Forecasting Stock Returns with Large Dimensional Factor Models

A. Giovannelli — Ministry of Environment, Rome, Italy

D. Massacci — Bank of England

S. Soccorsi

Department of Economics, Lancaster University Management School

December 12, 2017

Abstract

Motivated by the longstanding evidence that economic variables can be decomposed into common and idiosyncratic components, we study equity premium out-of-sample predictability extracting the information contained in a high number of predictors via large dimensional factor models. We compare the widepsread factor model with a static representation of the common components with a more general dynamic model known as the Generalized Dynamic Factor model. Using statistical and economic evaluation criteria, we show that the Generalized Dynamic Factor model helps predicting the equity premium. Furthermore, exploiting the well-known link between the business cycle and return predictability, we find more accurate predictions by combining rolling and recursive forecasts in real-time.

JEL classification: C38, C53, C55, G11, G17.

Keywords: Stock Returns Forecasting, Factor Model, Large Data Sets, Forecast Evaluation.

[0]The views in this paper are those of the authors and do not necessarily reflect those of the Bank of England or its policy committees. This paper benefits from comments from participants at the CORE@50 Conference (Louvain-la-Neuve, Belgium), the 36th International Symposium on Forecasting (Santander, Spain) and the Macroeconomics and Financial Time Series Analyis workshop (Lancaster, UK). Errors and omissions are the authors' responsibility.

E-mail addresses: alessandro.giovannelli@uniroma2.it (Alessandro Giovannelli); dm355@cantab.net (Daniele Massacci); s.soccorsi@lancaster.ac.uk (Stefano Soccorsi, corresponding author).

1 Introduction

As reflected in a voluminous literature, equity premium predictability plays a key role in several areas of finance such as asset pricing, portfolio allocation and evaluation of investment managers performance. (see Rapach and Zhou (2013) for a recent review). However, this is challenging task: under Lo (2004) adaptive markets hypothesis, equity premium predictability is short-lived due to traders' searches for forecasting patterns (Timmermann and Granger, 2004; Timmermann, 2008). Early contributions conclude that predictability is either confined to specific periods (Pesaran and Timmermann, 1995) or completely absent (Bossaerts and Hillion, 1999; Goyal and Welch, 2003; Welch and Goyal, 2008). Recently there is mounting evidence that returns are predictable by macroeconomic and financial variables (Campbell and Thompson, 2008; Rapach *et al.*, 2010; Ferreira and Santa-Clara, 2011; Pettenuzzo *et al.*, 2014), and by technical indicators (Neely *et al.*, 2014).

The majority of existing contributions study equity premium out-of-sample forecasting using a small set of predictors (see Rapach and Zhou, 2013): for example, the Welch and Goyal (2008) dataset is made of 14 and 15 variables at monthly and quarterly frequency, respectively. Nevertheless, there is a longstanding evidence of comovements and latent factor structure in large datasets of stock returns[1] for which the variables can be decomposed into common and idiosyncratic components, mutually orthogonal at all leads and lags. Common components are driven by a small number of common, latent factors, which determine comovements in the data. This paper studies equity premium out-of-sample forecasting using a high number of macroeconomic predictors in the estimation of such factors.

Early work on factor models considered small-scale datasets: see Geweke (1977), Sargent and Sims (1977), Engle and Watson (1981), Watson and Engle (1983), Sargent (1989), and Stock and Watson (1989) which are all based on an approach commonly referred to as *exact factor models* in the sense that it imposes no cross-sectional correlation among idiosyncratic terms — a property which is unlikely to hold in economic data. On the contrary, large dimensional factor models, as pioeneered by Chamberlain and Rothschild (1983) only rely upon an *approximate factor structure* in which idiosyncratic terms are allowed to exhibit a mild amount of cross-correlation — a weaker restiction. In this spirit, more recent contributions study large scale information sets: see Connor and Korajczyk (1986, 1988), Forni *et al.*

[1]See Connor and Korajczyk (1986, 1988), Diebold and Nerlove (1989) Ng *et al.* (1992), Jones (2001) among many others.

(2000), Bai and Ng (2002), Stock and Watson (2002a, b), Forni *et al.* (2004, 2005) and Forni *et al.* (2015, 2017).

We focus on the three classes of factor models differing on how they account for the time series dependence in the common components, the estimation strategy and forecasting equation.

(a) Stock and Watson (2002a) estimate common factors by principal components (a static method based on contemporaneous covariances only) and compute predictions with a simple projection onto the factor space — i.e. a static representation in which the factors are loaded contemporaneously.

(b) Forni *et al.* (2005) also compute predictions in a static way as projections onto the factor space, but they allow for a data generating process with a dynamic representation known as the Generalized Dynamic Factor Model (GDFM) — that is, one in which the common factors are loaded dynamically via one-sided filters (Forni *et al.*, 2000).

(c) Forni *et al.* (2015, 2017) extend the *dynamic method* of Forni *et al.* (2005) by allowing for an infinite dimensional factor space which relaxes any restriction on the leading-lagging relations among the variables and common factors, and a dynamic forecasting equation. In this sense, they provide a fully fledged dynamic approach to the estimation of the GDFM.

Within each class, we consider several models that differ from each other by the number of common factors.

Existing evidence on the prediction of stock returns with large factor models is mixed and still limited to the static method. Ludvingson and Ng (2007), using a large dataset of both macroeconomic and financial data, find evidence of predictability on quarterly returns. As reviewed by Rapach and Zhou (2013), quartely predictions are often found to be more accurate. Neely *et al* (2014), find that the information conatined in the small macroeconomic dataset of Welch and Goyal (2008) is not useful in predicting monthly returns unless it is complemented by that conveyed by technical indicators. Baetje and Menkoff (2016) find that, unlike the predictability in technical indicators, that contained in the Welch and Goyal (2008) dataset is unstable and declined over time. Çakmaklı and van Dijk (2016) succesfully exploit large macroeconomic information in the prediction of monthly returns via factor augmented regressions. In a similar exercise on monthly returns and based on macroeconomic factors extracted from a large dataset, Gonçalves *et al.* (2017) find little but statistically significant predictability provided by some, but not all, factors.

Yet, none of these works analyses the performance of the GDFM in the prediction of stock returns. Recently, Forni *et al.* (2018) found that extracting factors from a large macroeconomic dataset mostly similar to the one we consider in this work, the GDFM often yields more accurate prdictions than the more common factor model based on the static approach. Motivated by this encouraging results, we fill an important gap in the stock return forecasting literature by contributing with the very first evidence of predictability based on the GDFM.

We use the recently proposed monthly FRED-MD large dimensional macroeconomic database of McCracken and Ng (2016) to conduct a pseudo real-time one-step-ahead equity premium forecasting exercise. We consider several forecasting methods (Giacomini and White, 2006; Timmermann, 2008) comprising aspects such as: the specification of the underlying factor model (Stock and Watson, 2002a; Forni *et al.*, 2005; Forni *et al.*, 2015, 2017); recursive or rolling estimation windows (Timmermann, 2008); statistical and economic evaluation criteria (Leicht and Tanner, 1991; Pesaran and Timmermann, 1995). In order to facilitate comparison with the existing literature and assess the role of the macroeconomic information contained in our large dataset, we also consider the updated, small-dimensional Welch and Goyal (2008) monthly dataset.

We obtain three main results. First, the information contained in large macroeconomic datasets leads to more accurate predictions which are superior both in statistical and economic terms. In fact, factor models estimated using the large dimensional McCracken and Ng (2016) database outperform those that employ the small-dimensional Welch and Goyal (2008) dataset and a range of medium to small datasets obtained via a LASSO-driven variable selection. Second, predictions based on the GDFM — either with the estimator of Forni *et al.* (2005) and that of Forni *et al.* (2015, 2017) — prevail over those based on the static method of Stock and Watson (2002*a*). Third, we also contribute by proposing a novel method selection criterion and select the best performing method in pseudo real-time (Pesaran and Timmermann, 2005; Timmermann, 2008): this allows us to pick a model within a given class at each point in time (Pesaran and Timmermann, 1995; Bossaerts and Hillion, 1999) and timely switch between estimation windows (Clark and McCracken, 2009). Our results show the benefits of the proposed method selection criterion, in particular when it is applied to the models of Forni *et al.* (2005) and Forni *et al.* (2015, 2017). This confirms the widespread idea that return predictability is a cyclical phenomenon closely related with the business cycle conditions.

Finally, we study the linkages between statistical and economic measures of forecast accuracy (Leicht and Tanner, 1991; Pesaran and Timmermann, 1995). We consider a risk-averse investor with mean-variance preferences and relative risk aversion parameter γ (see Rapach and Zhou, 2013 and references therein). Our results favour the factor models of Forni *et al.* (2005) and Forni *et al.* (2015, 2017); they also show that statistical and economic measures of forecast accuracy are positively correlated (Cenesizoglu and Timmermann, 2012), and that the strength of the correlation increases with γ.

Similar conclusions on return predictability are drawn by recent works related to our and based on a variety of alternative approaches. Ohno and Ando (2014) propose factor augmented regressions based on a shrinkage estimator. Lima and Meng (2017) also employ shrinkage in a quantile combination approach. Similarly, Pettenuzzo e Ravazzolo (2016) apply Bayesian model combination.

The remainder of the paper is organized as follows. Section 2 provides an overview of the factor models we use in our work. Section 3 describes the data. Section 4 assesses the out-of-sample predictive ability of the factor models. Finally, Section 5 concludes.

2 Latent Factor Models

Agents and policy makers form expectations and make decisions based on a large amount of information, virtually all the information available. Conditioning predictions on such an information set poses a challenge for standard estimation methods since projecting variables of interest on a large panel implies the estimation of a huge number of parameters and too few degrees of freedom (the *curse of dimensionality*). For example, in an n-dimensional vector autoregressive (VAR) model the number of parameters increases with the square of n so that considering more and more variables quickly becomes unfeasible. On the contrary, dynamic factor models retain parsimony and solve the curse of dimensionality by aggregating variables in order to obtain a few latent factors accounting for the bulk of the dynamics in the panel. The methods we consider differ in the way such aggregation is done.

Let $\boldsymbol{x}_t = (x_{it}, \ldots, x_{nt})'$ be a panel of covariance stationary time series x_{it} (with cross-section $i = 1, \ldots, n$; and time $t = 1, \ldots, T$), $\boldsymbol{\Gamma}_k = E\boldsymbol{x}_t\boldsymbol{x}'_{t-k}$ its covariance matrix, and $\boldsymbol{\Sigma}(\theta)$ its spectral density matrix at frequency $\theta \in [-\pi, \pi]$. Define $\{v_j, z_j\}_{j=1}^{n}$ and $\{\lambda_j(\theta), p_j(\theta)\}_{j=1}^{n}$ the eigenvalues (sorted in decreasing order) and the corresponding eigenvectors of $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Sigma}(\theta)$, respectively. Factor models imply the orthogonal decomposition

$$x_{it} = \chi_{it} + \xi_{it},$$

where χ_{it} is x_{it}'s *common* component in the sense that is driven by common factors, and ξ_{it} is its *idiosyncratic* component. Since the dynamics of the common components are driven by relatively few latent factors, the number of parameters in factor models does not increase with n. Furthermore, consistent estimation is typically achieved as $n \to \infty$ and, for this reason, factor models are deemed to be gifted with the *blessing of dimensionality.*

As the two components are mutually orthogonal at all leads and lags, the same decomposition holds true for both $\mathbf{\Gamma}_k$ and $\mathbf{\Sigma}(\theta)$, that is

$$\begin{aligned} \mathbf{\Gamma}_k &= \mathbf{\Gamma}_k^{\chi} + \mathbf{\Gamma}_k^{\xi}, \\ \mathbf{\Sigma}(\theta) &= \mathbf{\Sigma}^{\chi}(\theta) + \mathbf{\Sigma}^{\boldsymbol{\xi}}(\theta), \end{aligned}$$

where $\mathbf{\Gamma}_k^{\chi}$ and $\mathbf{\Gamma}_k^{\xi}$ are common and idiosyncratic covariances, and $\mathbf{\Sigma}^{\chi}(\theta)$ and $\mathbf{\Sigma}^{\xi}(\theta)$ are common and idiosyncratic spectral densities.

Approximate factor structures are inferred both in the time and frequency domain. In fact, as $n \to \infty$ we have:

(i) the number $r \ll n$ of *static* common factors corresponds to the number of diverging eigenvalues v_j of $\mathbf{\Gamma}_0$ (Bai and Ng, 2002),

(ii) the number $q \ll n$ of *dynamic* common factors is equal to the number of spectral eigenvalues $\lambda_j(\theta)$ diverging almost everywhere in $[-\pi, \pi]$ (Hallin and Liska, 2007).

In the same way, as $n \to \infty$, idiosyncrasy is characterized by bounded idiosyncratic eigenvalues[2] and spectral eigenvalues.

In the rest of the paper, we refer to a *dynamic estimation method* for models with dynamic factors estimated considering the common-idiosyncratic decomposition of the spectral density matrix (Brillinger, 1981) and therefore account for the whole covariance structure of the data. To this category belong the

[2]Therefore a limited amount of cross-correlation between the idiosyncratic terms is allowed. This is the distinctive feature of the approximate factor models described here with the *exact* factor model studied by Geweke (1977) and Sargent and Sims (1977). Furthermore, serial correlation in the idiosyncratic terms is not dismissed.

factor models described in subsections 2.2 and 2.3. On the other hand, the *static estimation method* refers to models employing static factors which are estimated considering contemporaneous covariances only — i.e. Γ_0 — rather than the whole covariance structure of the data. To this category belongs the factor model described in subsection 2.1.

The most general factor model, known as the *Generalized Dynamic Factor Model* (GDFM), involves the following *dynamic representation* for the common components

$$\chi_{it} = \frac{\boldsymbol{c}_i(L)}{\boldsymbol{d}_i(L)} \boldsymbol{f}_t, \tag{1}$$

where $\boldsymbol{c}_i(L)$ and $\boldsymbol{d}_i(L)$ are one-sided polynomials in the lag operator L with square-summable coefficients, and $\boldsymbol{f}_t$ is a $q-$dimensional orthonormal white noise (see Forni and Lippi, 2001; Forni *et al.* 2000, 2004). The main advantage with respect to competing factor models is that, beyond stationarity and regularity conditions for the existence of spectral density matrix $\boldsymbol{\Sigma}(\theta)$, the GDFM does not require assumptions on the dynamics of factor structure to achieve consistent estimation of the common components. On the contrary, the widespread *static representation*

$$\chi_{it} = \boldsymbol{\lambda}_i' \boldsymbol{F}_t, \tag{2}$$

where $\boldsymbol{\lambda}_i$ are factor loadings and the factors $\boldsymbol{F}_t$ are only cross-sectionally uncorrelated, imposes strong restrictions on the data generating process if consistency is to be achieved (see Hallin and Lippi, 2013 and Forni *et al.* 2015, 2017). Nevertheless, factor models with static representation should still be considered dynamic time series models because they can accommodate some forms of dynamics. For instance,

$$\chi_{it} = \quad a_{i1} \underbrace{\frac{f_{1t}}{1-\alpha L}}_{F_{1t}} \quad + \quad a_{i2} \underbrace{f_{2t}}_{F_{2t}} \quad + \quad a_{i3} \underbrace{f_{2t-1}}_{F_{3t}} \tag{3}$$

allows for a static representation with 3 static factors (F_{1t}, F_{2t}, F_{3t}) and a dynamic representation with 2 dynamic factors (f_{1t}, f_{2t}). On the contrary,

$$\chi_{it} = \quad a_i \frac{f_{1t}}{1-\alpha_i L} = \quad a_i \left(f_{1t} + \alpha_i f_{1t-1} + \alpha_i^2 f_{1t-2} + \dots\right) \tag{4}$$

does not allow for a (finite-order) static representation. Notice that, while in equation (3) the impulse response functions of all common components χ_{it} to f_{1t} only differ up to the scaling term a_{i1}, in equation (4) heterogeneity is allowed in the shape of the impulse responses to f_{1t}. To such dynamic heterogeneity does not correspond a finite-order static representation, as it can be seen in the term in brackets in equation (4).

The model in subsection 2.3 is based on the dynamic representation (1) while those in subsections 2.1 and 2.2 are based on the static representation (2).

2.1 Static method, static representation (SW)

The method proposed by Stock and Watson (2002*a*) — henceforth SW — involves static principal components and projections on the factor space.

Static factors are extracted from $\boldsymbol{x}_t$ by taking r principal components of $\widehat{\boldsymbol{\Gamma}}_0$ (the sample counterpart of $\boldsymbol{\Gamma}_0$) solving the eigenvalue problem

$$z_j \widehat{\boldsymbol{\Gamma}}_0 = v_j z_j, \qquad j = 1, \dots, r.$$

The estimated factors are $\boldsymbol{F_t^{SW}} = \widehat{\boldsymbol{z}} \boldsymbol{x}_t$, where $\widehat{\boldsymbol{z}} = (\widehat{z}_1, \dots, \widehat{z}_r)'$, and the $h-$step ahead forecast of χ_{it} is

$$\chi_{it+h|t}^{SW} = \widehat{\boldsymbol{\Gamma}}_h \widehat{\boldsymbol{z}}' \left(\widehat{\boldsymbol{z}} \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{z}}' \right)^{-1} \boldsymbol{F_t^{SW}}. \tag{5}$$

2.2 Dynamic method, static representation (FHLR)

The dynamic method proposed by Forni *et al.* (2005) — henceforth FHLR — is a two-step procedure based on the dynamic estimation method and predictions formed from a static representation via a *constrained* projection onto the factor space.

Step one: estimation

The spectral density matrix of the data at frequency $\theta \in [-\pi, \pi]$ is estimated through discrete Fourier transforms of the sample covariance matrix

$$\widehat{\boldsymbol{\Sigma}}(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} e^{-ik\theta} w_k \widehat{\boldsymbol{\Gamma}}_k,$$

where w_k are the weights of a window function[3] and M is a truncation parameter.

Letting $\widehat{p_j}(\theta)$ and $\widehat{\lambda_j}(\theta)$ be the eigenvector and eigenvalues of $\widehat{\boldsymbol{\Sigma}}(\theta)$, the spectral density matrices of the common and idiosyncratic components are computed as

$$\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\chi}}(\theta) = \sum_{j=1}^{q} \widehat{\lambda_j}(\theta)\, \widehat{p_j}'(\theta)\, \widehat{p_j}(\theta), \tag{6}$$

$$\widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\xi}}(\theta) = \sum_{j=q+1}^{n} \widehat{\lambda_j}(\theta)\, \widehat{p_j}'(\theta)\, \widehat{p_j}(\theta), \tag{7}$$

respectively; the covariances via inverse Fourier transforms are

$$\widehat{\boldsymbol{\Gamma}}_k^{\chi} = \frac{2\pi}{2H+1} \sum_{j=-H}^{H} e^{ik\theta} \widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\chi}}(\theta_j), \tag{8}$$

$$\widehat{\boldsymbol{\Gamma}}_k^{\xi} = \frac{2\pi}{2H+1} \sum_{j=-H}^{H} e^{ik\theta} \widehat{\boldsymbol{\Sigma}}^{\boldsymbol{\xi}}(\theta_j), \tag{9}$$

with Fourier frequencies $\theta_j = \frac{2\pi j}{2H+1}$.

Step two: forecasting equation

The so-called *generalized principal components* of the couple $\left(\widehat{\boldsymbol{\Gamma}}_0^{\chi}, \widehat{\boldsymbol{\Gamma}}_0^{\xi}\right)$ solve the eigenvalue problem

$$\widehat{z}_j^g \widehat{\boldsymbol{\Gamma}}_0^{\chi} = \widehat{v}_j^g \widehat{z}_j^g \widehat{\boldsymbol{\Gamma}}_0^{\xi}, \qquad j = 1, \ldots, r$$

$$\text{subject to} \quad \begin{cases} \widehat{z}_j^g \widehat{\boldsymbol{\Gamma}}_0^{\xi} \widehat{z}_j^g = 1 \\ \widehat{z}_i^g \widehat{\boldsymbol{\Gamma}}_0^{\xi} \widehat{z}_j^g = 0, \quad i \neq j. \end{cases}$$

Letting $\widehat{\boldsymbol{z}}^g = (\widehat{z}_1^g, \ldots, \widehat{z}_r^g)'$ be a vector of the first r generalized principal components[4], the estimated factors are $\boldsymbol{F_t^{FHLR}} = \widehat{\boldsymbol{z}}^g \boldsymbol{x}_t$. Finally, the $h-$step ahead forecast of χ_{it} is given by

$$\chi_{it+h|t}^{FHLR} = \widehat{\boldsymbol{\Gamma}}_h^{\chi} \widehat{\boldsymbol{z}}^{g\prime} \left(\widehat{\boldsymbol{z}}^g \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{z}}^{g\prime}\right)^{-1} \boldsymbol{F_t^{FHLR}},$$

[3] All empirical results in Section 4 are obtained using a standard triangular window.

[4] Generalized principal components correspond to principal components of data weighted according to their signal to noise ratio.

which is a *constrained* projection onto the factor space since it imposes the dynamic restrictions of factor structure by using $\widehat{\Gamma}_h^{\chi}$ rather than $\widehat{\Gamma}_h$ as in the unconstrained projection (5) employed by SW.

2.3 A fully fledged dynamic method (FHLZ)

The method proposed by Forni *et al.* (2015, 2017) — henceforth FHLZ — shares with FHLR the decomposition of the spectral density matrix (6) - (7) and that of the covariances (8) - (9) estimated as in subsection 2.2, Step one.

Given that, as pointed out by Anderson and Deistler (2011), "tall" filters — i.e. with more rows than columns — are *generically*[5] full column rank, then any $q+1-$dimensional vector[6] of common components $\boldsymbol{\chi}_t^{(i)}$ allows for a zeroless moving average representation which is *fundamental*[7] and can be inverted into an autoregressive representation

$$\boldsymbol{A}^{(i)}(L)\,\boldsymbol{\chi}_t^{(i)} = \mathbf{R}^{(i)}\boldsymbol{f}_t,$$

where $\boldsymbol{A}^{(i)}(L)$ is $(q+1)\times(q+1)$ and $\mathbf{R}^{(i)}$ is $(q+1)\times q$. Stacking all $q+1-$dimensional vectors[8] of common components we have an autoregressive representation in which the dynamic factors $\boldsymbol{f}_t$ are loaded only contemporaneously in $\boldsymbol{A}^{(i)}(L)\,\boldsymbol{\chi}_t^{(i)}$ and therefore can be consistently estimated[9] via principal components of filtered data

$$\boldsymbol{Z}_t = \underline{\boldsymbol{A}}(L)\,\boldsymbol{x}_t = \underline{\boldsymbol{R}}\boldsymbol{f}_t + \underline{\boldsymbol{A}}(L)\,\boldsymbol{\xi}_t,$$

where, with obvious notation, $\mathbf{Z}_t$ collects stacked vectors $\mathbf{z}^{(i)} \equiv \boldsymbol{A}^{(i)}(L)\,\boldsymbol{\chi}_t^{(i)}$, $\underline{\boldsymbol{R}}$ is a tall $n\times q$ matrix and the $n\times n$ autoregressive filter takes the form

$$\underline{\boldsymbol{A}}(L) = \begin{pmatrix} \boldsymbol{A}^{(1)}(L) & 0 & \dots & 0 \\ 0 & \boldsymbol{A}^{(2)}(L) & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \boldsymbol{A}^{(g)}(L) \end{pmatrix}.$$

Letting $(\omega_1,\dots,\omega_q)$ be the first q eigenvalues of the covariance matrix of $\boldsymbol{Z}_t$ and $\Psi = (\psi_1,\dots,\psi_q)'$ be

[5] That is, the property holds everywhere in the parameter space apart from a measure zero subset.

[6] E.g. $\boldsymbol{\chi}_t^{(1)} = (\chi_{1t},\dots,\chi_{q+1t})$.

[7] That is, it has no zeros in the complex unit disk. Non-zeroless moving averages admit a multitude of nonfundamental representations which cannot be inverted into causal vector autoregressive representations (see e.g. Soccorsi, 2016).

[8] In order to avoid heavier notation, we are assuming without loss of generality that $g = n/(q+1)$; no special challenge arises when n is not a multiple of $q+1$.

[9] In Section 4 results are obtained by determining the lag order of $\boldsymbol{A}^{(i)}(L)$ via a BIC information criterion.

a $q \times n$ matrix collecting the associated eigenvectors, the estimated dynamic factors are $\boldsymbol{f_t^{FHLZ}} = \Psi \boldsymbol{z}_t$.

The autoregressive coefficients in $\underline{\mathbf{A}}(L)$ are computed from the estimated common covariances (8) and $\widehat{\underline{\boldsymbol{R}}} = \Psi$. Given these quantities we have estimated impulse responses to the dynamic factors

$$\mathbf{w}(z) = \widehat{\underline{\boldsymbol{A}}(\boldsymbol{z})}^{-1} \widehat{\underline{\boldsymbol{R}}} \quad .$$

Finally, $h-$step ahead predictions are:

$$\chi_{it+h}^{FHLZ} = \mathbf{w}_{ih} \boldsymbol{f_t^{FHLZ}} + \mathbf{w}_{ih-1} \boldsymbol{f_{t-1}^{FHLZ}} + \dots \quad . \tag{10}$$

3 Data

Letting T_e be the end point of the estimation period, our aim is to produce one-step-ahead out-of-sample forecasts for the sequence of stock returns at each point in time $\tau + 1$ given the information available at time τ, for $\tau = T_e, \dots, T - 1$. We use monthly data on stock returns along with a large set of predictors from which we estimate the factors: these are 122 variables included in the FRED-MD database described in McCracken and Ng (2016). Our data sample spans the period January 1960 to December 2014: this is dictated by the availability of the series included in the FRED-MD database. We also use the 14 predictors originally proposed in Welch and Goyal (2008) and subsequently extended up to 2014 by the same authors: this allows for comparison with existing studies using low dimensional sets of predictors (see Rapach and Zhou, 2013 and references therein).

Stock returns are continuously compounded returns on the S&P 500 index in excess of a short T-bill rate and include dividends. Formally, let P_t denote the value of the S&P 500 at time t. The excess log return ρ_{t+1} at period $t+1$ is defined as $\rho_{t+1} = \ln(P_{t+1}) - \ln(P_t) - r_{ft}$, for $t = 1, \cdots, \tau - 1$, where r_{ft} is the known return on the risk-free asset from period t to period $t + 1$. The goal is to produce one-step-ahead out-of-sample forecasts of $\rho_{\tau+1}$ given the information set available at time τ, for $\tau = T_e, \dots, T - 1$.

The FRED-MD database organizes the variables into eight groups: (1) output and income; (2) labor market; (3) consumption and orders; (4) orders and inventories; (5) money and credit; (6) interest rate and exchange rates; (7) prices; (8) stock market. The choice of the 122 variables was based on data availability over the period of interest as reported in the Appendix. Given the sample under consideration, we use the January 2015 vintage.

The 14 predictors proposed in Welch and Goyal (2008) are: log dividend-price ratio (log (DP)), log dividend-yield (log (DY)), log earnings-price ratio (log (EP)), log dividend-payout ratio (log (DE)), stock variance (SVAR), book-to-market ratio (BM), net equity expansion (NTIS), treasury bill rate (TBL), long-term yield (LTY), long-term return (LTR), term spread (TMS), default yield spread (DFY), default return spread (DFR), lagged inflation[10] (INFL). The predictors fall into the following broad categories (see Pettenuzzo *et al.*, 2014): (*i*) valuation ratios (log (DP), log (DY), log (EP), BM); (*ii*) measures of bond yields (TBL, LTY, TMS, DFY, DFR); (*iii*) estimates of equity risk (LTR, SVAR); corporate finance variables (log (DE), NTIS); (*iv*) macroeconomic variables (INFL).

Table 1 about here

Table 1 provides summary statistics for the series of excess stock returns and for the variables included in the Welch and Goyal (2008) dataset. Despite the difference in the sample period of interest, the figures are aligned to those displayed in Table 1 in Pettenuzzo *et al.* (2014)[11].

A crucial issue in out-of-sample forecasting exercises is the choice of the sample-split between estimation and evaluation periods to avoid data mining (Hansen and Timmermann, 2012). As in Timmermann (2008), we use the first 10 years of data as a training sample and we evaluate the forecasts over the period January 1970 to December 2014[12]: a long evaluation sample allows for stronger power of forecast evaluation tests and minimizes the likelihood of spurious rejections (Hansen and Timmermann, 2012). The end point T_e of the estimation window thus is December 1969.

4 Out-of-Sample Analysis

4.1 Forecasting Methodology

As in Timmermann (2008), we explicitly follow Giacomini and White (2006) and distinguish between forecasting *model* and forecasting *method*. The former refers to the underlying econometric specification, in our case the three factor models described in Section 2. The latter includes the model and other

[10]As in Welch and Goyal (2008), we lag inflation by an extra period to allow for the delay in CPI releases.
[11]Table 1 in Pettenuzzo *et al.* (2014) covers the sample period January 1927 to December 2010.
[12]The sample used in Timmermann (2008) begins in 1959 : 12, ours in 1960 : 01.

choices made by the forecaster, such as the estimator for the model unknown parameters (as discussed in Section 2), the length of the estimation window, and the evaluation criteria.

4.1.1 Econometric Model

We estimate the factors from the small-dimensional Welch and Goyal (2008) dataset and from the large collection of FRED-MD variables. In the former case, we consider $r = 1, 2, 3$ static and $q = 1, 2$ dynamic factors. From the FRED-MD database, we estimate up to $r = 15$ and $q = 5$ static and dynamic factors, respectively: for ease of exposition and without loss of generality, we report results for $r = 1, 2, 3, 4, 5, 10, 15$ only.

4.1.2 Estimation Window

We consider recursive window and rolling window estimation schemes (Timmermann, 2008). Given the sample split described in Section 3, the first one uses data from 1960:01 up to the time the forecast is made to produce a series of one-step ahead forecasts — i.e. the first forecast uses data from 1960:01 to 1969:12 to obtain an out-of-sample prediction for 1970:01; the second uses data from 1960:01 to 1970:01 to produce a forecast for 1970:02, and so on. As in Timmermann (2008), the rolling window scheme employs a fixed-length window of the most recent ten years of data (i.e., 120 monthly observations) to estimate the models and to produce the sequence of one-step ahead forecasts. The recursive window scheme is commonly used in the empirical literature on out-of-sample stock return forecasting (see, e.g., Pesaran and Timmermann, 1995; Campbell and Thompson, 2008; Welch and Goyal, 2008; Rapach *et al.*, 2010; Pettenuzzo *et al.*, 2014). The rolling window scheme is common practice in the macroeconomic forecasting literature (see, e.g., D'Agostino and Giannone, 2012; Stock and Watson, 2012; and Forni *et al.*, 2018) concerned with structural breaks in macroeconomic data.

4.1.3 Evaluation Criteria

The first evaluation criterion we consider is the mean squared prediction error (MSPE), which assesses the absolute performance of a sequence of forecasts (Pesaran and Timmermann, 1995; Bossaerts and Hillion, 1999). We next compare the forecasts obtained from the factor models in relation to a given benchmark. Following common practice in the literature (Campbell and Thompson, 2008; Welch and

Goyal, 2008), we take as a benchmark the prevailing mean (PM), namely

$$\rho_{t+1} = \alpha + \varepsilon_{t+1}, \quad t = 1, \ldots, \tau - 1, \quad \tau = T_e, \ldots, T - 1, \tag{11}$$

where ε_{t+1} is a white noise error term with unpredictable mean. The equity premium forecast for period $\tau+1$ made at time τ is $\hat{\rho}_{\tau+1,rec} = \tau^{-1}_{t=1}{}^{\tau}\ \rho_t$ under recursive window; it is equal to $\hat{\rho}_{\tau+1,rol} = T_e^{-1}{}^{\tau}_{t=\tau-T_e+1}$ ρ_t under rolling window. The recursive window scheme produces the benchmark usually employed in the equity premium forecasting literature (Campbell and Thompson, 2008; Welch and Goyal, 2008). The choice of the estimation window is a function of the underlying assumption made about the mean of the equity premium (Timmermann, 2008): when estimated recursively, the model in (11) assumes the equity premium has a constant mean and it is not predictable; the rolling window scheme implies that the mean of the equity premium slowly changes over time.

The MSPE may be used to measure the out-of-sample goodness of fit of a sequence of forecasts. To this purpose, we next consider the out-of-sample R^2 (Campbell and Thompson, 2008; Timmermann, 2008; Welch and Goyal, 2008; Rapach *et al.*, 2010; Pettenuzzo *et al.*, 2014). Let MSPE_1 and MSPE_0 be the mean squared prediction errors from any factor model and from the prevailing mean in (11), respectively: the out-of-sample R^2 is $R^2_{OoS} = 1 - \text{MSPE}_1 / \text{MSPE}_0$. By construction, $R^2_{OoS} \leq 0$ if and only if $\text{MSPE}_1 \geq \text{MSPE}_0$, meaning that the benchmark is at least as good as the alternative model at forecasting $\rho_{\tau+1}$; conversely, $R^2_{OoS} > 0$ if and only if $\text{MSPE}_1 < \text{MSPE}_0$.

Finally, we assess the statistical significance of the improvement of the alternative model over the benchmark by testing the null hypothesis $R^2_{OoS} \leq 0$ against the one-sided alternative $R^2_{OoS} > 0$. We run Clark and West (2007) test (CW): this is robust to the different degrees of estimation error between models, which would otherwise favor the more parsimonious benchmark.

4.1.4 The Role of the Business Cycle

Rapach *et al.* (2010), Henkel *et al.* (2011), and Rapach and Zhou (2013) argue that stock returns predictability exhibit discernible patterns linked to business cycle dynamics. We then assess our forecasts over the entire evaluation period, as well as during NBER-dated expansions and recessions.

4.2 Empirical Results

4.2.1 Recursive Window

Table 2 displays the out-of-sample R^2 for the recursive window scheme.

Table 2 about here

When factors are extracted from Welch and Goyal (2008) small dimensional dataset (Panel A), the best performing model is FHLZ with $q = 1$ dynamic factor: the model outperforms all other specifications over the entire evaluation period, as well as during recessions and expansions; the out-of-sample R^2 is always positive and significant at 5% level or less[13]. The out-of-sample R^2 is equal to 0.95% over the whole evaluation period, and to 0.75% and 1.39% during expansions and recessions, respectively. The forecasts produced by FHLZ with $q = 1$ dynamic factor are thus more accurate during contractionary periods: this is consistent with Campbell and Cochrane (1999), Menzly *et al.* (2004), and Bekaert *et al.* (2009), who argue that risk premia are countercyclical and drive (at least part of) predictability; it also resembles Rapach *et al.* (2010), Henkel *et al.* (2011), and Rapach and Zhou (2013), who empirically show that out-of-sample stock returns predictability increases during recessions as compared to expansions.

Panel B in Table 2 shows that when the whole large dimensional FRED-MD dataset is used to estimate the factors, the best performing model is FHLR. The specifications with $r = 2, 3$ static factors and $q = 1$ dynamic factor overall produce the most accurate forecasts, with statistically significant improvements over the prevailing mean at 10% level or less. These two models also outperform FHLZ with $q = 1$ dynamic factor estimated from the small-dimensional Welch and Goyal (2008) dataset (see Panel A): with the data at hand, large dimensional factor models provide a hedge over smaller scale counterparts.

In mean square terms SW doesn't seem to provide more accurate forecasts with the best specification — i.e. $r = 2$ — being associate with a meagre $R^2 = 0.02$. Nevertheless, most of its specifications are still statistically more accurate according to Clark and West (2007) test results. Similar conclusions are drawn by Gonçalves *et al.* (2017) with an approach based on ad-hoc inference for static factors.

[13]As stressed in footnote 19 in Pettenuzzo *et al.* (2014), the p-values from the Clark and West (2007) test should be interpreted with caution and in line with Diebold (2015): those p-values should be intended to compare forecasts rather than models.

Boivin and Ng (2006) question whether adding series with little factor structure to estimate factors may result in factors being less useful for out-of-sample forecasting purposes. Panels C, D and E in Table 2 show results for the same models as in Panel B, but where factors are estimated from 25, 50 and 75 series from the FRED-MD dataset, respectively: the series are selected through a pseudo-real time LASSO procedure at each time τ the forecast is made. When only a subset of the series is used to estimate the factors, the forecasting ability of the models, as measured by the out-of-sample R^2, deteriorates. The performance of the forecasts improves as more series are used to estimate the factors. This result shows the usefulness of large data and is in line with basic asymptotic results on consistent factor estimation, which is achieved for growing cross-sections.

Given the data at hand, large dimensional datasets are more informative than small-dimensional counterparts in a recursive window framework; FHLR models overall produce the most accurate forecasts with $r = 2, 3$ static factors and $q = 1$ dynamic factor.

4.2.2 Rolling Window

Table 3 displays the out-of-sample R^2 for the rolling window scheme.

Table 3 about here

As with the recursive window, factor models estimated using the large dimensional FRED-MD dataset (see Panel B) generally produce more accurate forecasts than those based on the smaller scale Welch and Goyal (2008) data (see Panel A). Over the whole forecasting period, FHLZ forecasts are the most precise: the out-of-sample R^2 ranges between 1.17% and 2.02%, and it is always statistically greater than zero at least at 5% level. During expansionary periods, no large dimensional model unambiguously dominates any other specification: in particular, FHLZ with $q = 2$ dynamic factors as obtained from the Welch and Goyal (2008) dataset has the highest out-of-sample R^2 out of all models, which is equal to 0.54% and it is significant at 10% level. During economic recessions, FHLR and FHLZ models combined with the large dimensional FRED-MD dataset produce forecasts of similar accuracy: almost all out-of-sample R^2 are positive and significant at least at 10% level. The forecasts from SW models are statistically significant for low values of the number of static factors r, although they are always negative. When factors are estimated from a subset of 25, 50 and 75 variables from the FRED-MD dataset selected using a LASSO type procedure (see Panels C, D and E, respectively) the quality of the forecasts deteriorates; as in the case of the recursive window scheme, the precision of the forecasts improves as more variables are used to estimate the factors.

In conclusion, the results from the rolling window scheme generally favor large dimensional factor models, with FHLR and, especially, FHLZ having a hedge over SW.

4.2.3 The role of large macroeconomic information

In order to asses the role of our large cross-section of macroeconomic information, in Table 4 we resume the LASSO results in Tables 2 and 3 by picking the best specifications of SW, FHLR, FHLZ for each level of varible selection — namely, LASSO 25, LASSO 50, LASSO 75 — and under both estimation windows. Comparing these results obtained considering the restricted cross sections composed by the variables selected by the LASSO at each point in time, with those relative to the full FRED-MD dataset, we observe an almost monotonic improvement in the number of variables included. That is, the information in such large macroeconomic data is useful in the prediction of stock returns. In line with standard asymptotic results on factor models (references in Section 2), this result is to be interpreted as a clear sign that the underlying assuptions do hold in the data and the models we consider are not misspecified. In this sense, we confirm the conclusion of McCracken and Ng (2016) who proposed the FRED-MD dataset as a resource for factor analysis.

Also similar results on the performance of shrinkage methods applied to forecasting problems based on this kind of data are found by Stock and Watson (2012) and Giannone *et al.* (2017).

Table 4 about here

4.3 An Adaptive Method Selection Approach

The results discussed in Sections 4.2.1 and 4.2.2 and displayed in Tables 2 and 3, respectively, show two important findings: factor models estimated using large dimensional datasets tend to produce more reliable forecasts than those estimated using a smaller number of macroeconomic series; models based on the dynamic method (i.e., FHLR and FHLZ) outperform SW, based on the static method. These findings come from a high number of forecasting methods.

Furthermore, we find an empirical regularity along the business cycle: rolling forecast are more accurate during recessions while recursive forecasts are more accurate during expansions. However, decision makers require selecting the best performing method in real-time (Pesaran and Timmermann, 2005). To this purpose, we implement what we label a *method selection criterion*: this allows us to pick a model within a given class at each point in time (Pesaran and Timmermann, 1995; Bossaerts and Hillion, 1999) and to timely switch between estimation windows (in this spirit, see Clark and McCracken, 2009; Pesaran and Timmerman, 2007). So doing, we can exploit the (stylized) fact that return predictability is a cyclical phenomenon (see e.g. Rapach and Zhou, 2013).

4.3.1 Model Selection Strategy

In the spirit of Pesaran and Timmermann (1995) and Bossaerts and Hillion (1999), we study the model selection problem within SW, FHLR and FHLZ for a given estimation window.

When implementing model selection criteria using the recursive window scheme, we have to account for structural instability in the underlying factor model (see Baltagi *et al.* (2017) and references therein). Among others, Breitung and Eickmeier (2011) prove that ignoring breaks in the loadings leads to over-estimating the number of factors; this theoretical result is confirmed by the visual inspection of Figure 2 in McCracken and Ng (2016). Figure 1 shows the number of static and dynamic factors selected at each point in time by Bai and Ng (2002), and Hallin and Liška (2007) criteria, respectively, using the recursive window scheme. The figure clearly shows that the estimated number of static and dynamic factors is not constant over time: this provides evidence of structural breaks in the underlying factor models.

Figure 1 about here

We adopt the following strategy to tackle the problem of model selection in the presence of structural instability under the recursive window scheme. As suggested in Stock and Watson (2012), we *a priori* select $r = 5$ static factors and we keep this number fixed over the entire out-of-sample evaluation period. To the very best of our knowledge, no existing study allows us to *a priori* fix the number of dynamic factors. At each point in time, we choose $q = 4$ dynamic factors using Hallin and Liška (2007) criterion as applied to the rolling window scheme: model instability is less likely to affect this estimation scheme as the dynamic window effectively adapts to time variation in the loadings[14].

The empirical results with $r = 5$ and $q = 4$ show that FHLR and FHLZ outperform SW (see Panel B in Table 2): the former two produce forecasts with higher out-of-sample R^2 than the latter during the entire out-of-sample evaluation period, as well as during expansions and recessions. Between the two dynamic models, FHLZ generates more accurate forecasts than FHLR over the full evaluation period and in expansionary phases: the out-of-sample R^2 is equal to 0.75% and 0.56%, respectively, and in both cases it is significant at 5% level. During recessions, the out-of-sample R^2 from FHLZ forecasts is marginally higher than that from FHLR forecasts; in the latter case, statistical significance is achieved at 10% level. Overall, FHLZ has an edge over FHLR under the recursive window scheme.

Under rolling window estimation, at each point in time we choose the number of static and dynamic factors according to Bai and Ng (2002), and Hallin and Liška (2007) criteria, respectively: the problem of structural instability is likely to be less relevant in this case, as rolling window estimation accounts for time-variation in the parameters. As with the recursive window scheme, FHLR and FHLZ fare better than SW (see Panel B in Table 3). Between the two dynamic models, FHLZ forecasts are better during the whole evaluation period: the out-of-sample R^2 is equal to 1.71% and significant at 5% level. Rolling forecasts from FHLR are more accurate during recessions: the higher out-of-sample R^2 is 5.36% and significant at 5% level.

4.3.2 Switching the Estimation Window in pseudo real-time

Return prediction models are subject to structural instabilities (Paye and Timmermann, 2006; Rapach and Wohar, 2006). Clark and McCracken (2009) and Pesaran and Timmerman (2007) argue that forecast accuracy in the presence of structural breaks may be improved by combining recursive and rolling estimation windows in such a way that optimally handles the trade-off between variance (which decreases with the sample size and so in recursive windows) and bias (which is generated by the breaks and so is less harmful within rolling windows). We propose to select the estimation window as a function of the business cycle conditions so that forecast accuracy can enhance in the presence of instabilities linked to the business cycle. This is relevant as out-of-sample stock returns predictability is commonly believed to depend on the business cycle: Rapach *et al.* (2010) find improved predictive accuracy from forecast

[14] At each point in time, Hallin and Liška (2007) criterion selects between $q = 4$ and $q = 5$ dynamic factors. We choose $q = 4$ as a matter of parsimony. Results with $q = 5$ are very similar and are available upon request.

combinations with weights related to the business cycle.

Table 5 about here

In order to empirically motivate our strategy, Table 4 reports mean squared prediction errors (multiplied by 100) for the three large dimensional factor models we consider under recursive and rolling window estimation (Panels A and B, respectively). We first look at the whole set of available factor models. During economic expansions, all models produce better forecasts under recursive window estimation. The scenario changes during contractionary periods: forecasts from SW models have similar MSPE under recursive and rolling windows; FHLR and FHLZ models produce more accurate forecasts under rolling window. This finding is confirmed when the model selection strategy detailed in Section 4.3.1 is applied within each class of factor models: this *a priori* selects $r = 5$ and $q = 4$ under recursive window; it resorts to Bai and Ng (2002), and Hallin and Liška (2007) criteria under rolling window. The results confirm that FHLR and FHLZ generate better forecasts under recursive and rolling window during expansions and recessions, respectively. This last finding suggests that timely switching between estimation windows may improve the quality of the forecasts.

In order to select the estimation window, we employ a nowcasting procedure that tracks the current state of the economy (see Banbura *et al.*, 2011). At each point in time τ, we use the sequence $\{\text{ADS}_t\}_{t=1}^{\tau}$ of business cycle indicators of Aruoba *et al.* (2009) and select the estimation window by solving

$$\hat{\theta}_\tau = \arg\min_\theta \left| \left[\tau^{-1} \sum_{t=1}^{\tau} \mathbb{I}\left(\text{ADS}_t < \theta\right) \right] - R \right|, \quad \tau = T_e, \ldots, T-1,$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and $|\cdot|$ the absolute value of the argument: $R = 0.14$ is the approximate sample frequency of recessions over the $1946:01 - 1969:12$ period as identified by the NBER business cycle dates[15]. At each point in time τ, the threshold $\hat{\theta}_\tau$ minimizes the distance between the empirical frequency R and the one identified by Aruoba *et al.* (2009) business cycle indicator; for

[15]We also considered the case in which the empirical frequency of recessions at time τ is determined by the expanding window between $1946:01$ and τ. The results are quantitatively very similar to those from the case $R = 0.14$ and are available upon request.

each τ, we select recursive and rolling window if $\text{ADS}_\tau \geq \hat{\theta}_\tau$ and $\text{ADS}_\tau < \hat{\theta}_\tau$, respectively.

Table 6 about here

Results related to the proposed model selection criterion are displayed in Table 5. The table collects results for the PM model in (11), and for the large dimensional factor models SW, FHLR and FHLZ: under recursive and rolling windows, factor models are estimated as detailed in Sections 4.2.1 and 4.2.2, respectively; the model selection criterion chooses the estimation window according to the procedure previously described in this section. The results show that the mean squared prediction error (multiplied by 100) for FHLR and FHLZ is minimized when the method selection criterion is implemented and it is equal to 0.1981 and 0.1967, respectively (see Panel A): these value are lower than any other obtained from PM and SW. Table 5 also calculates the out-of-sample R^2 with respect to the most accurate benchmark, namely PM estimated by recursive window (Panel B): SW delivers negative values regardless of the estimation window; FHLR forecasts obtained from the method selection criterion outperform the benchmark with a positive out-of-sample R^2 equal to 1.11%; FHLZ produces the most accurate forecasts, which always deliver positive values for the out-of-sample R^2, with highest value equal to 1.80% achieved under the proposed method selection procedure.

In conclusion, the empirical evidence supports the proposed method selection criterion, which combines information stemming from both recursive and rolling estimation windows depending on the underlying state of the economy. Both dynamic factor methods, FHLZ and FHLR greatly benefit from the our estimation window selection.

4.3.3 Statistical Forecast Accuracy and Portfolio Choice

Following the pioneering work of Leicht and Tanner (1991), and Pesaran and Timmermann (1995), we finally study the economic value of equity premium forecasts. Our interest lies in understanding the linkages between statistical and economic measures of forecasting performance. This is an open issue: Leicht and Tanner (1991), and Cenesizoglu and Timmermann (2012) only find weak relationships between statistical and economic measures of forecast accuracy; at the same time, Pesaran and Granger (2000) advocate a closer link between decision theory and the forecast evaluation problem.

In line with Campbell and Thompson (2008), Rapach *et al.* (2010), Ferreira and Santa Clara (2011), Rapach and Zhou (2013) and Neely *et al.* (2014), we economically evaluate the forecasts by computing the certainty equivalent return for a risk-averse investor with mean-variance preferences and relative risk aversion parameter γ. At the end of each month, the investor allocates her wealth between stocks and a riskless asset. The choice depends on a dynamic trading strategy based on a benchmark and an alternative prediction method. As customary in the literature, our benchmark is the prevailing mean estimated with recursive window: it assumes the equity premium has a constant mean and it is not predictable (Timmermann, 2008); it also produces the lowest mean squared prediction error out of all methods based on the prevailing mean (see Panel A in Table 5).

Formally, let $j = 0$ and $j = 1$ denote the benchmark and the alternative method, respectively. If the investor opts for method j at period $\tau + 1$, at period τ she assigns to stocks a share $w_{j\tau}$ equal to

$$w_{j\tau} = \frac{1}{\gamma} \frac{\hat{\rho}_{j,\tau+1}}{\hat{\sigma}^2_{j,\tau+1}}, \quad j = 0, 1,$$

where $\hat{\rho}_{j,\tau+1}$ and $\hat{\sigma}^2_{j,\tau+1}$ are the point forecasts of $\rho_{\tau+1}$ and of its variance $\sigma^2_{\tau+1}$ made at time τ, respectively: as in Campbell and Thompson (2008), we compute the latter as the five-year moving window of past monthly returns, so that $\hat{\sigma}^2_{j,\tau+1} = \hat{\sigma}^2_{\tau+1}$ is independent of the underlying forecasting method j. The realized return on the investment portfolio from method j at time $\tau + 1$ then is

$$R^p_{j,\tau+1} = w_{j\tau}\rho_{\tau+1} + r_{f\tau}, \quad \tau = T_e, \ldots, T-1, \quad j = 0, 1.$$

The certainty equivalent return is the average realized utility from method j over the out-of-sample period and it is defined as

$$\bar{U}_j = \bar{\mu}^p_j - \frac{1}{2}\gamma \left(\hat{\sigma}^p_j\right)^2, \quad j = 0, 1,$$

where $\bar{\mu}^p_j$ and $\left(\hat{\sigma}^p_j\right)^2$ are the sample mean and variance, respectively, of the portfolio returns $R^p_{\tau,t+1}$ over the out-of-sample period. Following Campbell and Thompson (2008), we constrain the portfolio weights $w_{0\tau}$ and $w_{1\tau}$ such that $0 \leq w_{0\tau}, w_{1\tau} \leq 1.5$. We then compute the utility gain

$$\Delta = \bar{U}_1 - \bar{U}_0 :$$

the utility gain represents the portfolio management performance fee that a mean-variance investor is willing to pay to switch from the dynamic trading strategy based on the benchmark to the one based on the alternative method (Fleming et *al.*, 2001). In the empirical application, we set $\gamma = 3, 4, 5, 10$ (Rapach *et al.*, 2010; Rapach and Zhou, 2013; Cenesizoglu and Timmermann, 2012); we multiply Δ by 1200 to express it in average annualized percentage returns.

The results are displayed in Panels C, D, E and F of Table 5 for $\gamma = 3, 4, 5, 10$, respectively. The recursive window estimation scheme leads to less accurate forecasts than those produced by the benchmark uniformly across models as all utility gains are negative. Interesting results arise under the rolling window scheme and the method selection criterion. The most accurate forecasts are produced by FHLR and FHLZ: the former prevails for $\gamma = 3, 4$ and $\gamma = 5$ under rolling window and method selection, respectively; the latter is preferable for $\gamma = 10$ under method selection. The empirical results indicate that the link between statistical and economic measures of forecast accuracy increases with risk aversion. An analysis of the correlation between out-of-sample R^2 and utility gains confirms this first impression: the correlation is equal to -0.03, 0.10, 0.19 and 0.46 for $\gamma = 3, 4, 5, 10$, respectively, and thus monotonically increases in γ. This empirical regularity is further illustrated in Figure 2, which plots utility gains against out-of-sample R^2 for the values of γ of interest.

Figure 2 about here

In conclusion, our results show that for the empirically relevant values $\gamma = 3, 4, 5$, FHLR produces the most accurate forecasts as evaluated in economic terms: rolling window estimation is preferable for $\gamma = 3, 4$, whereas the proposed method selection criterion takes the lead for $\gamma = 5$. In the extreme case $\gamma = 10$, the method selection criterion as applied to FHLZ should be employed. We further show that statistical and economic measures of forecast accuracy tend to be positively correlated (Cenesizoglu and Timmermann, 2012), and the strength of the correlation increases with γ.

5 Conclusions

We study one-step-ahead out-of-sample predictability of the monthly equity premium using large dimensional factor models. We compare the widely used static method of Stock and Watson (2002a,b) with two different estimators proposed byForni *et al.* (2005) and Forni *et al.* (2015, 2017) for the most general approach known as Generalized Dynamic Factor model. We show that large dimensional factor models condense the information contained in a high number of predictors to accurately forecast stock returns, especially when the Generalized Dynamic Factor model is considered; this holds true when both statistical and economic forecast evaluation criteria are employed. Further improvements are found applying a pseudo real-time combination of recursive and rolling forecasts, which we label as *method selection*.

Our work may be extended in several ways. Recently, Barigozzi and Hallin (2015, 2017) and Barigozzi *et al.* (2017) proposed a two-step Generalized Dynamic Factor model for volatilities which also accounts for the factor structure in returns. In that framework, it is worth exploring whether applying our method selection criterion on returns' factor decomposition is useful to deliver more accurate volatility predictions. Bond return predictability is also worth considering as the focus of future research (see Gargano *et al.*, 2017; and references therein).

References

[1] Anderson, B. D. O. and M. Deistler (2011). "Properties of zero-free transfer function matrices," *SICE Journal of Control, Measurement, and System Integration* 1, 284 − 292.

[2] Aruoba, S. B., F. X. Diebold and C. Scotti (2009), "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, 27 (4), 417 − 427.

[3] Baetje, F. and L. Menkhoff (2016), "Equity premium prediction: Are economic and technical indicators unstable?", *International Journal of Forecasting*, 32, 1193–1207.

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

[5] Baltagi, B. H., C. Kao and F. Wang (2017), "Identification and Estimation of a Large Factor Model with Structural Instability", *Journal of Econometrics*, 197 (1), 87–100.

[6] Banbura, M. D., D. Giannone and L. Reichlin (2011), "Nowcasting," in M. P. Clements and D. F. Hendry (eds.), *The Handbook of Economic Forecasting*, 193 − 224, Oxford: Oxford University Press.

[7] Barigozzi, M. and M. Hallin (2015), "Generalized dynamic factor models and volatilities: recovering the market volatility shocks," *Econometrics Journal* , 19(1), 33–60.

[8] Barigozzi, M. and M. Hallin (2017), “Generalized dynamic factor models and volatilities: estimation and forecasting,” *Journal of Econometrics*. forthcoming.

[9] Barigozzi, M., M. Hallin and Soccorsi, S. (2017), “Identification of Global and Local Shocks in International Financial Markets via General Dynamic Factor Models,” http://dx.doi.org/10.2139/ssrn.2937243

[10] Bekaert, G., E. Engstrom and Y. Xing (2009), ”Risk, Uncertainty and Asset Prices,” *Journal of Financial Economics*, 91 (1), 59 – 82.

[11] Boivin, J. and S. Ng (2006), ”Are More Data Always Better for Factor Analysis?,” *Journal of Econometrics*, 132 (1), 169 – 194.

[12] Bossaerts, P. and P. Hillion (1999), ”Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?,” *Review of Financial Studies*, 12 (2), 405 – 428.

[13] Breitung, J. and S. Eickmeier (2011), ”Testing for Structural Breaks in Dynamic Factor Models,” *Journal of Econometrics*, 163 (1), 71 – 84.

[14] Brillinger, D.R. (1981), *Time Series: Data Analysis and Theory*, San Francisco: Holden-Day.

[15] Chamberlain, G. and M. Rothschild (1983), ”Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets,” Econometrica, 51 (5), 1281 1304.

[16] Çakmaklı, C. and D. van Dijk, (2016), “Getting the most out of macroeconomic information for predicting excess stock returns”. *International Journal of Forecasting* , 32, 650–668.

[17] Campbell, J. Y. and J. H. Cochrane (1999), ”By Force or Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy*, 107 (2), 205 – 251.

[18] Campbell, J. Y. and S. B. Thompson (2008), ”Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?” *Review of Financial Studies*, 21 (4), 1509 – 1531.

[19] Clark, T. E. and K. D. West (2007), ”Approximately Normal Tests for Equal Predictive Accuracy in Nested Models,” *Journal of Econometrics*, 138 (1), 291 – 311.

[20] Clark, T. E. and M. W. McCracken (2009), ”Improving Forecast Accuracy by Combining Recursive and Rolling Forecasts,” *International Economic Review*, 50 (2), 363 – 395.

[21] Connor, G. and R. A. Korajczyk (1986), "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics*, 15 (3), 373 − 394.

[22] Connor, G. and R. A. Korajczyk (1988), "Risk and Return in an Equilibrium APT: Application of a New Test Methodology," *Journal of Financial Economics*, 21 (2), 255 − 289.

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

[24] Diebold, F. X. (2015), "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *Journal of Business and Economic Statistics*, 33 (1), 1 − 24.

[25] Diebold, F. X. and M. Nerlove (1989), "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model", *Journal of Applied Econometrics* , 4, 1–21.

[26] Engle, R. F. and M. W. Watson (1981), "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates," *Journal of the American Statistical Association*, 76 (376), 774 − 781.

[27] Ferreira, M. A. and P. Santa-Clara (2011), "Forecasting Stock Market Returns: The Sum of the Parts Is More than the Whole," *Journal of Financial Economics*, 100 (3), 514 − 537.

[28] Fleming, J., C. Kirby and B. Ostdiek (2001), "The Economic Value of Volatility Timing", *Journal of Finance*, 56 (1), 329 − 353.

[29] Forni, M. A. Giovannelli, M. Lippi and S. Soccorsi (2018), "Dynamic Factor Model with Infinite Dimensional Factor Space: Forecasting," *Journal of Applied Econometrics*, Forthcoming.

[30] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2000), "The Generalized Dynamic Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82 (4), 540 − 554.

[31] Forni, M., M. Hallin, M. Lippi and L. Reichlin (2004), "The Generalized Dynamic Factor Model: Consistency and Rates," *Journal of Econometrics*, 119 (2), 231 − 255.

[32] 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 (471), 830 − 840.

[33] Forni, M., M. Hallin, M. Lippi and P. Zaffaroni (2015), "Dynamic Factor Models with Infinite-Dimensional Factor Spaces: One-Sided Representations," *Journal of Econometrics*, 185 (2), 359 – 371.

[34] Forni, M., M. Hallin, M. Lippi and P. Zaffaroni (2017), "Dynamic Factor Models with Infinite-Dimensional Factor Space: Asymptotic analysis," *Journal of Econometrics*, 199, 74–92

[35] Forni, M. and M. Lippi (2001). "The generalized dynamic factor model: Representation theory, " *Econometric Theory* 17, 1113 – 1141.

[36] Gargano, A., D. Pettenuzzo and A. Timmermann (2017), "Bond Return Predictability: Economic Value and Links to the Macroeconomy," Management Science. https://doi.org/10.1287/mnsc.2017.2829

[37] Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in D. J. Aigner and A. S. Goldberger (eds.), Latent Variables in Socio-Economic Models, North-Holland.

[38] Giacomini, R. and H. White (2006), "Tests of Conditional Predictive Ability," *Econometrica*, 74 (6), 1545 – 1578.

[39] Giannone, D., M. Lenza, and Primiceri, G.E. (2017), Economic Predictions with Big Data: The Illusion of Sparsity (August 2017). CEPR Discussion Paper No. DP12256.

[40] Gonçalves, S., McCracken, M.W., and B. Perron (2017), "Tests of equal accuracy for nested models with estimated factors", *Journal of Econometrics*, 198, 231–252.

[41] Goyal, A. and I. Welch (2003), "Predicting the Equity Premium with Dividend Ratios," *Management Science*, 49 (5), 639 – 654.

[42] Granger, C. W. J. and M. H. Pesaran (2000), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, 19 (7), 426 – 459.

[43] Hallin, M. and M. Lippi (2013). "Factor models in high-dimensional time seriesâĂŞa time-domain approach," *Stochastic Processes and their Applications* 123, 2678 – 2695.

[44] Hallin, M. and R. Liška (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102 (48), 603 – 617.

[45] Hansen, P. R. and A. Timmermann (2012), ”Choice of Sample Split in Out-of-Sample Forecast Evaluation,” Working Paper, European University Institute and University of California, San Diego.

[46] Henkel, S. J., J. S. Martin and F. Nardari (2011), ”Time-Varying Short-Horizon Predictability,” *Journal of Financial Economics*, 99 (3), 560 − 580.

[47] Jones, C. S. (2001), “Extracting factors from heteroskedastic asset returns”, *Journal of Financial Economics*, 62, 293–325.

[48] Leicht, G. and J. E. Tanner (1991), ”Economic Forecast Evaluation: Profits Versus the Conventional Error Measures”, *American Economic Review*, 81 (3), 580 − 590.

[49] Lo, A. W. (2004), ”The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective,” *Journal of Portfolio Management*, 30, 15 − 29.

[50] Ludvigson, S. C. and S. Ng (2009), “The empirical risk-return relation: A factor analysis approach”, *Journal of Financial Economics* , 83, 171–222.

[51] Menzly, L., T. Santos and P. Veronesi (2004), ”Understanding Predictability,” *Journal of Political Economy*, 112 (1), 1 − 47.

[52] McCracken, M. W. and S. Ng (2016), ”FRED-MD: A Monthly Database for Macroeconomic Research,” *Journal of Business and Economics Statistics*, 34(4), 574 – 589.

[53] Ng, V., R. F. Engle, and M. Rothschild (1992). “A multi-dynamic-factor model for stock returns”, *Journal of Econometrics*, 52, 245–266.

[54] Neely, C. J., D. E. Rapach, J. Tu and G. Zhou (2014), ”Forecasting the Equity Risk Premium: The Role of Technical Indicators,” *Management Science*, 60 (7) : 1772 − 1791.

[55] Ohno, S. and T. Ando (2014), ”Stock Return Predictability: A Factor-Augmented Predictive Regression System with Shrinkage Method”, *Econometric Reviews* 0, 1 − 43.

[56] Paye, B. S. and A. Timmermann (2006), ”Instability of Return Prediction Models,” *Journal of Empirical Finance*, 13 (3), 274 − 315.

[57] Pesaran, M. H. and A. Timmermann (1995), ”Predictability of Stock Returns: Robustness and Economic Significance,” *Journal of Finance*, 50 (4), 1201 − 1228.

[58] Pesaran, M. H. and A. Timmermann (2005), ”Real-Time Econometrics,” *Econometric Theory*, 21 (1), 212 – 231.

[59] Pesaran , M. H., and A. Timmermann (2007), “Selection of Estimation Window in the Presence of Breaks,” *Journal of Econometrics*, 137(1) , 134–61.

[60] Pettenuzzo, D., A. Timmermann and R. Valkanov (2014), ”Forecasting Stock Returns under Economic Constraints,” *Journal of Financial Economics*, 114 (3), 517 – 553.

[61] Rapach, D. E., J. K. Strauss and G. Zhou (2010), ”Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy,” *Review of Financial Studies*, 23 (2), 821 – 862.

[62] Rapach, D. E. and M. E. Wohar, ”Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns,” *Journal of Financial Econometrics*, 4 (2), 238 – 274.

[63] Rapach, D. E. and G. Zhou (2013), ”Forecasting Stock Returns,” in G. Elliott and A. Timmermann (eds.), Handbook of Economic Forecasting, Volume 2, Part A, 329 – 383, Amsterdam: North-Holland.

[64] Sargent, T. (1989), ”Two Models of Measurements and the Investment Accelerator,” *Journal of Political Economy*, 97 (2), 251 – 287.

[65] Sargent T. and C. A. Sims (1977), ” Business Cycle Modelling without Pretending to Have too Much a Priori Economic Theory,” in C. A. Sims (eds.), New Methods in Business Research, Federal Reserve Bank of Minneapolis.

[66] Soccorsi, S. (2016), “Measuring nonfundamentalness for structural VARs,” *Journal of Economic Dynamics and Control* 71, 86 – 101.

[67] Stock, J. H. and M. W. Watson (1989), ”New Indexes of Coincident and Leading Economic Indicators,” in O. J. Blanchard and S. Fischer (eds.), NBER Macroeconomics Annual (1989), vol. 4, MIT Press, 351 – 409.

[68] Stock, J. H. and M. W. Watson (2002*a*), ”Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97 (460), 1167 – 1179.

[69] Stock, J. H. and M. W. Watson (2002*b*). "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business Economic Statistics* 20, 147 – 162.

[70] Stock, J. H. and M. W. Watson (2012), "Generalized Shrinkage Methods for Forecasting Using Many Predictors," *Journal of Business and Economic Statistics*, 30 (4), 481 − 493.

[71] Timmermann, A. (2008), "Elusive Return Predictability," *International Journal of Forecasting*, 24 (1), 1 − 18.

[72] Timmermann, A. and C. W. J. Granger (2004), "Efficient Market Hypothesis and Forecasting," *International Journal of Forecasting*, 20 (1), 15 − 27.

[73] Watson, M. W. and R. F. Engle (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression models," *Journal of Econometrics*, 23 (3), 385 − 400.

[74] Welch, I. and A. Goyal (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21 (4), 1455 − 1508.

Table 1: Summary Statistics, 1960 - 2014

Notes. This table reports summary statistics for excess returns, calculated as continuously compounded returns on the S&P 500 index in excess of the treasury bill rate, and for the following 14 predictors proposed in Welch and Goyal (2008): log dividend-price ratio (log (DP)), log dividend-yield (log (DY)), log earnings-price ratio (log (EP)), log dividend-payout ratio (log (DE)), stock variance (SVAR), book-to-market ratio (BM), net equity expansion (NTIS), treasury bill rate (TBL), long-term yield (LTY), long-term return (LTR), term spread (TMS), default yield spread (DFY), default return spread (DFR), inflation (INFL). The sample period is $1960 - 2014$.

Variables	Mean	Std. Deviation	Skewness	Kurtosis
Excess Returns	0.001	0.043	-0.693	5.447
log (DP)	-3.577	0.397	-0.305	2.350
log (DY)	-3.571	0.397	-0.312	2.379
log (EP)	-2.829	0.432	-0.760	6.120
log (DE)	-0.748	0.317	2.667	17.390
SVAR	0.002	0.004	10.162	136.180
BM	0.509	0.259	0.690	2.646
NTIS	0.012	0.019	-0.883	3.984
TBL	0.049	0.031	0.698	4.036
LTY	0.067	0.026	0.804	3.264
LTR	0.006	0.029	0.425	5.827
TMS	0.018	0.015	-0.289	2.732
DFY	0.010	0.005	1.739	7.156
DFR	0.000	0.014	-0.366	9.763
INFL	0.003	0.003	0.008	6.693

Table 2: Out-Of-Sample Forecast Performance, Recursive Window, 1970 - 2014

Notes. This table shows the out-of-sample R^2 (%) of prediction models for the monthly excess return $\rho_{\tau+1}$ using the recursive window estimation scheme. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are estimated from the Welch and Goyal (2008) (WG) dataset in Panel A. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015) in Panel B, and from 25, 50 and 75 series selected with a LASSO estimator in Panels C, D and E, respectively. Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. The sample period is $1970 - 2014$.

Panel A: Welch and Goyal (2008) Data											
SW				FHLR				FHLZ			
Model	Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-1.33	-0.69	-2.73	$q=1$ $r=1$	-0.94	-0.48	-1.94	$q=1$	0.95***	0.75**	1.39**
$r=2$	-0.20*	-0.07**	-0.56	$r=2$	-2.11	-0.35	-5.98				
$r=3$	-2.98	-1.93*	-5.29	$r=3$	-2.68	0.08**	-8.74				
				$q=2$ $r=2$	-0.13	0.07	-0.57	$q=2$	0.07	0.38	-0.62
				$r=3$	-1.00	-0.55	-1.98				

Panel B: FRED-MD Data											
SW				FHLR				FHLZ			
Model	Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-0.79	0.25	-3.06	$q=1$ $r=1$	-0.80	-0.04	-2.47	$q=1$	0.61*	0.42*	1.03
$r=2$	0.02**	0.11***	-0.18	$r=2$	1.44***	0.84**	2.74*				
$r=3$	-0.87**	-0.83**	-0.93	$r=3$	1.43***	0.76**	2.91*				
$r=4$	-0.72**	-1.05**	-0.00	$r=4$	0.95***	0.51**	1.91				
$r=5$	-1.59**	-2.11*	-0.45	$r=5$	0.54**	0.28**	1.12				
$r=10$	-3.62*	-4.61	-1.44	$r=10$	0.80***	-0.73	4.17**				
$r=15$	-5.17*	-5.65	-4.13	$r=15$	-0.83*	-1.88	1.48				
				$q=2$ $r=2$	0.74***	0.55***	1.13	$q=2$	0.53*	0.47*	0.66
				$r=3$	0.78***	0.11**	2.27*				
				$r=4$	0.06**	-0.61**	1.56*				
				$r=5$	-0.27**	-0.81**	0.93				
				$r=10$	-1.52**	-2.60	0.87				
				$r=15$	-2.02*	-2.73	-0.44				
				$q=3$ $r=3$	0.44***	-0.21**	1.88*	$q=3$	0.51*	0.38*	0.79
				$r=4$	-0.144**	-0.67**	1.01				
				$r=5$	-0.29**	-1.07**	1.43*				
				$r=10$	-2.05*	-3.67	1.52*				
				$r=15$	-2.35*	-3.15	-0.61				
				$q=4$ $r=4$	-0.17**	-0.49**	0.53	$q=4$	0.75**	0.56**	1.15
				$r=5$	-0.46**	-1.18**	1.12*				
				$r=10$	-2.28*	-3.81	1.08*				
				$r=15$	-3.25	-3.88	-1.86				
				$q=5$ $r=5$	-0.31**	-1.13**	1.48	$q=5$	0.98**	0.72**	1.55
				$r=10$	-1.22**	-3.19	3.11**				
				$r=15$	-3.17	-4.00	-1.34				

Table 2-*Continued*: Out-Of-Sample Forecast Performance, Recursive Window, 1970 - 2014

Panel C: FRED-MD Data, LASSO, 25												
SW				FHLR					FHLZ			
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-0.75*	-0.27*	-2.23	$q=1$	$r=1$	-0.48	0.08*	-1.72	$q=1$	-0.16	0.15	-0.85
$r=2$	-1.42	-0.95*	-3.00		$r=2$	-1.17	-1.02	-1.52				
$r=3$	-2.33*	-3.79	0.94		$r=3$	-1.70	-1.22	-2.77				
$r=4$	-3.50	-4.23	-1.92		$r=4$	-2.62	-2.42	-3.06				
$r=5$	-3.02*	-3.48	-1.60		$r=5$	-1.58**	-1.76*	-1.19				
$r=10$	-7.17	-7.31	-7.28		$r=10$	-2.60	-3.02	-1.67				
$r=15$	-7.90	-9.16	-6.03		$r=15$	-2.17	-3.37	0.46*				
				$q=2$	$r=2$	-0.43*	-0.21**	-0.91	$q=2$	-0.33	-0.51	0.08
					$r=3$	-0.70**	-1.55*	1.17				
					$r=4$	-2.15*	-2.59	-1.18				
					$r=5$	-1.93*	-2.70	-0.24				
					$r=10$	-3.95	-4.93	-1.80				
					$r=15$	-4.92	-6.23	-1.97				
				$q=3$	$r=3$	-0.50**	-1.36*	1.40	$q=3$	-0.81	-1.20	0.04
					$r=4$	-2.11*	-2.23	-1.85				
					$r=5$	-2.52	-2.09*	-3.47				
					$r=10$	-5.48	-5.06	-6.39				
					$r=15$	-6.13	-6.72	-4.82				
				$q=4$	$r=4$	-2.89	-2.81	-3.07	$q=4$	-1.09	-1.69	0.22
					$r=5$	-3.29	-2.56	-4.92				
					$r=10$	-4.99	-4.88	-5.25				
					$r=15$	-5.30	-6.46	-2.75				
				$q=5$	$r=5$	-2.55	-2.71	-2.21	$q=5$	-1.95	-3.00	0.34
					$r=10$	-3.90	-4.15	-3.33				
					$r=15$	-5.93	-6.47	-4.72				

Table 2-*Continued*: Out-Of-Sample Forecast Performance, Recursive Window, 1970 - 2014

Panel D: FRED-MD Data, LASSO, 50											
SW				FHLR				FHLZ			
Model	Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-0.75	0.16	-2.75	$q=1$ $r=1$	-0.36	0.48**	-2.21	$q=1$	-0.91	-0.19	-2.48
$r=2$	-1.28	-1.51	-0.78	$r=2$	-1.56	-0.15*	-4.67				
$r=3$	-1.68**	-2.13**	-0.71	$r=3$	-0.39*	-0.39*	-0.39				
$r=4$	-2.57*	-2.83*	-2.00	$r=4$	-1.73	-0.92*	-3.51				
$r=5$	-2.78*	-3.36	-1.49	$r=5$	-1.20*	-1.25	-1.08				
$r=10$	-5.80	-5.27	-6.96	$r=10$	-1.90	-2.25	-1.14				
$r=15$	-6.90	-7.10	-6.45	$r=15$	-1.47	-2.02	-0.27				
				$q=2$ $r=2$	-0.85*	-0.87*	-0.82	$q=2$	-0.16	0.23	-0.95
				$r=3$	-0.92**	-1.74*	0.87				
				$r=4$	-1.29*	-1.80*	-0.19				
				$r=5$	-2.76	-2.38	-3.61				
				$r=10$	-4.09	-5.06	-1.96				
				$r=15$	-3.44	-3.80	-2.63				
				$q=3$ $r=3$	-0.82**	-1.27*	0.18	$q=3$	-0.65	-0.29	-1.41
				$r=4$	-1.06*	-0.81**	-1.61				
				$r=5$	-2.33	-1.91	-3.25				
				$r=10$	-3.79	-3.96	-3.41				
				$r=15$	-4.02	-5.58	-0.59				
				$q=4$ $r=4$	-1.38*	-1.18*	-1.81	$q=4$	-0.46	-0.53	-0.20
				$r=5$	-2.30	-1.79	-3.43				
				$r=10$	-2.93	-3.21	-2.33				
				$r=15$	-4.43	-7.32	1.95*				
				$q=5$ $r=5$	-1.79*	-1.67	-2.06	$q=5$	0.13*	-0.72	2.14*
				$r=10$	-2.55	-2.44	-2.80				
				$r=15$	-3.22*	-5.62	2.07*				

Table 2-*Continued*: Out-Of-Sample Forecast Performance, Recursive Window, 1970 - 2014

Panel E: FRED-MD Data, LASSO, 75												
SW				FHLR					FHLZ			
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r = 1$	-0.71	0.67**	-3.74	$q = 1$	$r = 1$	-0.62	0.72**	-3.57	$q = 1$	1.08**	0.65**	2.00
$r = 2$	-2.63	-0.85*	-6.54		$r = 2$	-0.65	0.31**	-2.77				
$r = 3$	-1.12**	-1.38*	-0.55		$r = 3$	-0.51	-0.25*	-1.06				
$r = 4$	-1.51**	-1.49**	-1.54		$r = 4$	-0.86	-0.25*	-2.18				
$r = 5$	-1.96*	-1.51**	-2.95		$r = 5$	-0.04**	0.19**	-0.54				
$r = 10$	-4.78	-4.39	-5.63		$r = 10$	-0.81*	-0.95*	-0.50				
$r = 15$	-6.38	-6.25	-6.66		$r = 15$	-1.17	-0.87*	-1.83				
				$q = 2$	$r = 2$	-0.95	0.11**	-3.29	$q = 2$	0.26	0.33*	0.09
					$r = 3$	0.10**	-0.29**	0.98				
					$r = 4$	-0.01**	-0.14**	0.28				
					$r = 5$	-0.64**	-0.34**	-1.29				
					$r = 10$	-0.81**	-1.51**	0.73				
					$r = 15$	-0.82**	-1.71**	1.15				
				$q = 3$	$r = 3$	-0.12**	-0.42**	0.55	$q = 3$	-0.08	0.32*	-0.97
					$r = 4$	-0.60**	-0.46**	-0.90				
					$r = 5$	-0.77**	-0.72**	-0.88				
					$r = 10$	-1.62*	-2.78	0.93				
					$r = 15$	-2.63*	-4.22	0.86				
				$q = 4$	$r = 4$	-0.50**	-0.14**	-1.29	$q = 4$	-0.20	0.05	-0.76
					$r = 5$	-1.05*	-0.72**	-1.78				
					$r = 10$	-2.32*	-2.56	-1.78				
					$r = 15$	-3.77	-4.32	-2.57				
				$q = 5$	$r = 5$	-1.15*	-0.86**	-1.79	$q = 5$	-0.37	-0.16	-0.81
					$r = 10$	-1.77*	-2.91	0.73				
					$r = 15$	-3.81	-5.65	0.24				

Table 3: Out-Of-Sample Forecast Performance, Rolling Window, 1970 - 2014

Notes. This table shows the out-of-sample R^2 (%) of prediction models for the monthly excess return $\rho_{\tau+1}$ using the rolling window estimation scheme. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are estimated from the Welch and Goyal (2008) (WG) dataset in Panel A. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015) in Panel B, and from 25, 50 and 75 series selected with a LASSO estimator in Panels C, D and E, respectively. Statistical significance is assessed through the Clarke and West (2007) CW statistic: *, ** and *** denote significance at 10%, 5% and 1%, respectively. BN and HL denote Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The sample period is 1970 − 2014.

Panel A: Welch and Goyal (2008) Data												
SW				FHLR					FHLZ			
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-3.73	-1.89	-7.84	$q=1$	$r=1$	-3.56	-1.74	-7.60	$q=1$	-0.07	-0.19	0.18
$r=2$	-3.22	-3.11	-3.46		$r=2$	-2.42	-1.88	-3.62				
$r=3$	-4.10*	-3.03**	-6.49		$r=3$	-4.56	-2.60	-8.94				
				$q=2$	$r=2$	-3.54	-3.71	-3.15	$q=2$	-0.69	0.54*	-3.43
					$r=3$	-4.10	-2.80	-7.01				

Panel B: FRED-MD Data												
SW				FHLR					FHLZ			
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-1.05	0.05*	-3.51	$q=1$	$r=1$	-1.83	-0.28	-5.10	$q=1$	1.17**	-0.55	5.12**
$r=2$	-1.78*	-2.16*	-0.92		$r=2$	0.12**	-0.48*	1.01				
$r=3$	-2.31**	-3.19*	-0.37*		$r=3$	0.69**	-0.84*	3.66**				
$r=4$	-3.81**	-5.35*	-0.38*		$r=4$	0.90***	-0.99**	4.41**				
$r=5$	-3.40***	-5.44*	1.14**		$r=5$	0.21**	-2.76	6.21**				
$r=10$	-11.74	-12.88	-9.21*		$r=10$	-2.14*	-5.23	4.32**				
$r=15$	-20.54	-22.01	-17.24		$r=15$	-4.46	-8.05	3.09*				
				$q=2$	$r=2$	0.93***	-0.10**	2.77*	$q=2$	1.63**	-0.42	3.75*
					$r=3$	0.43***	-0.65**	2.37*				
					$r=4$	0.15**	-1.81*	3.90**				
					$r=5$	-0.46**	-2.82	4.13**				
					$r=10$	-1.14**	-5.30	7.66***				
					$r=15$	-4.88	-7.43	0.19				
				$q=3$	$r=3$	0.43***	-1.00**	3.11**	$q=3$	1.34**	-0.45	3.88*
					$r=4$	0.01**	-2.26*	4.42**				
					$r=5$	-0.24**	-2.98	5.19**				
					$r=10$	-2.37**	-6.35	5.59**				
					$r=15$	-5.47	-7.58	-1.70				
				$q=4$	$r=4$	0.27***	-2.37	5.53**	$q=4$	1.75**	-0.37	5.22*
					$r=5$	0.05**	-2.89	5.97**				
					$r=10$	-2.47**	-6.43	5.52**				
					$r=15$	-4.32*	-7.51	2.00*				
				$q=5$	$r=5$	-0.51**	-3.26	5.00**	$q=5$	2.02***	-0.15	5.31**
					$r=10$	-3.21**	-7.25	4.95**				
					$r=15$	-6.34	-9.56	-0.12				
BN	-4.59**	-6.11	-1.19*	HL, BN		-0.09***	-2.75	5.36**	HL	1.71**	-0.34	4.85*

Table 3-*Continued*: Out-Of-Sample Forecast Performance, Rolling Window, 1970 - 2014

Panel C: FRED-MD Data, LASSO, 25													
SW				FHLR					FHLZ				
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-2.14	-2.40	-1.57	$q=1$	$r=1$	-1.17	-1.82	0.25	$q=1$	-1.13	-1.80	0.35
$r=2$	-2.93	-5.20	2.05*		$r=2$	-2.12	-3.99	2.01				
$r=3$	-2.93	-5.20	2.05*		$r=3$	-3.45	-5.09	0.18				
$r=4$	-7.02	-6.47	-8.22		$r=4$	-3.94	-6.30	1.23				
$r=5$	-8.75	-8.07	-10.25		$r=5$	-3.08	-6.19	3.78*				
$r=10$	-16.87	-18.29	-13.75		$r=10$	-2.96*	-7.64	7.34**				
$r=15$	-27.40	-31.24	-18.96		$r=15$	-3.06*	-8.50	8.91***				
				$q=2$	$r=2$	-1.91	-3.42	1.40*	$q=2$	-0.58	-2.22	3.02*
					$r=3$	-3.54	-4.96	-0.41				
					$r=4$	-2.09	-3.80	1.69*				
					$r=5$	-4.52	-6.62	0.09*				
					$r=10$	-5.35	-8.84	2.33*				
					$r=15$	-5.75*	-10.57	4.87**				
				$q=3$	$r=3$	-2.44*	-4.06	1.14*	$q=3$	-0.62	-2.84	4.26*
					$r=4$	-2.07*	-3.77	1.68*				
					$r=5$	-2.98*	-5.18	1.86*				
					$r=10$	-5.73*	-7.84	-1.07				
					$r=15$	-8.78	-14.69	4.22**				
				$q=4$	$r=4$	-2.82	-4.40	0.65*	$q=4$	-0.68	-3.26	4.99*
					$r=5$	-3.78	-5.83	0.73*				
					$r=10$	-6.75	-8.46	-3.00				
					$r=15$	-10.89	-16.06	0.48*				
				$q=5$	$r=5$	-1.911	-3.88	2.42*	$q=5$	-1.33	-3.31	3.03*
					$r=10$	-6.22	-7.81	-2.72				
					$r=15$	-10.94	-16.00	0.20				

Table 3-*Continued*: Out-Of-Sample Forecast Performance, Rolling Window, 1970 - 2014

Panel D: FRED-MD Data, LASSO, 50													
SW				FHLR					FHLZ				
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-2.58	-1.55	-4.85	$q=1$	$r=1$	-1.46	-0.78	-2.96	$q=1$	0.15	-1.23	3.18**
$r=2$	-4.75	-6.39	-1.13		$r=2$	-1.81	-2.46	-0.36				
$r=3$	-4.56	-6.50	-0.30*		$r=3$	-2.47	-3.58	-0.01				
$r=4$	-7.41	-9.17	-3.53*		$r=4$	-3.30	-4.49	-0.70				
$r=5$	-8.32	-8.85	-7.14		$r=5$	-3.73	-4.84	-1.29				
$r=10$	-13.98	-14.69	-12.43		$r=10$	-4.19	-6.26	0.36				
$r=15$	-20.06	-22.09	-15.57		$r=15$	-4.35	-7.66	2.94				
				$q=2$	$r=2$	-1.69	-3.80	2.96*	$q=2$	1.28**	-0.64	5.52**
					$r=3$	-1.37**	-3.85	4.09**				
					$r=4$	-2.21*	-4.83	3.54**				
					$r=5$	-3.02*	-4.42	0.06				
					$r=10$	-4.12*	-6.70	1.56*				
					$r=15$	-4.97*	-8.28	2.32*				
				$q=3$	$r=3$	-2.07*	-4.29	2.83*	$q=3$	0.84*	-1.08	5.04**
					$r=4$	-2.25*	-3.92	1.41*				
					$r=5$	-1.72**	-2.92	0.92*				
					$r=10$	-3.84**	-6.07	1.06*				
					$r=15$	-5.82*	-9.47	2.20*				
				$q=4$	$r=4$	-1.99**	-4.22	2.91**	$q=4$	0.28*	-1.91	5.12**
					$r=5$	-1.61**	-2.89	1.20*				
					$r=10$	-4.24*	-6.09	-0.17				
					$r=15$	-3.92**	-7.05*	2.96*				
				$q=5$	$r=5$	-1.87**	-3.30	1.27*	$q=5$	0.32*	-1.86	5.14**
					$r=10$	-4.90*	-6.72	-0.88				
					$r=15$	-5.22*	-9.11	3.34*				

Table 3-*Continued*: Out-Of-Sample Forecast Performance, Rolling Window, 1970 - 2014

Panel E: FRED-MD Data, LASSO, 75												
SW				FHLR					FHLZ			
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	-0.61*	-0.86	-0.06	$q=1$	$r=1$	-0.89	-0.72	-1.26	$q=1$	0.39	-0.85	3.12*
$r=2$	-5.46	-5.98	-4.31		$r=2$	-2.45	-2.18	-3.03				
$r=3$	-4.57*	-6.50	-0.33*		$r=3$	-2.42	-3.01	-1.10				
$r=4$	-5.18*	-7.72	0.42*		$r=4$	-3.13	-3.53	-2.27				
$r=5$	-5.35*	-7.09	-1.50*		$r=5$	-2.72	-3.35	-1.31				
$r=10$	-12.34	-15.80	-4.70*		$r=10$	-2.72	-4.95	2.17				
$r=15$	-18.56	-21.58	-11.92		$r=15$	-2.73	-5.23	2.78*				
				$q=2$	$r=2$	-1.52	-2.10	-0.26	$q=2$	0.44*	-0.98	3.55*
					$r=3$	-1.11**	-2.35	1.62*				
					$r=4$	-1.82*	-3.06	0.90				
					$r=5$	-1.01**	-3.71	4.94**				
					$r=10$	-1.35**	-5.42	7.62*				
					$r=15$	-2.64**	-7.11	7.19**				
				$q=3$	$r=3$	-0.90**	-2.09	1.71*	$q=3$	0.55*	-0.81	3.56*
					$r=4$	-1.21**	-2.89	2.50*				
					$r=5$	0.03**	-3.13	6.98**				
					$r=10$	-1.11**	-4.95	7.34*				
					$r=15$	-2.49**	-5.82	4.82*				
				$q=4$	$r=4$	-1.37*	-2.82	1.83*	$q=4$	0.38*	-0.96	3.34*
					$r=5$	-0.27**	-3.08	5.92**				
					$r=10$	-1.15**	-5.00	7.30**				
					$r=15$	-3.29**	-6.98	4.83**				
				$q=5$	$r=5$	0.31**	-2.94	7.46**	$q=5$	0.57*	-1.17	4.41*
					$r=10$	-2.62*	-7.07	7.18**				
					$r=15$	-5.17	-9.55	4.45*				

Table 4: Resume of small to medium LASSO Results, FRED-MD Data, MSPE, 1970 - 2014

Notes. This table shows the mean squared prediction error (MSPE ×100) of prediction models for the monthly excess return $\rho_{\tau+1}$ using recursive window (Panel A) and rolling window (Panel B) estimation schemes. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are extracted from 25, 50 and 75 series selected at each point in time with a LASSO estimator applied to the FRED-MD dataset detailed in MacCracken and Ng (2015). The sample period is 1970 − 2014.

		LASSO, 25	LASSO, 50	LASSO, 75	full FRED-MD dataset
	SW	−0.75	−0.75	−0.71	0.02
Panel A: Recursive Window	FHLR	−0.16	−0.36	0.10	1.44
	FHLZ	−0.50	0.13	1.08	0.98
	SW	−2.14	−2.58	−0.61	−1.05
Panel B: Rolling Window	FHLR	−1.17	−1.37	0.31	0.93
	FHLZ	−0.58	1.28	0.57	2.02

Table 5: Out-Of-Sample Results, FRED-MD Data, MSPE, 1970 - 2014

Notes. This table shows the mean squared prediction error (MSPE ×100) of prediction models for the monthly excess return $\rho_{\tau+1}$ using recursive window (Panel A) and rolling window (Panel B) estimation schemes. The factor models are Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ), described in Section 2. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015). BN and HL denote Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The sample period is $1970 - 2014$.

Panel A: Recursive Window												
SW				FHLR					FHLZ			
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r=1$	0.2019	0.1623	0.4196	$q=1$	$r=1$	0.2019	0.1628	0.4172	$q=1$	0.1991	0.1621	0.4030
$r=2$	0.2003	0.1626	0.4079		$r=2$	0.1974	0.1614	0.3960				
$r=3$	0.2020	0.1641	0.4109		$r=3$	0.1975	0.1615	0.3953				
$r=4$	0.2018	0.1645	0.4071		$r=4$	0.1984	0.1619	0.3994				
$r=5$	0.2035	0.1662	0.4089		$r=5$	0.1992	0.1623	0.4026				
$r=10$	0.2076	0.1703	0.4130		$r=10$	0.1987	0.1639	0.3902				
$r=15$	0.2107	0.1720	0.4239		$r=15$	0.2020	0.1658	0.4011				
				$q=2$	$r=2$	0.1988	0.1618	0.4025	$q=2$	0.1993	0.1620	0.4044
					$r=3$	0.1987	0.1626	0.3979				
					$r=4$	0.2002	0.1638	0.4008				
					$r=5$	0.2008	0.1641	0.4033				
					$r=10$	0.2033	0.1670	0.4036				
					$r=15$	0.2044	0.1672	0.4089				
				$q=3$	$r=3$	0.1994	0.1631	0.3995	$q=3$	0.1993	0.1621	0.4039
					$r=4$	0.2006	0.1638	0.4030				
					$r=5$	0.2009	0.1645	0.4013				
					$r=10$	0.2044	0.1687	0.4010				
					$r=15$	0.2050	0.1679	0.4096				
				$q=4$	$r=4$	0.2007	0.1635	0.4050	$q=4$	0.1988	0.1618	0.4024
					$r=5$	0.2012	0.1647	0.4026				
					$r=10$	0.2049	0.1690	0.4027				
					$r=15$	0.2068	0.1691	0.4147				
				$q=5$	$r=5$	0.2009	0.1646	0.4011	$q=5$	0.1984	0.1616	0.4008
					$r=10$	0.2028	0.1679	0.3945				
					$r=15$	0.2067	0.1693	0.4126				

Table 5-*Continued*: Out-Of-Sample Results, FRED-MD Data, MSPE, 1970 - 2014

Panel B: Rolling Window													
SW				FHLR					FHLZ				
Model	Full Sample	Expansions	Recessions	Model		Full Sample	Expansions	Recessions	Model	Full Sample	Expansions	Recessions
$r = 1$	0.2033	0.1639	0.4199	$q = 1$	$r = 1$	0.2048	0.1642	0.4218	$q = 1$	0.1988	0.1634	0.3822
$r = 2$	0.2047	0.1675	0.4094		$r = 2$	0.2009	0.1646	0.3973				
$r = 3$	0.2058	0.1692	0.4071		$r = 3$	0.1997	0.1652	0.3867				
$r = 4$	0.2088	0.1728	0.4072		$r = 4$	0.1993	0.1654	0.3837				
$r = 5$	0.2080	0.1729	0.4010		$r = 5$	0.2007	0.1683	0.3765				
$r = 10$	0.2248	0.1851	0.4430		$r = 10$	0.2055	0.1724	0.3840				
$r = 15$	0.2424	0.2001	0.4756		$r = 15$	0.2101	0.1770	0.3890				
				$q = 2$	$r = 2$	0.1993	0.1640	0.3903	$q = 2$	0.1979	0.1632	0.3878
					$r = 3$	0.2003	0.1647	0.3918				
					$r = 4$	0.2008	0.1667	0.3857				
					$r = 5$	0.2021	0.1684	0.3848				
					$r = 10$	0.2034	0.1725	0.3706				
					$r = 15$	0.2110	0.1760	0.4006				
				$q = 3$	$r = 3$	0.2003	0.1654	0.3889	$q = 3$	0.1984	0.1632	0.3872
					$r = 4$	0.2011	0.1675	0.3836				
					$r = 5$	0.2016	0.1687	0.3805				
					$r = 10$	0.2059	0.1742	0.3789				
					$r = 15$	0.2121	0.1762	0.4089				
				$q = 4$	$r = 4$	0.2006	0.1677	0.3792	$q = 4$	0.1976	0.1631	0.3818
					$r = 5$	0.2010	0.1685	0.3774				
					$r = 10$	0.2061	0.1743	0.3792				
					$r = 15$	0.2098	0.1761	0.3934				
				$q = 5$	$r = 5$	0.2022	0.1691	0.3813	$q = 5$	0.1971	0.1627	0.3815
					$r = 10$	0.2076	0.1757	0.3815				
					$r = 15$	0.2139	0.1795	0.4018				
BN	0.2104	0.1740	0.4105	HL, BN		0.2013	0.1683	0.3798	HL	0.1977	0.1630	0.3833

Table 6: Out-Of-Sample Forecast Performance, Method Selection Criterion, 1970 - 2014

Notes. This table shows mean squared prediction error (MSPE ×100), out-of-sample R^2 (%) and annualized utility gain (%) of prediction models for the monthly excess return $\rho_{\tau+1}$. The prevailing mean PM is defined in (11). Stock and Watson (2002*a*) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ) factor models are described in Section 2. Factors are extracted from the FRED-MD dataset detailed in MacCracken and Ng (2015). Under recursive window, $r = 5$ and $q = 4$ static and dynamic factors, respectively, are selected as detailed in Section 4.2.1. Under rolling window, the number of static and dynamic factors is selected according to Bai and Ng (2002), and Hallin and Liška (2007) model selection criteria, respectively. The method selection criterion allows to switch between estimation windows using the procedure described in Section 4.3.2. The sample period is $1970 - 2014$.

Panel A: MSPE											
Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
0.2003	0.2035	0.2012	0.1988	0.2011	0.2098	0.2013	0.1977	0.2007	0.2054	0.1981	0.1967

Panel B: Out-Of-Sample R^2 (%)											
Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.60	-0.46	0.75	-0.40	-4.74	-0.50	1.31	-0.20	-2.55	1.11	1.80

Panel C: Portfolio Choice, Δ (ann.) (%), $\gamma = 3$											
Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.18	-1.44	-2.01	0.72	0.42	1.54	0.79	-0.05	0.56	0.96	0.22

Panel D: Portfolio Choice, Δ (ann.) (%), $\gamma = 4$											
Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-1.89	-1.97	-1.93	0.37	-0.05	0.74	0.45	-0.04	0.25	0.69	0.32

Panel E: Portfolio Choice, Δ (ann.) (%), $\gamma = 5$											
Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-2.25	-2.33	-1.71	0.08	-0.48	0.14	0.26	-0.03	0.26	0.52	0.34

Panel F: Portfolio Choice, Δ (ann.) (%), $\gamma = 10$											
Recursive Window				Rolling Window				Method Selection			
PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ	PM	SW	FHLR	FHLZ
-	-2.83	-2.47	-0.85	0.00	-1.63	-1.19	-0.10	-0.01	-0.10	0.18	0.19

Figure 1: Estimated Number of Factors, Recursive Window Estimation Scheme

This figure plots the number of static and dynamic factors selected in pseudo real-time under the recursive window estimation scheme by Bai and Ng (2002), and Hallin and Liška (2007) information criteria, respectively.

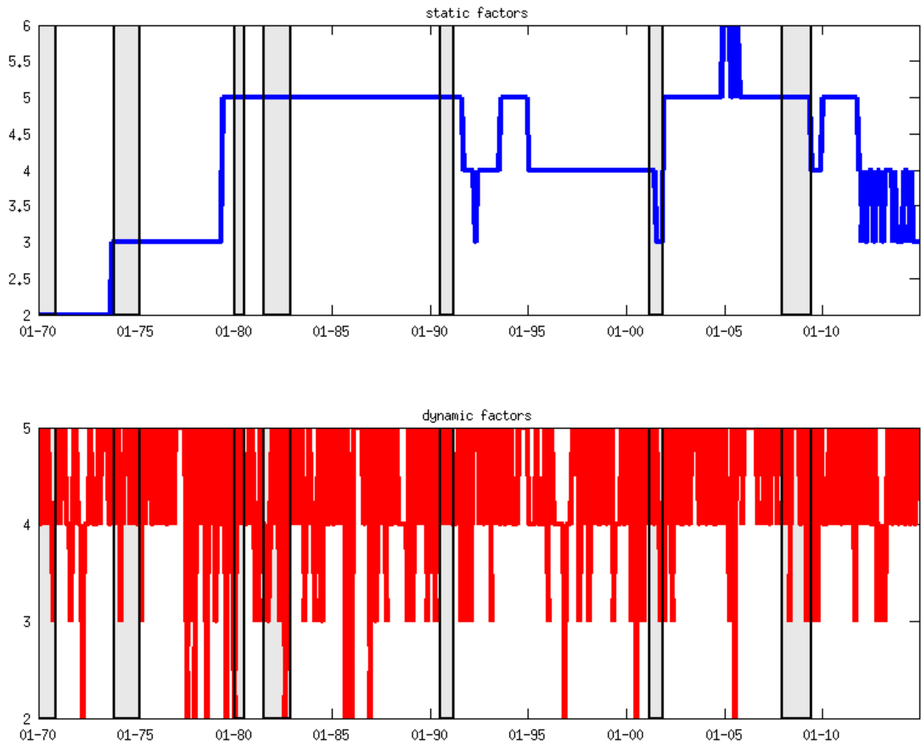
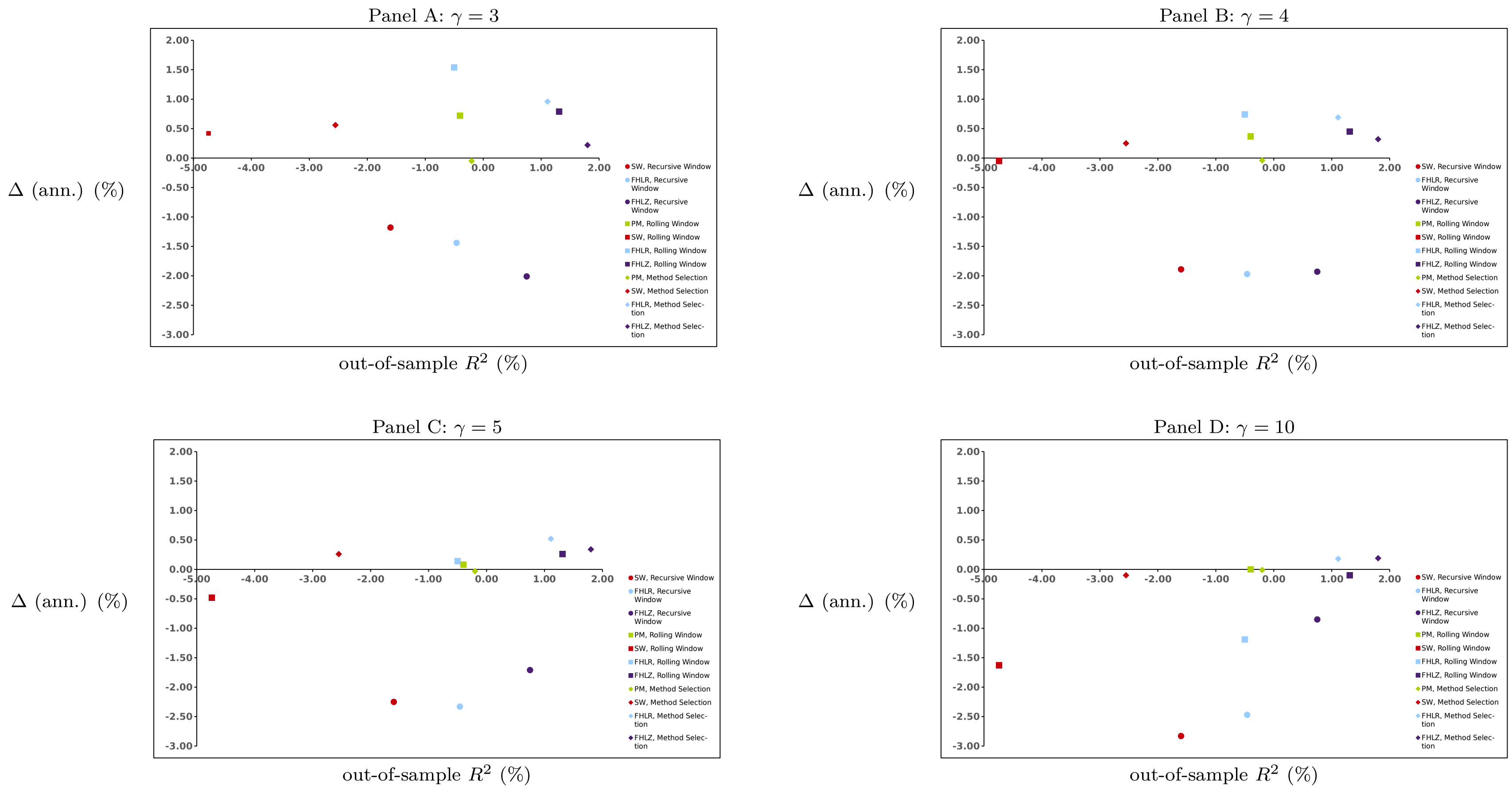


Figure 2: Statistical and Economic Out-of-Sample Forecast Accuracy

Notes. This figure illustrates the link between statistical and economic out-of-sample forecast accuracy as measured by out-of-sample R^2 (%) and annualized utility gain (%), respectively, of prediction models for the monthly excess return $\rho_{\tau+1}$. γ denotes the relative risk aversion parameter. Stock and Watson (2002) (SW), Forni *et al.* (2005) (FHLR), and Forni *et al.* (2015, 2017) (FHLZ) factor models are described in Section 2. Recursive (Rec) and rolling (Rol) window estimation schemes are described in Section 4.1.2. The method selection criterion (Sel) allows to switch between estimation windows using the procedure described in Section 4.3.2. The sample period is 1970 − 2014.



6 Appendix: FRED-DATASET

We adopt the balanced version of FRED dataset discarding the series with missing values at beginning of the sample. These are: PERMIT, PERMITNE, PERMITMW, PERMITS, PERMITW, ACOGNO, ANDENOx, TWEXMMTH, UMCSENTx.
Letting x_t be a raw series, the transformations adopted are:

(1) no transformation;

(2) Δx_t;

(3) $\Delta^2 x_t$;

(4) $\Delta log(x_t)$;

(5) $log(x_t)$;

(6) $\Delta^2 log(x_t)$;

(7) $\Delta \left(\frac{x_t}{x_{t-1} - 1} \right) log(x_t)$;

Table 1: List of the series

	mnemonic	description	tcode
1	RPI	Real Personal Income	5
2	W875RX1	RPI ex. Transfers	5
3	INDPRO	IP Index	5
4	IPFPNSS	IP: Final Products and Supplies	5
5	IPFINAL	IP: Final Products	5
6	IPCONGD	IP: Consumer Goods	5
7	IPDCONGD	IP: Durable Consumer Goods	5
8	IPNCONGD	IP: Nondurable Consumer Goods	5
9	IPBUSEQ	IP: Business Equipment	5
10	IPMAT	IP: Materials	5
11	IPDMAT	IP: Durable Materials	5
12	IPNMAT	IP: Nondurable Materials	5
13	IPMANSICS	IP: Manufacturing	5
14	IPB51222S	IP: Residential Utilities	5
15	IPFUELS	IP: Fuels	5
16	NAPMPI	ISM Manufacturing: Production	1
17	CAPUTLB00004S	Capacity Utilization: Manufacturing	2
1	HWI	Help-Wanted Index for US	2
2	HWIURATIO	Help Wanted to Unemployed ratio	2
3	CLF16OV	Civilian Labor Force	5
4	CE16OV	Civilian Employment	5
5	UNRATE	Civilian Unemployment Rate	2
6	UEMPMEAN	Average Duration of Unemployment	2
7	UEMPLT5	Civilians Unemployed <5 Weeks	5
8	UEMP5TO14	Civilians Unemployed 5-14 Weeks	5
9	UEMP15OV	Civilians Unemployed >15 Weeks	5
10	UEMP15T26	Civilians Unemployed 15-26 Weeks	5
11	UEMP27OV	Civilians Unemployed >27 Weeks	5
12	CLAIMSx	Initial Claims	5
13	PAYEMS	All Employees: Total nonfarm	5
14	USGOOD	All Employees: Goods-Producing	5
15	CES1021000001	All Employees: Mining and Logging	5
16	USCONS	All Employees: Construction	5
17	MANEMP	All Employees: Manufacturing	5
18	DMANEMP	All Employees: Durable goods	5
19	NDMANEMP	All Employees: Nondurable goods	5
20	SRVPRD	All Employees: Service Industries	5
21	USTPU	All Employees: TT&U	5
22	USWTRADE	All Employees: Wholesale Trade	5
23	USTRADE	All Employees: Retail Trade	5
24	USFIRE	All Employees: Financial Activities	5
25	USGOVT	All Employees: Government	5
26	CES0600000007	Hours: Goods-Producing	1
27	AWOTMAN	Overtime Hours: Manufacturing	2
28	AWHMAN	Hours: Manufacturing	1
29	NAPMEI	ISM Manufacturing: Employment	1
30	CES0600000008	Ave. Hourly Earnings: Goods	6
31	CES2000000008	Ave. Hourly Earnings: Construction	6
32	CES3000000008	Ave. Hourly Earnings: Manufacturing	6
1	HOUST	Starts: Total	4
2	HOUSTNE	Starts: Northeast	4
3	HOUSTMW	Starts: Midwest	4
4	HOUSTS	Starts: South	4
5	HOUSTW	Starts: West	4

- Continued on next page -

Table 1 – continued from previous page

1	DPCERA3M086SBEA	Real PCE	5
2	CMRMTSPLx	Real M&T Sales	5
3	RETAILx	Retail and Food Services Sales	5
4	NAPM	ISM: PMI Composite Index	1
5	NAPMNOI	ISM: New Orders Index	1
6	NAPMSDI	ISM: Supplier Deliveries Index	1
7	NAPMII	ISM: Inventories Index	1
8	AMDMNOx	Orders: Durable Goods	5
9	AMDMUOx	UnÞlled Orders: Durable Goods	5
10	BUSINVx	Total Business Inventories	5
11	ISRATIOx	Inventories to Sales Ratio	2
1	M1SL	M1 Money Stock	6
2	M2SL	M2 Money Stock	6
3	M3SL	MABMM301USM189S in FRED, M3 for the United States	6
4	M2REAL	Real M2 Money Stock	5
5	AMBSL	St. Louis Adjusted Monetary Base	6
6	TOTRESNS	Total Reserves	6
7	NONBORRES	Nonborrowed Reserves	6
8	BUSLOANS	Commercial and Industrial Loans	6
9	REALLN	Real Estate Loans	1
10	NONREVSL	Total Nonrevolving Credit	6
11	CONSPI	Credit to PI ratio	2
12	MZMSL	MZM Money Stock	6
13	DTCOLNVHFNM	Consumer Motor Vehicle Loans	6
14	DTCTHFNM	Total Consumer Loans and Leases	6
15	INVEST	Securities in Bank Credit	6
1	FEDFUNDS	Effective Federal Funds Rate	2
2	CP3M	3-Month AA Comm. Paper Rate	2
3	TB3MS	3-Month T-bill	2
4	TB6MS	6-Month T-bill	2
5	GS1	1-Year T-bond	2
6	GS5	5-Year T-bond	2
7	GS10	10-Year T-bond	2
8	AAA	Aaa Corporate Bond Yield	2
9	BAA	Baa Corporate Bond Yield	2
10	COMPAPFF	CP - FFR spread	1
11	TB3SMFFM	3 Mo. - FFR spread	1
12	TB6SMFFM	6 Mo. - FFR spread	1
13	T1YFFM	1 yr. - FFR spread	1
14	T5YFFM	5 yr. - FFR spread	1
15	T10YFFM	10 yr. - FFR spread	1
16	AAAFFM	Aaa - FFR spread	1
17	BAAFFM	Baa - FFR spread	1
18	EXSZUS	Switzerland / U.S. FX Rate	5
19	EXJPUS	Japan / U.S. FX Rate	5
20	EXUSUK	U.S. / U.K. FX Rate	5
21	EXCAUS	Canada / U.S. FX Rate	5
1	PPIFGS	PPI: Finished Goods	6
2	PPIFCG	PPI: Finished Consumer Goods	6
3	PPIITM	PPI: Intermediate Materials	6
4	PPICRM	PPI: Crude Materials	6
5	oilprice	Crude Oil Prices: WTI	6
6	PPICMM	PPI: Commodities	6
7	NAPMPRI	ISM Manufacturing: Prices	1
8	CPIAUCSL	CPI: All Items	6
9	CPIAPPSL	CPI: Apparel	6
10	CPITRNSL	CPI: Transportation	6
11	CPIMEDSL	CPI: Medical Care	6
12	CUSR0000SAC	CPI: Commodities	6
13	CUUR0000SAD	CPI: Durables	6
14	CUSR0000SAS	CPI: Services	6
15	CPIULFSL	CPI: All Items Less Food	6
16	CUUR0000SA0L2	CPI: All items less shelter	6
17	CUSR0000SA0L5	CPI: All items less medical care	6
18	PCEPI	PCE: Chain-type Price Index	6
19	DDURRG3M086SBEA	PCE: Durable goods	6
20	DNDGRG3M086SBEA	PCE: Nondurable goods	6
21	DSERRG3M086SBEA	PCE: Services	6
1	S&P 500	S&P: Composite	5
2	S&P: indust	S&P: Industrials	5
3	S&P div yield	S&P: Dividend Yield	2
4	S&P PE ratio	S&P: Price-Earnings Ratio	5