ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

FORECASTING BY FACTORS, BY VARIABLES, OR BOTH?

Jennifer L. Castle, Michael P. Clements and David F. Hendry

Number 600
April 2012

Manor Road Building, Oxford OX1 3UQ

Forecasting by factors, by variables, or both?

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.

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 Institute 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. 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, denoted EWMA, the closely related Holt–Winters approach (see Holt, 1957, and Winters, 1960), damped trend (see e.g., Fildes, 1992), and autoregressions, including the general time-series approach in Box and Jenkins (1970). Some members of this class were often found to dominate in forecasting competitions: see Makridakis, Andersen, Carbone, Fildes *et al.* (1982) and Makridakis and Hibon (2000).

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

The structure of the paper is as follows. Section 2 describes some of the issues that arise in any analysis of forecasting models or methods. Section 3 describes the statistical framework we use to analyse forecasting with factors or variables. Section 4 develops the analysis of forecasting from factor models with a taxonomy of sources of forecast errors in the empirically relevant case of non-stationary processes. Section 5 addresses the problem of systematic forecast failure to which equilibrium-correction formulations are prone in the face of location shifts. Section 6 discusses model selection with both factors and variables, and section 7 illustrates the analysis using US GDP forecasts. Section 8 concludes.

2 Setting the scene

Many interacting issues need addressed when analysing forecasting, the complexity of which mean 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.

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 (2011a) 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 data generating process (denoted LDGP), namely the DGP for the set of variables under consideration (see e.g., Hendry, 2009) by formulating a general unrestricted model (GUM) that nests the LDGP, and checking its congruence when feasible (estimable once $N \ll T$ and perfect collinearities are removed). Search thereafter ensures congruence, so all selected models are valid restrictions of the GUM, and should parsimoniously encompass the feasible GUM. Location shifts are removed in-sample by impulse-indicator saturation (IIS: see Hendry, Johansen and Santos, 2008, Johansen and Nielsen, 2009, and the simulation studies in Castle, Doornik and Hendry, 2011c), which also addresses possible outliers. Thus, if $\left\{1_{\{j=t\}}, t = 1, \ldots, T\right\}$ denotes the complete set of T impulse indicators, we allow for $\mathbf{f}_t, \ldots \mathbf{f}_{t-s}$, $\mathbf{z}_t, \ldots \mathbf{z}_{t-s}$ and $\left\{1_{\{j=t\}}, t = 1, \ldots, T\right\}$ all being included in the initial set of candidate variables to which multi-path search is applied. 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 (2011b). Here we are concerned with the application of models selected in this way to a forecasting context when the DGP is non-stationary due to structural breaks. Since there 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 6 discusses automatic model selection further.

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 are conflated with robustness to location shifts as most of the parsimonious models evaluated 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. However, the principles of robust-model formulation discussed in Clements and Hendry (2005b) 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 5 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), 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 (2011) provide some evidence on the efficacy of these strategies for forecasting US output growth and inflation, albeit using information sets consisting only of lags (and different vintage estimates) of the variable being forecast.

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, idea of stopping the model estimation period early, and attempting to predict the 'final' estimates of the most recent data), as well as during turbulent periods (Swanson and van Dijk, 2006), which might favour factor models over other approaches that do not explicitly attempt to take data revisions into account.

[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.

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, 1993b, 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 in 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 different forecasting models, the choice of levels versus differences can also 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 (1995) 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 of VARs in differences (for cointegrated systems) is attenuated when forecasts of growth rates are evaluated. When there are structural breaks, the evaluation of forecasts of growth rates may cloak the benefits of a better forecasting model, such as an intercept-corrected forecasting model as the benefits of using a robust forecasting device are potentially larger for 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.

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

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

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

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

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

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

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

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

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):

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

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

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

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:

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

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

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

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

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

Consequently, from (2) and (6):

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

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

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

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

3.2 Variable-based and factor-based models

Suppose the in-sample DGP for y_t is:

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

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

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

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

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

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

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

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 (7) and (11) we can obtain an equivalent representation to (11) in factor space:

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

Again, only a subset may matter, namely the $\mathbf{f}_{1,t-1}$ in (8), and the resulting parsimonious model in the space of relevant factors becomes:

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

where η_t need not be an innovation process against the omitted information: also, (12) and (14) are not equivalent representations in general even though (11) and (13) are.

Finally, we allow the possibility that when both variables and their principal components are allowed, some of the $\mathbf{z}_{a,t-1}$ and some of the $\mathbf{f}_{1,t-1}$ are retained to provide closer, yet more parsimonious, approximations to the behavior of y_t in-sample. In practice, there may well have been location shifts and outliers in-sample, so we also allow for IIS during model selection. Thus, a vector of deterministic terms (such as intercepts, location shifts, and indicator variables) denoted $\mathbf{q}_t$ with $\mathbf{Q}_t^1 = (\mathbf{q}_1 \ldots \mathbf{q}_t)$ is allowed, as well as longer lags, so the sequential conditional expectation of y_t at time t is denoted $\mathsf{E}_t[y_t|\mathbf{Z}^1_{t-1}, \mathbf{Y}^1_{t-1}\mathbf{q}_t]$ (when that exists).

An important special case is when the DGP for y_t is a simple autoregressive process, so that none of the $z_{i,t-1}$ have a role to play. When y_t is just an AR(1), say, then:

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

Searching over (or modeling) factors alone might lead to the retention of a large number of the elements of $\mathbf{f}_t$ to approximate y_{t-1}, especially if the $\mathbf{z}_t$ include y_t. When searching over variables, or both variables and factors, then the starting model includes y_{t-1}, and so should allow for simpler models that more closely resemble dynamic DGPs.

4 A forecast-error taxonomy for factor models

The aim is to forecast the scalar $\{y_{T+h}\}$ over a forecast horizon $h = 1, \ldots, H$, from a forecast origin at T, at which point the information set consists of $\mathbf{Z}^1_T = (\mathbf{z}_1 \ldots \mathbf{z}_T)$ and $(y_1 \ldots y_T)$. Forecast accuracy is to be judged by a criterion function $C_e\left(\widehat{u}_{T+1|T} \ldots \widehat{u}_{T+H|T}\right)$, which we take to depend only on the forecast errors $\widehat{u}_{T+h|T} = y_{T+h} - \widehat{y}_{T+h|T}$, and specialize further to squared-error loss in the empirical application, where we also consider the possible dependence of evaluation outcomes on the transformation of y_{T+h} being forecast.

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 (14):[2]

$$\widehat{y}_{T+1|T} = \widehat{\delta} + \widehat{\boldsymbol{\tau}}'_1\left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1\right) + \widehat{\rho}\left(\widehat{y}_T - \widehat{\delta}\right) \tag{15}$$

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

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

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

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

$$y_{T+h} = \delta^* + \boldsymbol{\beta}'\left(\mathbf{z}_{T+h-1} - \boldsymbol{\mu}^*\right) + \rho\left(y_{T+h-1} - \delta^*\right) + \epsilon_{T+h} \tag{17}$$

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

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

[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.

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

We derive the 1-step forecast-error taxonomy, which highlights the key factors, and allows us to separately distinguish 11 sources of forecast error. Calculating the forecast error as (18) minus (15), for $h = 1$, gives rise to:

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

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

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

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

Neglecting terms of $\mathsf{O}_p\left(T^{-1}\right)$ (including finite-sample biases in parameter estimates) to focus on the main sources, and taking expectations:

$$\mathsf{E}\left[\widehat{u}_{T+1|T}\right] \simeq (1-\rho)\left(\delta^* - \delta\right) - \boldsymbol{\tau}'\left(\boldsymbol{\kappa}^* - \boldsymbol{\kappa}\right) + \rho\left(y_T - \mathsf{E}[\widehat{y}_T]\right) + \boldsymbol{\tau}_1'(\mathbf{f}_{1,T} - \mathsf{E}[\widehat{\mathbf{f}}_{1,T}]) \tag{19}$$

which indicates that sources [A] and [B] in table 1 are the primary determinants of forecast bias, although data mismeasurement and factor estimation errors ([E] and [F]) also contribute. These last two and all the remaining terms contribute to the forecast-error variance. The factor approximation error does not enter (19) as $\mathsf{E}\left[\mathbf{f}_{2,T}\right] = \boldsymbol{\kappa}_2$. Even when [E] and [F] are negligible, the equilibrium-mean and factor-mean shifts could be large. For example, if in (1):

$$\boldsymbol{\pi}^* = \boldsymbol{\pi} + 1_{(t \geq T)}\boldsymbol{\theta} \;\text{ for }\; h = 1, \ldots, H \tag{20}$$

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

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

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

$$\boldsymbol{\kappa}^* = \boldsymbol{\Psi}^{-1}\left(\mathbf{I}_n - \boldsymbol{\Pi}\right)^{-1}\boldsymbol{\pi}^* = \boldsymbol{\kappa} + 1_{(t \geq T)}\boldsymbol{\Psi}^{-1}\left(\mathbf{I}_n - \boldsymbol{\Pi}\right)^{-1}\boldsymbol{\theta} \tag{21}$$

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

Consider now the possibility that $\boldsymbol{\tau}$ and ρ change value for the forecast period, so that in place of (18) the DGP is given by:

$$y_{T+1} = \delta^* + \boldsymbol{\tau}^{*\prime} \left(\mathbf{f}_T - \boldsymbol{\kappa}^* \right) + \rho^* \left(y_T - \delta^* \right) + \epsilon_{T+1} \tag{22}$$

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

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

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, it is illuminating to consider the principal sources of forecast error for an $\mathrm{AR}(1)$ model, as this model serves as the benchmark against which the *selected* factor-and-variable models in section 7 are to be compared. For the sake of brevity, we ignore factors of secondary importance, such as parameter estimation uncertainty and data mis-measurement, and construct the forecast error for the $\mathrm{AR}(1)$:

$$y_t = \delta + \alpha \left(y_{t-1} - \delta \right) + v_t \tag{23}$$

when the forecast period DGP is given by (18). Notice that the omission of the factors will typically change the autoregressive parameter α, so that α need not equal ρ, but the long-run mean is the in-sample period value of δ. Denoting the forecast error from the $\mathrm{AR}(1)$ model by $\widehat{v}_{T+1|T}$, we obtain:

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

with a forecast bias of:

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

matching the two leading terms in (19) for the bias of the factor-forecasting model. Hence whether we include the 'correct' set of factors, a subset of these, or none at all will have no effect on the bias of the forecasts (at the level of abstraction 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).

5 The equilibrium-correction problem

Section 4 assumes a single forecast origin, but forecasting is rarely viewed as a one-off venture, and of interest is the performance of the competing models as the origin moves through time. Although all models will fail when there is a location shift which is unknown when the forecast is made, 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 (15) belongs, is their lack of adaptability over time. To see this, note that (15) could be rewritten for 1-step forecasts as:

$$\Delta \widehat{y}_{T+1|T} = \widehat{\boldsymbol{\tau}}_1^{\prime} \left(\widehat{\mathbf{f}}_{1,T} - \widehat{\boldsymbol{\kappa}}_1 \right) + \left(\widehat{\rho} - 1 \right) \left(\widehat{y}_T - \widehat{\delta} \right)$$

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

$$\Delta y_{T+1} = \boldsymbol{\tau}'\left(\mathbf{f}_T - \boldsymbol{\kappa}^*\right) + (\rho - 1)\left(y_T - \delta^*\right) + \epsilon_{T+1} \tag{24}$$

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

$$\mathsf{E}\left[\Delta y_{T+1}\right] = \boldsymbol{\tau}'\mathsf{E}\left[\mathbf{f}_T - \boldsymbol{\kappa}^*\right] + (\rho - 1)\,\mathsf{E}\left[y_T - \delta^*\right] = \boldsymbol{\tau}_1'\left(\boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_1^*\right) + (\rho - 1)\left(\delta - \delta^*\right) \tag{25}$$

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

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

This generic difficulty for EqCMs suggests using a robust forecasting device approach which exploits (25), as in:

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

Again, under the simplifying assumptions (known in-sample parameters, accurate data, no approximation error), and denoting the forecast error by $\Delta\widetilde{\epsilon}_{T+2|T+1} = \Delta y_{T+2} - \Delta\widetilde{y}_{T+2|T+1}$, using (24) gives:

$$\begin{aligned}\Delta\widetilde{\epsilon}_{T+2|T+1} &= \boldsymbol{\tau}'\left(\mathbf{f}_{T+1} - \boldsymbol{\kappa}^*\right) + (\rho - 1)\left(y_{T+1} - \delta^*\right) + \epsilon_{T+2} - \Delta y_{T+1} - \boldsymbol{\tau}_1'\Delta\mathbf{f}_{1,T} - (\rho - 1)\,\Delta y_T \\ &= \Delta y_{T+2} - \Delta y_{T+1} - \left(\boldsymbol{\tau}_1'\Delta\mathbf{f}_{1,T} + (\rho - 1)\,\Delta y_T\right) \end{aligned} \tag{26}$$

which is less dependent on the location shifts.

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

However, breaks due to location shifts within any EqCM forecasting model will induce systematic mis-forecasting, and the above analysis applies equally to factor-based models (as illustrated in section 4). In the empirical forecasting exercise in section 7 below, the variables are already differenced once, so large shifts in equilibrium means are unlikely, and hence such formulations already embody a partial robustness to previous location shifts. Indeed, if in place of (23), the differenced-data version is used, then forecasting $T+2$ from $T+1$:

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

when:

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

we have:

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

which is close to (26).

6 Automatic Model Selection

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

One advantage of using an automatic model selection algorithm is that it enables us to remain agnostic 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 be used to forecast future outcomes. On the other hand, if the data are generated by individual disaggregated economic variables then these should form the forecasting model. By including both explanations jointly, the data can determine the most plausible structure.

A further advantage of model selection is that 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.*, 2011b). 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 tend not to be the first few PCs, so the correlation structure may differ from that between the dependent variable and the disaggregates.

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 (see Doornik, 2008). 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).

7 Forecasting US GDP and GDP growth

Our empirical forecasting exercise compares the forecast performance of regression models based on principal components, variables, or both. We forecast quarterly GDP growth over the period 2000–2011, as well as considering the corresponding level forecasts for GDP. Models are selected in-sample using *Autometrics*, with all variables and their 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. One of the key studies is Stock and Watson (2009). They find 'considerable evidence of instability in the factor model; the indirect evidence suggests instability in all elements (the factor loadings, the factor dynamics, and the idiosyncratic dynamics).' (Stock and Watson, 2009, p.197). They suggest estimating the factors on the full historical period across the break (here, 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 IIS.

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

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, but is updated here to 2011:2. There are $n = 109$ disaggregates, used both as the candidate set of regressors and the set of variables to form 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.

Denote h as the $h-$step ahead direct forecast, where $h = 1, 4, 8$. 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 $\mathbf{x}^d$ $(= \Delta\mathbf{x})$ denote the $(T + m \times n)$ matrix of disaggregated variables which have been transformed to non-integrated by appropriate differencing, and $\widehat{\mathbf{M}}$ the $n \times n$ sample correlation matrix. The eigenvalue decomposition is:

$$\widehat{\mathbf{M}} = \widehat{\mathbf{H}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{H}}' \tag{27}$$

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 computed as:

$$\widehat{\mathbf{f}} = \widehat{\mathbf{H}}'\widetilde{\mathbf{x}^d} \tag{28}$$

where $\widetilde{\mathbf{x}^d} = \left(\widetilde{\mathbf{x}}_1^d, \ldots, \widetilde{\mathbf{x}}_T^d\right)'$ is the standardized data, $\widetilde{x}_{j,t}^d = \left(x_{j,t}^d - \overline{x}_j^d\right)/\widetilde{\sigma}_{x_j^d}\ \forall j = 1, \ldots, n$ where $\overline{x}_j^d = \frac{1}{T}\sum_{t=1}^T x_{j,t}^d$ and $\widetilde{\sigma}_{x_j^d} = \left[\frac{1}{T}\sum_{t=1}^T \left(x_{j,t}^d - \overline{x}_j^d\right)^2\right]^{1/2}$. 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 undertaking selection on the general unrestricted model (GUM):

$$\Delta y_t = \gamma_0 + \sum_{j=J_a}^{J_b} \rho_j \Delta y_{t-j} + \sum_{i=1}^{n}\sum_{j=J_a}^{J_b} \beta_{i,j}\Delta x_{i,t-j} + \sum_{k=1}^{n}\sum_{j=J_a}^{J_b} \gamma_{k,j} f_{k,t-j} + \sum_{l=1}^{T} \delta_l \mathbf{1}_{\{l=t\}} + \epsilon_t \tag{29}$$

where Δy_t is the first difference of log real gross domestic product.

Forecasting models are obtained by undertaking selection on (29) using *Autometrics* where 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;
where the intercept and lags of the dependent variable are included in all models. The intercept is forced in all specifications, i.e. not selected over.

Three forecast horizons are recorded, including 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.

Either the forecasting model is selected and estimated over $t = 1, \ldots, T$; or the forecasting model is selected and estimated recursively over the forecast horizon, $t = 1, \ldots, T+m$, including the eigenvalues of the principal components, so the model specification can change with each new forecast.

Intercept-corrected forecasts are also computed. The simplest form of intercept correction is used, whereby 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 aims to retain 4.4 regressors on average under the null that no regressors are relevant ($N\alpha \approx 4.4$) and the super-conservative strategy gives a null retention of approximately 0.4 regressors. Hence, overfitting is not a concern despite commencing with $N \gg T$. Parsimony can be achieved by controlling the significance level, with the cost a loss of power for regressors with a significance level close to the critical value. No selection is undertaken for the model PC1-4, other than IIS for which the conservative strategy significance level refers ($T\alpha \approx 1.5$). Table 2 summarizes the selection significance levels.

	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%	-

Table 2: Significance levels used for model selection.
Notes: Intercept is forced in selection; PCs and LDV is forced in 'PC1-4' model.

Three benchmark 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 to avoid the transformation bias) are computed as:

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

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. Evaluation in levels and growth rates need not result in the same ranking, because the RMSFE criterion is not invariant to non-singular, scale preserving, linear transformations: see Clements and Hendry (1993b). Further, in section 2.7, we argued that evaluation in terms of growth rates could downplay the benefit of using a robust forecasting device when there are breaks.

7.3 Results

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.

	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

Table 3: In-sample model fit for GDP growth forecasting models selected with IIS: $\widehat{\sigma}$ = 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 records the average RMSFE, trimmed RMSFE (trimming 10% for the full sample and two subsamples), 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. That said, always just using the first 4 PCs (PC1-4) is the dominant strategy, compared to 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 criteria. Variable models 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.

	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

Table 4: 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. $(\times 100)$.

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]

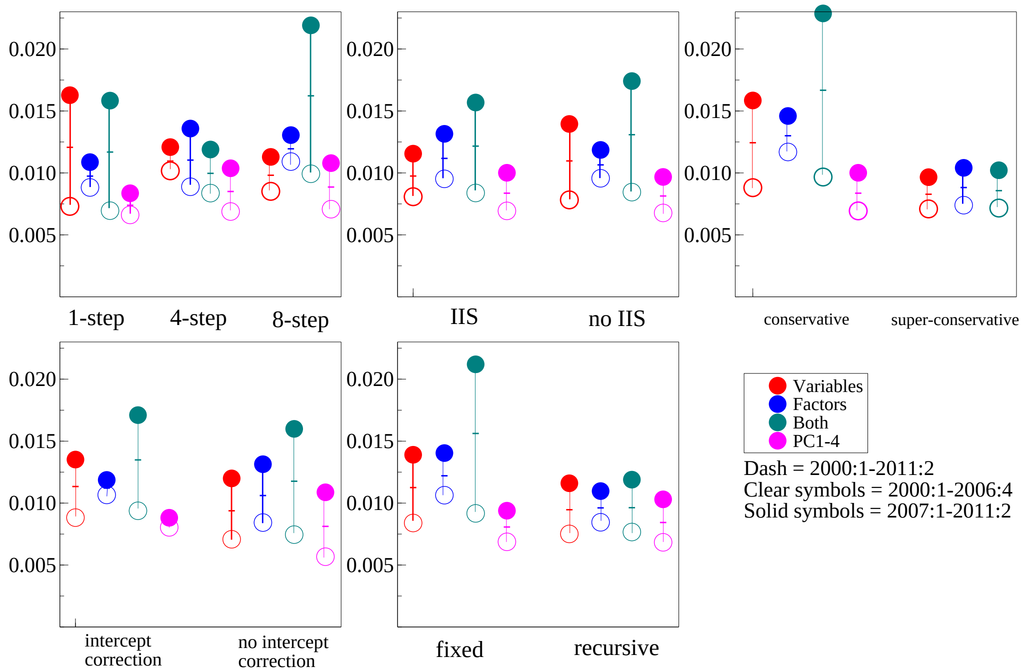

Figure 1: Average RMSFE for GDP growth $(\Delta\widehat{y}_{T+h})$

For GDP growth, RMSFE does not increase substantially as the forecast horizon grows. At the shortest horizon, the variable models 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 $h = 1, \ldots, 8$. 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 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.

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

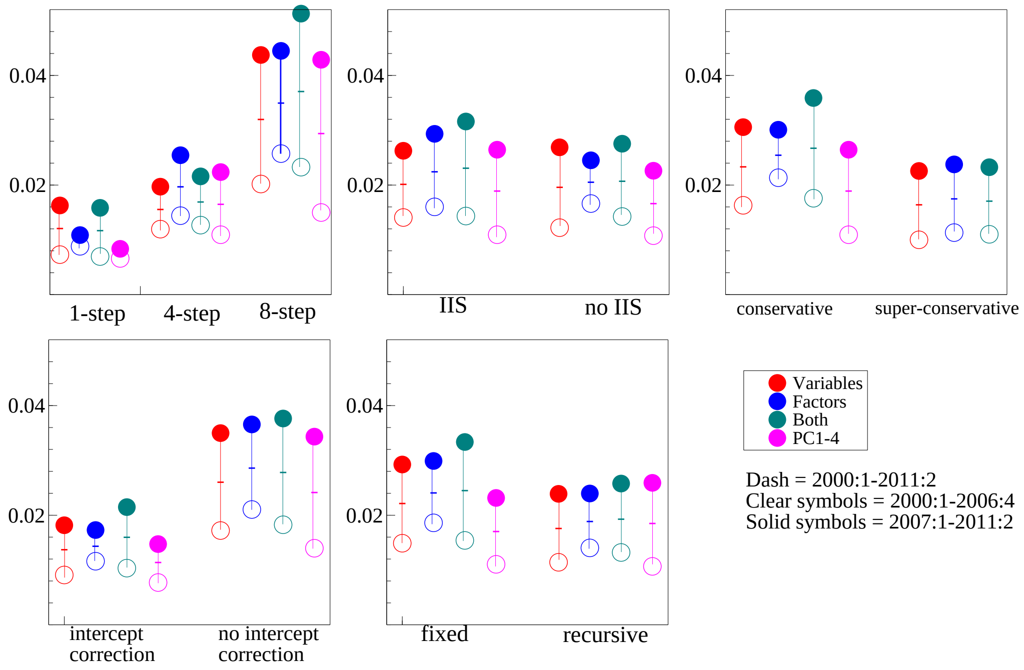

Figure 2: Average RMSFE for log GDP ($\widehat{y}_{T+h}$)

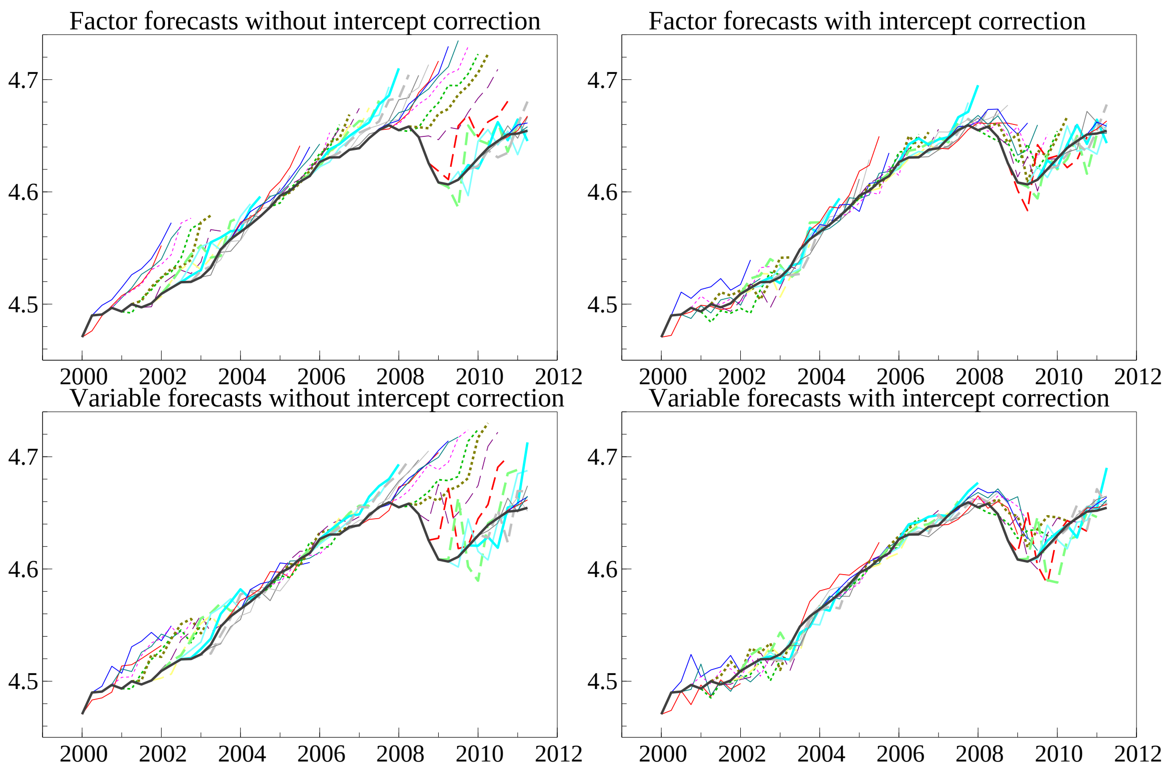

Figure 3: 1-step to 8-step ahead forecasts for GDP. Forecasts from factor models and variable models, both recursive super-conservative selection with IIS, with and without intercept correction.

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.

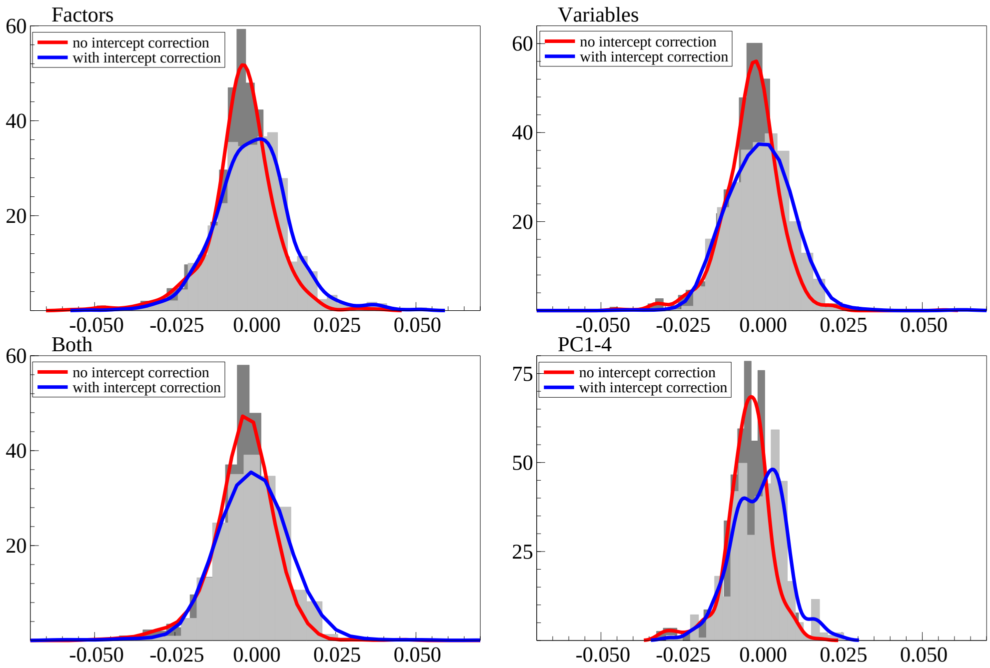

Figure 4: Distribution of forecast errors for GDP growth.

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.

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: the objective of this paper.

One of the key explanations for forecast failure is that of structural breaks. When the underlying data generating process shifts, but the forecasting model remains 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

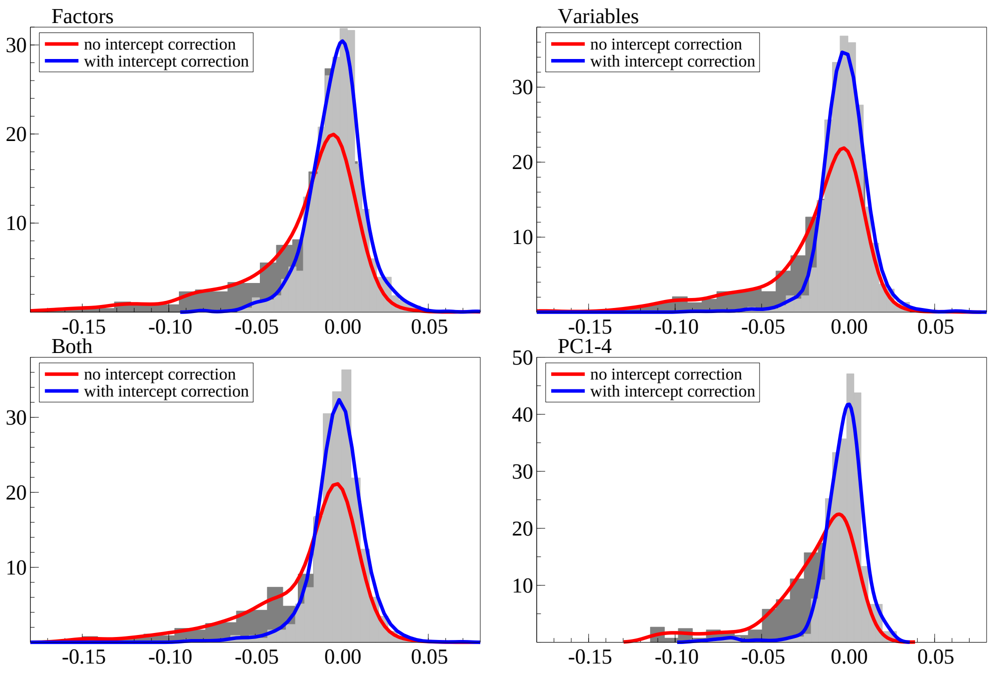

Figure 5: Distribution of forecast errors for log GDP.

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 the stationary principal components which may pick up 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 naive autoregressive benchmarks. These naive benchmarks are poor at forecasting levels though, 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 the forecast horizon is subject to structural change. Furthermore, 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. 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.

Whether the data are generated by latent factors or observable variables will depend on the phenomenon being analysed, so 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.

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

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

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

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

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

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

Castle, J. L., Doornik, J. A., and Hendry, D. F. (2011c). Model selection when there are multiple breaks. *Journal of Econometrics*, DOI:10.1016/j.jeconom.2012.01.026.

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. (2011a). 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 US. *Journal of Business and Economic Statistics*, **26**, 546–554.

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

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

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

Clements, M. P., and Hendry, D. F. (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics*, **10**, 127–146.

Clements, M. P., and Hendry, D. F. (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics*, **58**, 657–684.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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. DP, Federal Reserve Board of Governors, Washington, D.C.

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

Faust, J., and Wright, J. H. (2007). Comparing greenbook and reduced form forecasts using a large realtime dataset. Working paper series 13397, NBER.

Fildes, R. (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. (2011). Open-model forecast-error taxonomies. In Chen, X., and Swanson, N. R.(eds.), *Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions*, forthcoming, 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, Pennsylvania.

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

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

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

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

Kishor, N. K., and Koenig, E. F. (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.

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.

Tinbergen, J. (1930). Determination and interpretation of supply curves: An example [Bestimmung und Deutung von Angebotskurven: ein Beispiel]. *Zeitschrift fur Nationalökonomie*, **1**, 779–679.

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.

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.