,m}) + \varepsilon^M_{t,m}.$$

We then impose the constraint that the interpolated market price data are consistent with the quarterly totals.

The estimated standard errors of the interpolated data depend on the period in question. Table 6 presents the standard errors for GDP(O), 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.

Table 7, in turn, shows the mean of the standard errors of the estimates of the published measure of GDP, GDP(M). These figures reflect the extra components arising from the discrepancy between GDP(O) and GDP(B) as well as the gap between GDP(B) and GDP(M).

[15] See, for example, footnote 2, p. 28, in *Economic Trends*, July 1996.

[16] In any case estimation would be complicated by the fact that discrepancies are usually somewhat greater with more recent data and are subsequently revised away.

Month in Quarter	Levels (%)	Month on Month (%)	Quarter on Quarter (%)
1	0.1113	0.1308	0.1042
2	0.0869	0.1654	0.1042
3	0.1113	0.1660	0.0000

Table 6: Standard Errors in GDP(O)

Month in Quarter	Levels (%)	Month on Month (%)	Quarter on Quarter (%)
1	0.1295	0.1517	0.1214
2	0.1006	0.1922	0.1214
3	0.1295	0.1922	0.00

Table 7: Standard Errors in GDP(M)

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(M) are slightly larger than those of the output measure GDP(O) because the latter is used as an interpolator for the former. The standard error of GDP(M) therefore reflects the standard error of GDP(O) and an additional interpolation error.

8 A Time Series of Monthly and Rolling Quarterly Estimates of GDP

Figure 1 shows our estimates of the seasonally adjusted monthly rate of growth of the UK economy from January 1985 to March 2004. The underlying data were used by Artis (2002) in his analysis of the business cycle in the United Kingdom. A feature of the data that may cause concern is the short-term noise which they show. This arises from the volatility of industrial production as a direct contributor to GDP and also because it and retail sales (which can also be volatile) are used as indicators for the interpolation of Private Services output. Because of this volatility, users may find the rolling quarterly movement of the series more helpful. This is also shown in figure 1.

Programmes have been written in MATLAB to compute the estimates taking the source data from Datastream. The whole process takes about 20 minutes and the National Institute of Economic and Social Research releases its estimates of monthly GDP in the afternoon of the day on which ONS publishes its figures for industrial production. The time lag means that monthly estimates are produced about five weeks after the month to which they relate, for example, in early April we produce an estimate of GDP in February. The data are released to subscribers one and a half hours ahead of the press release. Academic users should approach mgdp@niesr.ac.uk for free access to the full time-series.

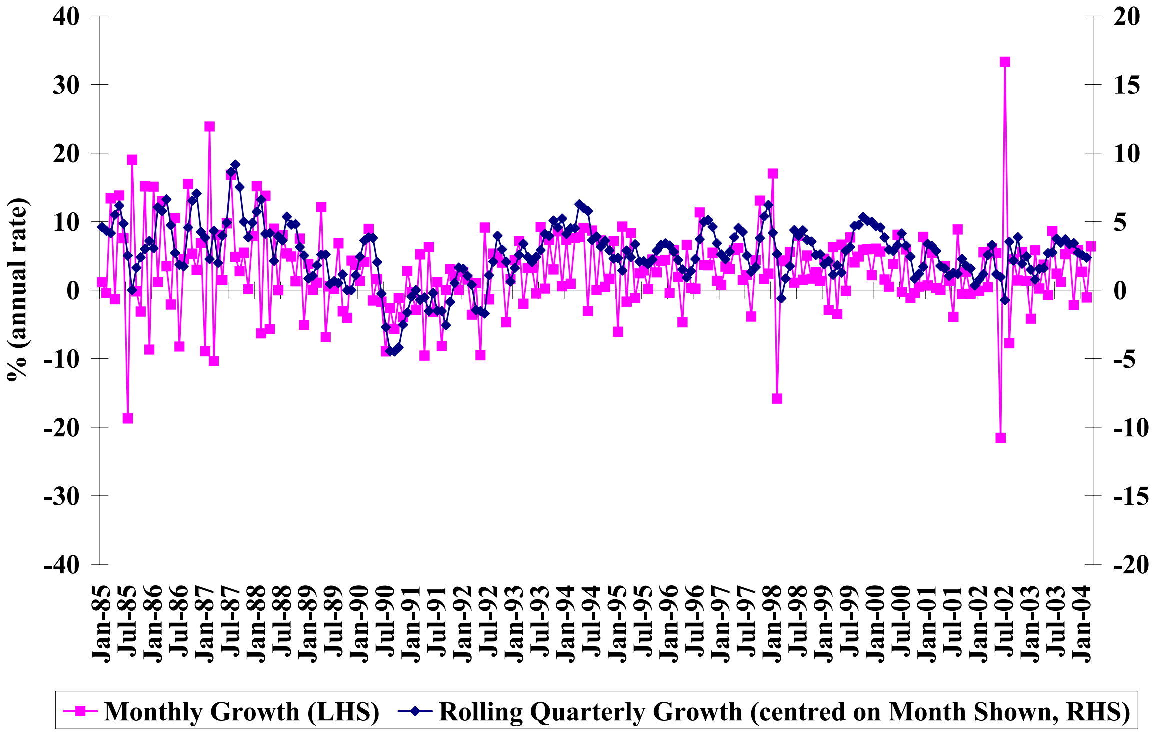

Figure 1: Monthly and Rolling Quarterly Estimates of UK GDP Growth

9 An Early Quarterly Estimate of GDP

Our emphasis on the rolling quarterly estimate raises the question whether an early estimate can be produced of the rolling quarterly GDP growth rate by combining two months of GDP estimates with forecasts for the third month. Even if the forecast for the third month is relatively poor, the estimate of the quarterly growth rate may be satisfactory. We link the rolling quarterly figures to the month in which the rolling quarter ends rather than the month on which it is centered so as to mesh with official quarterly data.

Regular production of such a quarterly estimate implies that after each calendrical quarter an estimate of quarterly GDP growth for the quarter will be produced at the same time as the official estimate of industrial production for the penultimate month of the quarter is published. This is about six days after the end of the quarter in question. It anticipates the first official estimate of quarterly GDP growth by about three weeks. Thus a comparison of our estimates of quarterly GDP growth with the first official data provides a means of assessing the performance of the framework in real time.[17]

There are a number of ways of producing the requisite forecast of monthly GDP. We forecast the monthly input series using autoregressive models in which, because of its explanatory power, the short-term interest rate is included as well as the variable of interest; only one month needs to be forecast so we do not need to concern ourselves with the prediction of the short-term interest rate. The models are re-estimated each month and the lag length of each model is given by the longest lag term which has a statistically significant t-statistic. The forecasts for the monthly variables are used directly in the case of industrial production or as inputs in the production of monthly estimates of Private Services output. Output of Construction and Agriculture is forecast by assuming no change. Output of Public Services is forecast by using a quarterly autoregression. A lag length of unity proved adequate although the resultant model does not have much explanatory power. The forecast generated is then used as an additional observation ahead of the application of the interpolation method of section 4.

We regard the appropriate means of verification of our estimates to be a comparison with the first official estimates on the grounds that the dataset available to us is a subset of the data available when the first ONS estimates are produced. The question of revisions and their predictability is an important but different issue. Figure 2 shows the comparison. We note that the R^2 between the current and lagged first estimates is 0.07 while the corresponding R^2 for the estimates generated by our method is 0.8. Thus our method has an explanatory power which is very much greater than that from a random walk forecasting model; there is therefore little need for a formal test of relative performance.

[17] The information used is a subset of the information available when the first official estimate is produced. Thus the appropriate comparision is with the first and not with the final estimate of GDP growth.

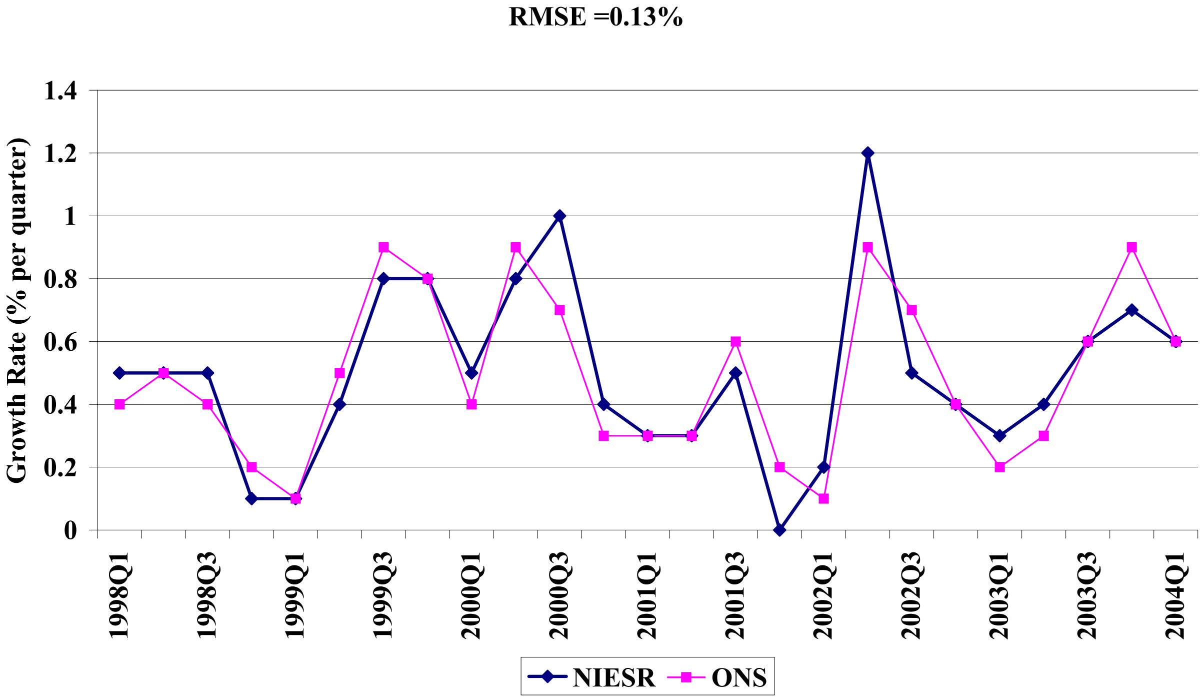

Figure 2: Quarterly Estimates of GDP Growth: a Comparison of NIESR Early Estimates with First Official Estimates

The RMSE is low compared with what the standard errors in table 7 might lead one to expect. Since the exercise is out of sample and the data for the third month are forecast the RMSE might be expected to be markedly higher than the RMSEs for rolling quarters which do not concide with the calendrical quarters used for the official data. However our early estimate uses a larger subset of the data available to the ONS when it produces its first estimate than when the data settle down to their final values. Thus measured against the first ONS estimate out-of sample real time projections it is not so surprising to find an RMSE lower than that deduced within sample from a single vintage of data. Koenig, Dolmas & Piger (2003) come to a similar conclusion for the United States. They find that predicting quarterly US GDP growth from real time estimates of employment, industrial production and retail sales their model has an RMSE of 0.20% points against the first estimate of GDP growth (at a quarterly rate) and a RMSE of 0.27% points against the final estimate

There are only two occasions in the period $1998 - 2003$ when the error was as large as 0.3 percentage points. When running the model in real time ahead of publication in 1997 there was a third occurrence of an error of 0.3 percentage points, in $1997Q3$. Two of these three occasions were associated with the

Royal Family. The first was because the funeral of the Princess of Wales took place in September, the third month of the quarter. This depressed economic activity for one day in September, in a manner that could not have been anticipated by a forecasting model. The second royal occasion was Queen Elizabeth's Golden Jubilee. This led to production being moved forward from June to May. Our forecasting model responded to the high level of output in May and predicted much of it would persist until June whereas in fact there was a sharp fall in output in June. This problem was only partly identified by the time the first official estimates were published. Our estimate of 1.2% growth in 2002Q2 was superseded by a first official estimate of 0.9% growth and a second official estimate of 0.5% growth. The third event which led to an error of 0.3% in our early estimates took place in September 2000. For about a week fuel depots were blockaded by protests about the price of road fuel which depressed output in a way that our forecasting models, again, could not be expected to anticipate.

A lesson to be drawn from these experiences is that there may be scope for improving on the performance of the early estimates of quarterly growth in real time by the introduction of "judgement", i.e. specific knowledge about events which have happened but which are not represented in the data. In 1997Q3 and 2000Q3 it was plain that there had been a negative shock to output in the third month of the quarter. It is not absolutely clear how the size of negative shocks of this sort should be measured but a reduction of 0.1% points to the quarterly growth rate would have been an improvement. In 2002Q2 careful questioning of ONS about working day adjustments might have indicated the problem. Such judgement is, however, a response to rare events not yet reflected in the data available rather than an arbitrary departure from a properly-articulated statistical framework. Hindsight certainly teaches that, in a real-time exercise of this sort, one should be awake to shocks in the forecasts used to produce the data estimates; it is unclear how far this would have improved the performance of the approach in real time.

Finally we note that the ONS first estimates are, like our own, constructed using forecasts of industrial production in the third month of the quarter. If they were to make their forecasting models and the forecasts themselves public it would be possible to improve the performance of the early estimates presented here by ensuring that the industrial production forecasts used were the same as those entering the ONS calculations.

10 Conclusions

Economic policy makers and people in occupations which involve anticipation of changes to economic policy need prompt indicators of the state of the economy and an efficient means of 'nowcasting' by aggregating the disparate pieces of data which appear ahead of estimates of real GDP change. The problem is of course open to formal statistical analysis and we have set out here a coherent and satisfactory means of estimating the current state of economic activity. It produces an economic indicator (monthly GDP) which, unlike some measures based on latent variables, relates directly to standard measures of economic activity. When allied with short-term forecasting techniques it can also be used to produce rolling quarterly estimates of GDP growth about six days after the end of the quarter concerned and therefore, once a calendar quarter, to anticipate official GDP data by about three weeks.

We have done this in real time for the last six years, making the results available each month by press release and with the full series available to academic users. Comparison of our rolling quarterly estimates with the first official estimates of GDP growth published by ONS offers a means of verifying the approach and of assessing the real-time reliability of these estimates. The R^2 with the first official estimates is 0.8 and the root mean square forecast error is 0.13 percentage points. In the light of this information on quality users such as the Monetary Policy Committee can decide how best to make use of the estimated data.

The real time assessment of the system suggests a degree of robustness; the pattern of errors against the official data shows no evidence of heteroscedasticity although one might be concerned that eventually, with the declining importance of the production sector as shown in Table 1, an indicator which relies heavily on the manufacturing output index both directly (in estimating private service output) and as a component of production must eventually lose its utility. The long-term solution to this is obviously for the ONS to improve its collection of data on the service sector to the point where there are timely official monthly indices of both construction and service output. The Statistics Commission has recommended exploring the first as a means of improving the ONS' own early estimates (Mitchell 2004, p.11) and the Allsop Review (Allsopp 2004) has recommended a range of improvements in service sector statistics which may lead to the latter. If or when these happen the problem will be largely reduced to that of forecasting the third month of each quarter from data which extend to the second month; solutions to this problem are well understood and widely discussed.

While it might be thought that the method could readily be extended to other countries, it must be pointed out that the production of UK estimates is greatly facilitated by the fact that value added in the production sector is measured by the industrial production index. This is not typical; in other countries industrial production would simply be an indicator used to produce early estimates of the value added by the industrial sector.[18] A conclusion from our efforts to apply the technique elsewhere is that work of this type is best done by someone with detailed knowledge of the particular characteristics of the data of the country in question and close contact with the statistical office concerned. Nevertheless, the proposed framework for interpolation and short-term forecasting provides a structure for an indicator of economic

[18]Nor is it appropriate. The UK approach assumes that intermediate inputs move in line with outputs in the short term. Producers of early estimates in other countries, e.g. *Deutsches Institut für Wirtschaftsforschung*, model output and intermediate input separately.

activity which is both interpretable and verifiable.

References

Allsopp, C. (2004), *Review of Statistics for Economic Policy-Making.* HM Treasury.

Anderson, B. & Moore, J. (1979), *Optimal Filtering*, Englewood Cliffs: Prentice-Hall.

Artis, M. (2002), 'Dating the business cycle in britain', *National Institute Economic Review* (182), 90–95.

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. (1987), 'Co-integration and error correction: Representation, estimation and testing', *Econometrica* **55**, 251–76.

Fernandez, R. (1981), 'A methodological note on the estimation of time series', *Review of Economics and Statistics* **63**, 471–478.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2001), 'Coincident and leading indicators for the euro area', *Economic Journal* **111**(471), 62–85.

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.

Harvey, A. & Chung, C.-H. (2000), 'Estimating the Underlying Change in Unemployment in the UK', *Journal of the Royal Statistical Society Series A* **163**, 303–339.

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.

Jazwinski, A. (1970), *Stochastic Processes and Filtering Theory*, New York: Academic Press.

Koenig, E., Dolmas, S. & Piger, J. (2003), 'The Use and Abuse of Real-Time Data in Economic Forecasting', *Review of Economics and Statistics* **LXXXV**, 618–628.

Mitchell, J. (2004), *Revisions to Official Statistics: Review by National Institute of Economic and Social Research*. Statistics Commission Report No 17, Volume 2.

National Accounts, Concepts, Sources and Methods (1998), Office for National Statistics, London.

O'Dea, D. (1975), *Cyclical Indicators for the Postwar British Economy*, C.U.P. NIESR Occasional Paper No. 28.

Office for National Statistics (2004), Gross Domestic Product: Preliminary Estimate 1st Quarter 2004. Press Release, 23rd April.

Palm, F. C. & Nijman, T. E. (1984), 'Missing observations in the dynamic regression model', *Econometrica* **52**, 1415–35.

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.

Stock, J. H. & Watson, M. (2002), 'Macro-economic Forecasting using Diffusion Indices', *Journal of Business and Economic Statistics* **20**, 147–162.

Stock, J. H. & Watson, M. W. (1989), 'New indices of coincident and leading economic indicators', *NBER Macroeconomics Annual* pp. 351–394.

Stone, J. (1947), 'On the interdependence of blocks of transactions', *Journal of the Royal Statistical Society Supplement* **9**(1), 1–32.

Stone, J., Meade, J. & Champernowne, D. (1942), 'The precision of national income estimates', *Review of Economic Studies* **9**, 111–125.

Weale, M. (2004), 'Business Surveys and Manufacturing Output', *National Institute Economic Review* (188), 48–49.

A Interpolation without Indicator Variables

A.1 Data Estimation

When there are no indicator variables available, we adopt the following model as in (4.1):

$$\Delta_1 h_{t,m} = \Delta_1 h(\tilde{y}_{t,m}) + \epsilon_{t,m}, \tag{A.1}$$

$m = 2, 3$, $t = 1$, $m = 1, 2, 3$, $t = 2, ..., T$, where $\Delta_1 = 1 - L$ is the first difference operator and $\tilde{y}_{t,m}$ denotes the monthly data constructed by the crude interpolation method described in section 3. Effectively, (A.1) corresponds to (3.1) with $\alpha = 1$ and $\beta_0 + \sum_{j=1}^{k} \beta_j(L) x_{t,m}^j$ set equal to $\Delta_1 h(\tilde{y}_{t,m})$. The solution for the interpolands, $y_{t,m}$, $m = 1, 2, 3$, $t = 1, ..., T$, may be obtained from equation (5.2).

A.2 An Approximate Variance for the Interpolands

In order to estimate the error variance σ_ϵ^2, we proceed as in section 3. Firstly, multiply (A.1) by the lag polynomial $1 + L + L^2$. Hence,

$$\Delta_3 h_{t,m} = \Delta_3 h(\tilde{y}_{t,m}) + u_{t,m}, \tag{A.2}$$

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

$$\sum_{m=1}^{3} h_{t,m} - \sum_{m=1}^{3} h_{t-1,m} = \sum_{m=1}^{3} h(\tilde{y}_{t,m}) - \sum_{m=1}^{3} h(\tilde{y}_{t-1,m}) + \sum_{m=1}^{3} u_{t,m}. \tag{A.3}$$

Applying the approximation (3.4) to (A.3) results in

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

where $u_t = \sum_{m=1}^{3} u_{t,m}$. 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 $[\sum_{t=2}^{T} u_t^2 + 2\sum_{t=3}^{T} u_t u_{t-1}]/[27(T-1) - 8]$.

B A State-Space Formulation

This appendix relates the regression-based approach of sections 3 and 5 to the state-space approach considered in Harvey & Pierse (1984) and Harvey (1989). For ease of exposition we confine attention to the case when $h(y_{t,m}) = y_{t,m}$ and continue to focus on an $AR(1)$ model for $y_{t,m}$.[19] Both approaches consist of

[19] When $h(y_{t,m}) \neq y_{t,m}$, just as our approach requires an approximation, see equations (3.4) and (3.5), the state-space approach would use the extended Kalman filter to approximate the likelihood function. See Harvey (1989, pp. 160-162).

two-steps, first of all eliminating the "incidental parameters" $\{y_{t,m}\}$ by aggregation in order to consistently estimate α, and, secondly, given the resultant estimator for α, to estimate the interpolands $\{y_{t,m}\}$.

First re-consider the estimable equation (3.3), ignoring for expositional ease any indicator variables,

$$y_t = \alpha^3 y_{t-1} + 3(1+\alpha+\alpha^2)\beta_0 + u_t \tag{B.1}$$

where $y_t = \sum_{m=1}^{3} y_{t,m}$ and u_t has the following moving average (MA) structure:

$$u_t = \epsilon_{t,3} + (1+\alpha)\epsilon_{t,2} + (1+\alpha+\alpha^2)\epsilon_{t,1} + (\alpha+\alpha^2)\epsilon_{t-1,3} + \alpha^2\epsilon_{t-1,2}. \tag{B.2}$$

We now show that equation (B.1) with MA errors (B.2) is in fact identical to that considered in the state-space approach.

To cast eq. (3.1) in state-space form, we define the measurement and transition equations respectively as

$$y_{t,m} = s_{t,m}, \tag{B.3}$$

$$s_{t,m} = \alpha s_{t,m-1} + \epsilon_{t,m}, \tag{B.4}$$

where $s_{t,m}$ denotes the (scalar) state variable and $y_{t,0} = y_{t-1,3}$, $t = 2, ..., T$; cf. Harvey (1989, eq. (6.3.1), p.310) in which $z = 1$, $T = \alpha$ and $R = 1$. Because $y_{t,m}$ is unobserved whereas $y_t = \sum_{s=1}^{3} y_{t,s}$ is, system (B.3)-(B.4) is not estimable. Hence, we define the cumulator $y_{t,m}^f = \sum_{s=1}^{m} y_{t,s}$, cf. Harvey (1989, eq. (6.3.13), p.313). Defining the indicator variable $\psi_m = 0$ if $m = 1$ and 1 otherwise, we augment the transition equation (B.4) with $y_{t,m}^f$ to obtain the revised measurement and transition equations from (B.3)-(B.4):

$$y_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} s_{t,m} \\ y_{t,m}^f \end{pmatrix}, \; m = 3, \; t = 1, ..., T, \tag{B.5}$$

$$\begin{pmatrix} s_{t,m} \\ y_{t,m}^f \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ \alpha & \psi_m \end{pmatrix} \begin{pmatrix} s_{t,m-1} \\ y_{t,m-1}^f \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{t,m} \\ 0 \end{pmatrix}; \tag{B.6}$$

see Harvey (1989, eq. (6.3.16), p.314). The consequent measurement and transition equations for the observables are obtained after repeated substitution for the state variable $s_{t,m}$ in (B.6):

$$y_t = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} s_t \\ y_t \end{pmatrix}, \tag{B.7}$$

$$\begin{pmatrix} s_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha^3 & 0 \\ \sum_{i=1}^{3} \alpha^i & 0 \end{pmatrix} \begin{pmatrix} s_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{r=0}^{2} \alpha^r \epsilon_{t,3-r} \\ \sum_{r=0}^{2} (\sum_{s=0}^{r} \alpha^s) \epsilon_{t,3-r} \end{pmatrix}, \tag{B.8}$$

where $s_{t,3}$ is re-defined as s_t, $t = 1, ..., T$; cf. Harvey eqs. (6.3.20a, 6.3.20b).

ML estimation of α is based on (B.7)-(B.8) *via* the prediction error decomposition; see Harvey (1989).

System (B.7)-(B.8) is easily seen to be equivalent to the observable regression equation (3.3) by substitution for the unobserved state variable $s_{t-1} = \alpha^3 s_{t-2} + \sum_{r=0}^{2} \alpha^r \epsilon_{t-1,3-r}$ in $y_t = (\sum_{i=1}^{3} \alpha^i) s_{t-1} + \sum_{r=0}^{2} (\sum_{s=0}^{r} \alpha^s) \epsilon_{t,3-r}$ of (B.8) noting $(\sum_{i=1}^{3} \alpha^i) s_{t-2} = y_{t-1} - \sum_{r=0}^{2} (\sum_{s=0}^{r} \alpha^s) \epsilon_{t-1,3-r}$. Re-arranging yields the following regression equation in term of observables:

$$y_t = \alpha^3 y_{t-1} + \epsilon_{t,3} + (1+\alpha)\epsilon_{t,2} + (1+\alpha+\alpha^2)\epsilon_{t,1} + (\alpha+\alpha^2)\epsilon_{t-1,3} + \alpha^2 \epsilon_{t-1,2},$$

which is equivalent to eq. (3.3). Of course, estimates of α may still differ in finite samples due to different initialisation assumptions.

Given an estimate for α the Harvey-Pierse approach then computes "smoothed" estimates for the missing observations, $y_{t,m}$, from (B.5)-(B.6); see Anderson & Moore (1979) for details of various smoothing algorithms. The smoothed estimator of the state $\mathbf{s}_{t,m|T} = E[\mathbf{s}_{t,m}|y_1, ..., y_T]$, where $\mathbf{s}_{t,m} = (s_{t,m},\ y_{t,m}^f)'$, exploits information from $t = 1, ..., T$, and delivers the minimum mean squared error estimator of $\mathbf{s}_{t,m}$ given the whole information set; see Anderson & Moore (1979) and Harvey (1989). Specifically (Jazwinski 1970, p. 151) the smoothed estimator $\mathbf{s}_{t,m|T}$, $t = 1, ..., T$, $m = 1, 2, 3$, is such that it minimises the following criterion:

$$\frac{1}{2\sigma_{1,1}^2}\epsilon_{1,1}^2 + \frac{1}{2\sigma^2}\sum_{m=2}^{3}\epsilon_{1,m}^2 + \frac{1}{2\sigma^2}\sum_{t=2}^{T}\sum_{m=1}^{3}\epsilon_{t,m}^2, \tag{B.9}$$

subject to (B.5), where $\sigma_{1,1}^2$ is the variance of the estimation error for the initial value $y_{1,1}$. The criterion (B.9) is identical to (5.2) of section 5 apart from the first term. Therefore, the "smoothed" estimator for $\{y_{t,m}\}$ based on the state-space formulation differs from the estimated interpolands of the main text only in terms of initialisation assumptions.