Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Forecasting 21 (2005) 377–389

international journal of forecasting

www.elsevier.com/locate/ijforecast

Business survey data: Do they help in forecasting GDP growth?

Jesper Hansson[a], Per Jansson[b,*], Mårten Löf[a]

[a]National Institute of Economic Research, SE-103 62 Stockholm, Sweden
[b]Sveriges Riksbank, Monetary Policy Department, SE-103 37 Stockholm, Sweden

Abstract

In this paper we examine whether data from business tendency surveys are useful for forecasting GDP growth in the short run. The starting point is a so-called dynamic factor model (DFM), which is used both as a framework for dimension reduction in forecasting and as a procedure for filtering out unimportant idiosyncratic noise in the underlying survey data. In this way, it is possible to model a rather large number of noisy survey variables in a parsimoniously parameterised vector autoregression (VAR). To assess the forecasting performance of the procedure, comparisons are made with VARs that either use the survey variables directly, use macro variables only, or use other popular summary indices of economic activity. Our DFM-based procedure turns out to outperform the competing alternatives in most cases.

Keywords: Business survey data; Dynamic factor models; Macroeconomic forecasting

1. Introduction

The interest in, and demand for, macroeconomic analyses at high frequencies, most notably forecasts, has increased substantially in recent years. But conducting analyses and making forecasts of high-frequency data is not an easy task. Compared with annual data, data that are observed daily, monthly, and quarterly typically display more complicated dynamics, are seasonal, and are—at least as concerns real variables—more frequently revised. One category of data that has the potential to be rather useful in this context is that produced by surveys. Survey data have the advantages of essentially being instantaneously accessible, never being revised, and, furthermore, having little or no measurement errors. The objective of the present paper is to explore whether such data can successfully be used for purposes of forecasting GDP growth. The forecasting performance for other key macro variables (e.g., inflation, unemployment, employment, wage inflation, interest rates, and the exchange rate) is investigated in a longer version of this paper, see Hansson, Jansson, and Löf (2003).

Our empirical application is based on the Swedish Business Tendency Survey (BTS), which is a large

* Corresponding author. Tel.: +46 8 787 01 55; fax: +46 8 21 05 31.
E-mail address: per.jansson@riksbank.se (P. Jansson).

0169-2070/$ - see front matter © 2004 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2004.11.003

business survey based on questions about economic activity posed to approximately 7000 different firms in various sectors of the Swedish economy. The sectors currently included are manufacturing, construction, and, since 1991, services. As a percentage of the total number of employed workers, these sectors cover around 50% of the Swedish economy. The full survey is undertaken quarterly, but a subset of the survey has also been available monthly since 1996. If the analysis is limited to the manufacturing and construction industries, then the survey provides continuous time series for most questions since the mid-1970s. In this case, the number of survey questions used is approximately 3000 and the coverage in terms of employed workers around 25%.

The questions of the survey are both coincident (regarding the situation in the current quarter) and forward-looking (regarding the outlook for the next quarter). When responding, the firms are merely asked to specify whether a particular activity (e.g., production or order flows) has increased, been unchanged, or decreased (or, in the forward-looking case, whether the activity is expected to increase, remain unchanged, or decrease). In some cases, the questions are dichotomous, just requiring a “yes” or a “no”. The final quantities used are “net balances” obtained by subtracting the weighted percentages of firms that have specified an increase from the weighted percentages of firms that have specified a decrease (or, just the weighted shares if the questions are dichotomous).[1]

Previous research into the forecasting properties of the Swedish BTS has focused on establishing direct relationships between the variable to be forecast (typically, the growth rate of industrial production) and the BTS data. The approaches range from simple single-equation models, e.g., Bergström (1992, 1993a) and Lindström (2000), to Kalman-filter-based updating schemes and sophisticated turning-point analyses, e.g., Kääntä and Tallbom (1993), Koskinen and Öller (2004), Öller and Tallbom (1996), and Rahiala and Teräsvirta (1993). A common finding is that only very few BTS variables are useful for macro forecasting, and that the information content in the forward-looking BTS series is particularly weak. One of the main results of the present paper is that the forecasting performance of the BTS data can be considerably improved if the BTS variables are appropriately filtered prior to forecasting, and thus indirectly, rather than directly, related to the variable to be forecast.

The benefits of making use of an indirect, rather than direct, link between the BTS data and the data to be predicted stem from the fact that changes in the BTS data cannot always be assumed to contain signals that are relevant for activity at the aggregate level. More specifically, it seems likely that idiosyncratic sector-specific changes in a particular series are largely unrelated to the overall state of economic activity. The filtering technique thus entails getting rid of this series-specific “noise” and only keeping those parts of the data that are common to the series under consideration. One previous analysis that supports the premise that the forecasting performance of the Swedish BTS may be enhanced by filtering techniques is that undertaken by Christoffersson, Roberts, and Eriksson (1992). Although these authors do not explicitly favour the kind of filtering procedure that we propose, they show, using methods in the frequency domain, that the BTS series are both noisy at high frequencies and highly collinear. This makes it difficult to directly include them as explanatory variables in a conventional forecasting equation.

As it happens, the proposed filtering procedure also has the property of implying a dimension-reduction framework for the BTS variables. From the forecasting literature, it is well known that forecasting approaches that use many explanatory variables, and thus many estimated parameters, generate forecasts that quickly become inefficient and unstable. Our proposed procedure addresses this issue by summarising the observable information of the large BTS data set in a single common-factor index.

To undertake the filtering of the BTS data, we employ a standard so-called dynamic factor model (DFM). Such models have previously been used for similar purposes, but have hardly been applied to survey data. Useful general references include

[1] Almost identical surveys, spanning similar time periods, exist in practically all other countries of the EU (and in the US). Therefore, there are good reasons to believe that the results obtained in this paper are indicative of the general value of using such survey data in forecasting economic activity. For full details of the Swedish BTS, see http://www.konj.se (the homepage of the National Institute of Economic Research, which publishes the survey).

Camba-Mendez, Kapetanios, Smith, and Weale (1999), Forni, Hallin, Lippi, and Reichlin (2000, 2003), Fukuda and Onodera (2001), and Stock and Watson (1989, 1991, 1999, 2002). Bruno and Malgarini (2002), European Commission (2000), and Goldrian, Lindbauer, and Nerb (2001) are examples of studies that make use of survey data. The models developed in Camba-Mendez et al. (1999), Forni et al. (2000, 2003), and Stock and Watson (1999) make use of some survey variables but are mainly based on macro variables. The DFM, and further issues related to the BTS data, are discussed in Section 2.

The forecasting performance of the BTS data filtered by the DFM is investigated using almost-real-time out-of-sample experiments. Here, the idea is that the forecaster takes the estimated common-factor index as given and computes the forecasts as if there is no knowledge about the generating mechanisms of the index. Thus, the forecaster fits a standard dynamic forecasting model, a vector autoregression (VAR). Each VAR is bivariate and consists of the estimated common-factor index and the variable that we wish to forecast, GDP growth.

To assess the relative accuracy of the DFM-based VAR forecasts, we compare the forecasts using three alternative approaches to forecasting GDP growth. These are VARs that use unfiltered BTS variables (i.e., that include the survey variables directly without employing the DFM filter); VARs that use information on macro variables only; and VARs that use other popular summary indices of economic activity. The DFM-based forecasts of the growth rate of real GDP are discussed in Section 3.1. Alternative forecasts of GDP growth are evaluated in Section 3.2.

By making comparisons with VARs based on the unfiltered survey data, we are able, in terms of forecast precision, to assess the gain derived from first applying the DFM to the BTS data (relative to not doing so). That is, we can quantify the effects on forecasting by parsimoniously modelling the noise-reduced BTS series rather than the original series themselves. The comparisons with macro VARs instead enable us to judge our performance relative to the "standard" forecasting model. Finally, the comparisons with VARs based on other summary indices of activity allow us to shed some light on the performance of our procedure when holding the gains of dimension reduction constant. Like the DFM procedure, such summary indices have the advantage of enabling the use of very parsimonious forecasting models, without having to give up too much of the relevant forecasting information.

2. The dynamic factor model

In this section, we discuss and estimate the dynamic factor model used to filter the business survey data. The output of this analysis is an estimate of a common-factor index which summarises the co-movements in a broad range of different economic activities such as production, order flows, time of deliveries, employment, and stocks of raw materials and goods. The index is constructed in such a way that it acknowledges that activities occur in different sectors of the economy and that there may be lead–lag relationships between them. Because the questions of the survey regard both activities in the current and next quarter, the whole analysis is undertaken for two different versions of the index: one coincident (current quarter) and one forward-looking (next quarter).

2.1. Specification

Let the n-dimensional vector that collects the relevant BTS series be denoted by X_t. It is assumed that the variables in X_t are (stochastically) stationary so that they can be normalised to have zero mean and unit variance. The assumption of stationarity is not restrictive: all BTS series are distinctively cyclical without trends (whether stochastic or deterministic). Standard unit-root tests confirm that all series are stationary $I(0)$.

In the model, X_t is driven by two stochastic components: the unobserved scalar index C_t, which is common to all the variables in X_t, and the equally unobserved n dimensional component I_t, which represents the idiosyncratic movements in the series. The model, in its general form, is:

$$X_t = \gamma(L)C_t + I_t, \tag{1}$$

$$\phi(L)C_t = \eta_t, \tag{2}$$

where L is the lag operator such that $L^j y_t = y_{t-j}$ for any vector or scalar variable y while $\gamma(L)$ and $\phi(L)$ are vector and scalar lag polynomials, respectively. The elements in I_t and η_t are the system's disturbances such that idiosyncratic shocks are purely temporary while common shocks may display some persistence. The assumption of a purely temporary process for the idiosyncratic component can be relaxed in favour of more general autoregressive specifications but was found to fit the data well in this particular application.

As it stands, model (1)–(2) is a standard DFM. As is well known, it is econometrically unidentified unless restrictions on its feasible set of parameter values are imposed. The following restrictions can be shown to be sufficient for identification: the disturbances in I_t and η_t are mutually and serially uncorrelated; the scalar C_t enters at least one of the variables in Eq. (1) only contemporaneously; and the standard deviation of η_t is normalised to unity (or, equivalently, one of the contemporaneous parameters in $\gamma(L)$ is normalised to unity).

To estimate the model, it is cast in so-called state-space form. One can then apply the Kalman filter together with a numerical optimisation routine to obtain Maximum Likelihood estimates of the unknown parameters and the unobserved components, i.e., the common-factor index C_t and the idiosyncratic noise processes in I_t; for details see, e.g., Harvey (1989). The analysis permits calculation of both one- and two-sided estimates of C_t (and I_t). The former are conditional on the information available at time t (written $C_{t|t}$) while the latter are conditional on the information available at end of sample (written $C_{t|T}$, with T being the last observation of the sample). The two-sided estimates $C_{t|T}$ may thus be regarded as being the "best" possible guesses of C_t, given a particular sample t=1,. . .,T.

In the forecasting exercises presented below, we make use of the one-sided estimates of C_t only. The reason is that we wish to simulate a recursive out-of-sample forecasting experiment without having to update the estimated DFM in each recursion. Given that the BTS data are not revised over time, the one-sided estimates will be approximately real time provided the DFM is empirically stable.[2] Ideally, the experiments should be carried out using the two-sided estimates generated by a recursively updated DFM but, since the model has to be solved numerically, this approach is infeasible.[3]

Having discussed the technical aspects of the DFM, we now turn our attention to the BTS variables included in the vector X_t. Although model (1)–(2) is quite flexible, it is parametric and thus has limitations as regards the number of variables that it can handle. All in all, the quarterly Swedish BTS at present includes 39 variables related to the manufacturing sector, and 19 variables related to the construction sector. Of these, roughly 25% are available as forward-looking (8 for manufacturing and 4 for construction). When deciding which of the variables to use in the DFM the following circumstances have been important. First, some of the variables have rather short time series and

Table 1
The BTS variables[a]

Activity	Coincident	Forward-looking
Manufacturing industries		
Production	BTVI101	BTVI301
Orders received (domestic)	BTVI105	BTVI305
Orders received (exports)	BTVI106	BTVI306
Time of deliveries	BTVI108	
As-of-now judgement of order books	BTVI201	
Number of workers employed	BTVI203	BTVI308
As-of-now judgement of stocks of raw materials	BTVI208	
As-of-now judgement of stocks of finished goods	BTVI210	
Construction industries		
Construction	BBOA101	BBOA201
Stocks of offers accepted	BBOA102	BBOA202
As-of-now judgement of order books	BBOA104	
Number of workers employed	BBOA106	BBOA204

[a] Each entry gives the code used by the National Institute of Economic Research to denote the particular survey question. The sample runs from 1978:1–2001:4 in the case of coincident variables and from 1978:2–2002:1 in the case of forward-looking variables.

[2] For the coincident data, the correlation between the full-sample and pre-forecast-period estimates of the common-factor index is 0.98. For the forward-looking data, the corresponding correlation is 1.0. This suggests that the estimated DFMs are indeed empirically stable.

[3] Although it may not be possible to update the DFM itself recursively, the two-sided estimates of the common-factor index are easily updated recursively given the full-sample estimate of the DFM. The empirical results when using these estimates instead of the one-sided ones were approximately the same. The paper uses the one-sided estimates because they are somewhat easier to compute and more common.

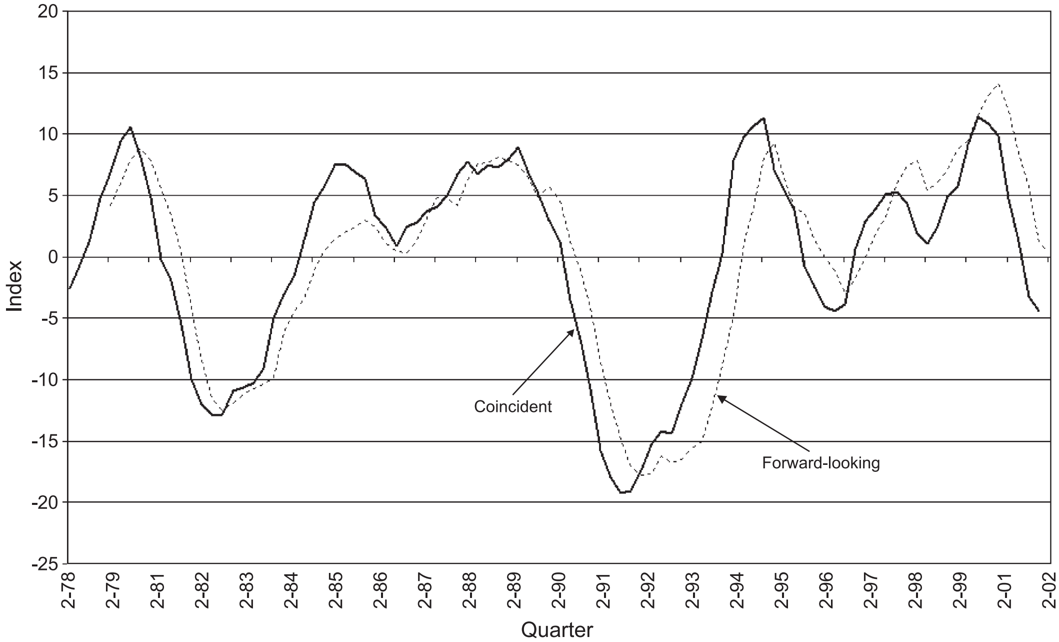

Fig. 1. One-sided estimates of common factors.

are therefore not well suited for econometric analyses of the kind undertaken here. Second, a couple of variables give information about similar activities and are therefore redundant. Third, nominal variables, which only provide indirect information on the amount of real activity, are excluded altogether. Using these criteria, we end up with a feasible set consisting of 12 coincident variables and 7 forward-looking variables. The details are shown in Table 1.

2.2. Results

The estimates of the coincident and forward-looking indices obtained by estimating the DFM with the BTS variables appear in Fig. 1. As emphasised above, the indices are constructed directly from the one-sided estimates of the common factors using either coincident BTS data only or forward-looking BTS data only.[4] Because the DFM is a pure time-series filter, its parameters have no particular interpretation. For this reason, and for expository convenience, we do not explicitly present the estimation results here (although these are of course available upon request). Residual diagnostics (again not shown for purposes of saving space) suggest that the two estimated DFMs by and large have acceptable statistical properties.

The estimated common-factor indices appear to accord rather well with common interpretations of cyclical developments in the Swedish economy over the last two decades (for further discussions, see Hansson et al., 2003). The next issue to be dealt with is to investigate whether this information can also be used successfully for the purpose of out-of-sample forecasting.

3. Forecasts

In this section, we undertake an almost-real-time out-of-sample forecasting experiment that aims to shed light on how useful the two estimated common-factor indices are for forecasting GDP growth in the short run. The forecasting model used throughout is a standard VAR, and the forecast accuracy is evaluated for four different forecast

[4] We also experimented with models that were based on differences between the coincident and forward-looking variables but such models did not perform well. The final models include the variables BTVI101, BTVI105, BTVI301, BTVI305, BTVI306, BTVI308, BBOA101, BBOA102, BBOA106, BBOA201, BBOA202, and BBOA204 (cf. Table 1).

horizons: one quarter, two quarters, four quarters, and eight quarters. To assess the relative accuracy of the DFM-based VAR forecasts, we make forecast comparisons using three alternative approaches to forecasting GDP growth: VARs that use the unfiltered BTS variables, VARs that use macro variables only, and VARs that use other popular summary indices of economic activity. The macro VARs are based on the following feasible set of (stationary) macro variables (in addition to GDP growth): inflation according to headline CPI and UND1X, the percentage change of the effective exchange rate, unemployment, employment growth, short- and long-term interest rates, and wage inflation. UND1X inflation is one of the Riksbank's measures of "core inflation" and is defined as CPI inflation excluding household mortgage interest expenditure and indirect taxes and subsidies (for further details of the data, see the Data appendix). Non-stationary variables in levels are made stationary by using (log) fourth differences.[5] Concerning the arrival of information, we make the assumption that there is an information lag of one quarter for real macro variables (including GDP growth). Hence, when we observe the BTS variables, the other popular summary indices, and the nominal macro variables in quarter t, the real macro variables are only known up to and including quarter $t-1$.[6] This information lag corresponds approximately to the publication lag that prevails today for these particular variables.

The VARs that are estimated are as follows:

$$Y_t = \alpha(L)Z_t + \beta(L)Y_t + \varepsilon_t, \tag{3}$$

$$Z_t = \tilde{\alpha}(L)Z_t + \tilde{\beta}(L)Y_t + \tilde{\varepsilon}_t, \tag{4}$$

where constants are omitted for expository convenience. Here, Y denotes GDP growth and Z is a $q-1$-dimensional vector of predictors, such that the full VAR is q-variate, $q \geq 2$. When the BTS variables and the popular summary indices are used, the polynomial $\alpha(L)$ includes contemporaneous effects; i.e., Z enters Eq. (3) both contemporaneously and lagged. The macro VARs are, on the other hand, restricted such that Z only enters Eq. (3) lagged. The remaining polynomials $\beta(L)$, $\tilde{\alpha}(L)$, and $\tilde{\beta}(L)$ are always restricted to exclude contemporaneous effects. The lag length is determined by minimising the system-based Bayesian information criterion (BIC).

The VARs that are based on the estimated common factors and the popular summary indices are always bivariate (q=2). Thus, in these models, Z is a scalar containing a single leading indicator of Y. In the case of macro VARs and VARs based on unfiltered BTS variables, we impose the restriction that the variable dimension is at most of fifth order ($q \leq 5$).[7]

The forecasts are computed as follows. The full sample period has 2001:3 as its last quarterly observation. To undertake the out-of-sample experiments we exclude quarters 1995:3–2001:3 (25 observations). Since we wish to derive the forecasts recursively, the procedure entails repeated re-estimation of Eqs. (3) and (4) by successively adding observations from the excluded quarters. In each recursion, we generate forecasts of GDP growth at the one-, two-, four-, and eight-quarter horizons. The exact procedure in the case Z in Eqs. (3) and (4) is the estimated coincident or forward-looking common factor is outlined in Table 2.

An alternative to using VARs to generate the h-step-ahead forecasts is to regress the variable to be forecast on information dated $t-h$ and earlier (see, e.g., Stock & Watson, 1999). Such specifications have the advantage that they do away with the need of forecasting the right-hand-side variables in Eq. (3) when deriving dynamic forecasts of Y. On the other hand, they are based on the assumption that Z in period t is (conditionally) correlated with Y in period $t-h$; an assumption that successively becomes less appealing as h is growing. As it happens, the results in this paper do not change much if we use this alternative setup instead.

The analysis of forecast accuracy is mainly based on the out-of-sample (root) mean-squared (forecast)

[5] In Hansson et al. (2003), we also use (log) first differences. This is of some importance but does not affect our results qualitatively.

[6] Wage inflation is an exception, however. Since this variable is generated from wage sums and hours worked in the national accounts it is subject to the same publication lag as the real macro variables.

[7] It is of course also possible to summarise the information content of the macro variables using a DFM. Given the results obtained in this paper, it seems plausible that this would improve the forecasting performance of the macro VARs. However, because the primary focus here is on the forecasting performance of survey data, this is left for future work.

Table 2
Setup for recursive forecasts of GDP growth using the estimated common factors[a]

	Coincident index (C)	Forward-looking index (C)
Sample Eq. (3)	1978:4–s, s=1995:2,. . .,2001:2	1979:4–m, m=1995:2,. . .,2001:2
Sample Eq. (4)	1978:4–s, s=1995:3,. . .,2001:3	1979:4–m, m=1995:3,. . .,2001:3
Information set	$C_{78:4}$,. . .,C_s, $Y_{78:4}$,. . .,Y_{s-1}, s=1995:3,. . .,2001:3	$C_{79:4}$,. . .,C_{m+1},$Y_{79:4}$,. . .,Y_{m-1}, m=1995:3,. . .,2001:3

[a] For each type of recursive model (distinguished by time indices s and m), the table gives the estimation samples and the information set used in each forecasting recursion. For example, when C is the forward-looking index, the first recursion estimates Eq. (3) over the sample 1979:4−1995:2 and Eq. (4) over 1979:4−1995:3. The available information set in this recursion contains GDP growth up to and including 1995:2 and BTS data up to and including 1995:4. The forward-looking index has fewer observations than the coincident index due to differences in lag structures, procedures for initial values, etc. in the DFM.

error (RMSE). Under the hypothesis of unbiasedness, the RMSE is simply the standard deviation of the out-of-sample forecast errors. When analysing performance in relative terms, we compute ratios of RMSEs and test the hypothesis of equal forecast accuracy using the modified Diebold and Mariano (1995) test suggested by Harvey, Leybourne, and Newbold (1997). The test (which is based on ratios of MSEs rather than RMSEs) uses a small-sample correction and generates a t statistic with $n-1$ degrees of freedom, where n is the number of h-step forecasts available. By this measure, our DFM-based forecasts of GDP growth significantly outperform the competing alternatives in several cases.

In our benchmark estimations, we fit all VARs without paying any attention to the models' in-sample performance. To gain some insights into how the analysis is affected if the VARs are required to fulfil criteria of in-sample performance, we repeat all forecasting experiments conditional on the forecasting equations satisfying certain tests of error-term adequacy. We compute three standard tests of model misspecification: the Breusch–Godfrey LM test against autocorrelation; Engle's LM test against ARCH effects; and Chow's parameter stability test. For the in-sample conditioning filter not to be too restrictive, we choose to consider a particular model as having acceptable in-sample properties if it passes at least two of the three error-term tests (at the conventional 5% test error margin). The reason for conditioning the forecasting models on their in-sample performance is that forecasters in practice presumably would pay some attention to such aspects when deriving their models. On the other hand, there is evidence that suggests that the link between in-sample and out-of-sample performance is (at best) weak (see, e.g., Clements & Hendry, 1998, 1999). Therefore, using in-sample criteria that are too restrictive may entail eliminating models that do not perform very well in sample but nevertheless work well for purposes of out-of-sample forecasting. This constitutes the main reason for disregarding in-sample performance in the benchmark estimations, and only using an informal procedure when investigating robustness of results with respect to such considerations.

3.1. GDP-growth forecasts using the estimated common factors

With the two-out-of-three requirement concerning the misspecification tests, all VAR models that use the estimated common factors turn out to qualify for the conditional out-of-sample forecasting comparisons (for details, see Hansson et al., 2003). We note that, to the extent that the models do not pass the misspecification tests, it is the parameter stability requirement that seems to be the most difficult criterion to fulfil (this is a general finding that holds true for all forecasting models investigated in this paper). This is interesting because the models have been explicitly designed to be (very) parsimoniously parameterised. Thus, although parsimonious, the models still display tendencies of instability. Judging from previous evidence, it is unclear whether this is a problem or not when it comes to out-of-sample forecasting. The results in this paper (as shown below) support the previous finding that in-sample performance is largely unrelated to out-of-sample forecasting accuracy.

Table 3 summarises the properties of the recursive GDP-growth forecasts using various measures of forecast accuracy. Looking first at the reported RMSEs we conclude that forecasts, even at very narrow horizons, are surrounded by a considerable amount of uncertainty. A 95% con-

Table 3
Forecast-error analysis for bivariate DFM-based VAR models: GDP growth[a]

	RMSE	MAE	RMedSE	Bias	Rank (RMSE)
Coincident index					
One-quarter	0.98	0.79	0.59	0.07	1
Two-quarter	1.14	0.93	0.71	0.13	4
Four-quarter	1.80	1.58	1.65	0.37	6
Eight-quarter	2.09	1.96	2.05	1.56	7
Forward-looking index					
One-quarter	1.04	0.81	0.71	0.18	2
Two-quarter	1.11	0.87	0.79	0.23	3
Four-quarter	1.59	1.28	1.19	0.16	5
Eight-quarter	2.21	2.08	2.16	1.75	8

[a] Columns one (RMSE) and two (MAE) give conventional root mean-squared errors and mean absolute errors. Columns three (RMedSE) and four (Bias) contain median-based RMSEs and average errors (in absolute terms).

fidence interval for normally distributed forecast errors at the one- and two-quarter horizons has a width of roughly 4 percentage points. The size of the typical forecast error (as measured by the mean absolute error, MAE) at the one- and two-quarter horizons is in the range 0.8–0.9 percentage points, reflecting the high uncertainty associated with the forecasts. In absolute terms, the average error (bias) is smaller than the MAE, which mirrors the fact that positive and negative errors to some extent cancel each other. Although our experiments are based on a rather limited amount of forecasts, in most cases, the bias is appreciably small.[8] Not surprisingly, all forecasts become successively less accurate as one prolongs the forecasting horizon.

From a more detailed study of the time paths of the forecast errors, it becomes evident that particularly large errors occur when the growth rate is at, or close to, a "turning point" (in the sense that growth switches from being increasing to decreasing and vice versa). To discern the extent to which the measures of forecast accuracy are influenced by large prediction errors that occur relatively infrequently, Table 3 also gives median-based measures of forecast accuracy (called RMedSEs). Comparing the root of the median-squared errors with the usual RMSEs, it can indeed be seen that large errors in the tails of the forecast-error distributions (which translate to the distributions of squared errors being skewed to the right) contribute to significantly worsening the forecasting performance of the procedures. This "turning-point" problem is typical for linear forecasting models, which are highly influenced by the (average) persistence of the variables in the conditioning set. A non-linear alternative

[8] Over the forecasting period (1995:3–2001:3), the sample mean of GDP growth is 2.73% (with variance 2.35). Thus, as a fraction of the mean, the bias (at the one- and two-quarter horizons) is in the range 3–8%.

Table 4
Alternative forecasts of GDP growth: coincident index and forward-looking index vs. unfiltered BTS variables[a]

	One-quarter	Two-quarter	Four-quarter
Coincident index vs. unfiltered BTS variables			
Median	1.23*** (1.20)	1.24* (1.24)	1.08 (1.07)
Mean	1.25‡ (1.22)	1.25‡ (1.23)	1.08‡ (1.08)
Best one-quarter	0.96 (0.96)	1.19** (1.19)	1.09 (1.09)
Best two-quarter	1.06 (1.07)	0.89 (0.92)	0.77** (0.79)
Best four-quarter	1.06 (1.07)	0.89 (0.92)	0.77** (0.79)
Forward-looking index vs. unfiltered BTS variables			
Median	1.23*** (1.23)	1.27* (1.28)	1.13* (1.13)
Mean	1.23‡ (1.23)	1.28‡ (1.28)	1.13‡ (1.13)
Best one-quarter	1.11 (1.11)	1.14 (1.14)	0.96 (0.96)
Best two-quarter	1.11 (1.11)	1.14 (1.14)	0.96 (0.96)
Best four-quarter	1.11 (1.11)	1.15 (1.15)	0.95 (0.95)

[a] All figures are relative RMSEs. The models that make use of the common-factor index appear in the denominator so that figures greater (smaller) than unity mean that the models based on the common factor outperform (are outperformed by) the models based on the unfiltered BTS variables. *, **, and *** indicate that the one-sided (modified) Diebold and Mariano (1995) test is significant at the 20%, 10%, and 5% levels, respectively (‡ means that no test is available). In the rows labelled "Median" and "Mean", the RMSEs of the models based on the unfiltered BTS variables are central-tendency moments of empirical distributions. The distributions are generated from the forecasts of all possible VAR models of dimension five or less (including the variable to be forecast) using the feasible set of BTS variables outlined in Table 1. In the rows "Best x-quarter", the RMSEs of the models based on the unfiltered BTS variables are optimised such that they are at their minima at the x-quarter horizon (again making use of the empirical distributions). The figures in brackets are results that condition the forecasting models on satisfying certain residual diagnostics criteria (see the text for details). The shares of the models that are excluded when subjected to the diagnostics criteria are 31.1% in the case of coincident data and 5.0% in the case of forward-looking data. The lag lengths of the VARs are determined by minimising the system-based BIC.

provides a potential solution to the problem, but applying such a framework would go beyond the scope of this paper.

As mentioned previously, in Hansson et al. (2003) we give a more extensive evaluation of forecasting performance, considering, among other things, forecasts of a number of macro variables. In that analysis, we also compare the forecasts of the different macro variables taking into account that they are measured in different scales. This is done using Theil's U (defined as the usual RMSE divided by the uncentred standard deviation of the series to be forecast). This measure of forecasting accuracy does not only allow us to compare forecasts across series with different scales but also provides us with a benchmark against the naïve no-change forecast (the criterion being that Theil's U is strictly less than unity). The GDP-growth forecasts evaluated in Table 3 all turn out to have a Theil's U well below unity and thus imply that the DFM-based VAR forecasts always outperform the naïve no-change forecast.

3.2. Alternative GDP-growth forecasts

The DFM-based forecasts of GDP growth are compared with various alternative GDP-growth forecasts in Tables 4–6. The comparisons with unfiltered BTS variables appear in Table 4, with macro data in Table 5, and with other popular summary indices in Table 6. These tables have the same basic format: the entries are relative RMSEs computed for various forecast horizons at which the RMSEs of the DFM-based VARs appear in the denominator so that figures greater (smaller) than unity mean that the DFM-based VARs outperform (are outperformed by) the alternative forecasts. The (modified) Diebold–Mariano statistic is used to test whether the differences in forecasting performance are statistically significant.

Tables 4 and 5 have further similarities: in these tables, we compare the RMSEs of the DFM-based VARs with RMSEs of alternative forecasts generated by making use of empirical distributions. These distributions are obtained from the forecasts of all possible q-variate VARs given certain feasible conditioning sets (see the discussion above) and the restriction that $q \leq 5$ (including GDP growth). The first two rows in the upper and lower panels of Tables 4 and 5 construct the relative RMSEs by making use of the medians and means of the RMSE distributions of the alternative forecasts. The interpretation of these figures is thus that they serve as a yardstick for assessing the performance of the DFM-based VARs relative to a typical alternative forecast that would be obtained when using the unfiltered BTS data (i.e., without employing the DFM filter) or when making

Table 5
Alternative forecasts of GDP growth: coincident index and forward-looking index vs. macro variables[a]

	One-quarter	Two-quarter	Four-quarter	Eight-quarter
Coincident index vs. macro variables				
Median	1.27** (1.26)	1.21 (1.20)	0.92 (0.92)	0.75*** (0.76)
Mean	1.28‡ (1.27)	1.21‡ (1.21)	0.93‡ (0.95)	0.76‡ (0.77)
Best one-quarter	1.16* (1.16)	1.09 (1.09)	0.93 (0.93)	0.86** (0.86)
Best two-quarter	1.23* (1.23)	1.08 (1.08)	0.84* (0.84)	0.84*** (0.84)
Best four-quarter	1.31** (1.23)	1.19 (1.08)	0.79** (0.84)	0.82*** (0.84)
Best eight-quarter	1.30** (1.41)	1.26 (1.27)	0.93 (1.10)	0.60*** (0.63)
Forward-looking index vs. macro variables				
Median	1.20* (1.19)	1.24* (1.23)	1.04 (1.05)	0.71* (0.72)
Mean	1.20‡ (1.20)	1.25‡ (1.24)	1.05‡ (1.07)	0.72‡ (0.73)
Best one-quarter	1.10 (1.10)	1.12 (1.12)	1.05 (1.05)	0.81*** (0.81)
Best two-quarter	1.16 (1.16)	1.11 (1.11)	0.95 (0.95)	0.79* (0.79)
Best four-quarter	1.23* (1.16)	1.23 (1.11)	0.89 (0.95)	0.78** (0.79)
Best eight-quarter	1.23* (1.33)	1.29* (1.31)	1.05 (1.25)	0.57** (0.60)

[a] See the notes in Table 4 (except that the empirical RMSE distributions here are generated using macro rather than BTS variables). The share of the models that are excluded when subjected to the residual diagnostics criteria is 53.7%.

Table 6
Alternative forecasts of GDP growth: coincident index and forward-looking index vs. popular summary indices[a]

	One-quarter	Two-quarter	Four-quarter
Coincident index vs. popular summary indices			
BTS confidence indicator, manufacturing	1.20**	1.15	0.94
BTS confidence indicator, construction	1.27**	1.23***	1.03
Consumer survey, personal economy†	1.15*	1.05	0.82*
Consumer survey, whole economy	1.14*	1.08	0.91
Consumer survey, unemployment	1.15**	1.18*	1.04
Consumer survey, past personal economy†	1.18*	1.11	0.89*
Activity index (four-quarter change)	1.17*	1.05	0.98
Forward-looking index vs. popular summary indices			
BTS confidence indicator, manufacturing	1.13	1.18	1.07
BTS confidence indicator, construction	1.19*	1.26**	1.16*
Consumer survey, personal economy†	1.09	1.08*	0.92
Consumer survey, whole economy	1.08	1.11*	1.03
Consumer survey, unemployment	1.09**	1.21***	1.18*
Consumer survey, past personal economy†	1.12	1.14	1.01
Activity index (four-quarter change)	1.11	1.08	1.11**

[a] See the notes in Table 4. † means that the models based on the other popular summary indices do not fulfil the residual diagnostics criteria. All models are bivariate VARs, whose lag lengths have been determined using the system-based BIC.

use of macro variables only. The remaining rows in the upper and lower panels of these tables ("Best x-quarter") substitute the central-tendency moments for the optimised RMSEs at the x-quarter horizon, x=1, 2, 4, 8, and report the relative RMSEs that are obtained when residually using the same model to derive the forecasts at the other horizons. These figures thus allow us to make a comparison with the best possible forecast at a certain horizon that can be derived using the unfiltered BTS variables or the macro variables (and also to see how stable the model used to compute this forecast is across different forecast horizons).

Turning first to the results in Table 4, we can see that the DFM filter generally improves the forecasting performance of the VARs. This holds true especially in the case of the forward-looking variables. For these variables, the DFM-based VARs outperform the rival models at the one- and two-quarter horizons in all cases, even if the models based on the unfiltered variables are optimised (with respect to the particular forecast horizons).

From the results in brackets, we see that the picture is unaltered when conditioning the forecasting models on their in-sample performance. In the case of coincident data, this obtains even though more than 30% of the models are discarded. The finding in previous studies that in-sample performance is largely unrelated to out-of-sample performance is thus confirmed here.

Next, turning to the comparison with macro data (Table 5), we find that the DFM approach again outperforms the rival models at all one- and two-quarter horizons. As before, the results are enhanced for the forward-looking data.[9] As expected, the gains from using macro variables increase with the length of the forecast horizon: at the eight-quarter horizon, the DFM forecasts never give lower RMSEs than the forecasts based on macro variables. However, in several cases, the DFM-based models do surprisingly well even at the four-quarter horizon (see the lower panel in Table 5).

Finally, the results that compare the DFM-based GDP-growth forecasts with the forecasts of GDP growth based on popular summary indices of activity are shown in Table 6. They are qualitatively similar to those in the previous tables: the DFM forecasts dominate at short horizons and the improvement is somewhat larger for forward-looking variables. One particularly interesting feature of the comparisons with the popular summary indices is that the DFM—except in the case of coincident data at the four-quarter horizon—outperforms the so-called activity index of Statistics Sweden by a relatively large margin (see the last row in each panel in Table 6). This is interesting because the activity index is

[9] This statement relates to the fact that the RMSE ratios generally are higher for those data. However, as can be seen from Table 5, the statement does not take formal account of the varying degrees of statistical significance.

explicitly designed to be a short-run indicator of the growth rate of GDP and is used by many professional forecasters. Since both indicators are analysed and published continuously (see http://www.scb.se and http://www.konj.se), future work may make a further contribution by comparing the performance of these indicators in genuine real time.

4. Summary and concluding remarks

In this paper, we examine whether data from business tendency surveys are useful for forecasting GDP growth in the short run. The starting point is a so-called dynamic factor model (DFM), which is used both as a framework for dimension reduction in forecasting and as a procedure for filtering out unimportant idiosyncratic noise in the underlying survey data. In this way, it is possible to model a rather large number of noisy survey variables in a parsimoniously parameterised vector autoregression (VAR).

To assess the forecasting performance of the procedure, comparisons are made with VARs that either use unfiltered survey variables, macro variables only, or other popular summary indices of economic activity. By making comparisons with VARs based on the unfiltered survey data, we are able, in terms of forecast precision, to assess the gain derived from first applying the DFM to the BTS data (relative to not doing so). That is, we can quantify the effects on forecasting by parsimoniously modelling the noise-reduced BTS series rather than the original series themselves. The comparisons with macro VARs instead enable us to judge our performance relative to the "standard" forecasting model. Finally, the comparisons with VARs based on other summary indices of activity allow us to shed some light on the performance of our procedure when holding the gains of dimension reduction constant. Like the DFM procedure, such summary indices have the advantage of enabling the use of very parsimonious forecasting models, without having to give up too much of the relevant forecasting information.

The evaluations are undertaken by subjecting both the DFM-based VARs and the rival models to a recursive out-of-sample forecasting competition. Most of our analyses concern forecasts at the one- and two-quarter horizons but, in some cases, we also investigate performance at slightly longer horizons (one year and two years in the future). The main finding of our paper is that the DFM-based procedure works quite well and outperforms the competing alternatives in most cases. Its performance is particularly striking in the case of forward-looking survey data (firms' expectations regarding economic activity in the next quarter), where it consistently outperforms the rival alternatives at the one- and two-quarter horizons. As expected, the performance of the macro VARs improves as the forecast horizon is prolonged. These VARs almost never outperform the DFM-based VARs at the one- and two-quarter horizons, but generate growth forecasts that are reliably more accurate at the eight-quarter horizon.

The findings in this paper relate to the recent research that reports good forecasting performance results for dynamic factor models, e.g., Forni et al. (2000, 2003) and Stock and Watson (1999, 2002). However, in a recent paper, Stock and Watson (2004) find evidence that simple mean combination forecasts (derived from simple indicator regressions augmented with AR terms) outperform DFM-based forecasts in many cases. The simple mean forecasts are found to work well, although the underlying individual forecasts display substantial instability. Extending our analysis by making comparisons with such mean combination forecasts is an interesting topic for future research. This also applies to forecasting models that are non-linear and restricted using Bayesian priors. To allow for such alternatives would be interesting, especially since we find that many models suffer from problems of parameter instability and large forecast errors at, or around, turning points.

Another possible extension relates to the uncertainty that surrounds our estimated common-factor indices and the data used to generate them. As mentioned previously, the data that underlie the analyses in this paper are in the form of so-called net balances; that is, differences between the shares of firms that have specified an increase and a decrease of a particular economic activity. The main reason for basing our analyses on these quantities is that they are the officially published data in the Swedish BTS. However, it may be the case that the dispersion across the responding firms (which is discarded when constructing the net balances) contains important additional information about changes in economic activity. As it happens, taking this information into account could be another way of mitigating the problem of large forecast errors at, or around, turning points. The simple idea here is that the inter-firm response dispersion is low in periods of stable economic expansion and contraction, but increases as demand and supply conditions become more unstable in the neighbourhood of turning points. In a preliminary analysis, Lindström (1999) finds some evidence that this indeed is the case for the Swedish BTS.[10] We believe that this constitutes another interesting field for future work.

Acknowledgements

We thank two anonymous referees, Michael Clements, David Harvey, Tomas Lindström, Jimmy Miller, Chris Sims, Mattias Villani, seminar participants at the National Institute of Economic Research (NIER), Sveriges Riksbank, and the Ministry of Finance, as well as participants at the 4th Eurostat and DG ECFIN Colloquium on Modern Tools for Business Cycle Analysis (Luxembourg, 20–22 October, 2003) for helpful comments. Most of the research undertaken in this paper was accomplished while Per Jansson was Deputy Director of the Department of Forecasting at the NIER. The views expressed in this paper are those of the authors and do not necessarily reflect those of Sveriges Riksbank or the NIER.

Appendix A. Data appendix

The sources of data are as follows: GDP growth, price inflation (underlying UND1X and headline CPI), wage inflation, unemployment, and employment growth are from Statistics Sweden and the NIER. The short-term interest rate is from IMF Financial Statistics. The long-term interest rate is from OECD Main Economic Indicators and Sveriges Riksbank. The exchange-rate variable is from Sveriges Riksbank and the NIER. The survey data (Swedish Business Tendency Survey, BTS) are from the NIER (see Table 1). The popular summary indices are from Statistics Sweden and the NIER (see Table 6).

CPI inflation, underlying (UND1X) inflation, unemployment, short- and long-term interest rates, and the percentage change of the exchange rate are expressed as quarterly averages. The wage variable is obtained by dividing the wage sum by the number of hours worked. Unemployment is open (official) unemployment in the age group 16–64. The short-term interest rate is a 3-month rate while the long-term interest rate is a 10-year rate. The employment variable is based on the number of hours worked. The exchange-rate variable is the effective rate and computed using the IMF's Total Competitiveness Weights (TCW).

Variables expressed in growth rates are log four-quarter differences. Among the level variables (unemployment and the two interest rates), unemployment is seasonally adjusted. In addition, all BTS variables are seasonally adjusted. The method of seasonal adjustment is TRAMO/SEATS (with automatic BIC-based model selection).

[10] However, see also Bergström (1993b) who argues that the net-balance transformations are not very restrictive. Dasgupta and Lahiri (2003) report results that by and large parallel those of Lindström (1999) for the US NAPM survey.

References

Bergström, R. (1992). The relationship between manufacturing production and different business survey series in Sweden. *Working Paper no. 12, National Institute of Economic Research, Stockholm*.

Bergström, R. (1993a). Quantitative production series compared with qualitative business survey series for five sectors of the Swedish manufacturing industry. *Working Paper no. 30b, National Institute of Economic Research, Stockholm*.

Bergström, R. (1993b). The full trichotomous scale compared with net balances in quantitative business survey data. *Working Paper no. 30a, National Institute of Economic Research, Stockholm*.

Bruno, G., & Malgarini, M. (2002). An indicator of economic sentiment for the Italian economy. *Working Paper no. 28/02, Institute for Studies and Economic Analyses, Rome*.

Camba-Mendez, G., Kapetanios, G., Smith, R., & Weale, M. (1999). An automatic leading indicator of economic activity: Forecasting GDP growth for European countries. *NIESR Discussion Papers no. 149, London*.

Christoffersson, A., Roberts, R., & Eriksson, U. (1992). The relationship between manufacturing and various BTS (business tendency survey) series in Sweden illuminated by frequency and complex demodulate methods. *Working Paper no. 15, National Institute of Economic Research, Stockholm*.

Clements, M., & Hendry, D. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press.

Clements, M., & Hendry, D. (1999). *Forecasting non-stationary economic time series*. Cambridge, MA: MIT Press.

Dasgupta, S., & Lahiri, K. (2003). On the use of dispersion measures from NAPM surveys in business cycle forecasting. *Journal of Forecasting*, *12*, 239–253.

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

European Commission. (2000). Business climate indicator for the Euro area (http://www.europa.eu.int/comm/economy_finance/indicators/businessclimate_en.htm)

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized factor model: Identification and estimation. *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.

Fukuda, S., & Onodera, T. (2001). A new composite index of coincident economic indicators in Japan: How can we improve forecast performances? *International Journal of Forecasting*, *17*, 483–498.

Goldrian, G., Lindbauer, J., & Nerb, G. (2001). Evaluation and development of confidence indicators based on harmonised business and consumer surveys. *EC Economic Paper no. 151, Bruxelles*.

Hansson, J., Jansson, P., & Löf, M. (2003). Business survey data: Do they help in forecasting the macro economy? *Working Paper no. 84, National Institute of Economic Research, Stockholm, and Sveriges Riksbank working paper series no. 151, Sveriges Riksbank, Stockholm*.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

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

Kääntä, P., & Tallbom, C. (1993). Using business survey data for forecasting Swedish quantitative business cycle variables: A Kalman filter approach. *Working Paper no. 35, National Institute of Economic Research, Stockholm*.

Koskinen, L., & Öller, L.-E. (2004). A classifying procedure for signalling turning points. *Journal of Forecasting*, *23*, 197–214.

Lindström, T. (1999). Forecasting business cycle turning points using dispersion measures from survey data: Evidence from Swedish manufacturing. *Manuscript*. *Stockholm*: *Sveriges Riksbank.*

Lindström, T. (2000). Qualitative survey responses and production over the business cycle. *Sveriges Riksbank working paper series no. 116, Sveriges Riksbank, Stockholm*.

Öller, L.-E., & Tallbom, C. (1996). Smooth and timely business cycle indicators for noisy Swedish data. *International Journal of Forecasting*, *12*, 389–402.

Rahiala, M., & Teräsvirta, T. (1993). Business survey data in forecasting the output of Swedish and Finnish metal and engineering industries: A Kalman filter approach. *Journal of Forecasting*, *12*, 255–271.

Stock, J., & Watson, M. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, *4*, 351–393.

Stock, J., & Watson, M. (1991). A probability model of the coincident economic indicators. In K. Lahiri, & G. Moore (Eds.), *Leading economic indicators: New approaches and forecasting records*. Cambridge: Cambridge University Press.

Stock, J., & Watson, M. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*, 293–335.

Stock, J., & Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, *20*, 147–162.

Stock, J., & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*, 403–430.

Biographies: Jesper HANSSON is head of the Division of Macroeconomic Research and Simulations at the NIER. He received his PhD from Lund University in 2001.

Per JANSSON has been Associate Professor of Economics at Uppsala University since 2000. He is also Deputy Director of the Monetary Policy Department at Sveriges Riksbank. He received his PhD from Uppsala University in 1994.

Mårten LÖF, who also is with the NIER, is a statistician and received his PhD from the Stockholm School of Economics in 2001.