Forecasting by factors, by variables, by both, or neither?

Jennifer L. Castle[†], Michael P. Clements[♦] and David F. Hendry[*]
[†]Magdalen College and Institute for New Economic Thinking at the
Oxford Martin School, University of Oxford, UK
[♦]Economics Department, Warwick University, UK
[⋆]Economics Department and Institute for New Economic Thinking at the
Oxford Martin School, University of Oxford, UK

June 29, 2011

Abstract

We forecast US GDP and inflation over 1-, 4- and 8-step horizons using the dataset from Stock and Watson (2009), with factors, variables, both, and neither. *Autometrics* handles perfect collinearity and more regressors than observations, enabling all principal components and variables to be included for model selection, jointly with using impulse-indicator saturation (IIS) for multiple breaks. Empirically, factor models are more useful for 1-step ahead forecasts than at longer horizons, when selecting over variables tends to be better. Accounting for in-sample breaks and outliers using IIS is useful. Recursive updating helps, but recursive selection leads to greater variability, and neither outperforms autoregressions.

JEL classifications: C51, C22.
Keywords: Model selection; Factor models; Forecasting; Impulse-indicator saturation; *Autometrics*

1 Introduction and historical background

There are three venerable traditions in economic forecasting based respectively on economic-theory derived empirical econometric models, 'indicator' or 'factor' approaches combining many sources of information, and mechanistic approaches.[1]

Members of the first group are exemplified by early models like Smith (1927, 1929) and Tinbergen (1930), smaller systems in the immediate post-war period (such as Klein, 1950, Tinbergen, 1951, Klein, Ball, Hazlewood and Vandome, 1961), leading onto large macro-econometric models (Duesenberry, Fromm, Klein and Kuh, 1969, and Fair, 1970, with a survey in Wallis, 1989), and now including both dynamic stochastic general equilibrium (DSGE) models widely used at Central Banks (see e.g, Smets and Wouters, 2003), and global models, first developed by project Link (see e.g., Waelbroeck, 1976) and more recently, global vector autoregressions (GVARs: see Dees, di Mauro, Pesaran and Smith, 2007, Pesaran, Schuerman and Smith, 2009, and Ericsson, 2010).

[*]This research was supported in part by grants from the Open Society Institute and the Oxford Martin School. Contact details: jennifer.castle@magd.ox.ac.uk, m.p.Clements@warwick.ac.uk and david.hendry@nuffield.ox.ac.uk.

[1]It is a great pleasure to contribute a paper on economic forecasting to a *Festschrift* in honor of Professor Hashem Pesaran, who has made so many substantive contributions to this important topic. DFH has known Hashem for nearly 40 years, and has benefited in countless ways from his vast range of publications on almost every conceivable topic in econometrics, theory and applied. That output has received nearly 20,000 citations, achieved while first creating then editing the *Journal of Applied Econometrics* since its foundation in 1986. We all look forward to many more salient insights from him.

The second approach commenced with the ABC curves of Persons (1924), followed by leading indicators as in Zarnowitz and Boschan (1977) with critiques in Diebold and Rudebusch (1991) and Emerson and Hendry (1996). Factor analytic and principal component methods have a long history in statistics and psychology (see e.g., Spearman, 1927, Cattell, 1952, Anderson, 1958, Lawley and Maxwell, 1963, Joreskog, 1967, and Bartholomew, 1987) and have seen some distinguished applications in economics (e.g., Stone, 1947, for an early macroeconomic application; and Gorman, 1956, for a microeconomic one). Diffusion indices and factor models are now quite widely used for economic forecasting: see e.g., Stock and Watson (1989, 1999, 2009), Forni, Hallin, Lippi and Reichlin (2000), Peña and Poncela (2004, 2006), and Schumacher and Breitung (2008).

The third set includes methods like exponentially weighted moving averages, denoted EWMA, the closely related Holt–Winters approach (see Holt, 1957, and Winters, 1960), damped trend (see e.g., Fildes, 1992), and autoregressions, including the general time-series approach in Box and Jenkins (1970). Some members of this class were often found to dominate in forecasting competitions, such as Makridakis, Andersen, Carbone, Fildes *et al.* (1982) and Makridakis and Hibon (2000).

Until recently, while the first two approaches often compared their forecasts with various 'naive' methods selected from the third group, there was little direct comparison between them, and almost no studies included both. Here, we consider the reasons for that lacuna, and explain how it can be remedied.

The structure of the paper is as follows. Section 2 describes some of the issues that are likely to bear on the topic of this paper, including the role of measurement errors. Section 3 relates the 'external' variables, denoted $\{\mathbf{z}_t\}$, to factors $\{\mathbf{f}_t\}$. Section 4 compares variable-based and factor-based models. Section 5 develops the analysis of forecasting from factor models with a taxonomy of sources of forecast error in the empirically relevant case of non-stationary processes. Section 6 addresses the problem of systematic forecast failure to which equilibrium-correction formulations are prone in the face of location shifts. Section 7 discusses model selection with both factors and variables, and section 8 illustrates the analysis using US GDP and inflation forecasts. Section 9 concludes.

2 Setting the scene

There are a number of interacting issues that need to be addressed in an analysis of forecasting, whatever device is used. The complexity of these issues, and the way they interact, means that an answer to the question in the title of this paper 'forecasting by factors, by variables, both, or neither?' is likely to be context specific. Even though general guidelines might prove hard to come by, it is fruitful to consider these issues and how they affect our research question. We consider eight aspects: (i) the pooling of both variables and factors in forecasting models; (ii) the role of in-sample model selection in that setting; (iii) whether or not breaks over the forecast horizon are unanticipated; (iv) the role of more versus less information in forecasting; (v) the type of forecasting model being used, specifically whether or not it is an equilibrium-correction mechanism (EqCM); (vi) measurement errors in the data, especially near the forecast origin; (vii) how to evaluate the 'success or failure' of forecasts; (viii) the nature of the DGP itself. We briefly consider these in turn.

2.1 Pooling of information

Factor models are a way of forecasting using a large number of predictors, as opposed to pooling over the forecasts of a large number of simple, often single-predictor, models. When there are many variables in the set from which factors are formed (the 'external' variables), including both sets will often result in the number of candidate variables, N, being larger than the sample size, T. This problem may have seemed insurmountable in the past, but now is not. Let $\mathbf{z}_t$ denote the set of n 'external' variables' from which the factors $\mathbf{f}_t = \mathbf{B}\mathbf{z}_t$ (say) are formed, then $\mathbf{f}_t, \ldots \mathbf{f}_{t-s}$, $\mathbf{z}_t, \ldots \mathbf{z}_{t-s}$ comprise the initial set of candidate variables. Automatic model selection can use multi-path searches to eliminate irrelevant variables by exploring all feasible paths with mixtures of expanding and contracting block searches, so can handle settings with both perfect collinearity and $N > T$, as shown in Hendry and Krolzig (2005) and Doornik (2009b). The simulations in Castle, Doornik and Hendry (2011a) show the feasibility of such an approach when $N > T$ in linear dynamic models. Hence we are not forced at the outset to allow only a small number of factors, or just the factors and a few lags of the variable being forecast, say, as candidates. When the number of candidate variables exceeds the sample size, model selection is unavoidable, so we consider that issue next.

2.2 Model selection

The 'general-to-specific' (*Gets*) search algorithm in *Autometrics* within *PcGive* (see Doornik, 2007, 2009a, and Hendry and Doornik, 2009) seeks the local data generating process (denoted LDGP), namely the DGP for the set of variables under consideration (see e.g., Hendry, 2009) by formulating a general unrestricted model (GUM) that nests the LDGP, and checking its congruence when feasible (estimable once $N \ll T$ and perfect collinearities are removed). Search thereafter ensures congruence, so all selected models are valid restrictions of the GUM, and should parsimoniously encompass the feasible GUM. Location shifts are removed in-sample by impulse-indicator saturation (IIS: see Hendry, Johansen and Santos, 2008, Johansen and Nielsen, 2009, and the simulation studies in Castle, Doornik and Hendry, 2011c), which also addresses possible outliers. Thus, if $\left\{1_{\{j=t\}}, t = 1, \ldots, T\right\}$ denotes the complete set of T impulse indicators, we allow for $\mathbf{f}_t, \ldots \mathbf{f}_{t-s}$, $\mathbf{z}_t, \ldots \mathbf{z}_{t-s}$ and $\left\{1_{\{j=t\}}, t = 1, \ldots, T\right\}$ all being included in the initial set of candidate variables to which multi-path search is applied, so $N > T$ will always occur when IIS is used. The in-sample feasibility of this approach is established in Castle, Doornik and Hendry (2011b). Here we are concerned with the application of models selected in this way to a forecasting context when the DGP is non-stationary due to structural breaks. Since there seem to be few analyses of how well a factor forecasting approach would then perform (see however, Stock and Watson, 2009, and Corradi and Swanson, 2011), we explore its behavior facing location shifts at the forecast origin.

2.3 Unanticipated location shifts

Third, *ex ante* forecasting is fundamentally different from *ex post* modeling when unanticipated location shifts can occur. Breaks can always be modeled after the event (at worst by indicator variables), but will cause forecast failure when not anticipated. Clements and Hendry (1998, 1999) proposed a general theory of economic forecasting for a world of structural breaks using mis-specified models, and emphasized that it had radically different implications from a forecasting theory based on stationarity and well-specified models (as in Klein, 1971, say). Moreover, those authors also establish that breaks other than location shifts are less pernicious for forecasting (though not for policy analyses). Pesaran and Timmermann (2005) and Pesaran, Pettenuzzo and Timmermann (2006) consider forecasting time series subject to multiple structural breaks, and Pesaran and Timmermann (2007) examine the use of moving windows in that context. Castle, Fawcett and Hendry (2010, 2011) investigate how breaks themselves might be forecast, and if not, how to forecast during breaks, but draw somewhat pessimistic conclusions due to the limited information that will be available at the time any location shift occurs. Thus, we focus the analysis on the impacts of unanticipated location shifts in factor-based forecasting models.

2.4 Role of information in forecasting

Fourth, factor models can be interpreted as a particular form of 'pooling of information', in contrast to the 'pooling of forecasts' literature discussed in (e.g.) Hendry and Clements (2004). Pooling information ought to dominate pooling forecasts, each of which is based on limited information, except when all variables are orthogonal (see e.g, Granger, 1989). However, the taxonomy of forecast errors in Clements and Hendry (2005b) suggests that incomplete information by itself is unlikely to play a key role in forecast failure (except if that information would forecast breaks). Consequently, using large amounts of data may not correct one of the main problems confronting forecasters, namely location shifts, unless that additional information is directly pertinent to forecasting breaks. Moreover, although we use *Gets* model selection from a very general initial candidate set, embedding available theory specifications, combined with congruence as a basis for econometric modeling, when forecasting facing location shifts, it cannot be that causal models will dominate non-causal (see Clements and Hendry, 1998) nor that congruent modeling helps (see e.g., Allen and Fildes, 2001). Conversely, Makridakis and Hibon (2000) conclude that parsimonious models do best in forecasting competitions, but Clements and Hendry (2001) argue that such findings are conflated with robustness to location shifts because most of the parsimonious models evaluated also happened to be relatively robust to location shifts compared to their non-parsimonious contenders.[2] Since more information cannot lower predictability, and omitting crucial explanatory variables will both bias parameter estimates and lead to an inferior fit, the jury remains out on the benefits of more versus less information when forecasting.

2.5 Equilibrium-correcting behavior

Fifth, factor models are often equilibrium correction in form, so they suffer from the general non-robustness to location shifts of that class of model. However, the principles of robust-model formulation discussed in Clements and Hendry (2005b) apply, and any equilibrium-correction system, whether based on variables or factors (or both), could be differenced prior to forecasting, thereby embedding the resulting model in a second-differenced forecasting device. Castle *et al.* (2010) show that how a given model is used in the forecast period matters, and explore various transformations that reduce systematic forecast failure after location shifts. Section 6 provides a more extensive discussion.

2.6 Measurement errors

Sixth, many of the 'solutions' to systematic forecast failure induced by location shifts exacerbate the adverse effects of data measurement errors near the forecast origin: for example, differencing doubles their impact. Conversely, averaging mitigates the effects of random measurement errors, so as one method of averaging over variables, factors might help mitigate data errors. Forecasting models which explicitly account for data revisions offer an alternative solution. These include modeling the different vintage estimates of a given time-observation as a vector autoregression (see, e.g., Garratt, Lee, Mise and Shields, 2008, 2009, and Hecq and Jacobs, 2009, following on from Patterson, 1995, 2003), as well as the approach of Kishor and Koenig (2010) (building on earlier contributions by Howrey, 1978, 1984, and Sargent, 1989). For the latter, a VAR is estimated on post-revision data, necessitating stopping the estimation sample short of the forecast origin, and the model forecasts of the periods up to the origin are combined with lightly-revised data for these periods via the Kalman filter to obtain post-revision estimates. The forecast is then conditioned on these estimates of what the revised values of the latest data will be. Clements and Galvão (2011) provide some evidence on the efficacy of these strategies for forecasting US output growth and inflation, albeit using information sets consisting only of lags (and different vintage estimates) of the variable being forecast.

[2] Parsimonious models need not be robust–to see that the two characteristics are distinct, consider using as the forecast an estimate of the unconditional mean of the process to date. No model specification/selection/estimation is required, apart from the calculation of a sample mean, suggesting a simple forecasting device, which is nevertheless highly susceptible to location shifts.

The frequency of macroeconomic data can also affect its accuracy, as can nowcasting (see e.g., Castle, Fawcett and Hendry, 2009, and Bánbura, Giannone and Reichlin, 2011) and 'real time' (versus *ex post*) forecasting (on the latter, see e.g., Croushore, 2006, and Clements and Galvão, 2008). Empirical evidence suggests that the magnitudes of data measurement errors are larger in the most recent data, in other words, in the data on which the forecast is being conditioned (hence the Kishor and Koenig, 2010, idea of stopping the model estimation period early, and attempting to predict the 'final' estimates of the most recent data), as well as during turbulent periods (Swanson and van Dijk, 2006), which might favour factor models over other approaches that do not explicitly attempt to take data revisions into account.

2.7 Forecast evaluation

Next, there is a vast literature on how to evaluate the 'success or failure' of forecasts (see among many others, Leitch and Tanner, 1991, Pesaran and Timmermann, 1992, Clements and Hendry, 1993a, Granger and Pesaran, 2000a, 2000b, Pesaran and Skouras, 2002), as well as using forecasts to evaluate models (see e.g., West, 1996, West and McCracken, 1998, Hansen and Timmermann, 2011, with a sceptical view in Castle and Hendry, 2011b), forecasting *methods* (Giacomini and White, 2006), and economic theory (Clements and Hendry, 2005a). As a first exercise in forecasting from models selected from both variables and factors, below we just report descriptive statistics of forecast performance.

2.8 Nature of the DGP

Finally, the nature of the DGP itself matters greatly to the success of a specific forecasting model or method. In particular, the factor model would be expected to do well if the 'basic' driving forces are primarily factors, in the sense that a few factors account for a large part of the variance of the variables of interest. The ideal case for factor model forecasting is where the DGP is:

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\Upsilon}(L)\,\mathbf{f}_t + \mathbf{e}_t \\ \mathbf{f}_t &= \boldsymbol{\Phi}(L)\,\mathbf{f}_{t-1} + \boldsymbol{\eta}_t \end{aligned}$$

where $\mathbf{x}_t$ is $n \times 1$, $\mathbf{f}_t$ is $m \times 1$, $\boldsymbol{\Upsilon}(L)$ and $\boldsymbol{\Phi}(L)$ are $n \times m$ and $m \times m$, and $n \gg m$ so that the low-dimensional $\mathbf{f}_t$ drives the co-movements of the high-dimensional $\mathbf{x}_t$. The latent factors are assumed here to have a VAR representation. Suppose in addition that the mean-zero 'idiosyncratic' errors $\mathbf{e}_t$ satisfy $\mathsf{E}[e_{i,t}e_{j,t-k}] = 0$ all k unless $i = j$ (allowing the individual errors to be serially correlated), and that $\mathsf{E}[\boldsymbol{\eta}_t\mathbf{e}_{t-k}] = \mathbf{0}$ for all k.

It then follows that given the $\mathbf{f}_t$, each variable in $\mathbf{x}_t$, say $x_{i,t}$, can be optimally forecast using only the $\mathbf{f}_t$ and lags of $x_{i,t}$ ($x_{i,t-1}$, $x_{i,t-2}$ etc). If we let $\boldsymbol{\lambda}_i(L)'$ denote the i^{th} row of $\boldsymbol{\Upsilon}(L)$, then:

$$\begin{aligned} \mathsf{E}_t\left[x_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] &= \mathsf{E}_t\left[\boldsymbol{\lambda}_i(L)'\,\mathbf{f}_{t+1} + e_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\ &= \mathsf{E}_t\left[\boldsymbol{\lambda}_i(L)'\,\mathbf{f}_{t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\ &+ \mathsf{E}_t\left[e_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\ &= \mathsf{E}_t\left[\boldsymbol{\lambda}_i(L)'\,\mathbf{f}_{t+1} \mid \mathbf{f}_t, \mathbf{f}_{t-1}, \ldots\right] + \mathsf{E}_t\left[e_{i,t+1} \mid e_{i,t}, e_{i,t-1} \ldots\right] \\ &= \boldsymbol{\gamma}(L)'\,\mathbf{f}_t + \delta(L)\,x_{i,t} \end{aligned}$$

under the assumptions we have made (see Stock and Watson, 2011, for a detailed discussion). Absent structural breaks, the model with the appropriate factors and lags of x_i would deliver the best forecasts (in population: ignoring parameter estimation uncertainty). The results of Faust and Wright (2007), among others, suggest that the factor structure may not be a particularly good representation of the macroeconomy. Our empirical approach allows that the 'basic' driving forces may be variables or factors, and that there may be non-linearities (captured by linear approximations), as well as the many possible non-stationarities noted above. We assume the DGP originates in the space of variables, with factors being potentially convenient approximations that parsimoniously capture linear combinations of effects. Although non-linearity can be tackled explicitly along with all the other complications (see e.g., Castle and Hendry, 2011a), we only analyze linear DGPs here.

Thus, we consider forecasting from linear models selected in-sample from (a) a large set of variables, (b) over those variables' principal components (PCs), and (c) over a candidate set including both, in each case with IIS, so the initial model will necessarily have $N > T$, and in the third case will be perfectly collinear. We exploit the ability of automatic model selection to operate successfully in such a setting, as well as to select despite more candidate variables than observations.

3 Relating 'external' variables to factors

Consider a vector of n stochastic variables $\{\mathbf{z}_t\}$ that are weakly stationary over $t = 1, \ldots, T$. For specificity, we assume that $\mathbf{z}_t$ is generated by a first-order vector autoregression (VAR) with deterministic term $\boldsymbol{\pi}$:

$$\mathbf{z}_t = \boldsymbol{\pi} + \boldsymbol{\Pi}\mathbf{z}_{t-1} + \mathbf{v}_t \tag{1}$$

where $\boldsymbol{\Pi}$ has all its eigenvalues inside the unit circle, and $\mathbf{v}_t \sim \mathsf{IN}_n[\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{v}}]$, where $n < T$. From (1):

$$\mathsf{E}[\mathbf{z}_t] = \boldsymbol{\pi} + \boldsymbol{\Pi}\mathsf{E}[\mathbf{z}_{t-1}] = \boldsymbol{\pi} + \boldsymbol{\Pi}\boldsymbol{\mu} = \boldsymbol{\mu}$$

where $\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Pi})^{-1}\boldsymbol{\pi}$. The principal-component description of $\mathbf{z}_t$ is:

$$\mathbf{z}_t = \boldsymbol{\Psi}\mathbf{f}_t + \mathbf{e}_t \tag{2}$$

where $\mathbf{f}_t \sim \mathsf{ID}_m[\boldsymbol{\kappa}, \mathbf{P}]$ is a latent vector of dimension $m \leq n$, so $\boldsymbol{\Psi}$ is $n \times m$, with $\mathbf{e}_t \sim \mathsf{ID}_n[\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{e}}]$, $\mathsf{E}[\mathbf{f}_t\mathbf{e}'_t] = \mathbf{0}$ and $\mathsf{E}[\mathbf{e}_t\mathbf{e}'_t] = \boldsymbol{\Omega}_{\mathbf{e}}$. When $\mathsf{E}[\mathbf{f}_t] = \boldsymbol{\kappa}$ and $\mathsf{E}[\mathbf{e}_t] = \mathbf{0}$, under weak stationarity from (2):

$$\mathsf{E}[\mathbf{z}_t] = \boldsymbol{\Psi}\mathsf{E}[\mathbf{f}_t] + \mathsf{E}[\mathbf{e}_t] = \boldsymbol{\Psi}\boldsymbol{\kappa} = \boldsymbol{\mu} \tag{3}$$

Then:

$$\begin{aligned}\mathsf{E}\left[(\mathbf{z}_t - \boldsymbol{\mu})(\mathbf{z}_t - \boldsymbol{\mu})'\right] &= \boldsymbol{\Psi}\mathsf{E}\left[(\mathbf{f}_t - \boldsymbol{\kappa})(\mathbf{f}_t - \boldsymbol{\kappa})'\right]\boldsymbol{\Psi}' + \mathsf{E}\left[\mathbf{e}_t\mathbf{e}'_t\right] \\ &= \boldsymbol{\Psi}\mathbf{P}\boldsymbol{\Psi}' + \boldsymbol{\Omega}_{\mathbf{e}} = \mathbf{M}\end{aligned} \tag{4}$$

say, where $\mathbf{P}$ is an $m \times m$ diagonal matrix and hence $\mathbf{z}_t \sim \mathsf{D}_n[\boldsymbol{\mu}, \mathbf{M}]$. Let:

$$\mathbf{M} = \mathbf{B}\boldsymbol{\Lambda}\mathbf{B}' \tag{5}$$

where $\mathbf{B}'\mathbf{B} = \mathbf{I}_n$, so $\mathbf{B}^{-1} = \mathbf{B}'$ and the eigenvalues are ordered from the largest downwards with:

$$\mathbf{B}' = \begin{pmatrix} \mathbf{B}'_1 \\ \mathbf{B}'_2 \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{22} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{\Lambda}_{11}$ is $m \times m$, with $\mathbf{B}'_1\mathbf{M}\mathbf{B}_1 = \boldsymbol{\Lambda}_{11}$ and:

$$\mathbf{B}\boldsymbol{\Lambda}\mathbf{B}' = \mathbf{B}_1\boldsymbol{\Lambda}_{11}\mathbf{B}'_1 + \mathbf{B}_2\boldsymbol{\Lambda}_{22}\mathbf{B}'_2.$$

Consequently, from (2) and (6):

$$\mathbf{B}'(\mathbf{z}_t - \boldsymbol{\mu}) = \mathbf{B}'(\boldsymbol{\Psi}(\mathbf{f}_t - \boldsymbol{\kappa}) + \mathbf{e}_t) = \mathbf{f}_t - \boldsymbol{\kappa} \tag{7}$$

If only m linear combinations actually matter, so $n - m$ do not, the matrix $\mathbf{B}_1'$ weights the $\mathbf{z}_t$ to produce the relevant principal components where:

$$\mathbf{B}_1' (\mathbf{z}_t - \boldsymbol{\mu}) = \mathbf{f}_{1,t} - \boldsymbol{\kappa}_1 \tag{8}$$

In (7), we allow for the possibility that $n = m$, so $\mathbf{f}_t$ is the complete set of principal components entered in the candidate selection set, of which only $\mathbf{f}_{1,t}$ are in fact relevant to explaining y_t.

4 Variable-based and factor-based models

The postulated in-sample DGP for y_t is:

$$y_t = \beta_0 + \boldsymbol{\beta}' \mathbf{z}_{t-1} + \rho y_{t-1} + \epsilon_t \tag{9}$$

where $|\rho| < 1$ and $\epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right]$. Integrated-cointegrated systems can be reduced to this framework analytically, albeit posing greater difficulties empirically. Under weak stationarity in-sample:

$$\mathsf{E}[y_t] = \beta_0 + \boldsymbol{\beta}' \mathsf{E}[\mathbf{z}_{t-1}] + \rho \mathsf{E}[y_{t-1}] = \beta_0 + \boldsymbol{\beta}' \boldsymbol{\mu} + \rho\delta = \delta \tag{10}$$

so $\delta = \left(\beta_0 + \boldsymbol{\beta}'\boldsymbol{\mu}\right) / (1 - \rho)$ and (9) can be expressed as:

$$y_t - \delta = \boldsymbol{\beta}' (\mathbf{z}_{t-1} - \boldsymbol{\mu}) + \rho (y_{t-1} - \delta) + \epsilon_t \tag{11}$$

or as an EqCM when that is a useful reparametrization. In general, only a subset of the $\mathbf{z}_{t-1}$ will matter substantively, and we denote that by $\mathbf{z}_{a,t-1}$, so the remaining variables are not individually significant at relevant sample sizes, leading to:

$$y_t - \delta \simeq \boldsymbol{\beta}_a' (\mathbf{z}_{a,t-1} - \boldsymbol{\mu}_a) + \rho_a (y_{t-1} - \delta) + u_t \tag{12}$$

However, that does not preclude that known linear combinations of the omitted variables might also be significant if added.

Alternatively, following up that last comment, from (7):

$$y_t - \delta = \boldsymbol{\beta}' \mathbf{B} (\mathbf{f}_{t-1} - \boldsymbol{\kappa}) + \rho (y_{t-1} - \delta) + \epsilon_t = \boldsymbol{\tau}' (\mathbf{f}_{t-1} - \boldsymbol{\kappa}) + \rho (y_{t-1} - \delta) + \epsilon_t \tag{13}$$

where again only a subset may matter, namely the $\mathbf{f}_{1,t-1}$ in (8), so that:

$$y_t - \delta \simeq \boldsymbol{\tau}_1' (\mathbf{f}_{1,t-1} - \boldsymbol{\kappa}_1) + \rho_1 (y_{t-1} - \delta) + w_t \tag{14}$$

Finally, there is the possibility that when both variables and their principal components are allowed, some $\mathbf{z}_{b,t-1}$ of the $\mathbf{z}_{a,t-1}$ and some $\mathbf{f}_{2,t-1}$ of the $\mathbf{f}_{1,t-1}$ are retained to provide closer, yet more parsimonious, approximations to the behavior of y_t in-sample:

$$y_t - \delta = \boldsymbol{\beta}_b' (\mathbf{z}_{b,t-1} - \boldsymbol{\mu}_b) + \boldsymbol{\tau}_2' (\mathbf{f}_{2,t-1} - \boldsymbol{\kappa}_2) + \rho_b (y_{t-1} - \delta) + \epsilon_t \tag{15}$$

In practice, there may well have been location shifts and outliers in-sample, so we also include IIS during model selection. Thus, a vector of deterministic terms (such as intercepts, location shifts, and indicator variables) denoted $\mathbf{q}_t$ with $\mathbf{Q}_t^1 = (\mathbf{q}_1 \ldots \mathbf{q}_t)$ is allowed, as well as longer lags, so the sequential conditional expectation of y_t at time t is denoted $\mathsf{E}_t[y_t|\mathbf{Z}_{t-1}^1, \mathbf{Y}_{t-1}^1 \mathbf{q}_t]$ (when that exists).

An important special case is when the DGP for y_t is an autoregression, so that none of the $z_{i,t-1}$ actually matter. When y_t is just an AR(1), say, then:

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + v_t.$$

Searching over factors alone, excluding y_{t-1}, might lead to the retention of many of the $\mathbf{f}_t$ to approximate y_{t-1}, especially if $\mathbf{z}_t$ includes y_t, so the starting model always includes y_{t-1}, allowing for simpler models when they are the DGP. When searching over $2n$ variables and factors at significance level α, then $2\alpha n$ will be adventitiously significant on average (one retained by chance when $n = 100$ and $\alpha = 0.005$, or $\alpha = 0.0025$ with IIS at $T = 150$, correpsonding to critical values of $c_\alpha \simeq 2.85$ and $c_\alpha \simeq 3.10$).

5 Forecasting from factors when variables matter

The aim is to forecast the scalar $\{y_{T+h}\}$ over a forecast horizon $h = 1, \ldots, H$, from a forecast origin at T, at which point the information set consists of $\mathbf{Z}_T^1 = (\mathbf{z}_1 \ldots \mathbf{z}_T)$ and $(y_1 \ldots y_T)$. Forecast accuracy is to be judged by a criterion function $C_e\left(\widehat{u}_{T+1|T} \ldots \widehat{u}_{T+H|T}\right)$, which we take to depend only on the forecast errors $\widehat{u}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T}$, where 'smaller' values of $C_e(\cdot)$ are preferable. Even so, unless the complete joint density is known, evaluation outcomes depend on the specific transformation considered.

Once in-sample estimates of the factors $\{\widehat{\mathbf{f}}_t\}$ are available, one-step forecasts can be generated from estimates of the selected equation:[3]

$$\widehat{y}_{T+1|T} = \widehat{\delta} + \widehat{\boldsymbol{\tau}}_1'\left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) + \widehat{\rho}\left(\widehat{y}_T - \widehat{\delta}\right) \tag{16}$$

where $\widehat{y}_T$ is the 'flash' estimate of the forecast origin value. Multi-step estimation can be used to obtain the values of the coefficients in the forecasting device (see e.g., Clements and Hendry, 1996, Bhansali, 2002, and Chevillon and Hendry, 2005, for overviews), so for h-step ahead forecasts:

$$\widehat{y}_{T+h|T} = \widehat{\delta}_{(h)} + \widehat{\boldsymbol{\tau}}_{1,(h)}'\left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) + \widehat{\rho}_{(h)}\left(\widehat{y}_T - \widehat{\delta}_{(h)}\right) \tag{17}$$

in which case $\widehat{u}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T}$ will generally be a moving average process of order $(h-1)$ (denoted MA$(h-1)$).

Existing taxonomies of sources of forecast errors have analyzed a range of open and closed models in variables, so here we consider the factor model when the DGP has a factor structure, as in (13). The DGP depends on $\mathbf{z}_{t-1}$ and y_{t-1}, although not all the variables $z_{i,t-1}$ need enter the DGP, and the forecasting model is allowed to incorporate a subset of the factors. Our taxonomy of forecast errors focuses attention on what are likely to be the principle sources of forecast bias and forecast-error variance. Following earlier work, we begin by allowing location shifts as the only source of instability over the forecast horizon, but then consider the impact of a shift in the parameter vector that determines the impact of the factors on y_t. Stock and Watson (2009) consider the effects of instabilities in the forecasting model–that is, in the effects of the factors on y_t–but as we show, a key determinant of forecasting performance is the impact of location shifts. We let the DGP change at T to:

$$y_{T+h} = \delta^* + \boldsymbol{\beta}'\left(\mathbf{z}_{T+h-1} - \boldsymbol{\mu}^*\right) + \rho\left(y_{T+h-1} - \delta^*\right) + \epsilon_{T+h} \tag{18}$$

for $h = 1, \ldots, H$. Mapping to principal components yields:

$$y_{T+h} = \delta^* + \boldsymbol{\tau}'\left(\mathbf{f}_{T+h-1} - \boldsymbol{\kappa}^*\right) + \rho\left(y_{T+h-1} - \delta^*\right) + \epsilon_{T+h} \tag{19}$$

where for now $\boldsymbol{\tau}$ and ρ remain at their in-sample values during the forecast period.

We derive the 1-step forecast-error taxonomy, which highlights the key factors, and allows us to separately distinguish 11 sources of forecast error. Calculating the forecast error as (19) minus (16), for $h = 1$, gives rise to:

$$\begin{aligned}\widehat{u}_{T+1|T} = &\left(\delta^* - \widehat{\delta}\right) + \boldsymbol{\tau}'\left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) - \widehat{\boldsymbol{\tau}}_1'\left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) \\ &+ \rho\left(y_T - \delta^*\right) - \widehat{\rho}\left(\widehat{y}_T - \widehat{\delta}\right) + \epsilon_{T+1}.\end{aligned}$$

Using $\boldsymbol{\tau}_1'\left(\boldsymbol{\kappa}_1^* - \boldsymbol{\kappa}_1\right) + \boldsymbol{\tau}_2'\left(\boldsymbol{\kappa}_2^* - \boldsymbol{\kappa}_2\right) = \boldsymbol{\tau}'\left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right)$, we derive the forecast error reported in table 1.

[3] Estimates $\widehat{\mathbf{f}}_{1,t}$ of $\mathbf{f}_t$ using principal components $\mathbf{B}_1'\left(\mathbf{z}_t - \overline{\boldsymbol{\mu}}\right)$ depend on the scaling of the $\mathbf{z}_t$, so are often based on the correlation matrix.

Table 1: Factor model taxonomy of forecast errors, $\widehat{u}_{T+1|T} = \ldots$

$(1-\rho)(\delta^* - \delta)$	[A] equilibrium-mean shift
$-\boldsymbol{\tau}'(\boldsymbol{\kappa}^* - \boldsymbol{\kappa})$	[B] factor-mean shift
$+(1-\rho)\left(\delta - \widehat{\delta}\right)$	[C] equilibrium-mean estimation
$-\boldsymbol{\tau}_1'(\boldsymbol{\kappa}_1 - \widehat{\boldsymbol{\kappa}}_1)$	[D] factor-mean estimation
$+\rho(y_T - \widehat{y}_T)$	[E] flash estimate error
$+\boldsymbol{\tau}_1'\left(\mathbf{f}_{1,T} - \widehat{\mathbf{f}}_{1,T}\right)$	[F] factor estimate error
$+\boldsymbol{\tau}_2'(\mathbf{f}_{2,T} - \boldsymbol{\kappa}_2)$	[G] factor approximation error
$+(\boldsymbol{\tau}_1 - \widehat{\boldsymbol{\tau}}_1)'\left(\widehat{\mathbf{f}}_{1,T} - \boldsymbol{\kappa}_1\right)$	[H] factor estimation covariance
$+(\rho - \widehat{\rho})\left(\widehat{y}_T - \widehat{\delta}\right)$	[I] flash estimation covariance
$+(\boldsymbol{\tau}_1 - \widehat{\boldsymbol{\tau}}_1)'(\boldsymbol{\kappa}_1 - \widehat{\boldsymbol{\kappa}}_1)$	[J] parameter estimation covariance
$+\epsilon_{T+1}$	[K] innovation error

Taking expectations assuming near unbiased parameter estimates, and neglecting terms of $\mathsf{O}_p\left(T^{-1}\right)$:

$$\mathsf{E}\left[\widehat{u}_{T+1|T}\right] \simeq (1-\rho)(\delta^* - \delta) - \boldsymbol{\tau}'(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}) + \rho(y_T - \mathsf{E}[\widehat{y}_T]) + \boldsymbol{\tau}_1'(\mathbf{f}_{1,T} - \mathsf{E}[\widehat{\mathbf{f}}_{1,T}]) \tag{20}$$

which indicates that sources [A] and [B] in table 1 are primary determinants of forecast bias, although data and factor estimation errors ([E] and [F]) also contribute. These last two and all the remaining terms contribute to the forecast-error variance. The factor approximation error does not enter (20) as $\mathsf{E}[\mathbf{f}_{2,T}] = \boldsymbol{\kappa}_2$. Even when [E] and [F] are negligible, the equilibrium-mean and factor-mean shifts could be large. For example, if in (1):

$$\boldsymbol{\pi}^* = \boldsymbol{\pi} + 1_{(t\geq T)}\boldsymbol{\theta} \;\text{ for }\; h = 1, \ldots, H \tag{21}$$

so that the intercept in the unmodeled variables representation undergoes a permanent shift at T, then as:

$$\boldsymbol{\pi} = (\mathbf{I}_n - \boldsymbol{\Pi})\,\boldsymbol{\Psi}\boldsymbol{\kappa}$$

when $\boldsymbol{\Pi}$ and $\boldsymbol{\Psi}$ are constant, $\boldsymbol{\kappa}$ will shift, and for $n = m$:

$$\boldsymbol{\kappa}^* = \boldsymbol{\Psi}^{-1}(\mathbf{I}_n - \boldsymbol{\Pi})^{-1}\boldsymbol{\pi}^* = \boldsymbol{\kappa} + 1_{(t\geq T)}\boldsymbol{\Psi}^{-1}(\mathbf{I}_n - \boldsymbol{\Pi})^{-1}\boldsymbol{\theta} \tag{22}$$

Thus, forecast-error biases are entailed by equilibrium-mean shifts within the forecasting model of y_{T+1} (i.e., $\delta^* \neq \delta$) or in the external variables entering its DGP ($\boldsymbol{\kappa}^* \neq \boldsymbol{\kappa}$) irrespective of the inclusion or exclusion of the associated factors, whereas the approximation error by itself does not induce such a problem. This outcome is little different from a model based directly on the $\mathbf{z}_t$ (rather than $\mathbf{f}_t$) where shifts in their equilibrium mean can also induce forecast failure yet omission does not exacerbate that problem (see Hendry and Mizon, 2011, for a general taxonomy of systems with unmodeled variables).

Consider now the possibility that $\boldsymbol{\tau}$ and ρ change value for the forecast period, so that in place of (19) the DGP is given by:

$$y_{T+1} = \delta^* + \boldsymbol{\tau}^{*\prime}(\mathbf{f}_T - \boldsymbol{\kappa}^*) + \rho^*(y_T - \delta^*) + \epsilon_{T+1} \tag{23}$$

Without constructing a detailed taxonomy, the key impacts can be deduced. Relative to the baseline case illustrated in table 1, the change in $\boldsymbol{\tau}$ induces an additional error term:

$$\boldsymbol{\tau}^{*\prime}(\mathbf{f}_T - \boldsymbol{\kappa}^*) - \boldsymbol{\tau}'(\mathbf{f}_T - \boldsymbol{\kappa}^*) = \left(\boldsymbol{\tau}^{*\prime} - \boldsymbol{\tau}'\right)(\mathbf{f}_T - \boldsymbol{\kappa}^*)$$

so that the slope change will interact with the location shift, but in its absence will be relatively benign– this additional term will not contribute to the bias when $\boldsymbol{\kappa}^* = \boldsymbol{\kappa}$, suggesting the primacy of location shifts. In a similar fashion, the change in persistence of the process (the shift in ρ) only affects the forecast bias if the mean of y_t also changes over the forecast period. To see this, the additional term in the forecast error when ρ shifts is:

$$(\rho^* - \rho)(y_T - \delta^*)$$

which has a zero expectation when the shift in ρ does not cause a shift in δ, so $\delta^* = \delta$.

Finally, it is illuminating to consider the principal sources of forecast error for an $\mathrm{AR}(1)$ model, as this model serves as the benchmark against which the *selected* factor-and-variable models in section 8 are to be compared. For the sake of brevity, we ignore factors of secondary importance, such as parameter estimation uncertainty and data mis-measurement, and construct the forecast error for the $\mathrm{AR}(1)$:

$$y_t = \delta + \gamma (y_{t-1} - \delta) + u_t \tag{24}$$

when the forecast period DGP is given by (19). Notice that the omission of the factors will typically change the autoregressive parameter γ, so that γ need not equal ρ, but the long-run mean is the in-sample period value of δ. Denoting the forecast error from the $\mathrm{AR}(1)$ model by $\widehat{v}_{T+1|T}$, we obtain:

$$\widehat{v}_{T+1|T} = (1-\rho)(\delta^* - \delta) - \boldsymbol{\tau}'(\boldsymbol{\kappa}^* - \mathbf{f}_T) + (\rho - \gamma)(y_T - \delta)$$

with a forecast bias of:

$$\mathsf{E}\left[\widehat{v}_{T+1|T}\right] = (1-\rho)(\delta^* - \delta) - \boldsymbol{\tau}'(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}),$$

matching the two leading terms in (20) for the bias of the factor-forecasting model. Hence whether we include the 'correct' set of factors, a subset of these, or none at all will have no effect on the bias of the forecasts (at the level of abstraction we are operating at here). This affirms the importance of location shifts and the relative unimportance of forecasting model mis-specification (as in e.g., Clements and Hendry, 2006).

6 The equilibrium-correction problem

Section 5 assumes a single forecast origin, but forecasting is rarely viewed as a one-off venture, and of interest is the performance of the competing models as the origin moves through time. Although all models will fail when there is a location shift which is unknown when the forecast is made, of interest is the speed and extent to which forecasts recover as the origin moves forward in time from the break point. A feature of the 'equilibrium-correction' class of models, to which (16) belongs, is their lack of adaptability over time. To see this, note that (16) could be rewritten for 1-step forecasts as:

$$\Delta\widehat{y}_{T+1|T} = \widehat{\boldsymbol{\tau}}_1' \left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) + (\widehat{\rho} - 1)\left(\widehat{y}_T - \widehat{\delta}\right)$$

so that $\mathsf{E}\left[\Delta\widehat{y}_{T+1|T}\right] \simeq 0$, whereas the DGP is given by:

$$\Delta y_{T+1} = \boldsymbol{\tau}'(\mathbf{f}_T - \boldsymbol{\kappa}^*) + (\rho - 1)(y_T - \delta^*) + \epsilon_{T+1} \tag{25}$$

with an expected value which is non-zero when there are locations shifts:

$$\mathsf{E}[\Delta y_{T+1}] = \boldsymbol{\tau}'\mathsf{E}[\mathbf{f}_T - \boldsymbol{\kappa}^*] + (\rho - 1)\mathsf{E}[y_T - \delta^*] = \boldsymbol{\tau}_1'(\boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_1^*) + (\rho - 1)(\delta - \delta^*) \tag{26}$$

Thus shifts in the deterministic terms will induce forecast failure, principally because they are embedded in Δy_{T+1}, but not in forecasts of this quantity. The class of equilibrium-correction models is such that this problem persists as the origin is extended forward. For example, forecasting $T+2$ from $T+1$ even for known in-sample parameters, accurate data and no approximation error, we find:

$$\Delta\widehat{u}_{T+2|T+1} = \Delta y_{T+2} - \Delta\widehat{y}_{T+2|T+1} = \boldsymbol{\tau}_1' \left(\boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_1^*\right) + (\rho - 1)\left(\delta - \delta^*\right) + \epsilon_{T+1}$$

This generic difficulty for EqCMs suggests using a robust forecasting device approach which exploits (26), as in:

$$\Delta\widetilde{y}_{T+2|T+1} = \Delta y_{T+1} + \widehat{\boldsymbol{\tau}}_1' \Delta\widehat{\mathbf{f}}_{1,T} + (\widehat{\rho} - 1)\,\Delta\widehat{y}_T$$

Again, under the simplifying assumptions (known in-sample parameters, etc), and denoting the forecast error by $\Delta\widetilde{u}_{T+2|T+1} = \Delta y_{T+2} - \Delta\widetilde{y}_{T+2|T+1}$, using (25) gives:

$$\begin{aligned}\Delta\widetilde{u}_{T+2|T+1} &= \boldsymbol{\tau}' \left(\mathbf{f}_{T+1} - \boldsymbol{\kappa}^*\right) + (\rho - 1)\left(y_{T+1} - \delta^*\right) + \epsilon_{T+2} \\ &\quad - \boldsymbol{\tau}' \left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) - (\rho - 1)\left(y_T - \delta^*\right) - \epsilon_{T+1} \\ &= \boldsymbol{\tau}_1' \Delta\mathbf{f}_{1,T+1} + (\rho - 1)\,\Delta y_{T+1} + \Delta\epsilon_{T+2} \end{aligned} \tag{27}$$

which is less dependent on the location shifts.

To the extent that most factor models are also EqCMs, location shifts could have two impacts. The first is when breaks affect the mapping between the original variables' information and the derived factors (i.e., changes in the weights). This is addressed in Stock and Watson (2009), who find a relatively innocuous effect. Breaks in the coefficients of zero-mean variables or factors in forecasting models also appear less problematic.

However, breaks due to location shifts within any EqCM forecasting model will induce systematic mis-forecasting, and the above analysis applies equally to factor-based models (as illustrated in section 5). In the empirical forecasting exercise in section 8 below, the variables are already differenced once, so large shifts in equilibrium means are unlikely, and hence such formulations already embody a partial robustness to previous location shifts. Indeed, if in place of (24), the differenced-data version is used, then forecasting $T+2$ from $T+1$:

$$\Delta\widetilde{y}_{T+2|T+1} = \gamma\Delta y_{T+1}$$

when:

$$\Delta y_{T+2} = \boldsymbol{\tau}^{*\prime}\Delta\mathbf{f}_{T+1} + \rho^*\Delta y_{T+1} + \Delta\epsilon_{T+2}$$

we have:

$$\widetilde{v}_{T+2|T+1} = \boldsymbol{\tau}^{*\prime}\Delta\mathbf{f}_{T+1} + (\rho^* - \gamma)\,\Delta y_{T+1} + \Delta\epsilon_{T+2}$$

which is close to (27).

7 Automatic Model Selection

The primary comparison of interest is between automatic selection over variables as against PC-based factor models in terms of forecasting. Factors are often regarded as necessary to summarize a large amount of information, but automatic selection procedures show this is unnecessary. Selection will place a zero weight on variables that are insignificant in explaining variation in the dependent variable according to a pre-specified critical value, whereas principal components will place a small, but non-zero weight on variables that have a low correlation with other explanatory variables.

One advantage of using an automatic model selection algorithm is that it enables us to remain agnostic initially about the LDGP. If the data are generated by a few latent factors that capture underlying movements in the economy such as business cycles, then principal components should be used to forecast future outcomes. On the other hand, if the data are generated by individual disaggregated economic variables then these should form the forecasting model. By including both explanations jointly the data can determine the most plausible structure.

A further advantage of model selection is that arbitrary methods to select the relevant principal components are not needed. Various methods have been proposed in the literature but most take the principal components that explain the most variation between the set of explanatory variables, not the most variation between the explanatory variables and the dependent variable. This would require the correlation structure between the regressors and the dependent variable to be similar to the correlation structure within the regressors (see e.g., Castle *et al.*, 2011b). Instead, by selecting PCs based on their statistical significance, we capture the latter correlation. In the empirical application we find the retained PCs tend not to be the first few PCs, suggesting that the correlation structure does differ between the dependent variable and the disaggregates.

The model selection algorithm used is *Autometrics*, which undertakes a multi-path tree search, commencing from the general model with all potential regressors including variables, factors and lags of both as well as impulse indicators, and eliminates insignificant variables while ensuring a set of pre-specified diagnostic tests are satisfied in the reduction procedure, checking the subsequent reductions with encompassing tests. Variables are eliminated if they are statistically insignificant at the chosen criterion whilst ensuring the resulting model is still congruent and encompassing (see Doornik, 2008). There are various methods to speed up the search procedure which involve joint testing.

The multi-path tree search enables perfectly collinear sets of regressors to be included jointly. While the general model is not estimable initially, the search procedure proceeds by excluding one of the perfectly-collinear variables initially so selection is undertaken within a subset of the candidate set, but the multi-path search allows that excluded variable to be included in a different path search, with another perfectly-singular variable being dropped. *Autometrics* uses expanding as well as contracting searches which enables regressors initially excluded to return within different candidate sets. This 'sieve' continues until $N < T$ and there are no perfect singularities. The standard tree search selection can then be applied: see Doornik (2009a, 2009b).

8 Forecasting US GDP and Inflation

The empirical forecasting exercise compares the forecast performance of regression models based on principal components, variables, or both. We forecast quarterly GDP growth and the quarterly change in inflation over the period 1997–2006, as well as considering the corresponding level forecasts for GDP and quarterly inflation. Models are selected in-sample using *Autometrics*, with all variables and their principal components included in the candidate set.

The AR benchmark models against which factor model forecasts are often compared have typically been difficult to beat systematically over this forecast period. For example, in terms of forecasting inflation, Stock and Watson (2010) argue that simple univariate models, such as a random-walk model, or the time-varying unobserved components model of Stock and Watson (2007), are competitive with models with explanatory variables. Stock and Watson (2003) are relatively downbeat about the usefulness of leading indicators for predicting output growth: also see Clements and Galvão (2009) for evidence using higher-frequency data.

8.1 Data

The data are 144 seasonally adjusted quarterly time series for the United States, over 1959q1–2006q4 taken from Stock and Watson (2009). There are $n = 109$ disaggregates in the dataset which are used as the candidate set of regressors and are also the set of variables used to form the principal components. All data (usually in logs) are transformed to remove unit roots by taking first or second differences, as described in Stock and Watson (2009) Appendix Table A1. The estimation sample spans T =1962q3–2006q4, after transformations and lags, with the forecast horizon spanning H =1997q1–2006q4.

8.1.1 Principal Components

Let $\mathbf{Z}$ denote the $(T+H)\times n$ matrix of transformed disaggregated variables, and $\widehat{\mathbf{\Omega}}$ the $n\times n$ sample correlation matrix. The eigenvalue decomposition is:

$$\widehat{\mathbf{\Omega}} = \widehat{\mathbf{B}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{B}}' \tag{28}$$

where $\widehat{\mathbf{\Lambda}}$ is the diagonal matrix of ordered eigenvalues $(\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_n \geq 0)$ and $\widehat{\mathbf{B}} = (\widehat{\mathbf{b}}_1, \ldots, \widehat{\mathbf{b}}_n)$ is the corresponding matrix of eigenvectors, with $\widehat{\mathbf{B}}'\widehat{\mathbf{B}} = \mathbf{I}_n$. The sample principal components are computed as:

$$\widehat{\mathbf{f}}_t = \widehat{\mathbf{B}}'\widetilde{\mathbf{z}}_t \tag{29}$$

where $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_T)'$ denotes the standardized data, $\widetilde{z}_{j,t} = (z_{j,t} - \overline{z}_j)/\widetilde{\sigma}_{x_j}\ \forall j = 1, \ldots, n$ where $\overline{z}_j = \frac{1}{T}\sum_{t=1}^{T} z_{j,t}$ and $\widetilde{\sigma}_{z_j} = \left[\frac{1}{T}\sum_{t=1}^{T}(z_{j,t} - \overline{z}_j)^2\right]^{1/2}$. When the principal components are estimated in-sample, $h = 0$, whereas $h = 1, \ldots, H$ for recursive estimation of the principal components.

8.2 Impulse-indicator saturation

A number of authors have assessed the forecast performance of factor models over this period, and Stock and Watson (2011) review studies which explicitly consider the impact of breaks on factor model forecasts. One of the key studies is Stock and Watson (2009) who find 'considerable evidence of instability in the factor model; the indirect evidence suggests instability in all elements (the factor loadings, the factor dynamics, and the idiosyncratic dynamics).' (Stock and Watson, 2009, p. 197). They identify a break in 1984, associated with the Great Moderation (see, e.g., McConnell and Perez-Quiros, 2000), and find the coefficients of the factors in the forecasting models are not constant across this period.

However, they argue that the factors can be reasonably well estimated by PCs even when the individual loadings are subject to instability, and more accurate factor model forecasts are found to result from estimating the factors on the whole sample, but the forecasting models only on the period after 1984. As such, we estimate the principal components over the full sample period, but rather than restricting the estimation sample to post-1984, we use IIS to account for parameter stability. Instead of imposing a single break at this point, we test for the presence of multiple breaks and outliers jointly with the selection procedure by applying IIS. Although this procedure adds an impulse indicator for every observation to the candidate regressor set, there is little efficiency loss under the null of no breaks, yet the procedure has power to detect both outliers and location shifts when there are breaks: see Castle *et al.* (2011c).

8.3 Forecasting models

The forecasting models are obtained by undertaking selection on the general unrestricted model (GUM):

$$\Delta y_t = \gamma_0 + \sum_{j=J_a}^{J_b} \gamma_j \Delta y_{t-j} + \sum_{i=1}^{n}\sum_{j=J_a}^{J_b} \beta_{i,j}\Delta z_{i,t-j} + \sum_{k=1}^{n}\sum_{j=J_a}^{J_b} \phi_{k,j} f_{k,t-j} + \sum_{l=1}^{T} \delta_l \mathbf{1}_{\{l=t\}} + \epsilon_t \tag{30}$$

where Δy_t is the first difference of log real gross domestic product or the quarterly change in quarterly inflation.

Forecasting models are obtained by undertaking selection on (30) using *Autometrics* where we set:
(i) $\boldsymbol{\phi} = \mathbf{0}$, i.e. select over variables only;
(ii) $\boldsymbol{\beta} = \mathbf{0}$, select over factors only; and
(iii) $\boldsymbol{\phi} \neq \mathbf{0}$ and $\boldsymbol{\beta} \neq \mathbf{0}$, i.e. jointly select variables and factors;
where the intercept and lags of the dependent variable are included in all models. For the three forecasting specifications we consider:
(a) $\boldsymbol{\delta} = \mathbf{0}$, no IIS; and
(b) $\boldsymbol{\delta} \neq \mathbf{0}$, with IIS,
resulting in six forecasting model specifications.

Three forecast horizons are recorded, including 1-step, 4-step and 8-step ahead direct forecasts. For the 1-step ahead forecasts we set $J_a = 1$ and $J_b = 4$, allowing for 4 lags of the dependent and exogenous regressors. For 4-step ahead direct forecasts we set $J_a = 4$ and $J_b = 7$, and 8-step ahead forecasts set $J_a = 8$ and $J_b = 11$.

One of the problems that factor forecasts face is the need to difference to stationarity in order to compute the principal components. This implies that any structural breaks in the levels will be differenced out. Many variables are second differenced to obtain stationarity, so breaks in growth rates will also be removed. A consequence of differencing is that the resulting forecasts will be robust to breaks after they occur. However, accurate forecasting of the differences is insufficient for levels, as an insignificant one-direction error can cumulate to an important under or over forecast: e.g., 1% per quarter cumulates to 4% over a year. Hence, we also report the implied log-level forecasts. As the forecasts are direct h-step forecasts rather than dynamic forecasts, we let the forecast origin, T, roll forward rather than remain fixed at 1996q4. The levels forecasts for log GDP and quarterly inflation are computed as:

$$\widehat{y}_{T+k+h} = \sum_{i=1}^{h} \Delta\widehat{y}_{T+k+h} + y_{T+k} \quad \text{for} \quad k = 0, \ldots, H - h \tag{31}$$

where $h = 4$ and 8 for the 4-step and 8-step ahead forecasts respectively in (31) (1-step forecast errors are the same for levels and differences). For the 4-step forecasts, we evaluate over 37 forecasts as we require forecasts for the period 1997q1–1997q4 to obtain the level forecast of 1997q4, and likewise for the 8-step ahead forecasts we evaluate over 33 forecasts.

There are $T = 138$ observations in-sample. For the variables only or factor forecasts, there are $N = 441$ regressors in (30) with no IIS and $N = 579$ with IIS. We set the significance level for selection at $\alpha = 1\%$ with no IIS, so approximately 4 regressors would be retained on average under the null that no regressors are relevant, and set $\alpha = 0.5\%$ for IIS. When both variables and factors are included, there are $N = 877$ regressors with no IIS, so we set $\alpha = 0.5\%$, and with IIS $N = 1015$, so $\alpha = 0.25\%$. Table 2 summarizes the selection significance levels, with null retention numbers in parentheses. Overfitting is not a concern despite commencing with $N \gg T$. Parsimony can be achieved by a tighter significance level, with the cost being a loss of power for regressors with a significance level less than the critical value. As a check, we consider a super-conservative strategy by setting the significance levels even tighter, as shown in the last row of table 2.

	Variables		Factors		Both	
	No IIS	IIS	No IIS	IIS	No IIS	IIS
Number of regressors	441	579	441	579	877	1015
Conservative (null)	1% (4.4)	0.5% (3)	1% (4.4)	0.5% (3)	0.5% (4.4)	0.25% (2.5)
Tighter (null)	0.5% (2.2)	0.1% (0.6)	0.5% (2.2)	0.1 (0.6)%	0.1% (0.9)	0.05% (0.5)

Table 2: Significance levels used for model selection.

We also compute three benchmark forecasts including the random walk and AR(1) forecasts computed directly and iteratively:

$$\Delta\widehat{y}^{RW}_{T+k+h} = \Delta y_{T+k} \tag{32}$$

$$\Delta\widehat{y}^{AR(D)}_{T+k+h} = \widehat{\beta}_0 + \widehat{\beta}_1 \Delta y_{T+k} \tag{33}$$

$$\Delta\widehat{y}^{AR(I)}_{T+k+h} = \sum_{i=1}^{h-1} \widehat{\gamma}_0 \widehat{\gamma}_1^i + \widehat{\gamma}_1^h \Delta y_{T+k} \tag{34}$$

for $k = 0, \ldots, H-h$ and $h = 1, 4$ and 8.

8.4 Results

We first consider in-sample selection and estimation, where the forecasting model is selected and estimated over $t = 1, \ldots, T$, with 40 forecasts computed over $k = 1, \ldots, H$, resulting in 90 forecasting models. We evaluate the forecasts on RMSFE, with the *caveat* that the MSFE criterion may not result in a definitive ranking, as that measure is not invariant to non-singular, scale preserving linear transformations, see Clements and Hendry (1993a, 1993b).

8.4.1 In-sample estimation and selection for GDP

First we consider the forecast performance for GDP of factors, variables or both when the forecasting model is selected and estimated in-sample. Table 3 records the in-sample model fit and number of retained regressors for selection with IIS. In-sample, the model fit is generally better for the variables than factors. Tightening the significance level results in fewer retained regressors and a worse model fit as expected. Few dummies are retained on average, but the retained dummies are clustered around 1984, supporting the identification of a break there by Stock and Watson (2009).

	1-step		4-step		8-step	
	Cons	Super	Cons	Super	Cons	Super
<u>Variables</u>						
$\widehat{\sigma}$	0.559	0.669	0.629	0.711	0.702	0.799
No. regressors	9	5	14	7	9	5
No. dummies	2	1	1	2	5	4
<u>Factors</u>						
$\widehat{\sigma}$	0.657	0.702	0.718	0.798	0.671	0.813
No. regressors	5	4	8	4	12	5
No. dummies	2	1	6	2	5	2
<u>Both</u>						
$\widehat{\sigma}$	0.553	0.753	0.712	0.819	0.767	0.788
No. regressors	9	2	9	4	9	6
No. dummies	2	0	1	0	2	4

Table 3: In-sample results for GDP growth models selected with IIS: $\widehat{\sigma}$ = equation standard error, No. regressors and No. dummies record the number of regressors and, as a subset, the number of dummies retained, and Cons and Super are the conservative and super-conservative strategies respectively.

Table 4 records the average RMSFE for each of the forecasting models, averaged across horizon, whether IIS is applied, and the selection significance level for GDP and GDP growth. The RMSFEs are closely similar for the first three methods, but forecasting with factors is slightly preferable to forecasting with variables for GDP growth, although it performs worse for the levels forecasts. However, the AR(1) model, either iterative or direct, has the lowest average RMSFE.

	Variables	Factors	Both	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{T+k}$	0.666	0.600	0.650	0.666	0.485	0.491
$\widehat{y}_{T+k}$	1.931	1.971	1.925	3.509	1.324	1.336

Table 4: RMSFE $(\times 100)$ for GDP and quarterly GDP growth, with benchmark Random Walk, direct AR(1) [AR(D)] and iterative AR(1) [AR(I)] forecasts.

We deconstruct the results for the first 3 methods in figures 1 and 2 by analyzing forecast performance over the forecast horizon in panel (a), whether IIS was applied or not in panel (b), and the tightness of the selection criterion in panel (c). We plot the results for the 1-step ahead levels forecasts in figure 2 for comparison, despite their being identical to those in figure 1.

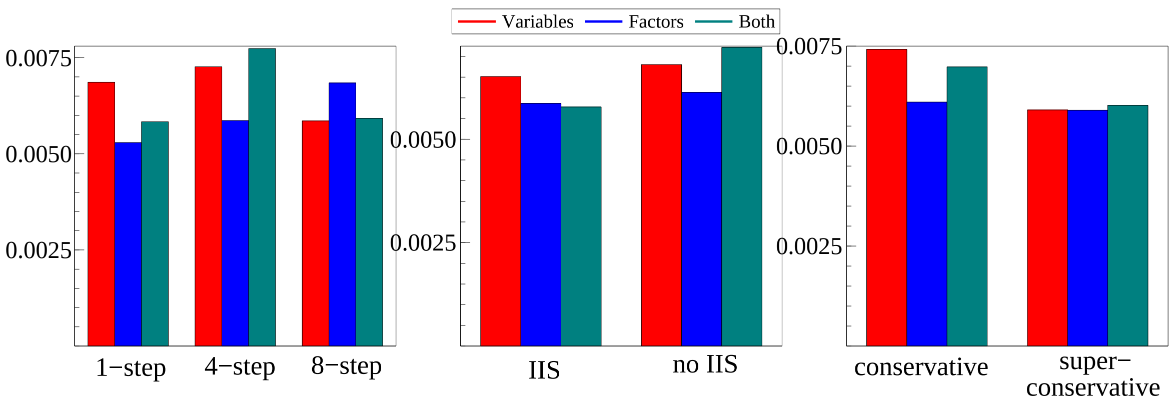

Figure 1: Average RMSFE for GDP growth $(\Delta\widehat{y}_{T+h})$

For GDP growth, the factor model performs best at short horizons but worst at longer horizons. The variable model improves in forecast accuracy at the 8-step horizon. There is some benefit to IIS but few indicators are retained on average given the tight significance level used for selection (no more than six), and the indicators retained are roughly the same for the factors and variables models. A tighter significance level does improve the variable model forecasts, suggesting parsimonious models are preferable here, but there is little improvement to the forecasts in the factor model.

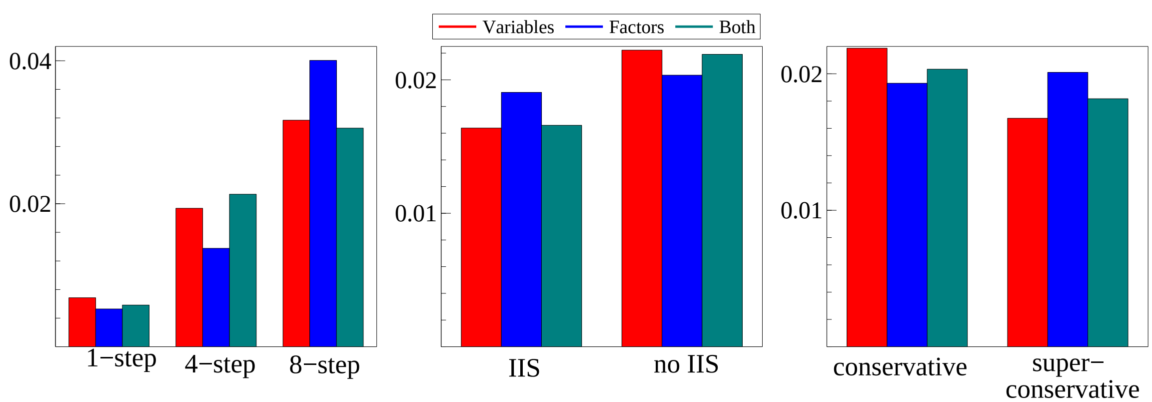

Figure 2: Average RMSFE for log GDP $(\widehat{y}_{T+h})$

In levels, worsening factor forecasts as the horizon increases is evident. IIS yields greater improvements as the few dummies that are retained will be translated to level shifts capturing the shift in 1984.

Figures 3 and 4 record the distributions of forecast errors for variables (panel a), factors (panel b) and both (panel c) for GDP growth and the level of GDP respectively.[4] In growth rates, the forecast errors are close to normal for variables and factors models. The levels forecasts have a fatter upper tail and some evidence of bimodality, but there are no significant differences between the variables and factors model forecast errors.

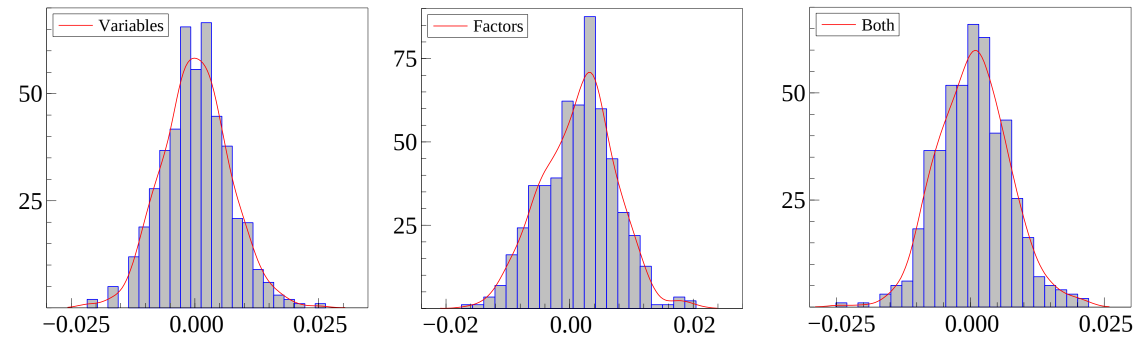

Figure 3: Distributions of forecast errors for GDP growth averaging across horizon, IIS/no IIS and selection strategy.

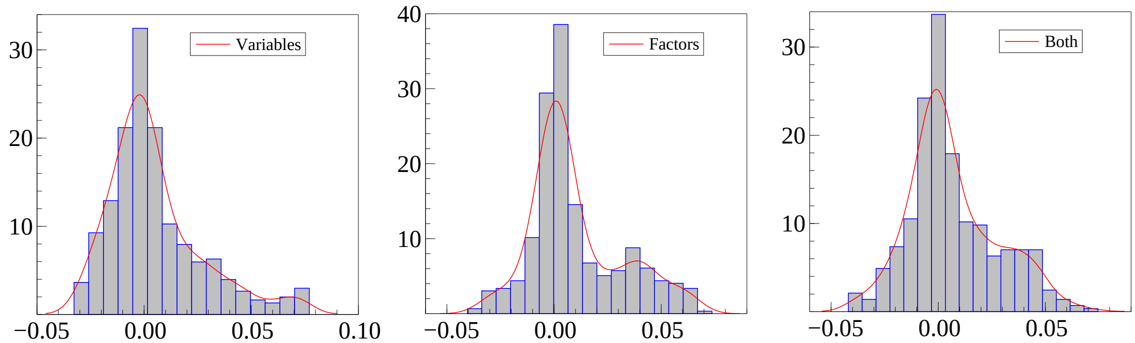

Figure 4: Distributions of forecast errors for log GDP averaging across horizon, IIS/no IIS and selection strategy.

8.4.2 In-sample estimation and selection for inflation

We next consider the results for inflation. Table 5 records the in-sample model fit and number of retained variables for selection with IIS. In-sample, selecting over variables results in a better model fit than factors or both at all horizons, in keeping with the GDP results. Direct models at longer horizons do not always result in a worse fit, and neither does a tighter significance level, despite fewer regressors retained for the super-conservative strategy. Few dummies are retained on average suggesting that the differencing to stationarity has removed most breaks, although more dummies are retained at longer horizons as direct models exclude factors and variables at lags shorter than the forecast horizon.

Table 6 reports the forecast results for inflation, averaged across horizon, selection significance level and whether IIS is applied. In differences, including both variables and factors outperforms the individual

[4]For GDP growth there are 480 forecast errors for each model specification. For the levels distributions, we include the 1-step levels forecasts (which are identical to the difference forecast errors), resulting in 440 forecast errors.

models, but not the AR(1) models, either direct or iterative, as in GDP growth. Variables perform better than factors, particularly in levels, due to worsening forecast performance of factors at longer horizons.

	1-step		4-step		8-step	
	Cons	Super	Cons	Super	Cons	Super
Variables						
$\widehat{\sigma}$	0.240	0.283	0.293	0.290	0.227	0.296
No. regressors	20	12	17	13	21	12
No. dummies	2	1	4	8	8	3
Factors						
$\widehat{\sigma}$	0.263	0.336	0.299	0.318	0.311	0.327
No. regressors	15	7	20	14	21	20
No. dummies	5	1	4	13	7	6
Both						
$\widehat{\sigma}$	0.266	0.284	0.302	0.356	0.284	0.334
No. regressors	14	11	12	10	13	8
No. dummies	4	3	4	3	1	3

Table 5: In-sample results for inflation forecasting models selected with IIS: $\widehat{\sigma}$ = equation standard error, No. regressors and No. dummies record the number of regressors and, as a subset, the number of dummies retained, and Cons and Super are conservative and super-conservative strategies respectively.

	Variables	Factors	Both	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{T+k}$	0.535	0.582	0.491	0.577	0.416	0.415
$\widehat{y}_{T+k}$	0.687	0.981	0.656	0.469	0.444	0.464

Table 6: RMSFE for quarterly inflation and the quarterly change in inflation, with benchmark Random Walk, direct AR(1) [AR(D)] and iterative AR(1) [AR(I)] forecasts.

To examine the results more closely, figures 5 and 6 record results for the forecast horizon in panel (a), whether IIS was applied or not in panel (b), and the selection significance level in panel (c).

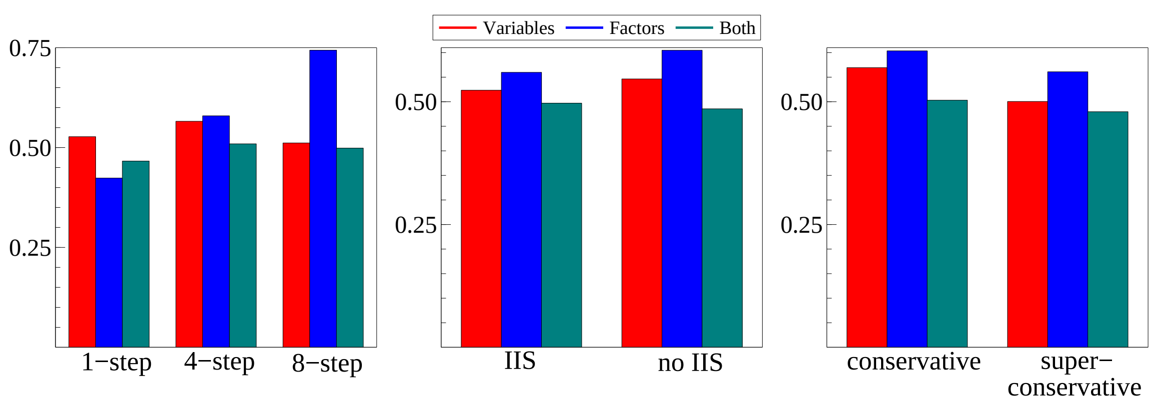

Figure 5: Average RMSFE for annual change in inflation $(\Delta\widehat{y}_{T+k})$

In differences, variables models and both factors and variables models are roughly constant over the forecast horizon but the factor models perform worst on average as the horizon increases. There are some gains to IIS for variable models or factor models, but the gains are small. Tightening the selection criterion improves the forecasts, with the more parsimonious models for the variable models and factor

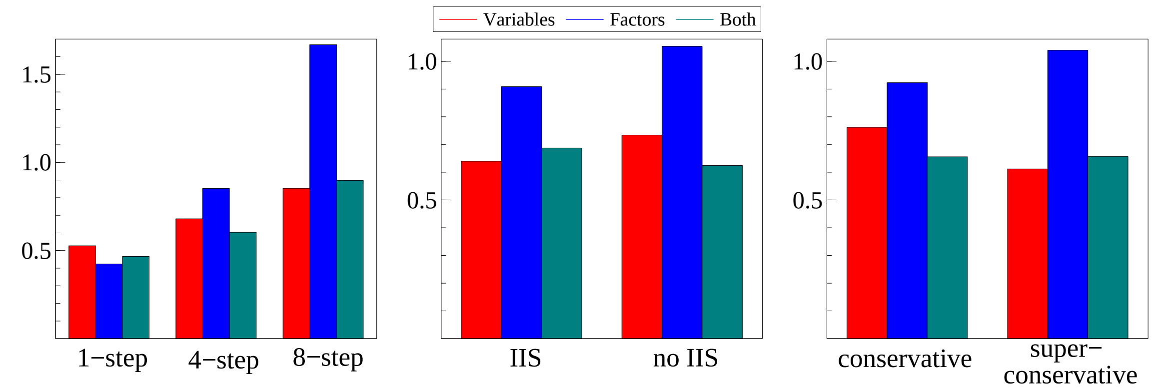

Figure 6: Average RMSFE for quarterly inflation $(\widehat{y}_{T+k})$

models yielding a smaller RMSFE. Factors models selected using a super-conservative strategy are best on a RMSFE criterion at a 1-step horizon, both with and without IIS, but also the worst over the longer 8-step horizon. Thus, factor models appear more useful for the shorter forecasting horizons here.

Transforming to levels highlights the worsening forecast accuracy of the factor model over the forecast horizon. The transformation can result in differing rankings, as can be seen at the 8-step horizon where the variable model is preferred in levels but the combined model is preferred in differences, albeit by small magnitudes. Also, a looser strategy is preferred for the factor models as opposed to the other models. The differences between the combined and variable models are small because most of the retained regressors for the combined model are variables. However, including many additional factors in selection is not costly when undertaken at a conservative or super-conservative strategy.

Figures 7 and 8 record the distributions of forecast errors for the variables, factors, and combined models. The forecast errors are approximately normally distributed for the variables models. In levels, the factors model has a fatter lower tail, but again there is little systematic difference in forecast errors for the different models.

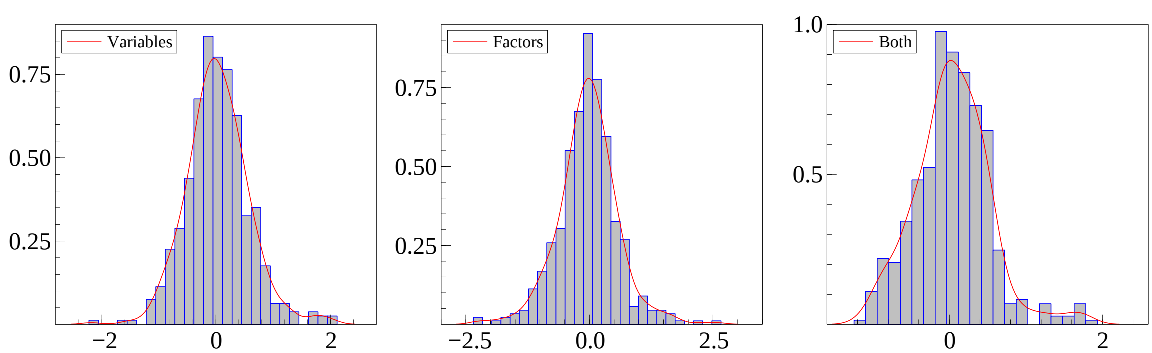

Figure 7: Distributions of forecast errors for the quarterly change in inflation averaging across horizon, IIS/no IIS and selection strategy.

8.4.3 Selection and estimation method

Finally, we assess whether the selection and estimation method used impacts on forecast accuracy. We compare four methods:

(1) the forecasting model is fixed after selection and estimation in-sample over $t = 1, \ldots, T$;

(2) the model is selected over $t = 1, \ldots, T$, but both parameters and eigenvalues for the principal components are recursively re-estimated over each forecast horizon, $T = 1, \ldots, T + h$;

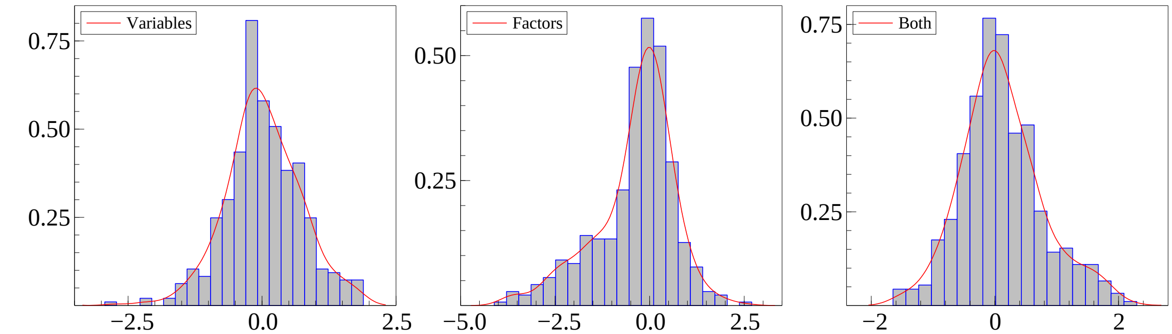

Figure 8: Distributions of forecast errors for quarterly inflation averaging across horizon, IIS/no IIS and selection strategy.

(3) the models is both selected and re-estimated recursively over each forecast horizon, $t = 1, \ldots, T + h$, and the eigenvalues re-calculated, so the model specification can change with each new observation; and (4) partial recursive selection and estimation, where the previous selected model is forced, in that selection occurs only over the additional variables or factors, so previously retained variables are not dropped even if insignificant.

Figure 9 records the RMSFEs for GDP growth with IIS using the conservative strategy for each estimation method. There are no clear advantages to using any specific selection and estimation method. Recursive estimation marginally outperforms in-sample estimation but there are cases where in-sample estimation is preferred, for example the factor model at 4-step and the variable model at 8-step. Recursive selection and estimation can be worse at longer horizons, with partial recursive selection providing a slightly more robust selection method. In general, the differencing and IIS has accounted for the in-sample breaks and so there is little difference between selection and estimation methods for GDP growth.

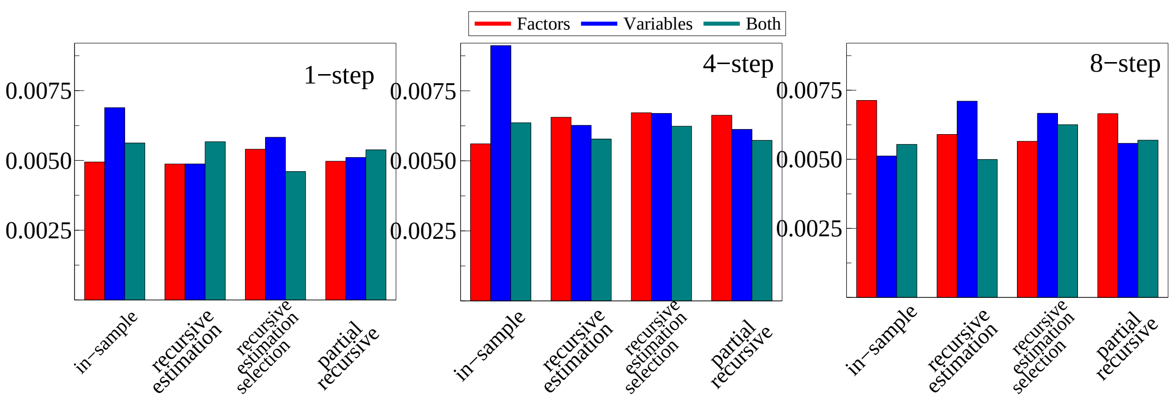

Figure 9: Average RMSFE for GDP growth with IIS using the conservative strategy.

Figure 10 records the RMSFEs for the quarterly change in inflation with IIS using the conservative strategy for each estimation method. At 1-step and 4-step there is very little difference in RMSFEs between the selection and estimation strategies, but at 8-step there is weak evidence to suggest recursive estimation or partial recursive estimation and selection is preferred, particularly for variables.

9 Conclusion

There have been many analyses of the forecast performance of either factor models or regression models, but few examples of the joint consideration of factors and variables. Automatic model selection can allow

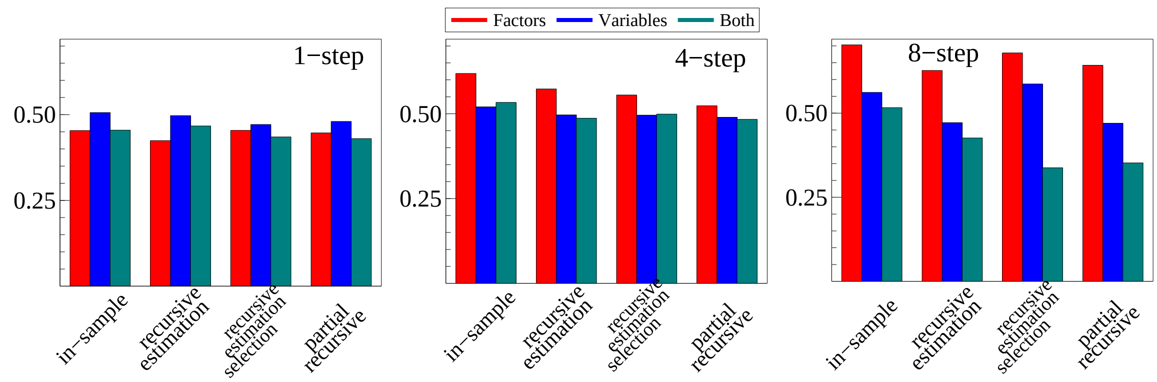

Figure 10: Average RMSFE for the quarterly change in inflation with IIS using the conservative strategy.

for more regressors than observations, perfect collinearities and multiple breaks and outliers, so the set of candidate regressors can include both factors, as measured by their static principal components, and variables. We consider which methods perform best in a forecasting context, including naive devices.

One of the key explanations for forecast failure is that of structural breaks. When the underlying data generating process shifts, but the forecasting model remains unchanged, forecast failure will result. As both variables and factor models are in the class of equilibrium-correction models, they both face the problem of non-robustness to location shifts. In our empirical example, we use impulse-indicator saturation to account for breaks and outliers in-sample, noting that IIS could be used to implement intercept corrections if an indicator variable were retained for the last in-sample observation. We find there is some advantage to using IIS for forecasting in differences as the unconditional mean is better estimated, but as the data are differenced to estimate the principal components, few impulse-indicators are retained. Backing out levels forecasts does highlight the non-stationarity due to level shifts (e.g., the structural break in 1984), and a further extension would be to consider modeling and selecting variables in levels, augmented by the stationary principal components which may pick up underlying latent variable dynamics.

The empirical application considered both GDP and inflation, and their differences, using *Autometrics* to select forecasting models that include either principal components (PCs), individual variables, or both. The results are mixed but suggest that factor models are more useful for 1-step ahead forecasting, but their relative performance declines as the forecast horizon increases. For direct multi-step forecasting, *Autometrics* selection over variables (or variables and PCs) tends to forecast better than factor forecasts, suggesting that there are benefits to selecting the weights based on the correlation with y_{t+h}. There is little evidence to suggest that recursive estimation or recursive estimation and selection is better. A more robust alternative of partial recursive estimation and selection is proposed. There are gains to using IIS, but a tight significance is needed to control the retention rate.

The ability to undertake model selection jointly on factors and variables avoids imposing a model specified in either just variables or factors, and circumvents the need for arbitrary selection of factors. Thus it is a useful tool, both for in-sample modeling and for forecasting out-of-sample. Whether the data are best described by latent factors or observable variables will depend on the phenomena being analysed, and can be determined by the data itself using model selection techniques.

References

Allen, P. G., and Fildes, R. A. (2001). Econometric forecasting strategies and techniques. In Armstrong, J. S. (ed.), *Principles of Forecasting*, pp. 303–362. Boston: Kluwer Academic Publishers.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley & Sons.

Bánbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. in Clements, and Hendry (2011), Ch. 7.

Bartholomew, D. J. (1987). *Latent Variable Models and Factor Analysis*. New York: Oxford University Press.

Bhansali, R. J. (2002). Multi-step forecasting. in Clements, and Hendry (2002), pp. 206–221.

Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011a). Evaluating automatic model selection. *Journal of Time Series Econometrics*, **3 (1)**, DOI: 10.2202/1941–1928.1097.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011b). Model selection in equations with many 'small' effects. Discussion paper 528, Economics Department, Oxford University.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011c). Model selection when there are multiple breaks. *Journal of Econometrics*, forthcoming.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2009). Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review*, **210**, 71–89.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, **158**, 25–36.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting Breaks and During Breaks. In Clements, and Hendry (2011), Ch. 11.

Castle, J. L., and Hendry, D. F. (2011a). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T.(eds.), *System Identification, Environmental Modelling and Control*, p. forthcoming. New York: Springer.

Castle, J. L., and Hendry, D. F. (2011b). On not evaluating economic models by forecast outcomes. Discussion paper, 538, Economics Department, Oxford University.

Castle, J. L., and Shephard, N.(eds.)(2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Cattell, R. B. (1952). *Factor Analysis*. New York: Harper.

Chevillon, G., and Hendry, D. F. (2005). Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting*, **21**, 201–218.

Clements, M. P., and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, **26**, 546–554.

Clements, M. P., and Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, **24**, 1187–1206.

Clements, M. P., and Galvão, A. B. (2011). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. Discussion paper, Department of Economics, University of Warwick.

Clements, M. P., and Hendry, D. F. (1993a). On the limitations of comparing mean squared forecast errors (with discussion). *Journal of Forecasting*, **12**, 617–637.

Clements, M. P., and Hendry, D. F. (1993b). On the limitations of comparing mean squared forecast errors: A reply. *Journal of Forecasting*, **12**, 669–676.

Clements, M. P., and Hendry, D. F. (1996). Multi-step estimation for forecasting. *Oxford Bulletin of*

Economics and Statistics, **58**, 657–684.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

Clements, M. P., and Hendry, D. F. (2001). Explaining the results of the M3 forecasting competition. *International Journal of Forecasting*, **17**, 550–554.

Clements, M. P., and Hendry, D. F.(eds.)(2002). *A Companion to Economic Forecasting*. Oxford: Blackwells.

Clements, M. P., and Hendry, D. F. (2005a). Evaluating a model by forecast performance. *Oxford Bulletin of Economics and Statistics*, **67**, 931–956.

Clements, M. P., and Hendry, D. F. (2005b). Guest editors' introduction: Information in economic forecasting. *Oxford Bulletin of Economics and Statistics*, **67**, 713–753.

Clements, M. P., and Hendry, D. F. (2006). Forecasting with breaks. In Elliott, G., Granger, C., and Timmermann, A.(eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 605–657: Elsevier, Horth-Holland.

Clements, M. P., and Hendry, D. F.(eds.)(2011). *Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

Corradi, V., and Swanson, N. R. (2011). Testing for factor model forecast and structural stability. Working paper, Economics Department, Warwick University.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In Elliott, G., Granger, C., and Timmermann, A.(eds.), *op.cit.*, pp. 961–982.

Dees, S., di Mauro, F., Pesaran, M. H., and Smith, L. V. (2007). Exploring the international linkages in the EURO area: A global VAR analysis. *Journal of Applied Econometrics*, **22**, 1–38.

Diebold, F. X., and Rudebusch, G. D. (1991). Forecasting output with the composite leading index: An ex ante analysis. *Journal of the American Statistical Association*, **86**, 603–610.

Doornik, J. A. (2007). *Object-Oriented Matrix Programming using Ox* 6th ed. London: Timberlake Consultants Press.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics*, **70**, 915–925.

Doornik, J. A. (2009a). Autometrics. In Castle, and Shephard (2009), pp. 88–121.

Doornik, J. A. (2009b). Econometric model selection with more variables than observations. Working paper, Economics Department, University of Oxford.

Duesenberry, J. S., Fromm, G., Klein, L. R., and Kuh, E.(eds.)(1969). *The Brookings Model: Some Further Results*. Amsterdam: North-Holland.

Emerson, R. A., and Hendry, D. F. (1996). An evaluation of forecasting using leading indicators. *Journal of Forecasting*, **15**, 271–291.

Ericsson, N. R. (2010). Improving GVARs. Discussion paper, Federal Reserve Board of Governors, Washington, D.C.

Fair, R. C. (1970). *A Short-Run Forecasting Model of the United States Economy*. Lexington: D.C. Heath.

Faust, J., and Wright, J. H. (2007). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *NBER Working Paper Series 13397*.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Fore-*

casting, **8**, 81–98.

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

Garratt, A., Lee, K., Mise, E., and Shields, K. (2008). Real time representations of the output gap. *Review of Economics and Statistics*, **90**, 792–804.

Garratt, A., Lee, K., Mise, E., and Shields, K. (2009). Real time representations of the UK output gap in the presence of model uncertainty. *International Journal of Forecasting*, **25**, 81–102.

Giacomini, R., and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, **74**, 1545 – 1578.

Gorman, W. M. (1956). Demand for related goods. Discussion paper, Agricultural Experimental Station, Iowa.

Granger, C. W. J. (1989). Combining forecasts - Twenty years later. *Journal of Forecasting*, **8**, 167–173.

Granger, C. W. J., and Pesaran, M. H. (2000a). A decision-theoretic approach to forecast evaluation. In Chon, W. S., Li, W. K., and Tong, H.(eds.), *Statistics and Finance: An Interface*, pp. 261–278. London: Imperial College Press.

Granger, C. W. J., and Pesaran, M. H. (2000b). Economic and statistical measures of forecasting accuracy. *Journal of Forecasting*, **19**, 537–560.

Hansen, P. R., and Timmermann, A. (2011). Choice of sample split in forecast evaluation. Working paper, Economics Department, Stanford University.

Hecq, A., and Jacobs, J. P. A. M. (2009). On the VAR-VECM representation of real time data. Discussion paper, mimeo, University of Maastricht, Department of Quantitative Economics.

Hendry, D. F. (2009). The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In Mills, T. C., and Patterson, K. D.(eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Basingstoke: Palgrave MacMillan.

Hendry, D. F., and Clements, M. P. (2004). Pooling of forecasts. *Econometrics Journal*, **7**, 1–31.

Hendry, D. F., and Doornik, J. A. (2009). *Empirical Econometric Modelling using PcGive: Volume I*. London: Timberlake Consultants Press.

Hendry, D. F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, **33**, 317–335. Erratum, 337–339.

Hendry, D. F., and Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, **115**, C32–C61.

Hendry, D. F., and Mizon, G. E. (2011). An open-model forecast-error taxonomy. Working paper 552, Economics Department, Oxford University.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

Howrey, E. P. (1978). The use of preliminary data in econometric forecasting. *Review of Economics and Statistics*, **60**, 193–200.

Howrey, E. P. (1984). Data revisions, reconstruction and prediction: an application to inventory investment. *The Review of Economics and Statistics*, **66**, 386–393.

Johansen, S., and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, and Shephard (2009), pp. 1–36.

Joreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**.

Kishor, N. K., and Koenig, E. F. (2010). VAR estimation and forecasting when data are subject to revision. *Journal of Business and Economic Statistics*. Forthcoming.

Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921–41*. No. 11 in Cowles Commission Monograph. New York: John Wiley.

Klein, L. R. (1971). *An Essay on the Theory of Economic Prediction*. Chicago: Markham Publishing Company.

Klein, L. R., Ball, R. J., Hazlewood, A., and Vandome, P. (1961). *An Econometric Model of the UK*. Oxford: Oxford University Press.

Lawley, D. N., and Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method*. London: Butterworth and Co.

Leitch, G., and Tanner, J. E. (1991). Economic forecast evaluation: Profits versus the conventional error measures. *American Economic Review*, **81**, 580–590.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., *et al.* (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1**, 111–153.

Makridakis, S., and Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

McConnell, M. M., and Perez-Quiros, G. P. (2000). Output fluctuations in the United States: What has changed since the early 1980s?. *American Economic Review*, **90**, 1464–1476.

Patterson, K. D. (1995). An integrated model of the data measurement and data generation processes with an application to consumers' expenditure. *Economic Journal*, **105**, 54–76.

Patterson, K. D. (2003). Exploiting information in vintages of time-series data. *International Journal of Forecasting*, **19**, 177–197.

Peña, D., and Poncela, P. (2004). Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, **119**, 291–321.

Peña, D., and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, **136**, 1237–1257.

Persons, W. M. (1924). *The Problem of Business Forecasting*. No. 6 in Pollak Foundation for Economic Research Publications. London: Pitman.

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, **73**, 1057–1084.

Pesaran, M. H., Schuerman, T., and Smith, L. V. (2009). Forecasting economic and financial variables with global VARs. *International Journal of Forecasting*, **25**, 642–675.

Pesaran, M. H., and Skouras, S. (2002). Decision-based methods for forecast evaluation. in Clements, and Hendry (2002), pp. 241–267.

Pesaran, M. H., and Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics*, **10**, 461–465.

Pesaran, M. H., and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, **129**, 183–217.

Pesaran, M. H., and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, **137**, 134–161.

Sargent, T. J. (1989). Two models of measurements and the investment accelerator. *Journal of Political Economy*, **97**, 251–287.

Schumacher, C., and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, **24**, 386–398.

Smets, F., and Wouters, R. (2003). An estimated stochastic dynamic general equilibrium model of the Euro Area. *Journal of the European Economic Association*, **1**, 1123–1175.

Smith, B. B. (1927). Forecasting the volume and value of the cotton crop. *Journal of the American Statistical Association*, **22**, 442–459.

Smith, B. B. (1929). Judging the forecast for 1929. *Journal of the American Statistical Association*, **24**, 94–98.

Spearman, C. (1927). *The Abilities of Man*. London: Macmillan.

Stock, J. H., and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macro-Economic Annual*, 351–409.

Stock, J. H., and Watson, M. W. (1999). A comparison of linear and nonlinear models for forecasting macroeconomic time series. In Engle, R. F., and White, H.(eds.), *Cointegration, Causality and Forecasting*, pp. 1–44. Oxford: Oxford University Press.

Stock, J. H., and Watson, M. W. (2003). How Did Leading Indicator Forecasts Perform During the 2001 Recession. *Federal Reserve Bank of Richmond, Economic Quarterly*, **89/3**, 71–90.

Stock, J. H., and Watson, M. W. (2007). Why has U.S. Inflation Become Harder to Forecast?. *Journal of Money, Credit and Banking*, **Supplement to Vol. 39**, 3–33.

Stock, J. H., and Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. In Castle, and Shephard (2009), Ch. 7.

Stock, J. H., and Watson, M. W. (2010). Modelling Inflation after the Crisis. *NBER Working Paper Series*, **16488**.

Stock, J. H., and Watson, M. W. (2011). Dynamic factor models. In Clements, and Hendry (2011), Ch. 2.

Stone, J. R. N. (1947). On the interdependence of blocks of transactions. *Journal of the Royal Statistical Society*, **8**, 1–32. Supplement.

Swanson, N. R., and van Dijk, D. (2006). Are statistical reporting agencies getting it right? data rationality and business cycle asymmetry. *Journal of Business and Economic Statistics*, **24**, 24–42.

Tinbergen, J. (1930). Determination and interpretation of supply curves: An example [Bestimmung und Deutung von Angebotskurven: ein Beispiel]. *Zeitschrift fur Nationalökonomie*, **1**, 669–679. Reprinted in Hendry, D. F. and Morgan, M. S. (1995), *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.

Tinbergen, J. (1951). *Business Cycles in the United Kingdom 1870–1914*. Amsterdam: North-Holland.

Waelbroeck, J. K. (ed.)(1976). *The Models of Project LINK*. Amsterdam: North-Holland Publishing Company.

Wallis, K. F. (1989). Macroeconomic forecasting: A survey. *Economic Journal*, **99**, 28–61.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica*, **64**, 1067–1084.

West, K. D., and McCracken, M. W. (1998). Regression-based tests of predictive ability. *International Economic Review*, **39**, 817–840.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**, 324–342.

Zarnowitz, V., and Boschan, C. (1977). Cyclical indicators: An evaluation and new leading indexes. In *Handbook of Cyclical Indicators*, pp. 170–183. Washington, USA: U.S. Department of Commerce.