# Forecasting industrial production: the role of information and methods

Guido Bulligan,[1] Roberto Golinelli[2] and Giuseppe Parigi[3]

## 1. Introduction

One of the main tasks of the economy watcher is to extract reliable signals from high-frequency indicators to provide the decision-maker with an early picture of the short-term economic situation. The index of industrial production (IPI) is probably the most important and widely analysed high-frequency indicator, given the relevance of manufacturing activity as a driver of the whole business cycle. This can be seen by the extensive comments and reactions of business analysts as soon as the IPI is published. Indeed, the IPI is a crucial variable in the forecasting process of the short-term evolution of GDP in most countries (see Golinelli and Parigi (2007) for an application to the G7 countries).

However, the IPI itself is characterised by a significant publication delay, which limits its usefulness and motivates the great efforts to compute reliable and updated forecasts. The efforts of statistical institutes to shorten the delay of the first release imply a greater degree of revision of the early estimates, which leads to the usual problem of assessing the ability of alternative forecasting methods using real-time data (see, for example, Croushore and Stark (2001, 2002), Diebold and Rudebusch (1991)).

The aim of this paper is to explore the real-time performance of alternative ways of forecasting the monthly dynamics of the Italian IPI, ie different "forecasting methods, which include the models as well as the estimation procedures and the possible choices of estimation windows" (Giacomini and White (2006), p 1549).

Our forecasting methods are defined through combinations of the following three sets of options. First, the degree of model complexity (ie. the amount of information exploited). If randomness is the main feature of the indicator's information content, simpler models may be more suitable. On the other hand, complex models are preferable in order to reduce the noise stemming from the partial information of each indicator. In this case, two options are available: (i) disaggregate models, which entail forecasting errors that might compensate at a more aggregate level (see, for example, Hendry and Hubrich (2007)); and (ii) factor-based models, where a few predictors summarise the information content of a large number of indicators (the so-called "common factors" of the information set; see Stock and Watson (2006)).

Second, the estimation method. We apply both the ordinary least squares (OLS) and, given the disaggregate nature of the models (see point (i) above), the seemingly unrelated regression (SUR) procedures in order to increase the efficiency of parameter estimates by accounting for possible simultaneity of the random shocks to different equations. In the context of factor-based models (see point (ii) above), the choice is between static and dynamic principal components and different ways of selecting the appropriate number of factors.

---

[1] Research Department, Bank of Italy.

[2] Department of Economics, University of Bologna
(e-mail: roberto.golinelli@unibo.it, web page: www2.dse.unibo.it/golinelli/).

[3] Research Department, Bank of Italy.

Third, the length of the estimation window. If too wide a window is chosen it may affect the precision of the estimates (and, therefore, of the forecasts) because of the likely occurrence of breaks (see, for example, Stock and Watson (1996), Clements and Hendry (2002)). As there is no a priori indication of the appropriate length of the estimation window, we consider three cases: a window of seven years (chosen according to the average length of the Italian industrial cycle); a shorter window of four years; and a longer one of more than 10 years.

The relevance of data revisions (see Kozicki (2002) for a discussion) is assessed by comparing the results of our analysis with both real-time and the latest available data.

## 2. Alternative modelling approaches exploiting information sets of different sizes

The problem of extracting reliable signals from high-frequency indicators is not new. Klein and Sojo (1989) suggest two alternative ways of classifying the literature.

According to the "selected indicator model" (SM) approach, the monthly IPI (either aggregate or disaggregate) is regressed on some dynamic terms and a number of pre-selected indicators which are characterised by the same frequency. Out-of-sample SM forecasts are obtained by filling right-hand side explanatory indicators with their values (if known), or with extrapolations where necessary. In the case of the disaggregate SM, the IPI forecasts are obtained by aggregating (in alternative ways) the predictions for each sub-sector.

The "unstructured empirical indicator model" approach can be developed in two ways. First, each of the $n$ indicators in the dataset is used in an autoregressive distributed lag regression (ARDL) and the IPI forecast is obtained from the average of the $n$ forecasts (in our case, $n = 110$). Alternatively, the IPI forecast is computed through the principal components of the $n$ indicators in the dataset (the approximate factor-based model (FM)). This allows not only the information content of the single variables to be exploited but also their covariance, without incurring the "curse of dimensionality" as seen in unrestricted vector autoregressive models (see Stock and Watson (2006) for an updated survey).[4]

The specification of the SM is very similar to that of the bridge model (BM) used to forecast (quarterly) GDP with indicators which are generally available at a higher frequency: the choice of the most suitable indicators depends on several statistical testing procedures as well as on the skill and experience of the researcher (on this point, see Golinelli and Parigi (2007)). In the FM approach, the indicators are, instead, automatically reweighted so that greater weight is given to those variables that are most important in determining the common movements of the whole information set.[5] In other words, the SM approach allows greater flexibility in the specification strategy at the expense of lower automation (see Golinelli and Parigi (2008) for early attempts to automate BM specifications). However, both the SM and the FM require a number of modelling settings (regarding the estimation method or the sample size) which, in turn, imply alternative choices with different effects on the predictive ability.

---

[4] Though originally used to extrapolate cyclical conditions (see Zarnowitz (1992), Altissimo et al (2001)), the FM may also be used to forecast single variables such as GDP or the inflation rate (see Marcellino et al (2003), Cristadoro et al (2005, 2008), Altissimo et al (2007), Schumacher and Breitung (2008), Giannone et al (2008), Angelini et al (2008), Barhoumi et al (2008)).

[5] The claim of neutrality and generality of the FM is questioned by Boivin and Ng (2005, 2006), who stress the relevance of assessing the model forecasting ability of both the size/composition of the dataset and the way in which the factor-based forecasts are formulated.

Table 1

**The models used in this paper**

| | |
|---|---|
| | **A Univariate models** |
| ARIMA | Univariate ARIMA model |
| | **B Dynamic single-equation models with few indicators** |
| ESM | Early single-equation model based on electricity consumption, temperature and trend |
| USM | Updated single-equation model (with only a few more indicators than the ESM) |
| FSM | Final single-equation model (with additional indicators with respect to the USM) |
| | **C Dynamic multiple-equation models disaggregated in sub-sectors**[1] |
| ASGR | Aggregation of sectoral annual growth rates |
| ASLWE | Aggregation of sectoral levels excluding the energy sector |
| ASWL | Aggregation of sectoral levels using the official weights of the statistical agency |
| | **D Unstructured empirical indicator approaches** |
| FA-ARDL | Average of bivariate autoregressive distributed lag model forecasts[2] |
| SW-D | Direct h-step (multi-step) forecasts of the static factor model |
| SW-S | Sequential one-step forecasts of the static factor model |
| FHLR-F | Generalised dynamic factor model with fixed rule to determine the factor number |
| FHLR-O | Generalised dynamic factor model with optimal criteria to determine the factor number |

[1] The sectoral forecasts of these models are aggregated to obtain the IPI. [2] Each ARDL forecasting model uses the information from only one indicator of the whole dataset.

The models used in this paper are listed in Table 1: the SM in points B and C and the ARDL/FM in point D. They are reported according to the historical development of the Italian IPI short-term forecasting analysis, which is characterised chiefly by the use of an increasing number of indicators: from the early selected indicator model (ESM) (single equations with only a few specifically selected indicators, such as electricity consumption, temperature and trend, possibly non linear) to the updated selected indicator model (USM) and, more recently, the final selected indicator model (FSM). With large datasets of timely information, the IPI modelling issue may also be tackled at a more disaggregate level, by specifying equations for different manufacturing sub-sectors, such as those producing consumption, equipment, intermediate and energy goods (sectoral SM). The aggregate IPI predictions are then obtained with three different "aggregator" functions: ASGR, ASLWE and ASWL (Table 1, point C). An even larger number of (timely) time series are exploited when forecasting via the unstructured empirical indicator approach (Table 1, point D), where neither accounting nor economic relationships are postulated between the indicators and the variable to be forecast.

Each SM was obtained by applying the general-to-specific modelling approach to monthly levels of raw data[6] over the period up to January 2003. Besides the motivations of Wallis'

[6] The choice of levels instead of logarithms follows from the results in Marchetti and Parigi (2000) and has been confirmed by pre-processing with the program TRAMO as in Osborn et al (1999). Quantitative variables are mechanically adjusted only for trading day variations: if $x_r$ is the raw variable, the adjustment is given by $x_a = x_r \, td_{base}/td_t$, where $td_{base}$ is the average monthly number of trading days in the base year (2000 in our

seminal work (1974), we have decided to use raw data because they are directly available and avoid filtering problems that may be exacerbated in real-time datasets.[7] The resulting SMs contain a number of unusual data transformations and restrictions, which are suggested by searches over the historical period up to January 2003 and are not influenced by subsequent events.[8]

The forecast error of alternative models can be decomposed into: (i) idiosyncratic elements (such as future random shocks and data revisions) that cannot be forecast; (ii) the misspecification bias, which may be reduced with complex models (ie with many parameters); and (iii) the difference between population and estimated parameters, which is related to the number of parameters to be estimated and the length of the estimation sample. The last two cases are characterised by a double trade-off: (a) complex versus simple models; and (b) long versus short estimation samples. On the one hand, a low number of degrees of freedom due to too many parameters or too few data (or both) may affect the precision of estimates and forecasts; on the other, long samples may be associated with the presence of heterogeneity and structural change.

The models in Table 1 approximate the conditional expectation function with alternative mixes of the trade-offs discussed above. With regard to the simple/complex trade-off, we consider the univariate ARIMA model (Table 1, point A) as a benchmark, the alternative single-equation SM for the aggregate IPI (with only a few indicators, see Table 1, point B), the multiple-equation disaggregate SM (with a large number of indicators, see Table 1, point C) and the unstructured empirical indicator approach based on a very large dataset of indicators (Table 1, point D). The SM parameters are estimated with OLS and SUR techniques.[9] In the case of the FM, we use both static (Stock and Watson (2002a, 2002b)) and dynamic (Forni et al (2000, 2005)) representation estimates. The issue of the length of the estimation sample is dealt with by using rolling estimates with different windows.

The forecast errors obtained from the different forecasting methods, ie the combinations of different models with alternative estimators over rolling samples of different sizes, are compared by using the Giacomini and White test (GW) (2006). Its null hypothesis implies that alternative forecasting methods are equally accurate at the forecast target date using the available information set at the time the forecast is computed.[10] The GW test is suitable for our analysis because it is valid under very general data assumptions (including non-constant data-generating processes, which are common in the context of forecasting with indicators) and for both nested and non-nested models (eg the single-equation SM clearly nests with the ARIMA specifications) estimated with different techniques over different samples, and with both revised and unrevised data.[11]

---

case) and $td_t$ is the number of trading days in month $t$ (see Bodo and Signorini (1987) and the appendix in Bodo et al (1991) for more details).

7 With monthly US series, Ghysels et al (2002), find that monetary policy rules based on raw data have more predictive power than those based on seasonally adjusted data. Swanson and van Djik (2007) note that the seasonal adjustment process (highly nonlinear) weakens the linkage between first available and final data.

8 It is worth observing that the SMs in this paper are different from those currently used at the Bank of Italy to forecast the IPI, since the latter are finely tuned over a more recent period 2001–07 (a sample can be made available on request).

9 The SUR estimates are not reported because they performed worse when compared to the OLS estimates.

10 Under the null hypothesis, the GW test is distributed as a $\chi^2$ with q degrees of freedom, where q is the dimension of the test function. With q = 1, as in our paper, only a constant is considered; with q > 1, other variables are used in order to help distinguish between the forecast performance of the two methods.

11 Other similar tests, such as Diebold and Mariano (1995), are not normally distributed for nested models (see West (1996)) and in the presence of data revisions (on this, see Clark and McCracken (2008)).

## 3. Main results

Table 2 reports the comparison of SM and FM forecasts. Five alternative forecasting methods are shown for each prediction horizon: the ARIMA model; the average of the single-equation SM; the average of the multiple-equation SM; the average of the FM; and the overall average of the SM and FM models. All forecasts are computed with the latest available data, given the unavailability of a real-time dataset for some indicators (specifically two-digit Ateco data for the IPI). The first two columns report the RMSEs and their ratios with respect to the ARIMA model.

The picture is quite clear-cut: short-term information always matters. Both the SM and the FM models always significantly outperform the ARIMA model, suggesting that the short-term indicator signal dominates the noise, independently from the different methods used to extract it. In this context, however, the SM model significantly outperforms the FM model in terms of efficiency: the researcher can increase the amount of signal extracted from the available indicators and improve up to 30–40% the factor-based RMSE model. Though FM models are appealing because they can cope with many variables and capture the business cycle component of the target variable, it appears that they somehow fail to fully anticipate the highly idiosyncratic component which is characteristic of short-term dynamics.

Table 2 also reports the Fair and Shiller (FS) (1990) t-statistics for the null hypothesis that the forecast of the model in the row contains no information relevant to future IPI forecasts not already contained in the model in the column (ie the model in the row is encompassed by the model in the column). According to this test, ARIMA forecasts are generally encompassed by all models based on indicators: the FS t-statistics in the ARIMA row are never significant, contrary to those in the ARIMA column. The parsimonious use of indicators leads SMs to outperform all other forecasting approaches, as their FS t-statistics always reject the null hypothesis against all other forecasts. Among alternative SMs, the multiple-equation approach performs best in terms of the FS test: the parsimonious (ie with restrictions) exploitation of 30 indicators allows the IPI predictions to contain all relevant information.

Overall, the findings in Table 2 lend support to the superiority of the SM approach in terms of forecasting performance. This seems at odds with the results reported in similar recent papers by Angelini et al (2008) and Barhoumi et al (2008), which show that FMs outperform bridge models (BM) in predicting the euro area GDP short-term evolution (only at aggregate level in the former paper, also by country in the latter). The exercises performed in those papers are, however, fairly different from the one presented here. Barhoumi et al define the BM as a large number of bivariate regressions whose forecasts are averaged in a similar way to what we have defined as ARDL forecasts, and which we have proved perform less well.

Overall, “horse races” and comparisons of different forecasting methods lead to a better understanding of the advantages and disadvantages of the alternative approaches. BM/SM models generally provide very precise forecasts which are also very easy to interpret. Indicators that appear to be unrelated or only loosely linked to the target variable are ignored. This has two positive implications. First, SM/BM predictions can “tell the story” of the forecast on the basis of the evolution of the explanatory indicators. This is a very important feature in periods characterised by deep and rapid changes, when it is not only necessary to quantify the relevance of specific events, but also to understand their origin (recent advances in this topic in the field of FM are shown in Banbura and Runstler (2007)). Second, SM/BM datasets are smaller and less costly to update. The claim that all relevant information is used in FMs because nothing is a priori discarded implies that their datasets are very large and include indicators from many sources, with very different characteristics and quality standards.

Table 2

**Comparison of alternative forecasting approaches**

| | | | | Fair and Shiller (1990) test outcomes[3] | | | | |
|---|---|---|---|---|---|---|---|---|
| **One month ahead** | **RMSE[1]** | **Ratio[2]** | | **ARIMA** | **SM single** | **SM multiple** | **FM (SHIFT)** | **Avg (excl ARIMA)** |
| ARIMA | 1.39 | 1.000 | | – | 0.27 | 0.15 | 1.49 | 0.83 |
| SM single-equation | 0.57 | 0.408 | *** | 7.04 | – | 1.30 | 7.33 | 5.15 |
| SM multiple-equation | 0.53 | 0.381 | *** | 8.79 | 3.23 | – | 9.52 | 6.10 |
| FM (SHIFT) | 0.91 | 0.658 | ** | 2.04 | –0.56 | –0.32 | – | –7.80 |
| Average (excl ARIMA) | 0.71 | 0.510 | *** | 3.65 | –0.03 | –0.12 | 10.30 | – |
| **Two months ahead** | | | | | | | | |
| ARIMA | 1.88 | 1.000 | | – | 1.08 | 0.93 | 1.39 | 1.18 |
| SM single-equation | 1.10 | 0.582 | *** | 3.65 | – | –0.74 | 2.48 | 1.54 |
| SM multiple-equation | 0.99 | 0.524 | *** | 5.67 | 2.76 | – | 4.12 | 3.14 |
| FM (SHIFT) | 1.30 | 0.689 | ** | 1.48 | –0.01 | –0.48 | – | –2.42 |
| Average (excl ARIMA) | 1.14 | 0.604 | ** | 2.60 | 0.31 | –0.56 | 3.30 | – |
| **Three months ahead** | | | | | | | | |
| ARIMA | 2.09 | 1.000 | | – | 0.72 | 0.46 | 1.09 | 0.96 |
| SM single-equation | 1.45 | 0.691 | ** | 2.77 | – | –1.41 | 2.23 | 1.48 |
| SM multiple-equation | 1.24 | 0.595 | *** | 4.23 | 3.28 | – | 3.44 | 2.93 |
| FM (SHIFT) | 1.58 | 0.754 | * | 1.37 | –0.13 | –0.79 | – | –2.21 |
| Average (excl ARIMA) | 1.41 | 0.674 | ** | 2.15 | 0.19 | –0.94 | 2.82 | – |

[1] Root mean squared forecasting error of the seasonally adjusted forecast growth rates with respect to the previous month. [2] Ratios of the RMSE with respect to the ARIMA model (*, ** and *** reject at 10%, 5% and 1% the null of the GW test for equal predictive ability). [3] t-statistics of the estimates of $\beta_R$ and $\beta_C$ parameters in the regression using White (1980) standard errors: $\frac{y_t - y_{t-h}}{y_{t-h}} = \alpha + \beta_R \frac{\hat{y}_{R,t} - y_{t-h}}{y_{t-h}} + \beta_C \frac{\hat{y}_{C,t} - y_{t-h}}{y_{t-h}}$, where $\hat{y}_{R,t}$ and $\hat{y}_{C,t}$ are the forecasts of the two models being compared (and respectively listed along the rows and the columns), and $h$ is the forecast horizon ($h$ = one, two, three months ahead). The null hypothesis is that the forecast of model $R$ (in the row) contains no information relevant to the IPI forecast not in model $C$ (in the column). The results are robust to the use of Newey-West t-statistics.

However, the construction of SMs/BMs is more difficult and arbitrary and requires a number of subjective choices – entailing crucial trade-offs – about the model specification and the size of the estimation sample. We have shown that the IPI forecasts from multiple-equation models are often significantly better than those from single-equation models, especially at longer forecast horizons, suggesting the likely presence of some leading indicators in the information set used by multiple equations. Contrary to the FM, the SM forecasting performance appears to be more dependent on the size of the estimation sample, confirming the results in the literature about the greater stability of FM forecasts.

Both the SM and FM approaches appear to be complementary, as the strengths of one correspond to the weaknesses of the other. Factor-forecasting performance is less efficient because it cannot pre-select the “best” indicators from large datasets and is less interpretable. However, this weakness reduces the risk of omitting important predictors, allows new information to be exploited as soon as it becomes available, prevents uncertainty about the modeller’s skill/experience and delivers forecasts that are less prone to regime-shift biases. Thus, the challenge for future research is to find out how the pros and cons of the two approaches may be fruitfully merged in forecasting practice.

## References


Altissimo, F, A Bassanetti, R Cristadoro, M Forni, M Hallin, M Lippi, L Reichlin and G Veronese (2001): “Eurocoin: a real time coincident indicator of the euro area business cycle”, *CEPR Discussion Paper*, no 3108.

Altissimo, F, R Cristadoro, M Forni, M Lippi and G Veronese (2007): “New eurocoin: tracking economic growth in real time”, *Temi di Discussione*, no 631, Banca d’Italia and forthcoming in *Review of Economics and Statistics*.

Angelini, E, G Camba-Mendez, D Giannone, L Reichlin and G Runstler (2008): “Short-term forecasts of the euro area GDP growth”, *CEPR Discussion Paper*, no 6746.

Banbura, M and G Runstler (2007): “A look into the factor model black box – publication lags and the role of hard and soft data in forecasting GDP”, *Working Paper*, no 751, European Central Bank.

Barhoumi, K, S Benk, R Cristadoro, A den Reijer, A Jakaitiene, P Jelonek, A Rua, G Runstler, K Ruth, and C van Nieuwenhuyze (2008): “Short-term forecasting of GDP using large monthly datasets: a pseudo real-time forecast evaluation exercise”, *Occasional Paper Series*, no 84, European Central Bank.

Bodo, G, A Cividini and L F Signorini (1991): “Forecasting the industrial production index in real time”, *Journal of Forecasting*, vol 10, pp 285–99.

Bodo, G, R Golinelli and G Parigi (2000): “Forecasting industrial production in the euro area”, *Empirical Economics*, vol 25, pp 541–61.

Bodo, G and L F Signorini (1987): “Short-term forecasting of the industrial production index”, *International Journal of Forecasting*, vol 3, pp 245–59.

Boivin, J and S Ng (2005): “Understanding and comparing factor-based forecasts”, *International Journal of Central Banking*, vol 1, no 3, pp 117–51.

——— (2006): “Are more data always better for factor analysis?”, *Journal of Econometrics*, vol 132, pp 169–94.

Clark, T E and M W McCracken (2008): “Test of equal predictive ability with real-time data”, *Working Paper Series*, no 029A, Federal Reserve Bank of St Louis.

Clements, M P and D F Hendry (2002): “Modelling methodology and forecast failure”, *Econometrics Journal*, vol 5, pp 319–44.

Corradi, V, A Fernandez and N R Swanson (2008): “Information in the revision process of real-time datasets”, *Journal of Business and Economic Statistics*, forthcoming.

Cristadoro, R, M Forni, L Reichlin and G Veronese (2005): “A core inflation indicator for the euro area”, *Journal of Money, Credit, and Banking*, vol 37, pp 539–60.

Cristadoro, R, F Venditti and G Saporito (2008): “Forecasting inflation and tracking monetary policy in the euro area: does national information help?”, *Temi di Discussione*, no 677, Banca d’Italia.

Croushore, D and T Stark (2001): “A real-time data set for macroeconomists”, *Journal of Econometrics*, vol 105, pp 111–30.

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

Diebold, F X and G D Rudebusch (1991): “Forecasting output with the composite leading index: a real time analysis”, *Journal of the American Statistical Association*, vol 86, pp 603–10.

Fair, R C and R J Shiller (1990): “Comparing information in forecasts from econometric models”, *The American Economic Review*, vol 80, no 3, pp 375–89.

Forni, M, M Hallin, M Lippi and L Reichlin (2000): “The generalized dynamic factor model: identification and estimation”, *Review of Economics and Statistics*, vol 82, no 4, pp 540–54.

——— (2005): “The generalized factor model: one sided estimation and forecasting”, *Journal of the American Statistical Association*, vol 100, no 471, pp 830–40.

Ghysels, E, N R Swanson and M Callan (2002): “Monetary policy rules with model and data uncertainty”, *Southern Economic Journal*, vol 69, no 2, pp 239–65.

Giacomini, R and H White (2006): “Tests of conditional predictive ability”, *Econometrica*, vol 74, pp 1545–78.

Giannone, D L Reichlin and D Small (2008): “Nowcasting: the real-time informational content of macroeconomic data”, *Journal of Monetary Economics*, vol 55, pp 665–76.

Golinelli, R and G Parigi (2007): “The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries”, *Journal of Forecasting*, vol 26, pp 77–94.

——— (2008): “Real-time squared: a real time data set for real-time GDP forecasting”, *International Journal of Forecasting*, vol 24, pp 368–85.

Hendry, D F and K Hubrich (2007): “Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate”, mimeo.

Klein, L R and E Sojo (1989): “Combinations of high and low frequency data in macroeconometric models”, in L R Klein and J Marquez (eds), *Economics in theory and practice: an eclectic approach*, Kluwer Academic Publishers, pp 3–16.

Kozicki, S (2002): “Comments on forecasting with a real-time data set for macroeconomists”, *Journal of Macroeconomics*, vol 24, pp 541–57.

Marcellino, M, J H Stock and M W Watson (2003): “Macroeconomic forecasting in the euro area: country specific versus area-wide information”, *European Economic Review*, vol 47, pp 1–18.

Marchetti, D J and G Parigi (2000): “Energy consumption, survey data and the prediction of industrial production in Italy: a comparison and combination of different models”, *Journal of Forecasting*, vol 19, pp 419–40.

Osborn, D R, S Heravi and C R Birchenhall (1999): “Seasonal unit roots and forecasts of two-digit European industrial production”, *International Journal of Forecasting*, vol 15, pp 27–47.

Schumacher, C and J Breitung (2008): “Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data”, *International Journal of Forecasting*, vol 24, pp 386–98.

Stark, T and D Croushore (2002): “Forecasting with a real-time data set for macroeconomists”, *Journal of Macroeconomics*, vol 24, no 4, pp 507–68.

Stock, J H and M W Watson (1996): “Evidence on structural instability in macroeconomic time series relations”, *Journal of Business Economics and Statistics*, vol 14, pp 11–30.

——— (2002a):, "Macroeconomic forecasting using diffusion indexes", *Journal of Business and Economic Statistics*, vol 20, pp 147–62.

——— (2002b): "Forecasting using principal components from a large number of predictors", *Journal of the American Statistical Association*, vol 97, pp 1167–79.

——— (2006): "Forecasting with many predictors", in G Elliott, C W J Granger and A Timmermann (eds), *Handbook of economic forecasting*, North Holland, Chapter 10, pp 515–54.

Swanson, N R and D van Dijk (2007): "Are statistical agencies getting it right? Data rationality and business cycle asymmetry", *Journal of Business and Economic Statistics*, vol 24, no 1, pp 24–42.

Wallis, K F (1974): "Seasonal adjustment and relations between variables", *Journal of the American Statistical Association*, vol 69, pp 18–31.

West, K D (2006): "Forecast evaluation", in G Elliott, C W J Granger and A Timmermann (eds), *Handbook of economic forecasting*, North Holland, pp 99–134.

White, H (1980): "A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity", *Econometrica*, vol 48, pp 817–38.

Zarnowitz, V (1992): *Business cycles*, University of Chicago Press.