Econometrics Journal (2001), volume **4**, pp. S56–S90.

An automatic leading indicator of economic activity: forecasting GDP growth for European countries*

GONZALO CAMBA-MENDEZ[†], GEORGE KAPETANIOS[‡**], RICHARD J. SMITH[§] AND MARTIN R. WEALE[‖¶]

[†]*European Central Bank and National Institute of Economic and Social Research (NIESR)*
[‡]*Bank of England*
[§]*University of Bristol and NIESR*
[‖]*National Institute of Economic and Social Research, 2 Dean Trench Str., Smith Square, London SW1P 3HE, UK*
E-mail: mweale@niesr.ac.uk, Web: http://www.niesr.ac.uk/

Received: July 1999

Summary In the construction of a leading indicator model of economic activity, economists must select among a pool of variables which lead output growth. Usually the pool of variables is large and a selection of a subset must be carried out. This paper proposes an automatic leading indicator model which, rather than preselection, uses a dynamic factor model to summarize the information content of a pool of variables. Results using quarterly data for France, Germany, Italy and the United Kingdom show that the overall forecasting performance of the automatic leading indicator model appears better than that of more traditional VAR and BVAR models.

Keywords: *Dynamic factor model, Forecasting, Kalman filter, AR, VAR and BVAR models.*

1. INTRODUCTION

The demand for forecasts of major economic variables remains as strong as ever. In a number of countries monetary policy is set with reference to expectations of output growth and thus prospects for the gap between actual and potential output even though the final target of monetary policy is the control of inflation. This policy framework creates a need for forecasts of economic activity which is then accentuated because of the desire by those dealing in financial markets to anticipate interest rate movements. Traditional econometric models, constructed as systems of equations, remain one important means of producing such forecasts, although academic interest in such models is less than it was. However, as Diebold (1998) argues, non-structural models, which have a tradition reaching back to the 1920s, offer an alternative means of producing such forecasts. Perhaps the most prominent of the non-structural approaches is still exemplified by leading indicators of the type initially constructed by Burns and Mitchell (1946). Both the United Kingdom and the United States statistical offices stopped publishing these in the mid 1990s.

*The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank or the Bank of England.

¶Author to whom correspondence should be addressed.

**Contributed to this work while at NIESR.

However, the OECD continues to publish them despite damning criticism from authors such as Emerson and Hendry (1996) that the statistical methods used are almost certain to deliver inefficient economic forecasts.

One statistical method of non-structural forecasting is via regression equations. Regression equations offer an appropriate means for assessing the forecasting performance of individual variables; see, for example, Davis and Fagan (1997). It is less clear that they offer a satisfactory means of combining a number of indicators to produce a forecast. It is always possible to explain past economic growth reasonably well, using a relatively small number of carefully chosen variables with carefully chosen lags. But there is no reason to expect that such equations will necessarily be good forecasting tools. With a reasonably large number of potential variables, a relatively small number of observations, the ever-present risk of structural change, and without the constraints of any structural framework, the aim of the researcher must be to describe and evaluate a forecasting strategy rather than simply to find an equation which happens to fit the data.

The fundamental difficulty here is that there are typically too few degrees of freedom, which makes it impossible to move from a general to a specific model in the manner familiar from structural modelling. The alternative to relatively arbitrary restrictions is to employ some means of data reduction. Thus, Stock and Watson (1989) used a dynamic factor model to extract a latent variable that could be identified as the state of the economy; they then forecasted this variable using a quite separate vector autoregressive (VAR) model, focussing on the chance of a recession.[1] However, they did not test their model out-of-sample and interest in it declined after its failure to anticipate the recession in the United States in 1991.[2] Our study builds on Stock and Watson's work but regards the final aim of the exercise as being the production of forecasts of real GDP growth for the major European countries. We set out here a method of data reduction using dynamic factor analysis. This ensures that the salient information can be extracted from our indicator variables and used to forecast economic growth. We then test the performance of our forecasts against those generated by two other non-structural forecasting methods. The first comparator is provided by a univariate autoregressive (AR) model of output growth. The second comparator comes from the set of all possible VAR models which could reasonably be constructed from our indicator variables; we also consider Bayesian VAR (BVAR) models of the type described by Doan *et al.* (1984) as part of this comparison. This allows us to assess the performance of our forecasting method against competitors on an out-of-sample basis.[3] Although some of the leading indicator variables employed in this paper appear to display a single unit root, we do not consider cointegrated VAR models among those used for comparison purposes, preferring to work in first differences for reasons discussed in Section 2.

The structure of our paper is as follows. In Section 2 we describe our method of data reduction and our modelling approach. This is followed by an account of how we set up the comparator forecasting methods. Section 4 describes the data and Section 5 the basis for our comparison of

[1]Earlier applications of dynamic factor models to macroeconomic research include Sargent and Sims (1977) and Geweke (1977) who attempted to identify a single force underlying macroeconomic fluctuations. Geweke and Singleton (1981) used a dynamic factor model with two latent variables (factors) to explain the business cycle. At the same time Engle and Watson (1981) used a traditional dynamic factor model to forecast sectorial wage rates in Los Angeles.

[2]Although one could reasonably make the point that the recession was a consequence of the invasion of Kuwait and the Gulf War, which the CIA failed to forecast. Perhaps this demonstrates the difficulty of a system which is designed to forecast particular events such as recessions rather than the continuous variable economic growth.

[3]The tests are out-of-sample, but not 'real time', in that we use the most recent data vintage throughout, which means that we test how well our approach does in a world of random shocks and possible structural changes. We do not examine the separate issue of how well it deals with inaccuracies in earlier vintages of data.

the different models. In Section 6 we discuss the implementation of our method. The results for the four countries we study, France, Germany, Italy and the United Kingdom, appear in Section 7. Finally our findings are summarized in Section 8.

2. THE AUTOMATIC LEADING INDICATOR METHODOLOGY

Our objective is to build an automatic leading indicator (ALI) model for the m-vector of the variables of interest $\mathbf{y}_t$ based on the n-vector of observable exogenous variables $\mathbf{x}_t$. However, rather than use all of the variables $\mathbf{x}_t$ in a VAR model for $\mathbf{y}_t$, it is assumed that their influence on $\mathbf{y}_t$ may be conveniently summarized via k driving forces or *factors* $\mathbf{s}_t$ which underly the fluctuations of the variables $\mathbf{x}_t$; typically one would expect that $k < n$. Consequently, we propose to use the following two-stage procedure to obtain a forecasting equation for $\mathbf{y}_t$. The identification of the dynamic factor model for $\mathbf{x}_t$ is discussed in the Appendix.

2.1. Stage 1: factor extraction

Consider the following dynamic factor model for the n-vector of exogenous variables $\mathbf{x}_t$:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{B}\mathbf{s}_t + \mathbf{u}_t \\ \mathbf{C}(L)\mathbf{s}_t &= \eta_t, \end{aligned} \tag{1}$$

where $\mathbf{B}$ is an (n, k) matrix of unknown parameters and $\mathbf{u}_t$ an n-vector of disturbances with $k \leq n$. The k-vector of factors $\mathbf{s}_t$ follows a stationary AR(p) process described *via* the pth order matrix polynomial $\mathbf{C}(z) = \mathbf{I}_k - \sum_{i=1}^{p} \mathbf{C}_i z^i$ and the k-vector of disturbances η_t; thus, $\det \mathbf{C}(z) = 0$ has all roots outside the unit circle $|z| = 1$. It is assumed that the error processes $\{\mathbf{u}_t\}_{t=1}^{\infty}$ and $\{\eta_t\}_{t=1}^{\infty}$ are mutually and serially uncorrelated, conditionally homoscedastic with mean zero and positive definite variance matrices. The estimation of the unknown parameters in (1) and extraction of the factors $\mathbf{s}_t$ may be combined in the following step-wise fashion. Firstly, given knowledge of $\mathbf{B}$, $\mathbf{C}(L)$ and the variance matrices of $\mathbf{u}_t$ and η_t, (1) may be written in state space form with the Kalman filter used to extract $\mathbf{s}_t$ from observations on $\mathbf{x}_t$. Secondly, given the factors $\mathbf{s}_t$, the parameter matrices $\mathbf{B}$, $\mathbf{C}(L)$ and the variance matrices of $\mathbf{u}_t$ and η_t may be estimated by quasi-maximum likelihood. This step-wise procedure may be iterated until convergence; see Harvey (1993) for further details.

2.2. Stage 2: forecasting $\mathbf{y}_t$

The factors $\mathbf{s}_t$ obtained from Stage 1 above are incorporated into a VAR model to forecast $\mathbf{y}_t$ *via*

$$\mathbf{A}_y(L)\mathbf{y}_t = \mathbf{A}_s(L)\mathbf{s}_t + \varepsilon_t, \tag{2}$$

where $\mathbf{A}_y(z) = \mathbf{I}_m - \sum_{i=1}^{p_y} \mathbf{A}_{yi} z^i$, $\mathbf{A}_s(L) = \sum_{i=1}^{p_s} \mathbf{A}_{si} z^i$ and $\{\varepsilon_t\}_{t=1}^{\infty}$ an m-vector zero mean, conditionally homoscedastic and serially uncorrelated error process with positive definite variance matrix uncorrelated with the error processes $\{\mathbf{u}_t\}_{t=1}^{\infty}$ and $\{\eta_t\}_{t=1}^{\infty}$. Given the estimated factors $\{\mathbf{s}_t\}_{t=1}^{T}$ from Stage 1 above, the (m, m) and (m, k) matrices of unknown coefficients $\{\mathbf{A}_{yi}\}_{i=1}^{p_y}$

and $\{\mathbf{A}_{si}\}_{i=1}^{p_s}$ may be estimated over the sample period, $t = 1, \ldots, T$, by least squares (quasi-maximum likelihood). Note that when forecasting more than one period ahead it will also be necessary to forecast $\mathbf{s}_t$ from the estimated counterpart of (1). This approach differs from that suggested by Stock and Watson (1989) who treat estimation of the (single) factor and its forecast as two quite separate exercises. Subsection 6.3 details tests used to specify the structure of the forecasting equation (2).

2.3. Cointegration, error correction and structural breaks

As detailed above, our analysis requires that the variables $\mathbf{y}_t$ and $\mathbf{x}_t$ are integrated of order zero ($I(0)$) and that the ALI relationships (1) and (2) are structurally stable. Before implementing (1), conventional tests are used to ascertain the probable order of integration of the indicator variables which are employed in Stage 1, in particular, whether they are $I(2)$, $I(1)$ or $I(0)$. Variables which appear to be integrated are then differenced to an appropriate order to produce the indicators $\mathbf{x}_t$ for use in (1). Similarly, at Stage 2, the regressand $\mathbf{y}_t$ in (2) is the growth rate of output which appears to be $I(0)$ rather than the log level of output which standard tests indicate is $I(1)$. Details are provided in Section 4. There are a number of reasons for our working with differenced series rather than also including possible error–correction relationships between integrated variables. Evidence presented by Hoffman and Rasche (1996) demonstrates that the forecasting performance of VAR models may be better than that of error correction models over the short forecasting horizons which concern us. Only over long horizons are error correction models shown to have an advantage. Christoffersen and Diebold (1998) cast doubt on the notion that error correction models are better forecasting tools even at long horizons, at least with respect to the standard root mean square forecasting error criterion. They also argue that although unit roots are estimated consistently, modelling non-stationary series in (log) levels is likely to produce forecasts which are suboptimal in finite samples relative to a procedure that imposes unit roots, a phenomenon exacerbated by small sample estimation bias. Developing this argument, they suggest that for cointegrated series it is better to overestimate rather than underestimate the number of common trends. In a subsequent analysis Diebold and Killian (2000) suggest that the relative performance of models in differences to those in levels is sensitive to the forecast horizon, the strength of any time trend and the magnitude of the autoregressive root and that, in a univariate framework, it is difficult to come to general conclusions about the merits of working in levels rather than differences. However, Clements and Hendry (1996) suggest that VAR models estimated using differenced variables may offer greater protection than error correction models against unforeseen structural breaks by removing deterministic components. For this reason we have decided to ignore the possible contribution of cointegration terms in the construction of forecasts. The framework we describe is not generically robust to the sort of structural breaks which Emerson and Hendry (1996) argue will eventually lead to any system of indicators breaking down. However, because we test a strategy for the production of leading indicators, rather than any single indicator, we can assess how well the strategy copes with those structural breaks which do occur.

3. MODEL SELECTION WITH VAR AND BVAR SYSTEMS

We compare the forecasts generated by the ALI model (1) and (2) with those produced from VAR systems of leading indicators. Each VAR model includes up to five of the indicator variables described in Section 4 together with output growth.[4] VAR system lag selection is determined using the multivariate Bayesian information criterion (BIC); see Schwarz (1978). In VAR systems with three or fewer indicator variables, the maximum lag length employed is eight whereas for more than three variables it is six. All VAR models include an intercept.[5] After lag selection using BIC, VAR models are subjected to the following two specification tests.[6] Only those VAR models which pass them are compared with the ALI model (1) and (2).

- *Granger Causality*. A likelihood ratio statistic is used to test the null hypothesis that the included indicator variables do not Granger cause output growth; see Hamilton (1994). If Granger non-causality cannot be rejected, we may conclude that the indicator variables have no power in explaining output growth.[7]
- *Parameter Constancy*. The CUSUM-OLS test of Ploberger and Kramer (1992) for parameter constancy is employed. Evidence of parameter non-constancy indicates that the VAR model is not well specified and we reject such models.

The use of VAR models for forecasting purposes is often criticized on the basis of overparameterization. It is argued that consequent parameter uncertainty will necessarily induce a poor forecasting performance. BVAR models have been suggested and used as alternatives; see Doan *et al.* (1984) and Litterman (1986). In assessing the forecasting performance of the ALI model (1) and (2), we also examine forecasts obtained from BVAR models. The results below show that the ALI model both avoids the risk of choosing, on the one hand, an incorrect subset of the indicator variables intrinsic in VAR models and, on the other hand, a wrong prior associated with BVAR models.

4. THE DATA

We use a range of indicator variables similar to that adopted by the OECD in its construction of 'traditional' leading indicators. This range is somewhat narrower than that considered for the United States by Stock and Watson (1991, 1993) but reflects the data availability for the countries concerned.

Considerable weight is put on financial indicators. The power of the term structure of interest rates (i.e. the first of the yield spreads defined in Table 1) to forecast real economic activity has been documented by Bernanke (1990), Estrella and Hardouvelis (1991) and Plosser and Rouwenhorst (1994). Davis and Fagan (1997) suggested that a range of financial spreads might be used, because different financial variables may anticipate different aspects of future developments in the real economy. However, they considered only univariate models rather than VAR models, leaving the latter for future research.

Plots of four of the series revealed the presence of outliers for which, in all cases, there was an economic explanation and therefore no formal tests were conducted. The GDP series for Germany

[4] We note that the framework we suggest bears similarities to that discussed by Sullivan *et al.* (2000), Sullivan *et al.* (1999) and White (forthcoming) regarding the effects of data snooping on the evaluation of forecasting performance when large numbers of models are considered.

[5] The restrictions on indicator variable inclusion and maximum lag length are dictated by computational feasibility.

[6] We also checked whether the eigenvalues of the VAR models lay within the unit circle. All models displayed this property.

[7] If a subgroup of the indicators are significantly Granger causal even though the full set appears insignificant, because we search over all possible VAR systems, the VAR model with this subset as indicator variables would be identified.

displays a clear shift in 1989 as a result of German unification. In addition, UK GDP growth, new car registrations and the productivity index present clear outliers for the period 1979Q2. The annualized rate of GDP growth for that period is 20.77%. This was due to a consumption boom in anticipation of the introduction of the unified rate of VAT, which rose from 8% and 12.5% to 15% in June 1979. An intervention model was used in order to filter out these events from the series.[8] All series were tested for the presence of a unit root with first differences taken of those series (either in levels or in logs) where a unit root was indicated. Details of the series used per country are given in Table 1 together with augmented Dickey–Fuller (ADF) statistics. In all cases where the ADF statistic suggested the presence of a unit root, the null hypothesis of a further unit root was rejected at the 5% level by an ADF statistic using the differenced series.

5. ASSESSMENT OF FORECASTING METHODOLOGY VIA FORECAST PERFORMANCE

The basis for comparison of the ALI forecasting methodology (1) and (2) and other competing forecasting techniques is the statistic proposed by Diebold and Mariano (1995), which tests the null hypothesis of equal predictive ability in two competing forecasts generated by different forecasting techniques. Assuming a quadratic loss function for evaluating forecasting performance, the mean of the differences of squared prediction errors of the two competing models is the appropriate reference variable. We make use of the small sample correction proposed by Harvey *et al.* (1997) which results in the revised statistic

$$S^*_{\mathrm{DM}} = N^{-1}\{N + 1 - 2h + N^{-1}h(h-1)\}\hat{V}(\bar{d})^{-1/2}\bar{d},$$

where $\bar{d} = N^{-1}\sum_{i=1}^{N}\hat{d}_i$, $\hat{d}_i = \hat{\eta}^2_{\mathrm{ALI},i} - \hat{\eta}^2_i$, $i = 1, \ldots, N$, $\hat{\eta}_{\mathrm{ALI},i}$ and $\hat{\eta}_i$ are, respectively, the prediction errors from the ALI model (2) and the competing model, N is the number of prediction errors, $\hat{V}(\bar{d}) = N^{-1}(\hat{\gamma}_0 + 2\sum_{i=1}^{h-1}\hat{\gamma}_i)$, $\hat{\gamma}_i$, $i = 0, 1, \ldots, h-1$, are the estimated autocovariances of the series of squared prediction error differences $\{\hat{d}_i\}_{i=1}^{N}$ and h is the prediction horizon.[9] Critical values taken from the t distribution with $N - 1$ degrees of freedom are appropriate under the null hypothesis of equal predictive ability.[10]

The specification tests outlined above resulted in a large number of VAR models capable of producing forecasts for comparison with those generated by the ALI model. Tables A1, A4, A7 and A10 of Appendix 2 display the empirical distribution of the probability values of the predictive accuracy statistic S^*_{DM} for the four countries. For a 5% significance level, values smaller than 0.05 imply that the performance of the ALI forecasts is significantly better, and for values larger than 0.95 that it is significantly worse than the competing VAR forecasts. Hence, if the figure

[8]The structure of the intervention model is based on the structural time series model used by Harvey and Jaeger (1993), augmented to account for outliers. This method does not prejudge the issue of whether each series incorporates a unit root or a deterministic trend and is sufficiently flexible to accomodate an $I(2)$ process. Details are available from the authors on request.

[9]Note that $\hat{V}(\bar{d})^{-1/2}\bar{d}$ is the original Diebold–Mariano statistic. When $n = 1$, only $\hat{\gamma}_0$ is used in $\hat{V}(\bar{d})$.

[10]Note that the ALI and rival models compared in this paper are non-nested. Strictly speaking, the critical values for S^*_{DM} should make an allowance for both pretesting arising from model selection and parameter estimation. See West and McCracken (1998) and McCracken (2000), who assumed that N/T tends to a constant, where $N, T \to \infty$. If $N/T \to 0$, the influence of parameter estimation on the distribution of S^*_{DM} is asymptotically negligible. This latter assumption is more natural here as it accords with the main aim of this paper in comparing the qualities of forecasting methodologies rather than the specific models themselves.

Table 1. Variables used in each country model.[a]

	France	Germany	Italy	UK
Short interest rate	−2.72	−3.12	−2.94	−3.07
Log of real effective exchange rate		−2.94	−2.32	−3.21
Log of dwellings started/authorized	−1.85	−1.93		−1.88
Log share price index	−1.83	−2.80	−2.51	−2.32
Long bond–short rate	−3.23	−4.03	−3.53	−3.29
Long bond–corporate bonds	−6.38			−2.45
Long bond–German long rate	−2.95		−1.87	−2.20
Long bond–US long rate	−3.13	−1.88	−2.57	−2.16
Long bond–mortgage rate		−3.55		
Long bond–Dividend yield				−3.82
Survey: Production tendency	−2.08		−3.32	
Survey: Industrial prospects	−2.74	−4.21	−2.85	
Survey: Order book level		−2.48	−3.02	−2.81
Survey: Capacity utilization	−6.34			
Survey: Optimism index				−2.95
Log of new car registrations	−3.24	−3.47	−3.16	−3.15
Change in hours, services	−3.95			
Consumer confidence		−3.90		−4.29
Log of change in construction costs			−0.42	
Productivity index				−0.07

[a] For each entry the ADF statistic is presented. The critical value is −3.45. The BIC criterion is used for lag order selection.

below the column title, for example, 75%, in a table is less than 0.05, this is evidence that the predictive accuracy of the ALI forecasts is significantly better than that of 75% of the forecasts from VAR models for that country at the 5% level and, conversely, is significantly worse than that of 25% of the forecasts from VAR models if the figure is greater than 0.95. Overall, we interpret a probability value of less than 0.50 against the forecasts from the median (50%) VAR model as evidence in favour of the forecasts from the ALI model and vice versa if it exceeds 0.50.

Because $\hat{V}(\bar{d})$ is indefinite, it may be negative, in which case Diebold and Mariano (1995) suggest that the null hypothesis of equal predictive ability be rejected. Although these cases are likely to be slightly pathological, in such circumstances, we favour methods with the smaller root mean squared forecast error (RMSFE), without being able to attach any probability to whether such methods are better forecasting tools than the alternatives against which they are compared. In other circumstances, the use of RMSFE as a means of assessing forecasts is criticized, for example, by Clements and Hendry (1993). However, users may nevertheless wish to know how the forecasts produced by different tools stand up on that basis. For this reason and in order to deal with the situation when the Diebold–Mariano statistic cannot be calculated, we also provide indications of how RMSFEs arising from the ALI model compare with those generated by a univariate autoregression of order p (AR(p)) and also by VAR models. We display RMSFEs relative to the forecasting performance of an AR(p) model as a benchmark with the lag order p also selected by BIC. Thus, whenever the relative RMSFE of a particular model is greater (less) than 1, its forecasting performance is worse (better) than that of the AR(p) model. VAR models are ranked in terms of their forecasting performance and we establish the proportion of

Table 2. Test failure in VAR models.[a]

	Total	Par. Inst	Granger Causality
		FR (1585 models)	
Period 0	496	1	496
Period 1	236	41	196
Period 2	57	57	0
		GER (1023 models)	
Period 0	673	0	673
Period 1	236	68	168
Period 2	7	7	0
		IT (1023 models)	
Period 0	457	384	152
Period 1	100	65	35
Period 2	126	110	16
		UK (1470 models)	
Period 0	789	33	760
Period 1	12	3	9
Period 2	106	105	1

[a] Period 0, fail to pass test in 90Q2; Period 1, fail test during 90Q2–94Q2; Period 2, fail test during 94Q2–98Q2.

the acceptable VAR models which produce forecasts better than the ALI model; these results are shown along with relative RMSFEs in Tables A2, A5, A7 and A10 of Appendix 2.

We also computed BVAR models in the spirit of Doan *et al.* (1984) and Litterman (1986) based on the Minnesota prior which fixes the prior mean of the VAR parameters to zero and their prior variance depends on two hyperparameters, θ and λ.[11] The values of θ are fixed as 0.2 or 0.8 while the value of λ varies from 0.1 to 0.9. While some practitioners regard BVAR models as ideally suited for forecasting, others point to the risks associated with choosing the wrong priors as a major drawback. It is this latter aspect of BVAR models which we wish to address in the empirical results. The BVAR models chosen for examination are constructed using that set of variables which constitutes the median VAR model obtained above.[12] Results in Tables A3, A6, A9 and A11 compare the forecasting performance of the median VAR models against the BVAR model under alternative priors. Values of the corrected Diebold–Mariano probability values less than 0.05 reject the null of equal predictive accuracy of both models in favour of the median VAR model at the 5% level; similarly, values for relative RMSFEs less than unity favour the median VAR model.

[11] Let $\phi_{i,jk}$ denote the (j,k)th element of the ith lag VAR coefficient matrix. The Minnesota prior variance matrix for the matrices of VAR coefficients is diagonal with $\mathrm{var}\{\phi_{i,jk}\} = (\lambda/i)^2$ if $j = k$ and $(\lambda\theta\sigma_j/i\sigma_k)^2$ if $j \neq k$, $j,k = 1,\ldots,m+n$, where σ_j^2 is the variance of the innovation error in the jth equation of the VAR model, for which the unrestricted VAR estimator is substituted, $j = 1,\ldots,m+n$. See the above references and Lütkepohl (1993) for further details.

[12] Note that the median VAR model differs for different forecast periods and for different forecasting horizons. There is a total of 3 (forecasting periods) $\times$ 4 (forecasting horizon) median VAR models.

6. IMPLEMENTATION

6.1. Estimation period and lag order selection

All models are initially estimated over a sample period which concludes in 1990Q2 with the start being 1975Q2 or slightly later. The remainder of the sample, 1990Q3–1998Q2, provides a period for testing forecasting performance. This out-of-sample period is split into the two subperiods, 1990Q3–1994Q2 and 1994Q3–1998Q2. The first subperiod was one of considerable volatility in most of the countries comprising our sample, while economic growth was less volatile in the second which allows an assessment of model performance in differing economic circumstances. All models are recursively estimated over the forecasting periods. Alternative criteria were employed for the lag order selection in the ALI model which included the Akaike and Hannan-Quinn information criteria. Using these criteria, higher lag orders were suggested as might be expected; moreover, the forecasting performance of the resultant ALI models deteriorated.[13] For this reason and that of parsimony, lag orders were chosen according to BIC. For similar reasons, lag order selection in both VAR and AR models is done using BIC which is re-calculated at each stage of the out-of-sample forecasting process.

6.2. Forecasting horizon

We examine h-step ahead forecasts referred to as 'lead h' in the tables, where $h = 1, \ldots, 4$ denotes the number of quarters ahead. In computing the forecast at time $t + h$, only data to time t are used. Note that in constructing h-step ahead forecasts, $h - 1$ forecast errors over the forecasting period are lost.

6.3. Specification of the automatic leading indicator model

We assumed a VAR(p) structure for the factors $\mathbf{s}_t$ in (1), where the maximum lag was fixed at 4 for reasons of computational feasibility. Selection of the lag order p is achieved using BIC as noted above. The initialization $\mathbf{s}_0$ for the Kalman filter was fixed at the unconditional mean with the variance matrix of $\mathbf{s}_0$ the unconditional variance of the factors. Similarly, at the second stage, the selection of lag orders p_y and p_s in the forecasting equation (2) also uses BIC. The only remaining undetermined order parameter is k, the number of factors. Forecasting results are produced for $k = 1, \ldots, 4$. To distinguish between the various ALI models, we employ the following specification tests.

Firstly, we examine the serial correlation structure of the errors $\mathbf{u}_t$ in the equations linking the dynamic factors $\mathbf{s}_t$ to the indicator variables $\mathbf{x}_t$ to establish whether the factor structure is sufficiently rich enough to absorb the serial correlation in the data. We test the null hypothesis of no serial correlation using the Ljung–Box statistic, computed from the standardized prediction error residuals obtained from the Kalman filter. The Ljung–Box statistic is computed for both the first 12 ($Q(12)$) and 24 ($Q(24)$) autocorrelations. A satisfactory dynamic factor model (1) is obtained if the null hypothesis of no serial correlation is not rejected.

[13] Similar results were obtained when sequential F-tests were employed. Further details are available from the authors upon request.

Secondly, we test whether the factors $\mathbf{s}_t$ Granger cause output growth $\mathbf{y}_t$ in the forecasting equation (2) and also test for parameter constancy.[14] The significance of each factor is assessed separately in a single estimated forecasting equation (2), after the lag lengths p_y and p_s have been estimated, *via* F-statistics with insignificant factors being deleted. This procedure is asymptotically equivalent to removing the least significant of the insignificant factors, re-estimating the equation and repeating the process until the surviving factors are statistically significant because the individual factors in $\mathbf{s}_t$ are independent of each other as the lag matrix $\mathbf{C}(L)$ and the variance matrix of the factor errors η_t are assumed to be diagonal; see the Appendix. A satisfactory forecasting equation (2) is obtained if at least one significant factor is found and parameter constancy is not rejected. Consequently, an ALI model with k factors is said to be well specified if it performs satisfactorily in both Stages 1 and 2.

Resultant well-specified ALI models were then recursively estimated to obtain RMSFEs over the forecasting periods. Figures 1– 4 display the specification test results for each quarter in the out-of-sample period.

7. RESULTS

7.1. France

Figure 1 (upper panels) shows that the Ljung–Box statistic fails to reject the hypothesis of no serial correlation only for dynamic factor models with $k = 3$ and 4 factors. However, this is the case for $k = 3$ only when estimated up to 94Q2; when estimated beyond that period, $k = 4$ is required to capture the serial correlation in the series adequately. In Stage 2, none of the forecasting equations display parameter instability (lower right panel) whereas Granger non-causality (lower left panel) is rejected in all recursive estimations for models with $k = 2$ and 4 factors. However, as the dynamic factor model with $k = 2$ factors is badly specified at Stage 1, the preferred automatic leading indicator for France employs $k = 4$ factors.

Table 2 provides information on the competing VAR models. Out of a possible total of 1585 models estimated over the period up to 1990Q2, 496 fail to pass the tests described in Section 3; all of these models fail to reject the null hypothesis for the Granger non-causality test with one also showing parameter instability. Of the acceptable models, a further 236 fail during the period 1990Q3–1994Q2, with 196 failing to reject the null hypothesis for the Granger non-causality test and 41 showing parameter instability. During the period 1994Q3–1998Q2, an additional 57 models fail because of parameter instability.

Table A1 displays the outcome of the corrected Diebold–Mariano statistic S^*_{DM} which compares the forecast performance of the VAR models with ALI models constructed using $k = 1, \ldots, 4$ factors. As explained in Section 6.3, when more than one factor is available, econometric criteria are used to decide how many factors are actually used in constructing the forecast at each point. The table ranks the VAR models by the corrected Diebold–Mariano statistic S^*_{DM} and shows the corresponding probability values for the minimum, the quartiles and the maximum of this ranking for forecasts of $h = 1, \ldots, 4$ periods ahead.

With such a large number of forecasts generated by the VAR models, it is plainly highly implausible that the ALI models will significantly out-perform all or most of them. However,

[14] Other specification tests for the second-stage equation which are not reported in this paper were LM tests for serial correlation and ARCH errors. All models pass the former but not, generally, the latter tests, which may be because the volatility in GDP is dependent on the state of the cycle.

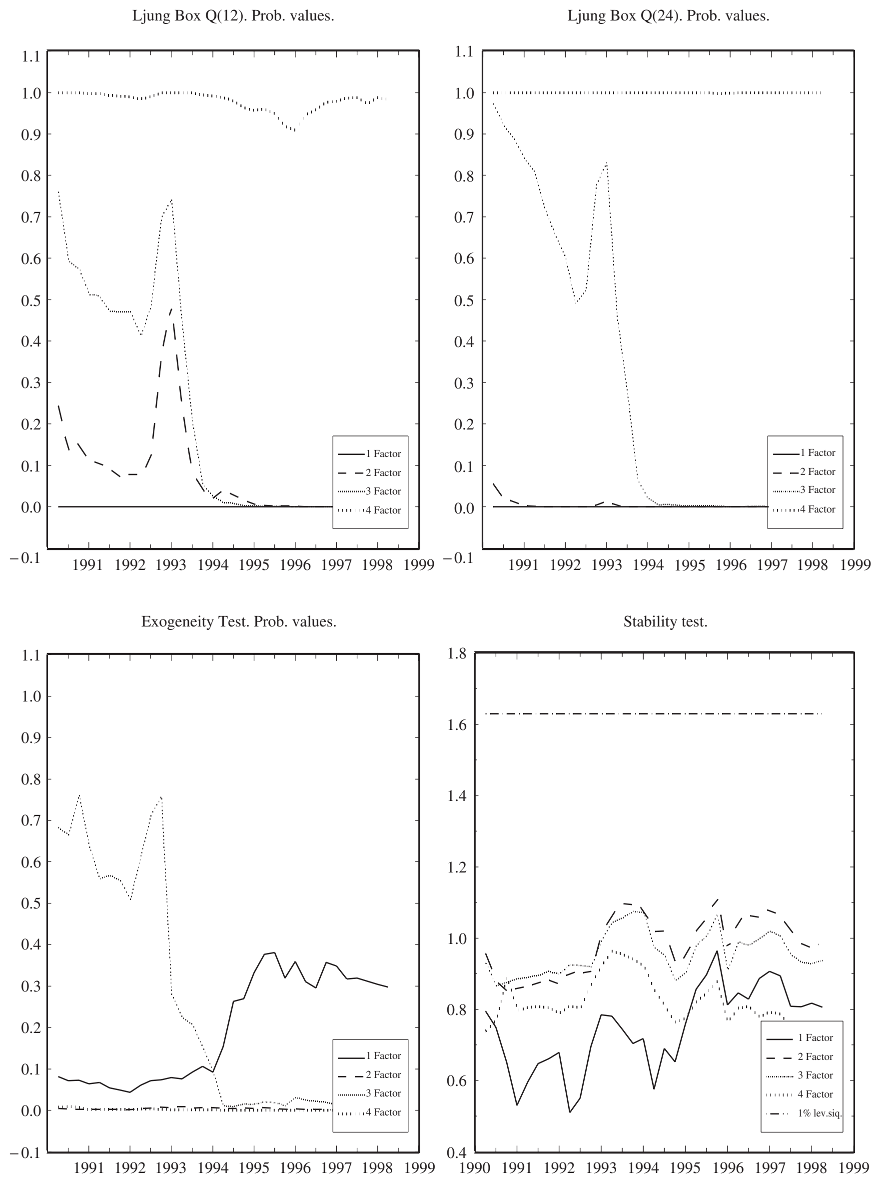

Figure 1. Specification tests for factor models. France

there are no conventional criteria comparable to the usual significance level testing approach by which to identify successful models. But, over the period 1990Q3–1994Q2, we see that, with $k = 4$ factors, for one-step ahead $h = 1$ forecasts, the ALI model significantly out-performs the median VAR at the 0.20 level and the third quartile VAR model at the 0.45 level. It is also the case that the best VAR significantly out-performs the ALI model at the 0.05 level, but this is of no real help to the forecaster without some means of identifying this VAR model *ex* *ante*. Overall, therefore, one might conclude that there is some evidence that the four-factor ALI model dominates the VAR modelling approach for one-step ahead forecasts over the period 1990Q3–1994Q2. Looking further down the table, we can see that this model retains its advantage for one-step ahead forecasts. It is significantly better than the median VAR model at the 0.15 level and the third quartile VAR model at the 0.26 level in the period 1994Q3–1998Q2. On the other hand, for two-step $h = 2$ ahead forecasts, randomly selected VARs would seem likely to do better. The first quartile VAR model is a significantly better forecasting model than the ALI model with $k = 4$ factors at the 0.42 level. Indeed, this part of the table would suggest that, the ALI model with $k = 3$ factors has a better record, at least when forecasting two or three periods ahead.

However, because the aim of this exercise is to assess forecasting performance out-of-sample, for an overall assessment, it is possibly better to examine the performance of the various models over the whole of the available period 90Q3–98Q2. On this basis, the ALI model with $k = 4$ factors has a significantly better forecasting performance than the median VAR for leads $h = 1, \ldots, 4$ at the 0.10, 0.36, 0.28 and 0.56 levels, respectively.

We also compare the performance of the ALI and univariate AR models, again evaluated using the corrected Diebold–Mariano statistic, which is shown in the last column of Table A1. For the full period, these results suggest that the four-factor ALI model is likely to be a better forecasting instrument than the AR model at leads of one to three periods, with the reverse being true for four periods ahead. However, the ALI model with only three factors performs better than the four-factor model for all leads except that of one period ahead. In the first subperiod, the four-factor ALI model dominates those with fewer factors while in the second subperiod, as in the sample as a whole, the three-factor model outperforms the AR model, except for a one-period lead.

Table A2 shows for the two subperiods 1990Q3–1994Q2 and 1994Q3–1998Q2, as well as for the whole period 1990Q3–1998Q2, the percentage of accepted VAR models which have RMSFEs lower than the ALI models. We can see that, in the first subperiod, the four-factor model is beaten by 23% of the VAR models when forecasting one period ahead, by 9% when looking two periods ahead and by only 1.5% when looking three periods ahead, whereas, when forecasting four periods ahead, it is beaten by just over half of the VAR models. With a two-period lead, 91% of the VAR models have RMSFEs smaller than that of the four-factor ALI model; the AR model also performs better for leads greater than one. However, assessing the out-of-sample period as a whole, the four-factor ALI model has RMSFEs lower than those of most of the VAR models and the AR model when leads of up to three periods are considered.

Table A3 details the probability values associated with the use of the corrected Diebold–Mariano statistic S^*_{DM} to test the null hypothesis that the median VAR and various BVAR models have equal forecasting performance together with the ratio of the RMSFE of the median VAR model to each of those of the BVAR models. Generally speaking, the median VAR model performs worse relative to all BVAR models for the first subperiod 1990Q3–1994Q2, whereas this situation is reversed for the second subperiod and the overall forecast period. Consequently, because the forecast performance of the four-factor ALI model was quite good relative to that of the median VAR model for the first subperiod (see Tables A1 and A2), the use of BVAR models does not appear to offer any particular advantage over the four-factor ALI model.

7.2. Germany

The specification tests for the ALI model are displayed in Figure 2. This points to dynamic factor models with either $k = 3$ or 4 factors since the Ljung–Box statistic (upper panels) cannot reject the null hypothesis of no serial correlation for these models whereas this hypothesis is rejected if $k = 1$ or 2. Only a forecasting equation with $k = 2$ factors rejects Granger non-causality in Stage 2, which means that none of the ALI models for Germany are well specified. Consequently, there is no preferred ALI model.

Rather fewer VAR models (Table 2) estimated for Germany satisfy our criteria. Of the 1023 possible models, 673 are rejected using data up to 1990Q2 because they fail to reject Granger non-causality. In the first subperiod, a further 236 are lost, with 168 failing to reject for the Granger non-causality test and 68 the test for parameter constancy. In the second subperiod, a further seven fail the test for parameter constancy.

The corrected Diebold–Mariano statistics (Table A4) again indicate that the $k = 4$ four-factor ALI model is the best performer although the record is not as good as for France. In the first subperiod, the four-factor model is significantly better than the median VAR model for $h = 1, 2$ and 4 steps ahead at around the 0.50 level, which would appear to suggest that there is little to choose between the ALI and VAR modelling approaches for these leads. At $h = 2$ steps ahead, the situation is much worse with the median VAR model significantly outperforming the four-factor ALI model at the 0.09 level. In the second subperiod, the relative forecast performance of the four-factor $k = 4$ ALI model is much improved, being significantly better than the median VAR model at levels varying between 0.10 and 0.28. For leads of one, two and four periods ahead, the four-factor ALI model also does reasonably well against the third quartile VAR model, but for $h = 2$ steps ahead, the third quartile VAR model is significantly better at the 0.36 level.

Examining the full out-of-sample period 1990Q3–1998Q2, the four-factor ALI model is satisfactory except at a lead of two periods ahead. For leads of one, three and four periods, the forecast performance of this ALI model is significantly better than the median and third quartile VAR models at around 0.25 and between the 0.31 and 0.47 levels, respectively. However, for two-step ahead forecasts, the median and third quartile VAR models are significantly different from the four-factor ALI model at the 0.22 and 0.07 levels, respectively.

There is some evidence that the comparison of the ALI models with the AR model also favours the four-factor model. Over the full out-of-sample period, it appears that this model is slightly better as a forecasting tool than the AR model for leads of one and two periods; ALI models with fewer than four factors are likely to be worse than the AR model at all leads. The four-factor ALI model is also the best performer in the first subperiod while in the second the AR dominates all ALI models at nearly all lag lengths.

The relative performance of the AR model is much better in the second subperiod than in the first, simply because GDP growth was much less volatile; the standard deviation of the quarterly growth rate was 2.06% in the second subperiod as compared with 3.63% in the first. Thus, these conclusions reflect the general point that, when GDP growth is stable, an AR model is likely to perform well as compared to models which try to take advantage of other information, whereas, in more volatile periods, richer models are likely to have a better record.

Moving to Table A5, the four-factor model is much the best performer and, relative to the VAR models, does better in the second subperiod than in the first. Thus, in the first subperiod it beats just over half of the VAR models at leads of one and three periods, but is beaten by nearly 80% and 66% at leads of two and four periods. Over the whole sample, it is beaten by only a small proportion of the VAR models (9%–19%) at leads of one, three and four periods, but by 57%

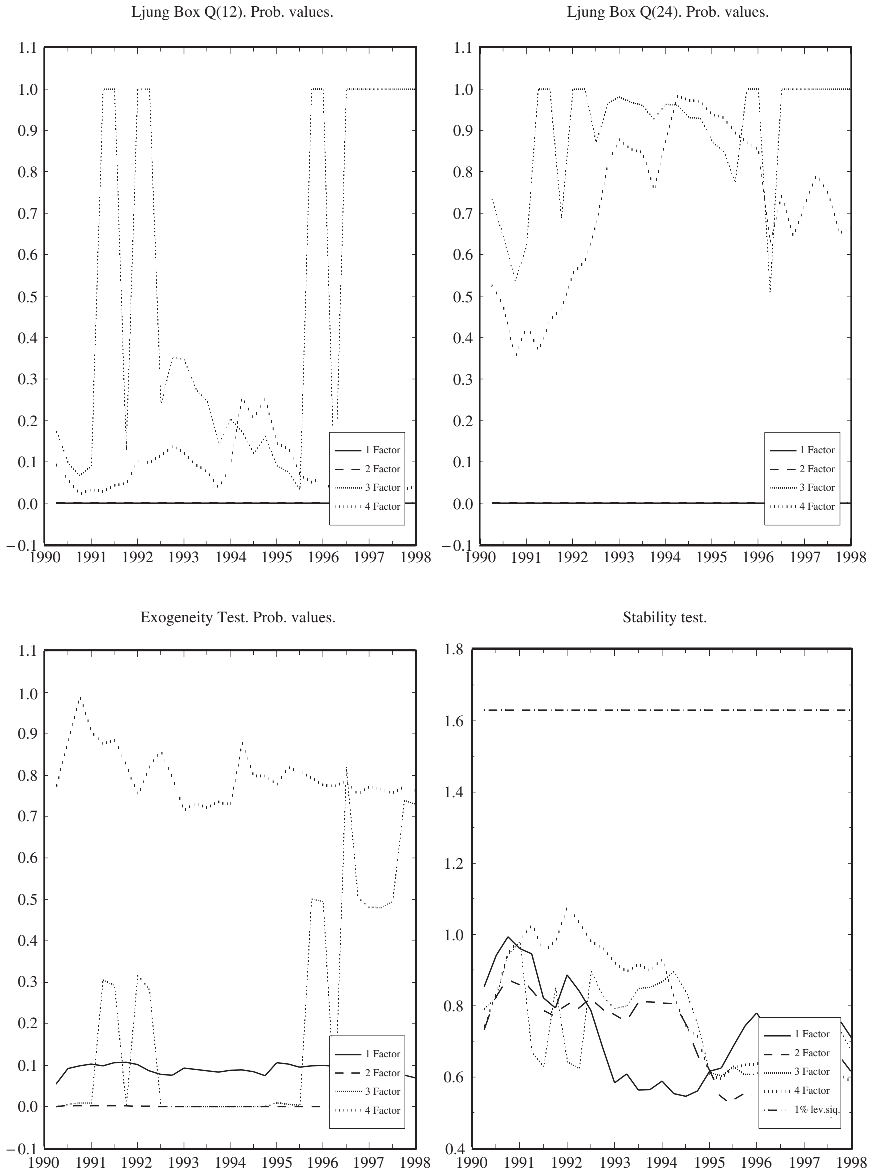

Figure 2. Specification tests for factor models. Germany

at a lead of two periods. Overall, the comparison suggests that the four-factor model performs much better than models with fewer factors which have higher RMSFEs than all or nearly all VAR models in both subperiods and in the full out-of-sample period.

Table A6 details the probability values associated with the use of the corrected Diebold–Mariano statistic S^{*}_{DM} to test the null hypothesis that the median VAR and various BVAR models have equal forecasting performance together with the ratio of the RMSFE of the median VAR model to each of those of the BVAR models. For the first subsample and one- and two-period leads, the median VAR model performs better than all BVAR models whereas for leads of three and four periods this situation is reversed. For the second subperiod 1994Q3–1998Q2, the median VAR model beats all BVAR models at leads of one, two and three periods. In the full forecast period, the median VAR model is much better than all BVAR models at one- and two-period leads and less so for four-step ahead forecasts with BVAR models slightly better for a three-period lead. These conclusions are also generally supported by the RMSFE ratios, whereas this situation is reversed for the second subperiod and the overall forecast period. Therefore, these results in conjunction with those in Tables A4 and A5 suggest that the BVAR models do not appear to offer the likelihood of a better forecast performance than that of the four-factor ALI model over either of the subperiods or the overall forecast period.

7.3. Italy

According to the Ljung–Box statistics reported in Figure 3 (upper panels), only dynamic factor models with $k = 3$ and 4 factors satisfactorily capture the serial correlation in the series; with $k = 1$ and 2 factors, the Ljung–Box statistic rejects the null hypothesis of no serial correlation. The preferred forecasting model is quite clearly a model with $k = 3$ factors in view of the results of the Granger non-causality test for Stage 2; see the bottom left graph in Figure 3. In view of these results, a three-factor ALI model would have been chosen for forecasting out-of-sample.

The VAR models are much more affected by parameter non-constancy (Table 2) than are those for France and Germany. Out of 1023 possible models, 457 are rejected when estimated on the data up to 1990Q2, with 384 displaying parameter non-constancy and 152 Granger non-causality. By the end of the first subperiod, a further 65 and 35 models, respectively, have been rejected on these grounds with none being rejected on both grounds. During the second subperiod, 110 more models show parameter non-constancy and 16 Granger non-causality, again with none being rejected on both grounds.

The corrected Diebold–Mariano statistics (Table A7) suggest that the two- and three-factor ALI models are the most satisfactory. In the first subperiod, the two-factor model is significantly better than the median VAR for leads of $h = 1, \ldots, 4$ periods at the 0.16, 0.17, 0.25 and 0.29 levels, respectively; for the three-factor ALI model, these levels change to 0.17, 0.47, 0.45 and 0.21. In the second subperiod, the performance of the two-factor model deteriorates substantially against the median VAR model with levels of 0.48, 0.47, 0.67 and 0.32, respectively. However, the forecast perfomance of the three-factor ALI model is significantly better than that of the median VAR model for one to four leads at levels varying between 0.10 and 0.26. For the full out-of-sample period, the three-factor ALI model dominates the two-factor model for leads of one, three and four periods with the situation reversed for two-step ahead forecasts.

This relative performance of the two- and three-factor ALI models is also reflected in the comparison between them and the AR model. Over the whole sample, for leads of one and two periods ahead, the corrected Diebold–Mariano statistics suggest that the three-factor model is significantly better than the AR model. In the first subperiod, the two-factor model is plainly the best while in the second subperiod the four-factor and three-factor ALI models are best for one and two leads. The corrected statistic cannot be calculated for leads of three and four periods, but a comparison of RMSFEs favours the three-factor ALI model at these leads both over the whole period and in the two subperiods.

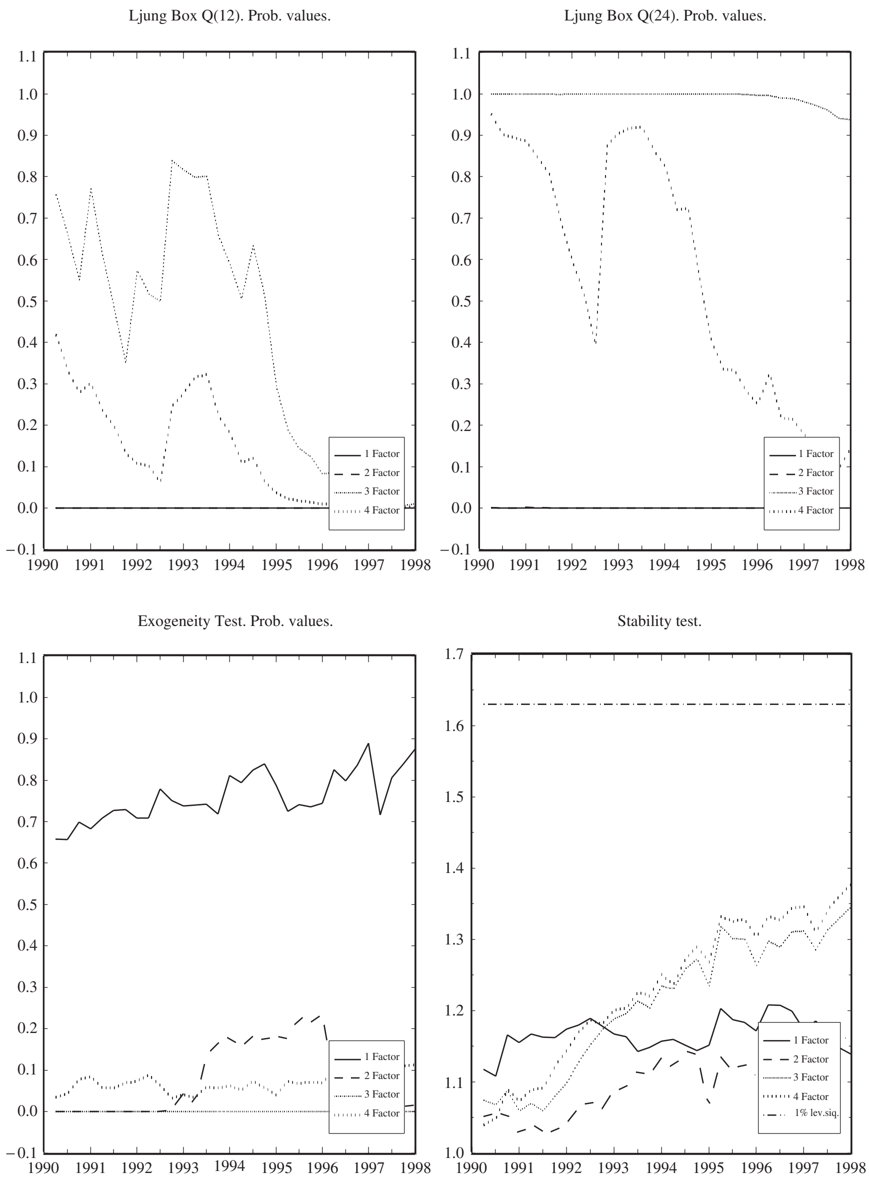

Figure 3. Specification tests for factor models. Italy

The above discussion suggests that there is little to choose between two- and three-factor ALI models, although the three-factor model appears to be better than the median VAR model for one-, two- and four-steps ahead. The three-factor model also gives a generally stronger performance when compared with the VAR models. Over the period as a whole, only 1.7% of VAR models do better with a one-period lead and none do better with a three-period lead, whereas 38% and 21% do better at leads of two and four periods. At one-, three- and four-steps ahead, these figures dominate those for the two-factor model. In the second subperiod, the three-factor ALI model beats more of the VAR models than the two-factor model for all lead times, while in the first subperiod the two-factor model does better at leads of one and two periods.

Table A9 details the probability values associated with the use of the corrected Diebold–Mariano statistic S^*_{DM} to test the null hypothesis that the median VAR and various BVAR models have equal forecasting performance together with the ratio of the RMSFE of the median VAR model to each of those of the BVAR models. For both the subsamples and the full forecast period and at one- and two-period leads, the median VAR model performs worse than all BVAR models, results which are supported at nearly all leads by the RMSFE ratios. However, because the three-factor ALI model outperforms the median VAR model in most cases as discussed above, Tables A7, A8 and A9 suggest that the forecast performance of the three-factor ALI model compares quite favourably with that of all BVAR models for all forecast periods and at all leads.

7.4. United Kingdom

The specification tests displayed in Figure 4 suggest that a dynamic factor model with $k = 4$ factors is the best specification. As above, only models with three and four factors account for the degree of serial correlation in the series. From 1997Q1 onwards, the Ljung–Box results (upper panels) show that even models with four factors do not account completely for the degree of serial correlation in the data, pointing perhaps to the need for a further factor. Parameter non-constancy, however, does not appear to present a problem in Stage 2. Only forecasting models with two and four factors display significant Granger causality over the full period, which suggests a preferred ALI model with four factors.

The main cause for the rejection of VAR models estimated over the period up to 1990Q2 is, as with France and Germany but not Italy, apparent Granger non-causality (see Table 2). Here, 760 out of 1470 models are rejected for this reason while 33 fail the test for parameter constancy; 10 VAR models fail both tests. Only 12 further models fail over the period 1990Q3–1994Q2, which might seem somewhat surprising because this period covers a period of both recession and recovery in the United Kingdom. Perhaps a reasonable explanation is that the pattern of output movements in this period, relative to possible indicators, did not change a great deal as compared with what had gone earlier. In the second subperiod, a further 105 models fail because of parameter non-constancy while only one further model displayed Granger non-causality.

Whether one looks at the first subperiod or the out-of-sample period as a whole, the corrected Diebold–Mariano statistics (Table A10) suggest that the four-factor ALI model is the best performer. In the first subperiod, the forecasting performance of the four-factor ALI model is significantly better than the median VAR model for $h = 1, 2$ and 3 leads at the 0.12, 0.11 and 0.39 levels, respectively, whereas for four leads the median VAR model is significantly better at the 0.41 level. For the period as a whole, these levels alter to 0.25, 0.19, 0.28 and 0.33; note that in this last case the two-factor model has a substantially better performance against the median VAR model. In the second subperiod, however, these levels change substantially indicating that the median VAR model is a better forecasting tool except at a lead of three periods. For this subperiod, the two-factor ALI model performs better than the four-factor model.

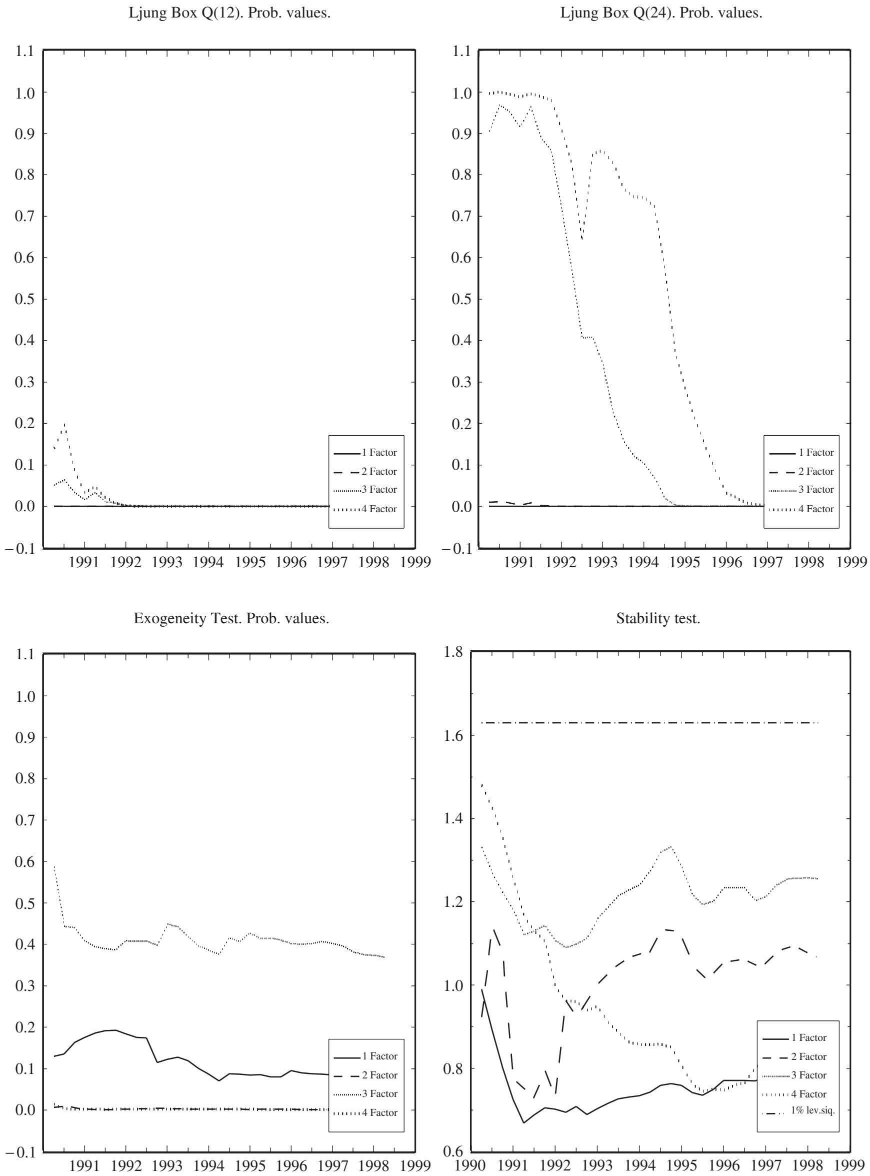

Figure 4. Specification tests for factor models. United Kingdom

The corrected Diebold–Mariano statistics for the forecast period as a whole suggest that the four-factor ALI model is the most strongly preferred relative to the AR model for a two-period lead, with the one-factor model for a lead of one period. The statistic cannot be calculated for three- and four-period leads but a comparison of the RMSFEs (Table A11) favours the four-factor model over the sample as a whole for a lead of three periods and the two-factor model for a four-period lead. In the second subsample, the AR model is generally preferred to the ALI models;

Table 3. Summary of the performance of the factor model.[a]

	France				Germany				Italy				United Kingdom			
	ex-ante $f = 4$				ex-ante $f = -$				ex-ante $f = 3$				ex-ante $f = 4$			
	Med VAR		AR(p)		Med VAR		AR(p)		Med VAR		AR(p)		Med VAR		AR(p)	
lead	best	DM	best	DM	best	DM	best	DM	best	DM	best	DM	best	DM	best	DM
1	4	0.10	4	0.04	4	0.25	4	0.49	3	0.08	3	0.07	4	0.25	1	0.23
2	4	0.36	4	0.22	4	0.77	4	0.39	2	0.23	2	0.16	4	0.19	4	0.10
3	4	0.28	3	0.27	4	0.24	4	0.62	3	0.16	3	*	4	0.28	4	*
4	2	0.42	2	0.01	4	0.26	4	0.69	3	0.34	3	*	3	0.41	3	*

[a] *Ex ante* indicates the number of factors preferred in the factor model on the basis of the specification tests. Under *Med VAR* the table reports the probability that the median VAR is a better forecasting tool than the factor model (under *DM*) together with the number of factors which gives the lowest value of this probability (under *best*). Similar values are reported with respect to the AR(p) process. The probabilities are calculated from the Diebold–Mariano statistics for the period 1990Q3–1998Q2. * Indicates that the Diebold–Mariano statistic cannot be calculated. For these cases the table shows the number of factors which produce forecasts with the lowest root mean square error.

only at a two-period lead and for the two-factor model does the significance level fall below 0.50. For leads of three and four periods, RMSFEs must be resorted to for comparison which suggest that the two- and three-factor models are preferable at leads of three and four periods, respectively; in both cases the models have RMSFEs less than those of the AR model.

The evidence suggests that the performance of the ALI models relative to the AR alternative is much better in the first subperiod than in the second. As with Germany, the explanation for this appears to be that output was more volatile in the first subperiod when the standard deviation of GDP growth was 2.79% per quarter between 1990Q3 and 1994Q2, but only 1.27% per quarter between 1994Q3 and 1998Q2.

Table A12 details the probability values associated with the use of the corrected Diebold–Mariano statistic S^*_{DM} to test the null hypothesis that the median VAR and various BVAR models have equal forecasting performance together with the ratios of the RMSFE of the median VAR model to each of those of the BVAR models. Both the corrected Diebold–Mariano statistics and the RMSFE ratios display a clear preference for the median VAR model over all BVAR models except for the first subperiod for one-step ahead forecasts. A comparison of Tables A10, A11 and A12 indicates little reason for the choice of a BVAR model over the four-factor ALI model on the basis of forecast performance. The median VARs have lower RMSFE than the BVARS (Table A12). In turn, for this whole out of sample period the Diebold–Mariano Fest suggests that at leads of one to three periods, the 4-factor model is preferred to the median VAR (Table A10) and Table 11 implies that the RMFSE of the 4-factor model is at these leads, lower than that of the median VAR.

8. SUMMARY AND CONCLUSIONS

Table 3 summarizes the forecasting results for the ALI model. Firstly, the number of factors is shown for the *ex ante* preferred model of each country selected on the basis of the specification tests applied to the initial estimation sample period (up to 1990Q2). Secondly, it indicates for each lead length, the probability value associated with the corrected Diebold–Mariano test for the hypothesis that the forecasting performance of the *best* ALI model is better than that of the median VAR model and the number of factors in the best ALI model, where the best model is defined as that ALI model which minimizes this probability value and the median VAR model is that VAR model at the mid-point in the ranking of the probability values corresponding to the number of factors of the best ALI model. Thirdly, the table shows the probability value associated with the corrected Diebold–Mariano test for the hypothesis that the forecasting performance of the *best* ALI model is better than that of the AR model and the number of factors in the best ALI model, where the best ALI model is defined as that ALI model which minimizes this probability value. The performance of the median VARs are not affected greatly if they are estimated using Bayesian priors.

There is no preferred ALI model for Germany and the performance of the best ALI model against the AR model is not very satisfactory with the probability values being just under 0.50 for leads of one and two periods, and over 0.50 for longer leads. Otherwise, however, the performance of the best ALI models is satisfactory. The number of the factors identified by the specification tests are generally also those associated with the best performers when assessed on the basis of out-of-sample performance. The probability values are also satisfactorily low.

Consequently, we may conclude that the ALI model framework provides a useful means of forecasting output growth. It avoids the arbitrary nature of many forecasting VAR models and the even more arbitrary nature of methods of constructing traditional leading indicators, as not only is it likely to avoid the risks of selecting the wrong set of indicators, but also the risks of using the wrong priors in the construction of BVAR models. ALI models also appear to perform well against comparator models. There is a satisfactory congruence between the number of factors identified *ex ante* on the basis of specification tests and the number of factors in the best-performing models. However, an important topic for future research concerns the length of the sample period and at what point one should start discarding early data points because they may no longer prove to be a useful guide to data movements in the immediate future. Resolution of this issue is likely to lead to further improvements in the performance of this type of forecasting model.

ACKNOWLEDGEMENTS

We would like to thank conference participants at the 1998 EC^2 meeting in Stockholm, Dominique Ladiray, Gian-Luigi Mazzi, the guest editor and two anonymous referees for their helpful comments. Financial support from the Economic and Social Research Council and Eurostat is gratefully acknowledged.

REFERENCES

Bernanke, B. (1990). On the predictive power of interest rates and interest rate spreads. *New England Economic Review (Federal Reserve Bank of Boston)*, 51–68.

Burns, A. F. and W. C. Mitchell (1946). *Measuring Business Cycles*. New York: National Bureau of Economic Research.

Christoffersen, P. and F. Diebold (1998). Co-integration and long-horizon forecasting. *Journal of Business and Economic Statistics 16*, 450–8.

Clements, M. and D. Hendry (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting 12*, 617–37.

Clements, M. P. and D. F. Hendry (1996). Intercept corrections and structural change. *Journal of Applied Econometrics 11*, 475–94.

Davis, E. P. and G. Fagan (1997). Are financial spreads useful indicators of future inflation and output growth in EU countries? *Journal of Applied Econometrics 12*, 701–14.

Diebold, F. (1998). The past, present and future of macroeconomic forecasting. *Journal of Economic Perspectives 12*, 175–92.

Diebold, F. and L. Killian (2000). Unit root tests are useful for selecting forecasting models. *Journal of Business and Economic Statistics 18*, 265–73.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–63.

Doan, T., R. B. Litterman, and C. A. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews 3*, 1–144.

Emerson, R. and D. Hendry (1996). An evaluation of forecasting using leading indicators. *Journal of Forecasting 15*, 271–92.

Engle, R. F. and M. Watson (1981). A one factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association 76*, 774–81.

Estrella, A. and G. A. Hardouvelis (1991). The term structure as a predictor of real economic activity. *Journal of Finance 46*, 555–76.

Geweke, J. F. (1977). The dynamic factor analysis of economic time series models. In Aigner, D. J. and A. S. Goldberger (eds), *Latent Variables in Socioeconomic Models*. Amsterdam: North-Holland.

Geweke, J. F. and K. J. Singleton (1981). Maximum likelihood confirmatory factor analysis of economic time series. *International Economic Review 22*, 37–54.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. C. (1993). *Time Series Models*. Hemel Hempstead: Harvester and Wheatsheaf.

Harvey, A. C. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics 8*, 231–47.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean square errors. *International Journal of Forecasting 13*, 273–81.

Hoffman, D. L. and R. H. Rasche (1996). Assessing forecast performance in a cointegrated system. *Journal of Applied Econometrics 11*, 495–517.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics 4*, 25–38.

Lutkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

McCracken, M. (2000). Robust out-of-sample inference. *Journal of Econometrics 99*, 817–40.

Ploberger, W. and W. Kramer (1992). The CUSUM test with OLS residuals. *Econometrica 60*, 271–85.

Plosser, C. I. and K. Rouwenhorst (1994). International term structures and real economic growth. *Journal of Monetary Economics 33*, 135–55.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Sargent, T. J. and C. Sims (1977). Business cycle modelling without pretending to have too much a priori theory. In Sims, C. (ed.), *New Methods of Business Cycle Research*. Federal Reserve Bank of Minneapolis.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics 6*, 461–4.

Stock, J. H. and M. Watson (1989). New indexes (sic) of coincident and leading indicators. *NBER Macroeconomics annual 4*, 351–94.

Stock, J. H. and M. Watson (1991). A probability model of the coincident economic indicators. In Moore, G. and K. Lahiri (eds), *Leading Economic Indicators*. Cambridge: Cambridge University Press.

Stock, J. H. and M. Watson (1993). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. In Stock, J. H. and M. Watson (eds), *Business Cycles, Indicators and Forecasting*. Chicago: University of Chicago Press.

Sullivan, R., A. Timmermann, and H. White (1999). Data-snooping, technical trading rule performance and the bootstrap. *Journal of Finance 54*, 1647–91.

Sullivan, R., A. Timmermann, and H. White (2000). Dangers of data-driven inference: the case of calendar effects in stock returns. San Diego: University of California.

West, K. and M. McCracken (1998). Regression-based tests of predictive ability. *International Economic Review 39*, 817–40.
White, H. (2000). A reality check for data snooping. *Econometrica 68*, 1097–1126.

APPENDIX 1: IDENTIFICATION OF THE DYNAMIC FACTOR MODEL

The dynamic factor model (1) is not identified without the imposition of *a priori* restrictions on the parameter matrices $\mathbf{B}$, $\mathbf{C}(L)$ and the variance matrices of $\mathbf{u}_t$ and η_t.

In order to detail the identification problem, consider the spectral decomposition of the observable process $\{\mathbf{x}_t\}_{t=1}^{\infty}$. Given the infinite moving average representation for the unobservable factor process $\{\mathbf{s}_t\}_{t=1}^{\infty}$, $\mathbf{s}_t = \widetilde{\mathbf{C}}(L)\eta_t$, where $\widetilde{\mathbf{C}}(z)[= \mathbf{C}(z)^{-1}] \equiv \sum_{j=0}^{\infty} \widetilde{\mathbf{C}}_j z^j$, $\widetilde{\mathbf{C}}_0 \equiv \mathbf{I}_k$, system (1) may be written as a multivariate linear process with added noise:

$$\mathbf{x}_t = \mathbf{B}\widetilde{\mathbf{C}}(L)\eta_t + \mathbf{u}_t.$$

Hence, given that the error processes $\{\mathbf{u}_t\}_{t=1}^{\infty}$ and $\{\eta_t\}_{t=1}^{\infty}$ are uncorrelated, the spectral decomposition of $\{\mathbf{x}_t\}_{t=1}^{\infty}$ is given by

$$\mathbf{F}_x(\omega) = \mathbf{B}\widetilde{\mathbf{C}}(\omega)\mathbf{F}_\eta(\omega)\widetilde{\mathbf{C}}(\omega)'\mathbf{B}' + \mathbf{F}_u(\omega),$$

where, for example, $\mathbf{F}_x\,(\omega) \equiv \sum_{j=-\infty}^{\infty} \Gamma_x(j)\exp(-ij\omega)$ is the spectral density matrix of $\{\mathbf{x}_t\}_{t=1}^{\infty}$ at frequency $\omega \in [0, \pi]$, $\widetilde{\mathbf{C}}(\omega) \equiv \sum_{j=0}^{\infty} \widetilde{\mathbf{C}}_j \exp(-ij\omega)$, $\{\Gamma_x(j)\}_{j=-\infty}^{\infty}$ are the autocovariances of $\{\mathbf{x}_t\}_{t=1}^{\infty}$ and $i = \sqrt{-1}$; see, for example, Priestley (1981).

The identification problem concerns whether the spectral decomposition for $\{\mathbf{x}_t\}_{t=1}^{\infty}$ given above by $\mathbf{F}_x(\omega)$, $\omega \in [0, \pi]$, implies unique values for the matrix $\mathbf{B}$, the matrix polynomial $\mathbf{C}(z)$ and the variance matrices of the error processes $\{\mathbf{u}_t\}_{t=1}^{\infty}$ and $\{\eta_t\}_{t=1}^{\infty}$. For example, in the absence of any *a priori* restrictions on $\mathbf{C}(z)$ and $\mathbf{F}_\eta(\omega)$ then $\mathbf{F}_x(\omega)$, $\omega \in [0, \pi]$, is consistent with both of the pairs $\widetilde{\mathbf{C}}(\omega)$, $\mathbf{F}_\eta(\omega)$ and $\widetilde{\mathbf{C}}^*(\omega) = \widetilde{\mathbf{C}}(\omega)\mathbf{M}(\omega)$, $\mathbf{F}^*_\eta(\omega) = \mathbf{M}(\omega)^{-1}\mathbf{F}_\eta(\omega)\mathbf{M}(\omega)'^{-1}$ for some nonsingular matrix polynomial $\mathbf{M}(\omega) \equiv \sum_{j=0}^{\infty} \mathbf{M}_j \exp(-ij\omega)$.

Definition 1. *The dynamic factor model* (1) *is identified if and only if* $\mathbf{B}^*$, $\widetilde{\mathbf{C}}^*(\omega)$ *and* $\mathbf{F}^*_\eta(\omega)$ *satisfy the same a priori restrictions as* $\mathbf{B}$, $\widetilde{\mathbf{C}}(\omega)$ *and* $\mathbf{F}_\eta(\omega)$, $\omega \in [0, \pi]$, *and the only values of* $\mathbf{B}^*$, $\widetilde{\mathbf{C}}^*(\omega)$ *and* $\mathbf{F}^*_\eta(\omega)$ *which satisfy*

$$\mathbf{B}\widetilde{\mathbf{C}}(\omega)\mathbf{F}_\eta(\omega)\widetilde{\mathbf{C}}(\omega)'\mathbf{B}' + \mathbf{F}_u(\omega) = \mathbf{B}^*\widetilde{\mathbf{C}}^*(\omega)\mathbf{F}^*_\eta(\omega)\widetilde{\mathbf{C}}^*(\omega)'\mathbf{B}^{*\prime} + \mathbf{F}_u(\omega)$$

are such that $\mathbf{B} = \mathbf{B}^*$, $\widetilde{\mathbf{C}}(\omega) = \widetilde{\mathbf{C}}^*(\omega)$ *and* $\mathbf{F}_\eta(\omega) = \mathbf{F}^*_\eta(\omega)$, $\omega \in [0, \pi]$.

Therefore, (1) is identifiable if and only if $\mathbf{B}^*$, $\widetilde{\mathbf{C}}^*(\omega)$ and $\mathbf{F}^*_\eta(\omega)$ satisfy the same *a priori* restrictions as $\mathbf{B}$, $\widetilde{\mathbf{C}}(\omega)$ and $\mathbf{F}_\eta(\omega)$, $\omega \in [0, \pi]$, and the only admissible (k, k) matrix $\mathbf{K}$ and (k, k) matrix polynomial $\mathbf{M}(\omega) \equiv \sum_{j=0}^{\infty} \mathbf{M}_j \exp(-ij\omega)$ such that $\mathbf{B}^* = \mathbf{B}\mathbf{K}^{-1}$, $\widetilde{\mathbf{C}}^*(\omega) = \mathbf{K}\widetilde{\mathbf{C}}(\omega)\mathbf{M}(\omega)$ and $\mathbf{F}^*_\eta(\omega) = \mathbf{M}(\omega)^{-1}\mathbf{F}_\eta(\omega)\mathbf{M}(\omega)'^{-1}$, $\omega \in [0, \pi]$, are $\mathbf{K} = \mathbf{I}_k$ and $\mathbf{M}(\omega) = \mathbf{I}_k$, $\omega \in [0, \pi]$. Proposition 1 presents a set of sufficient conditions for (1) to be identifiable.

Proposition 1. *The dynamic factor model* (1) *is identifiable if:*

(a) *the error processes* $\{\mathbf{u}_t\}_{t=1}^{\infty}$ *and* $\{\eta_t\}_{t=1}^{\infty}$ *are mutually and serially uncorrelated, conditionally homoscedastic with mean zero and positive definite variance matrices;*
(b) *the error process* $\{\mathbf{u}_t\}_{t=1}^{\infty}$ *has a diagonal variance matrix* ($\mathbf{F}_u(\omega)$ *is diagonal,* $\omega \in [0, \pi]$)*;*
(c) *the error process* $\{\eta_t\}_{t=1}^{\infty}$ *has identity variance matrix* ($\mathbf{F}_\eta(\omega) = \mathbf{I}_k$, $\omega \in [0, \pi]$)*;*
(d) $\mathbf{C}(z)$ *is a diagonal matrix;*
(e) *the elements of* $\mathbf{B} = \{b_{ij}\}$ *are such that* $b_{ii} = 1$, $i = 1, \ldots, k$.

Proof. Initially, assume $\mathbf{B}$ is known. Therefore, $\mathbf{K} = \mathbf{I}_k$ and we may confine our attention to $\widetilde{\mathbf{C}}^*(\omega) = \widetilde{\mathbf{C}}(\omega)\mathbf{M}(\omega)$, $\omega \in [0, \pi]$. Given $\mathbf{F}_\eta(\omega) = \mathbf{I}_k$ from (c), $\mathbf{M}(\omega) = \mathbf{I}_k$ if $\widetilde{\mathbf{C}}(\omega)$ is lower triangular and $\tilde{c}_{tt}(\omega) \neq 0$, $t = 1, \ldots, k$, by Proposition 1 of Geweke and Singleton (1981). Now, suppose $\mathbf{B}$ is unknown. If $\widetilde{\mathbf{C}}(\omega)$ is lower triangular, then $\mathbf{K}$ must also be lower triangular. Because $\widetilde{\mathbf{C}}(\omega)$ is further restricted to be diagonal by (d), then so is $\mathbf{K}$. Now, $b^*_{it} = b_{it}k^{tt}$, $i = 1, \ldots, n$, $t = 1, \ldots, k$, where $\mathbf{K}^{-1} = \mathrm{diag}(k^{11}, \ldots, k^{kk})$. The (i, i)th element of $\mathbf{B}$ is normalized to unity by (e), that is, $b_{ii} = 1$, $i = 1, \ldots, k$. Hence, $k^{ii} = k_{ii} = 1$, $i = 1, \ldots, k$, and $\mathbf{K} = \mathbf{I}_k$ as required. □

APPENDIX 2: COUNTRY RESULTS

Table A1. France. Predictive accuracy tests. Distribution of p-values.

Factors		Diebold–Mariano (corrected) Min	25%	50%	75%	Max	AR Model
		Period: 90Q3–94Q2					
	1 lead	0.282	0.634	0.846	0.939	0.994	0.662
1	2 leads	0.324	0.575	0.788	0.845	0.963	0.483
	3 leads	0.218	0.506	0.676	0.850	1.000	0.493
	4 leads	0.372	0.690	0.756	0.900	0.972	0.656
	1 lead	0.028	0.388	0.655	0.873	0.990	0.340
2	2 leads	0.000	0.493	0.711	0.755	1.000	0.317
	3 leads	0.080	0.368	0.582	0.606	0.876	0.305
	4 leads	0.002	0.531	0.557	0.575	0.603	N/A
	1 lead	0.268	0.700	0.846	0.925	0.992	0.686
3	2 leads	0.046	0.613	0.712	0.745	0.980	0.304
	3 leads	0.002	0.593	0.628	0.647	1.000	N/A
	4 leads	0.001	0.554	0.592	0.638	0.946	N/A
	1 lead	0.039	0.109	0.201	0.452	0.948	0.117
4	2 leads	0.000	0.221	0.259	0.323	0.997	0.213
	3 leads	0.032	0.261	0.274	0.327	0.626	0.269
	4 leads	0.041	0.466	0.514	0.733	0.861	0.454
		Period: 94Q3–98Q2					
	1 lead	0.059	0.617	0.704	0.834	0.988	0.354
1	2 leads	0.355	0.497	0.580	0.746	1.000	N/A
	3 leads	0.155	0.636	0.730	0.921	1.000	0.935
	4 leads	0.064	0.424	0.491	0.716	0.991	0.682
	1 lead	0.006	0.135	0.284	0.525	0.904	0.102
2	2 leads	0.491	0.770	0.852	0.885	1.000	0.717
	3 leads	0.000	0.213	0.332	0.515	0.720	0.519
	4 leads	0.000	0.035	0.186	0.299	0.960	N/A
	1 lead	0.006	0.391	0.497	0.685	0.931	0.149
3	2 leads	0.000	0.206	0.341	0.624	0.996	0.293
	3 leads	0.001	0.109	0.201	0.434	0.999	0.299
	4 leads	0.000	0.138	0.470	0.636	0.985	0.259
	1 lead	0.000	0.067	0.150	0.264	0.792	0.085
4	2 leads	0.026	0.584	0.668	0.755	0.921	0.608
	3 leads	0.000	0.231	0.412	0.607	0.752	0.629
	4 leads	0.000	0.356	0.456	0.537	1.000	N/A

Table A1. Continued.

Factors		Diebold–Mariano (corrected) Min	25%	50%	75%	Max	AR Model
				Period: 90Q3–98Q2			
	1 lead	0.263	0.713	0.870	0.964	0.999	0.481
1	2 leads	0.314	0.636	0.789	0.853	0.970	0.478
	3 leads	0.279	0.635	0.670	0.710	0.964	0.613
	4 leads	0.253	0.677	0.756	0.916	1.000	0.852
	1 lead	0.010	0.262	0.504	0.852	0.980	0.123
2	2 leads	0.220	0.726	0.802	0.851	1.000	0.511
	3 leads	0.000	0.389	0.482	0.519	0.792	0.271
	4 leads	0.000	0.379	0.419	0.447	0.806	0.007
	1 lead	0.214	0.656	0.829	0.928	0.987	0.383
3	2 leads	0.027	0.629	0.721	0.771	0.978	0.196
	3 leads	0.002	0.491	0.533	0.578	0.888	0.072
	4 leads	0.086	0.436	0.472	0.544	0.854	0.046
	1 lead	0.008	0.049	0.103	0.332	0.911	0.035
4	2 leads	0.112	0.279	0.362	0.577	1.000	0.228
	3 leads	0.000	0.147	0.284	0.355	0.574	0.344
	4 leads	0.042	0.488	0.563	0.589	0.712	0.544

Table A2. France. RMSFE performance of the factor model.

	% of VAR models[a]				RMSFE relative to AR(p)				RMSFE
	$f = 1$	$f = 2$	$f = 3$	$f = 4$	$f = 1$	$f = 2$	$f = 3$	$f = 4$	AR(p)
Period: 90Q3–94Q2									
1 lead	90.387	66.589	93.904	22.860	1.025	0.960	1.037	0.804	2.56
2 leads	86.166	74.678	81.712	9.496	0.995	0.977	0.988	0.862	2.82
3 leads	83.353	57.562	86.753	1.524	0.998	0.984	0.999	0.920	2.89
4 leads	99.883	52.052	93.083	51.583	1.049	0.983	0.998	0.981	2.88
Period: 94Q3–98Q2									
1 lead	90.201	25.879	48.995	7.663	0.954	0.850	0.896	0.807	2.09
2 leads	76.131	99.623	33.668	91.583	1.019	1.065	0.983	1.036	1.96
3 leads	95.729	26.256	19.975	42.839	1.088	1.001	0.995	1.018	2.01
4 leads	50.000	43.342	30.151	60.176	1.039	1.007	0.994	1.051	2.07
Period: 90Q3–98Q2									
1 lead	94.598	51.005	89.950	17.462	0.997	0.917	0.983	0.805	2.33
2 leads	90.201	93.844	83.417	33.794	0.997	1.001	0.986	0.920	2.41
3 leads	98.116	38.568	70.603	0.879	1.018	0.990	0.997	0.968	2.44
4 leads	99.372	2.136	32.663	72.739	1.041	0.989	0.996	1.008	2.42

[a] % of VAR models performing better than the model with f factors.

Table A3. France. Median VAR models versus BVAR models.

	Diebold–Mariano (corrected)					RMSFE				
	λ					λ				
	0.10	0.30	0.50	0.70	0.90	0.10	0.30	0.50	0.70	0.90
$\theta = 0.2$										
Period: 90Q3–94Q2										
1 lead	0.724	0.729	0.730	0.730	0.730	1.157	1.161	1.162	1.162	1.162
2 leads	0.619	0.624	0.625	0.626	0.626	1.073	1.076	1.077	1.077	1.077
3 leads	0.573	0.573	0.573	0.573	0.573	1.089	1.089	1.089	1.088	1.088
4 leads	0.560	0.557	0.557	0.557	0.556	1.068	1.065	1.064	1.064	1.064
Period: 94Q3–98Q2										
1 lead	0.397	0.401	0.402	0.402	0.402	0.734	0.735	0.735	0.735	0.735
2 leads	0.341	0.343	0.344	0.344	0.344	0.783	0.788	0.789	0.789	0.789
3 leads	0.362	0.361	0.361	0.361	0.361	0.766	0.766	0.766	0.766	0.766
4 leads	0.374	0.372	0.371	0.371	0.371	0.761	0.763	0.763	0.763	0.764
Period: 90Q3–98Q2										
1 lead	0.459	0.466	0.467	0.468	0.468	0.927	0.930	0.931	0.931	0.932
2 leads	0.376	0.374	0.373	0.372	0.372	0.916	0.922	0.923	0.924	0.924
3 leads	0.395	0.401	0.402	0.403	0.403	0.898	0.901	0.901	0.902	0.902
4 leads	0.043	0.048	0.049	0.050	0.050	0.866	0.866	0.865	0.865	0.865
$\theta = 0.8$										
Period: 90Q3–94Q2										
1 lead	0.717	0.726	0.728	0.729	0.729	1.153	1.159	1.161	1.161	1.162
2 leads	0.619	0.624	0.625	0.626	0.626	1.074	1.076	1.077	1.077	1.078
3 leads	0.574	0.574	0.574	0.573	0.573	1.090	1.089	1.089	1.089	1.089
4 leads	0.559	0.557	0.557	0.557	0.556	1.068	1.065	1.064	1.064	1.064
Period: 94Q3–98Q2										
1 lead	0.393	0.399	0.401	0.401	0.402	0.735	0.736	0.736	0.736	0.735
2 leads	0.341	0.343	0.344	0.344	0.344	0.775	0.784	0.786	0.787	0.787
3 leads	0.362	0.362	0.361	0.361	0.361	0.762	0.764	0.765	0.765	0.765
4 leads	0.374	0.372	0.371	0.371	0.371	0.761	0.763	0.763	0.764	0.764
Period: 90Q3–98Q2										
1 lead	0.455	0.464	0.466	0.467	0.467	0.928	0.931	0.932	0.932	0.932
2 leads	0.373	0.372	0.371	0.371	0.371	0.911	0.920	0.923	0.923	0.924
3 leads	0.388	0.398	0.400	0.401	0.402	0.898	0.901	0.901	0.901	0.902
4 leads	0.033	0.044	0.047	0.048	0.049	0.865	0.865	0.865	0.865	0.865

See footnote 11, Section 5 for a definition of λ and θ.

Table A4. Germany. Predictive accuracy tests. Distribution of p-values.

Factors		Diebold–Mariano (corrected)					
		Min	25%	50%	75%	Max	AR Model
				Period: 90Q3–94Q2			
	1 lead	0.249	0.827	0.845	0.859	0.886	0.933
1	2 leads	0.308	0.841	0.897	0.930	1.000	0.955
	3 leads	0.609	0.917	0.935	0.991	1.000	0.983
	4 leads	0.396	0.897	0.906	0.924	1.000	0.960
	1 lead	0.063	0.614	0.652	0.695	0.749	0.513
2	2 leads	0.708	0.961	0.979	0.987	1.000	0.987
	3 leads	0.503	0.671	0.694	0.734	0.781	0.678
	4 leads	0.570	0.827	0.838	0.846	0.866	0.817
	1 lead	0.062	0.690	0.734	0.774	0.831	0.590
3	2 leads	0.592	0.834	0.871	0.887	0.948	0.856
	3 leads	0.591	0.704	0.720	0.749	0.788	0.710
	4 leads	0.643	0.845	0.849	0.851	0.859	0.842
	1 lead	0.044	0.350	0.491	0.674	0.937	0.330
4	2 leads	0.001	0.582	0.912	0.992	1.000	0.410
	3 leads	0.030	0.341	0.520	0.807	0.994	0.103
	4 leads	0.171	0.474	0.537	0.594	0.955	0.424
				Period: 94Q3–98Q1			
	1 lead	0.058	0.184	0.232	0.284	0.409	0.738
1	2 leads	0.264	0.490	0.589	0.705	0.816	0.685
	3 leads	0.000	0.109	0.207	0.260	0.962	0.681
	4 leads	0.163	0.163	0.163	0.163	0.163	N/A
	1 lead	0.795	0.857	0.868	0.884	0.899	0.887
2	2 leads	0.841	0.873	0.876	0.882	0.894	0.883
	3 leads	0.687	0.724	0.730	0.737	0.755	0.742
	4 leads	N/A	N/A	N/A	N/A	N/A	N/A
	1 lead	0.707	0.826	0.842	0.859	0.882	0.863
3	2 leads	0.758	0.792	0.818	0.829	0.838	0.793
	3 leads	0.724	0.776	0.785	0.792	0.803	0.785
	4 leads	0.798	0.798	1.000	1.000	1.000	1.000
	1 lead	0.013	0.041	0.101	0.381	0.827	0.781
4	2 leads	0.008	0.057	0.287	0.638	0.999	0.477
	3 leads	0.101	0.178	0.201	0.237	0.946	0.751
	4 leads	0.022	0.233	0.276	0.334	0.936	0.642

Table A4. Continued.

Factors		Diebold–Mariano (corrected)					
		Min	25%	50%	75%	Max	AR Model
				Period: 90Q3–98Q1			
	1 lead	0.404	0.740	0.770	0.801	0.840	0.955
1	2 leads	0.241	0.836	0.901	0.933	0.995	0.958
	3 leads	0.402	0.821	0.862	0.895	0.959	0.962
	4 leads	0.041	0.850	0.870	0.898	0.971	0.950
	1 lead	0.702	0.869	0.880	0.886	0.914	0.843
2	2 leads	0.920	0.992	0.995	0.997	0.998	0.996
	3 leads	0.703	0.847	0.855	0.870	0.880	0.850
	4 leads	0.655	0.996	0.997	0.998	1.000	0.984
	1 lead	0.654	0.872	0.884	0.891	0.925	0.841
3	2 leads	0.758	0.933	0.956	0.966	0.976	0.949
	3 leads	0.759	0.892	0.899	0.912	0.921	0.897
	4 leads	0.831	0.995	0.997	0.997	0.999	0.992
	1 lead	0.060	0.149	0.245	0.420	0.667	0.485
4	2 leads	0.013	0.233	0.774	0.927	1.000	0.394
	3 leads	0.002	0.146	0.237	0.467	0.853	0.615
	4 leads	0.045	0.211	0.255	0.306	0.897	0.689

Table A5. Germany. RMSFE performance of the factor model.

	% of VAR models[a]				RMSFE relative to AR(p)				RMSFE
	$f = 1$	$f = 2$	$f = 3$	$f = 4$	$f = 1$	$f = 2$	$f = 3$	$f = 4$	AR(p)
Period: 90Q3–94Q2									
1 lead	97.36	92.10	96.49	49.12	1.146	1.009	1.056	0.924	3.69
2 leads	99.12	100.0	100.0	79.82	1.135	1.346	1.289	0.982	3.47
3 leads	100.0	100.0	100.0	43.86	1.167	1.146	1.226	0.983	3.37
4 leads	98.24	100.0	100.0	65.78	1.153	1.339	1.453	0.998	3.50
Period: 94Q3–98Q1									
1 lead	0.000	100.0	100.0	13.08	1.125	1.899	1.728	1.207	2.05
2 leads	72.89	100.0	100.0	28.03	1.078	1.717	1.459	0.996	2.13
3 leads	20.56	100.0	100.0	20.56	1.099	1.702	1.688	1.110	2.14
4 leads	6.542	100.0	100.0	7.477	1.013	1.549	1.636	1.036	2.00
Period: 90Q3–98Q1									
1 lead	99.06	100.0	100.0	9.346	1.141	1.265	1.239	0.995	3.01
2 leads	99.06	100.0	100.0	57.00	1.121	1.456	1.335	0.986	2.85
3 leads	98.13	100.0	100.0	15.88	1.140	1.332	1.363	1.012	2.77
4 leads	98.13	100.0	100.0	18.69	1.137	1.381	1.480	1.013	2.82

[a] % of VAR models performing better than the model with f factors.

Table A6. Germany. Median VAR models versus BVAR models.

	Diebold–Mariano (corrected)					RMSFE				
	λ					λ				
	0.10	0.30	0.50	0.70	0.90	0.10	0.30	0.50	0.70	0.90
					$\theta = 0.2$					
					Period: 90Q3–94Q2					
1 lead	0.087	0.071	0.064	0.060	0.057	0.875	0.865	0.859	0.856	0.854
2 leads	0.232	0.229	0.228	0.227	0.227	0.879	0.878	0.877	0.876	0.876
3 leads	0.503	0.500	0.499	0.499	0.499	1.002	1.000	1.000	0.999	0.999
4 leads	0.532	0.532	0.532	0.532	0.533	1.030	1.030	1.031	1.031	1.031
					Period: 94Q3–98Q1					
1 lead	0.087	0.084	0.083	0.082	0.081	0.944	0.935	0.931	0.929	0.927
2 leads	0.018	0.018	0.018	0.018	0.018	0.750	0.745	0.742	0.740	0.739
3 leads	0.184	0.178	0.176	0.174	0.174	0.913	0.915	0.917	0.917	0.918
4 leads	N/A	N/A	N/A	N/A	N/A	0.744	0.746	0.747	0.747	0.747
					Period: 90Q3–98Q1					
1 lead	0.051	0.045	0.041	0.039	0.038	0.903	0.893	0.887	0.884	0.882
2 leads	0.011	0.011	0.011	0.011	0.011	0.832	0.829	0.827	0.826	0.825
3 leads	0.560	0.548	0.542	0.539	0.537	0.922	0.923	0.923	0.923	0.923
4 leads	0.354	0.352	0.351	0.350	0.350	0.935	0.935	0.935	0.935	0.935
					$\theta = 0.8$					
					Period: 90Q3–94Q2.					
1 lead	0.071	0.057	0.053	0.051	0.050	0.865	0.854	0.850	0.848	0.847
2 leads	0.230	0.227	0.226	0.226	0.226	0.878	0.876	0.875	0.875	0.875
3 leads	0.501	0.499	0.498	0.498	0.498	1.001	0.999	0.999	0.999	0.999
4 leads	0.532	0.533	0.533	0.533	0.533	1.030	1.031	1.031	1.031	1.031
					Period: 94Q3–98Q1					
1 lead	0.085	0.081	0.080	0.079	0.079	0.934	0.927	0.924	0.923	0.923
2 leads	0.018	0.018	0.018	0.018	0.018	0.744	0.739	0.737	0.737	0.736
3 leads	0.179	0.174	0.173	0.172	0.172	0.916	0.918	0.918	0.919	0.919
4 leads	N/A	N/A	N/A	N/A	N/A	0.746	0.747	0.748	0.748	0.748
					Period: 90Q3–98Q1					
1 lead	0.045	0.038	0.036	0.035	0.035	0.891	0.881	0.878	0.877	0.876
2 leads	0.011	0.011	0.011	0.011	0.011	0.828	0.825	0.823	0.823	0.822
3 leads	0.546	0.536	0.533	0.531	0.530	0.923	0.923	0.923	0.923	0.923
4 leads	0.353	0.350	0.349	0.348	0.348	0.935	0.935	0.935	0.935	0.935

See footnote 11, Section 5 for a definition of λ and θ.

Table A7. Italy. Predictive accuracy tests. Distribution of p-values.

		Diebold–Mariano (corrected)					
Factors		Min	25%	50%	75%	Max	AR model
				Period: 90Q3–94Q2			
	1 lead	0.127	0.297	0.628	0.819	0.976	0.651
1	2 leads	0.144	0.601	0.773	0.909	0.986	0.851
	3 leads	0.000	0.666	0.925	0.946	0.992	N/A
	4 leads	0.063	0.404	0.478	0.535	0.825	N/A
	1 lead	0.030	0.100	0.160	0.255	0.834	0.189
2	2 leads	0.000	0.051	0.167	0.355	0.934	0.185
	3 leads	0.000	0.023	0.252	0.658	1.000	N/A
	4 leads	0.000	0.207	0.285	0.740	1.000	N/A
	1 lead	0.025	0.121	0.172	0.285	0.875	0.193
3	2 leads	0.055	0.260	0.466	0.774	1.000	0.351
	3 leads	0.000	0.083	0.445	0.657	0.834	N/A
	4 leads	0.023	0.164	0.207	0.252	1.000	N/A
	1 lead	0.113	0.262	0.520	0.788	0.976	0.484
4	2 leads	0.117	0.406	0.649	0.932	1.000	0.644
	3 leads	0.000	0.142	0.719	0.872	0.999	N/A
	4 leads	0.044	0.282	0.764	0.831	0.905	N/A
				Period: 94Q3–98Q1			
	1 lead	0.376	0.503	0.578	0.801	0.907	0.263
1	2 leads	0.005	0.447	0.499	0.674	0.825	0.353
	3 leads	0.320	0.451	0.563	0.615	0.974	N/A
	4 leads	0.316	0.397	0.463	0.538	0.571	N/A
	1 lead	0.166	0.368	0.483	0.794	0.881	0.241
2	2 leads	0.019	0.393	0.470	0.758	0.994	0.394
	3 leads	0.255	0.394	0.672	0.750	0.835	N/A
	4 leads	0.101	0.290	0.321	0.397	0.994	N/A
	1 lead	0.042	0.084	0.102	0.231	0.854	0.120
3	2 leads	0.001	0.197	0.255	0.347	0.864	0.268
	3 leads	0.054	0.114	0.173	0.238	0.509	N/A
	4 leads	0.000	0.121	0.172	0.341	0.404	N/A
	1 lead	0.058	0.084	0.094	0.260	0.835	0.099
4	2 leads	0.119	0.514	0.564	0.749	1.000	0.499
	3 leads	0.032	0.143	0.201	0.311	0.807	N/A
	4 leads	0.246	0.379	0.694	0.841	0.991	N/A

Table A7. Continued.

Factors		Diebold–Mariano (corrected) Min	25%	50%	75%	Max	AR Model
				Period: 90Q3–98Q1			
	1 lead	0.154	0.388	0.584	0.751	0.943	0.355
1	2 leads	0.078	0.480	0.771	0.912	0.989	0.766
	3 leads	0.111	0.519	0.889	0.920	0.996	N/A
	4 leads	0.104	0.409	0.892	0.924	0.944	N/A
	1 lead	0.060	0.152	0.245	0.432	0.839	0.120
2	2 leads	0.000	0.075	0.230	0.505	0.987	0.164
	3 leads	0.009	0.059	0.633	0.796	0.911	N/A
	4 leads	0.023	0.084	0.504	0.853	0.954	N/A
	1 lead	0.014	0.051	0.079	0.123	0.629	0.066
3	2 leads	0.009	0.110	0.374	0.636	0.984	0.205
	3 leads	0.000	0.012	0.164	0.295	0.469	N/A
	4 leads	0.016	0.084	0.335	0.432	0.625	N/A
	1 lead	0.029	0.104	0.163	0.263	0.917	0.100
4	2 leads	0.067	0.362	0.776	0.951	1.000	0.610
	3 leads	0.000	0.042	0.670	0.726	0.966	N/A
	4 leads	0.026	0.189	0.880	0.921	0.958	N/A

Table A8. Italy. RMSFE performance of the factor model.

	% of VAR models[a]				RMSFE relative to AR(p)				RMSFE
	$f=1$	$f=2$	$f=3$	$f=4$	$f=1$	$f=2$	$f=3$	$f=4$	AR(p)
Period: 90Q3–94Q2									
1 lead	53.6	7.9	12.0	50.8	1.031	0.835	0.854	0.995	2.31
2 leads	81.5	16.5	48.0	67.1	1.086	0.882	0.971	1.025	2.78
3 leads	86.2	37.5	37.7	57.9	1.036	0.830	0.830	0.921	3.31
4 leads	74.6	56.6	36.9	59.4	0.977	0.855	0.835	0.930	3.49
Period: 94Q3–98Q1									
1 lead	75.5	47.3	19.4	12.9	0.962	0.937	0.807	0.790	3.65
2 leads	49.1	43.5	17.9	82.3	0.983	0.977	0.940	1.000	3.22
3 leads	66.1	65.8	1.1	22.3	0.960	0.959	0.829	0.879	3.37
4 leads	40.2	34.7	0.0	61.1	0.956	0.947	0.892	0.997	3.20
Period: 90Q3–98Q1									
1 lead	62.0	19.7	1.7	11.4	0.983	0.908	0.821	0.857	3.03
2 leads	72.9	25.2	37.6	67.0	1.030	0.936	0.956	1.013	2.95
3 leads	77.0	60.8	0.0	61.1	1.004	0.899	0.836	0.908	3.23
4 leads	67.0	50.8	21.4	61.7	0.981	0.902	0.883	0.959	3.28

[a]% of VAR models performing better than the model with f factors.

Table A9. Italy. Median VAR models versus BVAR models.

	Diebold–Mariano (corrected)					RMSFE				
	λ					λ				
	0.10	0.30	0.50	0.70	0.90	0.10	0.30	0.50	0.70	0.90
$\theta = 0.2$										
Period: 90Q3–94Q2										
1 lead	0.730	0.727	0.726	0.726	0.726	1.157	1.155	1.154	1.154	1.154
2 leads	0.880	0.877	0.876	0.875	0.874	1.422	1.416	1.413	1.412	1.410
3 leads	N/A	N/A	N/A	N/A	N/A	1.225	1.221	1.220	1.220	1.220
4 leads	N/A	N/A	N/A	N/A	N/A	1.293	1.286	1.284	1.284	1.283
Period: 94Q3–98Q1										
1 lead	0.959	0.959	0.959	0.959	0.959	0.991	0.983	0.981	0.981	0.980
2 leads	0.959	0.959	0.958	0.958	0.958	1.036	1.035	1.035	1.034	1.034
3 leads	N/A	N/A	N/A	N/A	N/A	0.982	0.981	0.981	0.981	0.981
4 leads	N/A	N/A	N/A	N/A	N/A	1.101	1.102	1.103	1.103	1.103
Period: 90Q3–98Q1										
1 lead	0.968	0.967	0.966	0.966	0.966	1.034	1.027	1.025	1.024	1.024
2 leads	0.852	0.847	0.847	0.846	0.846	1.143	1.140	1.139	1.138	1.138
3 leads	N/A	N/A	N/A	N/A	N/A	1.049	1.049	1.049	1.049	1.049
4 leads	N/A	N/A	N/A	N/A	N/A	1.071	1.070	1.070	1.070	1.070
$\theta = 0.8$										
Period: 90Q3–94Q2										
1 lead	0.729	0.727	0.726	0.726	0.726	1.157	1.155	1.154	1.154	1.154
2 leads	0.871	0.872	0.872	0.872	0.872	1.406	1.407	1.406	1.406	1.406
3 leads	N/A	N/A	N/A	N/A	N/A	1.232	1.224	1.223	1.222	1.222
4 leads	N/A	N/A	N/A	N/A	N/A	1.296	1.287	1.285	1.284	1.284
Period: 94Q3–98Q1										
1 lead	0.960	0.959	0.959	0.959	0.959	0.992	0.983	0.981	0.981	0.980
2 leads	0.957	0.958	0.958	0.958	0.958	1.036	1.035	1.035	1.035	1.035
3 leads	N/A	N/A	N/A	N/A	N/A	0.982	0.981	0.981	0.981	0.981
4 leads	N/A	N/A	N/A	N/A	N/A	1.101	1.102	1.102	1.103	1.103
Period: 90Q3–98Q1										
1 lead	0.966	0.966	0.966	0.966	0.966	1.034	1.027	1.025	1.024	1.024
2 leads	0.855	0.850	0.848	0.848	0.847	1.139	1.138	1.138	1.138	1.137
3 leads	N/A	N/A	N/A	N/A	N/A	1.049	1.049	1.049	1.049	1.049
4 leads	N/A	N/A	N/A	N/A	N/A	1.071	1.070	1.070	1.070	1.070

See footnote 11, Section 5 for a definition of λ and θ.

Table A10. UK. Predictive accuracy tests. Distribution of p-values.

Factors		Diebold–Mariano (corrected)					
		Min	25%	50%	75%	Max	AR model
				Period: 90Q3–94Q2			
	1 lead	0.002	0.188	0.418	0.674	0.948	0.033
1	2 leads	0.002	0.349	0.543	0.724	0.989	0.124
	3 leads	0.000	0.471	0.597	0.724	0.999	N/A
	4 leads	0.060	0.558	0.646	0.720	0.954	N/A
	1 lead	0.002	0.115	0.329	0.637	0.960	0.200
2	2 leads	0.007	0.599	0.740	0.832	1.000	0.216
	3 leads	0.093	0.460	0.701	0.907	1.000	N/A
	4 leads	0.000	0.264	0.487	0.589	0.982	N/A
	1 lead	0.596	0.939	0.975	0.989	0.998	0.981
3	2 leads	0.587	0.855	0.900	0.916	0.956	0.870
	3 leads	0.075	0.691	0.845	0.928	0.999	N/A
	4 leads	0.189	0.665	0.814	0.911	0.947	N/A
	1 lead	0.002	0.034	0.123	0.318	0.964	0.159
4	2 leads	0.007	0.062	0.111	0.219	0.985	0.105
	3 leads	0.004	0.206	0.386	0.538	0.807	N/A
	4 leads	0.000	0.378	0.591	0.748	0.980	N/A
				Period: 94Q3–98Q2			
	1 lead	0.074	0.466	0.618	0.738	0.992	0.918
1	2 leads	0.231	0.514	0.718	0.906	0.994	0.893
	3 leads	0.038	0.448	0.719	0.911	1.000	N/A
	4 leads	0.000	0.520	0.595	0.786	1.000	N/A
	1 lead	0.076	0.548	0.704	0.814	0.987	0.894
2	2 leads	0.000	0.138	0.255	0.552	0.804	0.235
	3 leads	0.000	0.166	0.312	0.568	0.796	N/A
	4 leads	0.005	0.350	0.486	0.589	1.000	N/A
	1 lead	0.380	0.818	0.873	0.909	0.963	0.917
3	2 leads	0.233	0.744	0.780	0.809	0.866	0.754
	3 leads	0.000	0.429	0.570	0.621	0.699	N/A
	4 leads	0.003	0.233	0.357	0.530	1.000	N/A
	1 lead	0.012	0.673	0.821	0.914	0.987	0.927
4	2 leads	0.037	0.445	0.657	0.834	0.985	0.646
	3 leads	0.101	0.260	0.390	0.724	1.000	N/A
	4 leads	0.352	0.552	0.710	0.782	1.000	N/A

Table A10. Continued.

Factors		Diebold–Mariano (corrected) Min	25%	50%	75%	Max	AR model
				Period: 90Q3–98Q2			
	1 lead	0.006	0.208	0.391	0.690	0.994	0.234
1	2 leads	0.069	0.439	0.643	0.866	0.996	0.190
	3 leads	0.003	0.494	0.668	0.757	0.999	N/A
	4 leads	0.076	0.545	0.632	0.728	1.000	N/A
	1 lead	0.007	0.180	0.365	0.712	0.995	0.360
2	2 leads	0.001	0.453	0.660	0.813	0.958	0.133
	3 leads	0.002	0.415	0.636	0.713	0.997	N/A
	4 leads	0.021	0.271	0.409	0.568	0.973	N/A
	1 lead	0.708	0.966	0.987	0.995	0.999	0.995
3	2 leads	0.662	0.930	0.963	0.971	0.990	0.940
	3 leads	0.377	0.723	0.832	0.893	0.978	N/A
	4 leads	0.289	0.699	0.828	0.918	0.991	N/A
	1 lead	0.001	0.113	0.249	0.518	0.950	0.306
4	2 leads	0.009	0.096	0.187	0.398	0.931	0.102
	3 leads	0.000	0.160	0.284	0.400	0.919	N/A
	4 leads	0.227	0.549	0.671	0.767	0.943	N/A

Table A11. UK. RMSFE performance of the factor model.

	% of VAR Models[a]				RMSFE relative to AR(p)				RMSFE
	$f = 1$	$f = 2$	$f = 3$	$f = 4$	$f = 1$	$f = 2$	$f = 3$	$f = 4$	AR(p)
Period: 90Q3–94Q2									
1 lead	42.0	34.9	100.0	5.9	0.866	0.842	1.301	0.733	2.81
2 leads	59.1	81.1	100.0	7.7	0.818	0.910	1.193	0.639	2.37
3 leads	71.1	76.2	95.9	27.9	0.824	0.839	0.939	0.683	2.20
4 leads	87.8	47.8	95.2	53.8	0.875	0.728	0.946	0.771	2.38
Period: 94Q3–98Q2									
1 lead	68.2	83.1	99.4	94.6	1.230	1.283	1.559	1.373	1.41
2 leads	77.6	31.7	98.9	69.2	1.109	0.883	1.380	1.077	1.43
3 leads	69.8	39.4	65.0	55.7	1.010	0.843	0.987	0.950	1.49
4 leads	64.4	66.6	62.1	85.4	0.894	0.905	0.887	1.040	1.47
Period: 90Q3–98Q2									
1 lead	36.2	35.7	100.0	26.8	0.951	0.948	1.357	0.900	2.22
2 leads	70.8	71.9	100.0	15.2	0.910	0.916	1.254	0.768	1.99
3 leads	74.2	67.1	95.5	7.4	0.893	0.867	0.995	0.774	1.89
4 leads	85.7	36.7	94.6	87.5	0.884	0.807	0.949	0.891	1.98

[a] % of VAR models performing better than the model with f factors.

Table A12. UK. Median VAR models versus BVAR models.

	Diebold–Mariano (corrected)					RMSFE				
	λ					λ				
	0.10	0.30	0.50	0.70	0.90	0.10	0.30	0.50	0.70	0.90
$\theta = 0.2$										
Period: 90Q3–94Q2										
1 lead	0.521	0.555	0.563	0.567	0.570	1.010	1.026	1.030	1.032	1.033
2 leads	0.260	0.260	0.260	0.260	0.260	0.730	0.732	0.732	0.732	0.731
3 leads	N/A	N/A	N/A	N/A	N/A	0.625	0.625	0.625	0.624	0.624
4 leads	N/A	N/A	N/A	N/A	N/A	0.636	0.635	0.635	0.635	0.635
Period: 94Q3–98Q2										
1 lead	0.003	0.004	0.004	0.004	0.004	0.680	0.697	0.701	0.704	0.705
2 leads	0.121	0.120	0.120	0.120	0.120	0.469	0.475	0.477	0.478	0.478
3 leads	N/A	N/A	N/A	N/A	N/A	0.431	0.430	0.429	0.428	0.428
4 leads	N/A	N/A	N/A	N/A	N/A	0.410	0.407	0.405	0.404	0.404
Period: 90Q3–98Q2										
1 lead	0.225	0.242	0.246	0.247	0.248	0.883	0.901	0.906	0.908	0.910
2 leads	0.067	0.068	0.068	0.068	0.068	0.621	0.624	0.624	0.623	0.623
3 leads	N/A	N/A	N/A	N/A	N/A	0.541	0.544	0.544	0.545	0.545
4 leads	N/A	N/A	N/A	N/A	N/A	0.560	0.562	0.562	0.562	0.562
$\theta = 0.8$										
Period: 90Q3–94Q2										
1 lead	0.533	0.562	0.568	0.571	0.573	1.015	1.029	1.032	1.033	1.034
2 leads	0.258	0.259	0.259	0.259	0.259	0.727	0.729	0.730	0.730	0.730
3 leads	N/A	N/A	N/A	N/A	N/A	0.622	0.623	0.623	0.623	0.623
4 leads	N/A	N/A	N/A	N/A	N/A	0.635	0.635	0.635	0.635	0.635
Period: 94Q3–98Q2										
1 lead	0.004	0.004	0.004	0.004	0.004	0.680	0.698	0.702	0.704	0.705
2 leads	0.120	0.120	0.120	0.120	0.119	0.471	0.477	0.478	0.479	0.479
3 leads	N/A	N/A	N/A	N/A	N/A	0.426	0.426	0.426	0.426	0.426
4 leads	N/A	N/A	N/A	N/A	N/A	0.405	0.403	0.403	0.403	0.402
Period: 90Q3–98Q2										
1 lead	0.227	0.243	0.246	0.248	0.249	0.889	0.905	0.909	0.911	0.912
2 leads	0.066	0.067	0.068	0.068	0.068	0.616	0.620	0.621	0.621	0.622
3 leads	N/A	N/A	N/A	N/A	N/A	0.544	0.545	0.545	0.545	0.545
4 leads	N/A	N/A	N/A	N/A	N/A	0.562	0.562	0.562	0.562	0.562

See footnote 11, Section 5 for a definition of λ and θ.