Forecasting Business Cycles in a Small Open Economy: A Dynamic Factor Model for Singapore

HWEE KWAN CHOW[1*] AND KEEN MENG CHOY[2]

[1] *School of Economics and Social Sciences, Singapore Management University, Singapore*

[2] *Division of Economics, Nanyang Technological University, Singapore*

ABSTRACT

We apply multivariate statistical methods to a large dataset of Singapore's macroeconomic variables and global economic indicators with the objective of forecasting business cycles in a small open economy. The empirical results suggest that three common factors are present in the time series at the quarterly frequency, which can be interpreted as world, regional and domestic economic cycles. This leads us to estimate a factor-augmented vector autoregressive (FAVAR) model for the purpose of optimally forecasting the key macroeconomic and sectoral aggregates of Singapore. By taking explicit account of the common factor dynamics, we find that iterative forecasts generated by this model are significantly more accurate than direct multi-step predictions based on the identified factors as well as forecasts from univariate and vector autoregressions.

KEY WORDS business cycles; principal components; dynamic factor model; factor-augmented VAR; forecasting; Singapore

INTRODUCTION

The study of business cycles in small and open economies has always presented the empirical researcher with particular challenges. A fundamental reason for this lies in the

* Correspondence to: H.K. Chow, School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903. E-mail: hkchow@smu.edu.sg

vulnerability of such economies to the vagaries of international macroeconomic fluctuations, which accentuate the intrinsic volatility caused by domestically generated disturbances. Nonetheless, many papers appearing in academic journals are cognizant of the role played by international fluctuations when documenting the ‘stylized facts’ of business cycle co-movements in relatively open economies, for example Sweden, Switzerland, New Zealand and Korea (see respectively Englund *et al*., 1992, Danthine and Girardin, 1989, Kim *et al*., 1994, and Kim and Choi, 1997). Two recent articles that examined the nature of economic fluctuations in the small city-states of Hong Kong and Singapore also find that external factors contribute significantly to these economies’ internal gyrations (Leung and Suen, 2001; Choy, 2006). In addition to such industrialized country studies, Kose (2002) calibrated the impact of world price shocks on a group of small open developing economies while Kose, Otrok and Whiteman (2003) successfully used worldwide, regional and country-specific business cycles to explain the aggregate co-movements observed in a broad cross-section of countries.

As for published work on forecasting the economic cycles of highly open economies, attempts to include the impact of international events are often hampered by the need for parsimony. Typically, the forecasting problem is approached on an ad hoc basis, using only a limited number of foreign variables to capture external shocks to the economy. This is to avoid running into the degrees-of-freedom problem associated with a loss of efficiency in regression-type models such as single and multiple equations, large-scale macroeconometric models, and even statistical time series methods.

Unfortunately, an inadequate account of the influence of global causative factors on the genesis and propagation of local business cycles could well lead to sub-optimal predictions of economic variables. As a remedy, one could consider dynamic factor models that permit the incorporation of a large number of variables capturing the foreign disturbances which buffet small and open economies as well as impulses originating from domestic sources. This class of models is appealing from a theoretical standpoint since it views all macroeconomic fluctuations as being driven by a small number of common shocks and an idiosyncratic component that is peculiar to each economic time series—an idea that was already implicit in Burns and Mitchell’s (1946) early characterization of business cycles. In spite of the seminal paper by Sargent and Sims (1977), however, dynamic factor models have only lately been revived for the purpose of forecasting real economic activity in the US and larger European economies, partly because the statistical techniques and computing power needed to efficiently exploit the vast amount of information in large datasets were developed but recently. Whilst their application to small open economies remains unexplored, the results are promising so far and they suggest that these data-intensive models could outperform the standard approaches (see, *inter alia*, García-Ferrer and Poncela, 2002; Stock and Watson, 2002b; Forni *et al*., 2003; Artis *et al*., 2005; Schumacher, 2007).

In this paper, we illustrate the use of a dynamic factor model for macroeconomic forecasting of an archetypal small open economy with an empirical application to Singapore. The exercise begins with a multivariate analysis of a large collection of quarterly time series that includes foreign economic indicators such as the real GDPs and asset prices of Singapore's trading partners, global electronics series and world prices and interest rates. The domestic variables consist of GDP and its components, gross value-added by sectors, industrial production indices, sectoral indicators, trade series, general price indexes, labour market variables, monetary and financial series and business expectations surveys. Importantly, some of these series are known to be leading indicators of economic activity in Singapore.

It turns out that the bulk of the observed co-variation in the dataset can be explained by three uncorrelated factors representing world, regional and domestic business cycles. Once these factors have been estimated by a highly general factor model, they can be utilized for short-term forecasting. A novelty of the paper lies in the way forecasts are generated from the dynamic factors. In existing studies, predictions are routinely produced by a multi-step approach that entails the estimation of distinct forecasting equations at each horizon. To evaluate the empirical performance of these direct forecasts, we also generate predictions iteratively by specifying a factor-augmented vector autoregression (FAVAR) which includes the identified factors and the variables to be forecasted. Such a model was proposed by Bernanke, Boivin and Eliasz (2005) for the structural analysis of monetary policy shocks but as far as we know, has not been employed in forecasting real variables. This is surprising as projections from the FAVAR model can draw on information extracted from a much larger cross-section of macroeconomic time series than is feasible with small-scale VAR models.

Our objective is to compare the two types of factor forecasts and determine which represent more accurate predictions in practice—another empirical issue that has not received much attention in the literature. The multi-step approach has the advantage of mitigating specification error in the one-step ahead forecasting model while the iterated procedure offers potential gains in efficiency since fewer parameters are estimated. Marcellino *et al*. (2006) concluded from an analysis of autoregressive models that the latter dominates the former insofar as US economic time series are concerned. We ourselves find in the factor context that the FAVAR approach, by taking explicit account of the future evolution of dynamic factors, leads to substantial improvements in forecast accuracy when compared to the direct multi-step method based on the estimated factors.

The following section describes the data series used in the forecasting exercise and carries out a preliminary multivariate analysis by the method of principal components. Plausible economic interpretations of the largest estimated components are suggested by relating them to key variables. We then proceed to estimate a generalized dynamic factor model for the Singapore economy using the method of maximum likelihood and test for the number of common factors in the model with the help of information criteria. Next, we use the dynamic factors to generate multi-step as well as iterated out-of-sample predictions of the growth in real GDP and major sectoral output indicators, which are formally evaluated against univariate and multivariate time series models. Lastly, we present the paper's conclusions.

A PRINCIPAL COMPONENT ANALYSIS OF SINGAPORE DATA

Singapore has a large and reliable database of macroeconomic variables by the standards of newly industrialized economies. The national income accounts, in particular, are very rich in revealing sectoral details of the compilation of real GDP by the output approach. In view of this, we broaden our search for the set of indicators to be used in the multivariate analyses of this and subsequent sections to time series of the quarterly frequency, hence providing a more comprehensive coverage of the many facets of macroeconomic activity. Needless to say, monthly data is not excluded from the exercise, although these have to be aggregated or averaged to yield quarterly values.

If pertinent, we employ the seasonally adjusted time series supplied by data sources save for a few cases where we performed the adjustment ourselves using the X-12 software (these instances are noted in the appendix, which lists all the variables covered). Since our interest is in forecasting Singapore's *growth cycles*, almost all the variables we work with are transformed into approximate year-on-year growth rates by taking the fourth differences of their logarithms, thereby ensuring also that the data is covariance stationary (the exceptions are indicated in the appendix). In this respect, we depart from the conventional practice of modelling quarter-on-quarter growth rates since these are very volatile for a small open economy like Singapore.

To avoid overweighting any one series, all raw and transformed variables are normalized by subtracting their means and dividing by their standard deviations. A visual inspection of time plots revealed a handful of unusual occurrences during the sample period from 1993Q1 to 2006Q4 due to the Asian financial crisis of 1997 and the outbreak of the SARS disease in early 2003. As a robustness measure, the outlying observations are excluded in the computation of means and standard deviations. The choice of starting date was dictated by the availability of electronics data, compelling us to increase N (number of predictor variables) at the expense of T (length of time series).

As stated above, the importance of international and foreign economic indicators for short-term monitoring of the Singapore economy should not be underestimated, so it is imperative to consider external series that are known to co-move with—and sometimes lead—domestic variables, subject of course to data availability. After adding in the local macroeconomic series, we have a total of 177 quarterly indicators on hand (41 foreign and 136 local). The transnational and national indicators selected can be grouped as follows:

- Real GDPs of Singapore's major trading partners and their weighted average (10 countries and one region); composite leading indexes of the US and major European and Asian economies; foreign stock prices and interest rates

- Global semiconductor sales, US technology cycle index and electronics leading series; world oil price, non-fuel commodity prices and global consumer prices

- Singapore's real GDP and expenditure components; gross value-added output in manufacturing and major service sectors; industrial production indices; investment commitments and business expectations surveys; official composite leading index

- Construction and housing related series e.g. residential investment, building contracts awarded and property prices

- Sectoral indicators such as retail sales, new car registrations, tourist arrivals, air and sea cargo handled, electricity generation and new company formations in different sectors

- Foreign trade series: exports and imports of goods and services, domestic exports and re-exports—all disaggregated into oil and non-oil categories

- Export and import price indices; terms of trade; consumer and producer price indices; GDP and sectoral deflators

- Labour market variables: changes in employment, retrenchments, overall and resident unemployment rates, unit labour and business costs

- Financial series such as share prices, interest rates and exchange rates; monetary aggregates and bank credit

Even though empirical dynamic factor models for macroeconomic time series are of recent vintage, there has long existed techniques for data simplification in the statistical literature. All these have in common the aim of reducing high-dimensional data into a smaller and more manageable set of linear combinations, ranging from the classical methods of principal component and static factor analyses to more sophisticated techniques like cluster analysis and projection pursuit. Given its close affinity with the dynamic factor model, we decided to carry out a principal component analysis of the Singapore dataset so as to gain preliminary insights into the interrelationships between the selected economic indicators (the results are later used as pre-estimates to initialize the estimation procedure for the common factors).

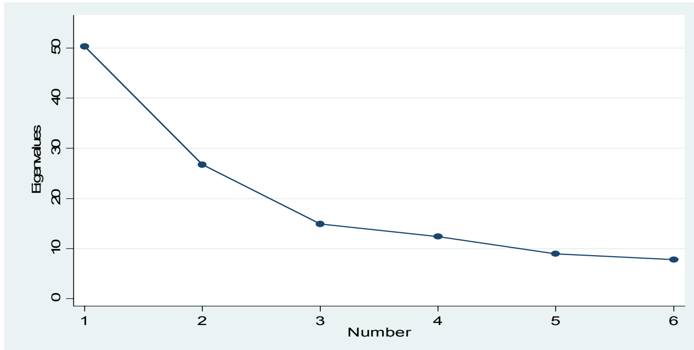

Figure 1. Scree plot of eigenvalues from principal component analysis

Figure 1 shows the outcome of the principal component analysis by way of a *scree plot*, which graphs the largest six eigenvalues of the data matrix. There appears to be a natural break at the third value, with the remaining eigenvalues flattening out. The first three principal components explain on average 28%, 15% and 8% of the total variance in our economic series, making for a cumulative proportion of 52%. This is remarkable in view of the large number and diversity of the time series included in the analysis. By contrast, the fourth, fifth and sixth components account for only 7%, 5% and 4% respectively and they are also less amenable to economic interpretation.

To see what sort of interpretations, if any, could be assigned to the first three principal components, we execute an orthogonal rotation of the estimated eigenvectors using the popular varimax method. The rotated components are graphed as bar charts in Figures 2–4, where the numbers on the horizontal axis refer to the ordering of the series (see appendix listing) and the principal component 'loadings' are shown on the vertical axis. Table I provides a summary by indicating whether the variables in a particular grouping have predominantly positive, negative or zero loadings on a given component. Scanning down the table, we might aptly label, in the typical manner of principal component analysis, the first component as 'regional services', the second as 'domestic construction', and the third as 'global manufacturing'.

The first rotated principal component is a linear combination that places heavy weights on regional economic series and world electronics indicators. In terms of sectoral breakdown, domestic services and the semiconductor-related industries are strongly emphasized. This is very much in line with the regional orientation of Singapore's exportable services and her role as a producer of high value-added parts and accessories in the electronics supply chain based in Asia. In contrast, the second principal component clearly picks out the indicators associated with the domestic construction cycle, property prices and supporting services such as real estate and bank lending. Labour market variables and effective exchange rates also load highly on this component. The third rotated eigenvector seems to be linked to global business cycles as it has large and positive coefficients for US and European variables, producer prices and financial series. Naturally, local manufacturing output is more strongly aligned to these cycles than services production. In the light of these findings, we interpret the driving forces behind short-term fluctuations in Singapore's macroeconomic variables as world, regional and country-specific business cycles. The sections that follow describe how these factors could be estimated and utilized for forecasting business cycles.

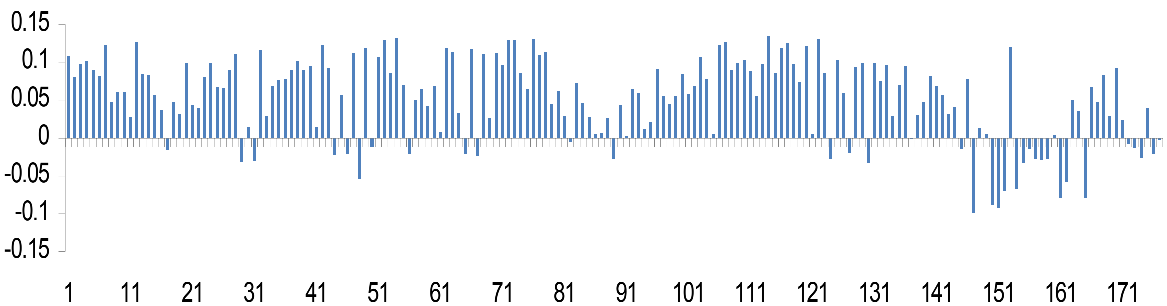

Figure 2. Loadings on first principal component

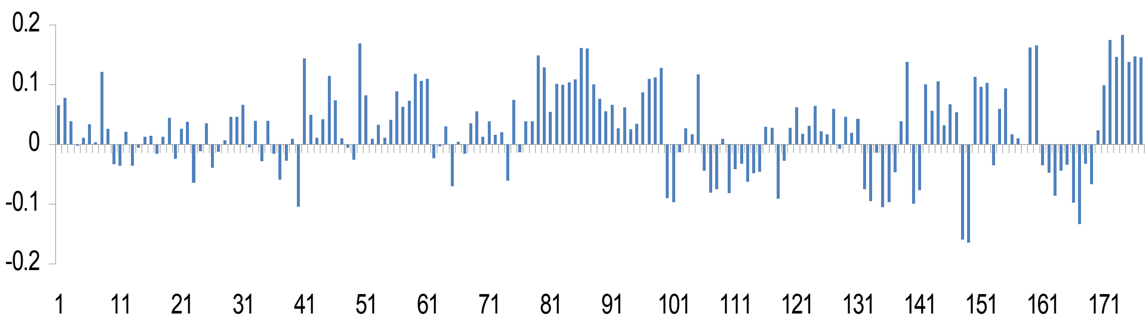

Figure 3. Loadings on second principal component

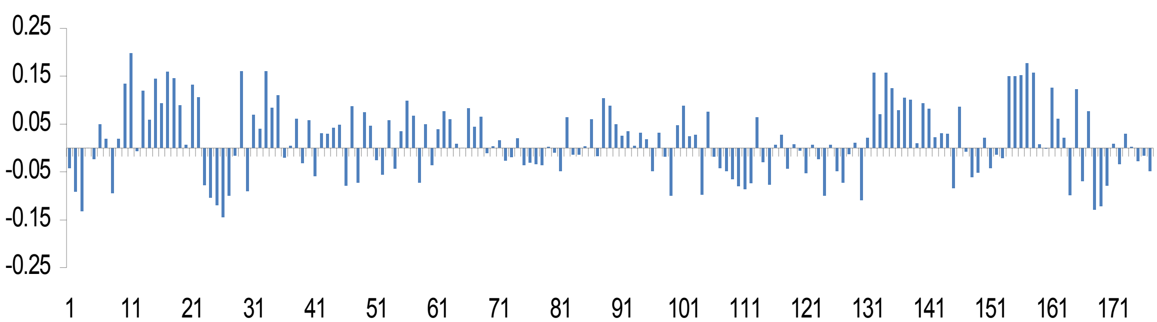

Figure 4. Loadings on third principal component

Table I. Loadings on leading principal components

Variable Groupings	Component 1	Component 2	Component 3
Foreign GDPs and stock prices	+	0	+
Asia	+	0	–
US & Europe	0	0	+
World electronics	+	0	+
Singapore GDP components	+	+	+
Manufacturing	+	0	+
Services	+	+	0
Construction	0	+	+
Sectoral indicators	+	–	–
Trade	+	0	–
Prices	+	–	+
Labour market	–	+	0
Financial series	–	–	+
Monetary aggregates	0	+	0

Note: a plus sign indicates positive and large loadings and a minus represents negative loadings.

THE GENERALIZED DYNAMIC FACTOR MODEL

Representation

Variables cast in a factor model representation are characterized by the sum of two mutually orthogonal unobservable components: the common component driven by a small number of factors and the idiosyncratic component driven by variable-specific shocks. Let $X_t, t = 1, \ldots, T,$ be the N-dimensional vector of stationary time series. Since we have a cross-sectional panel of 177 predictor variables in our study on Singapore, $N = 177$ while the length of each time series is $T = 56$ quarters. The generalized dynamic factor model for these variables is given by

$$X_{it} = \lambda_i(L) f_t + \varepsilon_{it} \tag{1}$$

for $i = 1, \ldots, N$. The $(q \times 1)$ vector f_t contains the common dynamic factors and $\lambda_i(L) = \lambda_{i0} + \lambda_{i1}L + \cdots + \lambda_{is}L^s$ is an s th order polynomial in the lag operator L that represents a vector of dynamic factor loadings. Unlike in the exact factor model, the idiosyncratic disturbance ε_{it} is permitted to have limited serial and cross-correlation (see the discussion below and Forni *et al.*, 2000 and Stock and Watson, 2002b). However, the factors and idiosyncratic errors are assumed to be mutually uncorrelated at all leads and lags—an assumption that is essential for estimation of the factor model.

The central idea of the factor model is that information in a large dataset can be parsimoniously summarized by a small number of common factors i.e. $q \ll N$. As in principal component analysis, each factor is a weighted linear combination of the variables found in the information set. In other words, economic variables are pooled to average out noisy disturbances in the idiosyncratic component and to capture the relevant information in the common component. We assume that the latter explains the major part of the variation in observed time series regardless of the cross-sectional dimension. In the dynamic version of the factor model described by (1), current realizations of variables can also be affected by the lagged values of factors and this allows for a richer dynamic structure. Further, the *generalized* dynamic factor model does not restrict the order s of the factor loadings and at the same time relaxes the assumption of uncorrelated idiosyncratic terms utilized in traditional factor analysis. In particular, by allowing for both contemporaneous and lagged correlation between the idiosyncratic disturbances, it can accommodate the statistical features typically found in macroeconomic data for business cycle analysis and forecasting applications.

For estimation purposes, the model in (1) is reformulated as:

$$X_t = \Lambda F_t + \varepsilon_t \tag{2}$$

where $F_t = \left(f_t', \dots, f_{t-s}' \right)'$ is an $r = q\left(s+1\right)$-dimensional vector of stacked common factors and Λ is now an $(N \times r)$ matrix of factor loadings. The key advantage of this static representation is that the unobserved factors can be estimated consistently as $N, T \to \infty$ jointly by taking principal components of the covariance matrix of X_t, provided mild regularity conditions are satisfied (Stock and Watson, 2002a). Forni *et al.* (2000) among others showed that as N diverges, the principal components become increasingly collinear with the factors by virtue of a large number law. Hence, this non-parametric technique enables one to estimate the common factors from a potentially huge panel of related time series in a computationally convenient manner.[1]

[1] In fact, Stock and Watson (2002a) proved a stronger result: even if there is parameter instability caused by, say, structural change, the principal component estimates are still consistent because their precision improves with N, thus making it possible to compensate for short panels where T is relatively small, as exemplified by our set of data.

Determination of the number of dynamic factors

Bai and Ng (2007) recently demonstrated that the dynamic factor model (1) always has a static factor representation (2) in which the dynamics of F_t are characterized by a vector autoregression. In the same paper, they showed how the number of dynamic factors, q, can be determined from a knowledge of the number of static factors, r. Since some factors in the static model are dynamically dependent—being lags of the others—it follows that $q \le r$. This observation forms the basis of Bai and Ng's method to determine the value of q, which the authors interpret as a test for the number of primitive shocks driving economic fluctuations. Specifically, q is the number of non-zero eigenvalues in the residual correlation matrix of the static factor VAR.

The Bai-Ng procedure proceeds in two steps. In the first, the static factors are estimated by the principal component method and r is consistently selected using one of the six variants of information criteria developed in their earlier work (Bai and Ng, 2002). All the criteria are asymptotically equivalent but their small sample properties vary due to different specifications of the penalty term. The most widely used criterion and one of the best in terms of performance in simulations is the following:

$$IC(r) = \ln\left(V\left(r,F\right)\right) + r\left(\frac{N+T}{NT}\right)\ln\left(\min\left\{N,T\right\}\right) \tag{3}$$

$$\ln\left(V\left(r,F\right)\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \Lambda_i F_t\right)^2 \tag{4}$$

The penalty imposed by the second term in (3), which is an increasing function of N and T as well as the number of factors, serves to counter-balance the minimized residual sum of squares, thereby effecting an optimal trade-off between over-fitting and goodness of fit. Evidently, the criterion can be viewed as an extension of the Akaike information criterion (AIC) with consideration for the additional cross-sectional dimension to the time series. Applying it to our dataset with a pre-specified upper bound of 12 suggests that around 10 common factors should be included in the static model.

In the second step, the principal component estimators of F_t conditional on $r = 10$ are used to fit a pth-order VAR model and the least squares residuals obtained.

As mentioned above, the procedure to determine q is based on the estimated eigenvalues of the VAR residual correlation matrix. Let these be denoted as $\hat{c}_1 \geq \hat{c}_2 \geq \ldots \geq \hat{c}_r \geq 0$ in descending order. The marginal contribution of the k th eigenvalue is given by:

$$\hat{D}_k = \left(\frac{\hat{c}_{k+1}^2}{\sum_{j=1}^{r} \hat{c}_j^2} \right)^{\frac{1}{2}} \tag{5}$$

Under the assumption that the number of dynamic factors is q, $c_k = 0$ for $k > q$. Bai and Ng (2007) showed that $\hat{D}_k$ converges asymptotically to zero for $k \geq q$ at a rate depending on the sampling error induced by estimation of the VAR correlation matrix. Hence, the smallest integer k that satisfies the bounded set $\left\{ k : \hat{D}_k < m / \min\left[N^{\frac{1}{2}-\delta}, T^{\frac{1}{2}-\delta} \right] \right\}$ is the estimated number of dynamic factors in the model. For values of $m = 1.25$ and $\delta = 0.1$, the eigenvalue test selected $q = 3$ for the Singapore data.[2] Therefore, we utilize three dynamic factors in the forecasting exercises below.

Quasi-maximum likelihood estimation

The static principal components method outlined above is merely one approach to the estimation of dynamic factor models. Alternative methods have been used in the literature, such as dynamic principal components (Forni *et al.*, 2003) and quasi-maximum likelihood (Doz *et al.*, 2006), both of which estimate the dynamic factor model in (1) directly. The simulation results in Boivin and Ng (2005) suggest that the factor model estimated via static principal components is relatively more robust than its dynamic counterpart, partly because the latter require the specification of many auxiliary parameters—including the truncation lag parameter for spectral density estimation and the number of frequency grids—and is therefore more prone to specification errors. Indeed, an empirical application to US data shows that static principal components outperformed the dynamic factor estimates in terms of forecast accuracy. By contrast,

[2] These settings follow the design of the simulation experiments in Bai and Ng (2007). Like them, we employed a second order VAR model; in any case, the selection criterion mostly indicate three dynamic factors when we used VAR(1) and VAR(3) models instead.

Schumacher (2007) carried out forecast simulations on German data and found dynamic estimation to be superior, on the whole, to the static approach.

These mixed findings notwithstanding, we choose to perform quasi-maximum likelihood (QML) estimation of the dynamic factor model for Singapore as this method takes explicit account of the common factor dynamics through a VAR representation i.e.

$$\Gamma(L)F_t = \eta_t \tag{6}$$

where $\Gamma(L)$ is a matrix lag polynomial of finite order and η_t is multivariate white noise. By considering the joint estimation of the entire system, the QML approach has been shown to lead to modest efficiency improvements over static principal components (Doz *et al*., 2006). As it is based in the time domain, the method also requires fewer auxiliary parameters to be specified compared to dynamic principal components. Despite these advantages, maximum likelihood estimation of the dynamic factor model in large panels is thought to be infeasible and it has only been used to estimate the parameters of low-dimensional models (Stock and Watson, 1989).[3]

However, Doz *et al*. (2006) showed that QML estimation of the generalized factor model becomes computationally tractable as the cross-section enlarges. To carry it out, the dynamic factor model of (1) and (6) is first cast into state space form with the states being the r static factors. The Kalman filter can then be applied to evaluate the Gaussian likelihood and the likelihood maximized using the EM algorithm. Good initial estimates of model parameters and factors to initialize the numerical algorithm as well as a small number of states are important for the QML method to be feasible, however. The principal component estimates are used as they are good approximations to the common factors, particularly when N is large. It is also unsurprising that the number of iterations required for convergence is inversely related to the size of the panel, as shown by the simulations in Doz *et al.* (2006). In the final stage of the QML procedure, the dynamic factors are equated to their expected values, which are in turn computed from the Kalman smoother.

The true factors can be consistently estimated in this way as long as $N,T \to \infty$ under two conditions. First, the common component has to be pervasive even as the

[3] The exception is Quah and Sargent (1993), who implemented the technique for a large cross-section where the idiosyncratic errors are assumed to be independent.

cross-sectional dimension increases and second, the cross-correlation of the idiosyncratic components must be weak in an asymptotic sense. These are the same assumptions we made earlier when introducing the generalized dynamic factor model. The property of consistency in large samples explains why the maximum likelihood estimates for an exact factor model, where the idiosyncratic terms are assumed to be orthogonal and normally distributed, can be viewed as the QML estimates for the generalized factor model.

After extracting them from the Singapore and global time series, the QML estimates of the common factors are used to forecast the growth rates of overall GDP and value-added for the manufacturing (MFG), construction (CON) and services (SER) sectors in the next section. As a preliminary check, we regress these four series on their corresponding estimated common components. The regressions yielded R^2 values of 0.97, 0.96, 0.85 and 0.95 respectively. Such high coefficients of determination suggest that the estimated dynamic factor model provides a very good in-sample fit to the data, especially for these variables.

FORECASTING WITH FACTOR MODELS

We employ a common framework for generating pseudo-out-of-sample forecasts from the factor-based and other competing models. Initially, each forecasting model is estimated using data over the period 1993Q1 to 2004Q4 and its h-step ahead predictions calculated for $h = 1, \ldots, 4$ quarters (given the volatility of Singapore's economic growth, we eschew longer forecast horizons). Thereafter, the sample is augmented by one quarter, the parameters of the individual models are re-estimated and the corresponding h-step forecasts computed by moving the forecast window forward.[4] This recursive procedure is continued until the sample's end date reaches 2006Q4, at which point the final set of forecasts for the four quarters of 2007 are made, resulting in a combined total of 36 out-of-sample predictions at each forecast horizon for the four variables of interest.

[4] Although we would have liked to re-estimate the dynamic factors recursively too, this proved to be infeasible as the QML algorithm becomes unstable when the sample size is reduced. Consequently, the full sample period was used to estimate the factors.

Forecasting models

A distinctive feature of the recent work on forecasting with factor models lies in the way multi-period predictions are produced. Let the macroeconomic variable to be forecasted be denoted as X_{it} and the three dynamic factors identified in the previous section as $\hat{f}_t$. Then the h-step ahead forecast is computed directly by projecting $X_{i,t+h}$ onto its observable past and the estimated factors as follows:

$$X_{i,t+h} = \mu + \alpha(L)X_{it} + \beta(L)\hat{f}_t + e_{t+h} \tag{7}$$

At each prediction horizon, a separate forecasting equation is estimated by ordinary least squares techniques and the uniform order of the lag polynomials for the autoregressive component and the factors selected by minimizing the Bayes information criterion (BIC), starting with a maximum of 4 lags. In simulations, Stock and Watson (2002b) found that the BIC performs satisfactorily when used to select the optimal number of factors and their lags to be included in the forecasting equation.

The direct multi-step forecasting methodology prescribed by (7) differs from the usual approach whereby future predictions are generated dynamically by repeatedly iterating the one-step ahead forecasting model and replacing unknown values by their forecasts. The purported benefit of the direct method is that it obviates the need to model the evolution of the dynamic factors. Furthermore, any misspecification of the one-step ahead model will not be transmitted to the other forecast horizons since distinct models are estimated at each step of prediction. On the other hand, multi-step forecasting entails the estimation of a larger number of model parameters, thus reducing efficiency. Given this trade-off between bias and efficiency, which type of forecasts turn out to be more accurate in practice is largely an empirical question.

Boivin and Ng (2005) concluded that the direct approach works well with factor models. In contrast, Marcellino *et al*. (2006) reported for US data that the iterated method produces better predictions from autoregressive models. Here, we provide further empirical evidence by generating both multi-step and iterated projections based on the estimated common factors. Also known as 'plug-in' forecasts, the latter require the specification and estimation of a subsidiary model for the dynamic factors, the natural choice being the VAR model in (6). To facilitate comparison with the multi-step approach in (7), we include the variable of interest into the VAR and determine its optimal lag length with the BIC again:

$$\boldsymbol{y}_t = \boldsymbol{\tau} + \Gamma_1 \boldsymbol{y}_{t-1} + \cdots + \Gamma_p \boldsymbol{y}_{t-p} + \boldsymbol{\eta}_t \tag{8}$$

where $\boldsymbol{y}_t = \left[X_{it}, \hat{f}_t \right]$. The model in (8) has been called a factor-augmented VAR (FAVAR) by Bernanke, Boivin and Eliasz (2005) and can be used in principle for both policy analysis and forecasting. For the second purpose, predictions at the various horizons are computed simply by iterating the FAVAR model h steps ahead.

On leaving the dynamic factors out from the FAVAR, we get a univariate autoregression for each variable of interest. This constitutes the benchmark model with which the performances of the factor forecasting models are compared. We pick the lag length of the individual autoregressions through the Box-Jenkins methodology, making sure the residuals in the equations are white noise. It turns out that the AR models selected are all of order 5, thus allowing the complex roots in them to capture the cyclical behaviour of the data. Such models are therefore not as 'naïve' as they have been made out to be in the literature.

The multivariate competitor to the above models which we employ in the forecast comparison is small-scale VARs. The vector of stationary time series in these models is always a subset of the full dataset employed in the factor analysis, but it changes with the variable being forecast. When attempting to predict aggregate output growth, the vector on the left side of (8) is replaced by $\boldsymbol{y}_t = \left[GDP_t, FGDP_t, CLI_t, ELI_t, NEER_t, CA_t\right]'$, where $FGDP_t$ is a weighted average of incomes in Singapore's major trading partners, CLI_t is the official composite leading index, ELI_t is an electronics leading index, $NEER_t$ is the nominal effective exchange rate and CA_t represents total construction contracts awarded (see Department of Statistics, 2004 and Chow and Choy, 2006 for further details on the leading indexes).

In the models for manufacturing and services value-added growth, GDP_t is replaced by MFG_t or SER_t as the case may be and CA_t is dropped. A priori reasoning suggests, and practical experimentation confirms, that the leading indexes are not useful predictors of Singapore's construction sector; in fact, the building industry requires its own leading series (Chow and Choy, 1995). Consequently, the vector of endogenous variables for forecasting the growth rate of the construction sector is $\boldsymbol{y}_t = [CONSTR_t, CA_t, PPIRES_t]'$, the last entry representing the residential property price index. Before turning to the forecasting results, it should be noted that the predictions from the AR and VAR models are generated iteratively rather than directly, as this is what is usually done in practice.

Forecast comparison

The results of the pseudo-out-of-sample forecasting exercises are shown in Table II in the form of relative root mean square error (RMSE) measures.[5] Each statistic is computed as the ratio between the respective RMSEs of the model under consideration and the benchmark AR(5) model. A number exceeding unity indicates that the model's forecasting performance at a given horizon is worse than that of the benchmark and vice versa.

Putting aside the autoregressive forecasts, the models in competition can be ranked in this order: the FAVAR performs best overall, followed by the multi-step factor model, and then the small-scale VAR. That said, a few additional remarks are warranted. First, none of the models could beat the benchmark predictions in the case of manufacturing sector growth although they are able to do so for the other variables. Second, the direct multi-step and VAR forecasts occasionally produce forecast errors that are smaller than those of the iterated factor model, particularly in predicting construction growth. Third, the superiority of the FAVAR projections becomes more pronounced as the forecast horizon lengthens. In other words, it pays to explicitly model the future movements of dynamic factors in macroeconomic forecasting.

[5] We also calculated analogous statistics for the mean absolute error (MAE) but do not report them as they reveal similar findings. The results can be requested from the authors.

Table II. Forecast comparisons

Horizon (Quarters)	GDP			MFG			SER			CON		
	Factor	FAVAR	VAR	Factor	FAVAR	VAR	Factor	FAVAR	VAR	Factor	FAVAR	VAR
1	0.84	0.82	1.16	1.52	1.18	1.63	0.72	0.73	0.79	1.41	1.07	0.91
2	1.28	0.77	1.55	1.79	1.39	2.81	0.80	0.84	1.12	0.92	1.08	0.93
3	3.15	0.68	2.14	1.93	1.63	2.63	0.56	0.53	1.35	0.83	0.90	0.84
4	2.46	1.32	2.03	2.57	1.80	1.84	0.99	0.72	1.83	0.63	0.71	0.86

Notes: The numbers represent the RMSE statistics for the FAVAR, multi-step factor forecasting model and small VAR respectively, all relative to the benchmark AR model.

Of course, some of the observed differences between the RMSE ratios could just be attributed to chance. Table III assesses the influence of sampling variability on the prediction errors by presenting the Diebold-Mariano (1995) test statistics for the null hypothesis of equal forecast accuracy between the FAVAR and other models on a pairwise basis. In view of the relatively small number of observations involved, the small sample version due to Harvey *et al*. (1997) is reported. A negative Diebold-Mariano statistic in the table implies that the FAVAR model shows an improved forecast performance; if the difference in accuracy is statistically significant at the 10% level or lower, the statistic appears in bold.[6] For the one-step ahead forecasts, the hypothesis of equal predictive accuracy is hard to reject except for manufacturing output, where the AR model is clearly better. At the 2-step horizon, the iterated factor forecasts are significantly more accurate than direct forecasts in predicting the growth rates of GDP and the manufacturing sector. When $h \geq 3$, however, the FAVAR model tends to dominate its rivals for all the variables of interest. The gains in forecast accuracy over small-scale VARs is especially noteworthy, confirming that the cyclical information contained in the factors is more comprehensive.

[6] In several instances, the Diebold-Mariano statistic cannot be calculated in the usual way because the estimated spectral density at the origin is not guaranteed to be non-negative. When this happens, we use the Bartlett window to estimate the density and set the truncation lag equal to 4, as suggested by Diebold and Mariano (1995).

Table III. Tests of predictive accuracy

Horizon (Quarters)	GDP			MFG			SER			CON		
	AR	Factor	VAR	AR	Factor	VAR	AR	Factor	VAR	AR	Factor	VAR
1	–0.48	–0.35	–1.07	**1.45**	–0.78	–1.01	–0.81	0.12	–0.26	0.20	–0.94	0.98
2	–0.78	**–2.01**	**–1.51**	0.94	**–2.03**	–1.37	–0.34	0.57	–1.07	0.47	**1.84**	**1.65**
3	**–1.44**	–1.33	**–1.73**	**2.10**	–0.66	**–1.42**	**–1.85**	–0.09	**–1.76**	–0.37	0.42	0.27
4	0.16	–1.01	**–1.57**	1.34	**–57.09**	–0.20	–0.33	**–1.89**	**–3.94**	**–3.75**	0.47	**–3.78**

Notes: The numbers represent the small sample Diebold-Mariano statistics for the FAVAR models *vis-à-vis* the AR, multi-step factor, and VAR models. Bold figures denote statistical significance at the 10% level or lower.

CONCLUSIONS

Forecasting business cycles in small open economies is no easy task due to the myriad economic shocks that besiege such economies from time to time. Fortunately, recent developments in factor analysis have provided a parsimonious solution to this problem: first summarize the relevant information in a large macroeconomic dataset—including time series that capture external disturbances—through a small number of dynamic factors, and then use these to improve on *ex-ante* forecasts of economic aggregates.

In this endeavor, an important parameter to determine is the number of 'optimal' factors to exploit, which can also be interpreted as the number of primitive shocks driving business cycles. The results in this paper suggest that three dynamic factors are sufficient to explain over half of the observed macroeconomic fluctuations in Singapore. This is a remarkable finding when put in international perspective—typically, five to six factors are needed to explain the same proportion of variance in larger economies. Put in another way, Singapore's business cycles seem to be caused by a small number of relatively large shocks originating from the world at large, her neighbours in Asia, and the domestic property market.

Regardless of the economic interpretations given to the dynamic factors, prediction based on them can be carried out in two ways. The direct multi-step approach restricts the information set to the estimated factors and makes no attempt to project them. An alternative method examined in this paper models the dynamic process of the common factors explicitly through a factor-augmented vector autoregression. We find that iteratively generated forecasts of macroeconomic and sectoral aggregates in

Singapore from the FAVAR model generally outperform those based on the multi-step approach as well as multivariate VAR models and univariate autoregressions. It appears that the gains in efficient estimation brought about by the iterated approach outweigh any misspecification of the forecasting model. Moreoever, the improvements in predictive accuracy are shown to be systematic and significant at forecast horizons of 2–4 quarters. In conclusion, we might say that the dynamic factor model has proven its worth in forecasting the gyrations of small open economies.

APPENDIX: THE DATASET

Series	Mnemonic	Source	Transformation	SA Status
Foreign real GDPs (12)				
1. USA	USAGDP			
2. Japan	JAPGDP			
3. Korea	KORGDP	*Econometric*		
4. Rest of the OECD	ROECDGDP	*Studies*		
5. Malaysia	MALGDP	*Unit,*	Growth	Source
6. Indonesia	INDOGDP	*National*	rates	SA
7. Thailand	THAIGDP	*University*		
8. Philippines	PHILGDP	*of Singapore*		
9. Taiwan	TAIGDP			
10. Hong Kong	HKGDP			
11. China	CHINGDP			
12. Foreign GDP	FORGDP			
Foreign leading indexes (6)				
13. USA	USACLI			
14. Japan	JAPCLI		Deviations	
15. Germany	GERCLI	*SourceOECD*	from	Source
16. UK	UKCLI		trend	SA
17. 4 big European	EUROCLI			
18. 5 major Asian	ASIACLI			
Foreign stock prices (10)				
19. US	USASPI			Own SA
20. Japan	JAPSPI			NSA
21. Germany	GERSPI			Own SA
22. UK	UKSPI			Own SA
23. Korea	KORSPI	*Bloomberg*	Growth	NSA
24. Malaysia	MALSPI		rates	NSA
25. Indonesia	INDOSPI			NSA
26. Thailand	THAISPI			NSA
27. Philippines	PHILSPI			NSA
28. Hong Kong	HKSPI			NSA
Foreign real interest rates (3)				
29. US (3-mth LIBOR – CPI Δ)	USAIR	*International*		
30. Japan (3-mth LIBOR – CPI Δ)	JAPIR	*Financial*	Levels	NSA
31. UK (3-mth LIBOR – CPI % Δ)	UKIR	*Statistics*		

Series	Mnemonic	Source	Transformation	SA Status
World electronics (7)				
32. Global semiconductor sales	CHIP	*SIA*	Growth rates	Own SA
33. US Tech Pulse Index	TECH	*New York Fed*	Growth rates	NSA
34. Nasdaq index	NASDAQ	*Bloomberg*	Growth rates	NSA
35. US new orders for electronics (excl. semiconductors)	USNO	*US Census Bureau*	Growth rates	Source SA
36. US shipments-to-inventories ratio for electronics (excl. semiconductors)	USSI	*US Census Bureau*	Growth rates	Source SA
37. Book-to-bill ratio	BBR	*SEMI*	Levels	NSA
38. Electronics Leading Index	ELI	Authors	Growth rates	NSA
World prices (3)				
39. Real oil price (Dubai Fateh – World CPI Δ)	OIL	) *International*)	Growth	NSA
40. Non-fuel commodity prices	NONOIL	) *Financial*)	rates	NSA
41. World CPI	WORLDCPI	) *Statistics*)		NSA
Real GDP components (7)				
42. Real GDP	GDP	))	)	
43. Private consumption	CON	))	)	
44. Government consumption	GCON	))	Growth)	
45. Gross fixed capital formation	GFCF	) *STS*)	rates)	Source
46. Transport equipment	GFCFTPT	))	)	SA
47. Machinery, equipment & software	GFCFMEQ	))	)	
48. Net exports	NX	))	)	
Gross value-added (13)				
49. Manufacturing	MFG	))	)	
50. Construction	CONSTR	))	)	
51. Services	SER	))	)	
52. Commerce	COMM	))	)	
53. Wholesale & retail trade	WRTRADE	))	)	
54. Hotels & restaurants*	HOTREST	) *STS*)	Growth)	Source
55. Transport & Communications*	TRANSCOM	))	rates)	SA
56. Transport & storage*	TRANSTOR	))	)	
57. Information & communications	INFOCOM	))	)	
58. Financial & Business Services	FINBIZ	))	)	
59. Financial services	FIN	))	)	
60. Business services	BIZ	))	)	
61. Other Services	OTHER	))	)	

Series	Mnemonic	Source	Transformation	SA Status
Industrial production (7)				
62. Total	IIP			
63. Electronics	IIPELEC			
64. Chemicals	IIPCHEM	*STS*	Growth	Source
65. Biomedicals	IIPBIO		rates	SA
66. Precision engineering	IIPPRE			
67. Transport engineering	IIPTPT			
68. General manufacturing	IIPGEN			
Business surveys (17)				
69. Mfg investment commitments	COMMIT		Growth rates	NSA
70. General expectations for mfg	EXPMFG			Source SA
71. Employment expectations for mfg	EMPMFG			Own SA
72. General expectations for services	EXPSER			Own SA
73. Wholesale & retail trade	EXPWRTRADE			Own SA
74. Hotels & catering	EXPHOTREST			Own SA
75. Transport & storage	EXPTRANSTOR			Own SA
76. Financial services	EXPFIN		Net	NSA
77. Business services	EXPBIZ	*STS*	balances	Own SA
78. Real estate	EXPESTATE		of	Own SA
79. Employment expectations for services	EMPSER		firms	Own SA
80. Wholesale & retail trade	EMPWRTRADE			Own SA
81. Hotels & catering	EMPHOTREST			Own SA
82. Transport & storage	EMPTRANSTOR			NSA
83. Financial services	EMPFIN			NSA
84. Business services	EMPBIZ			Own SA
85. Real estate	EMPESTATE			NSA
Construction (14)				
86. GFCF in construction & works	GFCFCONSTR			Source SA
87. Residential buildings	GFCFRES			Source SA
88. Non-residential buildings	GFCFNRES			Source SA
89. Others	GFCFOTHER	*STS*	Growth	Source SA
90. Contract awards	CA		rates	NSA
91. Public	CAPUB			NSA
92. Private	CAPTE			NSA
93. Residential buildings	CARES			NSA
94. Commercial buildings	CACOMM			NSA
95. Civil engineering & others	CACIVIL			NSA

Series	Mnemonic		Source		Transformation	SA Status
96. Property price index (residential)	PPIRES	)	*STS*	)	Growth	Own SA
97. Property price index (office)	PPIOFF	)		)	rates	NSA
98. Property price index (shop)	PPISHOP	)		)		NSA
99. Resale price index	RESALE	)		)		Own SA
Sectoral Indicators (16)						
100. Retail sales volume	RETAIL	)		)		Own SA
101. Car registrations	CAR	)		)		Own SA
102. Visitor arrivals*	VISIT	)		)		Source SA
103. Air cargo handled	AIR	)		)		Own SA
104. Sea cargo handled	SEA	)		)		Own SA
105. Electricity generation	ELECTRIC	)		)		Own SA
106. Composite leading index	CLI	)		)		NSA
107. Formation of companies	FORM	)	*STS*	)	Growth	Own SA
108. Manufacturing	FORMMFG	)		)	rates	NSA
109. Construction	FORMCONSTR	)		)		Own SA
110. Wholesale & retail trade	FORMWRTRADE	)		)		Own SA
111. Hotels & restaurants	FORMHOTREST	)		)		Own SA
112. Transport & storage	FORMTRANSTOR	)		)		Own SA
113. Information & comms	FORMINFOCOM	)		)		NSA
114. Financial & insurance	FORMFIN	)		)		Own SA
115. Real estate & leasing	FORMESTATE	)		)		NSA
External Trade (16)						
116. Exports of goods & services	X	)		)		Source SA
117. Imports of goods and services	M	)		)		Source SA
118. Exports of goods	GX	)		)		Source SA
119. Oil	OGX	)		)		Source SA
120. Non-oil	NOGX	)		)		Source SA
121. Imports of goods	GM	)		)		Source SA
122. Oil	OGM	)	*STS*	)	Growth	NSA
123. Non-oil	NOGM	)		)	rates	Source SA
124. Exports of services	SX	)		)		Own SA
125. Imports of services	SM	)		)		Own SA
126. Domestic exports	DX	)		)		Source SA
127. Oil	ODX	)		)		Source SA
128. Non-oil	NODX	)		)		Source SA
129. Re-exports	RX	)		)		Source SA
130. Oil*	ORX	)		)		NSA
131. Non-oil	NORX	)		)		Source SA

Series	Mnemonic	Source	Transformation	SA Status
Price Indices (14)				
132. Export price index	XPI	)	)	NSA
133. Oil	OXPI	)	)	NSA
134. Non-oil	NOXPI	)	)	NSA
135. Import price index	MPI	)	)	NSA
136. Oil	OMPI	)	)	NSA
137. Non-oil	NOMPI	)	)	NSA
138. Terms of trade	TOT	) *STS*	) Growth	NSA
139. Consumer price index	CPI	)	) rates	Source SA
140. Domestic supply price index	DSPI	)	)	NSA
141. Manufactured price index	SMPI	)	)	NSA
142. GDP deflator	PGDP	)	)	Own SA
143. Manufacturing deflator	PMFG	)	)	NSA
144. Construction deflator	PCONSTR	)	)	NSA
145. Services deflator	PSER	)	)	Own SA
Labour Market (7)				
146. Changes in employment	EMP	)	Levels	NSA
147. Retrenchments	RETRENCH	)	Growth rates	NSA
148. Unemployment rate (overall)	U	) *STS*	Levels	Source SA
149. Unemployment rate (resident)	URES	)	Levels	Source SA
150. Unit labour costs	ULC	)	Growth rates	Source SA
151. Manufacturing unit labour costs	MULC	)	Growth rates	Source SA
152. Manufacturing unit business costs	MUBC	)	Growth rates	Source SA
Financial (17)				
153. Stock prices	SES	)	Growth rates	NSA
154. 3-mth interbank rate	INTER	)	Levels	NSA
155. 1-yr treasury bill yield	TB	)	Levels	Own SA
156. 2-yr bond yield	BOND2	)	Levels	Own SA
157. 5-yr bond yield	BOND5	)	Levels	NSA
158. Prime lending rate	PLR	)	Levels	NSA
159. Nominal effective exchange rate	NEER	) *STS*	)	NSA
160. Real effective exchange rate	REER	)	)	NSA
161. Singapore dollar to US$	USD	)	)	NSA
162. Singapore dollar to Pound	POUND	)	) Growth	NSA
163. Singapore dollar to Yen	YEN	)	) rates	NSA
164. Singapore dollar to Malaysian $	RINGGIT	)	)	NSA
165. Singapore dollar to HK$	HKD	)	)	NSA

Series	Mnemonic	Source	Transformation	SA Status
166. Singapore dollar to Korean won	WON	)	)	NSA
167. Singapore dollar to Taiwan $	NTD	) *STS*	) Growth	NSA
168. Singapore dollar to Indo rupiah*	RUPIAH	)	) rates	NSA
169. Singapore dollar to Thai baht	BAHT	)	)	NSA
Monetary (8)				
170. M1	M1	)	)	Source SA
171. M3	M3	)	)	Source SA
172. Bank loans	LOAN	)	)	NSA
173. Manufacturing	LOANMFG	) *STS*	) Growth	NSA
174. Building & construction	LOANCONSTR	)	) rates	NSA
175. Commerce	LOANCOMM	)	)	Own SA
176. Financial institutions	LOANFIN	)	)	NSA
177. Professional & pte individuals	LOANPRO	)	)	NSA

Notes: Figures in parentheses represent the number of variables in each category. STS is the Singapore Department of Statistics online time series database. SA (NSA) indicates series that have (not) been deseasonalised. Time series adjusted for outliers are marked with an asterisk.

REFERENCES

Artis MJ, Banerjee A, Marcellino M. 2005. Factor forecasts for the UK. *Journal of Forecasting* **24**: 52–60.

Bai J, Ng S. 2002. Determining the number of factors in approximate factor models. *Journal of Business and Economic Statistics* **25**: 147–162.

Bai J, Ng S. 2007. Determining the number of primitive shocks in factor models. *Econometrica* **70**: 191–221.

Boivin J, Ng S. 2005. Understanding and comparing factor-based forecasts. *International Journal of Central Banking* **1**: 117–151.

Bernanke BS, Boivin J, Eliasz P. 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics* **120**: 387–422.

Burns AM, Mitchell WC. 1946. *Measuring Business Cycles*. National Bureau of Economic Research: New York.

Choy KM. 2006. Business cycles in Singapore: stylized facts for a small open economy. Mimeo.

Choy KM, Chow HK. 1995. Economic leading indicators for tracking Singapore's growth cycle. *Asian Economic Journal* **9**: 177–191.

Choy KM, Chow HK. 2006. Forecasting the global electronics cycle with leading indicators: a Bayesian VAR approach. *International Journal of Forecasting* **22**: 301–315.

Danthine JP, Girardin M. 1989. Business cycles in Switzerland: a comparative study. *European Economic Review* **33**: 31–50.

Department of Statistics, Singapore. 2004. Singapore's growth chronology, coincident and leading indicators. Information Paper on Economic Statistics.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253–263.

Doz C, Giannone D, Reichlin L. 2006. A quasi-maximum approach for large approximate dynamic factor models. European Central Bank Working Paper No. 674.

Englund P, Persson T, Svenson LEO. 1992. Swedish business cycles: 1861–1988. *Journal of Monetary Economics* **30**: 343–371.

Forni M, Hallin M, Lippi M, Reichlin L. 2000. The generalized dynamic factor model: identification and estimation *The Review of Economics and Statistics* **82**: 540–554.

Forni M, Hallin M, Lippi M, Reichlin L. 2003. Do financial variables help forecasting inflation and real activity in the Euro area? *Journal of Monetary Economics* **50**: 1243–1255.

García-Ferrer A, Poncela P. 2002. Forecasting European GNP data through common factor models and other procedures. *Journal of Forecasting* **21**: 225–244.

Harvey D, Leybourne S, Newbold P. 1997. Testing the equality of prediction mean squared errors. *International Journal of Forecasting* **13**: 281–291.

Kim K, Buckle RA, Hall VB. 1994. Key features of New Zealand business cycles. *The Economic Record* **70**: 56–72.

Kim K, Choi YY. 1997. Business Cycles in Korea: is there any stylized feature? J*ournal of Economic Studies* **24**: 275–293.

Kose MA. 2002. Explaining business cycles in small open economies: how much do world prices matter? *Journal of International Economics* **56**: 299–327.

Kose MA, Ostrok C, Whiteman CH. 2003. International business cycles: world, region, and country-specific factors. *American Economic Review* **93**: 1216–1239.

Leung CF, Suen W. 2001. Some preliminary findings on Hong Kong business cycles. *Pacific Economic Review* **6**: 37−54.

Marcellino M, Stock JH, Watson MW. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**: 499−526.

Quah D, Sargent TJ. 1993. A dynamic index model for large cross sections. In *Business Cycles, Indicators, and Forecasting*, Stock JH, Watson MW (eds). University of Chicago Press: Chicago.

Sargent TJ, Sims CA. 1977. Business cycle modeling without pretending to have too much *a priori* economic theory. In *New Methods in Business Cycle Research*, Sims CA (ed.). Federal Reserve Bank of Minneapolis: Minneapolis.

Schumacher. 2007. Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting* **26**: 271–302.

Stock JH, Watson MW. 1989. New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*.

Stock JH, Watson MW. 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* **97**: 1167–1179.

Stock JH, Watson MW. 2002b. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* **20**: 147–162.