ECARES

Forecasting with Approximate Dynamic Factor Models: the Role of *Non-Pervasive* Shocks

Matteo Luciani
ECARES, SBS-EM, Université Libre de Bruxelles

ECARES working paper 2011-022

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Forecasting with Approximate Dynamic Factor Models: the role of *non-pervasive* shocks

Matteo Luciani*
ECARES, Université libre de Bruxelles

July 2011

Abstract

In this paper we investigate whether accounting for *non-pervasive* shocks improves the forecast of a factor model. We compare four models on a large panel of US quarterly data: factor models, factor models estimated on selected variables, Bayesian shrinkage, and factor models together with Bayesian shrinkage for the idiosyncratic component. The results of the forecasting exercise show that the four approaches considered perform equally well and produce highly correlated forecasts, meaning that *non-pervasive* shocks are of no helps in forecasting. We conclude that comovements captured by factor models are informative enough to make accurate forecasts.

JEL Classification: C13, C32, C33, C52, C53
Keywords: Dynamic Factor Models, Penalized Regressions, Local Factors, Bayesian Shrinkage, Forecasting

1 Introduction

In recent years the literature has proposed two methodologies to cope with the curse of dimensionality problem, namely: factor models (Forni et al., 2000; Stock and Watson, 2002a) and Bayesian shrinkage (De Mol et al., 2008). Roughly speaking, the main idea of factor models is to *summarize* the information content of a large number of predictors in a few factors, while the idea of Bayesian shrinkage is to limit estimation uncertainty by *shrinking* the potentially complex model toward a simple *naive* prior model.[1]

*Address: ECARES, Solvay Brussels School of Economics and Management, Université libre de Bruxelles, 50 Av F.D. Roosevelt CP114/04, B1050 Brussels, Belgium. Phone:+32 (0)2 650 33 66. E-mail: *matteo.luciani@ulb.ac.be*. I am particularly indebted to Matteo Barigozzi, Domenico Giannone, and Marco Lippi for useful suggestions and comments. I have benefited also from discussions with Cecilia Frale, and Libero Monteforte. This paper is a revised version of the fourth chapter of my Ph.D. thesis at the University of Rome "La Sapienza". I wrote a large part of this paper while I was visiting at the Italy's Ministry of the Economy and Finance, Treasury Department - Office for Economic and Financial Analysis, whose hospitality is gratefully acknowledge. I also gratefully acknowledge financial support from the Belgian National Bank and the IAP P6/07 contract, from the IAP program (Belgian Scientific Policy), "Economic policy and finance in the global economy". Of course, any errors are my responsibility.

[1]Other methods, not used in this paper, able to forecast with a large number of predictors are Partial Least Squares (Groen and Kapetanios, 2008), Forecast Combination (Bates and Granger, 1969), Bayesian Model Average (Leamer, 1978), and Bagging (Breiman, 1996). For a review on forecasting with many predictors see Stock and Watson (2006).

Due to the strong comovement among macroeconomic time series, factor models offer a realistic and parsimonious representation of the data since they assume that a small number of *pervasive* shocks drives the bulk of comovement in the data. Moreover, factor models can be easily estimated with the method of principal components under general assumptions on the cross-correlation of the idiosyncratic errors (Bai and Ng, 2002; Bai, 2003; Forni et al., 2000, 2005; Stock and Watson, 2002a). That is, in factor models idiosyncratic errors are allowed to be "weakly" cross-sectionally-dynamically correlated (approximate factor structure), a likely feature in large macroeconomic database where *non-pervasive* (sectoral or regional) shocks might affect groups of variables (local factors).

Factor models have proved to be successful in predicting economic activity (see among others Boivin and Ng, 2005, and D'Agostino and Giannone, 2011, and Eickmeier and Ziegler, 2008, for a review). However, in particular when used in forecasting, traditional factor models do not model local factors, an issue that Bai and Ng (2008b) highlighted as one of the unresolved issue in factor models. This implies that if there is something predictable other than the common factors, then the forecast could be further improved.

In this paper we study the role of *non-pervasive* shocks in factor forecasting. Our aim is to understand whether there are local dynamics (i.e. local factors) strong enough to be exploited to improve the forecast. We argue that by augmenting the approximate dynamic factor model with a sparse model for the idiosyncratic component we are able to account for possible local factors. The goal of the paper is to quantify possible gains in terms of forecasting performance.

The rationale of our approach is to capture the bulk of the comovements in the data with the common factors, and to capture local dynamics by searching for those few variables driven by the same *non-pervasive* shocks that drive the target variable. Our method consists in mixing factor models and L_1 penalized regressions, which are equivalent to Bayesian shrinkage with double exponential priors. We choose L_1 penalized regressions because they perform both shrinkage and variable selection, and thus impose a sparse structure on the idiosyncratic component. This sparse structure is particularly appropriate for our purpose since we are interested in capturing *non-pervasive* shocks which, by definition, affect only a limited number of variables.

The literature recently suggested a different forecasting strategy also involving L_1 penalized regressions. This method is used by Bai and Ng (2008a) and De Mol et al. (2008). The former suggest extracting the factors only from those variables that are really informative for forecasting the target variable; while the latter suggest to select the predictors and to estimate the model by using only the selected predictors. Moreover, De Mol et al. (2008) show that L_1 penalized regressions and factor models are indeed intimately related: if the data are highly correlated, as it is the case in presence of a factor structure, a few variables, if appropriately selected, are able to capture the bulk of comovement in the data.

Although our approach uses the same method of Bai and Ng (2008a) and of De Mol et al. (2008), it is theoretically different since we impose a sparse structure only on the idiosyncratic component, while they impose a sparse structure on the whole dataset. That is, we first extract what is common, and then we impose sparsity on what is left.

An alternative method to account for local factors, used in forecasting by Bańbura and Modugno (2010), consists in estimating a factor model with global and local factors either with maximum likelihood techniques (Doz et al., 2011), or with Bayesian methods (Kose et al., 2008; Moench et al., 2009).[2] However, we do not consider this approach since, in order

[2]Hallin and Liška (2010) suggest a method involving dynamic principal components which is able to account

to identify local factors, it requires *a priori* information on the structure of the economy. In contrast, our method automatically identifies local factors by performing variable selection.

Our method is tested by means of a pseudo real time forecasting exercise on US quarterly data against the factor model of Forni et al. (2005), and the methods of Bai and Ng (2008a) and De Mol et al. (2008).

The rest of the paper is organized as follows: section 2 illustrates the methodology, and section 3 gives details of the forecasting exercise. In section 4 we present results of our empirical exercise, and in section 5 we conclude.

2 Methodology

Let x_t be an $N \times 1$ vector of stationary variables. Suppose we are interested in forecasting the i-th variable h steps ahead, $x_{i,t+h|t}$, by using all the N potential predictors. In this case, the best linear prediction defined as

$$x_{i,t+h|t} = Proj\{x_{i,t+h}|\Omega_t\},$$

where $\Omega_t = span\{x_{t-p}, p = 0, 1, \ldots\}$, might be extremely inefficient, or even impossible, due to lack of degrees of freedom. This is the well known curse of dimensionality problem. In recent years the literature has suggested two solutions: factor models (Forni et al., 2000; Stock and Watson, 2002a) and Bayesian shrinkage (De Mol et al., 2008).

If the comovements of the N variables in x_t can be well approximated by a small number $r \ll N$ of factors F_t, while the variable specific dynamics (ξ_t) are only mildly correlated, then the information set can be split in two orthogonal spaces: the space spanned by the factors and the space spanned by the idiosyncratic component ($\Omega_t = \Omega_t^F \cup \Omega_t^\xi$, where $\Omega_t^F = span\{F_{t-p}, p = 0, 1, \ldots\}$, and $\Omega_t^\xi = span\{\xi_{t-p}, p = 0, 1, \ldots\}$, with $\Omega_t^F \cap \Omega_t^\xi = \emptyset$). The idea of factor models is to approximate the linear projection on the whole information set by the sum of the linear projection on the space spanned by the factors, and of the linear projection on the space spanned by the present and past values of the variable specific dynamic:

$$x_{i,t+h|t} = Proj\{x_{i,t+h}|\Omega_t\} \approx Proj\{x_{i,t+h}|\Omega_t^F\} + Proj\{x_{i,t+h}|\Omega_{it}^\xi\}$$

where $\Omega_{it}^\xi = span\{\xi_{i,t-p}, p = 0, 1, \ldots\}$. That is, factor models solve the curse of dimensionality problem by *summarizing* the information content of a large number of predictors in a few factors.

The idea of Bayesian shrinkage is to limit estimation uncertainty by *shrinking* the potentially complex model toward a simple *naive* prior model. Namely, Bayesian shrinkage solve the curse of dimensionality by

$$x_{i,t+h|t} = Proj_s\{x_{i,t+h}|\Omega_t\},$$

where $Proj_s$ means linear projection with *shrinkage* of the parameter, and $\Omega_t = span\{x_{t-p}, p = 0, 1, \ldots\}$. Bayesian shrinkage are strictly related with penalized regression. In particular, Bayesian shrinkage with Gaussian priors are equivalent to L_2 norm penalized regressions, also known as ridge regressions; while Bayesian shrinkage with double exponential priors are equivalent to L_1 norm penalized regressions, also known under the labels of Lasso Regressions, or

for local factors but cannot be used in forecasting.

Least Angle Regressions (LARS).[3] Moreover, De Mol et al. (2008) show that both L_1 and L_2 penalized regressions are intimately related with factor models. In fact, on the one hand when the data are highly correlated, just as the factors, a few variables selected with an L_1 regression are able to capture the bulk of comovement in the data; on the other hand both L_2 penalized regressions and factor models produce linear combinations of all the variables in the dataset while differing only in the weighting scheme.

Factor models have proved successful in predicting economic activity (see Eickmeier and Ziegler, 2008, for a review). However, in particular when used in forecasting, traditional factor models do not model local factors. This implies that if there is something predictable other than the common factors, than the forecast could be further improved.

In this paper we augment the factor model with a sparse model for the idiosyncratic component thus accounting for local factors when forecasting. Our approach consists in mixing factor models and L_1 shrinkage. Our idea is to capture the bulk of the comovements in the data via the factor model, and to capture the local correlation via L_1 shrinkage, namely:

$$x_{i,t+h|t} = Proj\{x_{i,t+h}|\Omega_t^F\} + Proj_s\{x_{i,t+h}|\Omega_t^\xi\}$$

where $Proj\{x_{i,t+h}|\Omega_t^F\}$ is the linear projection onto the space spanned by the static factors, and $Proj_s\{x_{i,t+h}|\Omega_t^\xi\}$ is the linear projection with shrinkage onto the space spanned by the idiosyncratic components. In practice, we estimate $Proj\{x_{i,t+h}|\Omega_t^F\}$ by the two-step procedure of Forni et al. (2005), while we estimate $Proj_s\{x_{i,t+h}|\Omega_t^\xi\}$ by L_1 penalized regressions.[4] We suggest that L_1 penalty, which performs both shrinkage and variables selection, rather than L_2 penalty, which performs shrinkage and variables aggregation, is the appropriate penalization since, differently from L_2 penalty, it imposes a sparse structure on the idiosyncratic components.[5] This sparse structure is particularly appropriate for our purpose since we are interested in capturing local factors which, by definition, are related only to a limited number of variables.

Similarly to factor models, our approach summarizes the comovements in the data with a few factors; whereas, differently from factor models, our approach considers the whole space spanned by the idiosyncratic component rather then approximating it by the present and past values of the variable specific dynamic. Similarly to Bayesian shrinkage our approach considers the whole information set spanned by x_{t-p} (Ω_t), but, differently from Bayesian shrinkage, our approach exploits the fact that Ω_t can be split in two orthogonal spaces ($\Omega_t = \Omega_t^F \cup \Omega_t^\xi$).

3 The Forecasting Exercise

In the next section we will evaluate the forecasting performance of four methods: (1) factor models (Forni et al., 2005);[6] (2) factor models and selection of predictors (this paper); (3)

[3] To clarify, Lasso (Least Absolute Shrinkage and Selection Operator), and LARS (Least Angle Regressions), are the name of two algorithms proposed in Tibshirani (1996) and Efron et al. (2004) to estimate L_1 penalized regressions.

[4] In this paper, in order to estimate L_1 norm penalized regressions we use the LARS algorithm. A full description of the Forni et al. (2005) methodology, of penalized regressions, and of the LARS algorithm is available in the appendix.

[5] Another methodology which performs both shrinkage and variables selection, and which is gaining increasing attention in the economic literature, is Boosting (Freund and Schapire, 1997). In the appendix we also provide results obtained with this algorithm.

[6] Stock and Watson (2002a) suggest a method to forecast with factor models slightly different from the one of Forni et al. (2005) (for a comparison of the two methodologies see D'Agostino and Giannone, 2011). Result

factor models estimated on selected variables (Bai and Ng, 2008a); and (4) selection of predictors (De Mol et al., 2008, Bayesian shrinkage). Table 1 presents the complete list of estimated models.

Table 1: *Estimated Models*

Naive Approach

N°	Label	Forecast Equation
0	*naive*	$x^h_{i,t+h\|t} = \alpha_i + v_t, v_t \sim wn(0,1)$

Forni et al. (2005) Approach:

N°	Label	Forecast Equation	Forecast of the idiosyncratic component
1	FHLR	$x^h_{i,t+h\|t} = \chi^h_{i,t+h\|t} + \xi^h_{i,t+h\|t}$	Linear Projection

Luciani Approach:

N°	Label	Forecast Equation	Forecast of the idiosyncratic component
2	ML	$x^h_{i,t+h\|t} = \chi^h_{i,t+h\|t} + \xi^h_{i,t+h\|t}$	L_1 penalized regressions

Bai and Ng (2008a) Approach:

N°	Label	Forecast Equation	Forecast of the idiosyncratic component
3	$BN_1^{\dagger}$	$x^h_{i,t+h\|t} = \chi^h_{i,t+h\|t} + \xi^h_{i,t+h\|t}$	Linear Projection
4	$BN_2^{\ddagger}$	–	–

† Variables are selected with L_1 penalized regressions.
‡ Variables are selected according to economic theory.

De Mol et al. (2008) Approach:

N°	Label	Forecast Equation	Estimation Method
5	DGR	$x^h_{i,t+h\|t} = \beta(L)x_{i,t} + \varepsilon_t$	L_1 penalized regressions

Model 0 is our benchmark and consists in a *naive* forecast in which the forecast is simply a constant.

In model 1 to 4 the forecast of $x^h_{i,t+h}$ is obtained as the sum of the forecast of the common component ($\chi^h_{i,t+h|t}$) and of the forecast of the idiosyncratic component ($\xi^h_{i,t+h|t}$). In all four models, the forecast of the common component is obtained by the two-step method of Forni et al. (2005). In FHLR, BN_1, and BN_2, the idiosyncratic component is either neglected as in Forni et al. (2003), or a forecast is obtained as the linear projection of $\xi^h_{i,t+h}$ on the present and past values of x_i as suggested by D'Agostino and Giannone (2011). Differently, in model ML the forecast of the idiosyncratic component is obtained by an L_1 penalized regression estimated with the LARS algorithm. In BN_1 and BN_2 we run the two-step procedure on selected predictors: for the former model, the predictors are selected with L_1 penalized regressions; while for the latter model, predictors are selected with economic theory an approach also

obtained with the method of Stock and Watson (2002a) are not significantly different from those obtained with Forni et al. (2005) and are available in a supplementary appendix to this paper.

adopted by Bańbura et al. (2010a,b) both in forecasting with Large Bayesian VARs and in nowcasting with factor models.[7] Finally, in model DGR the forecast of the i-th variable is computed by L_1 penalized regressions.

We use the method of direct forecast (Stock and Watson, 2002b): let $X_{i,t}$ be the raw variable assumed to be integrated of order one, then $x^h_{i,t+h} = 100 \times (\log(X_{i,t+h}) - \log(X_{i,t}))$, that is the growth rate between period t and period $t+h$. When forecasting with models 1-4, given that both the forecast of the common component and the forecast of the idiosyncratic component are obtained on standardized data it is necessary to re-attribute the mean and the standard deviation of the series to obtain $x^h_{i,t+h}$. Hence, we have that $x^h_{i,t+h} = \sigma_i(\chi^h_{i,t+h} + \xi^h_{i,t+h}) + \mu_i$, where μ_i and σ_i are, respectively, the mean and the standard deviation of x_i.

For each methodology we select a benchmark specification that we will use for comparison throughout the paper. However, in a supplementary appendix we show results obtained with a large number of parameter configurations.

For model 1-3 our benchmark specification includes 2 dynamic factors ($q = 2$) and 4 static factors ($r = 4$), while for model 4 it includes 1 dynamic and 2 static factors.[8] For model FHLR, BN_1, and BN_2 we consider the case with no idiosyncratic component forecast.[9] For model ML we run LARS on a panel including the first four lags of the vector of idiosyncratic component (ξ_{t-p}, $p = 0, \dots, 3$) and we retain the $2r$-th iteration. Finally, in order to implement forecast with model 5 we run LARS on a database including the first four lags of each variable (x_{t-p}, $p = 0, \dots, 3$) and we retain the $2r$-th iteration.

4 Empirical Application

In this section we evaluate the forecasting performance of different models. The analysis is carried out on a panel of 104 quarterly series describing the US economy. The variables cover 12 different categories: Industrial Production, Consumer Price Indexes, Producer Price Indexes, Monetary Aggregates, Banking, GDP and Components, Housing Sector, Productivity & Cost, Interest Rates, Employment and Population, Survey, Financial Markets. All variables are transformed to reach stationarity and then standardized to have zero mean and unit variance thus preventing possible scale effects when extracting the factors. The complete list of variables and transformations is reported in the Appendix.

[7] In the empirical application for BN1 we select 52, while for BN2 we select 26 variables: Industrial Production (INDPRO), Consumer Price (CPIAUCSL), Producer Price (PPIACO), M1 (M1SL), Loans and Investments (LOANINV), GDP (GDPC), Residential Investment (PRFIC), Non Residential Investment (PNFIC), Imports (IMPGSC), Exports (EXPGSC), Government Consumption & Investment (GCEC), Private Inventories (CBIC), Consumption of Nondurable Goods (PCNDGC), Consumption of Services (PCESVC), Consumption of Durable Goods (PCDGCC), GDP Price Index (GDPCTPI), Building Permit (PERMIT), Fed Funds Rate (FEDFUNDS), 10-Year Treasury Rate (GS10), Employment (CE16OV), Unemployment Rate (UNRATE), Duration of Unemployment (UEMPMEAN), Oil Price (OILPRICE), Business Surveys Production (PRODUCTION), Standard & Poors, 500 (GSPC). For detailed informations on the database see next section and the appendix.

[8] The number of factors is selected with the Onatski (2009b) test, and the IC_1 criterion of Bai and Ng (2002, table 3)

[9] The choice of not considering the idiosyncratic component forecast is based on D'Agostino and Giannone (2011) according to which predicting the idiosyncratic component with linear projection do not improve the forecast. This result is confirmed also in our dataset (see the complementary appendix). Note that our method differs since it consist in predicting the idiosyncratic component with linear projection with *shrinkage* where the information set is the whole space spanned by the idiosyncratic component.

The forecasting exercise is performed on six variables: GDP (GDPC), Residential Investments (PRFIC), Non Residential Investments (PNFIC), Consumption of Non Durable Goods (PCNDGC), Consumption of Durable Goods (PCDGCC), and Consumption of Services (PCESVC).

Forecasts are produced according to a rolling scheme by using at each point in time observations for the last 20 years (80 observations): the first estimation is carried out on a sample from 1960:3 to 1980:2, while the last estimation on a sample from 1988:4 to 2008:3. For each variable and for each method we produce one step and four steps ahead forecasts. Overall we produced 114 forecasts for the one step ahead (1980:3-2008:4), and 111 for the four steps ahead (1981:2-2008:4).

Table 4 reports relative mean squared errors which are computed relative to a *naive* forecast. An entry lower than 1 means that the m-th model beats the *naive* forecast, while an entry greater than 1 means that model m does worse than the *naive* model. Asterisks indicate a rejection of the test of equal predictive accuracy between each model and the *naive* model.[10]

As we can see from table 4, all models outperform naive forecast but when forecasting PCDGCC and PCNDGCC,[11] and, overall, they perform very similar with a slight better performance for model BN_1, BN_2, and DGR, which confirms that *a priori* selection of predictors is a powerful strategy in forecasting. Finally, comparing ML with FHLR we can conclude that the there is no advantage in modeling the idiosyncratic component by accounting for local factors. This means that common factors capture not only the bulk of comovement, but also the dynamic of macroeconomic time series.

Table 5 shows the correlation between the forecasts obtained with FHLR and those obtained with models 2-5. To confirm the results of table 4, the forecast produced by FHLR and ML are always highly correlated (above 90%). Moreover, albeit some exceptions, the forecast obtained with both BN_1 and DGR (note that both models select predictors with L_1 regressions) are highly correlated with FHLR. As pointed out De Mol et al. (2008) if data are highly collinear the factors and few variables selected with L_1 regressions contain essentially the same information.

4.1 Sub-Sample Forecasts

As it is well known, parameter instability and structural breaks often occurs in economic time series. This suggests to study the forecasting performance of our model over time. Figure 2 shows the mean squared errors of FHLR, ML, BN_1, and DGR relative to the *naive* forecast with 10 years moving windows; while figure 3 shows at each point in time the difference between the squared forecasting errors of the m-th model and of the *naive* forecast. To make the graphs intelligible we do not show the MSE of BN_2. In both graph the shaded areas denote US recessions as dated by the NBER.

Figure 2 and figure 3 show that, for all variables but PNFIC, (*i*) after the starts of the so-called "Great Moderation" the forecasting performance of FHLR, ML, BN_1, and DGR has declined remarkably (see also D'Agostino et al., 2006; D'Agostino and Giannone, 2011); and, (*ii*) that the major gain from large database forecasts occur during recessions, in particular

[10]The test used here is the test of "Unconditional Predictive Ability" of Giacomini and White (2006) which is equivalent to the Diebold and Mariano (1995) test statistic. When using rolling window estimations, this test statistics has a standard normal limit distribution.

[11]The unpredictability of PCDGCC and of PCNDGCC is explained in figure 1 which shows how the growth rate of these two variables is nearly white noise.

the 1991 and the 2008 recessions. This is so because periods of high comovements, such as recessions, are those in which factor models produce the best representation of the economy (see also D'Agostino and Giannone, 2011).

Finally, by comparing the performance of FHLR and ML we can see that the two models perform very similar over time. This result strengthen the conclusion of the previous section: common factors capture not only the bulk of comovement, but also the dynamic of macroeconomic time series.

4.2 Are More Data Always Harmful for Factor Forecasting?

To test the robustness of our method we investigate whether its performance depends on the composition of the dataset, that is we evaluate how the method that we propose in this paper is affected by increased cross sectional correlations in the idiosyncratic components.

This robustness check links this paper also to another open issue of factor models, namely: excess cross correlation might harm factor estimates (Bai and Ng, 2008b). In fact, although the literature has shown that asymptotically "weakly" cross-sectionally correlated errors do not affect the factor forecast, in empirical application this might necessarily not be the case. Simulations in Boivin and Ng (2006) show that, as the cross correlation among idiosyncratic errors increases the estimation and the forecasting performance of the model deteriorates. Onatski (2009a) shows that if some factors are weakly pervasive so that *pervasive* and *non-pervasive* shocks cannot be clearly distinguished, then the principal component estimator is inconsistent, and simulation results in Bai and Ng (2008b) confirm that in this situation the factor estimates can be severely compromised.

How to construct a database is a practical problem for which it does not exists a cookbook, and which forecasters unavoidably face when forecasting with factor models. In principle, given that factor estimates are consistent as $N \to \infty$, including all available variables is a natural choice. However, as pointed out by Boivin and Ng (2006), if by adding an extra variable we are not adding information about the factors, rather simply extra cross sectional correlation among idiosyncratic errors, then the estimate of the factors deteriorates. This is indeed a concrete possibility. Think for example to Producer Price Indexes (PPI). In our dataset we included six PPIs, but from the FRED database one can easily retrieve data for more than fifteen PPIs. However, unless one is willing to comment on all of them, including all fifteen indexes is a bad choice since, as a matter of fact, this will only increase the cross-sectional correlation among idiosyncratic errors, and thus deteriorates the factor estimates.

To perform this exercise we mix up the dataset by adding many artificial variables with the goal of mimicking a practical situation as the one just described. We increase the number of variables in the dataset by copying all producer price indexes (6 variables), and all consumer price indexes (7 variables). To avoid perfect collinearity, every time we copy one variable we add some noise to this variable: $x_{it}^a = x_{it} + v_{it}$, where the superscript a stands for "artificial", and v_{it} is an error term. We do this four times, meaning that overall we add 52 variables, and each time we specify a different error term:

1. $v_{it} \sim N(0, \frac{1}{4}\sigma_i^2)$;
2. $v_{it} = 0.2v_{it-1} + \epsilon_{it}$, with $\epsilon_{it} \sim N(0, \frac{1}{5}\sigma_i^2)$;
3. $v_{it} = \epsilon_{it} + 0.2\epsilon_{it-1}$, with $\epsilon_{it} \sim N(0, \frac{1}{5}\sigma_i^2)$;
4. $v_{it} = 0.2v_{it-1} + \epsilon_{it} + 0.2\epsilon_{it-1}$, with $\epsilon_{it} \sim N(0, \frac{1}{10}\sigma_i^2)$.

By running this procedure we are sure that we are not adding any information rather that we are adding noise. In addition, in this way we heavily unbalance the database towards nominal variables: in the new database we have that 41% of the variables are price indexes, compared to 10% in the benchmark database.

Table 6 shows the cross correlation in the true and in the artificial database. As we can see the cross correlation is considerably increased: in the new database more than 25% of the variables have the highest (in absolute value) cross correlation coefficient greater than 0.9, while in the benchmark database the same percentage was less than 5%. Furthermore, in the new database more than 10% of the idiosyncratic components have the second (in absolute value) highest cross correlation coefficient greater than 0.9, while in the benchmark database the same percentage was 3%. In accordance with the conclusions in Boivin and Ng (2006) the performance of our model should deteriorate consistently.

A first consequence of the increased cross-correlation is that the tests for the number of factors detect a higher number of factors. Recall that in the benchmark database the Onatski (2009b) test and the IC_1 criteria of Bai and Ng (2002) detected respectively 2 dynamic, and 4 static factors. Differently, in the new database the Onatski (2009b) test does not find any factor, while IC_1 detects 6 static factors.[12] However, we know for sure that the number of factors has not increased since we have artificially added just noise.

Table 7 replicates the forecasting exercise of table 4 on the augmented database. Although we know it is wrong, we assume that there are 4 commons shocks, and 6 static factors. Asterisks indicate a rejection of the test of equal predictive accuracy between the m-th model estimated on the benchmark database, and the m-th model estimated on the augmented database.

The results in table 7 are quite surprising: the forecasting performance of models 1-5 is basically not affected by the increased noise in the database. That is, if we follow the indication of the tests, and we pick a higher number of factors the performance of models 1-5 is near identical to the one obtained in the benchmark database. We conclude that factor forecasts are robust to the composition of the database.

5 Conclusions

This paper studies the role of *non-pervasive* shocks in factor forecasting. Our aim is to understand whether there are local dynamics (i.e. local factors) strong enough to be exploited to improve the forecast and to quantify possible gains in terms of forecasting performance.

By means of a forecasting exercise we compare four approaches: (1) factor models (Forni et al., 2005); (2) factor models estimated on selected variables (Bai and Ng, 2008a); (3) selection of predictors (Bayesian shrinkage, De Mol et al., 2008); and (4) factor models and selection of predictors (this paper). In the latter case we augment the approximate dynamic factor model with a sparse model for the idiosyncratic component thus being able to account for possible local factors. The rationale of this approach is to capture the bulk of the comovements in the data with the factor model, and to capture local dynamics via L_1 penalized regressions by searching for those few variables driven by the same *non-pervasive* shocks that drive the target variable.

[12] In order to determine the number of dynamic factors in the new database we used the criterion of Hallin and Liška (2007) which indicates the presence of four dynamic factors. The same criteria applied to the benchmark database indicates 2 dynamic factors as the Onatski (2009b) test.

The analysis is carried out by means of a pseudo real-time exercise performed on quarterly US data for six real variables: GDP, Durables/Non Durables/Services Consumption, and Residential/Non Residential Investment.

The results of the forecasting exercise show that the four approaches considered perform equally well and produce highly correlated forecasts meaning that, after all, these models capture the same information. These results suggest that the local dynamics left over once controlling for common factors is not very substantial, and hence that the idiosyncratic component is of no helps in forecasting. We conclude that common factors not only capture the comovements, but also the bulk of the dynamics in the data.

Finally, we test the robustness of these models to the composition of the dataset. Results show that factor forecasts are extremely robust and that more data are not always harmful for factor forecasting.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*, 135–171.

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

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

Bai, J. and S. Ng (2008b). Large dimensional factor analysis. *Foundations and Trends® in Econometrics 3*(2), 89–163.

Bai, J. and S. Ng (2009). Boosting diffusion indices. *Journal of Applied Econometrics 24*(4), 607–629.

Bańbura, M., D. Giannone, and L. Reichlin (2010a). Large bayesian vector auto regressions. *Journal of Applied Econometrics 25*, 71–92.

Bańbura, M., D. Giannone, and L. Reichlin (2010b). Nowcasting. Discussion Papers 7883, CEPR.

Bańbura, M. and M. Modugno (2010). Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. Working Paper Series 1189, European Central Bank.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*, pp. 451–468.

Boivin, J. and S. Ng (2005). Understanding and Comparing Factor-Based Forecasts. *International Journal of Central Banking 1*, 117–151.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 127*, 169–194.

Breiman, L. (1996). Bagging predictors. *Machine Learning 24*, 123–140.

Bühlmann, P. (2006). Boosting for high-dimensional linear models. *Annals of Statistics 34*, 559–583.

Bühlmann, P. and B. Yu (2003). Boosting with the L_2 loss: Regression and classification. *Journal of the American Statistical Association 98*, 324–339.

D'Agostino, A. and D. Giannone (2011). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics*. forthcoming.

D'Agostino, A., D. Giannone, and P. Surico (2006). (Un)Predictability and macroeconomic stability. Working Paper Series 605, European Central Bank.

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

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*, 253–263.

Doz, C., D. Giannone, and L. Reichlin (2011). A quasi maximum likelihood approach for large approximate dynamic factor models. *Review of Economics and Statistics*. forthcoming.

Efron, B., T. Hastie, L. Johnstone, and R. Tibshirani (2004). Least Angle Regression. *Annals of Statistics 32*, 407–499.

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The Generalized Dynamic Factor Model: Identification and Estimation. *The Review of Economics and Statistics 82*, 540–554.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003). Do financial variables help forecasting inflation and real activity in the Euro Area? *Journal of Monetary Economics 50*, 1243–1255.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2004). The Generalized Dynamic Factor Model: Consistency and Rates. *Journal of Econometrics 119*, 231–255.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The Generalized Dynamic Factor Model: One Sided Estimation and Forecasting. *Journal of the American Statistical Association 100*, 830–840.

Forni, M. and M. Lippi (2001). The Generalized Dynamic Factor Model: Representation Theory. *Econometric Theory 17*, 1113–1141.

Forni, M. and M. Lippi (2010). The Unrestricted Dynamic Factor Model: One-sided Representation Results. *Journal of Econometrics*. forthcoming.

Freund, Y. and R. E. Schapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences 55*, 119–139.

Gelper, S. and C. Croux (2008). Least angle regression for time series forecasting with many predictors. Technical report, Katholieke Universiteit Leuven.

Giacomini, R. and H. White (2006). Tests of Conditional Predictive Ability. *Econometrica 74*, 1545–1578.

Groen, J. J. and G. Kapetanios (2008). Revisiting useful approaches to data-rich macroeconomic forecasting. Working Papers 624, Queen Mary, University of London, School of Economics and Finance.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*, 603–617.

Hallin, M. and R. Liška (2010). Dynamic factors in the presence of blocks. *Journal of Econometrics*. forthcoming.

Kose, A. M., C. Otrok, and C. H. Whiteman (2008). Understanding the evolution of world business cycles. *Journal of International Economics 75*, 110–130.

Leamer, E. E. (1978). *Specification Searches.* New York: Wiley.

Lutz, R. W. and P. Bühlmann (2006). Boosting for high-multivariate responses in high-dimensional linear regression. *Statistica Sinica 16*, 471–494.

Moench, E., S. Ng, and S. Potter (2009). Dynamic hierarchical factor models. Staff Reports 412, Federal Reserve Bank of New York.

Onatski, A. (2009a). Asymptotics of the principal components estimator of large factor models with weak factors. Manuscript, Columbia University.

Onatski, A. (2009b). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*, 1447–1479.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*, 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics 20*, 147–162.

Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting.* Elsevier.

Stock, J. H. and M. W. Watson (2010). Dynamic Factor Models. In M. P. Clements and D. F. Hendry (Eds.), *Oxford Handbook of Economic Forecasting.* Oxford: Oxford University Press.

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

Appendix A: Data Description and Data Treatment

N°	Series ID	Definition	F	SA	Unit	T	C
1	INDPRO	Industrial Production Index	M	1	2002=100	3	1
2	IPBUSEQ	Industrial Production: Business Equipment	M	1	2002=100	3	1
3	IPDCONGD	Industrial Production: Durable Consumer Goods	M	1	2002=100	3	1
4	IPDMAT	Industrial Production: Durable Materials	M	1	2002=100	3	1
5	IPNCONGD	Industrial Production: Nondurable Consumer Goods	M	1	2002=100	3	1
6	IPNMAT	Industrial Production: nondurable Materials	M	1	2002=100	3	1
7	CPIAUCSL	Consumer Price Index for All Urban Consumers: All Items	M	1	1982-84=100	4	2
8	CPIENGSL	CPIAUCs: Energy	M	1	1982-84=100	4	2
9	CPILEGSL	CPIAUCs: All Items Less Energy	M	1	1982-84=100	4	2
10	CPILFESL	CPIAUCs: All Items Less Food & Energy	M	1	1982-84=100	4	2
11	CPIUFDSL	CPIAUCs: Food	M	1	1982-84=100	4	2
12	CPIULFSL	CPIAUCs: All Items Less Food	M	1	1982-84=100	4	2
13	PPICRM	Producer Price Index: Crude Materials for Further Processing	M	1	1982 = 100	4	3
14	PPIENG	Producer Price Index: Fuels & Related Products & Power	M	0	1982 = 100	4	3
15	PPIFGS	Producer Price Index: Finished Goods	M	1	1982 = 100	4	3
16	PPIIDC	Producer Price Index: Industrial Commodities	M	0	1982 = 100	4	3
17	PPICPE	Producer Price Index: Finished Goods: Capital Equipment	M	1	1982 = 100	4	3
18	PPIACO	Producer Price Index: All Commodities	M	0	1982 = 100	4	3
19	PPIITM	Producer Price Index: Supplies & Components	M	1	1982 = 100	4	3
20	AMBSL	St. Louis Adjusted Monetary Base	M	1	Bil. of $	4	4
21	ADJRESSL	St. Louis Adjusted Reserves	M	1	Bil. of $	4	4
22	CURRSL	Currency Component of M1	M	1	Bil. of $	4	4
23	M1SL	M1 Money Stock	M	1	Bil. of $	4	4
24	M2SL	M2 Money Stock	M	1	Bil. of $	4	4
25	BUSLOANS	Commercial and Industrial Loans at All Commercial Banks	M	1	Bil. of $	3	5
26	CONSUMER	Consumer (Individual) Loans at All Commercial Banks	M	1	Bil. of $	3	5
27	LOANINV	Total Loans and Investments at All Commercial Banks	M	1	Bil. of $	3	5
28	LOANS	Total Loans and Leases at Commercial Banks	M	1	Bil. of $	3	5
29	REALLN	Real Estate Loans at All Commercial Banks	M	1	Bil. of $	3	5
30	TOTALSL	Total Consumer Credit Outstanding	M	1	Bil. of $	3	5
31	GDPC1	Real Gross Domestic Product, 1 Decimal	Q	1	Bil. of Ch. 2000 $	3	6
32	FINSLC1	Real Final Sales of Domestic Product, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
33	GPDIC1	Real Gross Private Domestic Investment, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
34	SLCEC1	Real State & Local Cons. Exp. & Gross Investment, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
35	PRFIC1	Real Private Residential Fixed Investment, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
36	PNFIC1	Real Private Nonresidential Fixed Investment, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
37	IMPGSC1	Real Imports of Goods & Services, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
38	GCEC1	Real Government Cons. Exp. & Gross Investment, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
39	EXPGSC1	Real Exports of Goods & Services, 1 Decimal	Q	1	Bil. of Ch. 2005 $	3	6
40	CBIC1	Real Change in Private Inventories, 1 Decimal	Q	1	Bil. of Ch. 2005 $	2	6
41	PCNDGC96	Real Personal Consumption Expenditures: Nondurable Goods	Q	1	Bil. of Ch. 2005 $	3	6
42	PCESVC96	Real Personal Consumption Expenditures: Services	Q	1	Bil. of Ch. 2005 $	3	6
43	PCDGCC96	Real Personal Consumption Expenditures: Durable Goods	Q	1	Bil. of Ch. 2005 $	3	6
44	PCECC96	Real Personal Consumption Expenditures	Q	1	Bil. of Ch. 2005 $	3	6
45	DGIC96	Real National Defense Gross Investment	Q	1	Bil. of Ch. 2005 $	3	6
46	NDGIC96	Real Federal Nondefense Gross Investment	Q	1	Bil. of Ch. 2005 $	3	6
47	DPIC96	Real Disposable Personal Income	Q	1	Bil. of Ch. 2005 $	3	6
48	PCECTPI	Personal Consumption Expenditures: Chain-type Price Index	Q	1	Index 2005=100	4	5
49	GPDICTPI	Gross Private Domestic Investment: Chain-type Price Index	Q	1	Index 2005=100	4	5
50	GDPCTPI	Gross Domestic Product: Chain-type Price Index	Q	1	Index 2005=100	4	5
51	HOUSTMW	Housing Starts in Midwest Census Region	M	1	Thous. of Units	3	7
52	HOUSTNE	Housing Starts in Northeast Census Region	M	1	Thous. of Units	3	7
53	HOUSTS	Housing Starts in South Census Region	M	1	Thous. of Units	3	7
54	HOUSTW	Housing Starts in West Census Region	M	1	Thous. of Units	3	7
55	PERMIT	New Private Housing Units Authorized by Building Permit	M	1	Thous. of Units	3	7
56	ULCMFG	Manufacturing Sector: Unit Labor Cost	Q	1	1992 = 100	3	8
57	COMPRMS	Manufacturing Sector: Real Compensation Per Hour	Q	1	1992 = 100	3	8
58	COMPMS	Manufacturing Sector: Compensation Per Hour	Q	1	1992 = 100	3	8
59	HOAMS	Manufacturing Sector: Hours of All Persons	Q	1	1992 = 100	3	8
60	OPHMFG	Manufacturing Sector: Output Per Hour of All Persons	Q	1	1992 = 100	3	8
61	ULCBS	Business Sector: Unit Labor Cost	Q	1	1992 = 100	3	8
62	RCPHBS	Business Sector: Real Compensation Per Hour	Q	1	1992 = 100	3	8
63	HCOMPBS	Business Sector: Compensation Per Hour	Q	1	1992 = 100	3	8
64	HOABS	Business Sector: Hours of All Persons	Q	1	1992 = 100	3	8
65	OPHPBS	Business Sector: Output Per Hour of All Persons	Q	1	1992 = 100	3	8
66	OILPRICE	Spot Oil Price: West Texas Intermediate	M	0	$ per Barrel	4	8
67	MPRIME	Bank Prime Loan Rate	M	0	%	2	9
68	FEDFUNDS	Effective Federal Funds Rate	M	0	%	2	9
69	TB3MS	3-Month Treasury Bill: Secondary Market Rate	M	0	%	2	9
70	GS1	1-Year Treasury Constant Maturity Rate	M	0	%	2	9
71	GS10	10-Year Treasury Constant Maturity Rate	M	0	%	2	9
72		GS1-FedFunds	M	0	%	2	9
73		GS10-FedFunds	M	0	%	2	9

74	EMRATIO	Civilian Employment-Population Ratio	M	2	%	2	10
75	CE16OV	Civilian Employment	M	2	Thous.	3	10
76	UNRATE	Civilian Unemployment Rate	M	2	%	2	10
77	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	M	2	Thous.	3	10
78	UEMP5TO14	Civilian Unemployed for 5-14 Weeks	M	2	Thous.	3	10
79	UEMP15T26	Civilians Unemployed for 15-26 Weeks	M	2	Thous.	3	10
80	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M	2	Thous.	3	10
81	UEMPMEAN	Average (Mean) Duration of Unemployment	M	2	Weeks	3	10
82	UNEMPLOY	Unemployed	M	2	Thous.	3	10
83	DMANEMP	All Employees: Durable Goods Manufacturing	M	2	Thous.	3	10
84	NDMANEMP	All Employees: Nondurable Goods Manufacturing	M	2	Thous.	3	10
85	SRVPRD	All Employees: Service-Providing Industries	M	2	Thous.	3	10
86	USCONS	All Employees: Construction	M	2	Thous.	3	10
87	USEHS	All Employees: Education & Health Services	M	2	Thous.	3	10
88	USFIRE	All Employees: Financial Activities	M	2	Thous.	3	10
89	USGOOD	All Employees: Goods-Producing Industries	M	2	Thous.	3	10
90	USGOVT	All Employees: Government	M	2	Thous.	3	10
91	USINFO	All Employees: Information Services	M	2	Thous.	3	10
92	USLAH	All Employees: Leisure & Hospitality	M	2	Thous.	3	10
93	USMINE	All Employees: Natural Resources & Mining	M	2	Thous.	3	10
94	USPBS	All Employees: Professional & Business Services	M	2	Thous.	3	10
95	USPRIV	All Employees: Total Private Industries	M	2	Thous.	3	10
96	USSERV	All Employees: Other Services	M	2	Thous.	3	10
97	USTPU	All Employees: Trade, Transportation & Utilities	M	2	Thous.	3	10
98	USWTRADE	All Employees: Wholesale Trade	M	2	Thous.	3	10
99	ORDERS	Business Surveys, ISM Manufacturing, New orders	M	1	index	2	11
100	PRODUCTION	Business Surveys, ISM Manufacturing, Production	M	1	index	2	11
101	EMPLOYMENT	Business Surveys, ISM Manufacturing, Employment	M	1	index	2	11
102	DELIVERIES	Business Surveys, ISM Manufacturing, Deliveries	M	1	index	2	11
103	INVENTORIES	Business Surveys, ISM Manufacturing, Inventories	M	1	index	2	11
104	GSPC	Standard & Poors, 500 Composite, Index, Price Return	M	0	index	3	12

NOTE: Outliers are detected as values differing from the median more than six times the interquantile difference. Outliers are replaced with the median of the five previous observations. All series are from the Fred II database of the Federal Reserve Bank of St. Louis with the exception of series 56, 71, 72 and of series 99-104. Series 56 is the Census Bureau House Price Index (*new one-family houses sold including value of lot*) deflated by the implicit price deflator for the nonfarm business sector (IPDNBS) taken from the Fred II database. Series 71 and 72 are own calculations. Series 99-103 are taken from ISM web site (http://www.ism.ws), while series 104 is from finance.yahoo.com.

Abbreviations

Categories	Frequency	Transformations	Seasonally Adjusted
1 = Industrial Production	Q = Quarterly	0 = none	0 = no
2 = Consumer Price Indexes	M = monthly*	1 = first difference	1 = yes
3 = Producer Price Indexes		2 = first difference of natural logarithm	2 = own SA⋆
4 = Monetary Aggregates			
5 = Banking			
6 = GDP and Components			
7 = Housing Sector			
8 = Productivity & Cost			
9 = Interest Rates			
10 = Employment and Population			
11 = Survey			
12 = Financial Markets			

* All monthly series are transformed into quarterly observation by simple averages.
⋆ Seasonal adjustment was obtained by regressing variables on a set of seasonal dummies.

Appendix B: Methodology

Factor Models

Let x_t be an $N \times 1$ vector of stationary variables. If x_t admits a *Factor Representation* (Forni et al., 2000; Forni and Lippi, 2001) it can be written as:

$$\begin{aligned} x_t &= \chi_t + \xi_t, \quad \text{for} \quad t = 1, \dots, T \\ \chi_t &= C(L)u_t, \end{aligned} \tag{1}$$

where u_t is a $q \times 1$ vector of *common shocks*, with $q \ll N$, $C(L) = \sum_{j=0}^{s} C_j L^j$ is an $N \times q$ matrix polynomial in the lag operator with square summable entries, s is the maximum allowed lag length, which, in principle, could be also infinite,[13] and χ_t and ξ_t are $N \times 1$ vectors containing respectively the *common component* and the *idiosyncratic component*. The common shocks and the idiosyncratic components are assumed to be uncorrelated at all leads and lags, while the idiosyncratic components are allowed to be both serially and cross-sectionally correlated albeit by a limited amount (approximate factor structure). For a complete review of factor models see Bai and Ng (2008b), and Stock and Watson (2010).

Forecasting with Factor Models: the Forni et al. (2005) Methodology

Suppose that our target variable is one of the entries of a vector x_t that admits a factor representation, say the i-th entry for simplicity, then a forecast of $x_{i,t+h|t}$ can be obtained as the sum of the forecast of the common and of the idiosyncratic component: $x_{i,t+h|t} = \chi_{i,t+h|t} + \xi_{i,t+h|t}$. Forni et al. (2005) (proposition 4) demonstrate that by means of a two-step estimator a forecast of the common component that converges to the best linear forecast of $\chi_{i,t+h|t}$ can be obtained, while they suggest that the idiosyncratic component can be neglected.[14]

Step 1: Let $\tilde{\Sigma}^{\chi}(\theta)$ and $\tilde{\Sigma}^{\xi}(\theta)$ be the estimated spectral density matrix of, respectively, the common and the idiosyncratic component obtained with the method of dynamic principal components, then the covariance matrices of χ_t, $\tilde{\Gamma}^{\chi}_k$, and ξ_t, $\tilde{\Gamma}^{\xi}(\theta)$, can be consistently estimated as the inverse Fourier transform of, respectively, $\tilde{\Sigma}^{\chi}(\theta)$ and $\tilde{\Sigma}^{\xi}(\theta)$.

Step 2: Let $\widehat{Z}$ be the $N \times r$ matrix containing the first normalized r eigenvectors of $\tilde{\Gamma}^{\chi}_0(\tilde{\Gamma}^{\xi}_0)^{-1}$, then the static factors can be estimated as the first r generalized principal components of x_t, $\widehat{F}_t = \widehat{Z}' x_t$. The factor loadings Λ can then be recovered as the linear projection of the static factors on x_t, $\Lambda = \tilde{\Gamma}^{\chi}_0 \hat{Z}(\hat{Z}'\tilde{\Gamma}^{\chi}_0 \hat{Z})^{-1}$. Having estimated both the factors and the loadings the forecast of the common components is obtained as: $\hat{\chi}_{t+h|t} = \tilde{\Gamma}^{\chi}_h \hat{Z}(\hat{Z}'\tilde{\Gamma}^{\chi}_0 \hat{Z})^{-1} \hat{Z}' x_t$.

Penalized Regressions

The main idea of penalized regressions is to *penalize* the coefficients of those variables that are less informative for predicting the target variable. Two type of penalizations are particularly used in the literature: the L_2 norm penalization, which is strictly connected with Bayesian regressions with Gaussian priors, and the L_1 norm penalization, which is strictly connected to Bayesian regression with double exponential priors. Let y be the $T \times 1$ vector containing the observations of our target variable, x be the $T \times N$ matrix containing the potential predictors, and suppose that we want to obtain a

[13] The case where $s = \infty$ is studied in Forni et al. (2000, 2004), Forni and Lippi (2001, 2010), Hallin and Liška (2007, 2010), and Onatski (2009b).

[14] A refinement of the Forni et al. (2005) procedure is proposed in D'Agostino and Giannone (2011) which suggest to forecast the idiosyncratic component as the linear projection of $\xi_{i,t+h|t}$ on $[x_{i,t}\ x_{i,t-1}\ \dots\ x_{i,t-p}]$.

forecast h step ahead. The L_2 penalization solves the problem:

$$\widehat{\beta} = \arg\min_{\beta} \sum_{t=h}^{T} (y_{t+h} - x_t\beta)^2 + \lambda \sum_{j=1}^{N} |\beta_j|^2 \tag{2}$$

which for $0 \leq \lambda < \infty$ shrink toward zero the coefficients of the uninformative predictors. However, as a result of the quadratic penalization, no parameter will exactly be set to zero, and hence uninformative predictors will still contribute to the forecast.

Alternatively, the L_1 penalized regression not only shrinks OLS coefficients towards zero, but also performs as a variable selection operator because it potentially sets $\beta_j = 0$, for some j. Formally, the L_1 penalization solves:

$$\widehat{\beta} = \arg\min_{\beta} \sum_{t=h}^{T} (y_{t+h} - x_t\beta)^2 + \lambda \sum_{j=1}^{N} |\beta_j| \tag{3}$$

where the tuning parameter λ controls for the amount of shrinkage, and thus for the number of parameters that are set to zero.

To gain intuition about L_1 penalized regressions let us briefly illustrate the case when the regressors in x are orthogonal. In this case the solution of equation (3) will be $\widehat{\beta}_j^{L_1} = \text{sign}\{\widehat{\beta}_j^{ols}\}(|\widehat{\beta}_j^{ols}| - \lambda/2)_+$ (see De Mol et al., 2008; Bai and Ng, 2008a), where $\widehat{\beta}_j^{ols}$ is the OLS estimate obtained by regressing $x_{j,t}$ on y_{t+h}, and $a_+ = a$ if $a_+ > 0$, and zero otherwise. This means that L_1 penalized regressions set to zero all coefficients which in absolute value are below the threshold $\lambda/2$, while they shrink all the other coefficients by an amount equal to the threshold.

A variant of L_1 penalized regressions, for which results are available in the complementary appendix, exploits the concept of hard thresholding. This method consists in first identifying the variables for which the coefficient is above the threshold, and then to use these variables to a linear regression of y_{t+h} thus obtaining β.

Least Angle Regressions

Least Angle Regressions (Efron et al., 2004, LARS) is an algorithm that estimates a linear relation between one variable and many potential predictors. The idea of Least Angle Regression is to build recursively an estimate of y by $x\hat{\beta}$ where at each stage a regressor is added. At the first stage the variable mostly correlated with y, say x_j, is selected, and an OLS regression of y on x_j is run. Define the residual of the first step as $v = y - \gamma\hat{\beta}_j x$, where γ is the step length, then the algorithm takes the largest step towards the direction of this residual until it finds another regressor, say x_l, as much correlated with v as x_j. Then the LARS algorithm search for the third variable equiangularly between x_j and x_l. At the k-th step, $\hat{\beta}$ has k non zero elements, and $N - k$ zero elements.

Formally, suppose that the algorithm was run k times.[15] Let $\hat{\mu}^k$ be the prediction obtained after k steps, and let $\hat{c} = x(y - \hat{\mu}^k)$ be the $(N - k) \times 1$ vector of current correlation among the $N - k$ excluded variables and the residual of the k-th step. Define $\mathcal{A}$ the *active set*, that is the set of indexes corresponding to the variables mostly correlated (in absolute value) with the residual: $\hat{C} = \max_j\{|\hat{c}_j|\}$ and $\mathcal{A} = \{j : \hat{c}_j = \hat{C}\}$. Let $s_j = sign(\hat{c}_j)$, then define the *active matrix* as $X_{\mathcal{A}} = (s_j x_j)_{j\in\mathcal{A}}$. Let $G_{\mathcal{A}} = X'_{\mathcal{A}}X_{\mathcal{A}}$, and $A_{\mathcal{A}} = (\mathbf{1}'_{\mathcal{A}}G_{\mathcal{A}}^{-1}\mathbf{1}_{\mathcal{A}})^{-1/2}$, where $\mathbf{1}_{\mathcal{A}}$ is a $(k + 1) \times 1$ vector of ones. Define the *equiangular vector* $\mathrm{u}_{\mathcal{A}} = X_{\mathcal{A}}w_{\mathcal{A}}$ where $w_{\mathcal{A}} = A_{\mathcal{A}}G_{\mathcal{A}}^{-1}\mathbf{1}_{\mathcal{A}}$, the vector making equals angle less than $90°$, meaning not orthogonal, with the columns of $X_{\mathcal{A}}$, such that $X'_{\mathcal{A}}\mathrm{u}_{\mathcal{A}} = A_{\mathcal{A}}\mathbf{1}_{\mathcal{A}}$. Then, the LARS algorithm updates $\hat{\mu}$ as:

$$\hat{\mu}^{k+1} = \hat{\mu}^k + \hat{\gamma}\mathrm{u}_{\mathcal{A}},$$

[15] This is the LARS algorithm presented in Efron et al. (2004). An extension not used in this paper of the LARS algorithm for time series problems is suggested in Gelper and Croux (2008).

where

$$\hat{\gamma} = \min_{j \in \mathcal{A}^c}{}^{+} \left\{ \frac{\hat{C} - \hat{c}_j}{A_{\mathcal{A}} - a_j}, \frac{\hat{C} + \hat{c}_j}{A_{\mathcal{A}} + a_j} \right\},$$

"+" indicates that the minimum is taken on positive values only, and a_j is the j-th element of the inner product vector $a = X'\mathrm{u}_{\mathcal{A}}$.

Some further detail: (i) at the first step we set $\hat{\mu}^0 = 0$, so that the first variable included in the active set is the one mostly correlated with the target variable. (ii) The direction of the search, and the updating rule are computed endogenously by the algorithm. (iii) At each iteration, the estimator $\hat{\beta}_k$ is obtained as $\hat{\beta}_k = \hat{\beta}_{k-1} + \gamma w'$.

Appendix C: Boosting

As said in footnote 5, another methodology which performs both shrinkage and variables selection, and which is gaining increasing attention in the economic literature, is Boosting (Freund and Schapire, 1997). Originating from the machine learning literature, it has been shown to be useful in regression with a large number of predictors (Bühlmann and Yu, 2003; Bühlmann, 2006; Lutz and Bühlmann, 2006). The idea of boosting is to build an estimate of y by recursively estimating regressions of y on x_j, where x_j is the variable most powerful in predicting y. At each step the prediction is update by $\hat{\mu} = \gamma \hat{b}_j x_j$, where γ is the step length, and $\hat{b}_j x_j$ is the linear projection of x_j on u, the residual obtained in the previous step.

Formally, let $Z_t = \{x_{1,t}, \ldots, x_{1,t-p_x}, \ \ldots, x_{N,t}, \ldots, x_{N,t-p}\}'$, be the $Np \times 1$ vector containing all the N potential predictors and their first p lags. Suppose that the algorithm was run k times, let $\hat{\mu}^k$ be the prediction obtained after k steps, and define the residual $u = y - \hat{\mu}^k$, then:[16]

1 for each $i = 1 \ldots \bar{N}$, regress u on Z_i, thus obtaining $\hat{b}_i$, and define $\hat{e}_i = u - x_i \hat{b}_i$, and $ssr_i = \hat{e}_i' \hat{e}_i$,

2 select the variable j such that $ssr_j = \min_{i=1,\ldots,\bar{N}} \{ssr_i\}$,

3 update the prediction as $\hat{\mu}^{k+1} = \hat{\mu}^k + \gamma x_j \hat{b}_j$, where $0 < \gamma \leq 1$ is the step length.

Some further detail: (i) at the first stage we set $\hat{\mu}^0 = \bar{y}$. (ii) At each iteration, the estimator $\hat{\beta}_k$ is obtained as $\hat{\beta}_k = \hat{\beta}_{k-1} + \gamma \hat{b}_k^{\dagger}$, where $\hat{b}_k^{\dagger}$ is an $\bar{N} \times 1$ vector where all entries are zero but the j-th entry. (iii) In order to fix the number of iterations, Bai and Ng (2009) suggest the following information criteria: $IC = \log \hat{\sigma}_m^2 + \frac{2}{T} \mathrm{df}_m + \frac{2}{N} c_m$, where $\mathrm{df}_m = \mathrm{trace}(B_m)$, $B_m = I_T - \prod_{j=0}^{m} (I_T - \gamma P^{(m)})$, $P^{(m)} = x_{j_m} (x_{j_m}' x_{j_m})^{-1} x_{j_m}$, $B_0 = \mathbf{1}_T \mathbf{1}_T' / T$, $c_m = \beta' \Sigma_x \beta / \sigma^2$, and $\hat{\sigma}_m^2 = \sum_{t=1}^{T} (y_t - \hat{\mu}_t^k)^2$. This criteria takes into account that some regressors might be estimated quantity as it is the case of factors. In this case, when we add an extra lag we are also adding an extra error component due to the fact that the factors are estimated. However, this error vanishes at a rate N, where N is the size of the panel from which factors are extracted.

Table 2: *Relative Mean Squared Errors*

	h	**FHLR**	**ML$_1$**	**ML$_2$**	**DGR$_1$**	**DGR$_2$**
GDPC	1	0.72**	0.72**	0.71**	0.67**	0.65**
	4	0.75*	0.76	0.78	0.79**	0.77
PRFIC	1	0.67**	0.68*	0.50**	0.37**	0.34**
	4	0.8	0.85	0.77	0.92	0.85
PNFIC	1	0.63**	0.62**	0.63**	0.68**	0.67**
	4	0.64**	0.65**	0.61**	0.83**	0.74*
PCNDGC	1	0.94	0.96	0.95	0.95	0.97
	4	0.86	0.9	0.88	0.85	0.84
PCESVC	1	0.88	0.84	0.9	0.89**	0.84**
	4	0.95	0.92	0.93	0.86*	0.83*
PCDGCC	1	0.91	0.95	0.82*	1	1.06
	4	0.9	1.01	0.92	0.78**	0.75**

Each cell reports the relative mean squared error of the m-th model compared to a *naive* forecast. An entry lower than 1 means that the m-th model beats the *naive* forecast, while an entry greater than 1 means that model m does worse than the *naive* model. Asterisks denote model forecasts that are statistically more accurate than the *naive* at 5% (**) and 10% (*) significance levels.

[16] This is the *component-wise* algorithm suggested by Bühlmann and Yu (2003). Bai and Ng (2009) suggest another algorithm labeled as *block-wise* boosting which is not used in this paper.

In table 2 we show forecasting results obtained with boosting. Model ML_2 applies boosting to forecast the idiosyncratic component, while model DGR_2 applies boosting directly on the variables. Model FHLR, ML_1 and DGR_1 are included just for comparison purposes. In model ML_2 and DGR_2 we run the "Component-Wise Algorithm" suggested by Bai and Ng (2009) on ξ_{t-p}, $p = 0, \ldots, 3$, we set the length step γ equal to 0.1, the maximum number of iterations to 30, and we select which iteration to retain with the modified BIC criterion suggested by Bai and Ng (2009).

As we can see from table 2 the performance of boosting is good no matter it is applied to the variables or to the idiosyncratic components. However, Boosting seems not able to improve with respect to L_1 penalized regressions estimated with the LARS algorithm.

Tables

Table 3: *Testing for the Number of Factors*

	DB$_0$				DB$_1$			
r	λ_i^x	R	μ_i^x	IC_1	λ_i^x	R	μ_i^x	IC_1
1	0.389	0.259	0.284	-0.262	0.372	0.260	0.239	-0.210
2	0.132	0.004	0.125	-0.369	0.145	0.450	0.118	-0.304
3	0.091	0.771	0.057	-0.385	0.090	0.600	0.083	-0.367
4	0.069	0.398	0.045	-0.389	0.072	0.255	0.066	-0.417
5	0.051	0.825	0.036	-0.381	0.049	0.471	0.049	-0.447
6	0.044	0.624	0.033	-0.373	0.042	0.590	0.045	-0.478
7	0.036	0.239	0.032	-0.369	0.036	0.377	0.027	-0.473
8	0.029	0.164	0.030	-0.367	0.030	0.388	0.024	-0.466

λ_i^x is the percentage of variance explained by i-th eigenvalue (in decreasing order) of the spectral density matrix of x. μ_i^x is the percentage of variance explained by i-th eigenvalue (in decreasing order) of the variance covariance matrix of x. R is the p-value of the Onatski (2009b) statistic for the null of $r-1$ dynamic factors against the alternative of r dynamic factors. IC_1 is the criteria of Bai and Ng (2002).

Table 4: *Relative Mean Squared Errors*

	h	**FHLR**	**ML**	**BN$_1$**	**BN$_2$**	**DGR**
GDPC	1	0.72**	0.72**	0.68**	0.55**	0.67**
	4	0.75*	0.76	0.75*	0.75	0.79**
PRFIC	1	0.67**	0.68*	0.54**	0.85	0.37**
	4	0.8	0.85	0.81	0.98	0.92
PNFIC	1	0.63**	0.62**	0.73**	0.72**	0.68**
	4	0.64**	0.65**	0.66*	0.81*	0.83**
PCNDGC	1	0.94	0.96	0.97	1.01	0.95
	4	0.86	0.9	0.97	1.03	0.85
PCESVC	1	0.88	0.84	0.81**	1	0.89**
	4	0.95	0.92	0.88*	1.08	0.86*
PCDGCC	1	0.91	0.95	0.98	1.09	1
	4	0.9	1.01	0.86	1.13	0.78**

Each cell reports the relative mean squared error of the m-th model compared to a *naive* forecast. An entry lower than 1 means that the m-th model beats the *naive* forecast, while an entry greater than 1 means that model m does worse than the *naive* model. Asterisks denote model forecasts that are statistically more accurate than the *naive* at 5% (**) and 10% (*) significance levels.

Table 5: *Correlation between Forecasts*

	h	**ML**	**BN_1**	**BN_2**	**DGR**
GDPC	1	0.98	0.86	0.84	0.7
	4	0.96	0.82	0.76	0.54
PRFIC	1	0.99	0.9	0.52	0.72
	4	0.97	0.84	0.5	0.45
PNFIC	1	0.98	0.88	0.7	0.85
	4	0.96	0.86	0.59	0.68
PCNDGC	1	0.95	0.65	0.33	0.41
	4	0.9	0.74	0.61	0.46
PCESVC	1	0.96	0.87	0.5	0.66
	4	0.93	0.9	0.7	0.73
PCDGCC	1	0.94	0.81	0.38	0.37
	4	0.87	0.85	0.52	0.61

Each cell report the correlation between the forecasts obtained with FHLR and the forecast obtained with model m.

Table 6: *Cross-Sectional Correlation between Idiosyncratic Components*

	DB_0			DB_{1a}			DB_{1b}		
	τ_1^*	τ_2^*	$\bar{\tau}$	τ_1^*	τ_2^*	$\bar{\tau}$	τ_1^*	τ_2^*	$\bar{\tau}$
25%	0.519	0.389	0.119	0.564	0.453	0.125	0.574	0.454	0.125
50%	0.609	0.467	0.132	0.767	0.630	0.148	0.788	0.709	0.149
75%	0.722	0.584	0.147	0.931	0.851	0.164	0.935	0.879	0.179
90%	0.811	0.685	0.160	0.992	0.982	0.191	0.987	0.984	0.204
95%	0.888	0.733	0.168	0.996	0.994	0.198	0.996	0.989	0.216
99%	0.980	0.924	0.185	0.999	0.996	0.206	0.999	0.995	0.238

DB_0 is the benchmark database, while DB_1 is the artificially increased database. To compute the idiosyncratic components for DB_0 and DB_{1a} we use 2 dynamic factors and 4 static factors, while for DB_{1b} we use 4 dynamic and 6 static factors.
Define τ_{ij} the cross correlation between ξ_i and ξ_j, then:
1) $\tau_1^*(i) = \max_j |\tau_{ij}|$ is the largest cross correlation for the i-th idiosyncratic component;
2) $\tau_2^*(i)$ is the second largest cross correlation for the i-th idiosyncratic component;
3) $\bar{\tau} = \frac{1}{N-1}\sum_j \tau_{ij}$ is the average cross sectional of the ξ_i with the whole database.

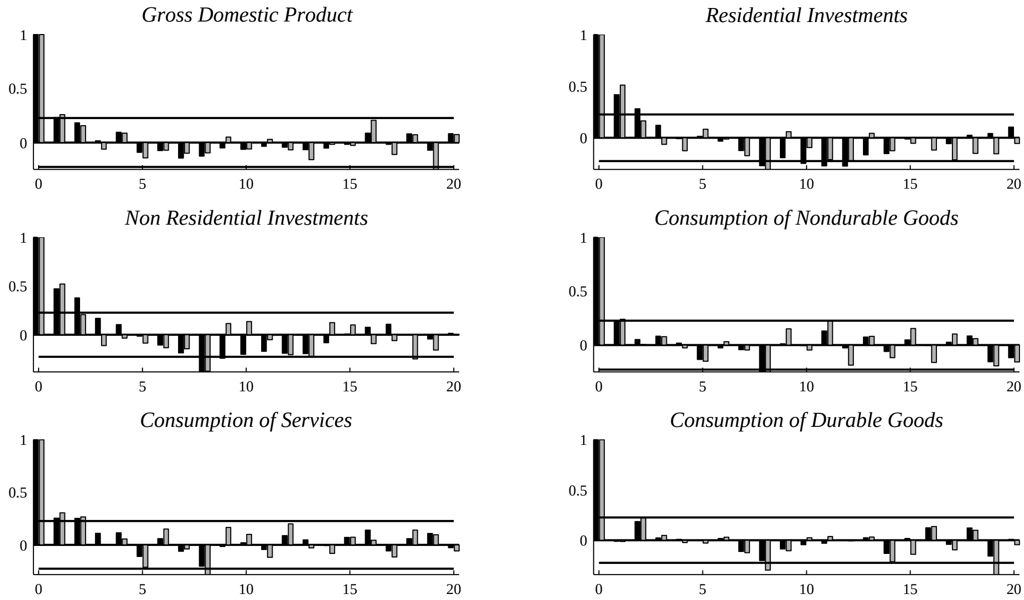
Table 7: *Relative Mean Squared Error*
Augmented Database
$q = 4$ - $r = 6$

	h	**FHLR**	**ML**	**BN$_1$**	**DGR**
GDPC	1	0.71	0.73	0.81*	0.66
	4	0.78	0.89	0.90*	0.79
PRFIC	1	0.8	0.78	0.76**	0.35
	4	0.88	0.89	0.85	0.88
PNFIC	1	0.59	0.62	0.79	0.68
	4	0.66	0.75	0.88*	0.79
PCNDGC	1	1	1.10**	0.99	0.95
	4	0.97	1.17**	0.91	0.85
PCESVC	1	0.93	0.96	0.94*	0.87
	4	0.95	0.99	1.01	0.85
PCDGCC	1	1.01	1.07	0.75**	1.04**
	4	0.94	1.13	0.85	0.73*

Each cell reports the relative mean squared error of the m-th model compared to a *naive* forecast. An entry lower than 1 means that the m-th model beats the *naive* forecast, while an entry greater than 1 means that model m does worse than the *naive* model. Asterisks denote rejection of the null of equal unconditional predictive ability between model m estimated on the benchmark database (table 4), and model m estimated on the augmented database (this table), respectively at 5% (**) and 10% (*) significance levels.

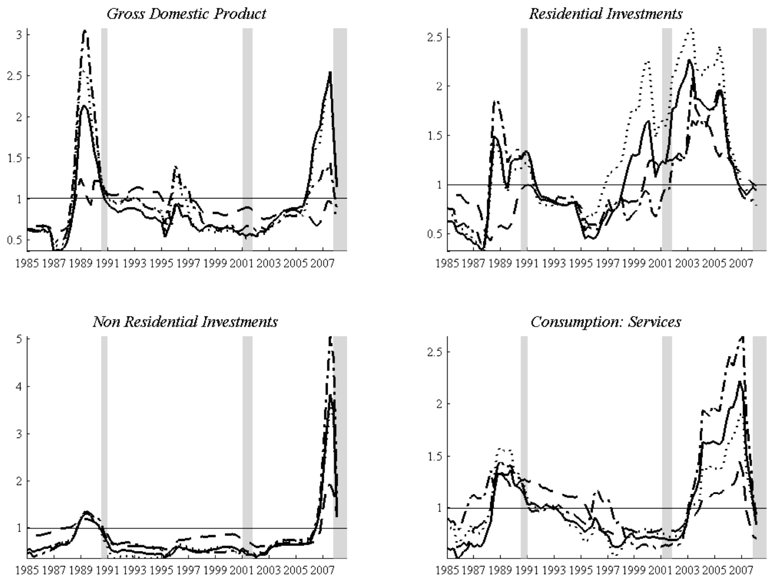
Graphs

Figure 1: *Autocorrelation Function for Target Variables*



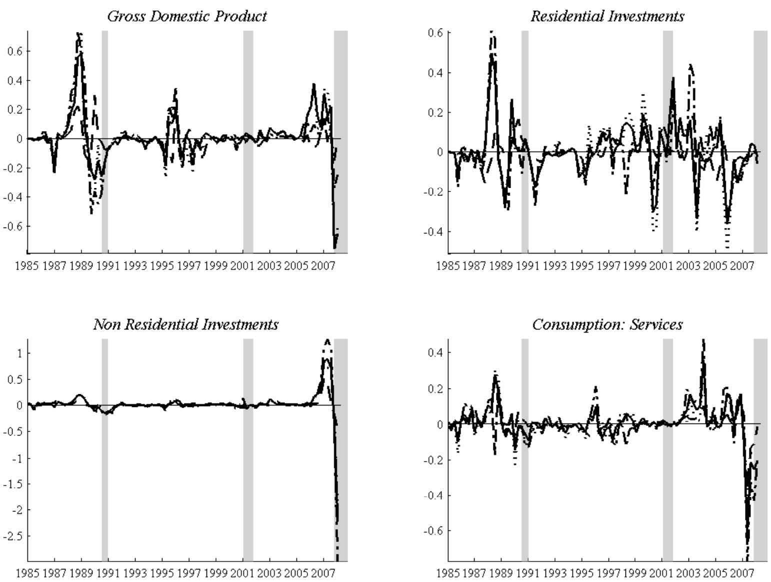
In each subplot, the black bar is the autocorrelation function of the growth rate of the variable computed on the sample 1960:3-2008:4, the gray bar is the partial autocorrelation function, while the straight lines are 95% confidence band.

Figure 2: *Evolution of Relative Mean Squared Error over time*
10 Years Window



Each line is the mean squared error of Forni et al. (2005) type models relative to the *naive* forecast with 10 years moving windows. Shaded areas denote the US recessions as dated by the NBER. Straight line = FHLR; dotted = ML_1; dashdot = BN_1; dashed = DGR_1.

Figure 3: *Difference Between the squared error of the m-th Model and the naive model*



Each line shows at each point in time the difference between the squared forecasting error of the m-th model and the *naive* forecast. At time t a value greater than 0 means that the naive forecast did better than the m-th model, while a value smaller than zero means that the m-th model was more accurate. Shaded areas denote the US recessions as dated by the NBER. Straight line = FHLR; dotted = ML_1; dashdot = BN_1; dashed = DGR_1.