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,
Oxford Martin School, University of Oxford, UK
[♦] Economics Department, Warwick University, UK
[⋆]Economics Department and Institute for New Economic Thinking,
Oxford Martin School, University of Oxford, UK

Abstract

We consider forecasting with factors, variables and both, modeling in-sample using *Autometrics* so all principal components and variables can be included jointly, while tackling multiple breaks by impulse-indicator saturation. A forecast-error taxonomy for factor models highlights the impacts of location shifts on forecast-error biases. Forecasting US GDP over 1-, 4- and 8-step horizons using the dataset from Stock and Watson (2009) updated to 2011:2 shows factor models are more useful for nowcasting or short-term forecasting, but their relative performance declines as the forecast horizon increases. Forecasts for GDP levels highlight the need for robust strategies, such as intercept corrections or differencing, when location shifts occur as in the recent financial crisis.

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, which are typically univariate.

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).

[*]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. Hashem has also published on virtually every conceivable topic in econometrics, both theory and applied, thereby acquiring almost 20,000 citations, as well as creating and editing the *Journal of Applied Econometrics* since its foundation in 1986. This research was supported in part by grants from the Open Society Foundations and the Oxford Martin School. We would like to thank seminar participants at the Computational and Financial Econometrics Conference, London 2011, the OxMetrics Conference, Washington 2012 and Leicester University Departmental Seminar for helpful discussions. We also acknowledge helpful comments from two anonymous referees and the editors. Contact details: jennifer.castle@magd.ox.ac.uk, M.P.Clements@warwick.ac.uk and david.hendry@nuffield.ox.ac.uk.

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) and Schumacher and Breitung (2008).

The third set includes methods like exponentially weighted moving averages, 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: see Makridakis, Andersen, Carbone, Fildes *et al.* (1982) and Makridakis and Hibon (2000), and are the 'neither' in the title.

Until recently, while the first two approaches often compared their forecasts with various 'naive' methods selected from the third group, there were few direct comparisons between them, and almost no studies included both. Here we consider models selected from very general initial specifications, which might be motivated as approximating the reduced forms of the models of the first group, and compare these directly with the factor models of the second group. Our automatic model selection algorithm permits the inclusion of variables and factors on an equal footing, allowing in-sample selection over both of these based on their explanatory power for the target variable. This remedies the dearth of direct comparisons of the two approaches in the literature. But, as emphasised in section 2, in-sample fit does not guarantee of a good out-of-sample forecast performance (see e.g., Clements and Hendry, 2005a) so that a detailed analysis of forecasting performance is undertaken.

The structure of the paper is as follows. Section 2 describes some of the issues that arise in any analysis of forecasting models or methods. Section 3 outlines the statistical framework used to analyse forecasting with factors or variables. Section 4 develops the analysis of forecasting from factor models when there are location shifts. Section 5 discusses model selection when including both factors and variables. Section 6 briefly reviews alternatives to principal components proposed in the statistics literature, including those which form part of the empirical analysis. Section 7 illustrates the analysis using US GDP forecasts. Section 8 concludes.

2 Setting the scene

Many interacting issues need to be addressed when analysing forecasting, the complexity of which entail that the answer to the title's question is likely to be context specific. Although general guidelines are rare, it is fruitful to 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) more versus less information in forecasting; (v) the type of forecasting model in use, specifically whether 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 data-generating process (DGP). We briefly consider these in turn, and indicate whether in principle our approach of selecting over variables and factors should be advantageous. An advantage of our approach is that it allows us to be agnostic as to the nature of the data generating process (DGP), especially whether the DGP can be usefully represented as having a factor structure.

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 the set of factors and the original variables will often result in the number of candidate variables, N, being larger than the sample size, T. Model selection when $N > T$ may have seemed insurmountable in the past, but is not now. Let $\mathbf{z}_t$ denote the set of n 'external' variables' from which the factors $\mathbf{f}_t = \mathbf{H}\mathbf{z}_t$ (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 with mixtures of expanding and contracting block searches, so can handle settings with both perfect collinearity and $N > T$: see Hendry and Krolzig (2005) and Doornik (2009b). The simulations in Castle, Doornik and Hendry (2011) show the feasibility of such an approach when $N > T$ in linear dynamic models. Investigators are, therefore, not forced to allow for only a small number of factors, or just the factors and a few lags of the variable being forecast, as candidates. Since model selection is unavoidable when $N > T$, we consider that next.

2.2 Model selection

The search algorithm in *Autometrics* within *PcGive* (see Doornik, 2009a, and Doornik and Hendry, 2009) seeks the local DGP (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, 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, 2012b), 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. Hence $N > T$ will always occur when IIS is used, but the in-sample feasibility of this approach is shown in Castle, Doornik and Hendry (2012a). 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 location shifts. Since there are 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 when faced with location shifts at the forecast origin. Section 5 discusses automatic model selection.

2.3 Unanticipated location shifts

Third, *ex ante* forecasting is fundamentally different from *ex post* modeling when unanticipated location shifts 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) proposed a general theory of economic forecasting using mis-specified models in a world of structural breaks, 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 show 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 (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

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 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, so using large data sets may not correct one of the main problems confronting forecasters, namely location shifts, unless that additional information is pertinent to forecasting breaks. Moreover, although we use model selection from a very general initial candidate set, combined with congruence as a basis for econometric modeling, it cannot be proved that congruent modeling helps for forecasting when facing location shifts (see e.g., Allen and Fildes, 2001). While Makridakis and Hibon (2000) conclude that parsimonious models do best in forecasting competitions, Clements and Hendry (2001) argue that such findings may conflate parsimony and robustness to location shifts: most of the parsimonious models were relatively robust to location shifts compared to their non-parsimonious contenders.[1] 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

Factor models are often equilibrium correction in form, so they suffer from the general non-robustness to location shifts of that class of model: see, e.g., Clements and Hendry (1998), Ch.8. However, the principles of robust-model formulation discussed in Castle *et al.* (2011) apply, and any EqCM, whether based on variables or factors (or both), could be differenced prior to forecasting, thereby embedding the resulting model in a more robust forecasting device. Castle *et al.* (2011) 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 4 provides a more extensive discussion.

2.6 Measurement errors

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 a 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 Patterson, 1995, 2003), as well as the approach of Kishor and Koenig (2011) (building on earlier contributions by Howrey, 1978, 1984, and Sargent, 1989),

[1]Parsimonious models need not be robust–just consider using an estimate of the unconditional historical mean of a process as its forecast. No model specification or selection are required, and estimation is just the calculation of the sample mean, but this parsimonious forecasting device is highly susceptible to location shifts.

who estimate a VAR on post-revision data. This necessitates stopping the estimation sample short of the forecast origin, so the model's forecasts of the periods up to the origin are combined with lightly-revised data via the Kalman filter to obtain post-revision estimates. The forecast is then conditioned on these estimates of what the revised latest data will be. Clements and Galvão (2012a) provide 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.

The frequency of macroeconomic data can also affect its accuracy, as can nowcasting (see e.g., 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, 2011, proposal to predict 'final' estimates of the latest 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

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, 1993, 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), 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, we report the traditional MSFE measure, and evaluate forecasts of the levels of (log) GDP and GDP growth. Both are of interest to the policy maker: the growth rate is a headline statistic; whereas the level of GDP is required for the calculation of output gaps, see e.g., Watson (2007). To judge the accuracy of alternative forecasting models, the choice of levels versus changes can matter, as differences between the accuracy of multi-step forecasts from correctly-specified models and models which impose 'too many' unit roots are typically diminished when forecasts are evaluated in terms of growth rates rather than levels. Clements and Hendry (1998), Ch.6 show this analytically for a cointegrated VAR, using the trace of the MSFE matrix as the measure of system-wide forecast accuracy, but the results specialize to the equivalent comparisons in terms of single equations. The impact of the mis-specification for cointegrated systems of VARs in differences is attenuated when forecasts of growth rates are evaluated. When there are location shifts, the evaluation of forecasts of growth rates may cloak the benefits of a better forecasting approach, such as intercept corrections, since robust forecasting devices then typically perform better on levels forecasts.

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.

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). Letting $\boldsymbol{\lambda}_i(L)'$ denote the i^{th} row of $\boldsymbol{\Upsilon}(L)$, then:

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

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 (2009), among others, suggest that the factor structure may not be a particularly good representation of the macroeconomy. Our empirical approach allows for the 'basic' driving forces to be variables or factors, 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, 2011), we only analyze linear DGPs here.

The implications of these eight considerations combined are to include both variables and factors based on a large information set, with dynamics, and indicators for location shifts, selecting at a stringent significance level, evaluating forecasts in both level and differences, and possibly adjusting EqCM forecasting models for shifts near the forecast origin. We only 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, but we exploit the ability of automatic model selection to operate successfully in such a setting.

3 Variables versus factors: the statistical framework

We begin by describing the relationship between the 'external' variables and the factors, and then the postulated in-sample DGP that relates the variable of interest to the factors or 'external' variables.

3.1 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 intercept $\boldsymbol{\pi}$:

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

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

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

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

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

so when $\mathsf{E}\left[\mathbf{f}_t\right] = \boldsymbol{\kappa}$ and $\mathsf{E}\left[\mathbf{e}_t\right] = \mathbf{0}$, under weak stationarity in-sample from (2):

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

where $\mathbf{f}_t \sim \mathsf{ID}_m\left[\boldsymbol{\kappa}, \mathbf{P}\right]$ is a latent vector of dimension $m \leq n$, so $\mathbf{\Psi}$ is $n \times m$, with $\mathbf{e}_t \sim \mathsf{ID}_n\left[\mathbf{0}, \mathbf{\Omega_e}\right]$, $\mathsf{E}[\mathbf{f}_t\mathbf{e}_t'] = \mathbf{0}$ and $\mathsf{E}\left[\mathbf{e}_t\mathbf{e}_t'\right] = \mathbf{\Omega_e}$. Then:

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

say, where $\mathbf{P}$ is an $m \times m$ diagonal matrix and hence $\mathbf{z}_t \sim \mathsf{D}_n\left[\boldsymbol{\mu}, \mathbf{M}\right]$. Let $\mathbf{M} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}'$ where $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$, so $\mathbf{H}^{-1} = \mathbf{H}'$ and the eigenvalues are ordered from the largest downwards with:

$$\mathbf{H}' = \begin{pmatrix} \mathbf{H}_1' \\ \mathbf{H}_2' \end{pmatrix} \text{ and } \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{22} \end{pmatrix}, \tag{5}$$

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

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

Consequently, from (2) and (5):

$$\mathbf{H}'\left(\mathbf{z}_t - \boldsymbol{\mu}\right) = \mathbf{H}'\left(\mathbf{\Psi}\left(\mathbf{f}_t - \boldsymbol{\kappa}\right) + \mathbf{e}_t\right) = \mathbf{f}_t - \boldsymbol{\kappa} \tag{6}$$

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

$$\mathbf{H}_1'\left(\mathbf{z}_t - \boldsymbol{\mu}\right) = \mathbf{f}_{1,t} - \boldsymbol{\kappa}_1 \tag{7}$$

In (6), 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.

3.2 Variable-based and factor-based models

Suppose the in-sample DGP for y_t is:

$$y_t = \beta_0 + \boldsymbol{\beta}'\mathbf{z}_{t-1} + \rho y_{t-1} + \epsilon_t \tag{8}$$

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

$$\mathsf{E}\left[y_t\right] = \beta_0 + \boldsymbol{\beta}'\mathsf{E}\left[\mathbf{z}_{t-1}\right] + \rho\mathsf{E}\left[y_{t-1}\right] = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\mu} + \rho\delta = \delta \tag{9}$$

so $\delta = \left(\beta_0 + \boldsymbol{\beta}'\boldsymbol{\mu}\right) / \left(1 - \rho\right)$ and (8) can be expressed in terms of deviations from means as:

$$y_t - \delta = \boldsymbol{\beta}'\left(\mathbf{z}_{t-1} - \boldsymbol{\mu}\right) + \rho\left(y_{t-1} - \delta\right) + \epsilon_t \tag{10}$$

or as an EqCM when that is a useful reparameterization. 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 the more parsimonious model:

$$y_t - \delta = \boldsymbol{\beta}'_a\left(\mathbf{z}_{a,t-1} - \boldsymbol{\mu}_a\right) + \rho_a\left(y_{t-1} - \delta\right) + \nu_t \tag{11}$$

However, that does not preclude that known linear combinations of the omitted variables might be significant, so ν_t need not be an innovation process.

Alternatively, given the mapping between variables and factors in §3.1, if a factor structure holds, from (6) and (10) we can obtain an equivalent representation to (10) in factor space:

$$y_t - \delta = \boldsymbol{\beta}'\mathbf{H}\left(\mathbf{f}_{t-1} - \boldsymbol{\kappa}\right) + \rho\left(y_{t-1} - \delta\right) + \epsilon_t = \boldsymbol{\tau}'\left(\mathbf{f}_{t-1} - \boldsymbol{\kappa}\right) + \rho\left(y_{t-1} - \delta\right) + \epsilon_t \tag{12}$$

Again, only a subset may matter, namely the $\mathbf{f}_{1,t-1}$ in (7), and the resulting parsimonious model in the space of relevant factors becomes:

$$y_t - \delta = \boldsymbol{\tau}'_1\left(\mathbf{f}_{1,t-1} - \boldsymbol{\kappa}_1\right) + \rho_1\left(y_{t-1} - \delta\right) + \eta_t \tag{13}$$

where η_t need not be an innovation process against the omitted information: also, (11) and (13) are not equivalent representations in general even though (10) and (12) are.

We are agnostic about the nature of the DGP, and whether a forecasting model based on a variant of (11) or (13) is superior from a forecasting perspective. Our proposed selection strategy allows either a variant of (11) or (13) to be the forecasting model, when we select over variables or factors, respectively, or some hybrid of variables and factors, when we select over variables and factors.

As discussed in §2.3, the DGP may not have constant parameters, but our approach allows for location shifts and outliers in-sample by implementing IIS simultaneously with model selection.

4 Factor models and location shifts

Once in-sample principal components estimates of the factors $\{\widehat{\mathbf{f}}_t\}$ are available, one-step forecasts can be generated from estimates of the selected equation (13):[2]

$$\widehat{y}_{T+1|T} = \widehat{\delta} + \widehat{\boldsymbol{\tau}}_1' \left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) + \widehat{\rho}\left(\widehat{y}_T - \widehat{\delta}\right) \tag{14}$$

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 for h-step forecasting (see e.g., Clements and Hendry, 1998, Ch.11, Bhansali, 2002, and Chevillon and Hendry, 2005): here we focus on 1-step ahead forecasts.

Suppose the DGP has a factor structure, as in (12), but that there is a location shift at T such that:

$$y_{T+1} = \delta^* + \boldsymbol{\tau}' \left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) + \rho\left(y_T - \delta^*\right) + \epsilon_{T+1} \tag{15}$$

where for now $\boldsymbol{\tau}$ and ρ remain at their in-sample values during the forecast period. Calculating the forecast error as (15) minus (14) gives rise to:

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

Using $\boldsymbol{\tau}_1' \left(\boldsymbol{\kappa}_1^* - \boldsymbol{\kappa}_1\right) + \boldsymbol{\tau}_2' \left(\boldsymbol{\kappa}_2^* - \boldsymbol{\kappa}_2\right) = \boldsymbol{\tau}' \left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right)$, the forecast error can be written in terms of a number of distinct components as in Table 1, which can be compared to non-factor model taxonomies in e.g., Clements and Hendry (2006) and Hendry and Mizon (2012).

Here we focus on the primary determinants of forecast bias, which are [A] and [B], the equilibrium-mean shift, and the factor-mean shift, respectively. To see this, ignore terms of $\mathsf{O}_p\left(T^{-1}\right)$ (including finite-sample biases in parameter estimates), and take expectations through (16):

$$\mathsf{E}\left[\widehat{u}_{T+1|T}\right] \simeq (1-\rho)\left(\delta^* - \delta\right) - \boldsymbol{\tau}' \left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right) + \rho\left(y_T - \mathsf{E}[\widehat{y}_T]\right) + \boldsymbol{\tau}_1'(\mathbf{f}_{1,T} - \mathsf{E}[\widehat{\mathbf{f}}_{1,T}]). \tag{17}$$

From (17), data mis-measurement and factor estimation errors ([E] and [F] in Table 1) can contribute to forecast bias. These last two, together with all remaining terms, also contribute to the forecast-error variance. The factor approximation error does not enter (17) as $\mathsf{E}\left[\mathbf{f}_{2,T}\right] = \boldsymbol{\kappa}_2$.

Consider now the possibility that $\boldsymbol{\tau}$ and ρ change value for the forecast period, so that in place of (15) the DGP is given by:

$$y_{T+1} = \delta^* + \boldsymbol{\tau}^{*\prime} \left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) + \rho^*\left(y_T - \delta^*\right) + \epsilon_{T+1} \tag{18}$$

Without constructing a detailed taxonomy, the key impacts can be deduced. Relative to the baseline case

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

illustrated in Table 1, the change in $\boldsymbol{\tau}$ induces an additional error term:

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

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:

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

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

Finally, we consider the principal sources of forecast error for an $\mathrm{AR}(1)$ model, as this serves as the 'neither' benchmark against which the *selected* factor-and-variable models in section 7 are to be compared. For brevity, we ignore influences of secondary importance, such as parameter estimation uncertainty and data mis-measurement, and construct the forecast error for the $\mathrm{AR}(1)$:

$$y_t=\delta+\phi\left(y_{t-1}-\delta\right)+v_t \tag{19}$$

when the forecast period DGP is (15). The omission of the factors entails that ϕ need not equal ρ, but the long-run mean remains the in-sample value of δ. Denoting the forecast error from the $\mathrm{AR}(1)$ by $\widehat{v}_{T+1|T}$:

$$\widehat{v}_{T+1|T}=\left(1-\rho\right)\left(\delta^*-\delta\right)-\boldsymbol{\tau}^{\prime}\left(\boldsymbol{\kappa}^*-\mathbf{f}_T\right)+\left(\rho-\phi\right)\left(y_T-\delta\right)$$

with a forecast bias of:

$$\mathsf{E}\left[\widehat{v}_{T+1|T}\right]=\left(1-\rho\right)\left(\delta^*-\delta\right)-\boldsymbol{\tau}^{\prime}\left(\boldsymbol{\kappa}^*-\boldsymbol{\kappa}\right),$$

matching the two leading terms in (17) for the bias of the factor-forecasting model. Hence, whether the 'correct' set of factors, a subset of these, or none at all is included, there is no effect on the bias of the forecasts (at the level of abstraction 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).

We have assumed a single forecast origin, but forecasting is rarely a one-off venture, so the performance of the competing models as the origin moves through time is of interest. Although all models will fail when there is a location shift which is unknown when the forecast is made, the speed and extent to which forecast accuracy recovers as the origin moves forward in time from the break point are important. A feature of the 'equilibrium-correction' class of models, to which (14) belongs, is their lack of adaptability over time, as established in Clements and Hendry (1998), Ch.8: see Castle *et al.* (2011) for some potential remedies.

5 Automatic Model Selection

The primary comparison of interest here is between automatic selection over variables as against PC-based factors in terms of forecasting. Factors are often regarded as necessary to summarize a large amount of information, but automatic selection procedures are an alternative. Given a pre-specified critical value, selection will place a zero weight on variables that are insignificant in explaining variation in the dependent variable y_t, whereas principal components will place a small, but non-zero, weight even on variables that have no correlation with y_t. Moreover, automatic model selection enables us to remain agnostic about the form of 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 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 separate selection of the relevant principal components is not needed. Various methods have been proposed in the literature, but most take the principal components that explain the maximum variation within the set of explanatory variables, not the most variation between the explanatory variables and the dependent variable, which 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.*, 2012a). Instead, by selecting PCs based on their statistical significance in the forecasting model, we capture the latter correlation. In the empirical application, the retained PCs are not always the first few PCs, so the correlation structure may differ from that between the dependent variable and the disaggregates. A number of approaches have been proposed to counter the potential deficiencies of PC factor forecasting, and these are briefly reviewed in section 6.

The model selection algorithm used is *Autometrics*, which undertakes a multi-path search using block expanding and contracting searches to eliminate insignificant variables, commencing from a general model defined by all potential regressors including variables, factors and lags of both, as well as impulse indicators. Once a feasibly estimable set is found, further reductions ensure pre-specified diagnostic and encompassing tests are satisfied. Variables are eliminated if they are statistically insignificant at the chosen criterion whilst ensuring the resulting model is still congruent. Various methods of joint testing can speed up the search procedure. *Autometrics* enables perfectly-collinear sets of regressors to be included jointly. While the general model is not estimable initially, the search proceeds by excluding some of the perfectly-collinear variables, so selection is undertaken within a subset of the candidate variables, but allows excluded variables to be included in a different path search with other perfectly-singular variables being dropped. 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).

6 Principal components and related approaches

The dataset we use consists of 109 variables. Some authors such as Boivin and Ng (2006) suggest that it may be better not to use all available data when constructing PCs, although Bernanke and Boivin (2003) find that the forecast performance of their factor models improves when the factors are calculated from a dataset consisting of 215 variables compared to 78 variables. Relatedly, the standard approach calculates the PCs of the same set of variables irrespective of the target variable. Bai and Ng (2008) calculate factors from a set of ‘targeted predictors’—variables which have been shown to have predictive power for the variable of interest, based on hard or soft thresholding. Our interest is in whether the results on selecting over variables and factors, where the factors are PCs based on the full set of 109 variables, are qualitatively unchanged if instead the PCs are based on smaller sets of targeted predictors. In the empirical section, we use the soft thresholding technique of LASSO (Tibshirani, 1996) to select the first 30 ‘most important’ variables, and calculate targeted factors $\widehat{\mathbf{f}}^*$ as the PCs of this set of targeted variables. We then apply selection as in our standard case, but replacing the 109 variables and their PCs by the targeted variables and targeted PCs.

An alternative is a block-factor approach, motivated by Moench, Ng and Potter (2009), where the data is divided up into a number of categories, and PCs are calculated for each category. We choose four categories: GDP components and industrial production (19 disaggregate variables); labour market (33 disaggregate variables); prices (29 disaggregate variables); financial variables (28 disaggregate variables). The block-factor approach ensures the factors summarise information from disparate sets of variables and are more readily interpretable. The first PC was computed for each block, and entered in the forecasting model:

$$\Delta\widehat{y}_{t+h} = \widehat{\beta}_0 + \widehat{\beta}_1 \Delta y_t + \sum_{j=1}^{4} \widehat{\gamma}_j z_{j,t} \tag{20}$$

for $h = 1, 4, 8$, where $z_{j,t}$ is the first PC from the j-th block. This approach is compared to simply using the first four PCs without blocking the variables.

Other approaches have been espoused in the literature, such as De Mol, Giannone and Reichlin (2008), who find that Bayesian shrinkage tends to perform as well as PC factor-model forecasting.

7 Forecasting US GDP and GDP growth

Our empirical forecasting exercise compares the forecast performance of regression models based on principal components, variables, both or neither. We forecast quarterly GDP growth and the corresponding level over the period 2000–2011. Models are selected in-sample using *Autometrics*, with all variables and principal components included in the candidate set jointly.

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. A key studies is Stock and Watson (2009) who find (p.197) ‘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)’. They suggest estimating the factors on the full historical period across the break (there, the Great Moderation around 1984, see, e.g., McConnell and Perez-Quiros, 2000), but only estimating the forecasting models that include the factors as explanatory variables on the post-break period. As an alternative strategy to handle instability in the forecasting models, we use the full estimation sample, but with IIS.

The ‘neither’ AR benchmark against which factor model forecasts are often compared have typically been difficult to beat systematically. In terms of forecasting (e.g.) 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 including explanatory variables. Stock and Watson (2003) are relatively downbeat about the usefulness of leading indicators for predicting output growth: see Clements and Galvão (2009) for evidence using higher-frequency data.

We begin by describing the data and forecasting models, and in section 7.3 present the results when factors are calculated in the standard way from all the available data series. Section 7.4 reports results for selection over targeted factors and variables (as described in §6).

7.1 Data

The data set, based on Stock and Watson (2009), consists of 144 quarterly time series for the United States over 1959:1–2006:4, updated here to 2011:2. There are $n = 109$ disaggregate variables, used both as the candidate set of regressors and the set for the principal components. All data are transformed to remove unit roots by taking first or second differences (usually in logs) as described in Stock and Watson (2009) Appendix Table A1. The data available for estimation span T =1962:3–2011:2, so there are 150 in-sample observations after transformations and lags, with the forecast horizon spanning 2000:1–2011:2, which is separated into two subsets; 2000:1–2006:4, and 2007:1–2011:2, to assess the performance of the forecasting models over the financial crisis period.

Let $h = 1, 4, 8$ denote the step-ahead direct forecasts. Let P denote the out-of-sample forecast period, where P_0 = 2000:1, P_1 = 2006:4, and P_2 = 2011:2. Forecasts are evaluated over $P_0 : P_2$ (full forecast sample of 46 observations); $P_0 : P_1$ (forecast subsample 1 of 28 observations); and $P_1 : P_2$ (forecast subsample 2 of 18 observations). N denotes the total number of regressors, which could include T impulse indicators, lags of variables or factors, and deterministic terms.

7.1.1 Principal Components

Let $\mathbf{x}^d$ denote the $(T+m)\times n$ matrix of disaggregated variables after transforming to non-integrated by appropriate differencing, and $\widehat{\mathbf{M}}$ the $n \times n$ sample correlation matrix. The eigenvalue decomposition is:

$$\widehat{\mathbf{M}} = \widehat{\mathbf{H}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{H}}' \tag{21}$$

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

$$\widehat{\mathbf{f}} = \widehat{\mathbf{H}}'\widetilde{\mathbf{x}}^d \tag{22}$$

where $\widetilde{\mathbf{x}}^d = (\widetilde{\mathbf{x}}^d_1, \ldots, \widetilde{\mathbf{x}}^d_T)'$ is the standardized data, $\widetilde{x}^d_{j,t} = (x^d_{j,t} - \overline{x}^d_j)/\widetilde{\sigma}_{x^d_j}\ \forall j = 1, \ldots, n$ where $\overline{x}^d_j = \frac{1}{T}\sum_{t=1}^{T} x^d_{j,t}$ and $\widetilde{\sigma}^2_{x^d_j} = \frac{1}{T}\sum_{t=1}^{T}(x^d_{j,t} - \overline{x}^d_j)^2$. When the principal components are estimated in-sample, $m = 0$, whereas $m = P_0, \ldots, P_2$ for recursive estimation of the principal components.

7.2 Forecasting models

The forecasting models are obtained by selection using *Autometrics* on the GUM:

$$\Delta y_t = \gamma_0 + \sum_{j=J_a}^{J_b} \rho_j \Delta y_{t-j} + \sum_{i=1}^{n}\sum_{j=J_a}^{J_b} \beta_{i,j}\Delta x_{i,t-j} + \sum_{k=1}^{n}\sum_{j=J_a}^{J_b} \gamma_{k,j} f_{k,t-j} + \sum_{l=1}^{T} \delta_l \mathbf{1}_{\{l=t\}} + \epsilon_t \tag{23}$$

where Δy_t is the first difference of log real gross domestic product. We set:
(i) $\boldsymbol{\gamma} = \mathbf{0}$, i.e. select over variables only;
(ii) $\boldsymbol{\beta} = \mathbf{0}$, select over factors only; and
(iii) $\boldsymbol{\gamma} \neq \mathbf{0}$ and $\boldsymbol{\beta} \neq \mathbf{0}$, i.e. jointly select variables and factors;
(iv) $\boldsymbol{\beta} = \mathbf{0}$, $n = 4$, $J_a = J_b = h$, i.e. the first four principal components only, with no selection;
intercepts and lags of Δy_t are included in all models, with intercepts always retained.

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

For the four forecasting specifications either:
(a) $\boldsymbol{\delta} = \mathbf{0}$, no IIS; or
(b) $\boldsymbol{\delta} \neq \mathbf{0}$, with IIS, applied in-sample.
The forecasting model is either selected and estimated over $t = 1, \ldots, T$ or recursively over the forecast horizon, $t = 1, \ldots, T+m$, including selecting the eigenvalues of the principal components, so the model specification can then change with each new forecast. Intercept-corrected forecasts are also computed, using the simplest form where the last in-sample residual is added to the forecast:

$$\Delta\widehat{y}^{IC}_{T+h+m} = \Delta\widehat{y}_{T+h+m|T+m} + \widehat{\epsilon}_{T+m} \quad \text{for } m = 0, \ldots, P_2.$$

Two selection strategies are considered; a conservative and a super-conservative strategy. The conservative strategy sets the significance level α so that $N\alpha \approx 4.4$ regressors are retained on average under the null that none are relevant; whereas the super-conservative strategy has a null retention of approximately 0.4 regressors. By controlling α, overfitting is not a concern despite commencing with $N \gg T$, with the cost of a loss of power for retaining regressors with retention test values close to the critical value. No selection is undertaken for the model PC1-4, other than IIS to which the conservative strategy significance level applies ($T\alpha \approx 1.5$). Table 2 summarizes the selection significance levels.

Three benchmark 'neither' forecasts are considered, the random walk (RW) and AR(1) forecasts computed directly and iteratively:

$$\begin{aligned}
\Delta\widehat{y}^{RW}_{T+h+m} &= \Delta y_{T+m} \\
\Delta\widehat{y}^{AR(D)}_{T+h+m} &= \widehat{\beta}_0 + \widehat{\beta}_1 \Delta y_{T+m} \\
\Delta\widehat{y}^{AR(I)}_{T+h+m} &= \sum_{i=0}^{h-1} \widehat{\gamma}_0 \widehat{\rho}_1^{i} + \widehat{\rho}_1^{h} \Delta y_{T+m}
\end{aligned}$$

for $m = P_0, \ldots, P_2$ and $h = 1, 4, 8$. As a result, there are 354 forecast models to compare. We evaluate the forecasts on root mean-square errors (RMSFEs), over the full sample and two subsamples, and for both levels and growth rates. The implied level forecasts for GDP (in logs) are:

$$\widehat{y}_{T+h+m|T+m} = \sum_{i=1}^{h} \Delta\widehat{y}_{T+i+m|T+m} + y_{T+m} \quad \text{for} \quad m = P_0, \ldots, P_2$$

for $h = 4, 8$. Although 1-step ahead forecast errors are identical for levels and differences, results are reported for comparison. 4-step forecasts are evaluated over from 2000:4 and 8-step forecasts are evaluated from 2001:4 as the h prior difference forecasts are required for computation, as evaluation in levels and growth rates need not result in the same ranking (see §2.7).

7.3 Results using PCs extracted from the whole dataset

Table 3 records the in-sample model fit and number of retained regressors for selection with IIS. At the looser significance level, selection over factors results in a better in-sample fit than selection over variables, while the ranking is reversed using the tighter significance level. The factor models retain a relatively large number of PCs under the conservative strategy, suggesting some overfitting, particularly at $h = 4$. The fit of the non-selected PC1-4 model is close to the fit for selecting over variables with the super-conservative strategy. Few dummies are retained on average.

Table 4 records the average RMSFE, trimmed RMSFE (trimming 10% in all samples), and average mean absolute error (MAE) for GDP and GDP growth for each of the forecasting models, averaged across: the forecast horizon, whether IIS is applied or not, the selection significance level, whether intercept correction is applied or not, and whether estimated in-sample or recursively. For GDP growth it is difficult to beat an AR(1) model, either iterative or direct, particularly in the earlier subsample ($P_0 : P_1$). In terms of selection over factors, variables, or both, the results generally favour selection over variables, although just using the first 4 PCs (PC1-4) dominates selection for both subperiods.

The rankings change dramatically for the levels forecasts. The AR(1) forecasts perform much worse over the second subsample, and PC1-4 is dominated by selection of factors or variables. The RW benchmark is preferred using the RMSFE criterion but is worse on trimmed RMSFE or MAE. Models with variables are preferred to factor models in both subsamples. There are huge differences in the forecast accuracy across the two subsamples reflecting the crisis period and the difficulty in forecasting GDP over this volatile period, emphasizing the role of location shifts.

The aggregate results using RMSFE are disentangled in figures 1 and 2 for GDP growth and log GDP respectively. Panel (a) averages across the variants for a given forecast horizon, panel (b) averages across all models with IIS and models without, panel (c) averages across the selection criterion, panel (d) averages across whether intercept correction was applied or not and panel (e) compares in-sample estimation versus recursive selection and estimation. The box plots record the subsample 1 and 2 average RMSFEs (where subsample 2 is always above subsample 1), with the dash denoting the full sample average.[3]

For GDP growth, RMSFE does not increase substantially as the forecast horizon grows. At the shortest horizon, the models with variables perform poorly in the second subsample relative to the first, whereas the factor model is less affected by greater turbulence of the second period. Using just the first four factors (PC1-4) is the dominant strategy. IIS yields more accurate forecasts when selection is over variables or 'both' for the recession period. A tighter selection strategy is preferred for all forecasting models, both in stable and volatile periods. The intercept-correction strategy yields more accurate factor model forecasts (both with and without selection) during the recession period, but little benefit during the earlier period. Recursive selection and estimation yield some gains.

In terms of forecasting the levels, the performance of all the models now deteriorates as the forecast horizon increases, as expected, but the stand-out finding is the gain from intercept correction for all models, especially in the second, more volatile, subperiod, so RMSFE-based evaluations of growth rates can hide the benefits of using a robust forecasting device. For the volatile subperiod, RMSFEs are approximately halved.

The reasons for the improvements in forecast accuracy from intercept correction are explored in figure 3, which records the h-step ahead forecasts of the levels of GDP for $h = 1, \dots, 8$. Panel (a) records the forecasts from the factor model without intercept correction, with recursive selection and estimation at the super-conservative significance level with IIS, panel (b) records the forecasts from the same model with intercept correction, and panels (c) and (d) record the forecasts from the corresponding variable

[3]Results for the 1-step ahead levels forecasts are recorded in figure 2 for comparison despite being identical to those in figure 1.

models. The benefits of intercept correction can be seen around the 2008/9 downturn, where the forecasts are pulled back on track. The simple correction is beneficial for both model's forecasts, indicating that the 'break' induced by the recession is the dominant feature affecting the forecast performance of the levels forecasts, and the choice of selecting over variables or factors is of secondary importance.

Figures 4 and 5 record the distributions of forecast errors for variables (panel a), factors (panel b), both (panel c) and the first 4 PCs (panel d), for GDP growth and the level of GDP, respectively. Separate distributions are plotted for the uncorrected and intercept-corrected forecasts. In growth rates, the forecast errors for the factor models are downward biased, but intercept correction corrects the bias. The variables and 'variable and factor' models contain some outliers resulting in very long tails. A closer examination reveals that the outlying forecast errors are mainly due to the retention of the second difference of the log monetary base as an explanatory variable. There was a dramatic increase in the monetary base following the financial crisis, which jumped from $863bn in 2008:3 to $1724bn in 2009:3. In practice intervention by the forecaster would likely attenuate such effects.

The levels forecasts demonstrate a substantial negative skew, and the benefits of intercept correction can be seen clearly in these distributions. Other than the couple of outliers in the variables, and variable and factor, models there are no major differences between the variable and factor model forecast errors.

7.4 Results using targeted factors and variables

Of interest is whether using targeted variables and factors as suggested by e.g., Bai and Ng (2008) affects the relative forecast performance of selection over factors, variables (or both). We use LASSO to select the 30 most important disaggregate variables, as described in section 6. Table 5 reports the significance levels used for selection after soft thresholding.

Table 6 reports in-sample summary statistics. The selected models are quite highly parameterized in many cases. For the targeted factors, IIS reduces the number of factors retained suggesting that more parsimonious specifications can be obtained if breaks and outliers are accounted for.

Table 7 reports the summary forecasting results for soft thresholding compared to standard selection, and Table 8 provides the breakdown by forecast horizon, where all the forecasts are based on fixed estimation schemes. Averaging across horizons (Table 7), there is little evidence that using targeted factors and variables leads to improvements across the board. For growth rates, selection over targeted factors improves on selection over (standard) factors, and selection over targeted factors and targeted variables improves on selection over factors and variables, but choosing the first 4 factors is the dominant factor-forecasting strategy. Table 8 shows that selection over targeted factors and targeted variables dominates selection over factors and variables at all horizons during the second forecast period, but that using the first 4 (standard) factors is the dominant strategy at $h = 4, 8$ for forecasting levels and growth rates.

7.4.1 Block-factor approach

In view of the good performance of the first 4 factors, we calculate forecasts using the block-factor approach of section 6. Without IIS, there is no selection, and with IIS there is selection only over the impulse-indicators at 1%, with the lagged dependent variable, intercept and four block factors (computed as the first factor from each block) always retained. The results are reported in Table 9, which averages over results obtained for fixed estimation, denoted 'Blocking 4 factors', allowing a direct comparison to 'First 4 factors' (the first 4 factors computed from the full variance-covariance matrix with 109 regressors). We also include the simple benchmark models.

The results in Table 9 generally are not supportive of the block-factor approach. Blocking does not lead to a forecast improvement for the growth rates. For the levels the rankings depend on the forecast error measure. Figure 6 plots the first four principal components from the full set of disaggregates and from blocking. There are some differences, particularly at the end of the sample for PC3, but the overall variation is similar.

8 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. Recent developments in automatic model selection now allow for more regressors than observations and perfect collinearities. This enables the set of regressors to be extended to include both factors, as measured by their static principal components, and variables, to be jointly included in regression models. The natural extension is to consider which methods perform best in a forecasting context, which is the objective of this paper.

One of the key explanations for forecast failure is that of location shifts. When the underlying data generating process shifts, but the forecasting model remains constant, forecast failure will often result. As both regression models 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 in-sample, and IIS could also be used to implement intercept corrections if an indicator variable was retained for the last in-sample observation. We find there is some advantage to using IIS for forecasting, as the unconditional mean is better estimated in differences. As the data are differenced to stationarity in order to estimate the principal components, few impulse-indicators are retained. Backing out levels forecasts highlights the non-stationarity due to level shifts, most notable over the 2008/9 recession, and a further extension would be to consider selection of the variables in levels, augmented by stationary principal components to capture underlying latent variable dynamics.

The empirical application considered GDP and GDP growth, computing forecasts using *Autometrics* to select forecasting models that include either principal components, individual variables, or both. When forecasting GDP growth, it is difficult to beat simple autoregressions, our 'neither' benchmarks. However, these naive benchmarks are poor at forecasting levels, when robust devices such as differencing (the random walk model) or intercept corrections are preferred. The empirical results are mixed, but suggest that selection over variables is preferable to selection over factors when breaks occur over the forecast horizon. Comparing Table 1 with that in Hendry and Mizon (2012) suggests this may be due to mean shifts in irrelevant variables that are given a non-zero weight in factors. There appears to be little empirical support for including both variables and factors jointly. The information set is identical between the two transformations of the data, but there is weak evidence to suggest that factor models are preferable for short horizons (nowcasting and 1-step ahead), but variable models are preferred at longer horizons during the second, volatile forecasting period. For direct multi-step forecasting, *Autometrics* selection over factors tends to forecast worse than imposing the first four factors, suggesting that there are no benefits to selecting the weights based on the correlation with y_{t+h}. While circumventing the need for off-line selection of factors, the empirical results suggest that this is of less importance than dealing with location shifts. The block-factor approach did not offer clear improvements relative to simply using the first four principal components, nor did selection over targeted factors and variables.

Whether the data are generated by latent factors or observable variables will depend on the phenomena being analyzed, but can be determined from the data using model selection techniques. Regardless of whether factor models or variable models are used for forecasting, the theory and evidence presented demonstrate the importance of robustifying the forecasts to location shifts.

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.

Bai, J., and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317.

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.

Bernanke, B. S., and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, **50**, 525–546.

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

Boivin, J., and Ng, S. (2006). Are more data always better for factor analysis?. *Journal of Econometrics*, *132*(1), 169–194.

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. (2011). 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. (2012a). Model selection in equations with many ‘small’ effects. *Oxford Bulletin of Economics and Statistics*, DOI: 10.1111/j.1468–0084.2012.00727.x.

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2012b). Model selection when there are multiple breaks. *Journal of Econometrics*, **169**, 239–246.

Castle, J. L., Fawcett, N. W. P., and Hendry, D. F. (2011). Forecasting Breaks and During Breaks. In Clements, and Hendry (2011), pp. 315–353.

Castle, J. L., and Hendry, D. F. (2011). Automatic selection of non-linear models. In Wang, L., Garnier, H., and Jackman, T.(eds.), *System Identification, Environmental Modelling and Control*, pp. 229–250. New York: Springer.

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. (2012a). Forecasting with vector autoregressive models of data vintages: US output growth and inflation. *International Journal of Forecasting*. Forthcoming.

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

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University 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). 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 *et al.* (2006), pp. 605–657.

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 *et al.* (2006), pp. 961–982.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?. *Journal of Econometrics*, **146**, 318–328.

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. (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.

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

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

Elliott, G., Granger, C. W. J., and Timmermann, A.(eds.)(2006). *Handbook of Econometrics on Forecasting*. Amsterdam: Elsevier.

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. (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset. *Journal of Business and Economic Statistics*, **27**, 468–479.

Fildes, R. A. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, **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., 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. (2012). Open-model forecast-error taxonomies. In Chen, X., and Swanson, N. R.(eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pp. 219–240. New York: Springer.

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

Howrey, E. P. (1978). The use of preliminary data in economic forecasting. *The Review of Economics and Statistics*, **60**, 193–201.

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. (2011). 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.

Moench, E., Ng, S., and Potter, S. (2009). Dynamic hierarchical factor models. Staff reports, no. 412, Federal Reserve Bank of New York.

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.

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.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, B*, **58**, 267–288.

Tinbergen, J. (1930). Determination and interpretation of supply curves: An example [Bestimmung und Deutung von Angebotskurven: ein Beispiel]. *Zeitschrift fur Nationalökonomie*, **1**, 669–679.

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.

Watson, M. W. (2007). How accurate are real-time estimates of output trends and gaps?. *Federal Reserve Bank of Richmond Economic Quarterly*, **93**, 143–161.

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.

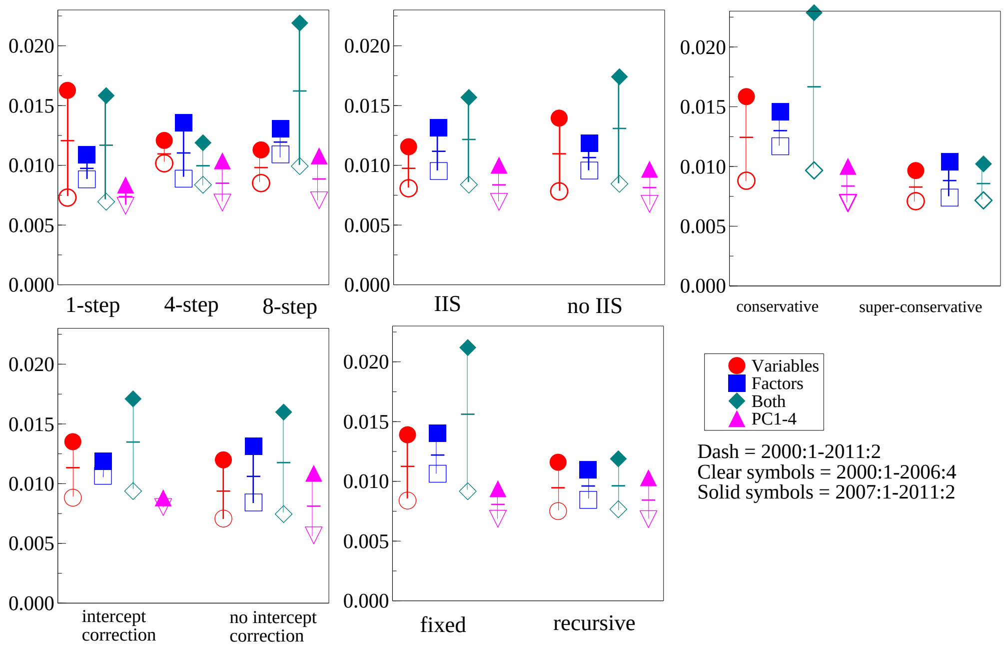

Figure 1: Average RMSFE for GDP growth $(\Delta\widehat{y}_{T+h})$

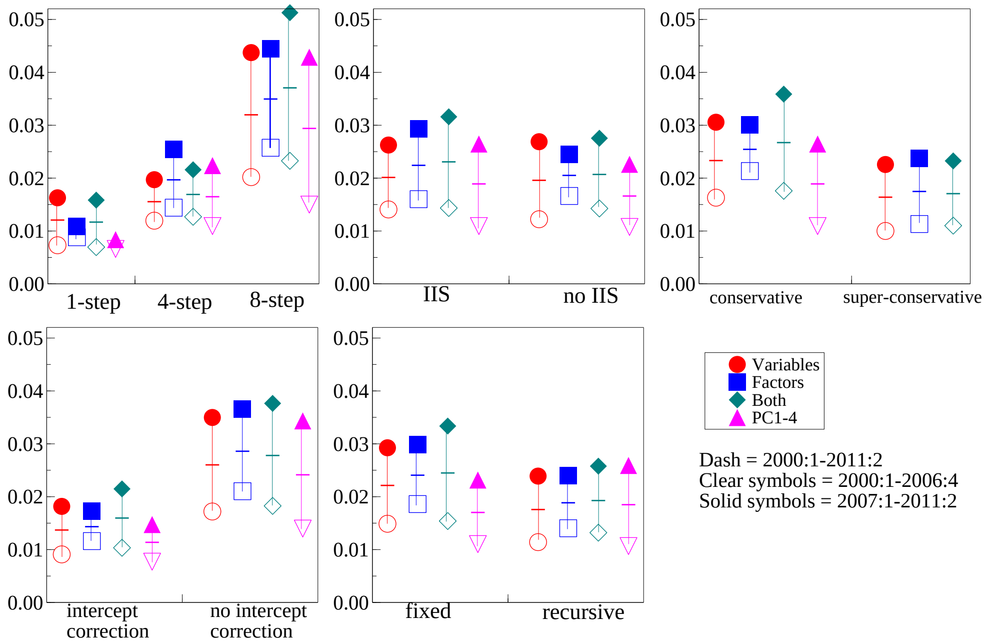

Figure 2: Average RMSFE for log GDP $(\widehat{y}_{T+h})$

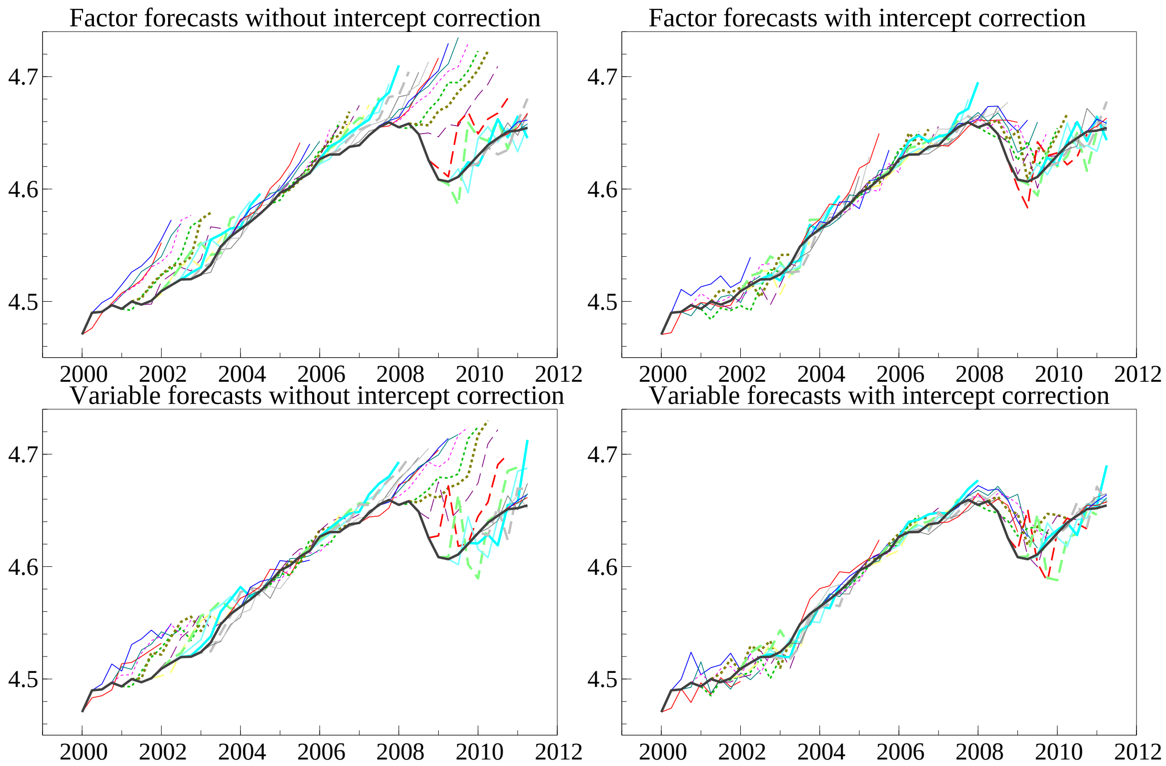

Figure 3: 1-step to 8-step ahead forecasts for GDP. Forecasts from factor models and variable models, both recursive super-conservative selection with IIS, with and without intercept correction.

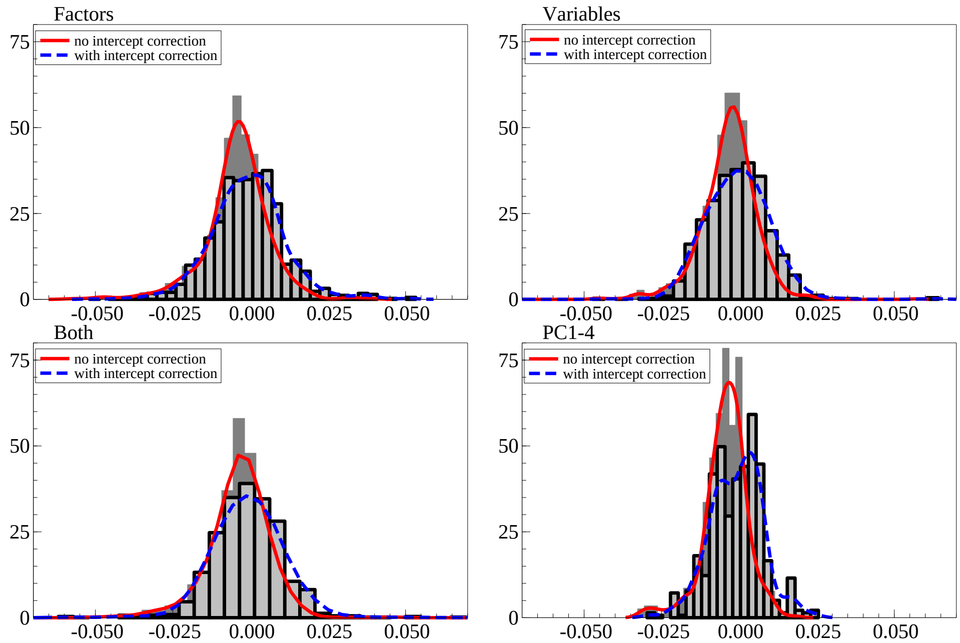

Figure 4: Distribution of forecast errors for GDP growth.

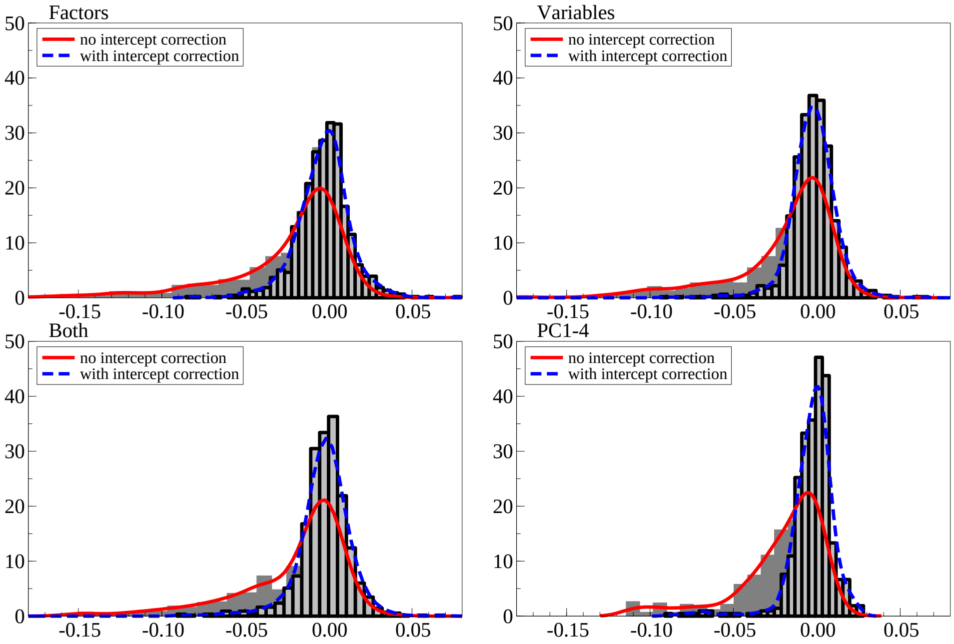

Figure 5: Distribution of forecast errors for log GDP.

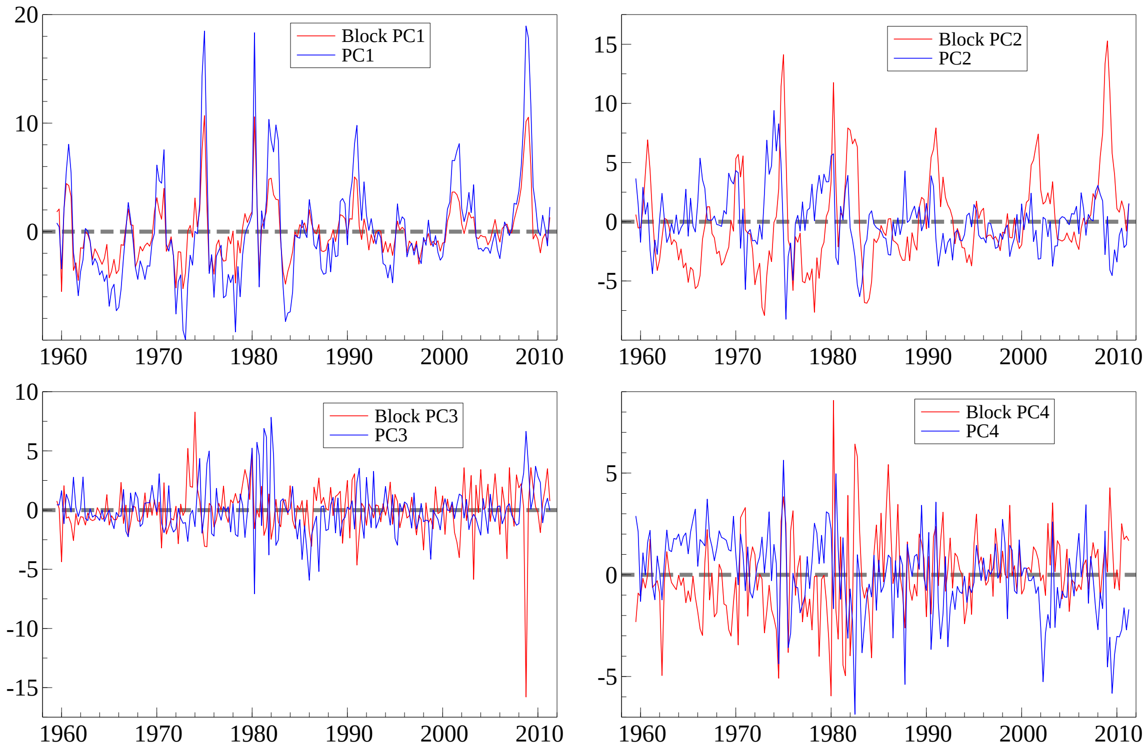

Figure 6: First 4 principal components from full set of disaggregates and from blocking.

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

Table 2: Significance levels used for model selection

	Variables		Factors		Both		PC1-4
	no IIS	IIS	no IIS	IIS	no IIS	IIS	IIS
N	441	591	441	591	877	1027	156
Conservative	1%	0.75%	1%	0.75%	0.5%	0.43%	1%
Super-conservative	0.1%	0.075%	0.1%	0.075%	0.05%	0.043%	-

Note: Intercepts are always retained in selection; PCs and LDV are retained in 'PC1-4' model.

Table 3: In-sample model fit for GDP growth forecasting models selected with IIS

	1-step		4-step		8-step	
	cons	super	cons	super	cons	super
Variables						
$\widehat{\sigma}$	0.49%	0.59%	0.58%	0.69%	0.51%	0.77%
No. regressors	15	6	16	8	26	6
No. dummies	2	2	5	1	7	3
Factors						
$\widehat{\sigma}$	0.40%	0.62%	0.33%	0.74%	0.50%	0.79%
No. regressors	24	7	35	6	27	5
No. dummies	6	2	11	4	5	2
Both						
$\widehat{\sigma}$	0.46%	0.64%	0.60%	0.74%	0.49%	0.69%
No. regressors	17	5	15	6	20	7
No. factors	2	1	4	0	3	2
No. dummies	4	1	4	4	13	4
PC1-4						
$\widehat{\sigma}$		0.59%		0.69%		0.74%
No. dummies		4		6		5

Notes: $\widehat{\sigma}$ is the equation standard error, No. regressors and No. dummies record the number of regressors (including the intercept) and, as a subset, the number of dummies retained, and 'cons' and 'super' are the conservative and super-conservative strategies respectively.

Table 4: Forecast-error outcomes

	Variables	Factors	Both	PC1-4	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{T+k}$							
	1.036	1.091	1.262	0.825	0.967	0.811	0.811
Full sample	0.697	0.806	0.741	0.588	0.700	0.495	0.489
	0.757	0.849	0.849	0.634	0.746	0.551	0.545
	0.795	0.954	0.842	0.686	0.768	0.545	0.551
2000:1-2006:4	0.644	0.758	0.669	0.536	0.619	0.427	0.431
	0.647	0.771	0.683	0.556	0.625	0.421	0.423
	1.275	1.250	1.654	0.984	1.191	1.101	1.097
2007:1-2011:2	0.865	0.923	0.946	0.714	0.909	0.692	0.677
	0.929	0.970	1.107	0.754	0.935	0.753	0.736
$\widehat{y}_{T+k}$							
	2.138	2.345	2.350	2.668	1.965	2.693	2.681
Full sample	1.505	1.767	1.611	1.380	1.977	1.725	1.712
	1.555	1.786	1.677	1.433	1.958	1.851	1.826
	1.400	1.745	1.507	1.828	1.156	1.289	1.310
2000:1-2006:4	1.134	1.461	1.216	0.942	1.379	0.997	1.013
	1.077	1.376	1.141	0.894	1.314	0.955	0.950
	2.873	2.967	3.187	3.554	2.750	3.978	3.947
2007:1-2011:2	2.286	2.482	2.472	2.315	3.083	3.318	3.282
	2.299	2.423	2.512	2.271	2.959	3.245	3.188

Notes: The three rows in each block correspond to (a) RMSFE; (b) trimmed RMSFE with 10% trimming; and (c) MAE for GDP and quarterly GDP growth, with benchmark Random Walk, direct AR(1) [AR(D)] and iterative AR(1) [AR(I)] forecasts. $(\times 100)$.

Table 5: Soft thresholding and selection significance levels

	Targetted factors	Targetted variables	Targetted factors & targetted variables	Targetted factors & all variables
Soft threshold	30 vars	30 vars	30 vars	30 vars
Selection with IIS	1%	1%	0.5%	0.5%
Selection without IIS	5%	5%	2.5%	1%

Table 6: In-sample summary statistics for soft thresholding using LASSO.

			$\widehat{\sigma}$	N	F^*	n	d	f
Targetted factors	1-step	IIS	0.501%	274	30	13	4	8
		no IIS	0.465%	124	30	25	-	24
	4-step	IIS	0.601%	274	30	12	8	3
		no IIS	0.642%	124	30	13	-	12
	8-step	IIS	0.512%	274	30	23	14	8
		no IIS	0.593%	124	30	28	-	25
Targetted variables	1-step	IIS	0.526%	274	30	14	2	-
		no IIS	0.463%	124	30	34	-	-
	4-step	IIS	0.605%	274	30	15	8	-
		no IIS	0.641%	124	30	18	-	-
	8-step	IIS	0.371%	274	30	41	28	-
		no IIS	0.710%	124	30	19	-	-
Targetted factors & targetted variables	1-step	IIS	0.492%	394	30	14	2	4
		no IIS	0.513%	244	30	14	-	6
	4-step	IIS	0.638%	394	30	9	4	2
		no IIS	0.665%	244	30	8	-	4
	8-step	IIS	0.500%	394	30	23	14	6
		no IIS	0.639%	244	30	15	-	5
Targetted factors & all variables	1-step	IIS	0.513%	710	30	11	1	9
		no IIS	0.577%	560	30	8	-	5
	4-step	IIS	0.641%	710	30	11	3	2
		no IIS	0.670%	560	30	11	-	3
	8-step	IIS	0.453%	710	30	28	9	7
		no IIS	0.595%	560	30	18	-	5

Notes: N = number of regressors in GUM (excluding the retained intercept); F^* = number of variables retained after soft thresholding, fixed at 30; n = number of retained regressors after selection with *Autometrics*; d = number of impulse indicators retained after IIS; f = number of principal components retained after selection, including lags.

Table 7: Summary forecast results for soft thresholding compared to standard selection.

	Factors	Target factors	Variables	Target variables	Factors & variables	Target factors & target variables	Target factors & all variables	First 4 factors	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{t+h}$											
Full sample	1.220	1.064	1.125	1.002	1.562	1.018	0.982	0.807	0.967	0.811	0.811
	0.916	0.823	0.759	0.770	0.804	0.759	0.758	0.582	0.700	0.495	0.489
	0.957	0.851	0.829	0.794	0.970	0.789	0.779	0.626	0.746	0.551	0.545
2000:1-2006:4	1.064	0.903	0.840	0.894	0.917	0.840	0.870	0.688	0.768	0.545	0.551
	0.846	0.735	0.684	0.724	0.729	0.673	0.711	0.538	0.619	0.427	0.431
	0.862	0.746	0.689	0.726	0.746	0.673	0.711	0.560	0.625	0.421	0.423
2007:1-2011:2	1.404	1.256	1.390	1.127	2.120	1.225	1.120	0.938	1.191	1.101	1.097
	1.065	1.005	0.999	0.879	1.056	0.981	0.864	0.690	0.909	0.692	0.677
	1.104	1.014	1.046	0.900	1.317	0.968	0.885	0.728	0.935	0.753	0.736
$\widehat{y}_{t+h}$											
Full sample	2.641	2.615	2.398	2.300	2.635	2.376	2.431	1.883	2.668	2.693	2.681
	2.033	2.150	1.725	1.779	1.764	1.847	1.945	1.371	1.977	1.725	1.712
	2.057	2.268	1.800	1.870	1.890	1.960	2.056	1.422	1.958	1.851	1.826
2000:1-2006:4	2.019	2.245	1.611	1.622	1.616	1.779	2.030	1.182	1.828	1.289	1.310
	1.733	1.967	1.355	1.383	1.352	1.492	1.750	0.973	1.379	0.997	1.013
	1.637	2.034	1.285	1.422	1.279	1.548	1.817	0.927	1.314	0.955	0.950
2007:1-2011:2	3.294	3.001	3.170	2.870	3.604	2.946	2.845	2.597	3.554	3.978	3.947
	2.780	2.607	2.583	2.482	2.729	2.538	2.429	2.191	3.083	3.318	3.282
	2.709	2.572	2.600	2.416	2.841	2.472	2.368	2.192	2.959	3.245	3.188

Notes: The three rows in each block correspond to (a) RMSFE; (b) trimmed RMSFE with 10% trimming; and (c) MAE for GDP and quarterly GDP growth, with benchmark Random Walk, direct AR(1) [AR(D)] and iterative AR(1) [AR(I)] forecasts. Averaged over fixed estimation. ($\times 100$).

Table 8: RMSFE for soft thresholding and standard selection by forecast horizon (fixed estimation)

		Factors	Target factors	Variables	Target variables	Factors & variables	Target factors & target variables	Target factors & all variables	First 4 factors	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{t+h}$												
Full sample	1-step	1.012	0.996	1.363	0.999	1.403	0.991	0.864	0.737	0.753	0.729	0.729
	4-step	1.214	0.965	0.963	0.897	1.019	0.918	0.942	0.819	1.045	0.834	0.851
	8-step	1.436	1.232	1.049	1.109	2.263	1.145	1.140	0.865	1.104	0.869	0.852
2000:1-2006:4	1-step	0.908	0.843	0.754	0.969	0.713	0.815	0.767	0.665	0.766	0.560	0.560
	4-step	0.974	0.743	0.842	0.798	0.874	0.818	0.781	0.685	0.681	0.536	0.546
	8-step	1.310	1.123	0.923	0.915	1.164	0.887	1.060	0.713	0.858	0.537	0.547
2007:1-2011:2	1-step	1.141	1.185	1.863	1.039	1.956	1.201	0.991	0.825	0.731	0.934	0.934
	4-step	1.497	1.216	1.118	1.025	1.197	1.052	1.137	0.967	1.439	1.153	1.177
	8-step	1.573	1.368	1.189	1.319	3.207	1.424	1.231	1.023	1.404	1.217	1.179
$\widehat{y}_{t+h}$												
Full sample	1-step	1.012	0.996	1.363	0.999	1.403	0.991	0.864	0.737	0.753	0.729	0.729
	4-step	2.764	1.699	1.863	1.697	1.983	1.743	1.974	1.935	2.922	2.615	2.719
	8-step	4.146	5.149	3.968	4.202	4.518	4.395	4.455	2.977	4.330	4.735	4.593
2000:1-2006:4	1-step	0.908	0.843	0.754	0.969	0.713	0.815	0.767	0.665	0.766	0.560	0.560
	4-step	1.843	1.311	1.412	1.230	1.437	1.420	1.508	1.182	1.534	1.250	1.290
	8-step	3.306	4.580	2.668	2.667	2.697	3.101	3.815	1.699	3.184	2.057	2.080
2007:1-2011:2	1-step	1.141	1.185	1.863	1.039	1.956	1.201	0.991	0.825	0.731	0.934	0.934
	4-step	3.737	2.113	2.365	2.183	2.538	2.104	2.470	2.713	4.262	3.878	4.038
	8-step	5.003	5.706	5.281	5.387	6.318	5.533	5.075	4.252	5.669	7.121	6.869

Note: ($\times 100$).

Table 9: Summary forecast results for blocking compared to including first four factors

	Blocking 4 factors	First 4 factors	RW	AR(D)	AR(I)
$\Delta\widehat{y}_{t+h}$					
Full sample	0.859	0.807	0.967	0.811	0.811
	0.603	0.582	0.700	0.495	0.489
	0.646	0.626	0.746	0.551	0.545
2000:1-2006:4	0.696	0.688	0.768	0.545	0.551
	0.541	0.538	0.619	0.427	0.431
	0.553	0.560	0.625	0.421	0.423
2007:1-2011:2	1.049	0.938	1.191	1.101	1.097
	0.759	0.690	0.909	0.692	0.677
	0.791	0.728	0.935	0.753	0.736
$\widehat{y}_{t+h}$					
Full sample	1.984	1.883	1.965	2.693	2.681
	1.439	1.371	1.977	1.725	1.712
	1.555	1.422	1.958	1.851	1.826
2000:1-2006:4	1.302	1.182	1.156	1.289	1.310
	1.067	0.973	1.379	0.997	1.013
	1.084	0.927	1.314	0.955	0.950
2007:1-2011:2	2.595	2.597	2.750	3.978	3.947
	2.148	2.191	3.083	3.318	3.282
	2.150	2.192	2.959	3.245	3.188

Notes: The three rows in each block correspond to (a) RMSFE; (b) trimmed RMSFE with 10% trimming; and (c) MAE for GDP and quarterly GDP growth, with benchmark Random Walk, direct AR(1) [AR(D)] and iterative AR(1) [AR(I)] forecasts. Averaging over fixed estimation results. ($\times 100$).