On the Selection of Common Factors for Macroeconomic Forecasting

Giovannelli, Alessandro and Proietti, Tommaso

30 November 2014

On the Selection of Common Factors for Macroeconomic Forecasting

Alessandro Giovannelli
Università di Roma “Tor Vergata”

Tommaso Proietti
Università di Roma “Tor Vergata” and CREATES

Abstract

We address the problem of selecting the common factors that are relevant for forecasting macroeconomic variables. In economic forecasting using diffusion indexes the factors are ordered, according to their importance, in terms of relative variability, and are the same for each variable to predict, i.e. the process of selecting the factors is not supervised by the predictand. We propose a simple and operational supervised method, based on selecting the factors on the basis of their significance in the regression of the predictand on the predictors. Given a potentially large number of predictors, we consider linear transformations obtained by principal components analysis. The orthogonality of the components implies that the standard t-statistics for the inclusion of a particular component are independent, and thus applying a selection procedure that takes into account the multiplicity of the hypotheses tests is both correct and computationally feasible. We focus on three main multiple testing procedures: Holm's sequential method, controlling the family wise error rate, the Benjamini-Hochberg method, controlling the false discovery rate, and a procedure for incorporating prior information on the ordering of the components, based on weighting the p-values according to the eigenvalues associated to the components. We compare the empirical performances of these methods with the classical diffusion index (DI) approach proposed by Stock and Watson, conducting a pseudo-real time forecasting exercise, assessing the predictions of 8 macroeconomic variables using factors extracted from an U.S. dataset consisting of 121 quarterly time series. The overall conclusion is that nature is tricky, but essentially benign: the information that is relevant for prediction is effectively condensed by the first few factors. However, variable selection, leading to exclude some of the low order principal components, can lead to a sizable improvement in forecasting in specific cases. Only in one instance, real personal income, we were able to detect a significant contribution from high order components.

Keywords: Variable selection; Multiple testing; p-value weighting.

JEL Codes: C22, C52, C58.

1 Introduction

The focus of much recent theoretical and applied econometric research has concentrated on the ability to predict key macroeconomic variables, such as output and inflation, using a large number of potential predictors, with little or no a priori guidance over their relevance. This theme, which developed contextually to the statistical and machine learning literature on data mining and discovery, has received a very distinctive solution, hinging upon the idea that the wealth of information on macroeconomic variables can be distilled by a limited number of common factors.

The common factors capture the comovements among the economic variables and can be consistently estimated by principal components analysis (PCA), as in the static factorial approach proposed by Stock and Watson (2002a), or by dynamic principal components analysis, using frequency domain methods, as proposed by Forni et al. (2005). Quoting from Stock and Watson (2006), the availability of a factor structure and of closed form inferences has turned the high dimensionality of the information set from a curse to a blessing.

Once the factors are extracted, they can be used for forecasting the variables of interest, by augmenting an observation driven model, such as an autoregression, by the estimated factors. This approach, known as the diffusion index (DI), or factor augmented autoregressive (FAR), forecasting methodology, has become mainstream, owing its success to the ability to incorporate information carried by a large number of potential predictors in a simple and parsimonious way. Applied economic forecasting has shown that the consideration of the factors as potential predictors has proved successful in macroeconomic forecasting using large datasets; it would be impossible to provide a list of references that could be representative of the research carried out in this field. The reviews in Stock and Watson (2006), Breitung and Eickmeier (2006) and Stock and Watson (2010), as well as Ng (2013), provide ample coverage of the main issues.

As it is well known, the principal components, arising from the spectral decomposition of the sample covariance matrix of the predictors, are ranked according to the size of the corresponding eigenvalue. The current forecasting practice selects the first components according to an information criterion, such as Bai and Ng (2002) and Onatski (2010), and uses them as explanatory variable in the forecasting model en lieu of the original predictors. A potential limitation of this procedure is that the selection of factors is blind on the predictive ability of the principal components, as no consideration is given to their relationship with the predictand by the information criteria commonly used.

The lack of supervision of the principal components in regression has been the matter of an old debate, which is echoed in Cox (1968), Joliffe (1982), Hadi and Ling (1998) and Cook (2007), among others. There are essentially two opposite views: the argument of the critics is that there is no logical reason why the predictand should not be related to the least important principal components, and secondly that different predictands, such as output and inflation, cannot depend on the same r principal components. The counter argument, using Mosteller and Tukey quotation of Einstein (Mosteller and Tukey (1977), pp. 397–398), is that "nature is tricky, but not downright mean": the first principal components capture the underlying common dimensions of the economy. If this was the case, the leading principal components, those corresponding to the largest eigenvalues, should carry the essential information for predicting economic variables.

The selection of the factors that are relevant for the prediction of macroeconomic variables has attracted a lot of interest and several solutions have been proposed in the literature for supervising the factors, taking into account their ability to predict a specific dependent variable. Bai and Ng (2008) propose distilling the factors, referred to as "targeted predictors", by performing a PCA on a subset of the original predictors, that are selected according to the strength of the relationship with the variable to be predicted in a marginal regression framework. This is an instance of the method of supervised PCA Bair et al. (2006), aiming at finding linear combinations of the predictors that have high correlation with the outcome. Bai and Ng (2009) considered bootstrap aggregation of the predictions arising from a FAR framework, that retains only the significant factors. A comprehensive review of variable selection in predictive regression is Ng (2013).

The research question addressed by this paper is whether many predictors can be replaced by a reduced number of principal components selected according to the strength of the relationship with the predictand, and whether components beyond the firsts carry useful information for improving the predictive ability. We propose a simple and operational supervised method based on selecting the factors on the basis of their significance in the regression of the predictand on the predictors. Given a potentially large number of predictors, we consider linear transformations obtained by principal components analysis. The orthogonality of the components implies that the standard t statistics for the inclusion of a particular component in the multiple regression framework are independent, and thus applying a multiple testing procedure to select the components that are significant at a particular level is both correct and computationally feasible.

The selection of the principal components can be seen as a decision problem involving multiple-testing, where a single null hypothesis claims that a specific component ought to be excluded from the model. There are several multiple testing procedures available that focus on controlling some type of error rate, namely the familywise error rate, such as the Bonferroni and Holm (1979) procedure, or the false discovery rate, which is the expected proportion of wrong rejections. Among the procedures controlling for the false discovery rate, we focus on the Benjamini-Hochberg procedure (see Benjamini and Hochberg (1995)) and on a weighted procedure that allows to incorporate prior information about the ordering of the components; see Genovese et al. (2006).

In summary, our methodology has three steps:

1. Orthogonalise the original N predictors by computing the N standardised PCs.

2. Select r principal components on the basis of their correlation with the predictand, taking into account the multiplicity of the testing problem and controlling for the error rate of the selection procedure.

3. Use the selected components in a factor augmented autoregressive predictive regression.

Our method can be nested within the shrinkage representation for forecasting using orthogonal predictors proposed by Stock and Watson (2012b) and has analogies with the idea of targeted predictors, although the object of the selection are the principal components, rather than the original predictors: this has the advantage of not having to consider the correlation of the tests statistics for the inclusion of the predictors.

We validate our procedure using a dataset consisting of 121 quarterly U.S. macroeconomic time series observed from 1959-I to 2011-II. A pseudo real-time rolling forecast experiment is conducted to compare the performance of our selection method to that of a benchmark autoregressive predictor, with order selected by an information criterion, and to the standard DI forecasts based on the first five components.

The paper is structured as follows. In section 2 we provide a brief review of the diffusion index methodology. Section 3 considers the issue of estimating supervised factors and reviews the main solutions available in the literature. In section 4 we present principal components regression as a shrinkage methods and discuss the issues posed by the selection of the components and the consequences in terms of forecasting accuracy. Section 5 exposes our supervised method using a multiple testing approach to the selection of the principal components in the FAR predictor.

2 Forecasting using principal components

Let $X_t = (X_{1t}, \dots, X_{Nt})'$ denote an $N \times 1$ vector of predictors observed at times $t = 1, \dots T$, and let $y_{t+h}^{(h)}$ denote the predictand, where $h > 0$ is the forecast lead, and y denotes a transformation of the original variable Y, which depends on its order of integration.

The Diffusion Index (DI) forecasting methodology, originally proposed by Stock and Watson (2002), provides a simple and parsimonious way of incorporating a highly dimensional information set; it is based on the assumption that the predictors have an approximate factor structure, such that the unobserved factors can be estimated consistently by principal component analysis (PCA). The factor model is formulated as follows:

$$X_t = \Lambda F_t + \xi_t, \tag{1}$$

where $F_t = (F_{1t}, \dots, F_{rt})$ with $r < N$ are the unobserved common factors, Λ is the $n \times r$ matrix of factor loadings and ξ_t is the idiosyncratic disturbance not explained by the factors.

Letting $S = T^{-1} \sum_t X_t X_t'$ and denoting the spectral decomposition of the covariance matrix by $S = VD^2V'$, where $V = (v_1, \dots, v_N)$ is the $(N \times N)$ matrix of eigenvectors, $V'V = VV' = I_N$, and $D = \text{diag}(d_1, \dots, d_N)$ is the matrix containing the square root of the ordered eigenvalues, $d_1 \geq d_2 \geq \dots \geq d_N$, as in Stock and Watson (2002a), the common factors F_t are estimated by the first r standardised principal components

$$\hat{F}_t = D_r^{-1/2} V_r' X_t, \tag{2}$$

where $D_r = \text{diag}(d_1, \dots, d_r)$ and $V_r = (v_1, \dots, v_r)$.

We assume that we are interested in predicting a variable y_t (which may as well be included in the set X_t) at the at horizon $h > 0$, by using all the information contained in X_t. For instance, if we are interested in forecasting quarterly industrial production, denoted Y_t, h quarters ahead, we set $y_{t+h}^{(h)} = 400(\ln Y_t - \ln Y_{t-1})$, which assumes that $\ln y_t$ is difference stationary. In predicting $y_{t+h}^{(h)}$ we also the estimated common factors and the lags of $y_t = (\ln Y_t - \ln Y_{t-1})$, according to factor augmented autoregressive (FAR) model:

$$y_{t+h}^{(h)} = \mu + \sum_{j=1}^{p} \phi_j^{(h)} y_{t-j+1} + \sum_{k=1}^{r} \beta_k^{(h)} \hat{F}_{kt} + \varepsilon_{t+h}, \tag{3}$$

where ε_{t+h} is the forecasting error with variance σ^2.

The DI forecasts are obtained according to a two step procedure: in the first step r factors are extracted from the set of predictors by performing a PCA and selecting the number of common factors according to information criteria proposed by Bai and Ng (2002), such as

$$IC_{p1}(r) = \ln V(r) + r\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right),$$

$$IC_{p2}(r) = \ln V(r) + r\left(\frac{N+T}{NT}\right)\ln\min\{N,T\},$$

where $V(r) = \frac{1}{NT}\sum_{t=1}^{T}(X_t - \hat{\Lambda}_r\hat{F}_t)'(X_t - \hat{\Lambda}_r\hat{F}_t), \hat{\Lambda} = V_r D_r^{1/2}$. Bai and Ng (2002) show that the value of r that minimizes $IC_{p1}(r)$ or $IC_{p1}(r)$ is a consistent estimator, for $N, T \to \infty$, of the number of common factors. In the second step, the estimated factors are used as predictors in (3). As shown in Bai and Ng (2006), we can treat $\hat{F}_t$ as observed regressors.

Since the factors are selected according to an information criterion that operates on the eigenstructure of X_t, then the method is unsupervised. The selection methodology assumes that the factors are ordered according to the size of the corresponding eigenvalue. However, there is no reason why a predictand should not depend on a higher order component or different predictand, such as output and inflation, should depend on the same factors.

3 Approaches to the supervision of the factors

Several proposals have been made for supervising the factors, so that the selected factors carry information that is useful for predicting the specific variable under consideration. In this section we sketch a brief survey of the literature, ignoring the shrinkage and model averaging approaches that are applied directly to the observable predictors, rather than the principal components. For an account of these approaches, see De Mol et al. (2008), Bai and Ng (2008) and Stock and Watson (2006).

In the supervised PC method proposed by Bair et al. (2006) a subset of predictors is first selected on the basis of their correlation with the response variable; more specifically all the predictors for which the estimated regression coefficients are larger than a threshold c are considered

$$\left|\sqrt{T}\frac{\sum_t X_{it} y_{t+h}^{(h)}}{\sum_t X_{it}^2}\right| > c, \quad i = 1, 2, \ldots, N,$$

and a principal component analysis is performed on the selected predictors to extract the factors to be used for prediction. The method clearly depends on the threshold c, which is estimated by cross-validation.

Bai and Ng (2008) construct supervised principal components, that they name *targeted predictors*, by pre-selecting a subset of predictors with predictive power for a specific predictand, and conducting a PCA on those. They explore hard thresholding rules constructed on the t-statistics of the regression of $y_{y+h}^{(h)}$ on a single predictor X_{it} (after controlling for a set of predetermined variables, such as the lags of the dependent variables), say t_i^*, selecting those variables for which $|t_i^*| > c$, c alternatively being equal to 1.28, 1.65 and 2.58. Their selection rule does not consider the issue of multiple testing. In page 306 they state however that application of Holm's procedure (see section 5) did not lead to different results. Other soft thresholding methods are considered, such as the lasso, least angle regression, and the elastic net are considered and compared. The paper concludes that targeting the predictors to the economic variable to be predicted, they consider inflation in particular, leads to a gain in forecasting accuracy.

Bai and Ng (2009) proposed componentwise and block-wise boosting algorithms for isolating the predictors in FAR models that are most helpful in predicting a variable of interest. The algorithms do not rely on the ordering of the variables (and in the componentwise case do not rely on the ordering of their lags). Starting from the null model including only a constant, the algorithms perform incremental forward stagewise fitting of the mean square prediction error, by a sequence of Newton-Raphson iterations that iteratively improve the fit. At each step, a single explanatory variable (e.g. a PC), or a block consisting of a regressor and its lags, is fitted by ordinary least squares regression, and selected according to the reduction in the residual sum of squares. The selected variable contributes to the current predictor with a coefficient which is shrunk towards zero by a fraction known as the learning rate. The algorithm is iterated until a stopping rule is found. Bai and Ng (2009) propose an information criterion for selecting the number of boosting iterations that takes into account the estimation error in the estimation of the factors, which is $O(N^{-1})$.

Inoue and Kilian (2008) present an application of bootstrap aggregation (bagging) of predictors of U.S. inflation using $N = 30$ variables. Among the predictors, they consider selecting by pretesting the PCs among the first K, where K ranges from 1 to 8. Several critical values for selection pretest are considered. Stock and Watson (2012b) and Kim and Swanson (2014) also consider averaging the FAR predictors obtained from independent bootstrap samples with factors selected according to the rule that their (robust) t-statistic must be larger than 1.96 in modulus.

Fuentes et al. (2014) propose the use of sparse partial least squares to select a small subset of supervised factors, extending to a dynamic setting the static methodology of Chun and Keleş (2010). The candidate factors arise from the spectral decomposition of the matrix $T^{-1}\tilde{X}'yy'\tilde{X}$, where y has generic element $y_{t+h}^{(h)}$ and X is a matrix with rows composed of the elements of X_t', augmented by the lags of the predictand. The loadings are shrunk towards zero by a LASSO type penalty, aiming at the extraction of sparse supervised components.

Finally, an important class of supervision methods is based on inverse regression. Let (y, X) denote the observable predictand and predictors and let $f(y, X)$ denote their joint density. DI forecasting starts from the factorization $f(y, X) = f(y|X)f(X)$, assuming a factor structure for X. Obviously, the factors are unsupervised, as only the marginal distribution $f(X)$ is considered. A different approach to the supervision of the factors deals with the factorization $f(y, X) =$

$f(X|y)f(y)$, using the first conditional density for obtaining a reduced set of predictors incorporating information concerning y, achieving a substantial dimensionality reduction. The reduced set is then used in the prediction of y, according to a maintained model for $f(y|X)$. One such methodology is sliced inverse regression (SIR, Li (1991)): the range of y is partitioned into slices, within which the centroids of the X's are computed; a singular value decomposition of the matrix of centroids is performed to obtain a few effective dimension-reduction directions. The method of principal fitted components (PFC) analysis, proposed by Cook (2007) and Cook and Forzani (2008), is based on inverse regression of X on y to obtain a dimension reduction that preserves the information that is relevant for predicting y; in Cook (2007) the conditional mean $E(X|Y = y)$ is estimated by projecting the X's on polinomial terms in y and a principal component analysis is conducted on the conditional mean estimates to obtain the PCFs.

4 Principal components regression and components selection as shrinkage methods

Our approach is a particular case of the generalised shrinkage model considered in Stock and Watson (2012b). In the sequel we will assume that the forecasting model does not contain lags of the predictand and that the DI predictor results exclusively from principal component regression. In particular, the data are generated as

$$y_{t+h}^{(h)} = x_t'\delta + \epsilon_{t+h},$$

where ϵ_{t+h} has mean zero and Var$(\epsilon_{t+h}) = \sigma^2$. We also assume that T observations are available for $y_{t+h}, t = 1, \dots, T$), and focus on the predictor

$$\hat{y}_{t+h|t}^{(h)} = \sum_{i=1}^{N} \psi_i \hat{\beta}_i \hat{F}_{it}, \tag{4}$$

where $\hat{F}_t = D^{-1/2}\Lambda' X_t$, denotes the $N \times 1$ vector containing the standardised PC scores $\hat{F}_{it}, i = 1, \dots, N$, such that $\frac{1}{T}\sum_t \hat{F}_t \hat{F}_t' = I$, ordered according to the eigenvalues of the matrix S. Moreover, $\hat{\beta}_i = \frac{1}{T}\sum_t F_{it} y_{t+h}$, is the least squares estimator of the regression coefficient of y on the i-th PC, and ψ_i is the indicator for the inclusion of the i-th PC. The decision whether to include it or not depends on the strength of the relationship with the predictand and will be discussed shortly.

As $\hat{F}_{it} = x_t' v_i / d_i$, where d_i^2 is the i-th eigenvalue of S and v_i is the corresponding eigenvector,

$$S = VD^2V', V = [v_1, \dots, v_i, \dots, v_N], D = \text{diag}(d_1, \dots, d_N),$$

the predictor (4) can be written $\hat{y}^{(h)}_{t+h|t} = x'_t\hat{\delta}$, for

$$\hat{\delta} = \sum_{i=1}^{N} \frac{v_i}{d_i}\hat{\beta}_i,$$

The lack of supervision of the ordering of the components, see cautionary note 2 in Hadi and Ling (1998), can be evidenced by a plot of d_i^2, the i-th eigenvalue, versus the increase of the regression residual sum of squares arising from the deletion of the i-th component, measured by $T\hat{\beta}_i^2 = T^{-1}\left(\sum_t \hat{F}_{it} y^{(h)}_{t+h}\right)^2$.

The mean square error (MSE) of the above predictor, treating the factors as observed variables, is

$$\text{MSE}(\hat{y}^{(h)}_{t+h|t}) = [\text{B}(\hat{y}^{(h)}_{t+h|t})]^2 + \text{Var}(\hat{y}^{(h)}_{t+h|t}),$$

where the bias and the variance are given respectively by

$$\text{B}(\hat{y}^{(h)}_{t+h|t}) = \sum_{i=1}^{N}(1-\psi_i)\hat{F}_{it} d_i v'_i\delta, \quad \text{Var}(\hat{y}^{(h)}_{t+h|t}) = \sigma^2\left(1 + \frac{1}{T}\sum_i \psi_i \hat{F}^2_{it}\right).$$

These simple expressions underlie the usual bias-variance trade-off: removing one factor from the set of predictors (i.e. setting $\psi_i = 0$) reduces the variance, but increases the bias. The bias term features the singular value d_i, which implies that the bias increase is potentially larger if cœteris paribus a component with high d_i is removed. The bias resulting from the omission of a particular PC further depends on $v'_i\delta$; this term depends on the population relationship between y and the x's and on the loadings of the i-th PC[1]. The main message, conveyed by the above expression for the MSE is that omitting a PC loading heavily on important variables ($(v'_i\delta)^2$ is large) will have more impact if the PC corresponds to a large eigenvalue. Note also that $\text{Var}(\hat{y}_{t+h|t})$ depends solely on σ^2 and the Mahalanobis squared distance of x_T from 0 in the x space, $x'_T\left(\sum_t X_t X'_t\right)^{-1} x_T = \sum_i \hat{F}^2_{it}$; recalling that $\hat{F}_{it} = x'_T v_i/d_i$, the variance will be inflated by the presence of components with small d_i and for which the inner product $x'_T v_i$ is large. The broad conclusion arising from this analysis is that discarding the first PCs is not in general a good idea, and the ordering of the components should be taken seriously. We see this simple result as a possible explanation for the failure of alternative shrinkage methods to outperform the DI approach documented in Stock and Watson (2012b).

In principle, we could determine the optimal set of indicators $\{\psi_i, i = 1, \ldots, N\}$, which minimises the above MSE of prediction, e.g. parameterising

$$\psi_i = \psi_i(\hat{\beta}_i; \gamma, c) = \frac{1}{1 + \exp[-\gamma(|\hat{\beta}_i| - c_i)]},$$

[1] Omitting a component loading on x variables with no effect on the predictand makes a zero contribution to the bias.

where e.g. $c_i = c/d_i$ for an unknown positive constant c, and thinking about replacing δ and σ by some estimate, perhaps iteratively. Stock and Watson (2012b) estimate $c_i = c$ by cross-validation and set $\gamma \to \infty$. Hwang and Nettleton (2003) propose a general approach to the problem. We do not pursue this here and consider strategies such that ψ_i is the indicator function that the p-value of the significance test for the i-th regression coefficient is below a given threshold.

Before exposing our methodology, it is perhaps useful to remark that PCR conducted using only the first r principal components, chosen according to an information criterion, poses $\psi_i = I(i \leq r)$. Another popular regularisation method, ridge regression, yields the predictor (4) where the shrinkage factor ψ_i vary with i: letting $\rho \geq 0$ denote the penalty parameter in the criterion $S(\rho, \delta) = \sum_t (y_{t+h} - x_t'\delta)^2 + \rho\delta'\delta$, then

$$\psi_i = \frac{d_i^2}{d_i^2 + \rho}.$$

See Hastie et al. (2009) for a general reference and discussion.

5 The selection of the common factors as a multiple testing problem

Consider the set of null hypotheses $H_{0i} : \beta_i = 0, i = 1, \ldots, N$, in the predictive regression model, $y_{t+h}^{(h)} = \sum_{i=1}^{N} \beta_i \hat{F}_{it} + \epsilon_t$, and let p_i denote the two-sided p-value, based on the t-statistic t_i associated to the i- th principal components regression coefficient,

$$t_i = \sqrt{T}\frac{\hat{\beta}_i}{\hat{\sigma}}. \tag{5}$$

An issue is posed by the estimation of the regression standard error σ by $\hat{\sigma}$ in the denominator. The usual estimator, the square root of

$$\hat{\sigma}^2 = \frac{RSS}{T - N}, \quad RSS = \sum_t y_{t+h}^{(h)2} - T\sum_i \hat{\beta}_i^2,$$

is either infeasible, if $N \geq T$, or severely downward biased due to the overfitting when N is large. We address the issue of estimating σ in subsection 5.1.

The testing strategy based on the t_i statistics is multivariate, i.e. it treats all the remaining variables as nuisance parameters, when testing for the significance of a particular effect. An alternative strategy, that avoids estimation of σ, is componentwise, being based on the marginal linear regression t_i^* statistics,

$$t_i^* = \sqrt{\frac{T(T-1)}{RSS_i}}\hat{\beta}_i, \quad RSS_i = \sum_t y_{t+h}^{(h)2} - T\hat{\beta}_i^2.$$

The two t-statistics are related by

$$t_i = t_i^* \sqrt{\frac{RSS_i/(T-1)}{RSS/(T-N)}},$$

i.e. the univariate t_i^* is multiplied by the root mean square ratio. The componentwise approach is fairly popular in genomics and in hyper-dimensional contexts such that $N > T$. It has been adopted by Bai and Ng (2008) for estimating the targeted predictors.

For T large, the null distribution of the statistic is $t_i \sim \mathrm{N}(0, 1)$, and the p-value is computed as $p_i = 2(1 - \Phi(|t_i|))$. Here, in fact, the alternative is two-sided, that is $H_{1i} : \beta_i \neq 0, i = 1, \ldots N$. Let us consider the re-ordering of the components according to the nonincreasing values of $|t_i|$, and let us denote by $\hat{F}_{(i),t}$ the (i)-th component in the new ordering. The corresponding ordered p-values will be denoted by

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(i)} \leq \cdots \leq p_{(K)}.$$

The inclusion of a the i-th PC is based on a multiple testing procedure, which provides a decision rule aiming at controlling overall error rates when performing simultaneous hypotheses tests. The procedures can be distinguished according to the error that is controlled. For this purpose, consider the following confusion matrix:

Actual	Decision Accept	 Reject	Total
$H_{0i} : \beta_i = 0$	TN	FP	N_0
$H_{1i} : \beta_i \neq 0$	FN	TP	N_1
Total	A	R	N

$R = FP + TP$ is the total number of hypotheses rejected; FP is the number of false rejections (Type I errors(, i.e. of falsely rejected hypothesis. There are N_0 true nulls and N_1 false nulls. TN in the number of correctly accepted true nulls, etc. FN is the number of false positive decisions, i.e. type II errors and TP (true positives) is the number of correct rejections.

The per comparison error rate (PCER) approach controls for the expected number of true H_{0i} rejected over N, the total number of tests,

$$\text{PCER} = \frac{\mathrm{E}(FP)}{N}.$$

It amounts to ignoring the multiplicity problem altogether, and uses the critical value corresponding to a preselected significance level α, and thus rejects all hypotheses for which $p_i < \alpha$. This guarantees that α is an upper bound for PCER, as when $N_0 = N$ (all null hypotheses are true) $\mathrm{E}(FP) = \alpha N$.

The problem with this approach is that the probability of a false rejection increases rapidly with N; in particular, if $N_0 = N$, $P(R \geq 1) = 1 - (1 - \alpha)^N$. This has led to developing multiple testing strategies requiring that the probability of one or more false rejections does not exceed a given level. Definining the family wise error rate (FWER) as the probability of rejecting any true H_{0i}, FWER $= P(FP \geq 1)$, the aim is defining decision rules guaranteeing FWER $\leq \alpha$. The simplest procedure controlling the FWER is the Bonferroni rule, rejecting all H_{0i} for which $p_i < \frac{\alpha}{N}$. However, the power of this rule is typically very low, when $N > N_0$ is large.

A more powerful method for controlling the FWER at level α is due to Holm (1979). Holm's method is a step-down procedure rejecting the (i)-th null hypothesis $H_{0,(i)}$ if

$$p_{(j)} \leq \frac{\alpha}{N-j+1}, j = 1, 2, \ldots, i.$$

Note that $p_{(1)} \leq \alpha/N$ and $p_{(N)} \leq \alpha$, so that at the initial step we apply Bonferroni's rule, while at the final step we get the PCER approach. Thus, if $p_{(1)} > \alpha/N$, all nulls are accepted and the procedure stops. Else, we reject $H_{0,(1)}$ and test the remaining hypotheses at level $\alpha/(N-1)$. In such case, we accept all $H_{0,(i)}, (i) > 1$ if $p_{(2)} > \alpha/(N-1)$, else we reject $H_{0,(2)}$ and test the remaining hypotheses at level $\alpha/(N-2)$, in which case we iterate the procedure, until all the remaining hypotheses are accepted.

Procedures that control for the FWER have unduly conservative when N is very large, despite the improvements offered by step-down procedures such as Holm's. A less conservative approach is offered by procedures that control for the false discovery rate (FDR), which is the expected proportion of falsely rejected nulls:

$$\text{FDR} = \text{E}\left(\frac{FP}{R}\right).$$

The main procedure is due to Benjamini and Hochberg (BH, Benjamini and Hochberg (1995)). Let α denote a control rate in the range (0,1). A decision rule that has FDR $= \alpha N_0/N \leq \alpha$ rejects all $H_{o,(i)}, i = 1, \ldots, r$, for which

$$p_{(i)} \leq \alpha\frac{i}{N}, i = 1, \ldots, r, \quad p_{(r+1)} > \alpha\frac{r+1}{N}.$$

In high dimensional settings BH has been proved to achieve a better balance between multiplicity control and power. Another advantage is that when $N = N_0$ BH controls also the FWER at level α. Adaptive variants and refinements are available in the literature that address the issue of correlation in the tests statistics. See Efron (2010) for a review.

As stated in in the introduction, if we can think that nature is benign and that the ordering of the factors carries important information, then the selection should incorporate information on the factor structure. This can be achieved by weighting the p-values, p_i, according to the index i of the factor. A procedure that allows for p-value weighting and achieves control over the FDR is due to Genovese et al. (2006), and it works as follows:

1. Assign weights $w_i > 0$ to each H_{0i} so that $\bar{w} = N^{-1}\sum_i w_i = 1$.

2. Compute $q_i = p_i/w_i$, $i = 1, \ldots, N$.

3. Apply the BH procedure at level α to q_i.

A natural choice for the weights in our context is setting $w_i = d_i^2$; as a matter of fact, if the predictors are standardised, then S is a correlation matrix with eigenvalues $0 \leq d_i^2 \leq N$, and $N^{-1} \sum_i d_i^2 = 1$.

5.1 Estimation of σ

Fan et al. (2012) propose a refitted cross-validatory (RCV) estimator of the error variance which is consistent when the number of factors grows at a faster rate than the number of observations. The procedure has two stages: the sample is split into two independent subsamples. In the first stage variable selection is carried out on the two subsamples by a consistent pretest selection procedure, based on the marginal t statistics $t_i^*, i = 1, \ldots, N$, yielding two sets of selected variables, denoted M_1 and M_2. In the second stage, the two models are estimated on the other subsample (i.e. M_1 is estimated on the second and M_2 on the first subsample), yielding two estimates, s_1^2 and s_2^2, of the error variance. The RCV estimator is the average of the two. Refitting aims at eliminating the influence of variable that have been spuriously selected in the first stage.

5.2 Controlling for the lags of the dependent variable

So far we have considered the regression of a predictand on the principal components. If lags of the predictand have to be incorporated, as in the FAR approach, the above variable selection procedures are applied to the regression of the residuals of the regression of $y_{t+h}^{(h)}$ on $\{y_t, y_{t-1}, \ldots, y_{t-p}\}$ on the principal components computed on the set $\{\tilde{X}_{it}, i = 1, \ldots, N\}$, where $\tilde{X}_{it}$ is the residual of the regression of X_{it} on $\{y_t, y_{t-1}, \ldots, y_{t-p}\}$.

6 Empirical Analysis

6.1 Data and methods

The dataset used in empirical analysis is derived from that employed by Stock and Watson (2012a), and consists of 211 U.S. macroeconomic time series, available at the quarterly observation frequency from 1959-I to 2011-II. Of the 211 series, 121 were considered for our empirical analysis[2]. The series are all transformed to induce stationarity by taking first or second differences, logarithms or first or second differences of logarithms. The series are grouped into 12 categories; for a complete list of variables and their transformation see the Appendix A.

We consider a transformation y_t of the original variable Y_t, depending on the order of integration. Real activity variables are typically integrated of order 1, denoted $Y_t \sim I(1)$; defining

[2]As in Stock and Watson (2012b), we exclude series at high aggregate level. We decided to exclude also series starting after 1959 or ending before 2011. As a result, our dataset can be considered as an update of that used in Stock and Watson (2012b), which ended in 2009:II.

$y_t = \Delta \ln Y_t$, the predictand is

$$y_{t+h}^{(h)} = 400 \times \frac{\Delta_h \ln Y_{t+h}}{h}, \quad \Delta_h \ln Y_{t+h} = \ln Y_{t+h} - \ln Y_t,$$

the h-period growth at an annual rate. For nominal price and wage series we assume $Y_t \sim I(2)$, and, in accordance to Stock and Watson (2002b), the series to be predicted is

$$y_{t+h}^{(h)} = 400 \times \left(\frac{\Delta_h \ln Y_{t+h}}{h} - \Delta_1 \ln Y_t \right).$$

The predictors are represented by the 121 standardized principal components obtained from the spectral decomposition of the covariance matrix of the original indicators. A heat map of the squared factor loadings for the 121 quarterly series is provided by figure 1. The vertical axis is the series categories reported in A; in the horizontal axis the order of the principal component. The plot evidences that the first factor loads principally on the growth rates of the the indicators of real activity; the second has rather sparse loadings on both real and nominal variables, whereas the third loads eminently on price and wage inflation rates.

The forecasts are obtained using a pseudo real-time forecasting procedure[3]. The data from 1960:II to 1984:IV are used as a training sample. A PCA on the N standardized indicators is conducted to extract the factors, represented by the standardized principal components. The PCA are selected and the estimated regression coefficients are used at time 1984:IV to forecast $\hat{y}_{T+h}^{(h)}$, $h = 1$, 2, and 4 quarters ahead. Then, the estimation sample is updated by one quarterly observations and downdated by removing the initial one, so that the second set of observations ranges from from 1960:III to 1985:I, and so forth. For each rolling window, consisting of $T = 100$ observations, a PCA is conducted to extract the N components, variable selection is performed and the predictor is formed. The process continues until the end of the sample is reached. In our case the last available data is 2010:II when $h = 4$. The experiment delivers around 100 forecasts for each forecasting method, that can be compared with the observed values.

The predictors that are compared are:

- The pure autoregressive (AR) predictor

$$\hat{y}_{t+h}^{(h)} = \hat{\phi}_1^{(h)} y_t + \cdots + \hat{\phi}_p^{(h)} y_{t-p},$$

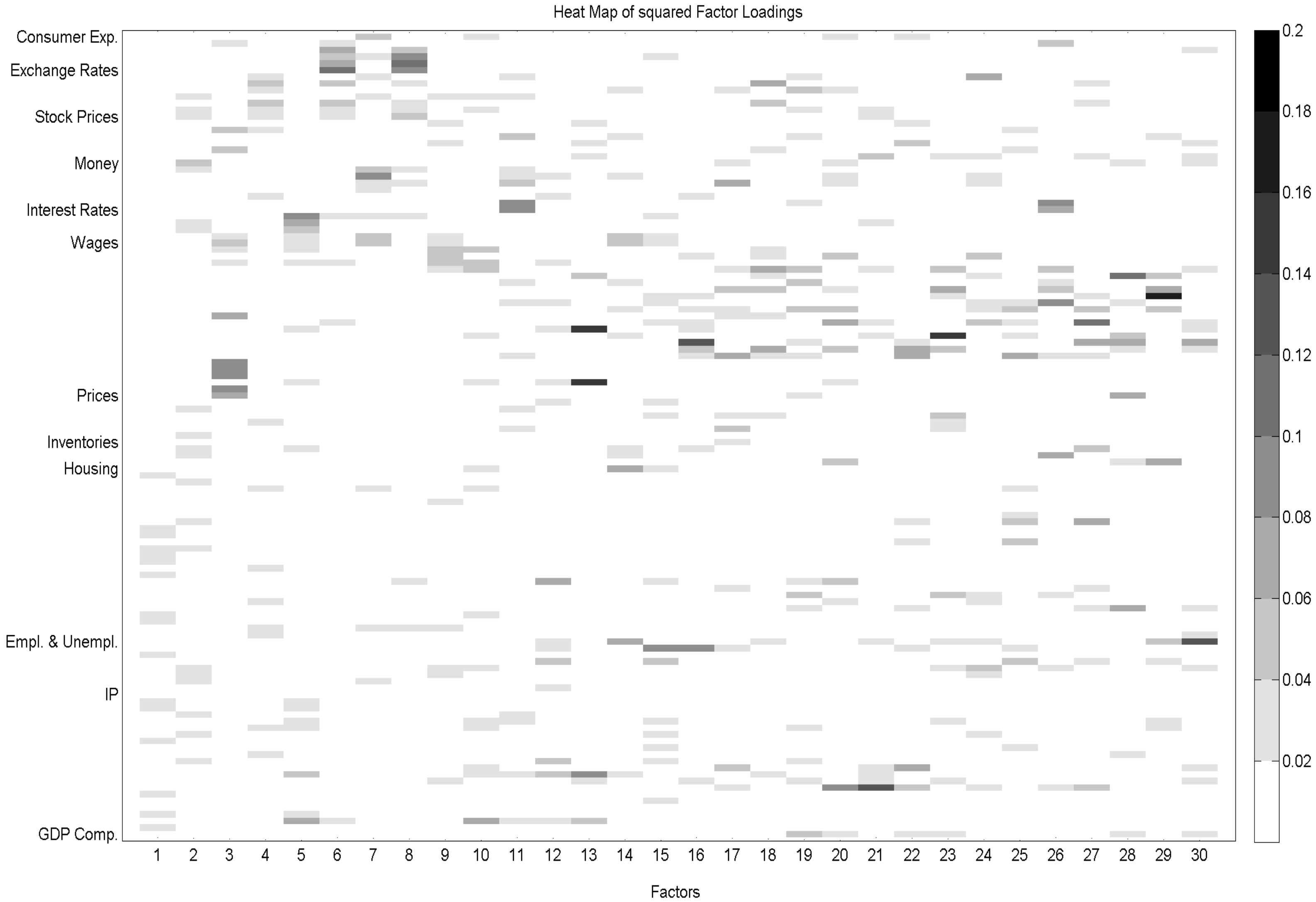
  where the order is selected according to the Schwarz Information criterion (SIC) and the AR coefficients are estimated by ordinary least squares.

- The dynamic factor model predictor (DFM5)

$$\hat{y}_{t+h}^{(h)} = \hat{\beta}_1^{(h)} \hat{F}_{1t} + \cdots + \hat{\beta}_r^{(h)} \hat{F}_{rt},$$

[3] As reported in D'Agostino and Giannone (2012), the exercise is pseudo real-time because we use the last vintage of data (for this dataset the last vintage is November 2011) and we do not consider the release available at the time of forecasting.

Figure 1: Heat map of the squared factor loadings for the 121 quarterly series as given in table 1. The vertical axis is the series categories as reported in A; the horizontal axis is the order of the principal component.



where $\hat{F}_{it}, i = 1, \dots, r$, are the first $r = 5$ principal components. This number is selected by the Bai and Ng (2002) criteria and coincides with the factor model benchmark proposed in Stock and Watson (2012b). When lags of the predictand are included, as in (4), the order p is that selected for the previous case (AR predictor).

- The supervised factor predictor

$$\hat{y}_{t+h}^{(h)} = \hat{\beta}_1^{(h)} \hat{F}_{(1)t} + \cdots + \hat{\beta}_r^{(h)} \hat{F}_{(r)t},$$

  with r factors, ranked according to their p-values, selected according to

  - Holm's multiple comparison procedure, controlling the FWER at the 5% level;
  - The Benjamini-Hochberg procedure, controlling the FDR at the 10% level[4];
  - The Genovese et al. (2006) procedure with p-values weighted according to the corresponding eigenvalue.

  If the lags of the predictand are considered, as in the FAR approach, then the factors are selected from the principal components computed on the residuals of the projection of the original predictors on the linear space spanned by the first p lags of the dependent variable.

  We consider two implementations of the variable selection procedure, the first based on the marginal t_i^* statistics and the second based on the t_i statistics computed using the Fan et al. (2012) estimator of the regression error variance.

The performance of the different methods is evaluated using the mean square forecast error (MSFE), defined as follows: let T_0 be the first point in time for out of sample evaluation and T_1 be the last point in time for which we compute the MSFE for $h = 1$, 2, and 4

$$MSFE = \frac{1}{T_1 - T_0} \sum_{\tau=T_0}^{T_1} \left(\hat{y}_{\tau+h}^{(h)} - y_{\tau+h} \right)^2.$$

The results are presented in terms of mean square error (MSFE) relative to the AR (SIC) benchmark

$$R_j = \frac{MSFE_j(h)}{MSFE_{AR}(h)}$$

where $j \in \{\text{DFM5}, \text{Holm}, \text{BH}, \text{GRW}\}$. A value below one indicates that the specified method is superior to the AR (SIC) forecast.

6.2 Empirical Results

The rolling forecast experiment was conducted for the following series:

[4]This is the target value most often considered in applications; see Efron (2010). Controlling the FDR at the 5% level leads to very similar results.

- Industrial Production Index (IPI)
- Total employment: Non farm Payroll (NPE)
- Unemployment Rate (UR)
- Housing Starts (HS)
- Consumer Price Index (CPI)
- Treasury Bill 10-years (TB)
- Real Personal Income (RPI)
- Gross National Product (GNP).

Tables 1 - 4 report the relative MSFEs for the five alternative forecasting models under consideration. In particular, Table 1 refers to the case when lags of the predictand are not considered for forecasting using DI and the selection of the PCs is based on the p-values computed on the marginal t_i^* statistics; Table 2 refers to the case when p lags are considered and the selection is based on the marginal t_i^* statistics. Tables 3 and 4 deal with the selection based on the t_i statistics with σ estimated according to the RCV method by Fan et al. (2012): in 3 no lags of the predictands were considered, whereas in 4 they were.

There are several conclusions that can be drawn from the empirical evidence summarised in the tables. The first broad consideration is that forecasting methods based on factor models provide accurate forecasts and improve over the AR benchmark in the majority of the cases across all horizons, when the lags of the dependent variable are not considered (which is the case considered in Tables 1 and 3).

The second general conclusion is that pre-selection of the components by the multiple testing procedures considered leads to several improvements in forecasting accuracy (when no lags of the predictand are considered). The three procedures show the best performances for 52% of the cases across horizons/variables combinations in tables 1 and 3, whereas DFM5 and AR(SIC) have the best performances only for 21% and 27% of the occurrences, respectively.

Thirdly, when the lags of the target variables are considered in the forecasting model, the predictors based on the factors, regardless of their selection, are more systematically outperformed by the benchmark AR predictor. The combined evidence of Tables 2 and 4 is that the AR predictor is ranked best in 58% of the cases. The last finding has already been reported and investigated in previous studies, among which we mention D'Agostino and Giannone (2012) and Stock and Watson (2002a). A possible explanation is that factor models have the ability to capture efficiently not only the information that is common to other cross-sectional variables, but also the specific dynamic features of each variable to predict. Also, after conditioning for the role of lagged values, the factors computed on the residual variation contribute more to the variability of the forecasts, leading to an increase in the MSFE.

The series for which the multiple testing procedures outperform the DFM5 predictor are Total employment: Non farm Payroll (NPE), Housing Starts (HS), Treasury Bill 10-years (TB), Real

Personal Income (RPI). For NPE, HS, TB and RPI they produce the minimum MSFEe across all horizons (panels A, B and C of table 1), with the exception of TB at horizon $h = 1$, for which the AR predictor ranks best. Finally, DFM5 is ranked best for UR, achieving a 20% reduction in the MSFE over the AR predictor and a 4% reduction over the multiple testing procedures, across all horizons. For the other variables the results are less sharp and depend basically on the forecast horizon. In Table 1 for IPX and GNP we observe a slight improvement of 4% of DFM5 and 15% of BH over AR only for $h = 1$, whereas for $h = 2$ and $h = 4$ both multiple testing procedures and DFM5 do not outperform the benchmark.

The choice of the reference test statistics (marginal t_i^* or multiple regression t_i, with RCV estimation of σ) does not seem to affect the results of the multiple testing procedures. The previous results are confirmed examining Table 2. We observe a further improvement only for GNP, where now the best performing predictor is GRW, achieving a 24% MSFE reduction with respect to the benchmark, also for $h = 2$ (Panel B) where the gain in forecasting accuracy amounts to about 10%.

Among the multiple testing procedures, weighting the p-values according to the eigenvalues does not lead to an improvement, with a few exceptions. Holm's sequential procedure clearly outperforms the other predictors in terms of MSFE in 27% of the cases when no lags of the dependent variable are in use, whereas BH and GRW are ranked best in 19% and 6% of the cases, respectively. This result seems to depend exclusively by the conservative nature of the Holm method, compared to the procedures controlling the FDR.

6.3 Assessment of real time performance

Following D'Agostino and Giannone (2012), we evaluate how the forecasting performance of the predictors evolved over time. In figure 2 we plot the time series pattern of the MSFEs of the DFM5 predictor (black solid line) and the predictor resulting from Holm's selection of the factors (blue dashed line), relative to the AR benchmark (red line). We consider only 3 series, namely NPE, TB and RPI, for which the Holm's selection provided sizable improvements. The relative MSFEs were smoothed over time with a centered moving window spanning 2 years. The shaded areas are the NBER recessions.

Interestingly, the factor based methods perform best during the great recession, and present no substantial gain during the great moderation. This empirical findings is consistent with the literature, as during the recession the comovements among economic variables are more prominent and thus the factors become more useful for forecasting. The selection of the factor leads to greater MSFE reductions in the last five years of the sample, including the great recession.

Further insight into the assessment of the performance of Holm's factor selection method can be gauged from the consideration of which factors are selected by the procedure. Figure 3 is a plot versus time of the index number of the selected factors arising as a by-product of the rolling forecast experiment. In the case of NPE (first row), the first principal components is always selected, and is the only relevant factors for forecasting one-step-ahead. The second and third factors enter the selection at horizons $h = 2$ and $h = 4$, with the second factor being switched off during the recession and the third propping up during the great recession. Hence, it may be concluded that

Table 1: Relative Mean Square Forecast Errors of 5 alternative predictors at horizons $h = 1, 2,$ and $h = 4$. The selection of the factors is based on the marginal t_i^* statistics and prediction occurs by principal component regression on the selected factors with no lags of the predictand.

	IPX	NPE	UR	HS	CPI	TB	RPI	GNP
				Panel A: Rolling, $h = 1$				
AR(SIC)	13.164	0.760	0.039	0.005	6.290	**0.172**	12.181	5.573
DFM5	**0.960**	0.910	**0.693**	0.943	**0.959**	1.207	0.898	0.942
Holm	1.158	**0.771**	0.728	0.914	1.014	1.079	**0.828**	0.900
BH	1.173	0.777	0.728	**0.896**	1.028	1.083	0.897	**0.848**
GRW	1.155	0.771	0.728	0.917	1.022	1.089	0.884	0.876
				Panel B: Rolling, $h = 2$				
AR(SIC)	**16.129**	1.103	0.041	0.003	2.068	0.121	5.816	**4.084**
DFM5	1.026	1.013	**0.680**	1.126	**0.873**	1.246	0.854	1.126
Holm	1.102	0.876	0.757	**0.995**	0.917	**0.985**	0.769	1.000
BH	1.102	**0.873**	0.754	1.041	0.949	1.064	0.774	1.000
GRW	1.102	0.876	0.759	0.995	0.917	0.985	**0.763**	1.000
				Panel C: Rolling, $h = 4$				
AR(SIC)	**16.184**	1.694	0.049	0.002	0.553	0.080	3.635	**3.595**
DFM5	1.124	1.072	**0.713**	1.084	0.869	1.198	0.801	1.112
Holm	1.047	**0.879**	0.759	**0.979**	0.928	0.974	**0.681**	1.008
BH	1.025	0.916	0.735	0.990	**0.859**	**0.963**	0.684	1.028
GRW	1.039	0.917	0.766	0.979	0.954	0.974	0.682	1.005

NOTE: Numerical entries are mean square forecast errors (MSFEs). Forecasts are quarterly, for the period 1985:IV - 2010:II for a total of 103 out of sample forecasts. Entries in the first row, corresponding to the AR(SIC) benchmark model, are actual MSFEs, while all other entries are relative MSFEs, such that an entry below one indicates that the specified method is superior to the AR(BIC) forecast.

Table 2: Relative Mean Square Forecast Errors of 5 alternative predictors at horizons $h = 1, 2$, and $h = 4$. The selection of the factors is based on the marginal t_i^* statistics and prediction occurs by the principal component regression on the selected factors including p lags of the predictand.

	IPX	NPE	UR	HS	CPI	TB	RPI	GNP
				Panel A: Rolling, $h = 1$				
AR(SIC)	13.164	0.760	0.039	0.005	**6.290**	**0.172**	12.181	5.573
DFM5	1.195	1.119	0.954	0.939	1.036	1.189	0.996	0.968
Holm	1.005	**0.968**	**0.912**	0.894	1.090	1.123	0.969	0.846
BH	**0.987**	0.968	0.912	**0.873**	1.093	1.115	**0.937**	**0.836**
GRW	1.004	0.971	0.912	0.880	1.075	1.095	0.971	0.847
				Panel B: Rolling, $h = 2$				
AR(SIC)	**16.129**	**1.103**	**0.041**	0.003	2.068	**0.121**	5.816	**4.084**
DFM5	1.230	1.215	1.158	0.984	**0.987**	1.046	0.875	1.067
Holm	1.166	1.204	1.064	**0.923**	1.035	1.033	0.851	1.021
BH	1.161	1.204	1.064	0.932	1.048	1.031	0.871	1.023
GRW	1.168	1.205	1.064	0.939	1.029	1.024	**0.846**	1.020
				Panel C: Rolling, $h = 4$				
AR(SIC)	**16.184**	**1.694**	**0.049**	**0.002**	0.553	**0.080**	3.635	**3.595**
DFM5	1.084	1.194	1.103	1.033	**0.935**	1.055	0.962	1.101
Holm	1.063	1.189	1.094	1.007	0.970	1.048	**0.912**	1.037
BH	1.064	1.185	1.093	1.009	0.959	1.054	0.928	1.040
GRW	1.077	1.189	1.094	1.009	0.972	1.051	0.919	1.037

Table 3: Relative Mean Square Forecast Errors of 5 alternative predictors at horizons $h = 1, 2$, and $h = 4$. The selection of the factors is based on the t_i statistics with RCV estimation of σ and prediction occurs by principal component regression on the selected factors and no lags of the predictand.

	IPX	NPE	UR	HS	CPI	TB	RPI	GNP
Panel A: Rolling, $h = 1$								
AR(SIC)	13.164	0.760	0.039	0.005	6.290	**0.172**	12.181	5.573
DFM5	**0.960**	0.910	**0.693**	0.943	**0.959**	1.207	0.898	0.942
Holm	1.185	0.744	0.721	0.901	1.040	1.114	**0.828**	0.816
BH	1.090	0.737	0.749	0.892	1.036	1.181	0.966	0.804
GRW	1.120	**0.728**	0.750	**0.833**	1.051	1.206	0.988	**0.786**
Panel B: Rolling, $h = 2$								
AR(SIC)	**16.129**	1.103	0.041	**0.003**	2.068	**0.121**	5.816	4.084
DFM5	1.026	1.013	**0.680**	1.126	**0.873**	1.246	0.854	1.126
Holm	1.123	**0.889**	0.734	1.065	0.917	1.098	**0.769**	**0.919**
BH	1.126	0.930	0.778	1.023	0.985	1.178	0.813	0.926
GRW	1.082	0.889	0.796	1.010	0.989	1.149	0.785	0.926
Panel C: Rolling, $h = 4$								
AR(SIC)	16.184	1.694	0.049	0.002	0.553	0.080	3.635	**3.595**
DFM5	1.124	1.072	0.713	1.084	**0.869**	1.198	0.801	1.112
Holm	**0.960**	1.032	**0.675**	0.988	0.896	**0.975**	0.686	1.145
BH	0.971	1.003	0.694	**0.963**	0.882	1.026	0.681	1.066
GRW	0.974	**0.993**	0.702	0.963	0.885	1.047	**0.678**	1.066

Table 4: Relative Mean Square Forecast Errors of 5 alternative predictors at horizons $h = 1, 2$, and $h = 4$. The selection of the factors is based on the t_i statistics with RCV estimation of σ and prediction occurs by principal component regression on the selected factors including p lags of the predictand.

	IPX	NPE	UR	HS	CPI	TB and	RPI	GNP
Panel A: Rolling, $h = 1$								
AR(SIC)	**13.164**	**0.760**	0.039	0.005	**6.290**	**0.172**	12.181	5.573
DFM5	1.195	1.119	0.954	0.939	1.036	1.189	0.996	0.968
Holm	1.149	0.957	0.921	0.944	1.099	1.192	**0.961**	0.922
BH	1.173	0.963	**0.861**	**0.924**	1.063	1.338	1.018	**0.821**
GRW	1.169	0.967	0.876	0.942	1.070	1.301	1.075	0.821
Panel B: Rolling, $h = 2$								
AR(SIC)	**16.129**	**1.103**	**0.041**	0.003	2.068	**0.121**	5.816	**4.084**
DFM5	1.230	1.215	1.158	0.984	**0.987**	1.046	0.875	1.067
Holm	1.206	1.184	1.062	**0.951**	1.045	1.059	0.874	1.013
BH	1.236	1.170	1.068	1.012	1.018	1.074	**0.863**	1.019
GRW	1.232	1.172	1.070	0.965	1.021	1.075	0.872	1.019
Panel C: Rolling, $h = 4$								
AR(SIC)	**16.184**	**1.694**	**0.049**	**0.002**	0.553	**0.080**	3.635	**3.595**
DFM5	1.084	1.194	1.103	1.033	**0.935**	1.055	0.962	1.101
Holm	1.070	1.178	1.091	1.000	0.959	1.034	**0.915**	1.054
BH	1.073	1.175	1.094	1.014	0.979	1.029	0.942	1.052
GRW	1.073	1.174	1.096	1.016	0.994	1.028	0.933	1.054

nature is benign in this case as the information that is essential for forecasting is well represented in the first three factors.

Nature is less benign in the TB case (second row panels). No factors is selected at the beginning and most noticeably at the end of the sample period. High order components are selected and the intermediate ones receive zero weight. Nature is even more bizarre for the RPI variable (bottom panels); the number of selected factors never exceeds the four, but order of selected factors is surprising. For predicting one-step-ahead the Holm's procedure, more or less regularly, selects the 16-th, 18-th and 24-th factors. This selection could never be contemplated in a classic factors model.

7 Conclusions

The paper has proposed a method for supervising the diffusion index methodology, originally proposed by Stock and Watson (2002a), which is based on the simple idea of selecting the relevant factors using a multiple testing procedure, achieving control over either the family wise error rate or the false discovery rate. Prior information about the order of the components may be introduced by weighting the p-values of the test statistics for variable exclusion with weights proportional to the eigenvalues.

Can we conclude that nature is tricky, but essentially benign? The answer is a qualified yes. The information that is needed for forecasting the eight macroeconomic variables considered in the paper is effectively condensed by the first few factors. However, variable selection, leading to exclude some of the low order principal components, can lead to a sizable improvement in forecasting in specific cases. Only in one instance, real personal income, we were able to detect a significant contribution from high order components.

Acknowledgements

This paper was presented at the 16th Annual Advances in Econometrics Conference, Aarhus, 15-16 November 2014. The authors gratefully acknowledge financial support by the Italian Ministry of Education, University and Research (MIUR), PRIN Research Project 2010-2011 - prot. 2010J3LZEN, Forecasting economic and financial time series. Tommaso Proietti gratefully acknowledges support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

Figure 2: Time varying performance of forecasting methods.

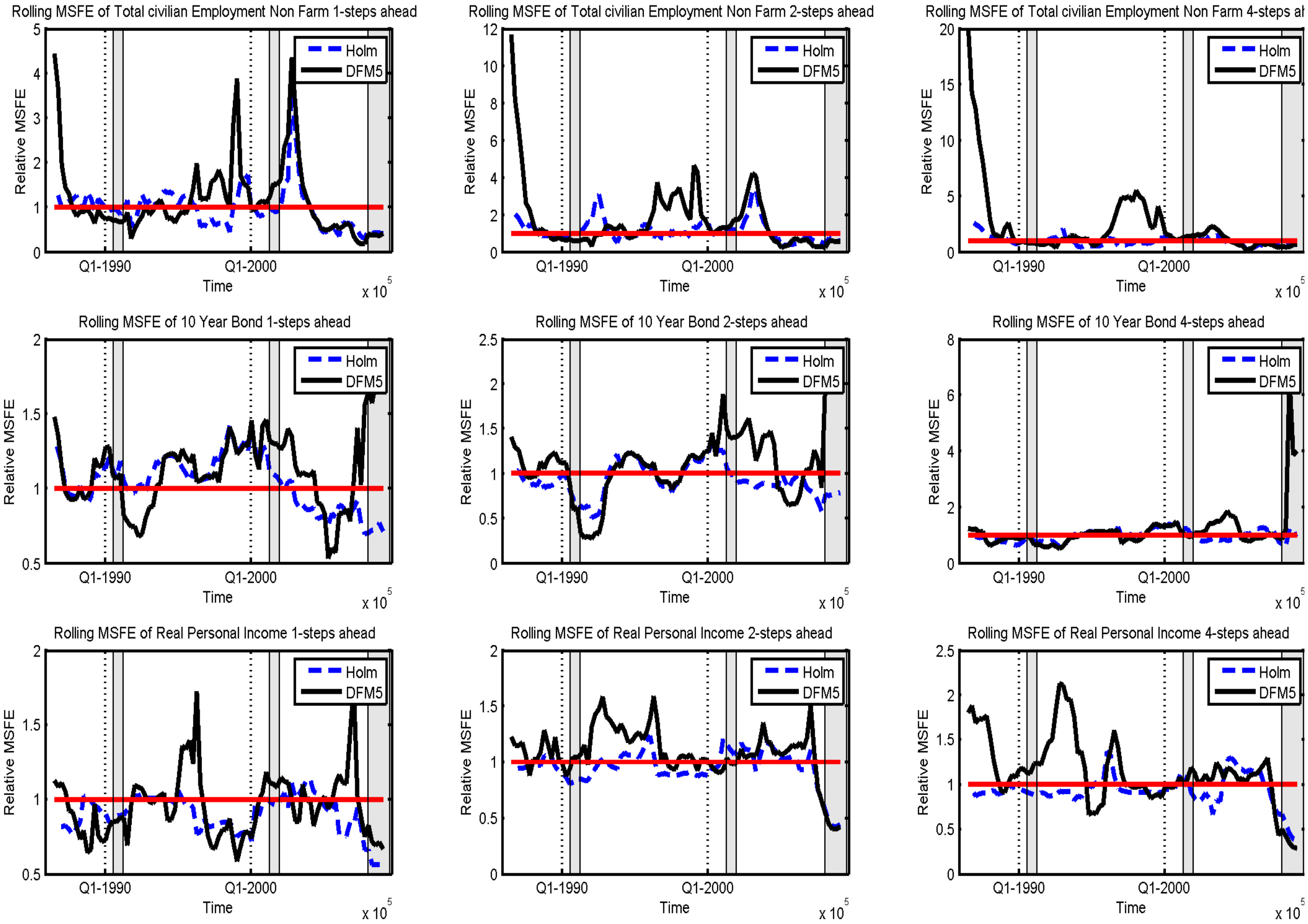
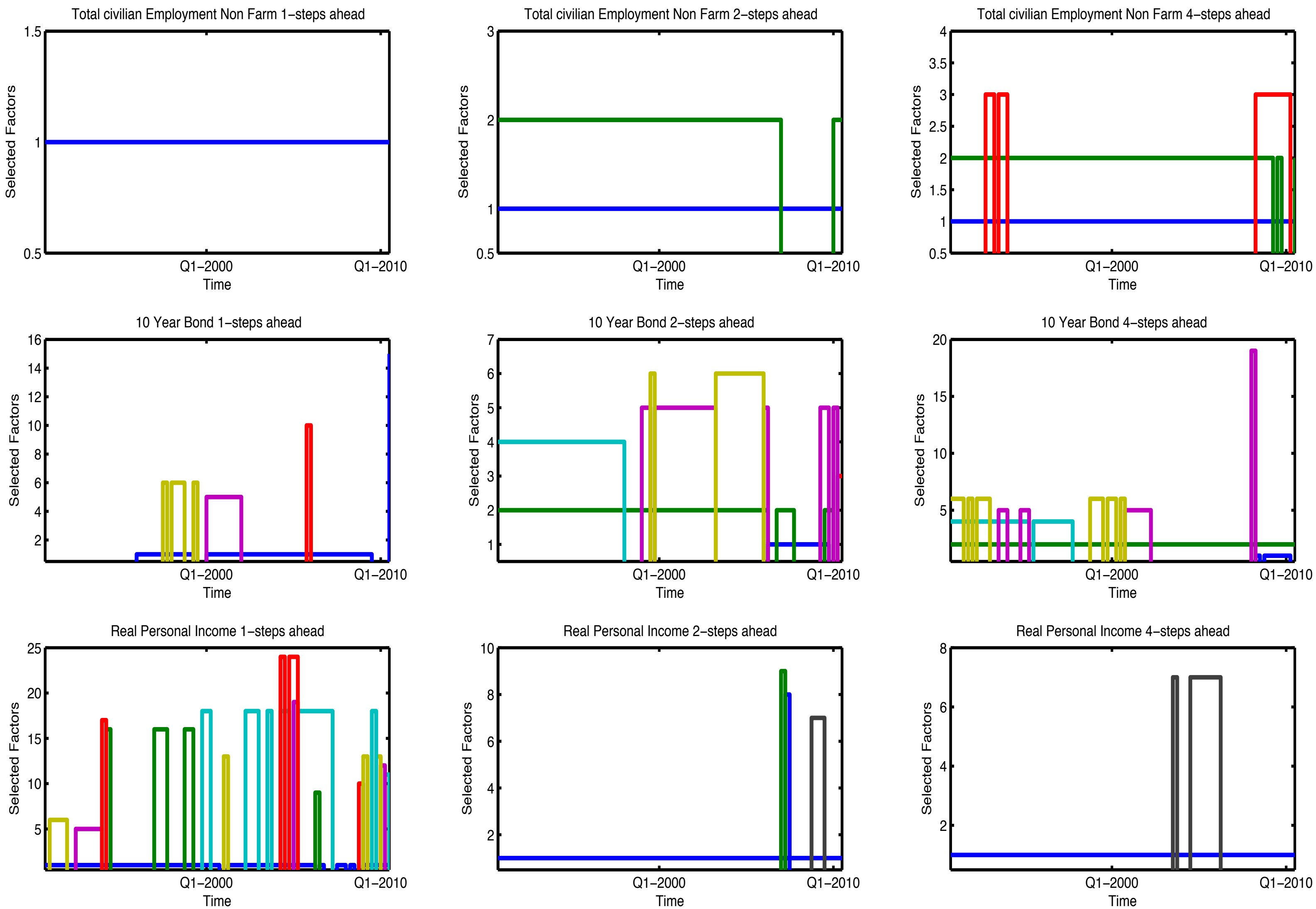


Figure 3: Selected factors using the Holm procedure.



References

Bai, J. and Ng, S. (2002). "Determining the Number of Factors in Approximate Factor Models." *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-Augmented Regressions." *Econometrica*, 74(4):1133–1150.

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. (2009). "Boosting Diffusion Indices." *Journal of Applied Econometrics*, 24(4):607–629.

Bair, E., Hastie, T., Debashis, P., and Tibshirani, R. (2006). "Prediction by Supervised Principal Components." *Journal of the American Statistical Association*, 101(473):119–137.

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Breitung, J. and Eickmeier, S. (2006). "Dynamic Factor Models." In *Modern Econometric Analysis*, 25–40. Springer.

Chun, H. and Keleş, S. (2010). "Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25.

Cook, R. D. (2007). "Fisher lecture: Dimension Reduction in Regression." *Statistical Science*, 1–26.

Cook, R. D. and Forzani, L. (2008). "Principal Fitted Components for Dimension Reduction in Regression." *Statistical Science*, 52:485–501.

Cox, D. (1968). "Notes on Some Aspects of Regression Analysis." *Journal of the Royal Statistical Society, Series A*, 131(3):265–279.

D'Agostino, A. and Giannone, D. (2012). "Comparing Alternative Predictors Based on Large-Panel Factor Models." *Oxford Bulletin of Economics and Statistics*, 74(2):306–326.

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.

Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Fan, J., Guo, S., and Hao, N. (2012). "Variance Estimation Using Refitted Cross-Validation in Ultrahigh Dimensional Regression." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):37–65.

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.

Fuentes, J., Poncela, P., and Rodríguez, J. (2014). "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting." *Journal of Applied Econometrics*.

Genovese, C. R., Roeder, K., and Wasserman, L. (2006). "False Discovery Control with p-value Weighting." *Biometrika*, 93(3):509–524.

Hadi, A. S. and Ling, R. F. (1998). "Some Cautionary Notes on the Use of Principal Components Regression." *The American Statistician*, 52(1):15–19.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer.

Holm, S. (1979). "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian journal of statistics*, 65–70.

Hwang, J. G. and Nettleton, D. (2003). "Principal Components Regression with Data Chosen Components and Related Methods." *Technometrics*, 45(1).

Inoue, A. and Kilian, L. (2008). "How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation." *Journal of the American Statistical Association*, 103:511–522.

Joliffe, I. T. (1982). "A Note on the Use of Principal Components in Regression." *Applied Statistics*, 31(3):300–303.

Kim, H. H. and Swanson, N. R. (2014). "Forecasting Financial and Macroeconomic Variables Using Data Reduction Methods: New Empirical Evidence." *Journal of Econometrics*, 178(2):352–367.

Li, K.-C. (1991). "Sliced Inverse Regression for Dimension Reduction." *Journal of the American Statistical Association*, 86:316–327.

Mosteller, F. and Tukey, J. W. (1977). "Data Analysis and Regression: a Second Course in Statistics." *Addison-Wesley Series in Behavioral Science: Quantitative Methods*.

Ng, S. (2013). "Variable Selection in Predictive Regressions." In Elliott, G. and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, volume 2, Part B, chapter 14, 752 – 789. Elsevier.

Onatski, A. (2010). "Determining the Number of Factors from Empirical Distribution of Eigenvalues." *The Review of Economics and Statistics*, 92(4):1004–1016.

Stock, J. H. and Watson, M. W. (2002a). "Forecasting Using Principal Components From a Large Number of Predictors." *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2002b). "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business & Economic Statistics*, 20(2):147–62.

Stock, J. H. and Watson, M. W. (2006). "Forecasting with Many Predictors." In Elliott, G., Granger, C., and Timmermann, A. (eds.), *Handbook of Economic Forecasting*, volume 1, chapter 10, 515–554. Elsevier.

Stock, J. H. and Watson, M. W. (2010). "Dynamic Factor Models." In Clements, M. P. and Hendry, D. F. (eds.), *Oxford Handbook of Economic Forecasting*, volume 1, chapter 2. Oxford University Press, USA.

Stock, J. H. and Watson, M. W. (2012a). "Disentangling the Channels of the 2007—09 Recession." *Brookings Papers on Economic Activity: Spring 2012*, 81.

Stock, J. H. and Watson, M. W. (2012b). "Generalized shrinkage methods for forecasting using many predictors." *Journal of Business & Economic Statistics*, 30(4):481–493.

A List of the time series used in the empirical illustration

This Appendix reports the time series in the dataset used in the application, the transformation type, the observations frequency (M= monthly and Q = quarterly) and the group to which they belong.

Letting Z_t denote the raw series, the following transformations are adopted:

$$X_t = \begin{cases} Z_t & \text{if Tcode=1} \\ \Delta Z_t & \text{if Tcode=2} \\ \Delta^2 Z_t & \text{if Tcode=3} \\ \ln(Z_t) & \text{if Tcode=4} \\ \Delta \ln Z_t & \text{if Tcode=5} \\ \Delta^2 \ln Z_t & \text{if Tcode=6} \end{cases}$$

Table 5: List of the predictors.

N	Short Description	Long Description	Tcode	Frequency	Category
		NIPA			
1	Disp-Income	Real Disposable Personal Income	5	Q	1
2	FixedInv	Real Private Fixed Investment	5	Q	1
3	Gov.Spending	Real Government Consumption Expenditures & Gross Investment	5	Q	1
4	GDP	Real Gross Domestic Product	5	Q	1
5	Investment	Real Gross Private Domestic Investment	5	Q	1
6	Consumption	Real Personal Consumption Expenditures	5	Q	1
7	Inv:Equip&Software	Real Nonresidential Investment: Equipment & Software	5	Q	1
8	Exports	Real Exports of Goods & Services	5	Q	1
9	Gov Receipts	Government Current Receipts (Nominal)	5	Q	1
10	Gov:Fed	Real Federal Consumption Expenditures & Gross Investment	5	Q	1
11	Imports	Real Imports of Goods & Services	5	Q	1
12	Cons:Dur	Real Personal Consumption Expenditures: Durable Goods	5	Q	1
13	Cons:Svc	Real Personal Consumption Expenditures: Services	5	Q	1
14	Cons:NonDur	Real Personal Consumption Expenditures: Nondurable Goods	5	Q	1
15	FixInv:NonRes	Real Private Nonresidential Fixed Investment	5	Q	1
16	FixedInv:Res	Real Private Residential Fixed Investment	5	Q	1
17	Gov:State&Local	Real State & Local Cons. Exp. & Gross Investment	5	Q	1
18	Inv:Inventories	Real Change in Private Inventories	5	Q	1
19	Inv:Inventories	Ch. Inv/GDP	1	Q	1
20	Output:Bus	Business Sector: Output	5	Q	1
21	Ouput:NFB	Nonfarm Business Sector: Output	5	Q	1
		Industrial Production			
22	IP: Dur gds materials	Industrial Production: Durable Materials	5	M	2
23	IP: Nondur gds materials	Industrial Production: nondurable Materials	5	M	2
24	Capu Man.	Capu Man. (Fred post 1972, Older series before 1972)	1	M	2
25	IP: Dur Cons. Goods	Industrial Production: Durable Consumer Goods	5	M	2
26	IP: Auto	IP: Automotive products	5	M	2
27	IP:NonDur Cons God	Industrial Production: Nondurable Consumer Goods	5	M	2
28	IP: Bus Equip	Industrial Production: Business Equipment	5	M	2
29	IP: Energy Prds	IP: Consumer Energy Products	5	M	2
		Employment and Unemployment			
30	Emp: Gov(Fed)	Federal	5	M	3
31	Emp: Gov (State)	State government	5	M	3
32	Emp: Gov (Local)	Local government	5	M	3
33	Emp: DurGoods	All Employees: Durable Goods Manufacturing	5	M	3
34	Emp: Const	All Employees: Construction	5	M	3
35	Emp: Edu&Health	All Employees: Education & Health Services	5	M	3
36	Emp: Finance	All Employees: Financial Activities	5	M	3
37	Emp: Infor	All Employees: Information Services	5	M	3
38	Emp:Leisure	All Employees: Leisure & Hospitality	5	M	3
39	Emp: Mining/NatRes	All Employees: Natural Resources & Mining	5	M	3
40	Emp: Bus Serv	All Employees: Professional & Bus. Services	5	M	3
41	Emp:OtherSvcs	All Employees: Other Services	5	M	3
42	Emp:Trade&Trans	All Employees: Trade, Transp. & Utilities	5	M	3
43	Emp:Retail	All Employees: Retail Trade	5	M	3
44	Emp:Wholesal	All Employees: Wholesale Trade	5	M	3

- Continued on next page -

Table 5 – continued from previous page

45	Urate: Age16-19	Unemployment Rate - 16-19 yrs	2	M	3
46	Urate:Age>20 Men	Unemployment Rate - 20 yrs. & over, Men	2	M	3
47	Urate: Age>20 Women	Unemployment Rate - 20 yrs. & over, Women	2	M	3
48	U: Dur<5wks	Number Unemployed for Less than 5 Weeks	5	M	3
49	U:Dur5-14wks	Number Unemployed for 5-14 Weeks	5	M	3
50	U:dur>15-26wks	Civilians Unemployed for 15-26 Weeks	5	M	3
51	U: Dur>27wks	Number Unemployed for 27 Weeks & over	5	M	3
52	Emp:SlackWk	Employment Level - Part-Time, All Industries	5	M	3
53	AWH Man	Average Weekly Hours: Mfg	1	M	3
54	AWH Overtime	Average Weekly Hours: Overtime: Mfg	2	M	3
55	Emp:nfb	Nonfarm Business Sector: Employment	5	Q	3
		Housing Starts			
56	Hstarts:MW	Housing Starts in Midwest Census Region	5	M	4
57	Hstarts:NE	Housing Starts in Northeast Census Region	5	M	4
58	Hstarts:S	Housing Starts in South Census Region	5	M	4
59	Hstarts:W	Housing Starts in West Census Region	5	M	4
		Inventories, Orders and Sales			
60	Orders (DurMfg)	Mfrs' new orders durable goods industries (bil. chain 2000 $)	5	M	5
61	Orders(Cons. Goods.	Mfrs' new orders, consumer goods and materials (mil. 1982 $)	5	M	5
62	UnfOrders(DurGds)	Mfrs' unfilled orders durable goods indus. (bil. chain 2000 $)	5	M	5
63	VendPerf	Index of supplier deliveries – vendor performance (pct.)	1	M	5
64	Orders(NonDefCap)	Mfrs' new orders, nondefense capital goods (mil. 1982 $)	5	M	5
65	MT Invent	Manufacturing and trade inventories (bil. Chain 2005 $)	5	M	5
66	Ret. Sale	Sales of retail stores (mil. Chain 2000 $)	5	M	5
		Prices			
67	Price:Oil	PPI: Crude Petroleum	5	M	6
68	PPI:FinGds	Producer Price Index: Finished Goods	6	M	6
69	PPI:FinConsGds(Food)	Producer Price Index: Finished Consumer Foods	6	M	6
70	PPI:FinConsGds	Producer Price Index: Finished Consumer Goods	6	M	6
71	PPI:IndCom	Producer Price Index: Industrial Commodities	6	M	6
72	PPI:IntMat	Producer Price Index: Interm. Materials: Supplies & Comp.	6	M	6
73	PCED_MotorVec	Motor vehicles and parts	6	Q	6
74	PCED_DurHousehold	Furnishings and durable household equipment	6	Q	6
75	PCED_Recreation	Recreational goods and vehicles	6	Q	6
76	PCED_OthDurGds	Other durable goods	6	Q	6
77	PCED_Food_Bev	Food and beverages purchased for off-premises cons.	6	Q	6
78	PCED_Clothing	Clothing and footwear	6	Q	6
79	PCED_Gas_Enrgy	Gasoline and other energy goods	6	Q	6
80	PCED_OthNDurGds	Other nondurable goods	6	Q	6
81	PCED_Housing-Utilities	Housing and utilities	6	Q	6
82	PCED_HealthCare	Health care	6	Q	6
83	PCED_TransSvg	Transportation services	6	Q	6
84	PCED_RecServices	Recreation services	6	Q	6
85	PCED_FoodServ_Acc.	Food services and accommodations	6	Q	6
86	PCED_FIRE	Financial services and insurance	6	Q	6
87	GDP Defl	Gross Domestic Product: Chain-type Price Index	6	Q	6
88	GPDI Defl	Gross Private Domestic Investment: Chain-type Price Index	6	Q	6
89	BusSec Defl	Business Sector: Implicit Price Deflator	6	Q	6
		Earnings and Productivity			
90	CPH:NFB	Nonfarm Business Sector: Real Compensation Per Hour	5	Q	7
91	CPH:Bus	Business Sector: Real Compensation Per Hour	5	Q	7
92	OPH:nfb	Nonfarm Business Sector: Output Per Hour of All Persons	5	Q	7
93	ULC:NFB	Nonfarm Business Sector: Unit Labor Cost	5	Q	7
94	UNLPay:nfb	Nonfarm Business Sector: Unit Nonlabor Payments	5	Q	7
		Interest Rates			
95	FedFunds	Effective Federal Funds Rate	2	M	8
96	TB-3Mth	3-Month Treasury Bill: Sec. Market Rate	2	M	8
97	AAA_GS10	AAA-GS10 Spread	1	M	8
98	BAA_GS10	BAA-GS10 Spread	1	M	8
99	tb6m_tb3m	tb6m-tb3m	1	M	8
100	GS1_tb3m	GS1_Tb3m	1	M	8
101	GS10_tb3m	GS10_Tb3m	1	M	8
		Money and Credit			
102	C&Lloand	Commercial and Ind. Loans at All Comm. Banks	5	M	9
103	ConsLoans	Consumer Loans at All Comm. Banks	5	M	9
104	NonBorRes	Non-Borr. Reserves of Dep. Inst. Auction Credit	5	M	9
105	NonRevCredit	Total Nonrevolving Credit Outstanding	5	M	9
106	LoansRealEst	Real Estate Loans at All Comm. Banks	5	M	9
107	TotRes	Total Reserves, Adj. for Chgs in Reserve Reqs.	5	M	9
108	ConsuCred	Total Consumer Credit Outstanding	5	M	9

- Continued on next page -

Table 5 – continued from previous page

		Stock Prices, Wealth and Household Balance Sheet			
109	S&P 500	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE	5	M	10
110	DJIA	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE	5	M	10
111	HHW:W	Total Net Worth	5	Q	10
112	HHW:TA_RE	TTABSHNO-REANSHNO	5	Q	10
113	HHW:RE	Real Estate - Assets - Households and Nonprofit Orgs	5	Q	10
114	HHW:Fin	Total Financial Assets - Assets - Households and Non Profits	5	Q	10
115	HHW:Liab	Total Liabilities - Households and Nonprofits	5	Q	10
		Echange Rates			
116	Ex rate: major	FRB Nominal Major Currencies Dollar Index	5	M	11
117	Ex rate: Switz	FOREIGN EXCHANGE RATE: SWITZERLAND	5	M	11
118	Ex rate: Japan	FOREIGN EXCHANGE RATE: JAPAN	5	M	11
119	Ex rate: UK	FOREIGN EXCHANGE RATE: UNITED KINGDOM	5	M	11
120	EX rate: Canada	FOREIGN EXCHANGE RATE: CANADA	5	M	11
		Other			
121	Cons. Expectations	Consumer expectations NSA	1	M	12