Real-Time Factor Model Forecasting and the Effects of Instability

Michael P. Clements*
ICMA Centre,
Henley Business School,
University of Reading,
Reading RG6 6BA

June 5, 2014

Abstract

We show that factor forecasting models deliver real-time gains over autoregressive models for US real activity variables during the recent period, but are less successful for nominal variables. The gains are largely due to the Financial Crisis period, and are primarily at the shortest (one quarter ahead) horizon. Excluding the pre-Great Moderation years from the factor forecasting model estimation period (but not from the data used to extract factors) results in a marked fillip in factor model forecast accuracy, but does the same for the AR model forecasts. The relative performance of the factor models compared to the AR models is largely unaffected by whether the exercise is in real time or is pseudo out-of-sample.

JEL classifications: C51, C22.
Keywords: Factor Models, Robust Approaches, Financial Crisis.

*Michael Clements is Professor of Econometrics at the ICMA Centre, Henley Business School, University of Reading, and a member of the Institute for New Economic Thinking at the Oxford Martin School, University of Oxford. Email: m.p.clements@reading.ac.uk. The computations were undertaken using code written in Gauss 14, Aptech Systems, and the graphics were produced using OxMetrics 6.30 (http://www.oxmetrics.net/).

1 Introduction

Diffusion indices and factor models have become popular in the economic forecasting literature in recent years: see e.g., Stock and Watson (1989, 1999, 2002, 2009a, 2011a), Forni, Hallin, Lippi and Reichlin (2000), Peña and Poncela (2004), Schumacher and Breitung (2008), Bai and Ng (2008) and Castle, Clements and Hendry (2013), *inter alia*. This can be explained by the greater availability of large databases, the desire to draw on all relevant data when analyzing the current state of the economy and scope for government intervention [1], and the favourable forecast performance of factor models, at least initially. Stock and Watson (2011a, p.54) suggest that the evidence indicates that 'factor forecasts perform well to very well relative to competitors for many, but not all, macroeconomic series. For U.S. real activity series, reductions in pseudo-out-of-sample mean squared forecast errors at the two- to four-quarter horizon are often in the range of 20% to 40%, although smaller or no improvements are seen for other series, such as U.S. inflation after 1990'. However, other assessments of the forecast performance of factor models are more equivocal, as indicated by Eickmeier and Ziegler (2008) in their meta-analysis of factor forecasting applications, and by D'Agostino, Giannone and Surico (2006), who suggest the predictability of US macroeconomic series has greatly diminished in the Great Moderation period (mid 1980s to the 2007 Financial Crisis).

Although Eickmeier and Ziegler (2008) tease out a number of the determinants of factor model forecast performance from a careful meta-analysis of 52 studies, a number of potentially important determinants are not explicitly addressed. Chief among these are the potential effects of parameter non-constancy or structural breaks on the factor forecasting models (as well as on the rival or bench-mark models). Structural breaks are sometimes viewed as the key culprit in causing forecast failure in general.[2] We find that the relative performance of factor models is markedly better than that of AR benchmarks during the recent crisis period, with factor models exhibiting a degree of adaptability in times of change. We also consider split-sample estimation of the factors and factor forecasting model following the promising results found by Stock and Watson (2009a) in their in-sample study. In addition we consider the adoption of rolling estimation windows, which is a common approach when it is felt that earlier observations may be less relevant.

We consider the usefulness of these way of mitigating the effects of breaks in a real-time forecasting exercise, as opposed to the 'pseudo-out-of-sample' exercises that typify the majority of the literature. That is, the vast majority of the evidence that underpins the conclusions of studies such as Eickmeier and Ziegler (2008) is based on forecasting exercises that have used fully-revised data, rather than the

[1]For example, Bernanke and Boivin (2003, p.526) refer to 'Central banker's reputations as data fiends'. Factor models are a formal way of allowing all the disparate data series available to the analyst to have some influence on the question of interest.

[2]See Clements and Hendry (2006) for a recent review.

vintages of data which would have been available at the time the forecasts were actually made. A focus of much of the recent literature has been on the implications of using fully-revised data versus real-time data for both historical analyses and forecasting. A number of papers have investigated whether the use of final-revised data (as in the pseudo out-of-sample) overstates the usefulness of various predictor variables relative to appraisals based solely on the data vintages available at the time the forecasts are constructed (see, for example, Diebold and Rudebusch (1991), Faust, Rogers and Wright (2003), and the review by Croushore (2006)). There are fewer than a handful of papers which report proper real-time evidence on factor forecasting performance, so that one contribution of our paper is to compare real-time and pseudo out-of-sample performance for forecasting a relatively large number of variables over the recent period.

Our study of the forecasting performance of factor models also acknowledges that factor models may work better for some macroeconomic variables than for others (as suggested by the quote above from Stock and Watson (2011a)), and we are careful to distinguish between real activity variables, and price/nominal variables in the way we present results.

We also investigate why it might be that factor models offer little improvement on AR models for forecasting.

In focusing on structural breaks and real-time effects, there are a number of aspects that we do not cover which might potentially be important, so our paper is complementary to the large body of work on factor forecasting models summarized in Eickmeier and Ziegler (2008). Among the aspects we neglect are the following. We estimate the factors by principal components, and do not consider alternative methods such as that of Forni, Hallin, Lippi and Reichlin (2005). We do not consider the role of the number of variables (N) used to estimate the factors, and because we adopt a fully real-time exercise, our N is necessarily at the lower end of the values used in pseudo out-of-sample forecasting exercises.[3] Furthermore, we do not calculate factors from a set of 'targeted predictors', as in Bai and Ng (2008). Targeted predictors are variables selected to have predictive power for the variable of interest, based on hard or soft thresholding (such as LASSO, see e.g., Tibshirani (1996)). Nor do we consider block-factor approaches, such as Moench, Ng and Potter (2009), where the data is divided up into a number of categories, and principal components are calculated for each category.

The plan of the remainder of the paper is as follows. Section 2 discusses the literature on factor forecasting using real-time data, and recent developments concerning factor models and instability. Section 3 describes the data set and our implementation of factor forecasting. Section 4 presents the results of the empirical forecasting exercises, section 5 investigates why factor models are only marginally better

[3]We have $N = 52$, but according to e.g., Bai and Ng (2002) and Boivin and Ng (2006), inter alia, this should not harm the forecasting performance of the factor models, although there is evidence to the contrary, e.g., Bern03.

than AR models in ‘normal times’, and section 6 concludes.

2 Related literature

There are few studies of factor model forecasting using real-time data, presumably because the requirements of amassing datasets consisting of a large number of variables covering reasonably long historical periods, for each of a number of data vintages, constitute a demanding data collection exercise. One such study is Bernanke and Boivin (2003), who calculate the accuracy of recursively-generated forecasts of 3 variables, CPI inflation, industrial production, and unemployment, over the period 1970–1998, for both factor models based on real-time data, and on fully-revised data. For factors calculated from a dataset comprising 78 variables, for which have real-time vintages and fully-revised data, they find forecast performance is broadly similar for the factor forecasting models estimated on fully-revised and real-time data, and the gains of the factor forecasts over an autoregressive model are modest: small for CPI, but larger for industrial production and unemployment. However, they find that if they use the 215 series fully-revised data set of Stock and Watson (2002) to construct factors in a pseudo-out-of-sample exercise (that is, estimating the models using the fully-revised data), they obtain sizeable gains for all three variables, including CPI. They suggest that the number of variables used to construct the factors may matter more than whether fully-revised or real-time data is used.

Faust and Wright (2009) use the real-time datasets associated with the Greenbook forecasts for the FOMC meetings from 1980 to 2000 to analyze forecasts of GDP-deflator inflation and GDP growth. Their interest goes beyond factor models, and they consider a raft of univariate and multivariate time-series models relative to the Greenbook forecasts.[4] They compare the relative performance of the Greenbook and model forecasts using *ex post* revised data and real-time data, and come to the same conclusion as Bernanke and Boivin (2003), that the relative performances are largely unaffected. Their results indicate that for output growth the factor model is 20% and 30% worse at current quarter and one-quarter horizons, respectively, and 10% worse at 4-quarters ahead. While for inflation the factor model is nearly 5% worse at the current-quarter horizon, and then roughly 10% better at the 1 and 4-quarter ahead horizons than the AR.[5] These findings are at odds with the summary of empirical findings in Stock and Watson (2011a) discussed in the introduction.

[4]A key question is whether the superiority of the Greenbook forecasts is solely due to the Fed staff’s knowledge of the current and state of the economy and of recent developments which affect the short-term outlook. By providing the atheoretical models with the same information as the Greenboook, they find that the Fed’s forecasts of the GDP-deflator measure of inflation remain superior, but their advantage forecasting GDP growth disappears.

[5]These figures are calculated from their table 2 which records the results of the real-time exercise. We divide their FAA (factor-augmented autoregression) by the RAR (recursive autoregression) for ‘Jump off -1’ (the cleanest of their comparisons between the AR and factor model in terms of adjusting for ‘current’ information) and their horizons of ‘0’, ‘1’ and ‘4’.

That Bernanke and Boivin (2003) and Faust and Wright (2009) report contradictory findings for the forecast performance of factor models for activity variables is not surprising, given that the results of forecast comparisons between models or methods are notoriously dependent on the sample period and the variable being forecast.[6] Neither study extends beyond the 1990s. An original feature of our study is the analysis of the performance of the factor forecasting model in the run-up to and during the onset of the financial crisis. A second is that we address the impact of breaks on factor model forecasts, extending the work of Stock and Watson (2009a) to out-of-sample forecasting. Stock and Watson (2009a) 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)'. However, their results suggest that the factors can be estimated on the full sample notwithstanding the instability in the factor loadings. Their analysis shows that forecasting equations estimated over the post-break period, but using estimates of the factors from the full historical period across the break, fit better in-sample than using subsample estimates of the factors.

Motivated by their findings, we assess the impact on forecast performance of different ways of allowing for the instability in the factor forecasting model given full sample estimates of the factors. The first is to use only post–break observations, realizing Stock and Watson's suggestion in an out-of-sample real-time forecasting exercise. We follow Stock and Watson, and use 1984 as the break period. This corresponds to the Great Moderation of output discussed by, e.g., McConnell and Perez-Quiros (2000). However, although a number of studies provide evidence of a break in the mid 1980s, our forecast period extends to 2011, and so includes the Financial Crisis, another obvious candidate for a break date. There are of course other less high-profile events and factors which may have contributed to non-constancies in the factor forecasting models. So in addition to considering an estimation sample consisting of post-break observations, we consider rolling forecasting schemes. .

3 The real-time dataset

The set of variables used to construct factors is described in table 1. The majority of these variables are subject to revision, and are taken from the Real-Time Data Set for Macroeconomists (RTDSM) of Croushore and Stark (2001), whereas the non-revised data come from FRED (see table for details and web addresses). Our selection is limited by the availability of data vintages (for variables which are revised) and of sufficiently long spans of historical data for model estimation. Our first 'vintage-origin' is 1996:Q2, and the last is 2011:Q1, so that we have 15 years of quarterly forecast origins. However, the

[6] The recent inflation forecasting literature provides an example of the former. A number of authors have emphasised the instability in inflation forecasting models: see e.g., the review by Stock and Watson (2009b).

1999:Q4 and 2009:Q3 vintages contain missing values for many series and these forecast origins are excluded. We also excluded these two origins from the pseudo-out-of-sample forecasting exercises which use fully-revised data, to ensure that the real-time and fully-revised data exercises are strictly comparable in terms of covering the same observations. Data at a monthly frequency are converted to quarterly observations by averaging the months. Similarly, where the RTDSM provides monthly vintages, we extract the February, May, August and November vintages which correspond to the quarterly vintages. Together with the not-revised series, we have a total of 51 variables. By beginning the estimation period in 1970 we are able to work with a balanced panel of quarterly data for factor estimation.

At each forecast origin the factors are calculated using observations on the variables from the vintage of data available at that date. (For example, for the first 'vintage-origin' of 1996:Q2, we use data from 1970 to 1996:Q1 from the 1996:Q2 data vintage). The data are made stationary, primarily by first or second differencing the logs of the real and nominal variables respectively (but see table 1[7]). The first n_f principal components are calculated, and used in the factor forecasting model for y_t:

$$y_t = \gamma_0 + \sum_{j=0}^{p-1} \rho_j y_{t-h-j} + \sum_{k=1}^{n_f} \sum_{j=0}^{k_l} \gamma_{k,j} f_{k,t-h-j} + \epsilon_t. \tag{1}$$

Note that the right-hand side variables are lagged h-periods or more relative to the predictand to allow direct forecasting (as opposed to generating multi-step ahead forecasts by iterating the forecasts of a '1-step' forecasting model). In most cases y_t is the difference of the log of the level Y_t, so that we forecast the quarterly-difference h-steps ahead (whereas some papers forecast the h-step difference)[8].

In principle, the number of autoregressive lags p, the number of factors n_f, and the number of lags of each of those factors (k_l) might be selected by the data, and Eickmeier and Ziegler (2008) describe a range of strategies that have been adopted.[9] We adopt the not uncommon approach of simply setting these to reasonable values. We set $n_f = 3$, $k_l = 0$ (i.e., no lags of the factors), and $p = 2$. The benchmark models are obtained by setting $n_f = 0$, so the factors are omitted. Hence we investigate the marginal contribution from including factors relative to a univariate model with two lags.

The forecast errors are calculated using data taken from a vintage available two quarters after the target quarter. So for the $h = 1$ forecast of 1996:Q2 from the first origin, for example, this would be the value available in the 1996:Q4 vintage, and for the $h = 4$ forecast of 1997:Q1, the value for this

[7]We followed Stock and Watson (e.g., Stock and Watson (2009a)) in choosing the transformations to apply to make these series stationary prior to estimating the factors.

[8]We also use the first difference specification when the variable is differenced twice prior to factor computation (as indicated in table 1). That is, we do not impose the second unit root in model specification.

[9]In addition, model selection with factors (and the original variables) was extensively investigated by Castle *et al.* (2013) for forecasting output growth. They find that simply using the first four factors (without lags, but including own-variable lags) dominates selection strategies involving factors.

period in the 1997:Q3 vintage. Although the 'true' values may be most accurately estimated by the latest-available vintage estimates of those observations (here, the vintage 2012:Q3 estimates), there are good reasons to choose 'real-time' actuals (often taken to be the vintage released two quarters after the observation quarter, as here). The latest available will incorporate the effects of annual revisions and periodic benchmark revisions which may be largely unpredictable. The second quarterly release may be better than the first release (available the following quarter) which is generally based on less complete data.

When we compare the findings of the real-time exercises and the pseudo out-of-sample, we use actual values taken from the latest-available vintage (2012:Q3) for the latter, because the defining characteristic of such exercises is that all the data are fully revised.

Finally, it should be noted that data revisions can create mismatches between the actuals and the forecasts in real-time contexts. The sequence of revisions of the Bureau of Economic Analysis (BEA) data is as follows. The 'advance' estimates are recorded in the RTDSM the quarter after the target observation, followed by the second estimate (the BEA 'final' estimate) a quarter after that. Thereafter BEA data are subject to three annual revisions which occur in the July of each year (see, e.g., Fixler and Grimm (2005, 2008) and Landefeld, Seskin and Fraumeni (2008)). In addition to the regular rounds of revisions, every five or ten years the BEA data are typically subject to 'benchmark revisions', reflecting methodological changes in measurement or collection procedures, and often include base year changes. The base year changes in particular are problematic for real-time assessments of the accuracy of forecasts of levels or log-levels of variables (and, though to a lesser extent, differences in non-log levels): when the forecast and actual value straddle such a change, an adjustment needs to be made (see, e.g., Clements and Galvão (2012, p.556) for one way of doing this). Generally growth rates are largely immune to re-basings of this sort. Rather than adjusting the levels data, we have chosen not to forecast a handful of variables which would have required some intervention to offset the effects of base-year changes (these variables are indicated in table 1). Note that these variables can be used as part of the dataset to calculate factors: problems only arise for evaluating real-time forecasts of these variables.

4 Results of empirical forecast comparisons

To aid interpretation, we divide the discussion of the results into a number of sub-sections, each presenting evidence on a possible determinant of factor model forecast performance. In each sub-section we present the relative accuracy of the factor model forecasts compared to an AR for each of the 44 variables we forecast. We also present the median relative performance across all 44 variables, and the medians separately across the 18 real activity variables and the 26 price/nominal variables (groupings defined in section 3). The base case is a recursive forecasting scheme, whereby parameter estimates are updated as we move through the forecast period. It might be argued that such a scheme in itself permits some adaptability of the models in the face of non-constancy (relative to a fixed forecasting scheme). However, there seems little reason not to update the model estimates as the forecast origin moves forward. Using instead a fixed scheme (results not shown) made little difference qualitatively to the relative accuracies of the models.

The last sub-section analyzes the results for forecasting real GDP, given that this is the main focus of much of the empirical material, and in some respects (especially forecasting over the Crisis period) the results are typical of those for many of the real variables.

4.1 Relative performance using real-time and pseudo real-time data

The results for the pseudo real-time data 1-step ahead indicate some benefit to using the factor model for real-activity variables 'on average': see table 2. The median ratio of the factor model RMSE to the AR across the real activity variables is 0.89. But factor models are slightly worse than the AR models on average for the nominal variables. For all variables taken together the median ratio is 0.98. These findings are consistent with the review by Eickmeier and Ziegler (2008) who find factor models are better for real variables than price variables, for the US, and who conclude that the performance of factor models is 'on average...slightly superior to that of other models' (p.261). Of course there are variables for which the factor model forecasts are markedly superior to the AR, and these include key indicators of the state of the economy, such as the rate of unemployment (ruc) and industrial production (ipt) which are over 20% and 15% more accurate on RMSE respectively.

Consider now the implications of replacing the fully-revised data with the vintages of data which would have been available at the time the forecasts were made, and of using 'preliminary' estimates of the outturns to assess forecast accuracy. The 'Real-time' columns of the table show that the (absolute) accuracy of the AR model generally worsens for the real variables, but that the relative accuracy of the factor model forecasts is largely unaffected. For the two variables we highlighted, unemployment and industrial production, the relative accuracy of the factor model worsens by just 2 percentage points, matching the average change. We conclude that the conclusions drawn from pseudo out-of-sample forecasting exercises are a good guide to the gains that would have been possible in real-time using the then available data vintages. This is consistent with the evidence presented in Bernanke and Boivin (2003, Table 1, p.532) for forecasting the unemployment rate and industrial production using real-time vintages and fully-revised data: their two forecasting exercises indicate similar gains of 10 to 20% for the period 1970–1998 for both real-time and fully-revised data. However, their forecast horizons are 6-

months and one year ahead, whereas our table 2 is for 1-step forecasts. D'Agostino *et al.* (2006) argue that there is little predictability beyond one-step ahead (relative to simple AR models) in the post-1984 period. The forecast period studied by Bernanke and Boivin (2003) includes the 1970s, which might be responsible for the good performance of the factor models. To address this claim, we next consider multi-step forecasts over our (1996-2010) forecast period.

4.2 Relative performance of factor model forecasts beyond a 1-quarter ahead horizon

Tables 3 and 4 replicate the comparisons presented in table 2 (for 1-step ahead) but for half-yearly ($h = 2$) and yearly horizons ($h = 4$). We find that the relative advantage of the factor model for forecasting the real variables diminishes in the forecast horizon: at $h = 4$ the median RMSE ratio for the real variables is 0.97 using fully-revised data, and 0.99 in real-time. Of the 17 separate real variables, for the unemployment rate alone is the gain to the factor model larger than 10%. In short, the benefit to using factor models is greatest at $h = 1$, and then primarily only for the real variables. Beyond one step ahead any gains are at best modest. Note that as at $h = 1$ the choice between fully-revised data and real-time vintages (in terms of relative performance) is largely inconsequential.

4.3 Full versus sub-sample estimation

In the presence of structural breaks, of which the onset of the Great Moderation circa 1984 is a prime example, it may be optimal in a MSE sense to drop earlier observations, as argued by Pesaran and Timmermann (2007) in the context of forecasting with AR models. Stock and Watson (2009a) suggest the same may be true of factor forecasting models: they find the in-sample fit of their factor models improves when they are estimated over the post-break period (but using full-sample estimates of the factors). The results to this point are based on a recursive forecasting scheme, so that the beginning of the estimation sample remains fixed at 1970, and the factors and forecasting models are estimated on expanding windows of data as the forecast origin is moved through the data (from 1996 to 2010). Table 5 presents the real-time out-of-sample analogue of the approach of Stock and Watson (2009a): it is a recursive exercise but with the model estimation sample beginning in 1985 (as opposed to 1970). Consider the left panel of the data, in which the factor model is estimated on the shortened samples, but the AR is not. The strategy of using only post-break observations yields marked improvements at 1-step ahead (but virtually nothing at $h = 4$). The median RMSE for the real variables is now 0.86, and for the nominal variables is 0.91. However, the right panel shows that shortening the AR estimation period improves the AR forecasts nearly as much as the factor model forecasts; the median RMSE ratio is now 0.89 for the real variables, compared to 0.91 in table 2 when the full-estimation period is used.

Using post-break observations improves forecast accuracy, but for both models, and and to a largely similar extent.

4.4 Rolling estimation

Table 6 reports the results of adopting a rolling estimation window having dropped the pre-1984 observations (but using recursive estimates of the factors with data beginning 1970). The results in table 6 are broadly similar to those in table 5, indicating that once the pre-1984 observations have been excluded, whether the forecasting models are estimated on fixed-length or expanding windows of observations is of little importance.

4.5 Constancy of relative forecast performance over time

One of the motivations of this paper was to assess whether the relative performance of factor models has deteriorated during the Great Moderation, as suggested by D'Agostino *et al.* (2006), *inter alia*. The results thus far consider average performance over the period 1996 to 2011, and suggest some gains to using factor models for forecasting real variables one-step ahead. Figures 1 to 3 plot 'local estimates' of RMSEs to assess whether performance has changed over time. Each point in the figure (from 2001:Q1 to 2011:Q1 inclusive) is based on the (real-time) one-step forecast errors for that point and the previous 19 observations (a one-sided 5-year window). We report the RMSE estimates of the FM forecasts and of the AR model forecasts. It is immediately apparent that for all but three of the eighteen real variables there is a marked improvement in FM relative to the AR from around 2008 onwards. [10] In many instances both models' performances worsen, but typically the deterioration in the FM forecasts is more muted. For many of the variables the performance of the FM and AR models is roughly comparable up to this point. Both models become more accurate from around 2004 up to 2008 for the majority of the real variables.[11] The estimates of the MSFE from 2008 onwards begin to include the financial crisis (the period 2007:Q4 – 2009:Q2 was designated a recession by the NBER), and indicate that the factor models were more successful at forecasting over this period than the AR models. The figures present the RMSEs for the two models individually, to highlight that the forecast performance of both models deteriorates towards the end of the period, and that the improvement in the factor model is relative.

If we only consider forecast origins up to 2007:Q3, and so omit the period 2008:Q1 to 2011:Q1, we find that on average (across variables) the FM forecasts are no better than the AR, consistent with the

[10]The exceptions are non-durable consumption, residential investment, and government expenditure

[11]This may occur as the 2001:Q1 to 2001:Q4 recession years drop out of the RMSEs, but the sample of rolling RMSFE estimates begins in 2001:Q1 so it is not clear what is happening prior to the earlier recession. What is clear is that after the 2001 recession the performances of the two models closely track each other, where as post 2008 they diverge (for many variables).

findings for activity variables in figures 1 to 3. We return to this in section 5.

4.6 Forecasting real GDP growth

We present a study of real GDP given that this variable is often the focus of interest, and because figures 1 to 3 indicate that the relative performances of the AR and FM models over the recent period for real GDP growth may be typical of a number of real variables. Figure 4 depicts the outturns and one-quarter ahead growth rate forecasts for AR and FM for 1996:Q2 to 2011:Q1. The dates refer to the periods being forecast. The results are based on the use of fully-revised data (2012:Q3 vintage data).[12] The figure indicates that large differences in forecast errors between the AR and FM mainly occur from 2008:Q3 onwards. In 2008:Q3 the economy shrinks by nearly one percentage point. The FM predicts around a half of this decline, but the AR predicts positive growth of around a half a percent. In the fourth quarter the economy contracts by over 2%, the FM again predicts a decline of around half a percent, and the AR positive growth of a third of a percent. In the following quarter the actual rate of decline eases, and the FM over-predicts the magnitude of the decline by half a percentage point, but is still markedly more accurate than the AR (which now predicts a small decline, but this is around $1\frac{1}{2}$ percentage points smaller than the actual drop in output). These three quarters, 2008:Q3 – 2009:Q1, account for the dramatic deterioration in accuracy of the AR relative to the FM depicted in figure 1. In the next two quarters the economy returns to modest growth, but FM continues to predict sizeable contractions ($1\frac{1}{2}$ and 1 percent respectively), and subsequently over-predicts for the next 6 quarters. The AR is more accurate over the period 2009:Q2 to 2011:Q1.

5 Why are FM forecasts only marginally more accurate than than AR model foecasts?

We have shown that FM forecasts are on par with AR model forecasts if we ignore the Crisis period, when they tend to perform better. In this section we shed some light on why FMs do not clearly outperform ARs, as might be expected given that FMs draw on large numbers of related variables, and ARs on only past values of the variable in question. To do so, we assume that the economy has a factor structure,

[12]Recall that the differences attributable to the use of fully-revised data and real-time were found to be small. Here we use fully-revised data and include results for forecasting the 1999:Q4 and 2009:Q3 periods. (In the results reported elsewhere in the paper these two periods are excluded. This is because the corresponding real-time data vintages contain short data histories for some series, so real-time forecasting is not feasible, and they were also excluded from the full-revised exercises for comparability of the results).

viz.

$$\mathbf{x}_t = \boldsymbol{\Upsilon}\mathbf{f}_t + \mathbf{e}_t \tag{2}$$

$$\mathbf{f}_t = \boldsymbol{\Phi}\mathbf{f}_{t-1} + \boldsymbol{\eta}_t \tag{3}$$

where $\mathbf{x}_t$ is $n \times 1$, $\mathbf{f}_t$ is $m \times 1$, $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Phi}$ 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 $\mathbf{x}_t$.[13] 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. This is a standard setup in the literature (see, e.g., Stock and Watson (2011b)).

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 $\boldsymbol{\lambda}_i{}'$ denote the i^{th} row of $\boldsymbol{\Upsilon}$, then:

$$\begin{aligned}
\mathsf{E}_t\left[x_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] &= \mathsf{E}_t\left[\boldsymbol{\lambda}_i'\mathbf{f}_{t+1} + e_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\
&= \mathsf{E}_t\left[\boldsymbol{\lambda}_i'\mathbf{f}_{t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\
&+ \mathsf{E}_t\left[e_{i,t+1} \mid \mathbf{x}_t, \mathbf{f}_t, \mathbf{x}_{t-1}, \mathbf{f}_{t-1}, \ldots\right] \\
&= \mathsf{E}_t\left[\boldsymbol{\lambda}_i'\mathbf{f}_{t+1} \mid \mathbf{f}_t, \mathbf{f}_{t-1}, \ldots\right] + \mathsf{E}_t\left[e_{i,t+1} \mid e_{i,t}, e_{i,t-1} \ldots\right] \\
&= \boldsymbol{\psi}_i'\mathbf{f}_t + \delta\left(L\right)x_{i,t} \qquad (4)
\end{aligned}$$

under the assumptions we have made. Hence when (2) and (3) hold, the optimal forecast of $x_{i,t+1}$ consists of the factors dated period t and the AR lags of $x_{i,t}$.

Abstracting from issues to do with the selection of the number of AR lags, the number of factors and the need to estimate the model's parameters, the relative accuracy of the FM to the AR for a variable x_i will obviously depend on the relative importance of the common factors compared to the idiosyncratic terms captured by the autoregressive lags. However, whereas there is a contemporaneous relationship effect of $\mathbf{f}_t$ on $x_{i,t}$ in the data generating process, a 'direct' forecasting model[14] based on (4) relates $x_{i.t}$ to $\mathbf{f}_{t-1}$ (for a one-step ahead forecasting horizon), so that the coefficient on $\mathbf{f}_{t-1}$ conflates the direct effect of $\mathbf{f}_t$ on $x_{i,t}$ and the parameter matrix $\boldsymbol{\Phi}$. Alternatively, one might estimate a model that relates $x_{i,t}$ to $\mathbf{f}_t$ (and lags of $x_{i,t}$) and generate forecasts of $x_{i,t+1}$ by replacing the unknown future values of $\mathbf{f}_{t+1}$ by forecasts obtained from estimating a VAR model such as (3). But in either case a low level of

[13] The form of (2) indicates that $\mathbf{x}_t$ responds only to current factors, $\mathbf{f}_t$, as assumed in the empirical work, but more generally $\boldsymbol{\Upsilon}$ would be replaced by $\boldsymbol{\Upsilon}(L)$. We also assume in (3) that the VAR for $\mathbf{f}_t$ is first order, but select the VAR order according to SIC in the empirical forecast comparisons.

[14] See, e.g., Clements and Hendry (1996) and Marcellino, Stock and Watson (2006).

predictability of $\mathbf{f}_t$ will adversely affect the accuracy of the FM forecasts. Consider the extreme case of $\mathbf{\Phi}$ being the null matrix, whence the FM is of no value (given the availability of an AR model) for forecasting.[15]

A direct way of gauging the impact on real-time forecasting of the predictability of the factors is to suppose the future values of the factors are known. The deterioration in forecast performance when the factors are replaced by forecasts (or when $x_{i.t}$ is related to $\mathbf{f}_{t-1}$ directly) measures the loss of accuracy from this avenue. We begin by establishing that estimating a model with the contemporaneous values of the factors, and then forecasting the factors (denoted FM$_f$)[16] is broadly similar (in terms of accuracy) to projecting the x_i on to lagged factors, our chosen approach thus far. In table 7 the column headed FM$_f$/FM reports the ratio of these two approaches. On average across the variables the two approaches are equivalent, and the variations across variables are relatively small - neither approach differs form the other by more than 10% across the real variables. So nothing is lost by treating the generation of the FM forecasts as a two stage procedure, using 'plug-in' forecasts of the factors at the second stage.

Suppose now the future values of the factors are known (denoted FM$_{Kn}$). The FM$_{Kn}$ forecasts are on average around 20% better than FM (column headed FM$_{Kn}$/FM in table 7): this is the loss in accuracy in forecasting the x_i's from forecasting the factors (or equivalently, from projecting the x_i's on past values of the factors). Hence for the real variables taken together, the need to forecast the factors reduces an average gain to the FM over the AR from around 30% (FM$_{Kn}$ relative to AR) to around 10% (FM/AR). Finally, we argued in section 4.5 that the overall gains to the FM models were solely due to the 2008-11 period. This is apparent from table 7 which reports the comparisons for the period excluding 2008-11. In terms of the averages, the median FM to AR RMSE ratios are now 1.05 for real activity variables, 1.00 for nominal variables, and 1.01 for all variables taken together, compared to full sample findings of 0.91, 1.00 and 0.97. However, even excluding the turbulent end years the FM forecasts would be superior to the AR forecasts (by around 20% on average for the real variables) *assuming* known factors (as evidenced by the RMSE ratios FM$_{Kn}$/FM and FM/AR in the right side of the table). Put another way, in the absence of large changes in the economy, the costs (in terms of forecast accuracy) of projecting the factors forward in time happens to largely offset the benefits that would otherwise accrue to using the factor model. When there are large movements in factors - as in times of change - the benefits to including the factors more than offset the negative effects of projecting the factors forward.

We conclude that the somewhat disappointing performance of the FM forecasts results is attributable

[15]The degree of predictability will also depend on the size of the unexplained component in the equation (3) for $\mathbf{f}_t$, that is, $E\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right)$.

[16]The forecasts of the factors are obtained from a VAR for the factors with lag order selected by SIC.

to a loss of accuracy from the need to forecast the future values of the factors (or projecting the x_i's on lagged factors).

6 Conclusions

Despite the widespread use of factor models for economic forecasting, and the belief that such models are on a par with other methods of handling large datasets (see De Mol, Giannone and Reichlin (2008)), there is less of a consensus regarding their value relative to simple benchmarks such as AR models. Our findings shed some light on the reasons for a lack of a consensus: the relative performance of factor models relative to AR models depends upon the historical period under study. The inclusion of the financial crisis period enhances the relative performance of the factor model. Earlier studies will of necessity report findings for different historical periods. For example, the sample period used by Bernanke and Boivin (2003) (1970-1988) includes the turbulent 1970's, which may explain the positive findings in favour of FMs. Studies that are based solely on the Great Moderation period may conclude that there is little benefit from using factor models relative to AR models. A more positive slant to these findings would be that FMs work relatively well when the AR model's forecast performance is at its worst, and in normal times are on a par with AR models.

In addition we have considered a number of issues that might be expected to bear on FM forecast performance. Any gains to FM forecasts appear to be short-lived - there are reduced benefits to using FMs at two-quarter ahead horizons, but no gains a year ahead on average across variables, although specific variables can be forecast more accurately (e.g., the unemployment rate). We have tended to stress the performance of the model for broad variable groups (real versus nominal variables), but the tables report result for 44 variables to indicate the variation around the average, as well as to aid comparison with earlier studies, and indicate any clear successes (such as the unemployment rate one-year ahead - 20% lower RMSE than the AR). We also present detailed results for real GDP growth, and find that the AR and FM forecasts closely track each other except for a few key observations during the Financial Crisis.

We also considered whether the pseudo out-of-sample forecasting comparisons (that comprise the vast majority of the factor forecasting literature) present an accurate reflection of the accuracy of FM forecasts in real time (that is, using the vintages of data which would have been available at the time the forecasts were made). By and large data vintage effects - using fully revised as opposed to real-time vintages - were small and could be ignored.

In terms of offsetting the effects of model non-constancy on forecast performance, we found evidence for estimating the factor forecasting model on data from 1985 onwards (but using the whole period from 1970 onwards to calculate the factors), consistent with the in-sample regression analysis of Stock and Watson (2009a). However the AR models were also improved by setting the beginning of the sample to 1985, so that the relative improvement to the FMs was more modest. Conditional on starting the estimation period in 1985, the use of a rolling or recursive forecasting scheme made little difference.

Overall, our results point to the primacy of the forecast environment as the key determinant of factor model forecast performance, and the finding that FM forecasts are relatively more accurate than the forecasts from AR models during major upheavals, such as the financial crisis. That this should turn out to be the case is far from obvious given the prevailing state of knowledge: simple, flexible models such as AR models are sometimes thought to offer greater adaptability at times of rapid change.

References

Aiolfi, M., Capistrán, C., and Timmermann, A. (2011). Forecast combinations, chapter 11. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 355–388: Oxford University Press.

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

Bai, J., and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bates, J. M., and Granger, C. W. J. (1969). The combination of forecasts. *Operations Research Quarterly*, **20**, 451–468. Reprinted in T.C. Mills (ed.), *Economic Forecasting*. Edward Elgar, 1999.

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

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

Castle, J. L., Clements, M. P., and Hendry, D. F. (2013). Forecasting by factors, by variables, by both or neither?. *Journal of Econometrics*, **177**, 305–319. DOI information: 10.1016/j.jeconom.2013.04.015.

Castle, J. L., Clements, M. P., and Hendry, D. F. (2014). Robust approaches to forecasting. *International Journal of Forecasting*. Forthcoming.

Chong, Y. Y., and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies*, **53**, 671–690. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

Clements, M. P., and Galvão, A. B. (2012). Improving real-time estimates of output gaps and inflation trends with multiple-vintage VAR models. *Journal of Business and & Economic Statistics*, **30(4)**, 554–562. DOI: 10.1080/07350015.2012/707588.

Clements, M. P., and Harvey, D. I. (2009). Forecasting combination and encompassing. In Mills, T. C., and Patterson, K. (eds.), *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*, pp. 169–198: Palgrave MacMillan.

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. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Mass.: MIT Press.

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.

Croushore, D. (2006). Forecasting with real-time macroeconomic data. In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting, Volume 1. Handbook of Economics 24*, pp. 961–982: Elsevier, Horth-Holland.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105(1)**, 111–130.

D'Agostino, A., Giannone, D., and Surico, P. (2006). (Un)Predictability and macroeconomic stability. Working paper series 0605, European Central Bank.

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*(2), 318–328.

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

Eickmeier, S., and Ziegler, C. (2008). How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach. *Journal of Forecasting*, *27*(3), 237–265.

Faust, J., Rogers, J. H., and Wright, J. H. (2003). Exchange rate forecasting: The errors we've really made. *Journal of International Economic Review*, **60**, 35–39.

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 (4)**, 468–479.

Fixler, D. J., and Grimm, B. T. (2005). Reliability of the NIPA estimates of U.S. economic activity. *Survey of Current Business*, **85**, 9–19.

Fixler, D. J., and Grimm, B. T. (2008). The reliability of the GDP and GDI estimates. *Survey of Current Business*, **88**, 16–32.

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.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, **100**, 830–840.

Hendry, D. F. (2006). Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics*, **135**, 399–426.

Landefeld, J. S., Seskin, E. P., and Fraumeni, B. M. (2008). Taking the pulse of the economy. *Journal* *of Economic Perspectives*, **22**, 193–216.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics*, **135**, 499–526.

McConnell, M. M., and Perez-Quiros, G. (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.

Newbold, P., and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society, A*, **137**, 131–146. Reprinted in Mills, T. C. (ed.) (1999), *Economic Forecasting. The International Library of Critical Writings in Economics*. Cheltenham: Edward Elgar.

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

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

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.

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

Stock, J. H., and Watson, M. W. (2009a). Forecasting in dynamic factor models subject to structural instability. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry. Chapter 7*, pp. 173–205. Oxford: Oxford University Press.

Stock, J. H., and Watson, M. W. (2009b). Phillips curve inflation forecasts. In Fuhrer, J., Kodrzycki, Y., Little, J., and Olivei, G. (eds.), *Understanding Inflation and the Implications for Monetary Policy*, pp. 101–186. Cambridge: MIT Press. Ch. 3.

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

D. F. (eds.), *Oxford Handbook of Economic Forecasting, Chapter 2*, pp. 35–60. Oxford: Oxford University Press.

Stock, J. H., and Watson, M. W. (2011b). Dynamic Factor Models, chapter 2. In Clements, M. P., and Hendry, D. F. (eds.), *The Oxford Handbook of Economic Forecasting*, pp. 35–59: Oxford University Press.

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

Table 1: Data Description

Mnemonic	T. code	Forecast	Description	Source
routputq	5	1	Real GNP/GDP	
rconq	5	1	Real Personal Consumption Expenditures Total	
rconndq	5	1	Real Personal Consumption Expenditures: Nondurable Goods	
rcondq	5	1	Real Personal Consumption Expenditures: Durable Goods	
rinvbfq	5	1	Real Gross Private Domestic Investment: Non-residential	
rinvresq	5	1	Real Gross Private Domestic Investment: Residential	
rexq	5	1	Real Exports of Goods and Services	
rimpq	5	1	Real Exports of Goods and Services	
rgq	5	1	Real Government Consumption and Gross Investment: Total	
ipt	5	1	Industrial Production Index: Total	
ipm	5	1	Industrial Production Index: Manufacturing	
ruc	1	1	Unemployment Rate	
employ	5	1	Nonfarm Payroll Employment	
hg	5	1	Indexes of Aggregate Weekly Hours: Goods-Producing	
hs	5	1	Indexes of Aggregate Weekly Hours: Service-Producing	
cut	1	1	Capacity Utilization Rate: Total	
cum	1	1	Capacity Utilization Rate: Manufacturing	
hstarts	4	1	Housing Starts	
pq	6	1	Price Index for GNP/GDP	
cpi	6	1	Consumer Price Index	
pconq	6	1	Price Index for Personal Consumption Expenditures	
pimpq	6	1	Price Index for Imports of Goods and Services	
noutputq	6	1	Nominal GNP/GDP	
nconq	6	1	Nominal Personal Consumption Expenditures	
wsdq	6	1	Wage and Salary Disbursements	
oliq	6	1	Other Labor Income	
propiq	6	1	Proprietors' Income	
divq	6	1	Dividends	
pintiq	6	1	Personal Interest Income	
tranrq	6	1	Transfer Payments	
sscontrq	6	1	Personal Contributions for Social Insurance	
npiq	6	1	Nominal Personal Income	
ptaxq	6	1	Personal Tax and Nontax Payments	
ndpiq	6	1	Nominal Disposable Personal Income	
pintpaiq	6	1	Interest Paid by Consumers	
tranpfq	6	1	Personal Transfer Payments to Foreigners	
ratesavq	1	1	Personal Saving Rate, Constructed	
m1	6	1	M1 Money Stock	
m2	6	1	M2 Money Stock	
aaa	2	1	Moody's Seasoned Aaa Corporate Bond Yield	FRED
baa	2	1	Moody's Seasoned Aaa Corporate Bond Yield	FRED
tb3ms	2	1	3-Month Treasury Bill: Secondary Market Rate	FRED
fedfunds	2	1	Effective Federal Funds Rate	FRED
gs10	2	1	10-Year Treasury Constant Maturity Rate	FRED
oilprice	5	0	Spot Oil Price: West Texas Intermediate	FRED
rinvchiq	1	0	Change in Private Inventories	
rentiq	3	0	Rental Income of Persons	
npsavq	3	0	Nominal Personal Saving	
trbasa	6	0	Total Reserves	
nbrbasa	3	0	Nonborrowed Reserves	
basebasa	6	0	Monetary Base	

The transformation code (column 2) denotes the transformation applied to the variable before principal components are calculated (1: level (no transformation); 2: difference; 3: second difference, 4: natural log; 5: first difference of log; 6: second difference of log. The third column indicates whether the variable is forecast (1) or not (0). We model and forecast the transformation of the variable given in column 2 except when this indicates the second difference of the level or of the log: in these cases we model and forecast the first difference of the level or log, respectively. In terms of pesenting the aggregate forecasting results, variables above the horizontal line are treated as 'real activity' variables, and those below the line as nominal or price variables. Sources: FRED (Federal Reserve Economic Data). http://research.stlouis.org/fred2. If left unspecified, data downloaded from the Real-Time Data Set for Macroeconomists (RTDSM), http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/, see Croushore and Stark (2001).

Table 2: Factor and AR models for 1-step forecasting using real- and pseudo real-time data

	Real-time		Pseudo Real-time	
	AR	FM/AR	AR	FM/AR
routputq	0.58	0.83	0.63	0.91
rconq	0.50	0.98	0.47	1.09
rconndq	0.82	0.92	0.69	0.92
rcondq	2.61	0.99	2.39	1.02
rinvbfq	2.39	0.86	1.80	0.86
rinvresq	3.10	1.05	2.91	1.13
rexq	2.71	0.93	2.58	0.92
rimpq	2.72	0.76	2.32	0.68
rgq	0.72	1.04	0.80	1.03
ipt	0.99	0.87	0.98	0.85
ipm	1.13	0.91	1.16	0.89
ruc	0.28	0.80	0.23	0.78
employ	0.27	0.90	0.26	0.86
hg	1.08	0.89	1.12	0.81
hs	0.38	0.91	0.35	0.90
cut	0.91	0.88	0.78	0.84
cum	1.01	0.91	0.84	0.85
hstarts	0.08	0.99	0.07	0.98
pq	0.27	1.00	0.21	0.99
cpi	0.66	1.05	0.67	1.06
pconq	0.46	1.03	0.45	1.05
pimpq	2.66	1.05	2.36	1.05
noutputq	0.68	1.04	0.72	1.00
nconq	0.72	1.00	0.70	1.03
wsdq	0.78	0.86	1.04	0.90
oliq	0.61	1.01	0.71	1.03
propiq	1.54	0.84	2.22	0.90
divq	5.85	0.92	6.19	0.95
pintiq	1.35	0.95	2.28	0.96
tranrq	1.74	0.95	1.82	0.93
sscontrq	2.00	0.93	2.00	0.95
npiq	0.70	0.94	0.93	0.95
ptaxq	5.37	0.83	5.50	0.85
ndpiq	1.03	1.00	1.07	0.99
pintpaiq	2.96	0.96	2.91	0.98
tranpfq	2.63	1.11	2.27	1.07
ratesavq	1.49	0.99	0.83	1.03
m1	1.33	1.04	1.16	1.02
m2	0.75	1.04	0.61	1.08
aaa	0.28	1.00	0.28	1.01
baa	0.38	1.00	0.38	1.01
tb3ms	0.45	1.08	0.45	1.11
fedfunds	0.47	1.11	0.47	1.20
gs10	0.37	1.00	0.37	1.01
Group and overall medians				
Real	0.95	0.91	0.82	0.89
Nominal	0.76	1.00	0.88	1.01
All	0.87	0.97	0.83	0.98

Notes. The table reports the RMSE of the AR model ('AR'), and the ratio of the factor model RMSE to that of the AR ('FM/AR'). Recursive forecasting scheme (forecasts of 1996Q1 to 2010Q4).

Table 3: Factor and AR models for 2-step forecasting using real- and pseudo real-time data

	Real-time		Pseudo Real-time	
	AR	FM/AR	AR	FM/AR
routputq	0.62	0.98	0.67	1.02
rconq	0.49	1.07	0.49	1.09
rconndq	0.79	0.98	0.67	1.00
rcondq	2.62	0.99	2.34	1.02
rinvbfq	2.61	0.90	2.23	0.88
rinvresq	3.36	1.06	3.41	1.10
rexq	2.82	1.00	2.64	0.99
rimpq	2.94	0.92	2.55	0.93
rgq	0.72	1.00	0.78	1.00
ipt	1.36	0.95	1.43	0.92
ipm	1.58	0.95	1.72	0.93
ruc	0.48	0.73	0.46	0.70
employ	0.38	0.93	0.39	0.90
hg	1.44	0.89	1.52	0.88
hs	0.48	0.96	0.51	0.93
cut	1.81	0.89	1.74	0.87
cum	2.02	0.91	1.91	0.89
hstarts	0.14	0.92	0.13	0.95
pq	0.25	1.00	0.23	1.03
cpi	0.72	1.11	0.72	1.13
pconq	0.51	1.06	0.52	1.09
pimpq	3.22	1.08	2.99	1.10
noutputq	0.75	1.00	0.79	1.00
nconq	0.74	1.00	0.79	1.02
wsdq	0.82	0.97	1.03	0.98
oliq	0.72	1.12	0.81	1.08
propiq	1.60	0.97	2.34	0.96
divq	5.02	0.96	6.38	0.95
pintiq	1.50	1.00	2.97	0.95
tranrq	1.74	0.98	1.84	0.97
sscontrq	2.01	0.98	2.01	0.97
npiq	0.74	0.99	1.04	1.01
ptaxq	5.31	0.96	5.17	0.96
ndpiq	0.99	1.00	1.11	1.04
pintpaiq	3.00	0.98	3.03	0.98
tranpfq	2.47	1.33	2.13	1.33
ratesavq	1.80	1.02	1.04	1.05
m1	1.50	1.03	1.38	1.04
m2	0.80	1.09	0.70	1.12
aaa	0.27	1.00	0.27	0.99
baa	0.39	1.01	0.39	1.01
tb3ms	0.48	1.09	0.48	1.09
fedfunds	0.53	1.17	0.53	1.20
gs10	0.35	1.03	0.35	1.02
Group and overall medians				
Real	1.40	0.95	1.48	0.93
Nominal	0.81	1.00	1.03	1.02
All	0.90	0.99	1.04	1.00

Notes. The table reports the RMSE of the AR model ('AR'), and the ratio of the factor model RMSE to that of the AR ('FM/AR'). Recursive forecasting scheme (forecasts of 1996Q1 to 2010Q4).

Table 4: Factor and AR models for 4-step forecasting using real- and pseudo real-time data

	Real-time		Pseudo Real-time	
	AR	FM/AR	AR	FM/AR
routputq	0.65	1.02	0.72	1.06
rconq	0.56	1.04	0.60	1.06
rconndq	0.84	0.99	0.74	0.99
rcondq	2.71	1.07	2.47	1.10
rinvbfq	2.83	0.99	2.64	0.96
rinvresq	3.63	1.12	3.81	1.12
rexq	2.69	1.05	2.61	1.06
rimpq	2.86	1.10	2.59	1.11
rgq	0.72	0.95	0.79	0.93
ipt	1.43	1.03	1.55	1.01
ipm	1.64	1.03	1.84	1.01
ruc	1.00	0.81	0.99	0.77
employ	0.51	0.98	0.55	0.95
hg	1.62	0.99	1.78	0.96
hs	0.58	0.97	0.65	0.96
cut	3.30	0.92	3.39	0.91
cum	3.68	0.93	3.75	0.91
hstarts	0.25	0.97	0.25	0.96
pq	0.30	1.02	0.28	1.18
cpi	0.71	1.08	0.72	1.12
pconq	0.53	1.10	0.55	1.16
pimpq	3.13	1.12	2.90	1.16
noutputq	0.80	1.00	0.87	1.02
nconq	0.79	1.00	0.86	1.00
wsdq	0.90	0.99	1.14	0.99
oliq	0.65	1.10	0.96	1.10
propiq	1.58	1.07	2.32	1.03
divq	4.79	1.01	5.92	0.99
pintiq	1.78	1.00	3.45	0.97
tranrq	1.56	1.00	1.62	0.99
sscontrq	1.72	1.08	1.60	1.14
npiq	0.81	0.99	1.21	1.03
ptaxq	5.32	1.00	5.34	1.01
ndpiq	1.03	1.01	1.19	1.02
pintpaiq	3.37	0.97	3.24	0.94
tranpfq	2.30	1.20	2.02	1.22
ratesavq	2.39	1.04	1.42	1.01
m1	1.75	1.01	1.68	1.02
m2	0.90	1.10	0.83	1.13
aaa	0.28	1.01	0.28	1.03
baa	0.39	0.99	0.39	0.98
tb3ms	0.46	1.03	0.46	1.04
fedfunds	0.51	1.06	0.51	1.05
gs10	0.37	1.01	0.37	1.02
Group and overall medians				
Real	1.53	0.99	1.66	0.97
Nominal	0.90	1.01	1.16	1.02
All	1.02	1.01	1.20	1.02

Notes. The table reports the RMSE of the AR model (‘AR’), and the ratio of the factor model RMSE to that of the AR (‘FM/AR’). Recursive forecasting scheme (forecasts of 1996Q1 to 2010Q4).

Table 5: Factor and AR models for 1 and 4-step forecasting using post-1984 observations

	FM - post 1984				AR & FM - post 1984			
	1-step		4-step		1-step		4-step	
	AR	FM/AR	AR	FM/AR	AR	FM/AR	AR	FM/AR
routputq	0.58	0.84	0.65	1.05	0.60	0.81	0.67	1.03
rconq	0.50	1.04	0.56	1.05	0.50	1.02	0.56	1.05
rconndq	0.82	0.97	0.84	1.08	0.83	0.95	0.86	1.05
rcondq	2.61	1.06	2.71	1.06	2.68	1.04	2.75	1.04
rinvbfq	2.39	0.85	2.83	1.06	2.46	0.83	2.91	1.03
rinvresq	3.10	0.99	3.63	1.00	3.15	0.98	3.72	0.98
rexq	2.71	0.92	2.69	1.04	2.58	0.97	2.76	1.01
rimpq	2.72	0.75	2.86	1.05	2.60	0.78	2.96	1.02
rgq	0.72	1.07	0.72	0.98	0.74	1.04	0.73	0.97
ipt	0.99	0.81	1.43	1.09	0.95	0.84	1.48	1.05
ipm	1.13	0.87	1.64	1.07	1.10	0.89	1.69	1.04
ruc	0.28	0.72	1.00	0.77	0.28	0.72	0.96	0.81
employ	0.27	0.77	0.51	0.90	0.26	0.80	0.45	1.01
hg	1.08	0.78	1.62	0.97	1.02	0.83	1.59	0.99
hs	0.38	0.82	0.58	0.98	0.36	0.87	0.57	1.01
cut	0.91	0.84	3.30	0.94	0.85	0.90	3.15	0.99
cum	1.01	0.87	3.68	0.96	0.96	0.92	3.59	0.98
hstarts	0.08	0.95	0.25	0.89	0.08	0.96	0.23	0.98
pq	0.27	0.89	0.30	0.84	0.24	0.99	0.28	0.92
cpi	0.66	0.95	0.71	0.83	0.60	1.04	0.59	1.00
pconq	0.46	0.97	0.53	0.88	0.43	1.04	0.44	1.04
pimpq	2.66	0.97	3.13	0.93	2.49	1.03	2.86	1.02
noutputq	0.68	0.78	0.80	0.88	0.61	0.87	0.70	1.00
nconq	0.72	0.89	0.79	0.92	0.68	0.94	0.72	1.00
wsdq	0.78	0.66	0.90	1.01	0.81	0.63	0.83	1.10
oliq	0.61	0.94	0.65	1.19	0.56	1.02	0.67	1.15
propiq	1.54	0.77	1.58	0.99	1.46	0.81	1.54	1.02
divq	5.85	0.96	4.79	1.01	6.13	0.92	4.78	1.01
pintiq	1.35	0.91	1.78	0.97	1.31	0.94	1.56	1.11
tranrq	1.74	0.86	1.56	0.96	1.67	0.90	1.51	1.00
sscontrq	2.00	0.76	1.72	0.96	1.64	0.93	1.61	1.03
npiq	0.70	0.76	0.81	0.96	0.67	0.79	0.74	1.05
ptaxq	5.37	0.80	5.32	1.00	5.33	0.80	5.26	1.01
ndpiq	1.03	0.83	1.03	0.92	0.88	0.97	0.95	0.99
pintpaiq	2.96	0.89	3.37	0.86	2.85	0.93	3.15	0.92
tranpfq	2.63	0.97	2.30	1.12	2.57	1.00	2.25	1.15
ratesavq	1.49	0.97	2.39	0.95	1.48	0.98	2.38	0.95
m1	1.33	1.16	1.75	1.06	1.38	1.12	1.76	1.05
m2	0.75	1.00	0.90	1.03	0.76	1.00	0.93	1.01
aaa	0.28	1.04	0.28	1.02	0.29	1.01	0.29	1.00
baa	0.38	1.05	0.39	1.00	0.37	1.07	0.39	1.00
tb3ms	0.45	0.91	0.46	1.11	0.36	1.13	0.47	1.09
fedfunds	0.47	0.88	0.51	1.08	0.37	1.12	0.50	1.09
gs10	0.37	1.03	0.37	1.03	0.38	0.99	0.37	1.02
Group and overall medians								
Real	0.95	0.86	1.53	1.02	0.90	0.89	1.53	1.01
Nominal	0.76	0.91	0.90	0.98	0.777	0.98	0.88	1.01
All	0.87	0.89	1.02	1.00	0.84	0.95	0.95	1.01

Notes. The table reports the RMSE of the AR model (‘AR’), and the ratio of the factor model RMSE to that of the AR (‘FM/AR’). Factors estimated on the whole sample, recursive forecasting scheme (forecasts of 1996Q1 to 2010Q4) using an estimation period of 1985 onwards for the FM (left panel) and FM and AR (right panel). Real-time forecasting exercise.

Table 6: Factor and AR models for 1 and 4-step forecasting using a rolling estimation window

	FM rolling only				FM and AR rolling			
	1-step		4-step		1-step		4-step	
	AR	FM/AR	AR	FM/AR	AR	FM/AR	AR	FM/AR
routputq	0.58	0.79	0.65	1.15	0.60	0.77	0.68	1.10
rconq	0.50	1.09	0.56	1.24	0.50	1.07	0.60	1.15
rconndq	0.82	1.04	0.84	1.25	0.89	0.95	0.95	1.10
rcondq	2.61	1.13	2.71	1.23	2.82	1.04	2.95	1.13
rinvbfq	2.39	0.86	2.83	1.16	2.56	0.81	3.00	1.09
rinvresq	3.10	1.05	3.63	1.03	3.36	0.97	3.81	0.98
rexq	2.71	0.93	2.69	1.08	2.57	0.98	2.78	1.05
rimpq	2.72	0.72	2.86	1.16	2.58	0.75	3.08	1.08
rgq	0.72	1.21	0.72	1.08	0.75	1.16	0.76	1.03
ipt	0.99	0.82	1.43	1.22	0.97	0.83	1.54	1.14
ipm	1.13	0.89	1.64	1.20	1.13	0.89	1.75	1.12
ruc	0.28	0.77	1.00	0.97	0.29	0.75	1.09	0.89
employ	0.27	0.72	0.51	0.97	0.26	0.73	0.45	1.09
hg	1.08	0.82	1.62	1.05	0.98	0.90	1.60	1.06
hs	0.38	0.84	0.58	1.09	0.35	0.91	0.57	1.12
cut	0.91	0.83	3.30	0.96	0.85	0.90	3.19	0.99
cum	1.01	0.89	3.68	1.00	0.94	0.95	3.61	1.02
hstarts	0.08	0.91	0.25	1.11	0.08	0.90	0.27	1.03
pq	0.27	0.90	0.30	0.82	0.24	1.01	0.28	0.91
cpi	0.66	0.97	0.71	0.84	0.63	1.02	0.62	0.96
pconq	0.46	0.95	0.53	0.84	0.43	1.02	0.43	1.03
pimpq	2.66	1.05	3.13	0.98	2.54	1.10	2.98	1.03
noutputq	0.68	0.73	0.80	0.93	0.60	0.82	0.70	1.06
nconq	0.72	0.90	0.79	0.98	0.67	0.96	0.72	1.08
wsdq	0.78	0.72	0.90	1.07	0.72	0.77	0.82	1.17
oliq	0.61	0.96	0.65	1.29	0.60	0.99	0.73	1.15
propiq	1.54	0.79	1.58	1.07	1.33	0.92	1.53	1.10
divq	5.85	1.11	4.79	1.08	7.12	0.91	4.80	1.08
pintiq	1.35	0.97	1.78	1.09	1.33	0.99	1.50	1.29
tranrq	1.74	0.93	1.56	1.06	1.73	0.93	1.60	1.03
sscontrq	2.00	0.76	1.72	0.96	1.57	0.97	1.59	1.04
npiq	0.70	0.77	0.81	1.04	0.64	0.83	0.73	1.15
ptaxq	5.37	0.78	5.32	1.12	5.45	0.77	5.58	1.07
ndpiq	1.03	0.81	1.03	0.94	0.86	0.97	0.95	1.02
pintpaiq	2.96	0.89	3.37	0.89	2.85	0.92	3.20	0.93
tranpfq	2.63	0.97	2.30	1.16	2.35	1.09	2.18	1.23
ratesavq	1.49	1.04	2.39	1.01	1.56	1.00	2.68	0.90
m1	1.33	1.26	1.75	1.09	1.42	1.18	1.83	1.04
m2	0.75	0.98	0.90	1.00	0.75	0.97	0.88	1.03
aaa	0.28	1.08	0.28	1.01	0.30	1.03	0.28	1.02
baa	0.38	1.15	0.39	0.99	0.39	1.11	0.39	0.98
tb3ms	0.45	0.99	0.46	1.18	0.37	1.19	0.46	1.18
fedfunds	0.47	0.95	0.51	1.16	0.38	1.18	0.50	1.18
gs10	0.37	1.01	0.37	1.08	0.38	0.98	0.37	1.07
Group and overall medians								
Real	0.95	0.88	1.53	1.10	0.92	0.90	1.57	1.08
Nominal	0.76	0.96	0.90	1.03	0.74	0.99	0.85	1.05
All	0.87	0.92	1.02	1.07	0.85	0.96	1.02	1.06

Notes. The table reports the RMSE of the AR model ('AR'), and the ratio of the factor model RMSE to that of the AR ('FM/AR'). Factors estimated on the whole sample, rolling forecasting scheme (forecasts of 1996Q1 to 2010Q4) using an estimation period of 1985 onwards for the FM (left panel) and FM and AR (right panel).
Real-time data. Factors estimated on whole sample.

Table 7: Explaining FM and AR relative 1-step forecasting performance: Infeasible FM forecasts

	Forecasts 1996-2011			Forecasts 1996-2007		
	FM/AR	FM_f/FM	FM_{Kn}/FM	FM/AR	FM_f/FM	FM_{Kn}/FM
routputq	0.83	0.99	0.74	1.10	0.98	0.53
rconq	0.98	1.04	0.67	1.04	1.06	0.60
rconndq	0.92	1.01	0.77	0.94	1.03	0.84
rcondq	0.99	1.00	0.76	1.05	1.01	0.75
rinvbfq	0.86	1.01	0.77	0.97	1.01	0.81
rinvresq	1.05	0.92	0.90	0.97	1.01	1.05
rexq	0.93	0.98	0.88	0.98	1.00	0.89
rimpq	0.76	1.07	0.95	1.02	0.93	0.80
rgq	1.04	0.97	0.94	1.02	0.98	1.06
ipt	0.87	1.00	0.75	1.10	0.96	0.66
ipm	0.91	1.00	0.66	1.16	0.96	0.69
ruc	0.80	0.95	0.82	0.84	0.91	0.91
employ	0.90	0.97	0.79	1.07	0.99	0.89
hg	0.89	0.99	0.76	1.17	0.96	0.73
hs	0.91	1.01	0.75	1.03	1.02	0.82
cut	0.88	0.96	0.59	1.07	0.97	0.61
cum	0.91	0.98	0.54	1.10	0.98	0.65
hstarts	0.99	1.06	0.91	1.08	0.97	0.83
pq	1.00	1.01	0.95	1.05	0.97	0.77
cpi	1.05	1.00	0.92	1.04	0.98	0.81
pconq	1.03	1.01	0.92	1.02	1.01	0.83
pimpq	1.05	0.98	0.87	1.01	0.99	0.92
noutputq	1.04	0.87	0.65	1.08	0.96	0.77
nconq	1.00	0.97	0.82	0.99	1.03	0.90
wsdq	0.86	0.97	0.81	1.02	0.96	1.04
oliq	1.01	0.92	0.83	0.99	1.02	0.89
propiq	0.84	1.02	0.95	1.01	0.97	0.94
divq	0.92	1.05	0.98	0.95	1.04	1.04
pintiq	0.95	1.04	0.81	0.97	1.00	0.75
tranrq	0.95	0.99	1.10	1.07	0.96	1.09
sscontrq	0.93	1.00	0.65	1.02	1.01	1.01
npiq	0.94	0.90	0.68	1.00	0.91	0.80
ptaxq	0.83	1.01	0.86	0.91	1.01	0.92
ndpiq	1.00	1.00	0.79	1.00	1.04	0.71
pintpaiq	0.96	0.99	0.80	0.96	1.00	0.83
tranpfq	1.11	0.87	0.67	1.09	0.91	0.69
ratesavq	0.99	1.00	0.50	0.99	1.01	0.48
m1	1.04	0.99	0.95	0.95	0.99	0.86
m2	1.04	1.00	0.81	1.00	1.00	0.89
aaa	1.00	1.08	0.76	0.99	1.08	0.81
baa	1.00	1.00	0.89	1.00	1.08	0.81
tb3ms	1.08	0.98	1.03	0.95	0.97	1.13
fedfunds	1.11	0.96	1.15	0.89	0.96	1.37
gs10	1.00	1.15	0.79	0.98	1.13	0.89
Group and overall medians of RMSE ratios						
Real	0.91	1.00	0.77	1.05	0.98	0.81
Nominal	1.00	1.00	0.83	1.00	1.00	0.87
All	0.97	1.00	0.81	1.01	0.99	0.83

Notes. The table reports the ratio of the FM to AR RMSE (the second column, replicating the third column of table 2, the ratio of the 'iterated' FM (FM_f), that uses forecasts of the factors, to the FM, and the FM with known factors (FM_{Kn}) to the FM. Real-time, recursive forecasting scheme from 1970 onwards.

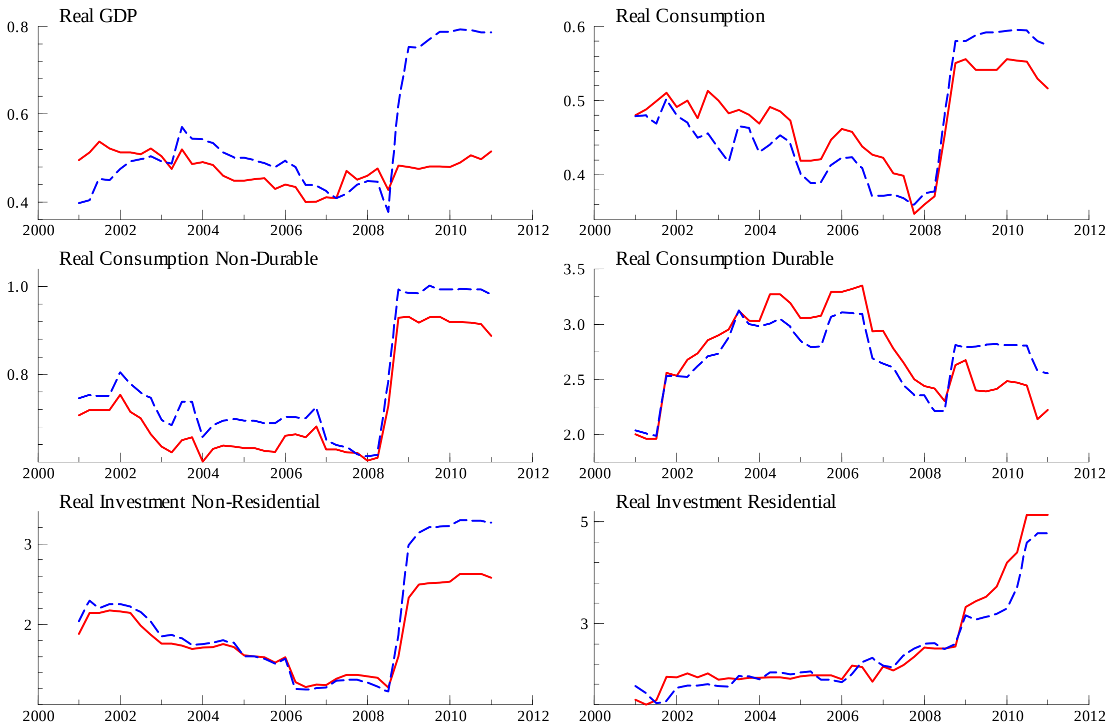

Figure 1: Rolling MSFEs (5 years to date shown) of the FM and AR model forecasts (solid and dashed lines, respectively. Recursive real-time forecasts.

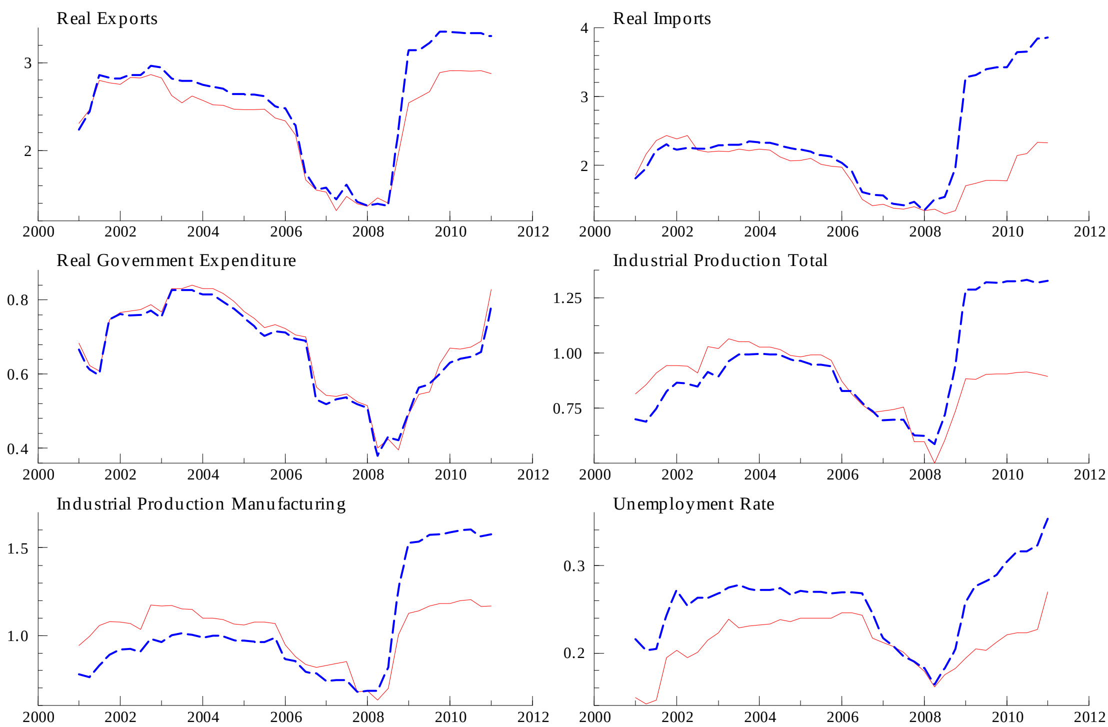

Figure 2: Rolling MSFEs (5 years to date shown) of the FM and AR model forecasts (solid and dashed lines, respectively. Recursive real-time forecasts.

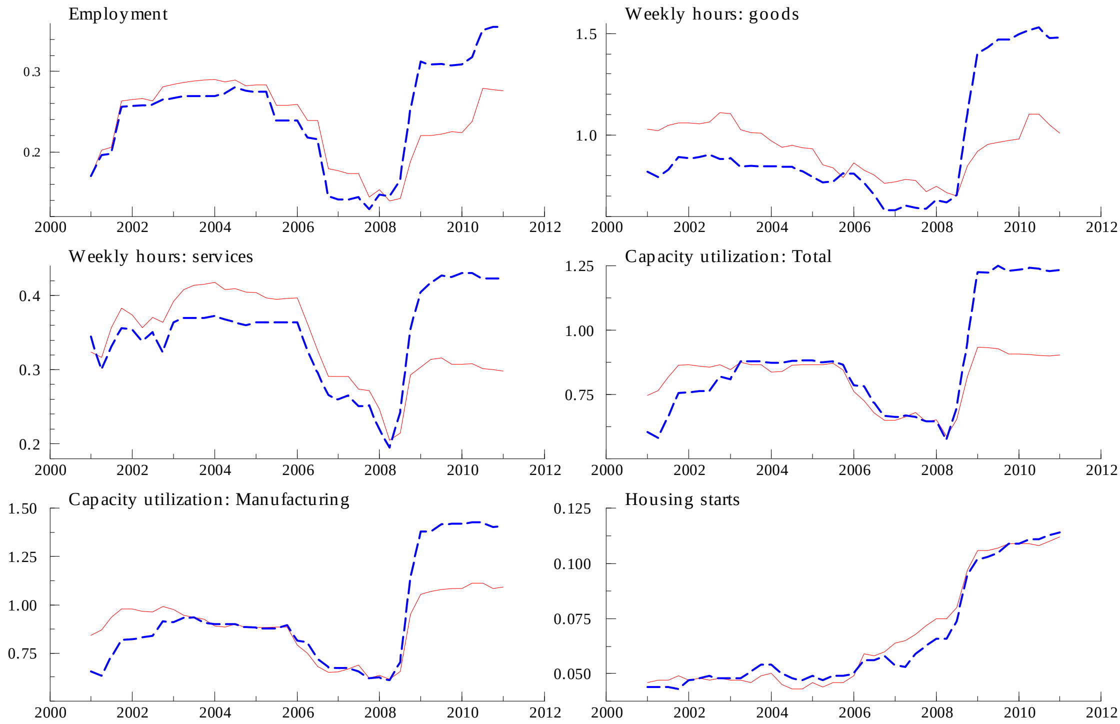

Figure 3: Rolling MSFEs (5 years to date shown) of the FM and AR model forecasts (solid and dashed lines, respectively. Recursive real-time forecasts.

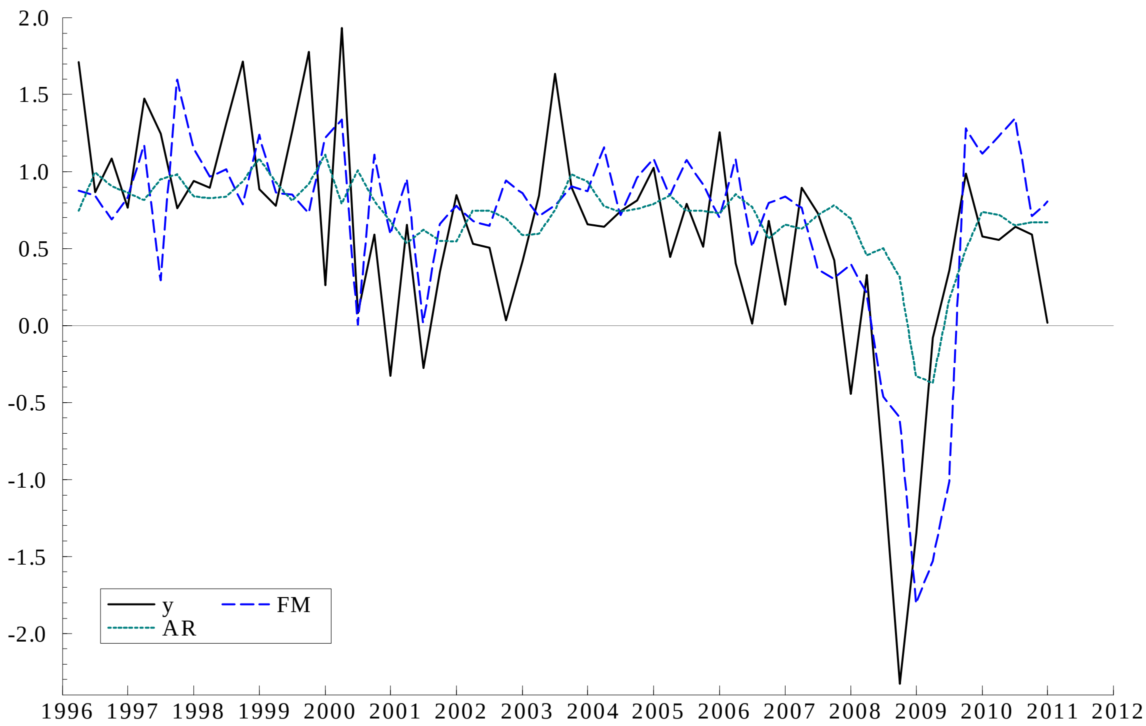

Figure 4: FM and AR forecasts of Real GDP growth, with the outturns (y). Using fully-revised data, and recursively-estimated models (estimation sample beginning in 1970).