Portfolio Efficiency with High-Dimensional Data as Conditioning Information

Caio Vigo Pereira†

†*Department of Economics - University of Kansas*

September 14, 2020

Abstract

In this paper, we build efficient portfolios using different frameworks proposed in the literature with several datasets containing an increasing number of predictors as conditioning information. We carry an extensive empirical study to investigate several approaches to impose sparsity and dimensionality reduction, as well as possible latent factors driving the returns of the risky assets. In contrast to previous studies that made use of naive OLS and low-dimension information sets, we find that (i) accounting for large conditioning information sets, and (ii) the use of variable selection, shrinkage methods and factors models, such as the principal component regression and the partial least squares provide better out-of-sample results as measured by Sharpe ratios.

JEL classification: G11, G17, C32, C38.

Keywords: Dimensionality reduction. Shrinkage. Efficient Portfolios. Principal Components Regression (PCR). Partial Least Squares (PLS). Three-Pass Regression Filter (3PRF). Ridge Regression, LASSO.

Corresponding author - 1460 Jayhawk Blvd - Snow Hall, The University of Kansas, USA. email - caiovigo@gmail.com

1 Introduction

The use of return predictability to form efficient portfolios is a cornerstone in the asset pricing literature. Theoretically, this has been addressed in the mean-variance efficiency framework where many ways of bringing the conditioning information during the optimization process was proposed. Over the last decades, the literature has recognized some variables that have predictive power and therefore could be used as conditioning information. However, the search for all these instruments produced an enormous amount of variables that supposedly has or had some predictability power. This applies to firm characteristics, whole market financial variables, and macroeconomic datasets as well.

In this paper, we combine the nascent high-dimensional literature in finance to the above portfolio theory problem. We incorporate conditioning information in the efficient portfolios framework proposed in the literature making use of the recently developed statistical techniques for high-dimensional predictive estimations. Specifically, we build efficient portfolios using high-dimensional information sets as conditioning information and conduct an extensive out-of-sample analysis comparing the high-dimensional statistical methods with a naive OLS approach of dealing with a large set of potential conditioning information.

Since the purpose of using conditioning information is to provide signals about the state of the economy or the condition of the risky assets, exploring a diverse, large set of possibilities for the predictive instruments, as well as flexible functional forms in which these signals are used in the optimization process can enhance the construction of efficient portfolios. Another motivation for working with high-dimensional information sets can be related to the Hansen-Richard critique (see Cochrane, 2009). Even though there might be unobservable conditioning information used by economic agents, when dealing with a large set of potential candidates, we could think that enlarging this set increases the chance of capturing more information from a wide variety of heterogeneous economic agents' information sets. On top of that, with the use of recently proposed measures of economic uncertainty (see Baker et al., 2016; Püttmann, 2018; Jurado et al., 2015), we could partially capture some of what used to be non-observable information.

We exploit the wealth set of data as predictive signals and condense them when estimating the conditional mean using high-dimensional data. We do this using many different techniques in order to impose sparsity and dimensionality reduction when finding the conditional mean, which is the most important driver in the formation of mean-variance efficient portfolios with conditioning information.

In short, we make a "bet in sparsity", in the sense that much of the information in the conditioning set can be summarized by few factors. We evaluate how penalized estimators, such as LASSO, Ridge and Elastic Net, as well as pure dimensionality reduction and latent factors approaches as Partial Least Squares (PLS) and Principal Components Regression (PCR), in addition to a generalization of the former (Three Pass Regression Filter) can produce different optimized portfolios.

This study makes a number of contributions to the literature. Assessing these methods in a high-dimensional setting for the collection of instruments, we show that it is possible to build efficient portfolios, for different efficient portfolios construction approaches, delivering on average higher out-of-sample Sharpe ratios, implied Sharpe ratios, and higher certainty equivalent returns (CER). We find that PLS and PCR can enhance the formation of optimized portfolios, generating large and significant alphas when evaluated in factor models, such as Fama-French 3 and 5. We also evaluate which one, among the large set of signals, have larger impacts in the excess returns conditional means.

From the theoretical implications derived by Hansen and Richard (1987) Hansen and Richard (1987), we could split in two main ways of defining the minimum variance efficient portfolios: the straightforward conditionally mean-variance (CMV), and the unconditionally mean-variance with conditioning information (UMV) portfolios. The first approach (CMV) is the most direct way to incorporate conditioning information. On the other hand, with roots on information asymmetry, UMV was studied in Ferson and Siegel (2001) who presented closed-form solutions. UMV is a more interesting and applicable way of working with conditioning information. For the sake of completeness, we also assess a modern portfolio management version of UMV proposed by Chiang (2015), known as mean-variance active tracking error (MVATE).

The dimension reduction literature is a growing topic having produced so far many methods to deal with high-dimensional datasets. One of the most straightforward and commonly used technique is based on the principal components analysis (PCA). The principal components regression (PCR) is a dimension reduction technique based on PCA. Several studies in macroeconomics made use of principal components to condense information (see Stock and Watson (2002a,b)). Kelly et al. (2019) proposed a dimension reduction based on PCA to explain the cross-section of average returns. Gu et al. (2018) made use of PCR, among other machine learning techniques, for measuring asset risk premia. Another classical dimension reduction technique is the partial least squares (PLS). PLS is a simple regression-based procedure designed to parsimoniously forecast a single time series using a large panel of predictors. Light et al. (2017) uses PLS to estimate the expected returns on individual stocks from a large set of firm characteristics.

A recent method was proposed by Kelly and Pruitt (2015). The Three-Pass Regression Filter (3PRF) is a generalization of PLS that can be represented as a set of ordinary regression filter. The 3PRF is a constrained least squares estimator and reduces to partial least squares as a special case. Kelly and Pruitt (2013) uses the 3PRF to forecast the conditional expectation of market returns and cash flow.

A growing and nascent literature in finance has started to apply penalized regression in asset pricing. Kozak et al. (2019) adapt Ridge and LASSO estimators to estimate the stochastic discount factor to summarize and impose sparsity of an extensive list of predictors in a high-dimension setting. Freyberger et al. (2017) use a variation of LASSO to select characteristics from a large number of characteristics and estimate a nonlinear function for expected returns. Chinco et al. (2019) use LASSO to make intraday return forecasts for a large number of stocks. Other papers made use of penalized estimators for macro variables and bond risk premia as well (see Bianchi et al. (2019); Huang and Shi (2011); Bai and Ng (2006)). In section 3 we present these estimators.

The structure of this paper is as follows. Next section introduces the general framework and presents the three approaches mentioned above, the optimization problems and the weights' solution of building mean-variance efficient portfolios. Section 3 summarizes the techniques to impose sparsity in a high-dimensional setting, presenting the penalized estimators, the PLS, PCR, and the 3PRF. Section 4 explains our research design, explaining the datasets used, and how the performance is evaluated. Section 5 presents the results. Finally, section 6 concludes. Additional results, tables and figures are presented in the Internet Appendix.

2 General Framework

Consider an investment set with N risky assets, indexed by $i = 1, \ldots, N$ and the existence of a risk-free asset. Let $R_{f,t}$ be the return of the risk-free asset at time t, and denote $\mathbf{R}_t$[1] as the N-dimensional vector with the gross returns of the N risky assets at time t. The vector of excess of returns $\mathbf{r}_t$ at time t is given by $\mathbf{r}_t = \mathbf{R}_t - \mathbf{1}R_f$, where $\mathbf{1}$ is a an N-dimensional vector of ones. Let $\mathbf{x}_t(\mathbf{Z}_t)$ denote the vector of portfolio weights on the N risky assets at time t, which is a function of the investor's conditioning information $\mathbf{Z}_t$ at time t. At each period of time t, using the conditioning information $\mathbf{Z}_t$ the investor set the weights of the vector $\mathbf{x}_t$, investing the remaining of the funds in the risk-free asset. Thus, the weight on the risk-free asset is given by $1 - \mathbf{x}_t(\mathbf{Z}_t)^\top \mathbf{1}$. Therefore, the final payoff of the portfolio at $t + 1$ is given by:

[1] A note on the notation throughout our analysis. Unless mentioned otherwise, we use lowercase boldface characters to represent vectors (except for $\mathbf{R}_t$), uppercase boldface characters for matrices, and regular characters for scalars.

$$R_{p,t+1} = \mathbf{R}_{t+1}^{\top}\mathbf{x}_t(\mathbf{Z}_t) + \left(1 - \mathbf{x}_t(\mathbf{Z}_t)^{\top}\mathbf{1}\right) R_{f,t+1} \tag{1}$$

$$= \left(\mathbf{R}_{t+1} - \mathbf{1}R_{f,t+1}\right)^{\top} \mathbf{x}_t(\mathbf{Z}_t) + R_{f,t+1} \tag{2}$$

$$= \mathbf{x}_t(\mathbf{Z}_t)^{\top}\mathbf{r}_{t+1} + R_{f,t+1} \tag{3}$$

In general, we assume that we can model the data generating process of the excess returns of each one the risky assets as

$$r_{i,t} = \mu_{i,t}\left(\mathbf{Z}_{t-1}\right) + \epsilon_{i,t} \tag{4}$$

where μ_t conditional expected excess returns of asset i at time t, and $\epsilon_{i,t}$ is the noise term with conditional mean equal to 0 given $\mathbf{Z}_{t-1}$. Equation (4) highlights that this conditional mean is a function of a set of variables that enters this function lagged one period, seeking to span the set of available information. The fundamental idea is that this conditional mean is a function that can be broadly represented as

$$\mu_{i,t}\left(\mathbf{Z}_{t-1}\right) = h_i\left(\mathbf{Z}_{t-1}\right) \tag{5}$$

where $h_i(\cdot)$ is any function to approximate the conditional mean. Previous studies that analyzed empirically efficient portfolios for a specific market and asset class, predominantly made use of OLS to estimate $\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)$ when deriving the efficient portfolios weights from the mean-variance optimization problem, setting $\mu_{i,t}\left(\mathbf{Z}_{t-1}\right) = \mathbb{E}\left(r_{i,t}|\mathbf{Z}_{t-1}\right)$ from a standard linear regression. This was mainly motivated by the fact they used a small number of instruments as conditioning information set. Most of them have used at most a handful of variables. The choices of these variables were mostly motivated by previous studies that have found some predictability power for a specific variable in a specific period in the past. In this fashion, they argued that the conditional expected return for each asset $i = 1, \ldots, N$, could be obtained by running the following regression:

$$\mathbf{r}_i = \mathbf{Z}_{t-1}\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i \tag{6}$$

where $\mathbf{Z}_{t-1}$ is a $T \times K$ matrix of lagged instruments and $\boldsymbol{\theta}_i$ is K-dimensional vector of parameters. Using the fitted values of the times series from equation (6) as the sample counterpart of $\boldsymbol{\mu}_t$, as follows

$$\widehat{\boldsymbol{\mu}}_t = \begin{bmatrix} \widehat{\mu}_{1,t}\left(\mathbf{Z}_{t-1}\right) \\ \widehat{\mu}_{2,t}\left(\mathbf{Z}_{t-1}\right) \\ \vdots \\ \widehat{\mu}_{N,t}\left(\mathbf{Z}_{t-1}\right) \end{bmatrix} \tag{7}$$

it is possible to find the solutions of the optimal portfolios weights.

However, in doing so, we face several issues. First, as Welch and Goyal (2007) shows in a comprehensive study, most of the variables taken as having predictability power and commonly used in academic research and by market practitioners perform badly out-of-sample, and in some cases even in the in-sample analysis. The authors selected variables from previously published articles and show that most of them behave poorly with unstable predictive performance.

Second, even in the case, a variable has some predictability power, this can only exist for a specific period (perhaps only in the past). Welch and Goyal (2007) sheds light on, among other things, to the problem of short-lived predictors, in contraposition to steady long-lived predictors. Some of the variables could have some predictive power that existed in a specific period. In a broad analysis McLean and Pontiff (2016) evaluated the out-of-sample post-publication return predictability of 97 signals from 79 different academic studies. They show that after an academic paper is published, the average predictor's long-short return shrinks post-publication[2]. They also show that the publication has an impact on the predictor portfolio returns. They argue that post-published correlation with already-published predictor portfolio returns increases.

Third, as Chinco et al. (2019) mentions, modern financial markets are big, fast, and complex. With access to so many sources, the process of data collection and processing reached the level of commonly known as big data. As Diebold et al. (2019) summarizes, in time-series regression involving T time periods, K regressors and N dimensions (multivariate), we need to deal with some different characteristic data. As K gets large (e.g. $K > T$) we have what is called as *wide data*, while as N gets large we have *high-dimensional data*. It is known that in such cases the ordinary least squares (OLS) estimates will have a poor behavior in forecasting, or not even exist for wide data for instance. An appropriate way to deal with these types of data is via some sort of shrinkage-type regularization. Sparsity and dimensionality reduction are the goals of these problems. There are compelling arguments that support that we cannot rely on a small set of conditioning information. In order to obtain

[2]Portfolio returns dropped 26% out-of-sample and 58%, yielding an estimated effect of 32% (58%–26%) lower return from publication-informed trading.

better approximations to the conditional mean, the set of instruments must be enriched to resemble the exponential growth of available information.

The financial literature has collected a large list of variables that could be part of the conditioning information set, as pointed by Cochrane (2011). Several recent studies tried to compile these possible predictors. Green et al. (2013) assessed 330 signals (stock-level) reported in the past years. Feng et al. (2017) studied a list of 150 risk factors, what they called a "zoo" of factors proposed in the literature in the past 30 years. Harvey et al. (2016) analyzed the statistical significance of more than 300 signals published over the last forty years.

Finally, finding the conditional mean to explain the variation of the risky assets, is a risk premia problem. By far, it is known that the problem of explaining the risk premia has been one of the most debated and scrutinized topics for a long period in the literature. Clearly, approximating $h_i(\cdot)$ is a hard task because market efficiency *forces* return variation to be dominated by unforecastable news that obscures risk premia Gu et al. (2018). We can say that not only the amount, but also the complexity of information driving the conditional mean of the risky assets could be enhanced. Modeling the conditional mean with few observable factors entering linearly, with no interactions, penalization, or dimension reduction techniques, could also be an incomplete structure to explain the stochastic discount factor.

Hence, in order to exploit the wealth of data and try to condense them when estimating the conditional mean using high-dimensional data, we seek to impose sparsity and dimensionality reduction when finding the conditional mean that enters in all functions to build efficient portfolios. Importantly, it is the fact that we can consider the function $h_i(\cdot)$ in equation 5 that seeks to approximate the conditional mean driving the DGP of the excess returns to have a rich and flexible format. In order to obtain better approximations, we assume that $h_i(\cdot)$ is asset dependent, in the sense that the factors that drive the excess returns of each risky asset i are built using a high-dimensional data for each different asset.

2.1 Optimal Portfolios with Conditioning Information

In this section, we briefly present some of the approaches that were proposed in the literature to build efficient portfolios with conditioning information, summarizing three frameworks.

2.1.1 Conditionally Efficient Portfolios

As in Hansen and Richard (1987), the conditionally efficient returns is the solution of the following problem:

$$\begin{aligned} \min_{x(Z)} \quad & \mathbb{E}\left(r_{p,t+1}^2 | \mathbf{Z}_t\right) \\ \text{s.t.} \quad & \mathbb{E}\left(r_{p,t+1} | \mathbf{Z}_t\right) = \alpha_{p,t+1} \end{aligned} \tag{8}$$

where $r_{p,t+1}$ is the portfolio excess returns at time $t+1$, and $\alpha_{p,t+1}$ is the target conditional expected excess return. The solution of the problem in equation (8) providing the weights allocated in each risky asset is given by:

$$\mathbf{x}_t(\mathbf{Z}_{t-1}) = \left(\frac{\mathbb{E}\left(r_{p,t+1} | \mathbf{Z}_t\right)}{\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^\top \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)} \right) \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right) \tag{9}$$

where $\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)$ is the n-dimensional vector of conditional expected excess returns at time t, and $\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)$ is the $N \times N$ conditional second moment matrix at time t.

2.1.2 Unconditionally Efficient Portfolios with Respect to the Information

The unconditionally mean-variance (UMV) efficient portfolio strategy takes another path. Ferson and Siegel (2001) study the properties of unconditional minimum-variance portfolios in the presence of conditioning information. As the authors argue, whenever the investor makes use of conditioning information, mean-variance efficiency can be either defined in terms of the two first conditional moments (as in section 2.1.1), or the unconditional moments as in the UMV approach. Here, the objective is to maximize the unconditional mean relative to the unconditional variance, being the strategies a function of the available information.

Hansen and Richard (1987) show that unconditionally efficient returns are a subset of conditionally efficient returns. Related with information asymmetry, this is a typical situation where a portfolio manager using conditioning information builds a conditionally efficient portfolio. However, for an uninformed investor, the portfolio does not appear efficient. As in Ferson and Siegel (2001), the solution of the UMV problem providing the weights allocated in each risky asset is given by:

$$\mathbf{x}_t(\mathbf{Z}_{t-1}) = \left(\frac{\mathbb{E}\left(r_{p,t+1}\right)}{\mathbb{E}\left(\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^\top \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)\right)} \right) \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right) \tag{10}$$

where $\mathbb{E}\left(r_{p,t+1}\right)$ is the unconditional expected return target, and $\mathbb{E}\left(\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^\top \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)\right)$ is the unconditional mean of $\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^\top \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)$.

2.2 Unconditionally Tracking Efficient Portfolios

Chiang (2015) follows a different approach to exploit conditioning information in the mean-variance framework. The author uses tracking error strategies, in which an active portfolio manager uses conditioning information to optimize unconditional performance measures relative to a benchmark. He calls this approach as *unconditional tracking efficiency.* The author mentions that this structure should be seen as a common practice between market practitioners, since a portfolio manager conducting an optimization uses more information than his clients, that possibly do not have access to the entire set being used as conditioning information, driving managers to seek higher performance from the clients' perspective.

In this framework, consider a portfolio manager who uses conditioning information $\mathbf{Z}_t$, which is not available to his clients, to build a vector $\mathbf{x}_t(\mathbf{Z}_{t-1})$ of unrestricted weights of the N available risky assets. Under the unconditional mean-variance tracking error approach (UMVTE), this active manager uses conditioning instruments to form portfolios that optimize unconditional performance measure. Thus, for a benchmark b, with returns given by $R_{b,t}$, the UMVTE problem is to minimize the unconditional error variance, $\mathbb{V}\text{ar}\left(R_{p,t+1} - R_{b,t+1}\right)$, for a given unconditional expected return target $\alpha_{p,t+1} = \mathbb{E}(R_{p,t+1} - R_{b,t+1})$,

$$\begin{aligned} \min_{\mathbf{x}(\mathbf{Z})} \quad & \mathbb{V}\text{ar}\left(R_{p,t+1} - R_{b,t+1}\right) \\ \text{s.t.} \quad & \mathbb{E}\left(R_{p,t+1} - R_{b,t+1}\right) = \alpha_{p,t+1} \end{aligned} \tag{11}$$

Chiang (2015) solves this minimization problem using calculus of variations. The unique solution to the problem given on equation 11 without constraint on portfolio risk is determined by the following function for the portfolios weights:

$$\mathbf{x}_t(\mathbf{Z}_{t-1}) = \lambda_1 \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right) + \left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\gamma}_t\left(\mathbf{Z}_{t-1}\right) \tag{12}$$

where, $\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)$ is the conditional mean, $\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)$ is the conditional second moment of the excess returns, $\boldsymbol{\gamma}_t\left(\mathbf{Z}_{t-1}\right) = \mathbb{E}\left(\mathbf{r}_t\left(r_{b,t} - \mu_{b,t}\right)|\mathbf{Z}_{t-1}\right)$, and λ_1 an scalar [3]. For a full proof of the solution of the UMVTE problem, see Chiang (2015). Note that the CMV, UMV and UMVTE are not the only frameworks proposed in the literature to build efficient portfolios taking into account conditioning information. Brandt and Santa-Clara (2006) Peñaranda (2016) also proposed different approaches to build mean-variance efficient portfolios with

[3] Precisely,

$$\lambda_1 = \frac{\left(\alpha_{p,t+1} + \mu_{b,t}\right) - \eta_2}{\eta_1} \tag{13}$$

where $\eta_1 = \mathbb{E}\left(\boldsymbol{\mu}_t^\top\left(\mathbf{Z}_{t-1}\right)\left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)\right)$, and $\eta_2 = \mathbb{E}\left(\boldsymbol{\mu}_t^\top\left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1} \boldsymbol{\gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)$

conditioning information.

3 Estimation

In this section we summarize some of the approaches proposed in the current state of the literature to deal with high-dimensional set of predictors, imposing sparsity through shrinkage and selection type techniques, and latent variables estimators as well.

3.1 Penalized Regression

3.1.1 LASSO

Consider the regression in equation (6). In general, the OLS regression to obtain the K parameters of $\boldsymbol{\theta}_i$ will provide nonzero estimates. When dealing with high-dimensional with a large number of predictors, usually we face two issues using OLS. First, whenever we have $K > T$, OLS cannot find a unique solution. Second, the infinite set of solutions have a tendency to overfit the data instead of extracting informative signals for the conditional mean. This fact, known as "curse of dimensionality" is well known in the literature. Gu et al. (2018) mention that this issue is more exasperating where the signal-noise ratio in return predictions is low. This causes the OLS estimations to have a poor behavior in forecasting.

To deal with the inconsistency and the inefficiency of the OLS in high-dimensional data, shrinking the number of estimated parameters is essential to avoid overfitting the data. Thus, some sort of shrinkage or variable selection might be necessary to deal with this type of data. This can be done penalizing the objective function being optimized. In general, penalized estimators share the following common structure:

$$\mathcal{L}(\boldsymbol{\theta}; \cdot) = \underbrace{\mathcal{L}(\boldsymbol{\theta})}_{\text{Loss Function}} + \underbrace{\phi(\boldsymbol{\theta}; \cdot)}_{\text{Penalty}} \tag{14}$$

An estimator to impose some sort of sparsity in the set of regressors is the Least Absolute Selection and Shrinkage Operator (LASSO), proposed in the seminal paper by Tibshirani (1996). This estimator regularizes the estimation process constraining the size of the estimates. Following Hastie et al. (2015); Friedman et al. (2001), the LASSO is a ℓ_1 regularized regression obtained by minimizing the following problem:

$$\hat{\boldsymbol{\theta}}_i^{LASSO} = \arg\min_{\theta_{0,i},\boldsymbol{\theta}_i} \left\{ \frac{1}{T}\sum_{t=1}^{T} \left(r_{i,t} - \theta_{0,i} - \sum_{k=1}^{K} Z_{k,t-1}\theta_{k,i} \right)^2 \right\} \quad (15)$$

$$\text{s.t.} \quad \|\theta\|_1 \leq c \quad (16)$$

where $\|\theta\|_1 = \sum_{k=1}^{K} |\theta_{k,i}|$ is the ℓ_1 norm of θ. We call c the "parameter budget", being possible to interpret $\boldsymbol{\theta}_i^{LASSO}$ as the best fit within this budget[4]. The parameter λ controls the amount the shrinkage, the larger it is, the greater is the shrinkage. The penalty imposed in equation (17) makes the solutions nonlinear in $r_{i,t}$. This is associate with the fact that there is no closed-form solution. The lasso problem is a convex program, and computing its solution is a quadratic programming problem. An important feature of LASSO is that making c sufficiently small, this estimator sets a subset of the estimated coefficients to exactly zero in order to impose sparsity.

3.1.2 Ridge

The ridge estimator is similar to the LASSO. However, ridge sets the penalty $\phi(\boldsymbol{\theta};\cdot)$ as $\|\theta\|_2^2 = \sum_{k=1}^{K} \theta_{k,i}^2$, i.e., as the penalty is a ℓ_2 norm. Thus, we we can it write as

$$\hat{\boldsymbol{\theta}}_i^{Ridge} := \left\{ \arg\min_{\theta_{0,i},\boldsymbol{\theta}_i} \frac{1}{T}\sum_{t=1}^{T} \left(r_{i,t} - \theta_{0,i} - \sum_{k=1}^{K} Z_{k,t-1}\theta_{k,i} \right)^2 + \lambda \sum_{k=1}^{K} \theta_{k,i}^2 \right\} \quad . \quad (18)$$

An important characteristic of Ridge approach is the one-to-one correspondence between the parameters λ and t. This fact makes Ridge estimation to attenuate multicollinearity, when it is present in the data. Since, when K is large, the large number of regressors may result in high correlation among some of them. With standard "naive" OLS, multicollinearity causes poorly determined coefficients with large variances (and large variance inflator factor - VIF), what Ridge can mitigate.

[4]Equivalently, in Lagrange-multiplier form we can it write as,

$$\hat{\boldsymbol{\theta}}_i^{LASSO} := \arg\min_{\theta_{0,i},\boldsymbol{\theta}_i} \left\{ \frac{1}{T}\sum_{t=1}^{T} \left(r_{i,t} - \theta_{0,i} - \sum_{k=1}^{K} Z_{k,t-1}\theta_{k,i} \right)^2 + \lambda \sum_{k=1}^{K} |\theta_{k,i}| \right\} \quad . \quad (17)$$

3.1.3 Elastic Net

Among many variants of LASSO, there is the Elastic Net (ENet) (Zou and Hastie, 2005). In summary, the ENet seeks to combine the penalties of both estimators above (LASSO and Ridge) via a convex combination. Therefore, the ENet estimator is given by,

$$\hat{\boldsymbol{\theta}}_i^{ENet} := \left\{ \arg\min_{\theta_{0,i},\boldsymbol{\theta}_i} \frac{1}{T}\sum_{t=1}^{T}\left(r_{i,t} - \theta_{0,i} - \sum_{k=1}^{K} Z_{k,t-1}\theta_{k,i}\right)^2 + \lambda \sum_{k=1}^{K}\left(\alpha\left|\theta_{k,i}\right| + (1-\alpha)\,\theta_{k,i}^2\right)\right\} \quad (19)$$

where $\alpha \in [0,1]$. Setting $\alpha = 1$ this estimator reduces to the LASSO regression with a ℓ_1-norm, with $\alpha = 0$ it reduces to the ridge regression with a ℓ_2-penalty. Among the positive features of the ENet, when adding both penalties, this estimator automatically controls for strong within-group correlations. The ENet is also a strictly convex problem, thus, providing a unique solution independently of correlations or duplications in the $Z_{k,t-1}$. However, in equation (19) we have an additional tuning parameter α that has to be determined or defined *ad hoc*.

3.2 Principal Components Regression (PCR)

Whenever we are dealing with a high dimensional set of explanatory variables, high collinearities may be an issue to generate good conditional means estimates. One technique to deal with multicollinearity is the Principal Components Regression (PCR). PCR is a dimension reduction approach based on the Principal Components Analysis (PCA). Being a regression technique, PCR regresses the dependent variable on the principal components generated by the PCA.

In summary, PCR first performs PCA on the matrix of predictors (here, the $\mathbf{Z}_{t-1}$ of lagged instruments) and obtains the principal components associate with it. Out of the M components, we can select a subset $m \in M$ and regress the dependent variable using the first m principal components as regressors[5]. Since PCR is essentially a shrinkage estimator that seeks to capture components with high variance[6], at the cost of those with low variance, this approach performs dimension reduction by selecting m components, and also has a shrinkage effect as it removes those components with the low variance to form the final model to obtain the conditional mean.

Precisely, consider equation (6). Using the singular value decomposition (SVD), we

[5] A final step is done to scale back to the original regressors using the PCA loadings.

[6] Those with higher eigenvalues in $\mathbf{Z}^\top\mathbf{Z}$.

can factorize $\mathbf{Z}$[7] as $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, where $\mathbf{U}_{T\times K}$ and $\mathbf{V}_{K\times K}$ are orthogonal matrices[8], and $\mathbf{D}_{K\times K} = \text{diag}\left[\delta_1, \delta_2, \dots, \delta_K\right]$, being each δ_m the singular values of $\mathbf{Z}$.

We can obtain the principal values of $\mathbf{Z}^\top\mathbf{Z}$ as $\boldsymbol{\Lambda}_{K\times K} = \text{diag}\left[\delta_1^2, \delta_2^2, \dots, \delta_K^2\right]$ through a spectral decomposition given by $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$. Finally, we can multiply the matrix $\mathbf{Z}$ by the first m principal components $\mathbf{V}^{(m)} = \left[\boldsymbol{\omega}^{(1)}, \dots, \boldsymbol{\omega}^{(m)}\right]$, where $m \in \{1, 2, \dots, K\}$, to obtain our derived covariates. Each m principal component $\boldsymbol{\omega}^{(m)}$ can be seen as m linear combinations of $\mathbf{Z}_{t-1}$. Thus, using the PCR approach, the conditional mean at time t of the excess returns of risky asset i can be approximated running the following regression

$$\mathbf{r}_i = \underbrace{\left(\mathbf{Z}_{t-1}\mathbf{V}_{t-1}^{(m)}\right)}_{\text{Derived Covariates}} \boldsymbol{\theta}_i^{(m)} + \boldsymbol{\epsilon}_i \quad , \tag{20}$$

so that $\widehat{\boldsymbol{\mu}}_i\left(\mathbf{Z}_{t-1}\right) = \left(\mathbf{Z}_{t-1}\mathbf{V}_{t-1}^{(m)}\right)\hat{\boldsymbol{\theta}}_i^{(m)}$, where $\hat{\boldsymbol{\theta}}_i^{(m)} = \left(\mathbf{V}_{t-1}^{(m)\top}\mathbf{Z}_{t-1}^\top\mathbf{Z}_{t-1}\mathbf{V}_{t-1}^{(m)}\right)\mathbf{V}_{t-1}^{(m)\top}\mathbf{Z}_{t-1}^\top r_{i,t}$. Equation (20) makes clear that PCR reduces dimensions by using m principal components that capture the largest common variation in the conditioning information set, and weights through $\boldsymbol{\omega}^{(m)}$ the original covariates. The m PCA components used to form the derived covariates in the final model is a parameter that needs to be chosen adaptively. In section 4.1 we discuss how this is done.

3.3 Partial Least Squares (PLS)

When PCR finds the components with the largest common variation in the conditioning information set, it takes into consideration in its objective function only the information from $\mathbf{Z}_{t-1}$. As Gu et al. (2018) argues, doing so it fails to incorporate the information from the dependent variable (excess returns) when performing dimension reduction, which can be suboptimal to generate good conditional means approximations. The Partial Least Squares (PLS), on the other hand, uses the common components from $\mathbf{Z}_{t-1}$ by conditioning on the joint distribution of $r_{i,t}$ and $\mathbf{Z}_{t-1}$. For this reason, PLS is considered a latent approach that models the covariance between the spaces generated by both matrices ($r_{i,t}$ and $\mathbf{Z}_{t-1}$).

As mentioned by Friedman et al. (2001) this method seeks to obtain the directions that provide the highest variance and highest correlation[9] between $r_{i,t}$ and $\mathbf{Z}_{t-1}$ Stone and Brooks (1990); Frank and Friedman (1993). Following Friedman et al. (2001), the idea of PLS is

[7] We dropped the time subscripts $t-1$ for simplicity.

[8] These matrices are known as left and right singular vectors of $\mathbf{Z}$ respectively.

[9] A clear contrast to PCR that only seeks directions that maximize only the variance between $r_{i,t}$ and $\mathbf{Z}_{t-1}$

to weight each vector $\mathbf{z}_{t-1,k} \in \mathbf{Z}_{t-1}$ by its partial least squares direction $\hat{\varphi}_k = \langle \mathbf{z}_{t-1,k}, r_{i,t} \rangle$. The derived input $\mathfrak{z}_k = \sum_k \widehat{\varphi}_k \mathbf{z}_{t-1,k}$ is then used to obtain the the estimated coefficient $\widehat{\theta}_k$ regressing $r_{i,t}$ on the derived input $\mathfrak{z}_k$. This is done from $k = 1, 2, \ldots, K$, where we orthogonalize the original predictors with respect to the previous component. This process repeats until $m \leq K$ desired directions to reduce dimensions have been found. Clearly, if we build $m = K$ directions, we return to the standard least squares estimates.

The idea of the partial least squares direction $\hat{\varphi}_k$ is to obtain the weight of the strength (or the partial sensitivity) of the univariate effect each variable in the conditioning set $\mathbf{Z}_{t-1}$ on $r_{i,t}$. As with the PCR, the m number of directions is a parameter that needs to be chosen appropriately.

3.4 Three-Pass Regression Filter (3PRF)

Kelly and Pruitt (2015) proposed the three-pass regression filter (3PRF), a generalization of the PLS estimator, in the sense that the latter is a special case of the former. In this setting, assuming that the data can be represented by an approximate factor model to reduce the dimension of the predictive information, the target variable are the time series of the excess returns of each risky asset i, having the following general form: $r^i_{t+h} = \beta^i_0 + \boldsymbol{\beta}^{i\top} \boldsymbol{F}^i_t + \eta^i_{t+h}$. Denoting by $\boldsymbol{r}^i$ the $T \times 1$ vector of the excess returns for asset i, we can write in matrix notation the target variable as:

$$\boldsymbol{r}^i = \begin{bmatrix} r^i_{t+1} \\ r^i_{t+2} \\ \vdots \\ r^i_{t+T} \end{bmatrix} \tag{21}$$

$$\boldsymbol{r}^i = \mathbf{1}\beta^i_0 + \boldsymbol{F}^i \boldsymbol{\beta}^i + \boldsymbol{\eta}^i \qquad i = 1, 2, \ldots, N \tag{22}$$

where $\mathbf{1}$ is again a T-dimensional vector of ones, β^i_0 is a scalar, $\boldsymbol{F}^i$ is a $T \times K_F$ matrix of factors, $\boldsymbol{\beta}^i$ is a K_F-dimensional vector of parameters, and $\boldsymbol{\eta}^i$ is a T-dimensional vector of errors, all defined for the risky asset i.

Denoting by $\mathbf{Z}$ the $T \times L$ matrix of conditioning information used as predictors, the cross-section of all the L variables forming the conditioning information is represented by $z_{l,t} = \phi_{l,0} + \boldsymbol{\phi}_l^\top \boldsymbol{F}_t + \epsilon_{l,t}$, for $l = 1, 2, \ldots, L$. Defining $\boldsymbol{z}_t = (z_{1,t}, z_{2,t}, \ldots, z_{L,t})^\top$, then the matrix representation of $\boldsymbol{Z}$ is given by

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix} = \begin{bmatrix} z_{1,1} & z_{2,1} & \dots & z_{3,1} \\ z_{1,2} & z_{2,2} & \dots & z_{3,2} \\ \vdots & \vdots & \dots & \vdots \\ z_{1,T} & z_{2,T} & \dots & z_{3,T} \end{bmatrix} \tag{23}$$

$$= \mathbf{1}\boldsymbol{\phi}_0^\top + \boldsymbol{F}\boldsymbol{\Phi}^\top + \boldsymbol{\varepsilon} \tag{24}$$

where $\mathbf{1}$ is a T-dimensional vector of ones, $\boldsymbol{\phi}_0$ is a $L \times 1$ vector of intercepts, $\boldsymbol{F}$ is the $T \times K_F$ matrix of latent factors, $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_L]^\top$ with dimension $L \times K_F$, and $\boldsymbol{\varepsilon}$ is a T-dimensional vector of errors.

For each risky asset i in $i = 1, 2, \dots, N$ there are three steps. The first pass of the estimator runs L separate time series regressions (for a fixed risky asset i), i.e., one for each variable forming our conditioning information on the forecast target:

$$z_{l,t}^i = \hat{\phi}_{l,0}^i + \hat{\phi}_l^i r_{t+h}^i + e_{l,t}^i \tag{25}$$

The estimated coefficients $\hat{\phi}_l$ describe the sensitivity of conditioning information (predictors) to the latent factor driving the forecast target.

In the second pass, we use the estimated first-pass coefficients in T separate cross section regressions, where our conditioning information variable l are the dependent variable, and the first-stage loadings $\hat{\phi}_l^i$ are the independent variables

$$z_{l,t}^i = \hat{c}_t^i + \boldsymbol{F}_t^i \hat{\phi}_l^i + u_{l,t}^i \tag{26}$$

The first-stage coefficient estimates map the cross-sectional distribution of predictors to the latent factors. Second-stage cross section regressions use this map to back out estimates of the factors at each point in time ((Kelly and Pruitt, 2015)).

Table 1: The Three Pass Regression Filter (3PRF)

Pass	Description
1	run time series regression of $\boldsymbol{z}_l^i$ on $\mathbf{r}^i$ for $j = 1, 2, \dots, L$ predictors
2	run cross section regression of $\boldsymbol{z}_l^i$ on $\hat{\boldsymbol{\phi}}_l^i$ for $t = 1, 2, \dots, T$
3	run time series regression of r_{t+1}^i on predictive factors $\hat{\mathbf{F}}_t^i$

.

The third stage is the final forecasting step. Running a single time series regression of the target variable r^i_{t+1} on the second-pass estimated factors $\hat{\boldsymbol{F}}^i_t$, it provides the forecast, which is the fitted value $r^i_{t+1} = \hat{\beta}^i_0 + \hat{\boldsymbol{F}}^{i\top}_t \hat{\boldsymbol{\beta}}^i$. The algorithm is summarized in table 1. Kelly and Pruitt (2015) prove that, under a set of assumptions[10] the 3PRF is a consistent estimator as L and T become large. The authors also show that the 3PRF has a one-step closed form:

$$\hat{\boldsymbol{r}}^i = \mathbf{1}_T \bar{\boldsymbol{r}}^i + \boldsymbol{J}_T \boldsymbol{Z} \boldsymbol{J}_L \boldsymbol{Z}^\top \boldsymbol{J}_T \boldsymbol{r}^i \left(\boldsymbol{r}^{i\top} \boldsymbol{J}_T \boldsymbol{Z} \boldsymbol{J}_L \boldsymbol{Z}^\top \boldsymbol{J}_T \boldsymbol{Z} \boldsymbol{J}_L \boldsymbol{Z}^\top \boldsymbol{J}_T \boldsymbol{r} \right)^{-1} \boldsymbol{r}^{i\top} \boldsymbol{J}_T \boldsymbol{Z} \boldsymbol{J}_L \boldsymbol{Z}^\top \boldsymbol{J}_T \boldsymbol{r}^i \quad (27)$$

where $\bar{\boldsymbol{r}}^i = \mathbf{1}^\top_T \boldsymbol{r}^i / T$, for $\mathbf{1}_T$ a T-dimensional vector of ones, and $\boldsymbol{J}_T \equiv \boldsymbol{I}_T - \frac{1}{T} \mathbf{1}_T \mathbf{1}^\top_T$, for $\boldsymbol{I}_T$ a T-dimensional identity matrix, being $\boldsymbol{J}_L$ analogous.

4 Empirical Strategy

Our focus is from July 1963 to December 2017 for U.S. data. The out-of-sample cut off starts in Jan-1996 and we use recursive windows to update the estimations with the new available information. Using different estimators to impose dimensionality reduction, we compare the performance of each approach to build CMV, UMV and MVATE efficient portfolios, as presented on section 2.1.

At each t we form the weights of our portfolios using equation (9) for the CMV, equation (10) for the UMV, and equation (12) for the UMVTE portfolio. For each one of them, we use the estimators presented in section 3 to obtain the conditional expected excess returns $\boldsymbol{\mu}_t$. To exploit the wealth of information available for predictors, we make use of high dimensional datasets as our conditioning information. Hence, we assess how PCR, PLS, 3PRF, LASSO, RIDGE, and the ENet behave, comparing them with our benchmark estimator, the OLS.

Thus, for all three methodologies we replace $\boldsymbol{\mu}_t(\mathbf{Z}_{t-1})$ by its sample counterpart $\widehat{\boldsymbol{\mu}}_t(\mathbf{Z}_{t-1})$ on $\mathbf{x}_t(\mathbf{Z}_t)$, being $\widehat{\boldsymbol{\mu}}_t(\mathbf{Z}_{t-1})$ the fitted values of the estimation of the excess returns of all N risky assets in the investment set on the L lagged instruments. Here we have an important difference from previous studies. Ferson and Siegel (2009), Fletcher and Basu (2016), Abhyankar et al. (2012), Chiang (2015) among others have used standard OLS to obtain the fitted values, and a small set of instruments selected in advance as conditioning information.

Given that returns of risky assets are assumed to be generated by the conditional mean plus a noise term as in equation (4), we can estimate the conditional second moment matrix as which is given by:

[10]For further details, please refer to the original paper.

$$\begin{aligned}\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right) &= \mathbb{E}\left(\mathbf{r}_t\mathbf{r}_t^{\top}|\mathbf{Z}_{t-1}\right) \\ &= \boldsymbol{\Sigma}_t\left(\mathbf{Z}_{t-1}\right) + \boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^{\top}\end{aligned} \tag{28}$$

where $\mathbf{r}_t$ is an $N \times 1$ vector of the N risky assets, and $\boldsymbol{\Sigma}_t\left(\mathbf{Z}_{t-1}\right)$ is the nonsigular conditional covariance matrix of the noise. Plugging the estimates of the conditional expected excess returns on $\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)$, we can estimate the second moment matrix, assuming that the conditional covariance matrix $\boldsymbol{\Sigma}_t\left(\mathbf{Z}_{t-1}\right)$ is constant[11]. Notice, that $\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)$ is time-varying due to the time-varying characteristic of $\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)$. We estimate the conditional covariance matrix $\boldsymbol{\Sigma}_t\left(\mathbf{Z}_{t-1}\right)$ as the estimate of the residual covariance matrix obtained with the regressions for $\hat{\boldsymbol{\mu}}_t\left(\mathbf{Z}_{t-1}\right)$.

Following Fletcher and Basu (2016), we also calculate $\mathbb{E}\left(\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^{\top}\left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1}\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)\right)$ on equation (10) as the average value of $\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)^{\top}\left(\boldsymbol{\Gamma}_t\left(\mathbf{Z}_{t-1}\right)\right)^{-1}\boldsymbol{\mu}_t\left(\mathbf{Z}_{t-1}\right)$ during the estimation window. Following Kirby and Ostdiek (2012) and Fletcher and Basu (2016) we set a small monthly target return. Thus, in equation (9) for the CMV, we set the conditional expected excess returns $\mathbb{E}\left(r_{p,t+1}|\mathbf{Z}_t\right)$ equal to 0.5%. Similarly, in equation (10) for the UMV framework, we also set the unconditional expected excess returns $\mathbb{E}\left(r_{p,t+1}\right)$ equal to 0.5%. For the UMVTE, we set α_p to be $\mathbb{E}\left(R_{p,t+1} - R_{b,t+1}\right) = 0.5\%$, where we consider the tracking portfolio to be the risk free asset of the economy.

4.1 Validation

All the techniques described in section 3 have hyperparameters that need to be chosen to define the final model. This process of determining the best hyperparameters is commonly known as tuning. As previously mentioned, penalized regressions do not have a closed form solution mainly because of the penalty λ imposed in the objective function. For ENet, λ and α need to be chosen, while for LASSO and RIDGE, only λ. For PCR, the number of principal components m in $\mathbf{V}^{(m)} = \left[\boldsymbol{\omega}^{(1)}, \ldots, \boldsymbol{\omega}^{(m)}\right]$ need to be defined to determine the final regression model. For PLS, the m number of principal directions should also be chosen[12].

To tune the parameters above, we recursively define the training and testing periods. The prevailing approach in this literature is to use the tunning subsample to adaptively determine the hyperparameters. At each $\tau \in \tau_{OOS}$, where τ_{OOS} is the OOS subsample, we use all the previous information up to $\tau - 1$ to build the training group. Within this group, the data ranges from $t = 1, 2, \ldots, \tau - 1$. As Bergmeir et al. (2018) point out the usage for

[11] Previous studies such as Ferson and Siegel (2009) and Fletcher and Basu (2016) have found that time-varying conditional covariance matrix leads to marginal changes.

[12] For the 3PRF, given its complexity, we pre-determined a small number of three latent factors to impose sparsity.

time-series applications, we decided to make use of K-fold cross-validation in the training subsamples. For a list of possible different hyperparameters of each estimator (i.e., whether λ, α, or the number of components m), the mean square error (MSE) is computed in this training sample. The hyperparameter(s) value(s) that produce(s) the lowest MSE is(are) chosen.

This recursion scheme allows us to incorporate the most recent data from the set of conditioning information $\mathbf{Z}$ when enlarging the in-sample period (training subsample) by one t (i.e., one month), while dynamically selecting the best hyperparameter(s) for a given objective function. Finally, the observation τ is used for testing using the best-tunned model estimates for all techniques.

4.2 Evaluating the Estimated Conditional Means

Since we use several different approaches to impose sparsity when forming the conditional means, we are interested to assess the characteristics of the generated conditional means by each estimator. We evaluate the predictive performance for each asset i using out-of-sample R^2 given by

$$R_i^2 = 1 - \frac{\sum_{t \in \tau_{OOS}} \left(r_{i,t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{t \in \tau_{OOS}} \left(r_{i,t+1}^2\right)} \tag{29}$$

where τ_{OOS} represents the OOS periods not used for testing or tuning. Usually, the OOS R^2 is computed demeaning the denominator. As Gu et al. (2018) argues, doing so can be a mistaken analysis, since the historical averages usually underperform a naive forecast of zero. A way to overcome this issue is to compare against a benchmark of zero, as in equation (29).

It is known that the R^2 are generally extremely low when dealing with forecasting excess returns. To assess if the OOS R^2 generated by different approaches deliver economically meaningful results, we follow Campbell and Thompson (2007); Gu et al. (2018) and calculate the implied SR (SR*). The implied SR seeks to measure the improvement produced in the SR when an investor can exploit signals from the conditioning information set $\mathbf{Z}$. Under mild assumptions[13], the implied SR is given by

$$SR^* = \sqrt{\frac{SR^2 + R^2}{1 - R^2}} \tag{30}$$

The interpretation is straightforward. In the case the conditioning information produces

[13] Assuming an investor with a mean-variance preferences, single-period horizon, and γ risk aversion coefficient.

valuable signal, the OOS R^2 is large relative to SR^2, so the investor can exploit the information in $\mathbf{Z}$ to obtain a large proportional increase in portfolio return.

4.3 Datasets

As it is standard in this literature, we use size- and value- sorted and grouped portfolios from Ken French's website to represent our universe of risky assets of U.S. stocks. The chosen number N of available portfolios are 25 and 100. In the Internet Appendix, we extend the analysis for other portfolios with a smaller number of assets. Precisely, we assess *5 Industry Portfolios* and *6 portfolios formed on size and boot-to-market*. Summary statistics tables for all portfolios are also presented in the Internet Appendix[14].

With high-dimensional datasets and a vast number of suggested predictors, we use several instruments and split them into groups. The first group is formed by the predictive candidates used in Welch and Goyal (2007) and made available on Amit Goyal's website. We consider a set of 10 variables: book-to-market (b/m), default return spread (dfr), default yield spread (dfy), inflation ($infl$), long-term yield (lty), long-term rate of returns (ltr), net equity expansion ($ntis$), stock variance ($svar$), term spread (tms) and treasury bill rate (tbl).

The second group makes use of the FRED-MD dataset as presented in McCracken and Ng (2016). The FRED-MD is a large macroeconomic database and monthly updated by the FRED[15] that shares the predictive content of the Stock-Watson dataset (Stock and Watson (1996)). It is a balanced panel consisting of 128 macroeconomic and financial variables. The variables are split into 8 groups: (1) output and income, (2) labor market, (3) housing, (4) consumption, orders, and inventories, (5) money and credit, (6) interest and exchange rates, (7) prices, and (8) stock market. In order to mitigate the consequences of the presence of cointegration, we apply the suggested transformations to obtain stationary time series. The transformations can be grouped in seven categories: (i) no transformation; (ii) Δx_t; (iii) $\Delta^2 x_t$; (iv) $log(x_t)$, (v) $\Delta log(x_t)$, (vi) $\Delta^2 log(x_t)$, and (vii) $\Delta(x_t/x_{t-1} - 1)$[16].

The third group is based on the economic policy uncertainty measure (EPU) from Baker et al. (2016) and related indexes. The EPU is an index that proxies for movements in policy-related economic uncertainty for U.S., being based on newspaper coverage frequency. In the sense of the EPU, we also use the financial stress indicator (FSI) for the U.S from Püttmann (2018). The essence of the FSI is being an indicator of negative financial sentiment. It is based on the reporting in five major US newspapers[17]. Püttmann (2018) shows that the FSI

[14]Tables C.2 to C.5 in the Internet Appendix.

[15]https://research.stlouisfed.org/econ/mccracken/fred-databases/

[16]Table C.1 in the Internet Appendix C presents the entire list of variables, groups, corresponding transformations, and sample means and standard deviations for the full sample, IS and OOS.

[17]Boston Globe, Chicago Tribune, Los Angeles Times, Wall Street Journal, and Washington Post.

is a robust indicator, such that an increase in negative financial sentiment is followed by a fall in output, higher unemployment, lower stock market returns, and rising corporate bond spreads.

5 Empirical Results

5.1 Out-of-Sample Analysis - Sharpe Ratios

5.1.1 Performance Evaluation

To answer whether we can exploit signals provided by a large set of conditioning information using penalized estimators to impose sparsity or latent factors models with small numbers of factors, we first evaluate how the estimators behave with different sets of conditioning information to provide mean-standard deviation ratios of the returns in the out-of-sample data.

Table 2 shows the Sharpe ratios of the seven estimators for all three mean-variance efficient portfolio frameworks. For the 25 Portfolios formed on size and BTM, we see that only the OLS failed to deliver a significant SR for the MVATE with large sets of conditioning information. Overall we note that all the other estimators provided similar SR, while the PCR and the PLS alternate to deliver a higher ratio to the remaining estimators. Using Goyal's instruments we that OLS, 3PRF, LASSO, RIDGE, and ENet produce similar SR. As expected, OLS cannot handle large set of covariates, what makes the SR drop whenever we increase the set of conditioning information.

For the 100 Portfolios formed on size and BTM, a similar pattern stands out. OLS cannot extract efficiently information from a large set of instruments, making its SR to be considerably lower and not statistically significant (in most cases) compared to the rest of the estimators. Again, we notice that the PCR and the PLS deliver higher SR in most cases. Differently from the case with a smaller number of risky assets, we see that all estimators, except PLS and PCR, failed to extract signals and deliver statistically significant SRs when using a small set of instruments (Goyal). An important point to notice is that, in general, as we move to a high-dimensional setting, increasing the number of instruments, the dimensionality reduction approaches deliver a higher SR.

Table 2: Out-of-Sample Sharpe ratios delivered by each estimator and set of conditioning information

	25 Portfolios Formed on Size and Book-to-Market						100 Portfolios Formed on Size and Book-to-Market					
	CMV		UMV		MVATE		CMV		UMV		MVATE	
Estimator	SR	p-val	SR	p-val	SR	p-val	SR	p-val	SR	p-val	SR	p-val
Panel A: Goyal												
OLS	0.307	0.000	0.293	0.000	0.270	0.000	0.076	0.221	0.072	0.245	0.003	0.962
3PRF	0.327	0.000	0.330	0.000	0.307	0.000	0.117	0.060	0.133	0.032	0.059	0.341
PLS	0.307	0.000	0.322	0.000	0.300	0.000	0.245	0.000	0.241	0.000	0.279	0.000
PCR	0.383	0.000	0.386	0.000	0.380	0.000	0.303	0.000	0.315	0.000	0.312	0.000
LASSO	0.276	0.000	0.279	0.000	0.278	0.000	0.087	0.158	0.100	0.107	0.084	0.173
RIDGE	0.313	0.000	0.311	0.000	0.299	0.000	0.060	0.333	0.083	0.178	0.044	0.472
ENET	0.315	0.000	0.316	0.000	0.259	0.000	0.097	0.117	0.103	0.096	0.043	0.483
Panel B: FRED-MD												
OLS	0.161	0.010	0.166	0.008	0.006	0.924	0.128	0.040	0.128	0.040	0.047	0.446
3PRF	0.282	0.000	0.262	0.000	0.261	0.000	0.210	0.001	0.197	0.002	0.135	0.029
PLS	0.296	0.000	0.299	0.000	0.285	0.000	0.350	0.000	0.346	0.000	0.230	0.000
PCR	0.371	0.000	0.361	0.000	0.374	0.000	0.203	0.001	0.209	0.001	0.198	0.002
LASSO	0.215	0.001	0.209	0.001	0.212	0.001	0.312	0.000	0.305	0.000	0.240	0.000
RIDGE	0.274	0.000	0.277	0.000	0.275	0.000	0.274	0.000	0.265	0.000	0.188	0.003
ENET	0.280	0.000	0.267	0.000	0.286	0.000	0.307	0.000	0.301	0.000	0.129	0.037
Panel C: All Instruments												
OLS	0.201	0.003	0.204	0.003	-0.054	0.419	0.118	0.078	0.118	0.078	-0.026	0.694
3PRF	0.290	0.000	0.278	0.000	0.254	0.000	0.240	0.000	0.228	0.000	0.145	0.020
PLS	0.331	0.000	0.303	0.000	0.314	0.000	0.309	0.000	0.315	0.000	0.263	0.000
PCR	0.307	0.000	0.293	0.000	0.310	0.000	0.269	0.000	0.272	0.000	0.271	0.000
LASSO	0.231	0.000	0.230	0.000	0.232	0.000	0.323	0.000	0.315	0.000	0.253	0.000
RIDGE	0.305	0.000	0.306	0.000	0.299	0.000	0.243	0.000	0.247	0.000	0.117	0.079
ENET	0.302	0.000	0.280	0.000	0.302	0.000	0.359	0.000	0.352	0.000	0.179	0.008

Table 2 summarises the OOS (Jan-1996 to Dec-2017) Sharpe ratios (SR) by estimator and optimal portfolio framework (CMV, UMV and MVATE) for both portfolios. Panel A reports the Sharpe ratios generated when the variables from Goyal's website are used as **Z**. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. Panel B presents the Sharpe ratios obtained using the FRED-MD variables. The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, panel C shows the Sharpe rations when all variables are used as conditioning information. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). The p-val is the p-value from the two-sided test of the SR.

5.1.2 Statistical Inference for the Difference of Sharpe Ratios

To answer if the Sharpe ratios generated by an estimator is statistically different from another SR generated by any other estimator, we need to test the difference between them. We follow Ledoit and Wolf (2008) to perform these tests. In short, consider the Sharpe ratios of the optimal portfolios generated by two different estimators, say a and b. The difference between SR_a and SR_b is given by

$$\Delta = SR^{(a)} - SR^{(b)} = \frac{\mu^{(a)}}{\sigma^{(a)}} - \frac{\mu^{(b)}}{\sigma^{(b)}} \tag{31}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the OOS unconditional mean and variance of the excess returns from the optimal portfolio generated by an estimator. Ledoit and Wolf (2008) suggest to construct a studentized time series bootstrap confidence interval for the difference of the SRs. This method has been shown to be robust when returns have tails heavier than the normal distribution or are of time series nature. The bootstrap data is generated using the circular block bootstrap of Politis and Romano (1992). The two-sided distribution function of the studentized statistic can be obtained via bootstrap as follows:

$$f\left(\frac{|\widehat{\Delta} - \Delta|}{se\left(\widehat{\Delta}\right)}\right) \approx f\left(\frac{|\widehat{\Delta}^{\text{boot}} - \Delta|}{se\left(\widehat{\Delta}^{\text{boot}}\right)}\right) \tag{32}$$

where $f(\cdot)$ is the distribution of a random variable, Δ is populational difference between SR_a and SR_b, $\widehat{\Delta}$ is the sample counterpart of this difference obtained in the data in the estimation window, and $\widehat{\Delta}^{\text{boot}}$ is the estimated difference computed from bootstrap. The standard errors are denoted by $se(\cdot)$. Out of the distribution obtained from the bootstrap in the equation (32), we can find the confidence interval for Δ and the p-values of the test[18].

Table 3 reports for the dataset with 100 portfolios formed on Size/BTM. Table 4 presents the test of difference of SRs for the dataset with 25 portfolios formed on Size/BTM[19]. In order to see that the Sharpe ratios generated using not only two different estimators, but also different set of conditioning information, these tables also present the tests of difference between these cases too.

Inspecting table 3, we notice that in the case with 100 risky assets PLS and PCR indeed generate different SRs to all the other 5 estimators when using the Goyal variables as conditioning information. The p-values of the difference of Sharpe ratios are always low, providing a strong evidence to this difference. This fact is valid for all three mean-variance

[18]For further details in this procedure, see Ledoit and Wolf (2008).

[19]In the Internet Appendix, tables B.2 and table B.3 report the same tests for the dataset with 6 portfolios formed on Size/BTM and the dataset with 5 industry portfolios respectively.

Table 3: Test for the difference of the Sharpe ratios - OOS (Jan-1996 - Dec-2017) - 100 Portfolios

Panel A: CMV

Goyal		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.092	0.001*	0.000*	0.594	0.428	0.358	0.023	0.000*	0.042	0.004*	0.059	0.002*	0.004*	0.000*	0.005*	0.001*	0.038	0.000*
	3PRF		0.001*	0.000*	0.105	0.002*	0.322		0.000*	0.092	0.030	0.165	0.021		0.003*	0.010	0.010	0.183	0.005*
	PLS			0.111	0.001*	0.000*	0.002*			0.204	0.596	0.819	0.579			0.872	0.654	0.806	0.451
	PCR				0.000*	0.000*	0.001*				0.928	0.800	0.959				0.847	0.371	0.877
	LASSO					0.087	0.407					0.113	0.013					0.116	0.002*
	RIDGE						0.022						0.003*						0.000*

FRED-MD		FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.166	0.001*	0.425	0.022	0.138	0.022	0.070	0.014	0.128	0.002*	0.109	0.002*
	3PRF		0.003*	0.930	0.102	0.469	0.112		0.122	0.515	0.046	0.705	0.023
	PLS			0.024	0.587	0.406	0.510			0.031	0.241	0.043	0.582
	PCR				0.317	0.561	0.291				0.309	0.818	0.194
	LASSO					0.564	0.901					0.034	0.670
	RIDGE						0.578						0.347

All Instr.		All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.037	0.010	0.107	0.003*	0.132	0.002*
	3PRF		0.343	0.850	0.200	0.847	0.101
	PLS			0.591	0.818	0.348	0.443
	PCR				0.585	0.815	0.390
	LASSO					0.087	0.187
	RIDGE						0.011

Panel B: UMV

Goyal		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.013	0.001*	0.000*	0.224	0.590	0.190	0.029	0.000*	0.023	0.005*	0.063	0.002*	0.005*	0.001*	0.002*	0.001*	0.030	0.000*
	3PRF		0.006*	0.001*	0.087	0.013	0.138		0.000*	0.137	0.069	0.255	0.048		0.028	0.022	0.035	0.254	0.017
	PLS			0.082	0.001*	0.000*	0.004*			0.311	0.614	0.856	0.594			0.978	0.690	0.860	0.476
	PCR				0.001*	0.000*	0.002*				0.917	0.632	0.869				0.660	0.284	0.961
	LASSO					0.349	0.788					0.169	0.029					0.165	0.005*
	RIDGE						0.184						0.016						0.002*

FRED-MD		FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.230	0.001*	0.380	0.027	0.161	0.025	0.097	0.013	0.108	0.004*	0.094	0.003*
	3PRF		0.001*	0.885	0.082	0.436	0.078		0.074	0.356	0.035	0.489	0.016
	PLS			0.037	0.559	0.358	0.477			0.023	0.178	0.040	0.493
	PCR				0.373	0.639	0.334				0.375	0.820	0.234
	LASSO					0.519	0.915					0.053	0.680
	RIDGE						0.530						0.337

All Instr.		All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.053	0.012	0.090	0.004*	0.116	0.003*
	3PRF		0.236	0.669	0.176	0.941	0.085
	PLS			0.608	0.992	0.280	0.546
	PCR				0.660	0.810	0.433
	LASSO					0.133	0.174
	RIDGE						0.017

Panel C: MVATE

Goyal		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.030	0.000*	0.000*	0.082	0.238	0.174	0.019	0.001*	0.020	0.000*	0.011	0.077	0.009*	0.001*	0.002*	0.000*	0.102	0.018
	3PRF		0.000*	0.000*	0.493	0.601	0.503		0.002*	0.053	0.009*	0.105	0.349		0.016	0.005*	0.006*	0.449	0.137
	PLS			0.413	0.000*	0.000*	0.000*			0.101	0.668	0.321	0.081			0.371	0.519	0.094	0.243
	PCR				0.001*	0.000*	0.000*				0.385	0.150	0.062				0.320	0.040	0.138
	LASSO					0.102	0.197					0.273	0.577					0.786	0.413
	RIDGE						0.963						0.275						0.176

FRED-MD		FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.081	0.137	0.282	0.008*	0.031	0.358	0.079	0.003*	0.130	0.005*	0.357	0.127
	3PRF		0.260	0.547	0.028	0.295	0.928		0.047	0.221	0.024	0.631	0.583
	PLS			0.537	0.901	0.662	0.275			0.986	0.853	0.185	0.421
	PCR				0.678	0.926	0.528				0.737	0.480	0.800
	LASSO					0.311	0.059					0.012	0.283
	RIDGE						0.374						0.796

All Instr.		All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET
	OLS	0.017	0.000*	0.041	0.001*	0.028	0.025
	3PRF		0.078	0.276	0.052	0.521	0.688
	PLS			0.951	0.856	0.020	0.274
	PCR				0.872	0.224	0.464
	LASSO					0.010	0.143
	RIDGE						0.321

p-val	$[0 \leq p \leq 0.05]$	$(0.05 < p < 0.1)$	$(0.1 \leq p \leq 1]$
Gradient color bounds			

Test for the differences of the Sharpe ratios of the OOS (Jan-1996 - Dec-2017) returns of the efficient portfolios formed from the dataset with 100 Size/BTM portfolios using 7 different estimators (OLS, 3PRF, PLS, PCR, LASSO, RIDGE and ENet) and three different set of conditioning information(Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios. Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

Table 4: Test for the difference of the Sharpe ratios - OOS (Jan-1996 - Dec-2017) - 25 Portfolios

Panel A: CMV

		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
Goyal	OLS	0.487	0.998	0.153	0.592	0.862	0.836	0.715	0.871	0.280	0.116	0.567	0.650	0.783	0.655	0.397	0.148	0.328	0.819
	3PRF		0.677	0.234	0.354	0.732	0.790		0.638	0.449	0.082	0.353	0.447		0.585	0.333	0.098	0.281	0.525
	PLS			0.066	0.673	0.916	0.898			0.149	0.151	0.518	0.643			0.273	0.180	0.312	0.875
	PCR				0.160	0.257	0.314				0.008*	0.034	0.053				0.009*	0.023	0.113
	LASSO					0.553	0.311					0.975	0.962					0.629	0.692
	RIDGE						0.949						0.608						0.790
FRED-MD	OLS							0.077	0.112	0.007*	0.454	0.099	0.111	0.053	0.116	0.176	0.436	0.123	0.047
	3PRF								0.833	0.092	0.220	0.797	0.971		0.582	0.837	0.340	0.783	0.693
	PLS									0.163	0.234	0.719	0.807			0.803	0.310	0.799	0.951
	PCR										0.008*	0.033	0.074				0.012	0.103	0.252
	LASSO											0.213	0.107					0.094	0.022
	RIDGE												0.900						0.406
All Instr.	OLS													0.139	0.139	0.190	0.352	0.127	0.079
	3PRF														0.697	0.976	0.265	0.998	0.879
	PLS															0.724	0.302	0.678	0.991
	PCR																0.415	0.975	0.670
	LASSO																	0.377	0.104
	RIDGE																		0.604

Panel B: UMV

		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
Goyal	OLS	0.190	0.510	0.074	0.783	0.611	0.539	0.652	0.931	0.267	0.165	0.785	0.665	0.802	0.572	0.496	0.233	0.535	0.714
	3PRF		0.847	0.211	0.301	0.612	0.718		0.639	0.608	0.071	0.382	0.330		0.364	0.327	0.096	0.311	0.307
	PLS			0.144	0.532	0.819	0.927			0.433	0.074	0.413	0.344			0.211	0.085	0.231	0.415
	PCR				0.116	0.199	0.250				0.006*	0.054	0.025				0.007*	0.033	0.039
	LASSO					0.581	0.312					0.975	0.870					0.775	0.967
	RIDGE						0.873						0.516						0.573
FRED-MD	OLS							0.189	0.093	0.012	0.564	0.109	0.181	0.106	0.223	0.240	0.507	0.106	0.120
	3PRF								0.568	0.074	0.383	0.662	0.939		0.604	0.677	0.588	0.345	0.745
	PLS									0.189	0.192	0.681	0.616			0.485	0.254	0.674	0.640
	PCR										0.013	0.049	0.078				0.018	0.119	0.167
	LASSO											0.183	0.190					0.057	0.063
	RIDGE												0.841						0.766
All Instr.	OLS													0.223	0.272	0.285	0.518	0.135	0.195
	3PRF														0.888	0.990	0.373	0.737	0.997
	PLS															0.892	0.454	0.964	0.994
	PCR																0.465	0.806	0.875
	LASSO																	0.217	0.247
	RIDGE																		0.948

Panel C: MVATE

		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
Goyal	OLS	0.262	0.585	0.070	0.890	0.499	0.863	0.903	0.831	0.114	0.376	0.935	0.796	0.803	0.935	0.885	0.495	0.736	0.747
	3PRF		0.891	0.150	0.599	0.850	0.455		0.764	0.289	0.157	0.590	0.740		0.721	0.673	0.232	0.504	0.792
	PLS			0.056	0.755	0.985	0.600			0.092	0.172	0.620	0.810			0.403	0.230	0.338	0.916
	PCR				0.175	0.198	0.143				0.009*	0.041	0.072				0.012	0.028	0.126
	LASSO					0.728	0.753					0.969	0.915					0.628	0.723
	RIDGE						0.484						0.843						0.959
FRED-MD	OLS							0.000*	0.001*	0.000*	0.016	0.000*	0.001*	0.001*	0.001*	0.001*	0.007*	0.001*	0.000*
	3PRF								0.720	0.057	0.417	0.736	0.683		0.521	0.511	0.625	0.499	0.472
	PLS									0.112	0.297	0.862	0.994			0.949	0.399	0.834	0.955
	PCR										0.007*	0.035	0.091				0.011	0.081	0.212
	LASSO											0.205	0.092					0.112	0.026
	RIDGE												0.828						0.420
All Instr.	OLS													0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
	3PRF														0.535	0.541	0.706	0.498	0.552
	PLS															0.950	0.436	0.811	0.874
	PCR																0.377	0.846	0.804
	LASSO																	0.434	0.131
	RIDGE																		0.610

p-val	$[0 \leq p \leq 0.05]$	$(0.05 < p < 0.1)$	$(0.1 \leq p \leq 1]$
Gradient color bounds			

Test for the differences of the Sharpe ratios of the OOS (Jan-1996 - Dec-2017) returns of the efficient portfolios formed from the dataset with 25 Size/BTM portfolios using 7 different estimators (OLS, 3PRF, PLS, PCR, LASSO, RIDGE and ENet) and three different set of conditioning information(Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios. Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

framework. Combined with the results from table 2, we can claim that the SRs of 0.245, 0.241 and 0.279 (CMV, UMV, MVATE respectively) for the PLS, and the SRs of 0.303, 0.315 and 0.312 (CMV, UMV, MVATE respectively) produced by the PCR are indeed higher than the remaining ones.

5.2 Characteristics and the Distribution of Optimized Portfolio Weights

In order to understand why a specific estimator can generate mean-variance optimized portfolio with higher return-volatility ratios we first analyze how the weights of the optimized portfolio behave for each estimator and set of conditioning information. For each month t in the out-of-sample period, each estimator produced a set of weights depending on the matrix $\mathbf{Z}$ used. We pooled the N weights for each t and plotted the distribution of these weights in separate figures for each dataset employed. Figure 1 plots the weights for the dataset with 25 portfolios formed on Size/BTM. Figure 2 shows the distribution of the weights $\mathbf{x}_t(\mathbf{Z}_t)$ for the dataset with 100 portfolios formed on Size/BTM[20]. Overall we see that across all three efficient strategies to build portfolios, PCR generates weights with higher variance than compared to the other estimators. Another point that can be inferred from these figures is that, in general, most of the estimators generate symmetric weights distributions. However, PCR does not, having in most cases a right-skewed distribution. We also notice that OLS is the estimator that does not respond much to predictive information, a fact that can be seen from its highly concentrated distribution around zero.

[20]In the Internet Appendix, we present the weights for the dataset with 6 portfolios formed on Size/BTM and the dataset with 5 industry portfolios.

Figure 1: Distribution of Optimized Portfolio Weights - 25 Portfolios Formed on Size and Book-to-Market

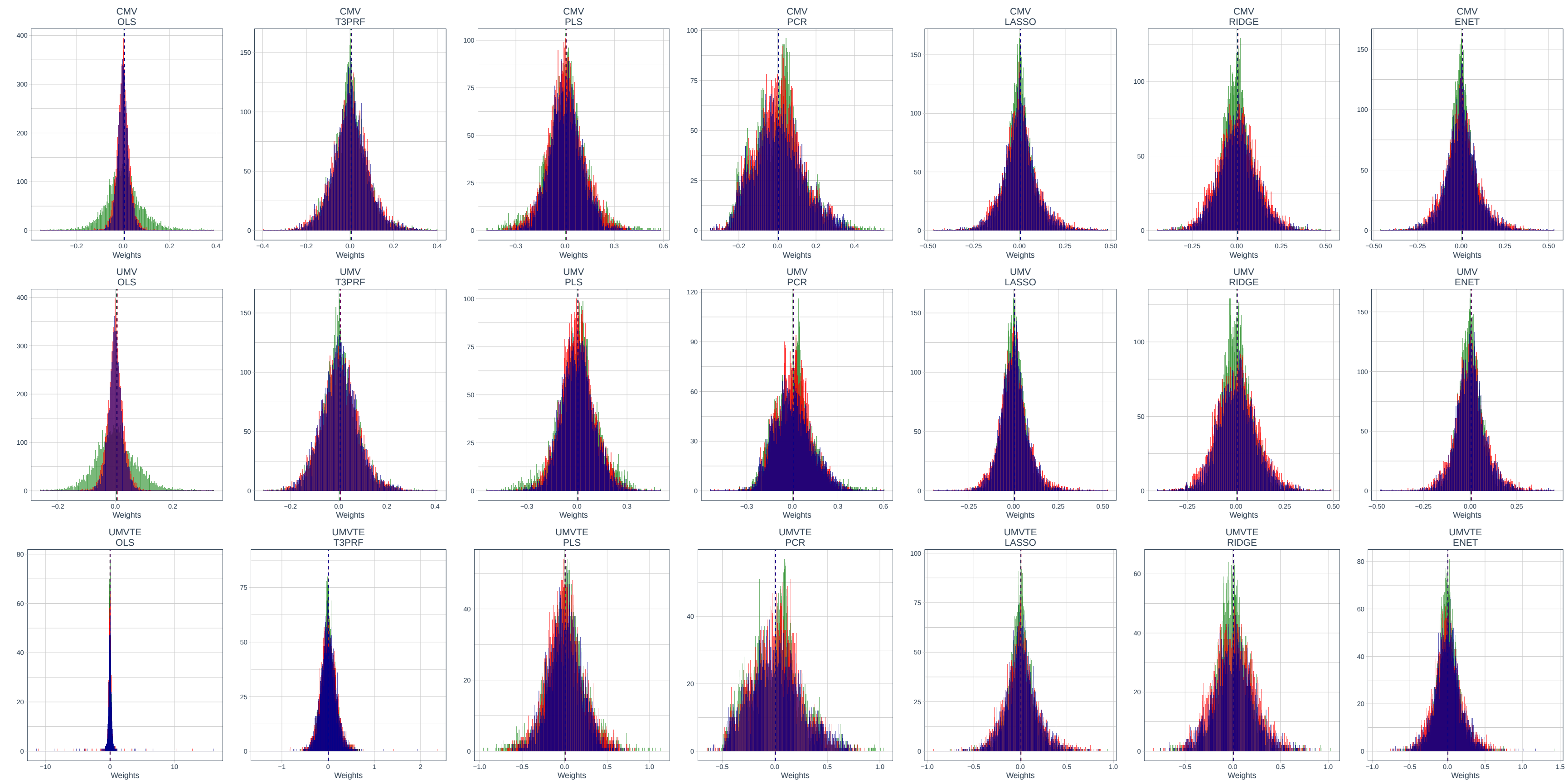


Figure 1 plots the distribution of the optimized portfolios weights generated by each estimator (columns) and mean-variance optimal framework (rows) for the 25 portfolios formed on Size/BTM. The first row reports the CMV strategies, the second row reports the UMV, and the third one plots the UMVTE. The set of conditioning information used in **Z** are plotted in different colors: (i) Goyal's in green, (ii) FRED-MD in red, and (iii) "All Instruments" in blue. Goyal's variables comprises: *b/m*, *dfr*, *dfy*, *infl*, *ltr*, *lty*, *ntis*, *svar*, *tms* and *tbl*. FRED-MD is a large dataset containing 128 macroeconomic and financial variables. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

Figure 2: Distribution of Optimized Portfolio Weights - 100 Portfolios Formed on Size and Book-to-Market

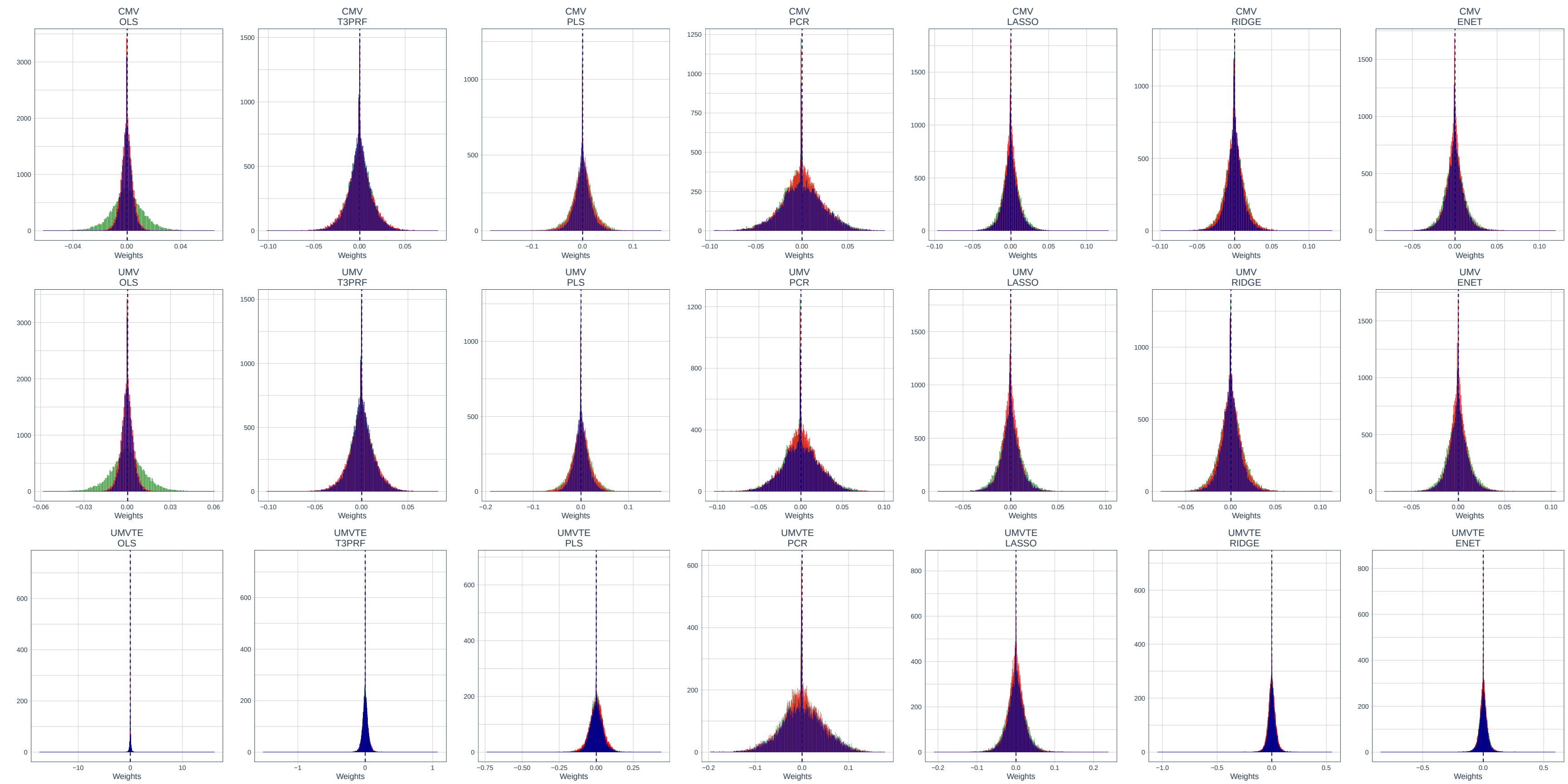


Figure 2 plots the distribution of the optimized portfolios weights generated by each estimator (columns) and mean-variance optimal framework (rows) for the 100 portfolios formed on Size/BTM. The first row reports the CMV strategies, the second row reports the UMV, and the third one plots the UMVTE. The set of conditioning information used in **Z** are plotted in different colors: (i) Goyal's in green, (ii) FRED-MD in red, and (iii) "All Instruments" in blue. Goyal's variables comprises: *b/m*, *dfr*, *dfy*, *infl*, *ltr*, *lty*, *ntis*, *svar*, *tms* and *tbl*. FRED-MD is a large dataset containing 128 macroeconomic and financial variables. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

5.3 Variable Contribution

The results from the previous sections raise the natural question of what variables are important to each estimator. To answer this question we evaluate how each lagged variable contributed to produce the estimated conditional means in the OOS. After we obtain the estimates of the conditional mean $\widehat{\mu}_i(\mathbf{Z}_{t-1}) = \widehat{\mathbb{E}}(r_{i,t}|\mathbf{Z}_{t-1})$ for each risky asset $i = 1, \ldots, N$ at each t, we evaluate how each lagged variable in our sets of conditioning information contribute to generate $\widehat{\boldsymbol{\mu}}_t$ to be used to form the weights in $\mathbf{x}_t(\mathbf{Z}_t)$.

We take a straightforward approach to assess this contribution. At each month t, we compute the absolute values of the estimated coefficients in equation (6) for each asset i and lagged variable k in the conditioning set. For the estimators that we standardized the instruments before the regression, we multiply each $\widehat{\theta}_{i,l}$ by its own standard deviation computed in the estimation window. Grouping by estimators and $\mathbf{Z}$ used, we pool all the estimates of the N assets and calculate the average of the absolute values. In order to make the comparison clear, we normalize these means to sum one. Doing so, we can rank the most influential covariates driving the conditional means on a percentage scale.

Figure 3 reports for the dataset with 100 portfolios formed on size and BTM the 10 most influential lagged variables for all the estimators[21]. Interestingly, the PLS and PCR have some similarities regarding variable contribution, however this is not consistent across all three sets of conditioning information. Using FRED-MD, PLS and PCR ranked similarly the variables. The five most influential are: *HWI* (Help-Wanted Index for US), *AAAFFM* (Moody's Aaa Corporate Bond minus FEDFUNDS Aaa-FF spread), *BAAFFM* (Moody's Baa Corporate Bond minus FEDFUNDS Baa-FF spread), *T10YFFM* (10-Year Treasury C Minus FEDFUNDS 10 yr-FF spread), *T5YMFFM* (5-Year Treasury C minus FEDFUNDS 5 yr-FF spread) for both estimators. On the other hand, using all conditioning information, there is a clear difference in weights. The majority of the influence for PCR is split by two variables (*AAAFFM* and *BAAFFM*), while for PLS these two variables respond less than 30%, with a labor market variable (*HWI*) also having a large impact (35%), and small contributions from interest and exchange rates variables (*T1YFFM*, *TB6SMFFM*, *TB3SMFFM*). Out of the ten variables from Goyal's dataset, only three contributes to build efficient portfolios when using PCR (*tbl*, *lty* and *tms*).

For LASSO, Ridge, and Enet we see a similar pattern for all three estimators depending on which set of conditioning information was used. Finally, for OLS we see a completely different pattern compared to the previous estimators. We notice that labor market variables such as *DMANEMP* (All Employees: Durable goods), *NDMANEMP* (All Employees: Nondurable

[21] Except for the 3PRF which is a latent factor model

Figure 3: Variable Contribution by Estimator and Set of Conditioning Information - 100 Portfolios Formed on Size and Book-to-Market

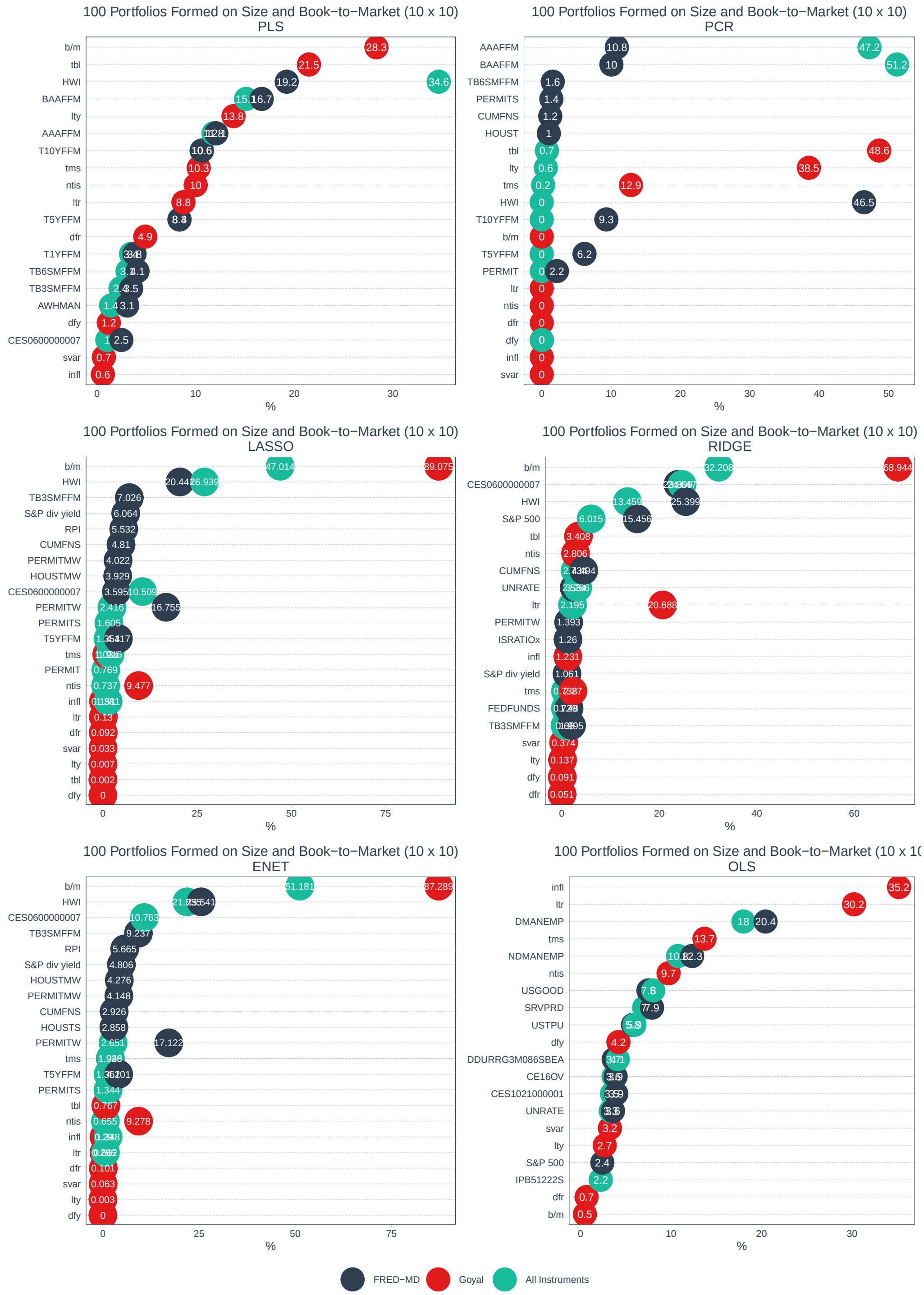


Figure 3 reports the 10 most influential variables by estimator (panels) and set of conditioning information used (colors) for the for the 100 portfolios formed on Size/BTM. We take a direct approach to obtain the contribution for each variable. At each month t, we compute the absolute values of the estimated coefficients in (6) for each asset i and lagged variable k in the conditioning set. We destandardize the variables whenever necessary. Grouping by estimators and $\mathbf{Z}$ used, we pool all the estimates of the N assets and calculate the average of the absolute values. In order to make the comparison clear, we normalize these means to sum one.

Figure 4: Variable Contribution by Estimator and Set of Conditioning Information - 25 Portfolios Formed on Size and Book-to-Market

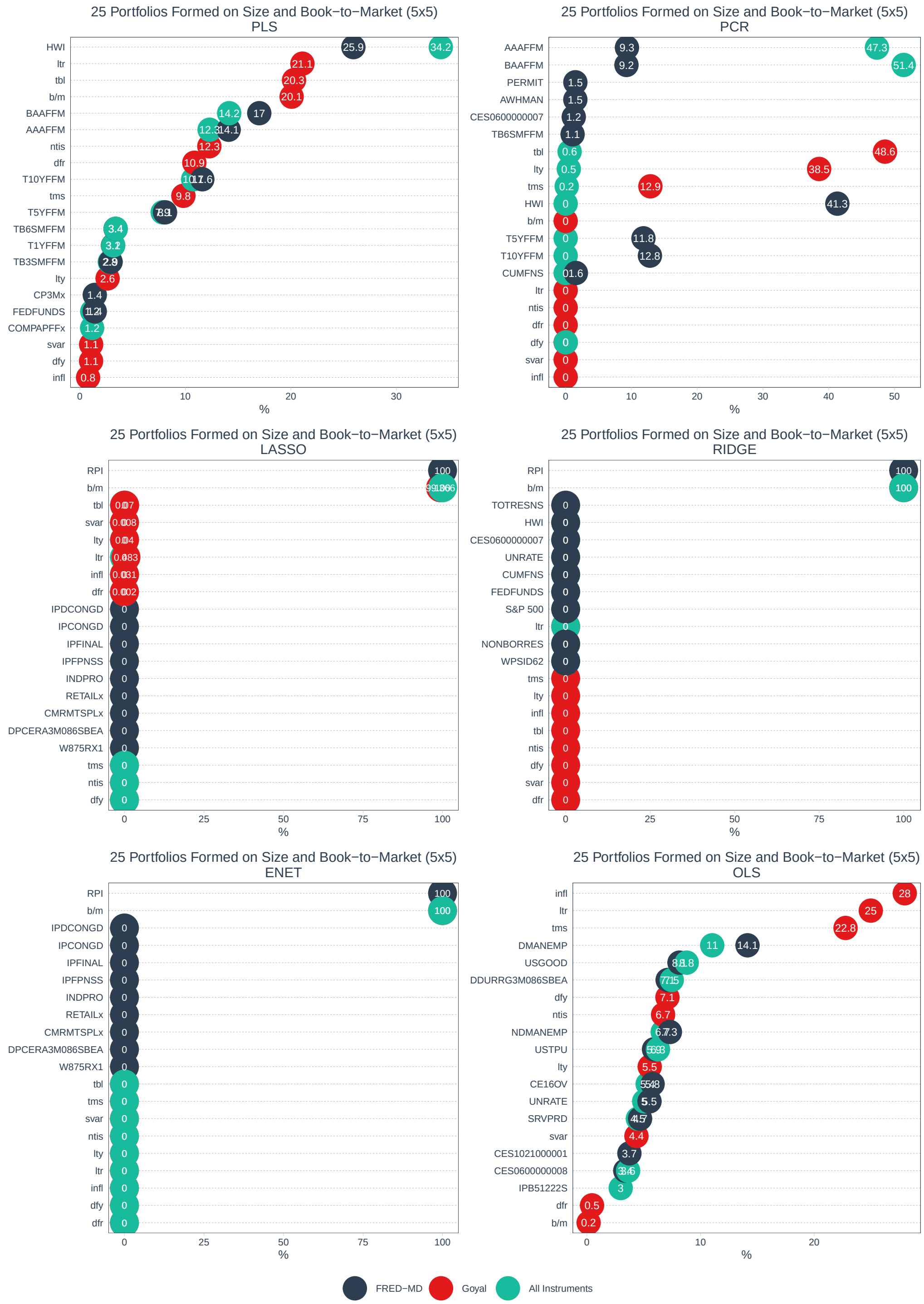


Figure 4 reports the 10 most influential variables by estimator (panels) and set of conditioning information used (colors) for the for the 25 portfolios formed on Size/BTM. We take a direct approach to obtain the contribution for each variable. At each month t, we compute the absolute values of the estimated coefficients in (6) for each asset i and lagged variable k in the conditioning set. We destandardize the variables whenever necessary. Grouping by estimators and $\mathbf{Z}$ used, we pool all the estimates of the N assets and calculate the average of the absolute values. In order to make the comparison clear, we normalize these means to sum one.

goods) and *USGOOD* (All Employees: Goods-Producing Industries) have a large impact when using large sets of predictors. These differences may provide us information on why some estimators produced better OOS mean-variance ratios.

Figure 4 illustrates the same analysis for the dataset with 25 portfolios formed on size and BTM[22]. The pattern for PCR and PLS is remarkably similar as seen for 100 portfolios formed on size and BTM. The same can be said regarding the OLS been driven mostly by labor market variables with high-dimensional conditioning information sets. However, for the penalized estimators we see a different story. We see that LASSO, Ridge, and ENet did a large shrinkage and selection in the set of lagged variables. For each one, the conditioning information set practically one variable was responsible for the full contribution: an output and income variable, *RPI* (Real Personal Income) using FRED-MD, and b/m (book-to-market) using Goyal and a combination of all lagged instruments.

5.4 Out-of-Sample Analysis - The Behavior of the Conditional Mean Estimates

To form the weights of the optimized portfolios, many factors could play a role to drive the formation of mean-variance efficient portfolios. Being the conditional mean one of these components, how the estimates behave? In order to answer this question, one way to evaluate the quality of the estimates is to assess the predictive performance of each estimator when using each set of lagged variables. Following Gu et al. (2018); Kelly et al. (2019) we calculate the out-of-sample R^2 as given in equation (29) for all 7 estimators and three different sets of conditioning information for each $\tau \in \tau_{OOS}$. Given the large number of results[23], we present them in different figures.

Figure 5 reports the OOS R^2 for the 100 and 25 portfolios formed on Size/BTM[24]. It is clear how PLS and PCR generate higher OOS R^2 compared to most of the estimators. This result is consistent across different sets of conditioning information and sets of risky assets. The figure also reports the pooled[25] OOS R^2, in which we pooled all risky assets and calculated the generated total OOS pooled R^2. In general, we see that again PLS and PCR

[22] In the Internet Appendix, we present similar analysis for the dataset with 6 portfolios formed on Size/BTM and the dataset with 5 industry portfolios.

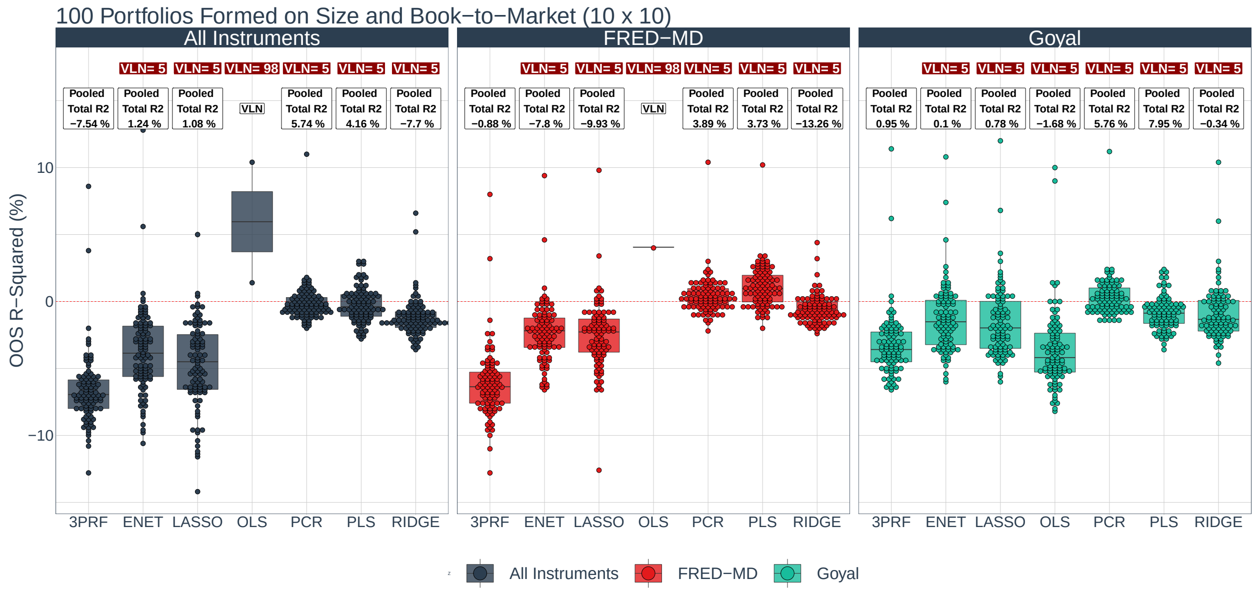
[23] One for each asset $i = 1, 2, \ldots, N$, estimator (7), and set of lagged variables (3).

[24] In the Internet Appendix, we present similar analysis for the dataset with 6 portfolios formed on Size/BTM and the dataset with 5 industry portfolios respectively.

[25]

$$R_{i,t}^{2^{(\text{pooled})}} = 1 - \frac{\sum_{i=1}^{N} \sum_{t \in \tau_{OOS}} \left(r_{i,t+1} - \hat{r}_{i,t+1}\right)^2}{\sum_{i=1}^{N} \sum_{t \in \tau_{OOS}} \left(r_{i,t+1}^2\right)} \tag{33}$$

Figure 5: Out-of-Sample R^2 - 100 and 25 Portfolios Formed on Size and Book-to-Market



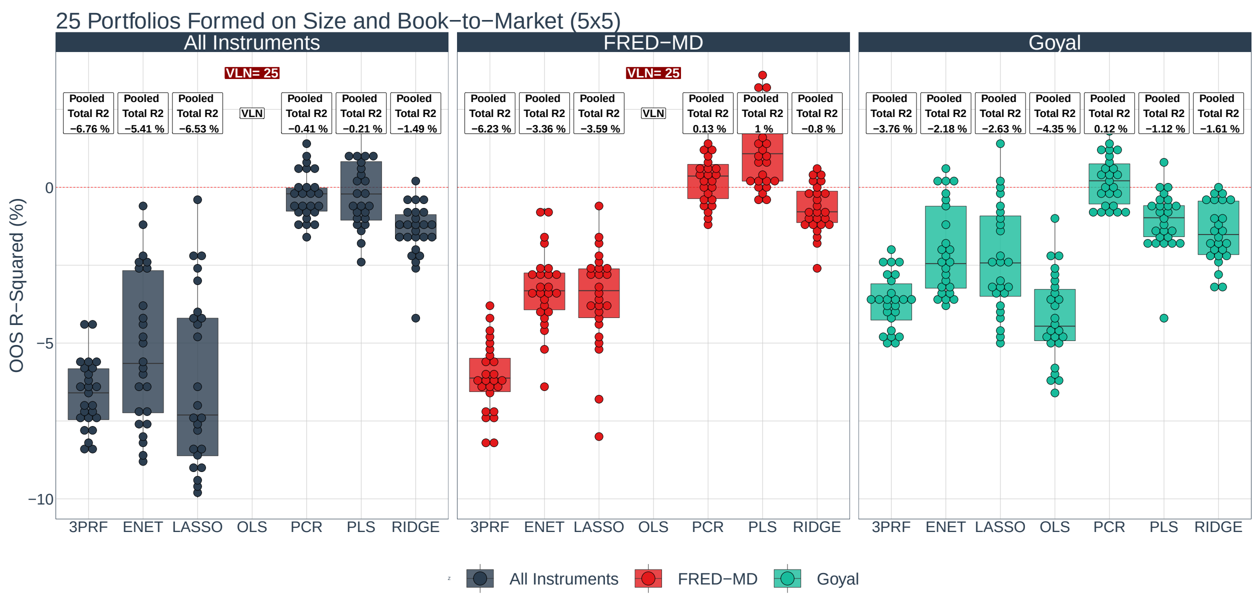


Figure 5 reports the OOS R^2 for the datasets with 100 and 25 portfolios formed on Size/BTM by estimator and set of conditioning information used, along with overlapping boxplots. The pooled OOS R^2 across all N risky assets in each dataset is also reported in a white box on the top of each plot. In order to make the plots readable and comparable, we filtered the OOS R^2 larger than -0.5 in absolute values and present the amount of assets that generated these very large numbers (VLN). The amount of VLN per estimator is reported in a red box on top of each plot.

can generate higher and in most cases positive R^2. Recall that these are monthly returns, and therefore positive R^2, even small ones, has a meaningful economic impact.

5.5 Economic Value

As argued in section 4.2, we can use the concept from Campbell and Thompson (2007) to evaluate how different estimators, using different predictive signals, can deliver economically meaningful results. Table 5 reports the implied Sharpe ratios for an investor with mean-variances preferences exploiting predictive information from the conditioning set. It is clear that PLS and PCR deliver higher monthly implied Sharpe ratios. Notice that when the OOS R^2 is negative, larger absolute R^2 values than the Sharpe ratio will cause the impossibility of the implied Sharpe ratio to be calculated.

5.5.1 Economic Gains

We use two measures to capture possible economic gains produced for optimal portfolios generated using high-dimensional data imposing sparsity in the different estimators. First, we consider the certainty equivalent return (CER) excess return of each optimal portfolio, which can be calculated as

$$\text{CER} = \hat{\mu}_p - \frac{\gamma}{2}\hat{\sigma}_p \tag{34}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p$ are the OOS unconditional mean and variance of the excess returns from the optimal portfolio generated by an estimator, γ is the risk aversion coefficient. We follow Brandt (2010); Goto and Xu (2015); Fletcher and Basu (2016) and set γ equal to 5. Higher CER values for a given optimal portfolio can be seen as that this portfolio has a better risk-return characteristic.

Table 6 reports the monthly CER in percentage for all cases. Inside brackets, we also report the ranking among all 7 estimators within strategy and set of conditioning information employed. We see that PLS and PCR generate a higher CER for an investor with risk aversion coefficient $\gamma = 5$ when N available risky assets is large (either 25 or 100)[26]. For 25 portfolios formed on Size/BTM we see that Ridge also generates high monthly CER. Overall, another fact that we can infer from the table is that the CER ranking change just marginally across the three approaches (CMV, UMV, MVATE).

Another way to infer possible economic gains of using a better approach to form mean-variance efficient portfolios is to measure the maximum fee an investor would pay to switch

[26] Table B.5 in the Internet Appendix presents the CER for 6 portfolios formed on Size/BTM and the dataset with 5 industry portfolios.

Table 5: Implied Sharpe Ratios

	25 Portfolios Formed on Size and Book-to-Market			100 Portfolios Formed on Size and Book-to-Market		
	CMV	UMV	MVATE	CMV	UMV	MVATE
Panel A: Goyal						
OLS	0.220	0.202	0.167	-	-	-
3PRF	0.258	0.262	0.234	0.153	0.166	0.115
PLS	0.286	0.302	0.279	0.389	0.387	0.413
PCR	0.384	0.388	0.382	0.398	0.408	0.406
LASSO	0.221	0.225	0.223	0.125	0.134	0.122
RIDGE	0.284	0.281	0.268	0.014	0.060	-
ENET	0.275	0.277	0.211	0.102	0.108	0.053
Panel B: FRED-MD						
OLS	-	-	-	-	-	-
3PRF	0.128	0.076	0.074	0.187	0.173	0.097
PLS	0.314	0.317	0.304	0.407	0.404	0.306
PCR	0.373	0.363	0.376	0.289	0.293	0.285
LASSO	0.099	0.086	0.094	-	-	-
RIDGE	0.258	0.261	0.259	-	-	-
ENET	0.208	0.191	0.216	0.123	0.109	-
Panel C: All Instruments						
OLS	-	-	-	-	-	-
3PRF	0.125	0.096	-	-	-	-
PLS	0.327	0.300	0.311	0.378	0.383	0.340
PCR	0.299	0.285	0.303	0.371	0.374	0.372
LASSO	-	-	-	0.341	0.333	0.275
RIDGE	0.278	0.279	0.271	-	-	-
ENET	0.187	0.152	0.187	0.378	0.371	0.212

Table 5 reports the implied SR^* as given in equation (30) for all seven estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge and ENet), three different sets of conditioning information (Goyal's, FRED-MD, and "All Instruments", wich is the combination of the previous two with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI)) and three different mean-variance approaches (CMV, UMV, MVATE) to build efficient portfolios. Notice that when the OOS R^2 is negative, larger absolute R^2 values than the Sharpe ratio will cause the impossibility of the implied Sharpe ratio to be calculated. These cases are omitted in the table.

Table 6: Certainty Equivalent Excess Returns (CER) (Monthly %)

	25 Portfolios Formed on Size and Book-to-Market						100 Portfolios Formed on Size and Book-to-Market					
	CMV		UMV		MVATE		CMV		UMV		MVATE	
Panel A: Goyal												
OLS	0.19	[5]	0.18	[5]	0.30	[5]	0.05	[6]	0.04	[7]	-0.13	[7]
3PRF	0.22	[4]	0.22	[3]	0.37	[4]	0.10	[3]	0.11	[3]	-0.01	[4]
PLS	0.29	[2]	0.32	[2]	0.49	[2]	0.22	[2]	0.22	[2]	0.40	[2]
PCR	0.38	[1]	0.38	[1]	0.67	[1]	0.26	[1]	0.25	[1]	0.46	[1]
LASSO	0.16	[7]	0.14	[7]	0.29	[7]	0.06	[5]	0.07	[5]	0.07	[3]
RIDGE	0.24	[3]	0.21	[4]	0.40	[3]	0.04	[7]	0.06	[6]	-0.05	[6]
ENET	0.18	[6]	0.16	[6]	0.29	[6]	0.07	[4]	0.07	[4]	-0.02	[5]
Panel B: FRED-MD												
OLS	0.03	[7]	0.03	[7]	-0.35	[7]	0.02	[7]	0.02	[7]	-0.86	[7]
3PRF	0.19	[5]	0.17	[5]	0.34	[5]	0.14	[3]	0.13	[3]	0.18	[5]
PLS	0.26	[3]	0.24	[3]	0.45	[3]	0.23	[1]	0.23	[1]	0.35	[1]
PCR	0.39	[1]	0.36	[1]	0.70	[1]	0.19	[2]	0.19	[2]	0.31	[2]
LASSO	0.16	[6]	0.14	[6]	0.28	[6]	0.10	[6]	0.09	[6]	0.18	[4]
RIDGE	0.26	[2]	0.25	[2]	0.46	[2]	0.13	[4]	0.12	[4]	0.23	[3]
ENET	0.21	[4]	0.17	[4]	0.38	[4]	0.12	[5]	0.11	[5]	0.12	[6]
Panel C: All Instruments												
OLS	0.03	[7]	0.03	[7]	-0.29	[7]	0.02	[7]	0.02	[7]	-2.04	[7]
3PRF	0.19	[5]	0.17	[4]	0.31	[5]	0.16	[3]	0.15	[3]	0.19	[4]
PLS	0.27	[3]	0.25	[3]	0.47	[3]	0.17	[2]	0.16	[2]	0.31	[2]
PCR	0.35	[1]	0.32	[1]	0.61	[1]	0.24	[1]	0.23	[1]	0.44	[1]
LASSO	0.17	[6]	0.14	[6]	0.30	[6]	0.13	[5]	0.13	[5]	0.23	[3]
RIDGE	0.29	[2]	0.27	[2]	0.49	[2]	0.12	[6]	0.12	[6]	0.12	[6]
ENET	0.22	[4]	0.17	[5]	0.40	[4]	0.14	[4]	0.14	[4]	0.18	[5]

Table 6 summarises the CER (monthly %) by estimator and optimal portfolio framework (CMV, UMV and MVATE) for both portfolios. Panel A reports the Sharpe ratios generated when the variables from Goyal's website are used as **Z**. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. Panel B presents the Sharpe ratios obtained using the FRED-MD variables. The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, panel C shows the Sharpe rations when all variables are used as conditioning information. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). Inside brackets, we also report the ranking among all 7 estimators within strategy and set of conditioning information employed from one approach to another. This management fee, motivated by Fleming et al. (2001), assumes a risk-averse investor with preferences given by a quadratic von Neumann-Morgenstern utility function[27]. This fee can be obtained by solving the following problem:

$$\mathbb{E}\left(r_{p,t}^{(a)}-\mathcal{F}\right)-\frac{\gamma}{2\left(1+\gamma\right)}\mathbb{E}\left(\left(r_{p,t}^{(a)}-\mathcal{F}\right)^{2}\right)=\mathbb{E}\left(r_{p,t}^{(b)}\right)-\frac{\gamma}{2\left(1+\gamma\right)}\mathbb{E}\left(r_{p,t}^{2^{(b)}}\right) \tag{36}$$

where $r_{p,t}^{(a)}$ is the OOS optimal portfolio return generated by a strategy making use of estimator (a), while $r_{p,t}^{(b)}$ is the benchmark OOS optimal portfolio return generated by estimator (b). Solving for $\mathcal{F}$ we can find the management fee that an investor would be willing to pay to have access to a better formation of an optimal portfolio.

Figure 6 reports the management fee, computed as the solution for $\mathcal{F}$ in equation (36). The comparison is made in pairs, assessing the management fee generated switching from (a) to (b). We can interpret the results in this figure, as the management fee of switching from a strategy given in the vertical axis, which shows the combination of the estimator employed and the mean-variance approach used, to another one plotted in different colors for all seven techniques and set of conditioning information. Essentially, we see that a positive management fee is generated when switching to either PLS or PCR. There are few cases in which switching from a penalized estimator to PLS or PCR produces a negative management fee.

5.6 Out-of-Sample Analysis - Portfolio Efficiency

Can these dimensionality reduction techniques generate significant alphas in a standard factor model? We answer this question in this section. We evaluate how the returns of the optimal portfolios produced by all seven estimators behave when evaluated in a standard factor model to explain the risk premia. We compare all possible combinations in the Fama-French three- and five-factor models, as well as with a Fama-French five-factor with momentum. The p-values of the generated alphas are plotted in figure 7, while the R^2 of the regressions are presented in figure 8. It is clear that in most cases, for N large PLS and PCR can generate statistically significant alphas. Table 7 present more details from the regression, reporting the alphas, t-statistics, p-values and the R^2 of the regressions using only the FRED-MD as the set of lagged instruments. In the Internet Appendix, we present the

[27] The expected utility function from the risk-averse investor is given by

$$U=W_0\left(\mathbb{E}\left(r_{p,t}\right)-\frac{\gamma}{2\left(1+\gamma\right)}\mathbb{E}\left(r_{p,t}\right)^{2}\right) \tag{35}$$

where W_0 represents the invertor's initial wealth, γ a constant risk aversion coefficient, and $r_{p,t}$ the portfolio return.

Figure 6: Management Fee

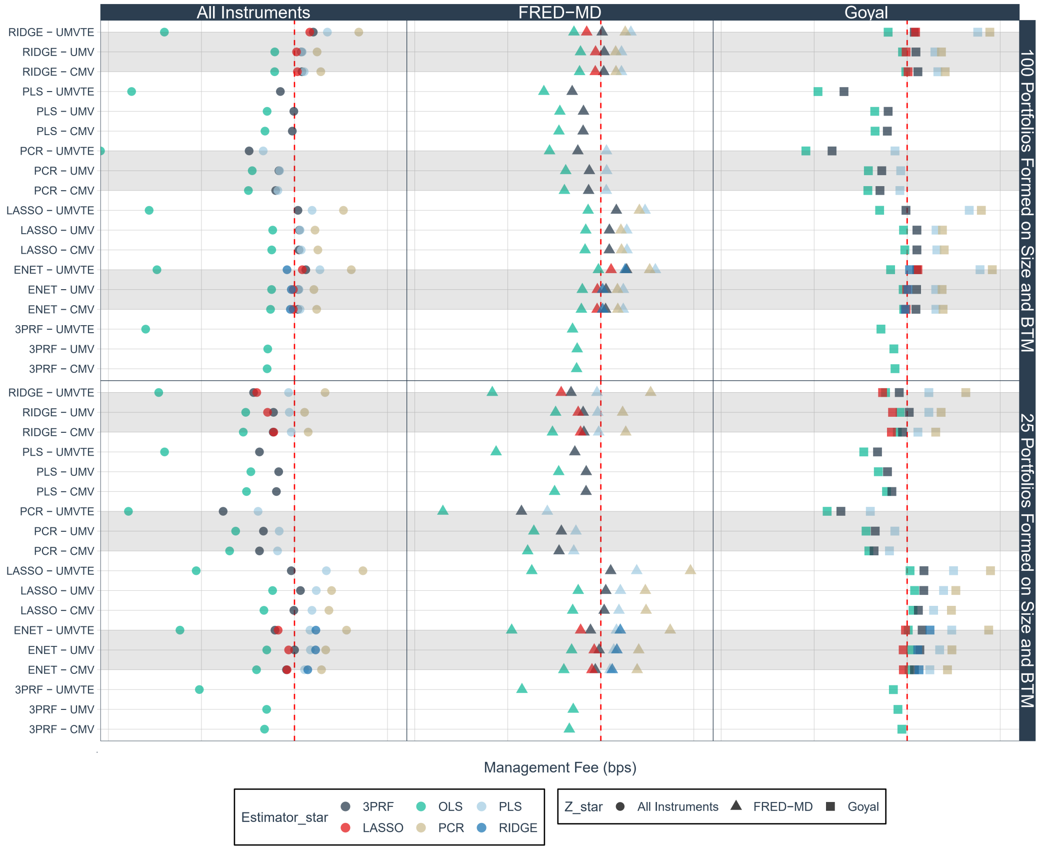


Figure 6 presents the management fee in bps, computed as the solution for $\mathcal{F}$ in equation (36), from an investor switching from an optimal portfolio formed by the estimator and mean-variance framework given in the left axis to another portfolio plotted in different colors and shape. The comparison is done in pairs of optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) and sets of conditioning information.

regressions for the remaining sets of conditioning information, as well as additional results.

Figure 7: p-values

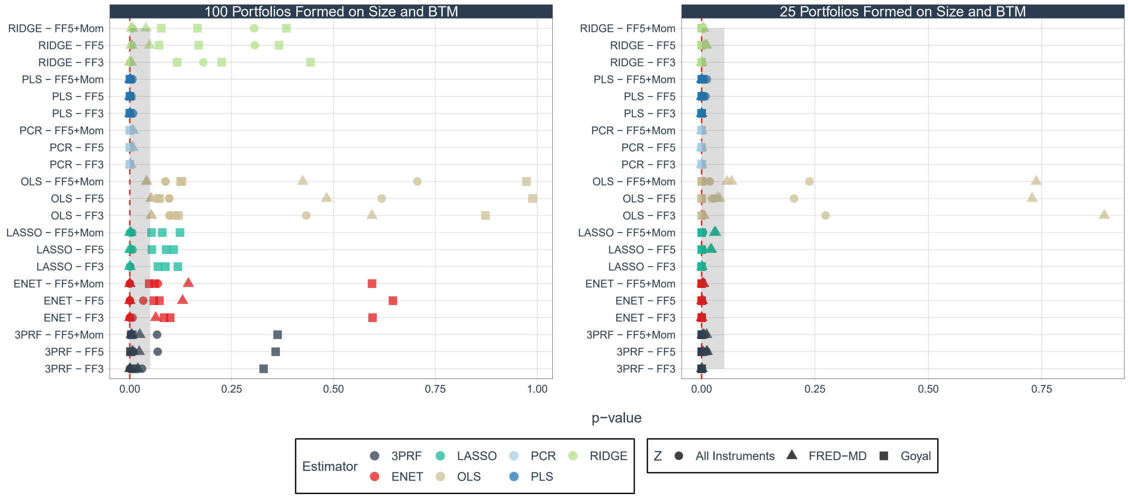


Figure 7 reports the p-values of alphas of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) and set of conditioning information used on the Fama-French 3, 5, and 5 + momentum factor models. The p-values are calculated from Newey–West t-statistics computed with one lag.

Figure 8: R-Squared

Figure 8 reports the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) and set of conditioning information used on the Fama-French 3, 5, and 5 + momentum factor models.

Table 7: Alphas (Monthly %)

		25 Portfolios Formed on Size and Book-to-Market									100 Portfolios Formed on Size and Book-to-Market								
		CMV			UMV			MVATE			CMV			UMV			MVATE		
		FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom
Z: FRED-MD																			
OLS	α (%)	0.027	0.021	0.019	0.027	0.021	0.019	-0.037	-0.082	-0.077	0.017	0.016	0.016	0.017	0.016	0.016	0.188	0.219	0.248
	t-statistics	[2.83]	[2.07]	[1.84]	[2.9]	[2.13]	[1.92]	[-0.14]	[-0.35]	[-0.34]	[1.95]	[1.95]	[2.05]	[1.94]	[1.95]	[2.05]	[0.53]	[0.7]	[0.8]
	p-val	0.005**	0.04*	0.066	0.004**	0.034*	0.056	0.888	0.729	0.738	0.052	0.052	0.041*	0.053	0.053	0.042*	0.594	0.483	0.425
	R^2	0.01	0.03	0.07	0.01	0.03	0.07	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.06	0.06
3PRF	α (%)	0.172	0.140	0.133	0.156	0.122	0.115	0.321	0.248	0.238	0.142	0.130	0.129	0.139	0.127	0.126	0.264	0.223	0.233
	t-statistics	[3.81]	[3.04]	[3.2]	[3.54]	[2.57]	[2.84]	[3.4]	[2.52]	[2.56]	[3.65]	[2.8]	[2.94]	[3.49]	[2.69]	[2.83]	[2.35]	[2.29]	[2.26]
	p-val	0**	0.003**	0.002**	0**	0.011*	0.005**	0.001**	0.012*	0.011*	0**	0.006**	0.004**	0.001**	0.008**	0.005**	0.02*	0.023*	0.025*
	R^2	0.12	0.15	0.18	0.09	0.13	0.16	0.14	0.17	0.19	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.06
PLS	α (%)	0.246	0.215	0.208	0.224	0.188	0.181	0.479	0.404	0.394	0.240	0.245	0.236	0.235	0.238	0.230	0.461	0.507	0.501
	t-statistics	[4.61]	[4.08]	[4.03]	[4.28]	[4.12]	[4.05]	[4.08]	[3.77]	[3.64]	[4.58]	[4.06]	[4.4]	[4.4]	[4.14]	[4.38]	[3.27]	[3.41]	[3.33]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.001**	0.001**	0.001**
	R^2	0.14	0.16	0.17	0.15	0.17	0.19	0.10	0.13	0.14	0.05	0.06	0.10	0.05	0.06	0.09	0.05	0.06	0.06
PCR	α (%)	0.379	0.321	0.300	0.346	0.291	0.271	0.742	0.622	0.582	0.217	0.205	0.195	0.211	0.196	0.187	0.427	0.397	0.379
	t-statistics	[6.19]	[5.47]	[5.32]	[5.99]	[5.02]	[4.99]	[6.09]	[5.53]	[5.19]	[3.1]	[2.84]	[2.79]	[3.23]	[2.9]	[2.77]	[3.07]	[2.69]	[2.64]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.002**	0.005**	0.006**	0.001**	0.004**	0.006**	0.002**	0.008**	0.009**
	R^2	0.25	0.28	0.37	0.28	0.31	0.40	0.25	0.28	0.37	0.00	0.01	0.03	0.00	0.01	0.04	0.00	0.01	0.03
LASSO	α (%)	0.157	0.114	0.103	0.138	0.101	0.092	0.296	0.213	0.193	0.099	0.084	0.081	0.096	0.080	0.077	0.190	0.155	0.149
	t-statistics	[3.35]	[2.35]	[2.21]	[3.39]	[2.31]	[2.19]	[3.34]	[2.32]	[2.17]	[4.87]	[4.2]	[3.99]	[4.75]	[4]	[3.8]	[3.55]	[3.11]	[3.01]
	p-val	0.001**	0.019*	0.028*	0.001**	0.022*	0.029*	0.001**	0.021*	0.031*	0**	0**	0**	0**	0**	0**	0**	0.002**	0.003**
	R^2	0.10	0.13	0.18	0.09	0.13	0.17	0.10	0.14	0.18	0.04	0.08	0.10	0.05	0.08	0.11	0.03	0.06	0.07
RIDGE	α (%)	0.247	0.171	0.150	0.238	0.169	0.149	0.470	0.325	0.290	0.127	0.106	0.096	0.122	0.101	0.091	0.263	0.207	0.189
	t-statistics	[4.19]	[2.56]	[3.04]	[3.53]	[2.56]	[2.86]	[4.5]	[2.75]	[3.01]	[4.32]	[2.92]	[2.9]	[4.21]	[2.83]	[2.74]	[2.91]	[1.98]	[2.07]
	p-val	0**	0.011*	0.003**	0**	0.011*	0.005**	0**	0.006**	0.003**	0**	0.004**	0.004**	0**	0.005**	0.007**	0.004**	0.049*	0.039*
	R^2	0.18	0.25	0.34	0.17	0.23	0.34	0.19	0.25	0.34	0.09	0.12	0.23	0.09	0.12	0.24	0.08	0.10	0.15
ENET	α (%)	0.195	0.159	0.149	0.163	0.128	0.121	0.395	0.315	0.302	0.117	0.102	0.098	0.112	0.097	0.093	0.151	0.117	0.106
	t-statistics	[4.3]	[3.47]	[3.34]	[4.32]	[3.16]	[2.84]	[4.21]	[3.51]	[3.03]	[4.63]	[4.11]	[3.89]	[4.58]	[3.93]	[3.8]	[1.86]	[1.52]	[1.47]
	p-val	0**	0.001**	0.001**	0**	0.002**	0.005**	0**	0.001**	0.003**	0**	0**	0**	0**	0**	0**	0.064	0.130	0.144
	R^2	0.13	0.16	0.20	0.13	0.17	0.19	0.09	0.12	0.14	0.05	0.07	0.11	0.05	0.07	0.11	0.00	0.02	0.05

Table 7 reports the alphas, t-statistics, p-values and the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) using the FRED-MD as conditioning information on the Fama-French 3, 5, and 5 + momentum factor models. FRED-MD is a large dataset containing 128 macroeconomic and financial variables.

5.7 Financial Metrics

As a final analysis, we also assess how sparsity and dimension reduction can affect common financial metrics for the returns generated by the efficient portfolios. We check how the turnover is affected by different estimators. Additionally, we also assess another common risk metric, which is the maximum drawdown. Related to it, we evaluate the pain ratio[28]. As in Kirby and Ostdiek (2012), we define the turnover as

$$\text{Turnover}_p = \frac{1}{T} \sum_{t \in \tau_{OOS}} \sum_{i}^{N} |x_{p,i,t} - x_{p,i,t-1}| + \left| \sum_{i}^{N} (x_{p,i,t} - x_{p,i,t-1}) \right| \tag{38}$$

where $x_{p,i,t}$ is the optimal weight of asset i for a portfolio p at t in the OOS, and $x_{p,i,t-1}$ is the optimal weight of asset i for a portfolio p at $t-1$. Following Gu et al. (2018), we define the maximum drawdown as

$$\text{MaxDD}_p = \max_{0 \leq Y_{p,t_1} \leq Y_{p,t_2} \leq T} (Y_{p,t_1} - Y_{p,t_2}) \tag{39}$$

where $Y_{p,t}$ is the cumulative log return for a portfolio p from the first OOS period to t. Given the large amount of results, we report only for the CMV approach[29].An inspection in table 8 show that overall all estimator generate high turnover, especially when the available number of risky assets N is low. Notice that PLS, in general, produces a lower rate of turnover compared to the penalized regressors for most of the cases. This behavior, as similarly noted by Gu et al. (2018), should be understood in light of the large role of price trend predictors selected by these dimension reduction techniques. The maximum drawdown for PLS and PCR are in line with the other estimators. It is clear that, with more risky assets, the diversification generates a lower drawdown. We see that the pain ratio also shows that, in general, for large N, PCR and PLS performs well compared to most estimators.

6 Conclusion

The financial literature has collected a large list of variables that potentially could be in investors' conditioning information sets. Standard approaches relying on a few lagged

[28]The Pain Ratio can be calculated as

$$\text{Pain Ratio}_p = \frac{r_{p,t}}{\sum_{t \in \tau_{OOS}} \frac{|D_t|}{T}} \tag{37}$$

where D_τ is the drawdown since previous peak.

[29]For the other two approaches, UMV and MVATE, the results are similar.

Table 8: Turnover and Financial Metrics (Monthly %)

	25 Portfolios Formed on Size and Book-to-Market			100 Portfolios Formed on Size and Book-to-Market		
	Turnover (%)	Max DD (%)	Pain Ratio	Turnover (%)	Max DD (%)	Pain Ratio
Panel A: Goyal						
OLS	92.35	3.50	0.323	61.26	15.20	0.072
3PRF	89.61	3.14	0.375	66.88	15.98	0.239
PLS	119.98	6.11	0.324	82.92	10.98	0.361
PCR	131.74	5.56	0.467	93.62	8.61	0.373
LASSO	126.59	4.36	0.277	75.93	14.22	0.124
RIDGE	120.74	3.44	0.368	73.03	22.05	0.053
ENET	129.82	2.39	0.418	75.89	14.26	0.132
Panel B: FRED-MD						
OLS	72.13	1.15	0.098	45.99	1.37	0.045
3PRF	162.75	4.66	0.197	121.68	4.22	0.289
PLS	121.16	7.82	0.308	87.83	4.57	0.528
PCR	126.98	4.98	0.543	100.89	13.13	0.154
LASSO	180.06	7.04	0.138	81.94	1.21	0.465
RIDGE	180.59	7.16	0.182	95.36	3.43	0.386
ENET	178.36	4.74	0.261	86.89	1.81	0.453
Panel C: All Instruments						
OLS	61.63	1.27	0.106	40.35	1.56	0.044
3PRF	144.62	4.51	0.229	100.64	4.21	0.336
PLS	120.02	9.78	0.225	82.15	3.46	0.354
PCR	154.23	6.06	0.335	123.91	9.84	0.273
LASSO	166.84	4.94	0.171	91.41	1.59	0.428
RIDGE	178.70	5.89	0.249	86.11	2.99	0.228
ENET	175.27	3.45	0.378	93.85	2.02	0.410

Table 8 reports several standard financial metrics to evaluate portfolios by estimator for the CMV framework for both portfolios. The Turnover is computed following equation (38). MaxDD is maximum drawdown as presented in equation (39). Pain Ratio is the standard metric, as shown in equation (37). Panel A reports the financial metrics generated when the variables from Goyal's website are used as **Z**. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl.* Panel B presents the financial metrics obtained using the FRED-MD variables. Finally, panel C shows the financial metrics when all variables are used as conditioning information. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

variables and in a naive OLS estimation can be problematic and inefficient, as shown in this work.

We exploit the wealth set of data as predictive signals and condense this information when estimating the conditional mean using high-dimensional data. We do this using many different techniques in order to impose sparsity and dimensionality reduction when finding the conditional mean, which is the most important item driving the formation of mean-variance efficient portfolios with conditioning information. We evaluate how penalized estimators, such as LASSO, Ridge and Elastic Net, as well as pure dimensionality reduction and latent factors approaches as Partial Least Squares (PLS) and Principal Components Regression (PCR), in addition to a generalization of the former (Three Pass Regression Filter) can produce different optimized portfolios.

We assess these methods in a high-dimensional setting for the predictive list of instruments. We see that it is possible to build efficient portfolios, for all three different approaches evaluated, delivering on average higher Sharpe ratios OOS, implied Sharpe ratios (SR^*), and higher certainty equivalent returns (CER). We find that PLS and PCR can enhance the formation of optimized portfolios, generating large and significant alphas when evaluated in factor models, such as Fama-French 3 and 5.

In short, exploiting a high-dimensional set of conditioning information, it is possible with a "bet on sparsity", employing shrinkage-type estimators, regularization, and allowing flexible forms for the approximation of the conditional mean of the excess returns.

References

Abhyankar, A., Basu, D., and Stremme, A. (2012). The optimal use of return predictability: An empirical study. *Journal of Financial and Quantitative Analysis*, 47(5):973–1001.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636.

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.

Bianchi, D., Büchner, M., and Tamoni, A. (2019). Bond risk premia with machine learning. *USC-INET Research Paper*, (19-11).

Brandt, M. W. (2010). Portfolio choice problems. In *Handbook of financial econometrics: Tools and techniques*, pages 269–336. Elsevier.

Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61(5):2187–2217.

Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Chiang, I.-H. E. (2015). Modern portfolio management with conditioning information. *Journal of Empirical Finance*, 33:114–134.

Chinco, A., Clark-Joseph, A. D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492.

Cochrane, J. H. (2009). *Asset Pricing*. Princeton University Press.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108.

Diebold, F. X., Ghysels, E., Mykland, P., and Zhang, L. (2019). Big data in dynamic predictive econometric modeling.

Feng, G., Giglio, S., and Xiu, D. (2017). Taming the factor zoo. *Fama-Miller Working Paper*, 24070.

Ferson, W. E. and Siegel, A. F. (2001). The efficient use of conditioning information in portfolios. *The Journal of Finance*, 56(3):967–982.

Ferson, W. E. and Siegel, A. F. (2009). Testing portfolio efficiency with conditioning information. *Review of Financial Studies*, 22(7):2735–2758.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance*, 56(1):329–352.

Fletcher, J. and Basu, D. (2016). An examination of the benefits of dynamic trading strategies in uk closed-end funds. *International Review of Financial Analysis*, 47:109–118.

Frank, L. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135.

Freyberger, J., Neuhierl, A., and Weber, M. (2017). Dissecting characteristics nonparametrically. Technical report, National Bureau of Economic Research.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Goto, S. and Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, 50(6):1415–1441.

Green, J., Hand, J. R., and Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, 18(3):692–730.

Gu, S., Kelly, B., and Xiu, D. (2018). Empirical asset pricing via machine learning. Technical report, National Bureau of Economic Research.

Hansen, L. P. and Richard, S. F. (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica*, pages 587–613.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. Chapman and Hall/CRC.

Huang, J.-z. and Shi, Z. (2011). Determinants of bond risk premia. Technical report, Citeseer.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kelly, B. and Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, 68(5):1721–1756.

Kelly, B. and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*.

Kirby, C. and Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, 47(2):437–467.

Kozak, S., Nagel, S., and Santosh, S. (2019). Shrinking the cross-section. *Journal of Financial Economics*.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Light, N., Maslov, D., and Rytchkov, O. (2017). Aggregation of information about the cross section of stock returns: A latent variable approach. *The Review of Financial Studies*, 30(4):1339–1381.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.

Peñaranda, F. (2016). Understanding portfolio efficiency with conditioning information. *Journal of Financial and Quantitative Analysis*, 51(3):985–1011.

Politis, D. N. and Romano, J. P. (1992). A circular block-resampling procedure for stationary data. *Exploring the limits of bootstrap*, 2635270.

Püttmann, L. (2018). Patterns of panic: Financial crisis language in historical newspapers.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stone, M. and Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(2):237–258.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Welch, I. and Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Internet Appendix

A Appendix - Statistical Inference for the Difference of Sharpe Ratios

To test for the difference of the Sharpe ratios of two different optimal portfolios, we follow Ledoit and Wolf (2008) approach to perform these tests. The main advantage of this method, is its robustness to (i) when returns have tails heavier than the normal distribution, and (ii) serial correlation of the actual returns or of the squared returns (volatility clustering). Following Ledoit and Wolf (2008) the difference between SR_a and SR_b, for two optimal portfolios produced by two different estimators, a and b, is given by

$$\Delta = SR^{(a)} - SR^{(b)} = \frac{\mu^{(a)}}{\sigma^{(a)}} - \frac{\mu^{(b)}}{\sigma^{(b)}} \tag{40}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the OOS unconditional mean and variance of the excess returns from the optimal portfolio generated by an estimator. Thus, the null hypothesis can be written as: $H_0 : \Delta = 0$. Denote the vector of uncentered population and sample moments of the excess returns generated by two different estimators by

$$\begin{aligned} \zeta &= \left[\mu^{(a)}, \mu^{(b)}, \mathbb{E}\left(r^{2^{(a)}}\right), \mathbb{E}\left(r^{2^{(b)}}\right)\right]^\top \\ \widehat{\zeta} &= \left[\widehat{\mu^{(a)}}, \widehat{\mu^{(b)}}, \widehat{\mathbb{E}\left(r^{2^{(a)}}\right)}, \widehat{\mathbb{E}\left(r^{2^{(b)}}\right)}\right]^\top \end{aligned} \tag{41}$$

Under the assumption that the differences of these moments converges in distribution to a Normal distribution with mean 0 and variance Ψ, using the delta method we have

$$\sqrt{T}\left(\widehat{\Delta} - \Delta\right) \xrightarrow{d} N\left(0, \nabla^\top f(\zeta)\Psi\nabla f(\zeta)\right) \tag{42}$$

where $f(\zeta) = \dfrac{\mu^{(a)}}{\sqrt{\mathbb{E}\left(r^{2^{(a)}}\right) - \mu^{2^{(a)}}}} - \dfrac{\mu^{(b)}}{\sqrt{\mathbb{E}\left(r^{2^{(b)}}\right) - \mu^{2^{(b)}}}}$ and the standard error of $\widehat{\Delta}$ is given by

$$se\left(\widehat{\Delta}\right) = \sqrt{\frac{\nabla^\top f(\widehat{\zeta})\widehat{\Psi}\nabla f(\widehat{\zeta})}{T}} \tag{43}$$

Since neither the returns, nor the squared returns of financial assets are generally an i.i.d. process, we need to use a robust estimator. One evident way to obtain $\widehat{\Psi}$ is using HAC kernel estimation. Another way is to use bootstrap inference. Ledoit and Wolf (2008) suggest to construct a studentized time series bootstrap confidence interval for the difference of the SRs. This method has been shown to be robust when returns have tails heavier than the normal distribution or are of time series nature. The bootstrap data is generated using the circular block bootstrap of Politis and Romano (1992). The two-sided distribution function of the studentized statistic can be obtained via bootstrap as follows:

$$f\left(\frac{|\widehat{\Delta}-\Delta|}{se\left(\widehat{\Delta}\right)}\right) \approx f\left(\frac{|\widehat{\Delta}^{\text{boot}}-\Delta|}{se\left(\widehat{\Delta}^{\text{boot}}\right)}\right) \tag{44}$$

where $f(\cdot)$ is the distribution of a random variable, Δ is populational difference between SR_a and SR_b, $\widehat{\Delta}$ is the sample counterpart of this difference obtained in the data in the estimation window, and $\widehat{\Delta}^{\text{boot}}$ is the estimated difference computed from bootstrap. The standard errors are denoted by $se(\cdot)$. Out of the distribution obtained from the bootstrap in the equation (44), we can find the confidence interval for Δ and the p-values of the test[30].

[30]For further details in this procedure, see Ledoit and Wolf (2008).

B Appendix - Additional Results

B.1 Out-of-Sample Analysis - Sharpe Ratios

B.1.1 Performance Evaluation

Table B.1: Out-of-Sample Sharpe ratios delivered by each estimator and set of conditioning information

	5 Industry Portfolios						6 Portfolios Formed on Size and Book-to-Market					
	CMV		UMV		MVATE		CMV		UMV		MVATE	
Estimator	SR	p-val	SR	p-val	SR	p-val	SR	p-val	SR	p-val	SR	p-val
Panel A: Goyal												
OLS	0.133	0.032	0.100	0.107	0.135	0.030	0.207	0.001	0.218	0.001	0.203	0.001
3PRF	0.125	0.044	0.083	0.181	0.125	0.043	0.222	0.000	0.258	0.000	0.223	0.000
PLS	0.121	0.051	0.109	0.078	0.119	0.054	0.247	0.000	0.246	0.000	0.247	0.000
PCR	0.089	0.149	0.076	0.217	0.085	0.168	0.298	0.000	0.299	0.000	0.300	0.000
LASSO	0.209	0.001	0.141	0.024	0.209	0.001	0.168	0.007	0.187	0.003	0.166	0.008
RIDGE	0.131	0.035	0.112	0.072	0.131	0.034	0.173	0.006	0.224	0.000	0.174	0.005
ENET	0.163	0.009	0.130	0.037	0.168	0.007	0.231	0.000	0.260	0.000	0.236	0.000
Panel B: FRED-MD												
OLS	0.005	0.934	0.009	0.881	-0.046	0.455	0.154	0.013	0.158	0.011	0.037	0.550
3PRF	0.072	0.246	0.009	0.883	0.073	0.239	0.186	0.003	0.197	0.002	0.188	0.003
PLS	0.102	0.099	0.046	0.455	0.101	0.102	0.210	0.001	0.198	0.002	0.201	0.001
PCR	0.115	0.064	0.019	0.764	0.112	0.070	0.272	0.000	0.255	0.000	0.275	0.000
LASSO	0.156	0.012	0.056	0.363	0.152	0.014	0.226	0.000	0.201	0.001	0.226	0.000
RIDGE	0.116	0.061	0.028	0.655	0.116	0.062	0.242	0.000	0.230	0.000	0.246	0.000
ENET	0.127	0.042	0.032	0.601	0.125	0.044	0.261	0.000	0.261	0.000	0.261	0.000
Panel C: All Instruments												
OLS	-0.035	0.594	-0.002	0.979	-0.078	0.240	0.160	0.017	0.168	0.012	-0.002	0.973
3PRF	0.067	0.281	0.029	0.640	0.068	0.271	0.193	0.002	0.213	0.001	0.191	0.002
PLS	0.002	0.976	0.037	0.581	-0.001	0.991	0.264	0.000	0.263	0.000	0.265	0.000
PCR	0.133	0.047	0.040	0.552	0.129	0.054	0.238	0.000	0.237	0.001	0.240	0.000
LASSO	0.122	0.067	0.052	0.435	0.118	0.078	0.271	0.000	0.228	0.000	0.274	0.000
RIDGE	0.113	0.091	0.035	0.602	0.112	0.094	0.248	0.000	0.229	0.001	0.247	0.000
ENET	0.069	0.301	0.047	0.483	0.069	0.300	0.255	0.000	0.234	0.000	0.266	0.000

Table B.1 summarises the OOS (Jan-1996 to Dec-2017) Sharpe ratios (SR) by estimator and optimal portfolio framework (CMV, UMV and MVATE) for the 5 industry portfolios and 6 portfolios formed on Size/BTM. Panel A reports the Sharpe ratios generated when the variables from Goyal's website are used as **Z**. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. Panel B presents the Sharpe ratios obtained using the FRED-MD variables. The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, panel C shows the Sharpe rations when all variables are used as conditioning information. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). The p-val is the p-value from the two-sided test of the SR.

B.1.2 Statistical Inference for the Difference of Sharpe Ratios

Table B.2: Test for the difference of the Sharpe ratios - OOS (Jan-1996 - Dec-2017) - 6 Portfolios Formed on Size and Book-to-Market

Panel A: CMV

		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
Goyal	OLS	0.734	0.505	0.116	0.316	0.492	0.593	0.695	0.974	0.241	0.747	0.507	0.351	0.780	0.332	0.642	0.251	0.457	0.485
	3PRF		0.621	0.077	0.298	0.339	0.860		0.825	0.234	0.938	0.655	0.431		0.621	0.990	0.309	0.843	0.443
	PLS			0.163	0.128	0.159	0.776			0.462	0.656	0.898	0.776			0.860	0.480	0.973	0.728
	PCR				0.012	0.012	0.221				0.079	0.071	0.344				0.714	0.034	0.535
	LASSO					0.905	0.182					0.120	0.084					0.045	0.175
	RIDGE						0.131						0.129						0.117
FRED-MD	OLS							0.655	0.525	0.156	0.334	0.252	0.175	0.591	0.244	0.395	0.126	0.286	0.170
	3PRF								0.659	0.038	0.437	0.052	0.120		0.035	0.091	0.056	0.011	0.151
	PLS									0.069	0.741	0.426	0.268			0.488	0.194	0.350	0.315
	PCR										0.233	0.281	0.757				0.785	0.300	0.956
	LASSO											0.676	0.228					0.685	0.209
	RIDGE												0.621						0.525
All Instr.	OLS													0.790	0.200	0.371	0.072	0.241	0.118
	3PRF														0.050	0.147	0.094	0.026	0.247
	PLS															0.586	0.477	0.641	0.709
	PCR																0.161	0.769	0.366
	LASSO																	0.248	0.666
	RIDGE																		0.459

Panel B: UMV

		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
Goyal	OLS	0.184	0.601	0.122	0.345	0.865	0.236	0.665	0.762	0.578	0.766	0.816	0.445	0.901	0.356	0.806	0.809	0.845	0.799
	3PRF		0.815	0.281	0.080	0.310	0.961		0.351	0.952	0.284	0.554	0.947		0.952	0.637	0.616	0.425	0.732
	PLS			0.218	0.308	0.646	0.804			0.855	0.376	0.686	0.777			0.745	0.788	0.565	0.930
	PCR				0.019	0.082	0.478				0.064	0.087	0.420				0.201	0.055	0.273
	LASSO					0.259	0.071					0.371	0.198					0.371	0.418
	RIDGE						0.294						0.505						0.770
FRED-MD	OLS							0.588	0.646	0.278	0.556	0.343	0.185	0.440	0.218	0.419	0.360	0.390	0.311
	3PRF								0.980	0.356	0.930	0.269	0.254		0.139	0.393	0.418	0.152	0.412
	PLS									0.184	0.957	0.494	0.233			0.384	0.534	0.531	0.469
	PCR										0.363	0.595	0.900				0.767	0.498	0.886
	LASSO											0.497	0.060					0.401	0.165
	RIDGE												0.472						0.724
All Instr.	OLS													0.608	0.226	0.448	0.372	0.413	0.364
	3PRF														0.258	0.621	0.605	0.462	0.631
	PLS															0.589	0.715	0.444	0.765
	PCR																0.944	0.867	0.901
	LASSO																	0.794	0.775
	RIDGE																		0.762

Panel C: MVATE

		Goyal						FRED-MD						All Instr.					
		3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET	3PRF	PLS	PCR	LASSO	RIDGE	ENET
Goyal	OLS	0.646	0.482	0.104	0.362	0.577	0.449	0.787	0.969	0.207	0.697	0.431	0.316	0.810	0.316	0.595	0.224	0.442	0.360
	3PRF		0.655	0.082	0.276	0.350	0.803		0.688	0.235	0.959	0.633	0.443		0.632	0.992	0.316	0.865	0.354
	PLS			0.156	0.121	0.169	0.840			0.437	0.655	0.962	0.782			0.877	0.463	0.998	0.646
	PCR				0.009*	0.013	0.235				0.074	0.070	0.303				0.720	0.029	0.621
	LASSO					0.853	0.125					0.091	0.073					0.040	0.111
	RIDGE						0.092						0.127						0.081
FRED-MD	OLS							0.045	0.046	0.002*	0.009*	0.006*	0.002*	0.041	0.009*	0.012	0.002*	0.009*	0.002*
	3PRF								0.808	0.034	0.459	0.037	0.124		0.030	0.090	0.055	0.011	0.117
	PLS									0.028	0.607	0.268	0.187			0.367	0.130	0.278	0.189
	PCR										0.213	0.297	0.710				0.787	0.263	0.955
	LASSO											0.631	0.236					0.698	0.143
	RIDGE												0.690						0.447
All Instr.	OLS													0.028	0.002*	0.003*	0.000*	0.002*	0.000*
	3PRF														0.038	0.129	0.080	0.022	0.177
	PLS															0.608	0.479	0.615	0.623
	PCR																0.167	0.823	0.328
	LASSO																	0.248	0.831
	RIDGE																		0.384

p-val	$[0 \leq p \leq 0.05]$	$(0.05 < p < 0.1)$	$(0.1 \leq p \leq 1]$
Gradient color bounds			

Test for the differences of the Sharpe ratios of the OOS (Jan-1996 - Dec-2017) returns of the efficient portfolios formed from the dataset with 6 Size/BTM portfolios using 7 different estimators (OLS, 3PRF, PLS, PCR, LASSO, RIDGE and ENet) and three different set of conditioning information (Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios. Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

Table B.3: Test for the difference of the Sharpe ratios - OOS (Jan-1996 - Dec-2017) - 5 Portfolios Formed on Size and Book-to-Market

Panel A: CMV

Goyal (rows) vs Goyal (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.888	0.841	0.539	0.105	0.961	0.528
3PRF		0.937	0.588	0.056	0.915	0.451
PLS			0.649	0.133	0.874	0.493
PCR				0.098	0.584	0.332
LASSO					0.035	0.223
RIDGE						0.442

Goyal (rows) vs FRED-MD (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.406	0.644	0.769	0.713	0.807	0.916
3PRF		0.733	0.872	0.642	0.899	0.986
PLS			0.931	0.604	0.947	0.937
PCR				0.381	0.713	0.605
LASSO					0.166	0.203
RIDGE						0.943

Goyal (rows) vs All Instr. (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.359	0.037	0.740	0.608	0.522	0.219
3PRF		0.088	0.949	0.817	0.732	0.360
PLS			0.991	0.899	0.805	0.430
PCR				0.746	0.831	0.710
LASSO					0.060	0.016
RIDGE						0.194

FRED-MD (rows) vs FRED-MD (columns)

FRED-MD	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.285	0.224	0.118	0.017	0.089	0.086
3PRF		0.701	0.524	0.163	0.305	0.371
PLS			0.851	0.458	0.856	0.722
PCR				0.515	0.982	0.850
LASSO					0.440	0.409
RIDGE						0.844

FRED-MD (rows) vs All Instr. (columns)

FRED-MD	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.302	0.874	0.076	0.074	0.073	0.312
3PRF		0.400	0.428	0.550	0.500	0.923
PLS			0.833	0.936	0.975	0.574
PCR				0.991	0.892	0.480
LASSO					0.431	0.114
RIDGE						0.390

All Instr. (rows) vs All Instr. (columns)

All Instr.	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.114	0.665	0.042	0.034	0.052	0.177
3PRF		0.395	0.388	0.536	0.433	0.937
PLS			0.127	0.142	0.227	0.403
PCR				0.892	0.742	0.413
LASSO					0.875	0.207
RIDGE						0.457

Panel B: UMV

Goyal (rows) vs Goyal (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.666	0.828	0.733	0.346	0.730	0.439
3PRF		0.603	0.918	0.237	0.532	0.328
PLS			0.609	0.572	0.954	0.687
PCR				0.400	0.612	0.466
LASSO					0.357	0.782
RIDGE						0.590

Goyal (rows) vs FRED-MD (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.193	0.548	0.221	0.446	0.271	0.287
3PRF		0.703	0.355	0.715	0.420	0.502
PLS			0.175	0.340	0.183	0.191
PCR				0.816	0.514	0.606
LASSO					0.083	0.129
RIDGE						0.244

Goyal (rows) vs All Instr. (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.281	0.282	0.331	0.366	0.231	0.351
3PRF		0.478	0.518	0.604	0.432	0.558
PLS			0.243	0.269	0.148	0.210
PCR				0.668	0.499	0.608
LASSO					0.063	0.116
RIDGE						0.223

FRED-MD (rows) vs FRED-MD (columns)

FRED-MD	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.997	0.696	0.892	0.440	0.751	0.723
3PRF		0.699	0.881	0.398	0.633	0.682
PLS			0.744	0.903	0.845	0.863
PCR				0.583	0.886	0.845
LASSO					0.569	0.365
RIDGE						0.929

FRED-MD (rows) vs All Instr. (columns)

FRED-MD	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.709	0.585	0.509	0.299	0.454	0.388
3PRF		0.697	0.555	0.454	0.489	0.495
PLS			0.843	0.941	0.806	0.899
PCR				0.688	0.842	0.738
LASSO					0.651	0.823
RIDGE						0.724

All Instr. (rows) vs All Instr. (columns)

All Instr.	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.640	0.635	0.582	0.378	0.554	0.468
3PRF		0.907	0.833	0.687	0.862	0.750
PLS			0.969	0.852	0.981	0.905
PCR				0.871	0.940	0.924
LASSO					0.742	0.869
RIDGE						0.820

Panel C: MVATE

Goyal (rows) vs Goyal (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.871	0.804	0.495	0.122	0.938	0.493
3PRF		0.912	0.551	0.060	0.915	0.405
PLS			0.628	0.128	0.853	0.433
PCR				0.093	0.553	0.285
LASSO					0.039	0.285
RIDGE						0.386

Goyal (rows) vs FRED-MD (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.409	0.611	0.724	0.778	0.789	0.883
3PRF		0.719	0.842	0.687	0.893	0.997
PLS			0.922	0.626	0.959	0.935
PCR				0.383	0.682	0.584
LASSO					0.172	0.204
RIDGE						0.925

Goyal (rows) vs All Instr. (columns)

Goyal	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.365	0.034	0.704	0.567	0.517	0.225
3PRF		0.083	0.911	0.769	0.725	0.369
PLS			0.990	0.873	0.816	0.451
PCR				0.747	0.798	0.760
LASSO					0.066	0.019
RIDGE						0.207

FRED-MD (rows) vs FRED-MD (columns)

FRED-MD	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.065	0.059	0.039	0.003*	0.016	0.015
3PRF		0.718	0.560	0.191	0.324	0.396
PLS			0.869	0.474	0.848	0.723
PCR				0.531	0.955	0.838
LASSO					0.483	0.448
RIDGE						0.859

FRED-MD (rows) vs All Instr. (columns)

FRED-MD	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.070	0.505	0.026	0.026	0.020	0.101
3PRF		0.375	0.477	0.606	0.533	0.909
PLS			0.855	0.964	0.981	0.595
PCR				0.989	0.912	0.510
LASSO					0.465	0.136
RIDGE						0.403

All Instr. (rows) vs All Instr. (columns)

All Instr.	3PRF	PLS	PCR	LASSO	RIDGE	ENET
OLS	0.035	0.364	0.017	0.013	0.016	0.060
3PRF		0.369	0.440	0.594	0.469	0.923
PLS			0.132	0.148	0.222	0.387
PCR				0.888	0.780	0.449
LASSO					0.922	0.256
RIDGE						0.476

p-val	$[0 \le p \le 0.05]$	$(0.05 < p < 0.1)$	$(0.1 \le p \le 1)$
Gradient color bounds			

Test for the differences of the Sharpe ratios of the OOS (jan-1996 - Dec-2017) returns of the efficient portfolios formed from the dataset with 5 industry portfolios using 7 different estimators (OLS, 3PRF, PLS, PCR, LASSO, RIDGE and ENet) and three different set of conditioning information (Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios. Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

B.2 Characteristics and the Distribution of Optimized Portfolio Weights

Figure B.1: Distribution of Optimized Portfolio Weights - 6 Portfolios Formed on Size and Book-to-Market

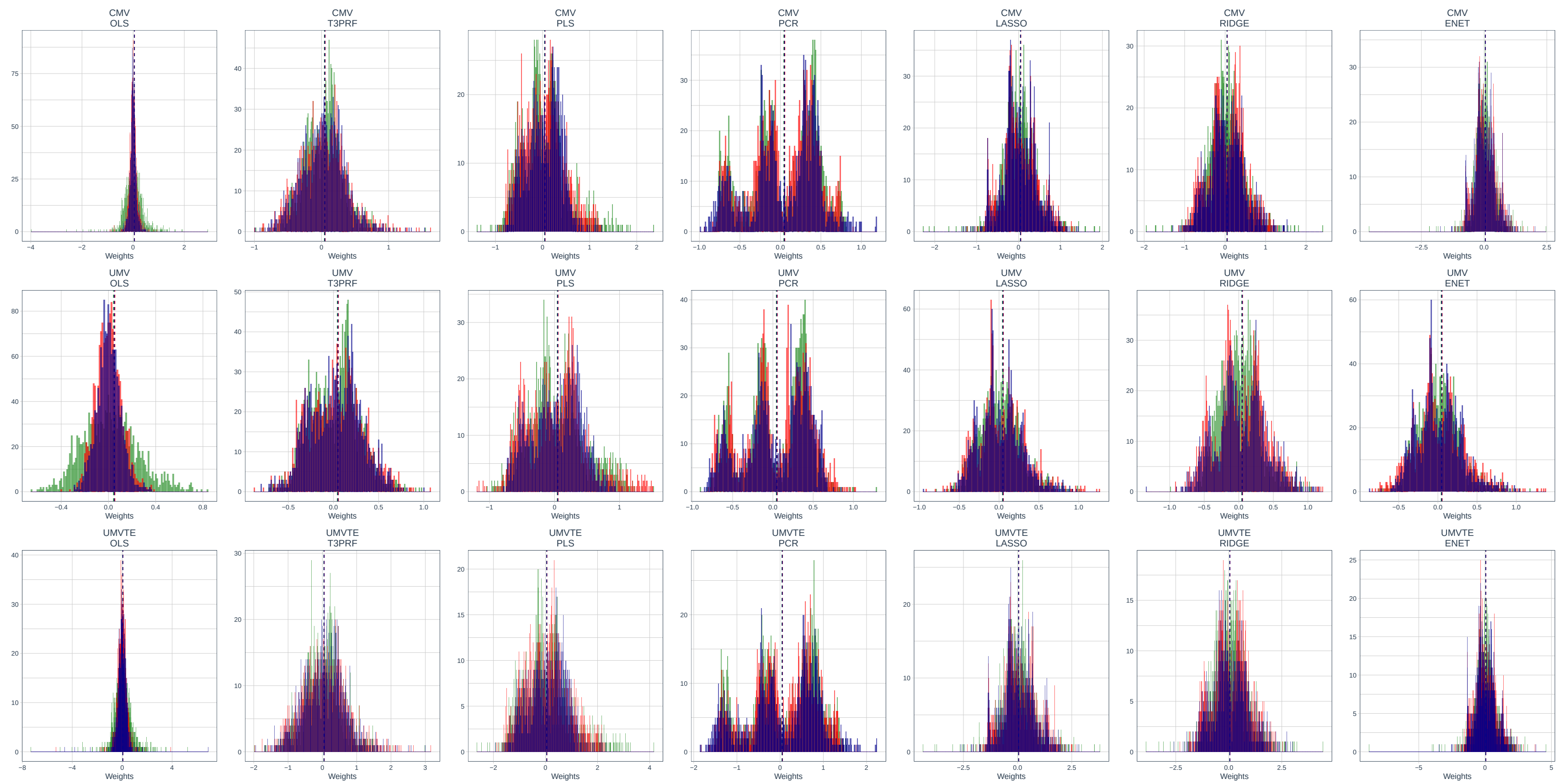


Figure B.1 plots the distribution of the optimized portfolios weights generated by each estimator (columns) and mean-variance optimal framework (rows) for the 6 portfolios formed on Size/BTM. The first row reports the CMV strategies,, the second row reports the UMV, and the third one plots the UMVTE. The set of conditioning information used in **Z** are plotted in different colors: (i) Goyal's in green, (ii) FRED-MD in red, and (iii) "All Instruments" in blue. Goyal's variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. FRED-MD is a large dataset containing 128 macroeconomic and financial variables. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

Figure B.2: Distribution of Optimized Portfolio Weights - 5 Industry Portfolios

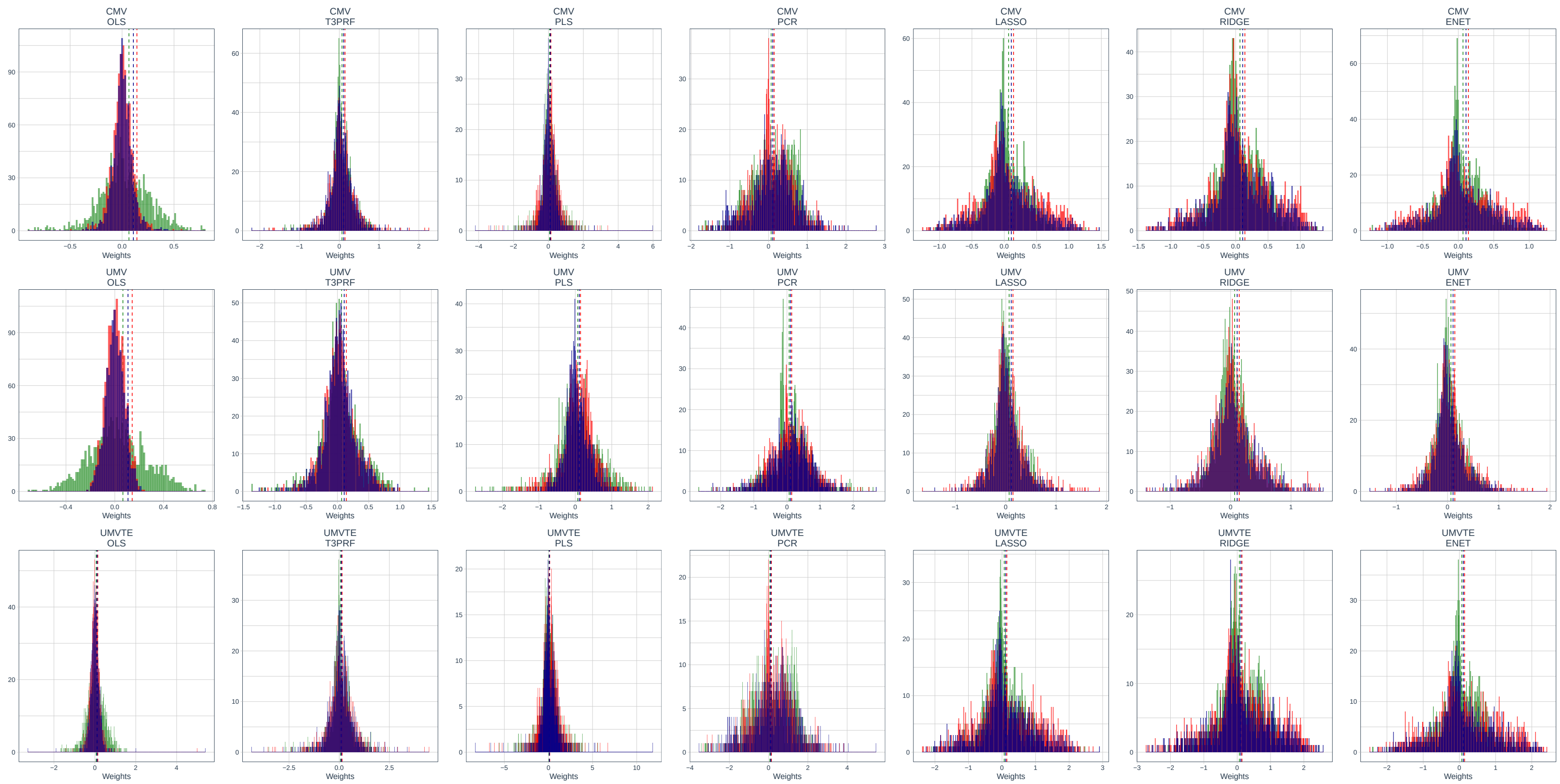


Figure B.2 plots the distribution of the optimized portfolios weights generated by each estimator (columns) and mean-variance optimal framework (rows) for the 5 industry portfolios. The first row reports the CMV strategies,, the second row reports the UMV, and the third one plots the UMVTE. The set of conditioning information used in **Z** are plotted in different colors: (i) Goyal's in green, (ii) FRED-MD in red, and (iii) "All Instruments" in blue. Goyal's variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. FRED-MD is a large dataset containing 128 macroeconomic and financial variables. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

B.3 Variable Contribution

Figure B.3: Variable Contribution by Estimator and Set of Conditioning Information - 6 Portfolios Formed on Size and Book-to-Market

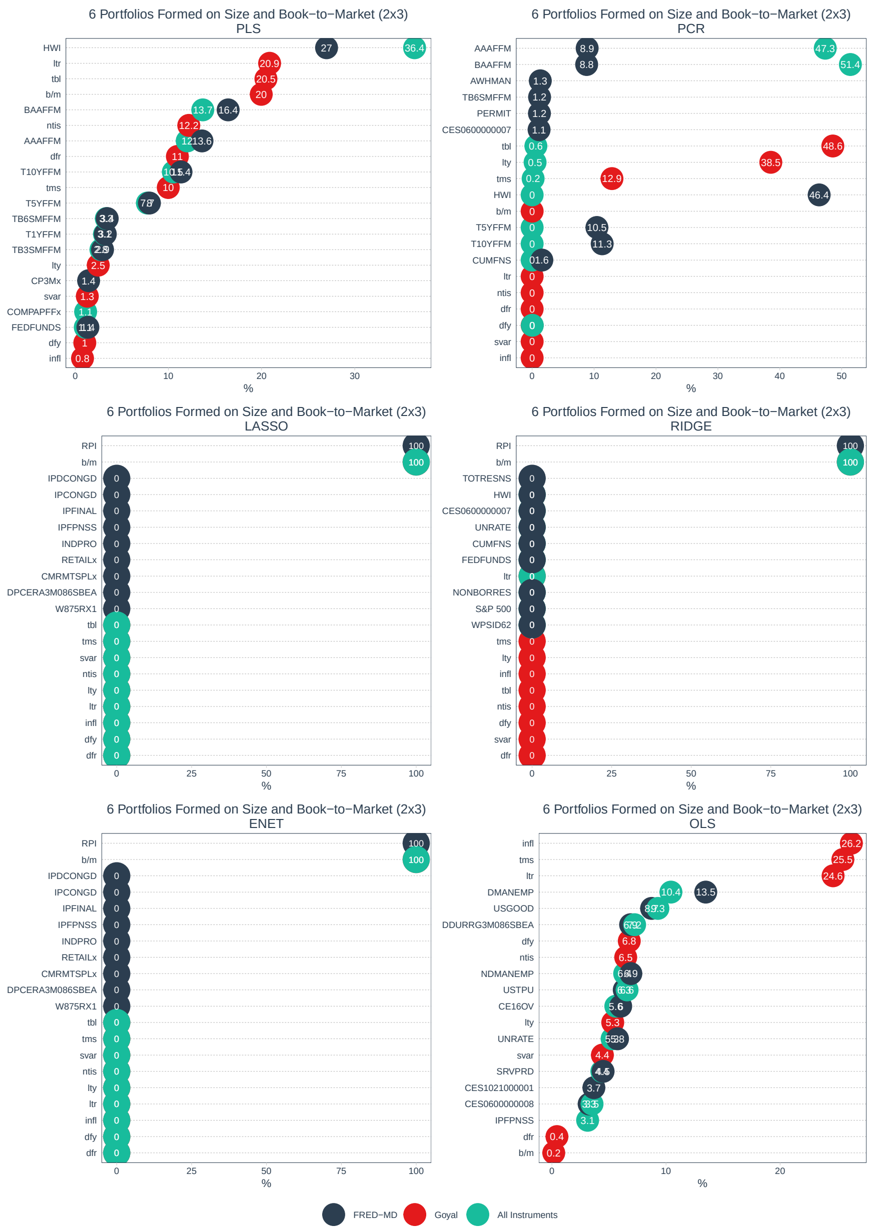


Figure B.3 reports the 10 most influential variables by estimator (panels) and set of conditioning information used (colors) for the for the 6 portfolios formed on Size/BTM. We take a direct approach to obtain the contribution for each variable. At each month t, we compute the absolute values of the estimated coefficients in (6) for each asset i and lagged variable k in the conditioning set. We destandardize the variables whenever necessary. Grouping by estimators and $\mathbf{Z}$ used, we pool all the estimates of the N assets and calculate the average of the absolute values. In order to make the comparison clear, we normalize these means to sum one.

Figure B.4: Variable Contribution by Estimator and Set of Conditioning Information - 5 Industry Portfolios

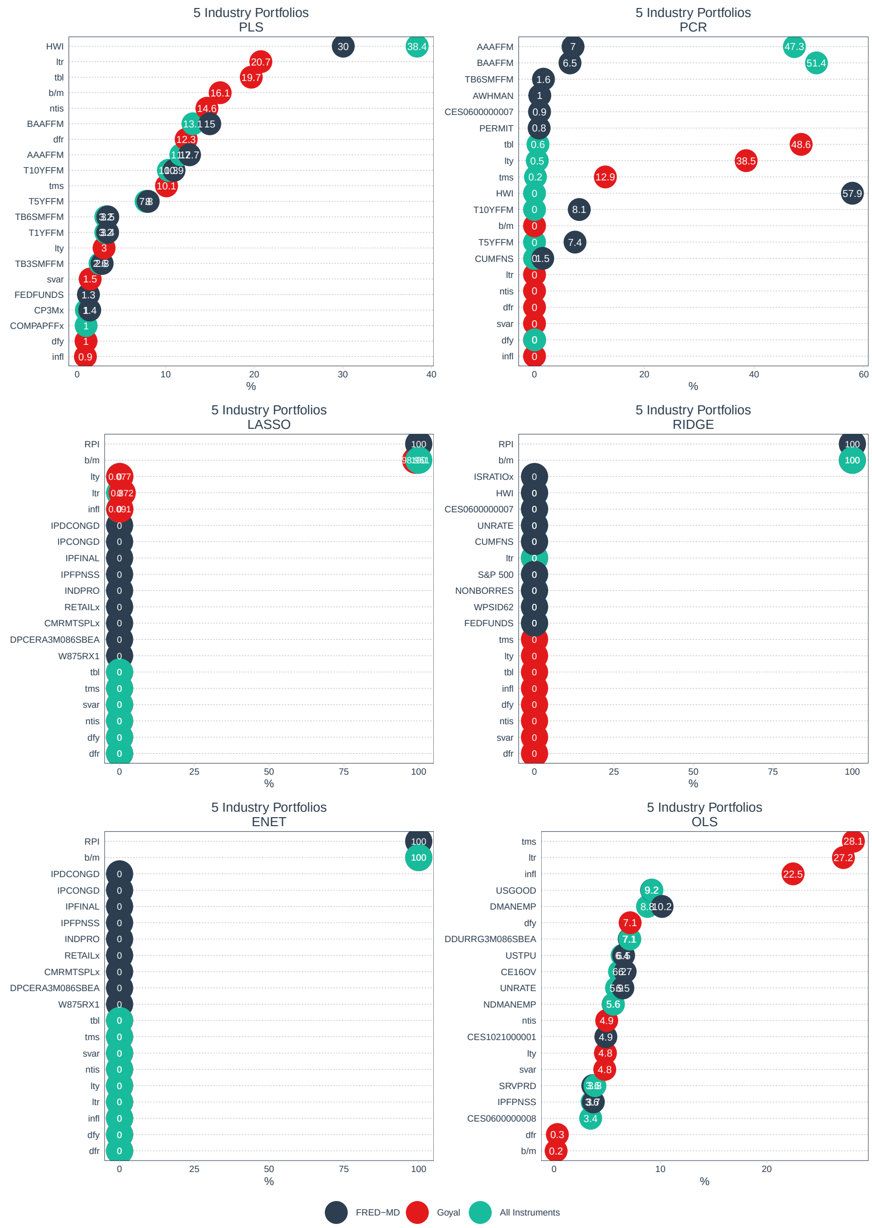


Figure B.4 reports the 10 most influential variables by estimator (panels) and set of conditioning information used (colors) for the for the 5 industry portfolios. We take a direct approach to obtain the contribution for each variable. At each month t, we compute the absolute values of the estimated coefficients in (6) for each asset i and lagged variable k in the conditioning set. We destandardize the variables whenever necessary. Grouping by estimators and $\mathbf{Z}$ used, we pool all the estimates of the N assets and calculate the average of the absolute values. In order to make the comparison clear, we normalize these means to sum one.

Figure B.5: Out-of-Sample R^2

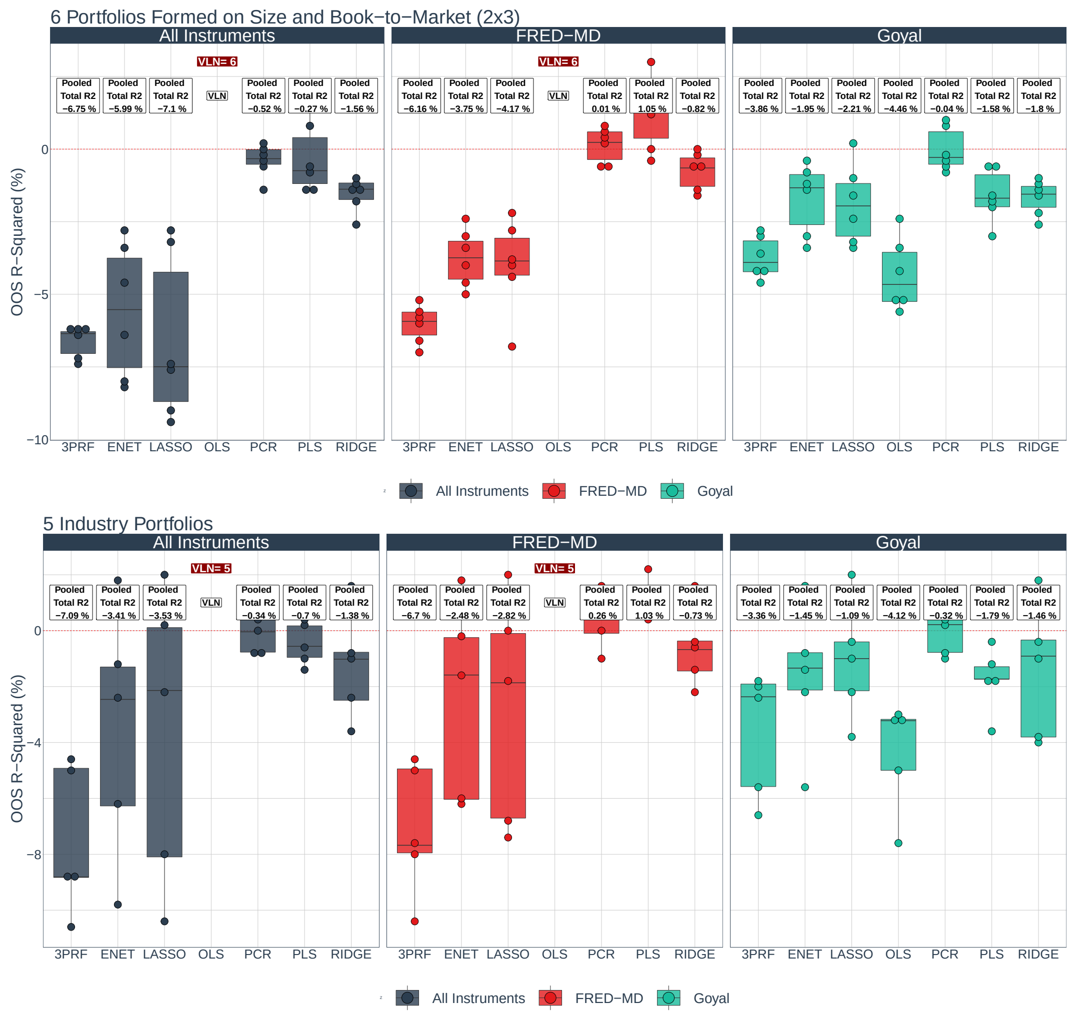


Figure B.5 presents the OOS R^2 for the datasets with 6 portfolios formed on Size/BTM and the 5 industry portfolios by estimator and set of conditioning information used, along with overlapping boxplots. The pooled OOS R^2 across all N risky assets in each dataset is also reported in a white box on the top of each plot. In order to make the plots readable and comparable, we filtered the OOS R^2 larger than -0.5 in absolute values and present the amount of assets that generated these very large numbers (VLN). The amount of VLN per estimator is reported in a red box on top of each plot.

B.4 Economic Value

Table B.4: Implied Sharpe Ratios

	5 Industry Portfolios			6 Portfolios Formed on Size and Book-to-Market		
	CMV	UMV	MVATE	CMV	UMV	MVATE
Panel A: Goyal						
OLS	-	-	-	-	0.055	-
3PRF	-	-	-	0.102	0.164	0.104
PLS	-	-	-	0.211	0.210	0.211
PCR	0.069	0.051	0.063	0.297	0.298	0.299
LASSO	0.181	0.094	0.180	0.076	0.112	0.073
RIDGE	0.051	-	0.051	0.109	0.177	0.110
ENET	0.109	0.048	0.116	0.182	0.217	0.188
Panel B: FRED-MD						
OLS	-	-	-	-	-	-
3PRF	-	-	-	-	-	-
PLS	0.145	0.112	0.144	0.235	0.224	0.226
PCR	0.126	0.055	0.124	0.273	0.255	0.275
LASSO	-	-	-	0.095	-	0.095
RIDGE	0.079	-	0.078	0.224	0.210	0.227
ENET	-	-	-	0.172	0.172	0.172
Panel C: All Instruments						
OLS	-	-	-	-	-	-
3PRF	-	-	-	-	-	-
PLS	-	-	-	0.259	0.258	0.259
PCR	0.119	-	0.115	0.227	0.225	0.228
LASSO	-	-	-	0.049	-	0.062
RIDGE	-	-	-	0.212	0.190	0.212
ENET	-	-	-	0.069	-	0.101

Table B.4 reports the implied SR^* for the 5 Industry portfolios and 6 portfolios formed on Size/BTM as given in equation (30) for all seven estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge and ENet), three different sets of conditioning information (Goyal's, FRED-MD, and "All Instruments", wich is the combination of the previous two with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI)) and three different mean-variance approaches (CMV, UMV, MVATE) to build efficient portfolios. Notice that when the OOS R^2 is negative, larger absolute R^2 values than the Sharpe ratio will cause the impossibility of the implied Sharpe ratio to be calculated. These cases are omitted in the table.

B.4.1 Economic Gains

Table B.5: Certainty Equivalent Excess Returns (CER) (Monthly %)

	5 Industry Portfolios						6 Portfolios Formed on Size and Book-to-Market					
	CMV		UMV		MVATE		CMV		UMV		MVATE	
Panel A: Goyal												
OLS	0.15	[5]	0.10	[4]	0.18	[3]	0.21	[4]	0.19	[6]	0.32	[5]
3PRF	0.15	[4]	0.05	[5]	0.09	[5]	0.20	[6]	0.25	[3]	0.34	[4]
PLS	0.13	[6]	-0.04	[7]	-0.22	[6]	0.32	[2]	0.35	[2]	0.50	[2]
PCR	0.01	[7]	-0.03	[6]	-0.57	[7]	0.42	[1]	0.41	[1]	0.69	[1]
LASSO	0.28	[1]	0.16	[1]	0.42	[1]	0.18	[7]	0.15	[7]	0.26	[7]
RIDGE	0.16	[3]	0.12	[3]	0.13	[4]	0.20	[5]	0.22	[5]	0.29	[6]
ENET	0.22	[2]	0.15	[2]	0.28	[2]	0.25	[3]	0.22	[4]	0.42	[3]
Panel B: FRED-MD												
OLS	-0.01	[6]	0.00	[2]	-0.12	[5]	0.07	[7]	0.06	[7]	0.00	[7]
3PRF	0.04	[5]	-0.06	[4]	-0.09	[4]	0.21	[6]	0.20	[5]	0.33	[6]
PLS	-0.11	[7]	-0.14	[6]	-1.39	[7]	0.27	[5]	0.25	[4]	0.39	[5]
PCR	0.11	[4]	-0.40	[7]	-0.28	[6]	0.40	[1]	0.36	[1]	0.64	[1]
LASSO	0.23	[1]	0.02	[1]	0.19	[1]	0.30	[4]	0.18	[6]	0.46	[4]
RIDGE	0.13	[3]	-0.09	[5]	-0.06	[3]	0.32	[3]	0.27	[3]	0.52	[3]
ENET	0.16	[2]	-0.04	[3]	0.07	[2]	0.34	[2]	0.27	[2]	0.55	[2]
Panel C: All Instruments												
OLS	-0.04	[6]	-0.01	[3]	-0.22	[5]	0.06	[7]	0.06	[7]	-0.09	[7]
3PRF	0.03	[4]	-0.03	[4]	-0.12	[3]	0.21	[6]	0.22	[4]	0.32	[6]
PLS	-0.27	[7]	-0.05	[5]	-1.05	[7]	0.32	[3]	0.32	[2]	0.54	[3]
PCR	0.14	[2]	-0.37	[7]	-0.43	[6]	0.37	[1]	0.33	[1]	0.55	[2]
LASSO	0.15	[1]	0.01	[1]	0.04	[1]	0.33	[2]	0.19	[6]	0.55	[1]
RIDGE	0.12	[3]	-0.10	[6]	-0.08	[2]	0.32	[4]	0.27	[3]	0.51	[5]
ENET	0.03	[5]	0.00	[2]	-0.17	[4]	0.30	[5]	0.19	[5]	0.52	[4]

Table B.5 summarises the CER (monthly %) by estimator and optimal portfolio framework (CMV, UMV and MVATE) for the 5 Industry portfolios and 6 portfolios formed on Size/BTM. Panel A reports the Sharpe ratios generated when the variables from Goyal's website are used as **Z**. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. Panel B presents the Sharpe ratios obtained using the FRED-MD variables. The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, panel C shows the Sharpe rations when all variables are used as conditioning information. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). Inside brackets, we also report the ranking among all 7 estimators within strategy and set of conditioning information employed

Figure B.6: Management Fee

Figure B.6 presents the management fee in bps for the 5 Industry portfolios and 6 portfolios formed on Size/BTM, computed as the solution for $\mathcal{F}$ in equation (36), from an investor switching from an optimal portfolio formed by the estimator and mean-variance framework given in the left axis to another portfolio plotted. The comparison is done in pairs of optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) and sets of conditioning information.

B.5 Out-of-Sample Analysis - Portfolio Efficiency

Figure B.7: p-values

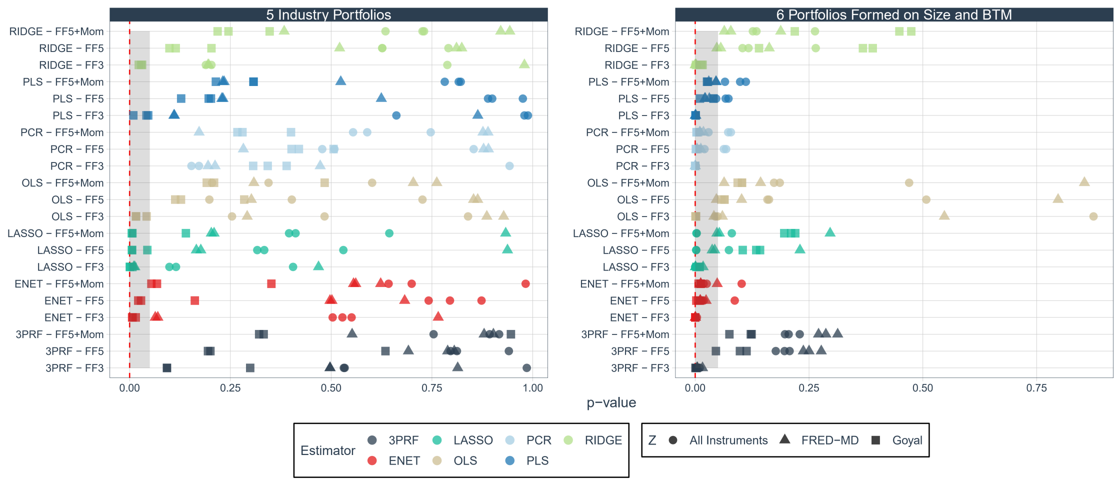


Figure B.7 reports for the 5 Industry portfolios and 6 portfolios formed on Size/BTM the p-values of alphas of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) and set of conditioning information used on the Fama-French 3, 5, and 5 + momentum factor models. The p-values are calculated from Newey–West t-statistics computed with one lag.

Figure B.8: R-Squared

Figure B.8 reports for the 5 Industry portfolios and 6 portfolios formed on Size/BTM the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) and set of conditioning information used on the Fama-French 3, 5, and 5 + momentum factor models.

Table B.6: Alphas (Monthly %) - 25 and 100 Portfolios Formed on Size and Book-to-Market - "All Instruments"

		25 Portfolios Formed on Size and Book-to-Market									100 Portfolios Formed on Size and Book-to-Market								
		CMV			UMV			MVATE			CMV			UMV			MVATE		
		FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom
Z: All Instruments																			
OLS	α (%)	0.034	0.027	0.025	0.034	0.027	0.025	-0.218	-0.246	-0.237	0.018	0.016	0.015	0.018	0.016	0.015	-0.420	-0.226	-0.181
	t-statistics	[3.07]	[2.23]	[2.39]	[3.11]	[2.3]	[2.4]	[-1.1]	[-1.27]	[-1.18]	[1.67]	[1.67]	[1.72]	[1.66]	[1.67]	[1.72]	[-0.79]	[-0.5]	[-0.38]
	p-val	0.002**	0.027*	0.018*	0.002**	0.022*	0.017*	0.273	0.204	0.238	0.097	0.096	0.087	0.098	0.097	0.088	0.433	0.618	0.706
	R^2	0.02	0.04	0.12	0.02	0.04	0.12	0.07	0.07	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.06	0.08
3PRF	α (%)	0.175	0.150	0.141	0.167	0.140	0.131	0.295	0.242	0.229	0.163	0.147	0.147	0.159	0.143	0.143	0.245	0.209	0.221
	t-statistics	[4.01]	[3.3]	[3.4]	[3.77]	[2.96]	[3.24]	[3.3]	[2.56]	[2.66]	[3.7]	[2.72]	[3.01]	[3.57]	[2.64]	[2.91]	[2.18]	[1.83]	[1.84]
	p-val	0**	0.001**	0.001**	0**	0.003**	0.001**	0.001**	0.011*	0.008**	0**	0.007**	0.003**	0**	0.009**	0.004**	0.03*	0.068	0.067
	R^2	0.09	0.11	0.16	0.07	0.09	0.15	0.12	0.15	0.17	0.04	0.05	0.05	0.04	0.05	0.05	0.07	0.08	0.08
PLS	α (%)	0.270	0.204	0.200	0.254	0.175	0.172	0.507	0.375	0.369	0.176	0.148	0.142	0.163	0.136	0.130	0.331	0.257	0.256
	t-statistics	[4]	[3.12]	[2.93]	[3.59]	[2.66]	[2.55]	[3.7]	[2.87]	[2.93]	[4.36]	[3.72]	[3.74]	[4.36]	[3.64]	[3.63]	[2.68]	[2.92]	[2.75]
	p-val	0**	0.002**	0.004**	0**	0.008**	0.011*	0**	0.005**	0.004**	0**	0**	0**	0**	0**	0**	0.008**	0.004**	0.006**
	R^2	0.08	0.16	0.18	0.04	0.15	0.16	0.08	0.17	0.18	0.03	0.07	0.13	0.04	0.08	0.15	0.07	0.11	0.11
PCR	α (%)	0.332	0.278	0.259	0.301	0.265	0.247	0.652	0.542	0.506	0.267	0.251	0.246	0.249	0.232	0.227	0.533	0.498	0.489
	t-statistics	[4.44]	[3.87]	[3.77]	[4.1]	[3.7]	[3.85]	[4.47]	[4.03]	[3.77]	[3.57]	[3.3]	[3.27]	[3.62]	[3.36]	[3.3]	[3.64]	[3.36]	[3.3]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.001**	0.001**	0**	0.001**	0.001**	0**	0.001**	0.001**
	R^2	0.24	0.25	0.38	0.25	0.25	0.37	0.24	0.26	0.38	0.00	0.01	0.02	0.00	0.01	0.03	0.00	0.01	0.02
LASSO	α (%)	0.178	0.139	0.132	0.139	0.114	0.108	0.336	0.260	0.248	0.134	0.121	0.119	0.128	0.116	0.114	0.244	0.221	0.215
	t-statistics	[3.81]	[3.22]	[3.18]	[3.65]	[3.21]	[2.98]	[3.84]	[3.21]	[3.21]	[3.75]	[3.18]	[3.51]	[3.72]	[3.07]	[3.3]	[3.3]	[2.74]	[2.8]
	p-val	0**	0.001**	0.002**	0**	0.002**	0.003**	0**	0.002**	0.002**	0**	0.002**	0.001**	0**	0.002**	0.001**	0.001**	0.007**	0.006**
	R^2	0.08	0.11	0.14	0.08	0.10	0.14	0.08	0.11	0.15	0.04	0.05	0.06	0.05	0.06	0.07	0.03	0.04	0.06
RIDGE	α (%)	0.277	0.207	0.193	0.271	0.208	0.193	0.513	0.380	0.353	0.119	0.099	0.094	0.118	0.099	0.094	0.144	0.108	0.103
	t-statistics	[4.31]	[2.92]	[3.32]	[3.89]	[2.87]	[3.26]	[4.37]	[2.9]	[2.95]	[3.4]	[2.83]	[2.75]	[3.38]	[2.87]	[2.76]	[1.34]	[1.02]	[1.03]
	p-val	0**	0.004**	0.001**	0**	0.004**	0.001**	0**	0.004**	0.004**	0.001**	0.005**	0.006**	0.001**	0.005**	0.006**	0.180	0.307	0.305
	R^2	0.12	0.17	0.28	0.11	0.16	0.29	0.13	0.18	0.28	0.09	0.11	0.15	0.09	0.11	0.15	0.11	0.11	0.12
ENET	α (%)	0.213	0.182	0.177	0.165	0.141	0.137	0.423	0.350	0.344	0.140	0.130	0.128	0.134	0.124	0.122	0.226	0.190	0.192
	t-statistics	[4.79]	[4.19]	[4.05]	[4.13]	[3.48]	[3.48]	[4.64]	[4]	[3.98]	[4.03]	[3.79]	[3.99]	[3.9]	[3.67]	[3.85]	[2.73]	[2.14]	[1.83]
	p-val	0**	0**	0**	0**	0.001**	0.001**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.007**	0.033*	0.069
	R^2	0.10	0.12	0.15	0.08	0.10	0.13	0.06	0.09	0.11	0.05	0.06	0.08	0.06	0.07	0.09	0.01	0.02	0.03

Table B.6 reports the alphas, t-statistics, p-values and the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) using the FRED-MD as conditioning information on the Fama-French 3, 5, and 5 + momentum factor models. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). Goyal's variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. FRED-MD is a large dataset containing 128 macroeconomic and financial variables.

Table B.7: Alphas (Monthly %) - 25 and 100 Portfolios Formed on Size and Book-to-Market - Goyal Variables

		25 Portfolios Formed on Size and Book-to-Market									100 Portfolios Formed on Size and Book-to-Market								
		CMV			UMV			MVATE			CMV			UMV			MVATE		
		FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom
Z: Goyal																			
OLS	α (%)	0.199	0.184	0.169	0.193	0.176	0.163	0.329	0.309	0.287	0.076	0.072	0.065	0.074	0.070	0.064	-0.023	0.002	0.004
	t-statistics	[5.06]	[4.68]	[4.71]	[4.7]	[4.42]	[4.25]	[4.18]	[3.92]	[4.03]	[1.59]	[1.82]	[1.54]	[1.56]	[1.8]	[1.52]	[-0.16]	[0.01]	[0.03]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.112	0.070	0.124	0.120	0.073	0.128	0.873	0.989	0.974
	R^2	0.01	0.05	0.18	0.02	0.06	0.18	0.01	0.05	0.14	0.01	0.03	0.04	0.01	0.03	0.04	0.01	0.02	0.02
3PRF	α (%)	0.226	0.210	0.194	0.244	0.224	0.210	0.403	0.381	0.357	0.143	0.128	0.121	0.146	0.129	0.122	0.134	0.102	0.101
	t-statistics	[4.99]	[5.06]	[4.97]	[4.89]	[4.71]	[4.53]	[4.63]	[4.54]	[4.61]	[2.46]	[3.04]	[2.74]	[2.68]	[3.29]	[2.99]	[0.98]	[0.92]	[0.91]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.015*	0.003**	0.007**	0.008**	0.001**	0.003**	0.328	0.358	0.363
	R^2	0.03	0.06	0.19	0.03	0.07	0.18	0.02	0.05	0.14	0.01	0.02	0.03	0.01	0.02	0.03	0.01	0.02	0.02
PLS	α (%)	0.292	0.254	0.235	0.352	0.303	0.285	0.537	0.472	0.435	0.242	0.217	0.206	0.243	0.217	0.207	0.462	0.421	0.403
	t-statistics	[4.35]	[3.78]	[3.51]	[3.73]	[3.52]	[3.15]	[4.64]	[3.87]	[3.78]	[3.9]	[3.97]	[3.67]	[3.85]	[3.96]	[3.62]	[4.38]	[3.97]	[4.14]
	p-val	0**	0**	0.001**	0**	0.001**	0.002**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**
	R^2	0.11	0.13	0.22	0.09	0.12	0.20	0.12	0.13	0.23	0.01	0.02	0.05	0.01	0.02	0.05	0.02	0.02	0.05
PCR	α (%)	0.367	0.326	0.311	0.381	0.329	0.318	0.700	0.624	0.594	0.277	0.267	0.260	0.272	0.259	0.252	0.529	0.510	0.496
	t-statistics	[5.38]	[4.64]	[4.52]	[4.62]	[4.07]	[3.98]	[5.36]	[4.65]	[4.56]	[4.21]	[3.98]	[3.73]	[4.21]	[3.78]	[3.77]	[4.23]	[3.83]	[3.83]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**	0**
	R^2	0.17	0.19	0.24	0.13	0.16	0.19	0.18	0.20	0.25	0.01	0.02	0.03	0.01	0.02	0.04	0.02	0.03	0.04
LASSO	α (%)	0.169	0.159	0.151	0.151	0.140	0.133	0.326	0.305	0.291	0.089	0.079	0.079	0.092	0.081	0.080	0.164	0.136	0.131
	t-statistics	[4.09]	[4.29]	[3.75]	[4.43]	[4.45]	[4.07]	[4.04]	[4.51]	[3.8]	[1.71]	[1.71]	[1.76]	[1.83]	[1.94]	[1.94]	[1.57]	[1.61]	[1.55]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.088	0.089	0.080	0.069	0.054	0.053	0.118	0.108	0.123
	R^2	0.06	0.07	0.11	0.07	0.08	0.13	0.07	0.08	0.12	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.02	0.02
RIDGE	α (%)	0.258	0.232	0.214	0.239	0.213	0.199	0.456	0.408	0.375	0.092	0.092	0.089	0.104	0.103	0.100	0.120	0.119	0.117
	t-statistics	[5.67]	[5.29]	[5.25]	[5.3]	[5.16]	[4.86]	[5.3]	[4.95]	[5.31]	[1.21]	[1.38]	[1.39]	[1.58]	[1.81]	[1.77]	[0.77]	[0.9]	[0.87]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.225	0.169	0.166	0.116	0.072	0.077	0.444	0.366	0.385
	R^2	0.05	0.07	0.19	0.08	0.11	0.22	0.05	0.07	0.19	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01
ENET	α (%)	0.192	0.179	0.168	0.176	0.157	0.148	0.329	0.300	0.283	0.094	0.085	0.085	0.095	0.084	0.085	0.076	0.052	0.060
	t-statistics	[5.49]	[5.47]	[5]	[5.42]	[5.61]	[5.17]	[4.1]	[4.02]	[3.9]	[1.66]	[1.8]	[1.88]	[1.73]	[1.9]	[1.99]	[0.53]	[0.46]	[0.53]
	p-val	0**	0**	0**	0**	0**	0**	0**	0**	0**	0.099	0.073	0.061	0.084	0.059	0.047*	0.596	0.646	0.595
	R^2	0.04	0.06	0.14	0.06	0.09	0.17	0.04	0.06	0.11	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01

Table B.7 reports the alphas, t-statistics, p-values and the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) using Goyal's variables as conditioning information on the Fama-French 3, 5, and 5 + momentum factor models. Goyal's variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl.*

Table B.8: Alphas from Factor Models (Monthly %) - 5 Industry Portfolios and 6 Portfolios Formed on Size and Book-to-Market - FRED-MD

		5 Industry Portfolios									6 Portfolios Formed on Size and Book-to-Market								
		CMV			UMV			MVATE			CMV			UMV			MVATE		
		FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom
Z: FRED-MD																			
OLS	α (%)	-0.006	-0.007	-0.012	-0.003	-0.006	-0.011	-0.090	-0.092	-0.095	0.070	0.062	0.058	0.062	0.048	0.044	0.039	0.018	0.013
	t-statistics	[-0.14]	[-0.17]	[-0.3]	[-0.09]	[-0.19]	[-0.38]	[-1.06]	[-1.03]	[-1.02]	[2.05]	[2]	[1.86]	[1.89]	[1.64]	[1.47]	[0.6]	[0.26]	[0.18]
	p-val	0.886	0.864	0.762	0.928	0.853	0.704	0.292	0.302	0.308	0.041*	0.047*	0.063	0.060	0.102	0.144	0.547	0.797	0.855
	R^2	0.04	0.05	0.07	0.04	0.05	0.08	0.03	0.04	0.04	0.05	0.06	0.07	0.06	0.08	0.10	0.04	0.04	0.04
3PRF	α (%)	0.091	0.039	0.019	-0.025	-0.045	-0.066	0.178	0.083	0.046	0.210	0.074	0.067	0.203	0.082	0.071	0.399	0.143	0.130
	t-statistics	[0.68]	[0.25]	[0.12]	[-0.24]	[-0.4]	[-0.6]	[0.68]	[0.27]	[0.15]	[2.84]	[1.15]	[1.07]	[2.42]	[1.09]	[1.01]	[2.91]	[1.18]	[1.1]
	p-val	0.497	0.805	0.902	0.814	0.692	0.552	0.497	0.789	0.879	0.005**	0.250	0.287	0.016*	0.277	0.313	0.004**	0.238	0.271
	R^2	0.10	0.12	0.15	0.10	0.10	0.14	0.10	0.12	0.14	0.35	0.49	0.50	0.24	0.38	0.40	0.38	0.52	0.53
PLS	α (%)	0.484	0.495	0.434	0.035	-0.111	-0.128	0.943	0.974	0.856	0.275	0.153	0.141	0.235	0.133	0.113	0.520	0.285	0.260
	t-statistics	[1.6]	[1.2]	[1.2]	[0.17]	[-0.49]	[-0.64]	[1.6]	[1.21]	[1.19]	[3.74]	[2.3]	[2.01]	[3.25]	[2.32]	[2.18]	[3.73]	[2.17]	[2.01]
	p-val	0.110	0.232	0.231	0.864	0.624	0.524	0.111	0.229	0.234	0**	0.022*	0.046*	0.001**	0.021*	0.03*	0**	0.031*	0.046*
	R^2	0.10	0.13	0.17	0.09	0.11	0.12	0.10	0.13	0.17	0.54	0.62	0.63	0.55	0.60	0.65	0.55	0.63	0.64
PCR	α (%)	0.251	0.028	-0.024	-0.150	-0.196	-0.248	0.468	0.050	-0.053	0.393	0.219	0.203	0.324	0.186	0.171	0.771	0.434	0.403
	t-statistics	[1.3]	[0.15]	[-0.14]	[-0.72]	[-1.08]	[-1.37]	[1.25]	[0.14]	[-0.16]	[4.3]	[2.9]	[2.58]	[4.37]	[2.52]	[2.38]	[4.27]	[2.92]	[2.59]
	p-val	0.195	0.878	0.889	0.473	0.282	0.172	0.212	0.890	0.877	0**	0.004**	0.011*	0**	0.012*	0.018*	0**	0.004**	0.01*
	R^2	0.12	0.19	0.26	0.24	0.24	0.28	0.12	0.18	0.25	0.42	0.55	0.57	0.47	0.56	0.58	0.42	0.55	0.57
LASSO	α (%)	0.310	0.201	0.176	0.081	0.010	-0.010	0.575	0.377	0.330	0.283	0.165	0.154	0.168	0.074	0.061	0.548	0.325	0.303
	t-statistics	[2.59]	[1.39]	[1.28]	[0.73]	[0.08]	[-0.08]	[2.51]	[1.35]	[1.26]	[3.47]	[2.04]	[1.94]	[2.39]	[1.2]	[1.05]	[3.49]	[2.09]	[1.99]
	p-val	0.01*	0.166	0.203	0.469	0.937	0.933	0.013*	0.177	0.210	0.001**	0.043*	0.053	0.017*	0.230	0.296	0.001**	0.037*	0.048*
	R^2	0.07	0.10	0.13	0.02	0.04	0.07	0.06	0.10	0.13	0.28	0.37	0.38	0.17	0.29	0.34	0.28	0.37	0.38
RIDGE	α (%)	0.218	0.043	0.012	-0.004	-0.095	-0.131	0.417	0.090	0.034	0.305	0.133	0.122	0.250	0.098	0.086	0.596	0.272	0.254
	t-statistics	[1.3]	[0.22]	[0.07]	[-0.03]	[-0.64]	[-0.87]	[1.3]	[0.24]	[0.1]	[3.97]	[1.92]	[1.77]	[3.15]	[1.4]	[1.32]	[3.96]	[1.99]	[1.86]
	p-val	0.196	0.824	0.943	0.979	0.521	0.383	0.194	0.811	0.921	0**	0.056	0.078	0.002**	0.163	0.188	0**	0.048*	0.064
	R^2	0.12	0.19	0.22	0.09	0.11	0.17	0.12	0.18	0.21	0.41	0.57	0.59	0.38	0.55	0.57	0.42	0.58	0.59
ENET	α (%)	0.246	0.119	0.099	0.037	-0.056	-0.070	0.466	0.235	0.196	0.313	0.203	0.194	0.254	0.151	0.139	0.606	0.399	0.383
	t-statistics	[1.86]	[0.67]	[0.58]	[0.3]	[-0.41]	[-0.49]	[1.82]	[0.68]	[0.59]	[3.82]	[2.57]	[2.53]	[3.61]	[2.27]	[1.98]	[3.72]	[2.55]	[2.53]
	p-val	0.064	0.502	0.561	0.766	0.682	0.623	0.070	0.497	0.556	0**	0.011*	0.012*	0**	0.024*	0.048*	0**	0.011*	0.012*
	R^2	0.06	0.11	0.13	0.02	0.05	0.07	0.06	0.10	0.12	0.33	0.41	0.42	0.27	0.37	0.40	0.32	0.40	0.40

Table B.8 reports the alphas, t-statistics, p-values and the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) using the FRED-MD as conditioning information on the Fama-French 3, 5, and 5 + momentum factor models. FRED-MD is a large dataset containing 128 macroeconomic and financial variables.

Table B.9: Alphas (Monthly %) - 5 Industry Portfolios and 6 Portfolios Formed on Size and Book-to-Market - “All Instruments”

		5 Industry Portfolios									6 Portfolios Formed on Size and Book-to-Market								
		CMV			UMV			MVATE			CMV			UMV			MVATE		
		FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom
Z: All Instruments																			
OLS	α (%)	-0.032	-0.042	-0.047	-0.006	-0.011	-0.015	-0.162	-0.176	-0.178	0.064	0.046	0.044	0.063	0.044	0.041	-0.016	-0.070	-0.070
	t-statistics	[-0.7]	[-0.84]	[-0.95]	[-0.2]	[-0.35]	[-0.52]	[-1.14]	[-1.29]	[-1.27]	[1.98]	[1.41]	[1.37]	[2.02]	[1.4]	[1.33]	[-0.16]	[-0.66]	[-0.72]
	p-val	0.484	0.402	0.345	0.840	0.727	0.602	0.254	0.198	0.205	0.049*	0.159	0.173	0.044*	0.162	0.186	0.875	0.508	0.470
	R^2	0.01	0.03	0.05	0.02	0.04	0.08	0.01	0.02	0.02	0.08	0.11	0.13	0.11	0.15	0.18	0.04	0.05	0.05
3PRF	α (%)	0.086	0.037	0.016	0.002	-0.009	-0.037	0.168	0.080	0.041	0.213	0.086	0.079	0.219	0.096	0.084	0.393	0.154	0.143
	t-statistics	[0.63]	[0.24]	[0.1]	[0.02]	[-0.07]	[-0.31]	[0.62]	[0.26]	[0.13]	[3.05]	[1.35]	[1.29]	[2.6]	[1.26]	[1.2]	[3.05]	[1.3]	[1.27]
	p-val	0.531	0.812	0.917	0.985	0.941	0.754	0.534	0.796	0.893	0.003**	0.177	0.197	0.01**	0.207	0.230	0.003**	0.196	0.205
	R^2	0.09	0.10	0.13	0.13	0.13	0.20	0.09	0.10	0.13	0.35	0.49	0.50	0.23	0.38	0.41	0.39	0.52	0.53
PLS	α (%)	0.003	-0.025	-0.042	0.063	-0.005	-0.041	-0.010	-0.055	-0.086	0.321	0.174	0.164	0.314	0.166	0.151	0.633	0.348	0.329
	t-statistics	[0.01]	[-0.13]	[-0.23]	[0.44]	[-0.03]	[-0.28]	[-0.03]	[-0.14]	[-0.23]	[3.3]	[1.8]	[1.6]	[3.4]	[2.01]	[1.85]	[3.34]	[1.84]	[1.66]
	p-val	0.988	0.900	0.821	0.662	0.976	0.782	0.979	0.889	0.816	0.001**	0.073	0.111	0.001**	0.046*	0.065	0.001**	0.067	0.099
	R^2	0.04	0.05	0.07	0.05	0.09	0.23	0.04	0.05	0.06	0.30	0.43	0.46	0.29	0.41	0.47	0.31	0.44	0.46
PCR	α (%)	0.376	0.168	0.131	0.020	-0.046	-0.083	0.697	0.302	0.231	0.359	0.171	0.160	0.298	0.187	0.178	0.708	0.343	0.321
	t-statistics	[1.43]	[0.71]	[0.59]	[0.07]	[-0.18]	[-0.32]	[1.37]	[0.66]	[0.54]	[3.23]	[1.83]	[1.77]	[3.88]	[2.34]	[2.19]	[3.24]	[1.87]	[1.81]
	p-val	0.153	0.478	0.554	0.943	0.854	0.747	0.172	0.508	0.589	0.001**	0.068	0.078	0**	0.02*	0.029*	0.001**	0.063	0.072
	R^2	0.11	0.16	0.21	0.16	0.17	0.21	0.11	0.16	0.21	0.44	0.55	0.57	0.53	0.58	0.59	0.44	0.56	0.58
LASSO	α (%)	0.227	0.152	0.127	0.085	0.075	0.052	0.419	0.284	0.236	0.336	0.227	0.216	0.199	0.112	0.100	0.652	0.446	0.426
	t-statistics	[1.66]	[1]	[0.85]	[0.83]	[0.63]	[0.46]	[1.58]	[0.97]	[0.82]	[4.42]	[3.09]	[2.96]	[3.15]	[1.79]	[1.76]	[4.45]	[3.11]	[3.04]
	p-val	0.099	0.317	0.396	0.405	0.530	0.644	0.115	0.334	0.412	0**	0.002**	0.003**	0.002**	0.074	0.080	0**	0.002**	0.003**
	R^2	0.11	0.12	0.19	0.08	0.09	0.17	0.10	0.12	0.19	0.21	0.30	0.32	0.12	0.24	0.30	0.21	0.30	0.33
RIDGE	α (%)	0.228	0.088	0.060	0.043	-0.046	-0.082	0.435	0.173	0.121	0.313	0.132	0.121	0.260	0.096	0.083	0.603	0.258	0.239
	t-statistics	[1.32]	[0.49]	[0.34]	[0.27]	[-0.27]	[-0.48]	[1.28]	[0.49]	[0.35]	[3.62]	[1.63]	[1.53]	[2.97]	[1.12]	[1.12]	[3.6]	[1.57]	[1.5]
	p-val	0.189	0.627	0.730	0.788	0.791	0.635	0.203	0.627	0.727	0**	0.104	0.127	0.003**	0.265	0.263	0**	0.117	0.135
	R^2	0.12	0.16	0.22	0.09	0.11	0.20	0.12	0.16	0.21	0.36	0.54	0.57	0.32	0.49	0.54	0.37	0.54	0.57
ENET	α (%)	0.086	-0.045	-0.067	0.062	0.017	-0.002	0.177	-0.071	-0.112	0.286	0.190	0.181	0.185	0.094	0.086	0.581	0.392	0.378
	t-statistics	[0.63]	[-0.33]	[-0.46]	[0.6]	[0.16]	[-0.02]	[0.67]	[-0.26]	[-0.39]	[3.33]	[2.37]	[2.25]	[2.93]	[1.72]	[1.64]	[3.39]	[2.47]	[2.37]
	p-val	0.528	0.742	0.643	0.551	0.873	0.982	0.504	0.795	0.700	0.001**	0.018*	0.025*	0.004**	0.087	0.102	0.001**	0.014*	0.018*
	R^2	0.10	0.14	0.20	0.06	0.07	0.14	0.10	0.13	0.18	0.24	0.32	0.33	0.16	0.28	0.31	0.21	0.29	0.30

Table B.9 reports the alphas, t-statistics, p-values and the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) using the FRED-MD as conditioning information on the Fama-French 3, 5, and 5 + momentum factor models. “All Instruments” is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). Goyal’s variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. FRED-MD is a large dataset containing 128 macroeconomic and financial variables.

Table B.10: Alphas (Monthly %) - 5 Industry Portfolios and 6 Portfolios Formed on Size and Book-to-Market - Goyal Variables

		5 Industry Portfolios									6 Portfolios Formed on Size and Book-to-Market								
		CMV			UMV			MVATE			CMV			UMV			MVATE		
		FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom	FF3	FF5	FF5 + Mom
Z: Goyal																			
OLS	α (%)	0.202	0.138	0.109	0.197	0.122	0.073	0.394	0.277	0.222	0.220	0.125	0.107	0.205	0.117	0.101	0.406	0.238	0.206
	t-statistics	[2.43]	[1.53]	[1.26]	[2.04]	[1.07]	[0.7]	[2.43]	[1.59]	[1.31]	[3.25]	[1.86]	[1.64]	[3.45]	[1.91]	[1.68]	[3.17]	[1.86]	[1.64]
	p-val	0.016*	0.127	0.209	0.042*	0.284	0.484	0.016*	0.113	0.191	0.001**	0.064	0.103	0.001**	0.057	0.093	0.002**	0.064	0.103
	R^2	0.09	0.12	0.22	0.14	0.17	0.35	0.09	0.11	0.20	0.11	0.21	0.27	0.09	0.21	0.29	0.13	0.21	0.27
3PRF	α (%)	0.203	0.153	0.109	0.153	0.072	0.010	0.391	0.304	0.219	0.209	0.113	0.103	0.272	0.147	0.134	0.401	0.222	0.207
	t-statistics	[1.69]	[1.28]	[0.97]	[1.04]	[0.48]	[0.07]	[1.69]	[1.3]	[0.99]	[3.04]	[1.66]	[1.54]	[3.46]	[2.01]	[1.79]	[3.14]	[1.59]	[1.55]
	p-val	0.093	0.201	0.333	0.299	0.635	0.946	0.093	0.195	0.321	0.003**	0.098	0.124	0.001**	0.045*	0.075	0.002**	0.113	0.122
	R^2	0.22	0.23	0.35	0.16	0.22	0.38	0.22	0.23	0.34	0.12	0.24	0.27	0.09	0.26	0.30	0.13	0.24	0.26
PLS	α (%)	0.352	0.236	0.185	0.528	0.408	0.289	0.676	0.462	0.362	0.310	0.204	0.190	0.363	0.219	0.193	0.595	0.399	0.373
	t-statistics	[2.04]	[1.3]	[1.02]	[2.62]	[1.53]	[1.25]	[2]	[1.28]	[1.02]	[3.35]	[2.06]	[2.23]	[3.64]	[2.58]	[2.21]	[3.31]	[2.06]	[2.24]
	p-val	0.042*	0.196	0.307	0.009**	0.128	0.214	0.046*	0.202	0.308	0.001**	0.04*	0.026*	0**	0.01*	0.028*	0.001**	0.04*	0.026*
	R^2	0.13	0.16	0.23	0.15	0.21	0.39	0.13	0.15	0.22	0.34	0.41	0.43	0.25	0.34	0.40	0.34	0.41	0.43
PCR	α (%)	0.184	-0.141	-0.198	0.164	-0.126	-0.164	0.329	-0.285	-0.396	0.405	0.245	0.236	0.401	0.233	0.224	0.781	0.480	0.462
	t-statistics	[1.02]	[-0.81]	[-1.08]	[0.86]	[-0.67]	[-0.84]	[0.95]	[-0.84]	[-1.11]	[4.68]	[3.12]	[2.97]	[4.42]	[2.83]	[2.75]	[4.7]	[3.24]	[3]
	p-val	0.307	0.420	0.280	0.390	0.506	0.400	0.343	0.402	0.267	0**	0.002**	0.003**	0**	0.005**	0.006**	0**	0.001**	0.003**
	R^2	0.11	0.22	0.30	0.12	0.22	0.26	0.11	0.21	0.29	0.41	0.53	0.54	0.39	0.53	0.54	0.42	0.54	0.55
LASSO	α (%)	0.341	0.300	0.260	0.241	0.178	0.139	0.659	0.588	0.510	0.186	0.107	0.092	0.161	0.080	0.069	0.356	0.206	0.179
	t-statistics	[3.43]	[2.76]	[2.73]	[2.72]	[2.02]	[1.48]	[3.41]	[2.78]	[2.81]	[2.6]	[1.47]	[1.23]	[3.05]	[1.63]	[1.3]	[2.63]	[1.5]	[1.25]
	p-val	0.001**	0.006**	0.007**	0.007**	0.044*	0.140	0.001**	0.006**	0.005**	0.01**	0.142	0.220	0.003**	0.104	0.196	0.009**	0.134	0.211
	R^2	0.11	0.11	0.26	0.08	0.10	0.25	0.10	0.11	0.26	0.24	0.29	0.33	0.22	0.32	0.36	0.26	0.31	0.34
RIDGE	α (%)	0.231	0.154	0.118	0.198	0.116	0.078	0.447	0.311	0.241	0.223	0.069	0.058	0.230	0.100	0.086	0.434	0.138	0.117
	t-statistics	[2.2]	[1.58]	[1.17]	[2.29]	[1.28]	[0.94]	[2.17]	[1.66]	[1.23]	[2.42]	[0.86]	[0.72]	[3.18]	[1.48]	[1.23]	[2.47]	[0.9]	[0.76]
	p-val	0.029*	0.114	0.245	0.023*	0.203	0.347	0.031*	0.099	0.218	0.016*	0.390	0.475	0.002**	0.140	0.219	0.014*	0.369	0.448
	R^2	0.13	0.15	0.24	0.16	0.21	0.34	0.12	0.14	0.23	0.26	0.40	0.42	0.13	0.31	0.35	0.27	0.42	0.43
ENET	α (%)	0.296	0.245	0.204	0.225	0.120	0.084	0.594	0.500	0.421	0.264	0.163	0.152	0.229	0.146	0.131	0.525	0.329	0.311
	t-statistics	[2.76]	[2.2]	[1.83]	[2.44]	[1.4]	[0.93]	[2.7]	[2.31]	[1.94]	[3.92]	[2.6]	[2.53]	[4.14]	[3.08]	[2.53]	[4.04]	[2.73]	[2.69]
	p-val	0.006**	0.029*	0.068	0.015*	0.162	0.352	0.007**	0.021*	0.054	0**	0.01**	0.012*	0**	0.002**	0.012*	0**	0.007**	0.008**
	R^2	0.13	0.14	0.26	0.06	0.11	0.23	0.13	0.13	0.25	0.21	0.30	0.32	0.15	0.26	0.33	0.22	0.31	0.32

Table B.10 reports the alphas, t-statistics, p-values and the R^2 of the regressions of the optimal portfolios generated by each mean-variance strategy (CMV, UMV, and MVATE), estimators (OLS, 3PRF, PLS, PCR, LASSO, Ridge, and ENet) using Goyal's variables as conditioning information on the Fama-French 3, 5, and 5 + momentum factor models. Goyal's variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*.

B.6 Financial Metrics

Table B.11: Turnover and Financial Metrics (Monthly %)

	5 Industry Portfolios			6 Portfolios Formed on Size and Book-to-Market		
	Turnover (%)	Max DD (%)	Pain Ratio	Turnover (%)	Max DD (%)	Pain Ratio
Panel A: Goyal						
OLS	59.49	10.16	0.073	122.09	7.02	0.146
3PRF	70.34	15.30	0.097	69.41	8.34	0.145
PLS	121.80	20.64	0.057	80.18	17.08	0.110
PCR	148.52	39.55	0.021	60.73	11.69	0.262
LASSO	96.34	8.70	0.199	176.69	14.29	0.109
RIDGE	96.63	13.93	0.080	126.70	18.74	0.096
ENET	95.01	10.47	0.125	176.96	9.84	0.193
Panel B: FRED-MD						
OLS	48.03	10.87	0.000	83.78	5.77	0.067
3PRF	139.96	20.32	0.013	113.50	14.10	0.134
PLS	143.27	32.41	0.051	91.14	22.45	0.120
PCR	150.09	36.17	0.038	60.61	12.37	0.238
LASSO	170.81	17.67	0.075	129.19	7.09	0.281
RIDGE	170.70	24.34	0.038	103.08	13.92	0.198
ENET	179.47	19.11	0.054	129.43	8.19	0.276
Panel C: All Instruments						
OLS	45.56	11.73	-0.004	68.81	5.08	0.084
3PRF	123.67	24.75	0.013	100.53	13.27	0.138
PLS	131.46	48.28	-0.002	73.32	15.26	0.198
PCR	194.51	44.80	0.033	78.35	11.85	0.187
LASSO	160.82	27.37	0.024	129.23	7.94	0.310
RIDGE	173.12	29.60	0.027	120.69	13.23	0.210
ENET	172.31	30.93	0.009	138.13	8.02	0.257

Table B.11 reports several standard financial metrics for the 5 Industry portfolios and 6 portfolios formed on Size/BTM to evaluate portfolios by estimator for the CMV framework. The Turnover is computed following equation (38). MaxDD is maximum drawdown as presented in equation (39). Pain Ratio is the standard metric, as shown in equation (37). Panel A reports the Sharpe ratios generated when the variables from Goyal's website are used as $\mathbf{Z}$. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl*. Panel B presents the Sharpe ratios obtained using the FRED-MD variables. The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, panel C shows the Sharpe rations when all variables are used as conditioning information. "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI).

C Appendix - Data

Table C.1: FRED-MD

				Full Sample		In-Sample		Out-of-Sample	
Fred Code	**tcode**	**Group**	**Description**	**Mean**	**sd**	**Mean**	**sd**	**Mean**	**sd**
RPI	5	Output and income	Real Personal Income	0.0020	0.0070	0.0020	0.0090	0.0020	0.0070
W875RX1	5	Output and income	Real personal income ex transfer receipts	0.0020	0.0070	0.0020	0.0110	0.0020	0.0070
INDPRO	5	Output and income	IP Index	0.0020	0.0060	0.0040	0.0040	0.0010	0.0070
IPFPNSS	5	Output and income	IP: Final Products and Nonindustrial Supplies	0.0010	0.0060	0.0030	0.0040	0.0010	0.0060
IPFINAL	5	Output and income	IP: Final Products (Market Group)	0.0010	0.0070	0.0030	0.0050	0.0010	0.0070
IPCONGD	5	Output and income	IP: Consumer Goods	0.0010	0.0070	0.0030	0.0060	0.0000	0.0070
IPDCONGD	5	Output and income	IP: Durable Consumer Goods	0.0020	0.0190	0.0060	0.0120	0.0010	0.0200
IPNCONGD	5	Output and income	IP: Nondurable Consumer Goods	0.0000	0.0070	0.0020	0.0050	0.0000	0.0070
IPBUSEQ	5	Output and income	IP: Business Equipment	0.0030	0.0120	0.0060	0.0070	0.0020	0.0130
IPMAT	5	Output and income	IP: Materials	0.0020	0.0080	0.0040	0.0040	0.0020	0.0090
IPDMAT	5	Output and income	IP: Durable Materials	0.0040	0.0110	0.0070	0.0070	0.0030	0.0110
IPNMAT	5	Output and income	IP: Nondurable Materials	0.0000	0.0110	0.0010	0.0070	0.0000	0.0120
IPMANSICS	5	Output and income	IP: Manufacturing (SIC)	0.0020	0.0070	0.0040	0.0050	0.0010	0.0070
IPB51222s	5	Output and income	IP: Residential Utilities	0.0010	0.0400	0.0030	0.0330	0.0010	0.0410
IPFUELS	5	Output and income	IP: Fuels	0.0010	0.0190	0.0010	0.0160	0.0010	0.0190
CUMFNS	2	Output and income	Capacity Utilization: Manufacturing	-0.0090	0.5000	0.0780	0.3790	-0.0240	0.5170
HWI	2	Labor market	Help-Wanted Index for United States	10.9680	220.7630	33.7610	123.4940	6.9960	233.5480
HWIURATIO	2	Labor market	Ratio of Help Wanted/No. Unemployed	0.0020	0.0330	0.0060	0.0180	0.0010	0.0350
CLF16OV	5	Labor market	Civilian Labor Force	0.0010	0.0020	0.0010	0.0020	0.0010	0.0020
CE16OV	5	Labor market	Civilian Employment	0.0010	0.0020	0.0010	0.0020	0.0010	0.0020
UNRATE	2	Labor market	Civilian Unemployment Rate	-0.0110	0.1560	-0.0390	0.1320	-0.0060	0.1590
UEMPMEAN	2	Labor market	Average Duration of Unemployment (Weeks)	0.0230	0.6950	-0.0090	0.4930	0.0280	0.7250
UEMPLT5	5	Labor market	Civilians Unemployed - Less Than 5 Weeks	-0.0010	0.0600	-0.0050	0.0520	-0.0010	0.0620
UEMP5TO14	5	Labor market	Civilians Unemployed for 43599 Weeks	-0.0010	0.0520	-0.0040	0.0500	-0.0010	0.0520
UEMP15OV	5	Labor market	Civilians Unemployed - 15 Weeks & Over	-0.0010	0.0450	-0.0070	0.0460	0.0000	0.0440
UEMP15T26	5	Labor market	Civilians Unemployed for 15-26 Weeks	-0.0020	0.0710	-0.0060	0.0820	-0.0010	0.0700
UEMP27OV	5	Labor market	Civilians Unemployed for 27 Weeks and Over	0.0000	0.0570	-0.0070	0.0500	0.0010	0.0590
CLAIMSx	5	Labor market	Initial Claims	-0.0020	0.0410	-0.0040	0.0380	-0.0020	0.0420
PAYEMS	5	Labor market	All Employees: Total Nonfarm	0.0010	0.0020	0.0020	0.0010	0.0010	0.0020
USGOOD	5	Labor market	All Employees: Goods-Producing Industries	0.0000	0.0040	0.0010	0.0020	-0.0010	0.0040
CES1021000001	5	Labor market	All Employees: Mining and Logging: Mining	0.0000	0.0090	-0.0030	0.0060	0.0010	0.0090
USCONS	5	Labor market	All Employees: Construction	0.0010	0.0060	0.0030	0.0050	0.0010	0.0060
MANEMP	5	Labor market	All Employees: Manufacturing	-0.0010	0.0030	0.0010	0.0010	-0.0010	0.0040
DMANEMP	5	Labor market	All Employees: Durable goods	-0.0010	0.0040	0.0010	0.0020	-0.0010	0.0050
NDMANEMP	5	Labor market	All Employees: Nondurable goods	-0.0010	0.0020	0.0000	0.0010	-0.0010	0.0030
SRVPRD	5	Labor market	All Employees: Service-Providing Industries	0.0010	0.0010	0.0020	0.0010	0.0010	0.0010
USTPU	5	Labor market	All Employees: Trade, Transportation & Utilities	0.0010	0.0020	0.0020	0.0020	0.0010	0.0020
USWTRADE	5	Labor market	All Employees: Wholesale Trade	0.0000	0.0020	0.0010	0.0020	0.0000	0.0020
USTRADE	5	Labor market	All Employees: Retail Trade	0.0010	0.0020	0.0020	0.0020	0.0000	0.0020
USFIRE	5	Labor market	All Employees: Financial Activities	0.0010	0.0020	0.0010	0.0020	0.0010	0.0020
USGOVT	5	Labor market	All Employees: Government	0.0010	0.0020	0.0010	0.0010	0.0010	0.0030
CES0600000007	1	Labor market	Avg Weekly Hours : Goods-Producing	40.6080	0.5840	40.7170	0.3270	40.5890	0.6170
AWOTMAN	2	Labor market	Avg Weekly Overtime Hours : Manufacturing	0.0020	0.1130	0.0150	0.1560	-0.0010	0.1040
AWHMAN	1	Labor market	Avg Weekly Hours : Manufacturing	41.1990	0.5940	41.2390	0.3630	41.1920	0.6250
CES0600000008	6	Labor market	Avg Hourly Earnings : Goods-Producing	0.0000	0.0030	0.0000	0.0020	0.0000	0.0030
CES2000000008	6	Labor market	Avg Hourly Earnings : Construction	0.0000	0.0060	0.0000	0.0060	0.0000	0.0050
CES3000000008	6	Labor market	Avg Hourly Earnings : Manufacturing	0.0000	0.0030	0.0000	0.0020	0.0000	0.0030
HOUST	4	Housing	Housing Starts: Total New Privately Owned	7.1220	0.3740	7.1890	0.0920	7.1110	0.4020
HOUSTNE	4	Housing	Housing Starts, Northeast	4.8070	0.3400	4.8340	0.1200	4.8020	0.3640
HOUSTMW	4	Housing	Housing Starts, Midwest	5.4140	0.4800	5.6980	0.1040	5.3640	0.5020
HOUSTS	4	Housing	Housing Starts, South	6.3780	0.3470	6.3630	0.1180	6.3800	0.3730
HOUSTW	4	Housing	Housing Starts, West	5.6850	0.4370	5.7590	0.1180	5.6720	0.4700
PERMIT	4	Housing	New Private Housing Permits (SAAR)	7.1370	0.3670	7.1320	0.1050	7.1380	0.3950
PERMITNE	4	Housing	New Private Housing Permits, Northeast (SAAR)	4.8700	0.3340	4.8650	0.1040	4.8710	0.3600
PERMITMW	4	Housing	New Private Housing Permits, Midwest (SAAR)	5.4310	0.4410	5.6480	0.0880	5.3930	0.4660
PERMITS	4	Housing	New Private Housing Permits, South (SAAR)	6.3790	0.3400	6.2720	0.1320	6.3980	0.3610
PERMITW	4	Housing	New Private Housing Permits, West (SAAR)	5.7120	0.4350	5.7260	0.1240	5.7090	0.4680
DPCERA3M086SBEA	5	Consumption, orders, and inventories	Real Personal Consumption Expenditures	0.0020	0.0030	0.0030	0.0030	0.0020	0.0030
CMRMTSPLx	5	Consumption, orders, and inventories	Real Manu. and Trade Industries Sales	0.0020	0.0080	0.0040	0.0080	0.0020	0.0080
RETAILx	5	Consumption, orders, and inventories	Retail and Food Services Sales	0.0040	0.0090	0.0050	0.0080	0.0030	0.0100
ACOGNO	5	Consumption, orders, and inventories	New Orders for Consumer Goods	0.0030	0.0190	0.0050	0.0150	0.0020	0.0190
AMDMNOx	5	Consumption, orders, and inventories	New Orders for Durable Goods	0.0020	0.0410	0.0080	0.0250	0.0020	0.0440
ANDENOx	5	Consumption, orders, and inventories	New Orders for Nondefense Capital Goods	0.0020	0.0810	0.0100	0.0480	0.0010	0.0850
AMDMUOx	5	Consumption, orders, and inventories	Unfilled Orders for Durable Goods	0.0030	0.0100	-0.0010	0.0050	0.0040	0.0100
BUSINVx	5	Consumption, orders, and inventories	Total Business Inventories	0.0030	0.0050	0.0040	0.0030	0.0020	0.0050
ISRATIOx	2	Consumption, orders, and inventories	Total Business: Inventories to Sales Ratio	-0.0010	0.0140	-0.0020	0.0130	0.0000	0.0150
UMCSENTx	2	Consumption, orders, and inventories	Consumer Sentiment Index	0.0870	3.8890	0.4830	3.5940	0.0190	3.9410

(Continued)

Table C.1: FRED-MD *(Continued)*

Fred Code	tcode	Group	Description	Full Sample		In-Sample		Out-of-Sample	
				Mean	sd	Mean	sd	Mean	sd
M1SL	6	Money and credit	M1 Money Stock	0.000	0.012	0.000	0.003	0.000	0.013
M2SL	6	Money and credit	M2 Money Stock	0.000	0.004	0.000	0.003	0.000	0.004
M2REAL	5	Money and credit	Real M2 Money Stock	0.003	0.005	-0.001	0.003	0.003	0.005
AMBSL	6	Money and credit	St. Louis Adjusted Monetary Base	0.000	0.025	0.000	0.003	0.000	0.027
TOTRESNS	6	Money and credit	Total Reserves of Depository Institutions	0.000	0.089	0.001	0.037	0.000	0.096
NONBORRES	7	Money and credit	Reserves Of Depository Institutions	0.000	1.618	0.001	0.036	0.000	1.754
BUSLOANS	6	Money and credit	Commercial and Industrial Loans	0.000	0.006	0.000	0.005	0.000	0.007
REALLN	6	Money and credit	Real Estate Loans at All Commercial Banks	0.000	0.007	0.000	0.003	0.000	0.007
NONREVSL	6	Money and credit	Total Nonrevolving Credit	0.000	0.008	0.000	0.007	0.000	0.008
CONSPI	2	Money and credit	Nonrevolving Consumer Credit to Personal Income	0.000	0.001	0.000	0.001	0.000	0.001
MZMSL	6	Money and credit	MZM Money Stock	0.000	0.005	0.000	0.004	0.000	0.005
DTCOLNVHFNM	6	Money and credit	Consumer Motor Vehicle Loans Outstanding	0.000	0.029	0.000	0.024	0.000	0.030
DTCTHFNM	6	Money and credit	Total Consumer Loans and Leases Outstanding	0.000	0.028	0.000	0.011	0.000	0.030
INVEST	6	Money and credit	Securities in Bank Credit at All Commercial Banks	0.000	0.012	0.000	0.012	0.000	0.012
FEDFUNDS	2	Interest and exchange rates	Effective Federal Funds Rate	-0.009	0.165	0.033	0.178	-0.016	0.162
CP3Mx	2	Interest and exchange rates	3-Month AA Financial Commercial Paper Rate	-0.008	0.194	0.033	0.207	-0.016	0.192
TB3MS	2	Interest and exchange rates	3-Month Treasury Bill	-0.008	0.177	0.028	0.180	-0.014	0.176
TB6MS	2	Interest and exchange rates	6-Month Treasury Bill	-0.008	0.175	0.026	0.209	-0.014	0.168
GS1	2	Interest and exchange rates	1-Year Treasury Rate	-0.008	0.195	0.022	0.260	-0.014	0.181
GS5	2	Interest and exchange rates	5-Year Treasury Rate	-0.014	0.236	-0.023	0.279	-0.013	0.228
GS10	2	Interest and exchange rates	10-Year Treasury Rate	-0.016	0.220	-0.035	0.239	-0.013	0.217
AAA	2	Interest and exchange rates	Moody's Seasoned Aaa Corporate Bond	-0.015	0.177	-0.032	0.179	-0.013	0.176
BAA	2	Interest and exchange rates	Moody's Seasoned Baa Corporate Bond	-0.016	0.201	-0.038	0.191	-0.012	0.203
COMPAPFFx	1	Interest and exchange rates	3-Month Commercial Paper Minus	0.144	0.242	0.255	0.229	0.125	0.239
TB3SMFFM	1	Interest and exchange rates	3-Month Treasury C Minus	-0.196	0.259	-0.099	0.210	-0.213	0.264
TB6SMFFM	1	Interest and exchange rates	6-Month Treasury C Minus	-0.090	0.307	0.081	0.341	-0.120	0.292
T1YFFM	1	Interest and exchange rates	1-Year Treasury C Minus	0.140	0.409	0.514	0.512	0.075	0.351
T5YFFM	1	Interest and exchange rates	5-Year Treasury C Minus	1.110	0.965	1.939	0.961	0.965	0.891
T10YFFM	1	Interest and exchange rates	10-Year Treasury C Minus	1.701	1.251	2.462	1.143	1.568	1.223
AAAFFM	1	Interest and exchange rates	Moody's Aaa Corporate Bond Minus	3.167	1.451	3.555	1.195	3.100	1.483
BAAFFM	1	Interest and exchange rates	Moody's Baa Corporate Bond Minus	4.115	1.658	4.250	1.256	4.092	1.720
TWEXMMTH	5	Interest and exchange rates	Trade Weighted U.S. Dollar Index: Major Currencies	0.000	0.016	-0.001	0.016	0.000	0.017
EXSZUSx	5	Interest and exchange rates	Switzerland / U.S. Foreign Exchange Rate	-0.001	0.025	-0.005	0.031	-0.001	0.024
EXJPUSx	5	Interest and exchange rates	Japan / U.S. Foreign Exchange Rate	0.000	0.026	-0.005	0.030	0.000	0.025
EXUSUKx	5	Interest and exchange rates	U.S. / U.K. Foreign Exchange Rate	-0.001	0.022	-0.003	0.029	-0.001	0.021
EXCAUSx	5	Interest and exchange rates	Canada / U.S. Foreign Exchange Rate	0.000	0.018	0.003	0.011	0.000	0.018
WPSFD49207	6	Prices	PPI: Finished Goods	0.000	0.008	0.000	0.003	0.000	0.008
WPSFD49502	6	Prices	PPI: Finished Consumer Goods	0.000	0.010	0.000	0.004	0.000	0.011
WPSID61	6	Prices	PPI: Intermediate Materials	0.000	0.009	0.000	0.003	0.000	0.009
WPSID62	6	Prices	PPI: Crude Materials	0.000	0.051	0.000	0.020	0.000	0.055
OILPRICEx	6	Prices	Crude Oil, Spliced WTI and Cushing	0.000	0.099	0.001	0.062	0.000	0.105
PPICMM	6	Prices	PPI: Metals and Metal Products:	0.000	0.039	-0.001	0.022	0.000	0.042
CPIAUCSL	6	Prices	CPI : All Items	0.000	0.003	0.000	0.001	0.000	0.003
CPIAPPSL	6	Prices	CPI : Apparel	0.000	0.006	0.000	0.006	0.000	0.006
CPITRNSL	6	Prices	CPI : Transportation	0.000	0.015	0.000	0.005	0.000	0.016
CPIMEDSL	6	Prices	CPI : Medical Care	0.000	0.002	0.000	0.001	0.000	0.002
CUSR0000SAC	6	Prices	CPI : Commodities	0.000	0.007	0.000	0.003	0.000	0.007
CUSR0000SAD	6	Prices	CPI : Durables	0.000	0.002	0.000	0.002	0.000	0.002
CUSR0000SAS	6	Prices	CPI : Services	0.000	0.001	0.000	0.001	0.000	0.001
CPIULFSL	6	Prices	CPI : All Items Less Food	0.000	0.003	0.000	0.001	0.000	0.003
CUSR0000SA0L2	6	Prices	CPI : All Items Less Shelter	0.000	0.004	0.000	0.002	0.000	0.004
CUSR0000SA0L5	6	Prices	CPI : All Items Less Medical Care	0.000	0.003	0.000	0.001	0.000	0.003
PCEPI	6	Prices	Personal Cons. Expend.: Chain Index	0.000	0.002	0.000	0.001	0.000	0.002
DDURRG3M086SBEA	6	Prices	Personal Cons. Exp: Durable goods	0.000	0.003	0.000	0.003	0.000	0.003
DNDGRG3M086SBEA	6	Prices	Personal Cons. Exp: Nondurable goods	0.000	0.008	0.000	0.003	0.000	0.008
DSERRG3M086SBEA	6	Prices	Personal Cons. Exp: Services	0.000	0.002	0.000	0.001	0.000	0.002
S&P 500	5	Stock market	S&P's Common Stock Price Index: Composite	0.006	0.035	0.009	0.017	0.006	0.038
S&P div yield	2	Stock market	S&P's Composite Common Stock: Dividend Yield	-0.004	0.077	-0.016	0.046	-0.002	0.081
S&P PE ratio	5	Stock market	S&P's Composite Common Stock: Price-Earnings Ratio	0.001	0.053	-0.004	0.018	0.001	0.057
VXOCLSx	1	Stock market	VXO	19.650	8.723	13.341	1.497	20.750	8.992

As in McCracken and Ng (2016) we transform the variables following the code presented in column 'tcode'. The transformation for a series x are: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $log(x_t)$, (5) $\Delta log(x_t)$, (6) $\Delta^2 log(x_t)$, and (7) $\Delta(x_t/x_{t-1} - 1)$. The column 'gsi' and 'gsi:description' present the comparable series in Global Insight.

Table C.2: Results for the 5 Industry Portfolios

A: CMV

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p-val	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.002	0.015	-0.108	0.018	0.133	0.062	0.462	0.214	2.157	0.032	0.038	12.087	0.000
3PRF	0.003	0.021	-0.106	0.089	0.125	0.062	0.434	0.214	2.028	0.044	0.035	12.386	0.000
PLS	0.004	0.032	-0.108	0.116	0.121	0.062	0.419	0.214	1.959	0.051	0.040	10.391	0.000
PCR	0.003	0.035	-0.037	-0.003	0.089	0.062	0.309	0.214	1.449	0.149	0.049	6.355	0.000
LASSO	0.004	0.017	-0.080	0.180	0.209	0.062	0.725	0.216	3.366	0.001	0.032	22.390	0.000
RIDGE	0.003	0.020	-0.185	0.239	0.131	0.062	0.454	0.214	2.123	0.035	0.040	11.321	0.000
ENET	0.003	0.019	0.004	0.216	0.163	0.062	0.565	0.215	2.633	0.009	0.042	13.361	0.000
FRED-MD													
OLS	0.000	0.006	0.003	0.102	0.005	0.062	0.018	0.213	0.083	0.934	0.046	0.384	0.702
3PRF	0.001	0.020	0.070	0.074	0.072	0.062	0.248	0.213	1.164	0.246	0.051	4.915	0.000
PLS	0.005	0.050	-0.027	-0.037	0.102	0.062	0.354	0.214	1.657	0.099	0.032	11.223	0.000
PCR	0.004	0.033	-0.106	0.218	0.115	0.062	0.398	0.214	1.861	0.064	0.043	9.335	0.000
LASSO	0.004	0.023	-0.084	0.077	0.156	0.062	0.540	0.214	2.517	0.012	0.041	13.134	0.000
RIDGE	0.003	0.026	0.019	0.013	0.116	0.062	0.402	0.214	1.881	0.061	0.049	8.250	0.000
ENET	0.003	0.023	-0.048	0.051	0.127	0.062	0.438	0.214	2.048	0.042	0.042	10.477	0.000
All Instruments													
OLS	0.000	0.007	-0.032	0.077	-0.035	0.066	-0.123	0.230	-0.534	0.594	0.046	-2.682	0.008
3PRF	0.001	0.020	0.081	0.081	0.067	0.062	0.231	0.213	1.081	0.281	0.053	4.380	0.000
PLS	0.000	0.033	0.001	-0.034	0.002	0.066	0.007	0.230	0.030	0.976	0.054	0.128	0.898
PCR	0.005	0.039	-0.038	0.157	0.133	0.067	0.461	0.231	1.998	0.047	0.058	7.949	0.000
LASSO	0.003	0.022	0.000	0.166	0.122	0.067	0.424	0.231	1.837	0.067	0.052	8.127	0.000
RIDGE	0.003	0.026	-0.002	0.035	0.113	0.067	0.392	0.231	1.698	0.091	0.056	6.944	0.000
ENET	0.002	0.023	-0.053	0.146	0.069	0.066	0.239	0.230	1.037	0.301	0.051	4.638	0.000

B: UMV

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.002	0.019	-0.149	0.086	0.100	0.062	0.345	0.214	1.616	0.107	0.038	9.176	0.000
3PRF	0.002	0.026	-0.051	0.096	0.083	0.062	0.286	0.214	1.341	0.181	0.043	6.590	0.000
PLS	0.005	0.047	-0.308	0.220	0.109	0.062	0.378	0.214	1.767	0.078	0.030	12.410	0.000
PCR	0.003	0.034	0.091	-0.072	0.076	0.062	0.264	0.214	1.239	0.217	0.055	4.795	0.000
LASSO	0.002	0.016	-0.132	0.061	0.141	0.062	0.488	0.214	2.278	0.024	0.036	13.381	0.000
RIDGE	0.002	0.018	-0.148	0.082	0.112	0.062	0.387	0.214	1.809	0.072	0.040	9.716	0.000
ENET	0.002	0.017	-0.150	0.192	0.130	0.062	0.449	0.214	2.098	0.037	0.041	11.050	0.000
FRED-MD													
OLS	0.000	0.005	-0.045	0.100	0.009	0.062	0.032	0.213	0.150	0.881	0.044	0.735	0.463
3PRF	0.000	0.017	0.007	0.088	0.009	0.062	0.031	0.213	0.147	0.883	0.046	0.676	0.499
PLS	0.002	0.034	0.079	-0.103	0.046	0.062	0.160	0.213	0.749	0.455	0.055	2.915	0.004
PCR	0.001	0.044	-0.109	0.190	0.019	0.062	0.064	0.213	0.301	0.764	0.043	1.480	0.140
LASSO	0.001	0.019	-0.125	0.012	0.056	0.062	0.194	0.213	0.911	0.363	0.040	4.893	0.000
RIDGE	0.001	0.025	-0.038	-0.011	0.028	0.062	0.095	0.213	0.448	0.655	0.046	2.056	0.041
ENET	0.001	0.021	-0.104	0.025	0.032	0.062	0.112	0.213	0.524	0.601	0.040	2.772	0.006
All Instruments													
OLS	0.000	0.005	-0.101	0.101	-0.002	0.066	-0.006	0.230	-0.027	0.979	0.049	-0.126	0.900
3PRF	0.001	0.018	0.020	0.162	0.029	0.062	0.100	0.213	0.468	0.640	0.046	2.151	0.032
PLS	0.001	0.023	-0.028	-0.049	0.037	0.066	0.127	0.230	0.552	0.581	0.055	2.303	0.022
PCR	0.002	0.047	-0.075	0.171	0.040	0.066	0.137	0.230	0.596	0.552	0.053	2.586	0.010
LASSO	0.001	0.019	-0.110	0.167	0.052	0.066	0.180	0.230	0.781	0.435	0.048	3.713	0.000
RIDGE	0.001	0.028	-0.097	0.003	0.035	0.066	0.120	0.230	0.522	0.602	0.049	2.463	0.015
ENET	0.001	0.018	-0.105	0.051	0.047	0.066	0.162	0.230	0.702	0.483	0.047	3.417	0.001

(Continued)

Table C.2: Results for the 5 Industry Portfolios *(Continued)*

C: MVATE

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.004	0.030	-0.121	0.015	0.135	0.062	0.467	0.214	2.181	0.030	0.037	12.513	0.000
3PRF	0.005	0.042	-0.107	0.091	0.125	0.062	0.435	0.214	2.030	0.043	0.035	12.492	0.000
PLS	0.007	0.062	-0.107	0.114	0.119	0.062	0.414	0.214	1.934	0.054	0.040	10.274	0.000
PCR	0.006	0.068	-0.038	-0.007	0.085	0.062	0.295	0.214	1.383	0.168	0.049	6.076	0.000
LASSO	0.007	0.033	-0.082	0.184	0.209	0.062	0.723	0.216	3.355	0.001	0.031	23.150	0.000
RIDGE	0.005	0.039	-0.192	0.243	0.131	0.062	0.455	0.214	2.125	0.034	0.040	11.503	0.000
ENET	0.006	0.037	-0.001	0.222	0.168	0.062	0.581	0.215	2.704	0.007	0.037	15.738	0.000
FRED-MD													
OLS	-0.001	0.014	0.019	0.115	-0.046	0.062	-0.160	0.213	-0.748	0.455	0.044	-3.608	0.000
3PRF	0.003	0.038	0.071	0.080	0.073	0.062	0.252	0.213	1.180	0.239	0.050	5.011	0.000
PLS	0.010	0.097	-0.026	-0.038	0.101	0.062	0.351	0.214	1.640	0.102	0.032	11.038	0.000
PCR	0.007	0.063	-0.104	0.221	0.112	0.062	0.389	0.214	1.821	0.070	0.043	9.116	0.000
LASSO	0.007	0.044	-0.081	0.074	0.152	0.062	0.528	0.214	2.461	0.014	0.041	12.803	0.000
RIDGE	0.006	0.051	0.022	0.013	0.116	0.062	0.401	0.214	1.873	0.062	0.049	8.174	0.000
ENET	0.005	0.044	-0.036	0.044	0.125	0.062	0.434	0.214	2.026	0.044	0.043	10.166	0.000
All Instruments													
OLS	-0.001	0.018	0.022	0.136	-0.078	0.066	-0.271	0.230	-1.179	0.240	0.042	-6.499	0.000
3PRF	0.003	0.039	0.084	0.082	0.068	0.062	0.235	0.213	1.102	0.271	0.053	4.459	0.000
PLS	0.000	0.065	0.002	-0.038	-0.001	0.066	-0.003	0.230	-0.011	0.991	0.053	-0.048	0.962
PCR	0.010	0.075	-0.035	0.154	0.129	0.067	0.447	0.231	1.936	0.054	0.058	7.669	0.000
LASSO	0.005	0.043	0.003	0.172	0.118	0.067	0.408	0.231	1.768	0.078	0.052	7.811	0.000
RIDGE	0.006	0.051	0.003	0.036	0.112	0.067	0.388	0.231	1.680	0.094	0.056	6.868	0.000
ENET	0.003	0.044	-0.045	0.147	0.069	0.066	0.239	0.230	1.038	0.300	0.052	4.586	0.000

Table C.2 reports some summary statistics of the OOS (Jan-1996 to Dec-2017) for the 5 Industry portfolios by estimator and three different set of conditioning information (Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios (UMV). Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). The table reports the monthly returns ($\hat{\mu}$), monthly returns ($\hat{\sigma}$), first and second autocorrelation ($\hat{\rho}_1$ and $\hat{\rho}_2$) of the monthly returns. The Sharpe ratios (SR) are presented monthly and annualized, as well as the standard error (se). For the sake of completeness, it is also reported the t-statistic (t), the standard p-value (p), as well as the HAC standard error ($se(HAC)$), the t-statistic obtained using HAC ($t(HAC)$) and its p-value ($p(HAC)$).

Table C.3: Results for the 6 Portfolios Formed on Size and Book-to-Market

A: CMV

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p-val	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.002	0.012	0.057	-0.024	0.207	0.062	0.719	0.215	3.335	0.001	0.049	14.647	0.000
3PRF	0.002	0.010	0.160	0.031	0.222	0.062	0.769	0.216	3.563	0.000	0.052	14.904	0.000
PLS	0.004	0.015	0.158	-0.016	0.247	0.062	0.856	0.216	3.954	0.000	0.059	14.417	0.000
PCR	0.005	0.016	0.132	0.024	0.298	0.063	1.032	0.218	4.738	0.000	0.048	21.518	0.000
LASSO	0.002	0.013	0.030	-0.002	0.168	0.062	0.580	0.215	2.704	0.007	0.046	12.726	0.000
RIDGE	0.003	0.015	0.131	-0.024	0.173	0.062	0.600	0.215	2.793	0.006	0.059	10.197	0.000
ENET	0.003	0.012	0.041	-0.097	0.231	0.062	0.799	0.216	3.698	0.000	0.045	17.752	0.000
FRED-MD													
OLS	0.001	0.005	0.050	0.024	0.154	0.062	0.534	0.214	2.491	0.013	0.049	10.802	0.000
3PRF	0.003	0.014	-0.018	-0.066	0.186	0.062	0.646	0.215	3.003	0.003	0.048	13.522	0.000
PLS	0.003	0.016	0.142	0.046	0.210	0.062	0.726	0.216	3.369	0.001	0.052	13.883	0.000
PCR	0.005	0.017	0.127	0.098	0.272	0.063	0.944	0.217	4.346	0.000	0.044	21.651	0.000
LASSO	0.004	0.016	-0.039	-0.057	0.226	0.062	0.783	0.216	3.627	0.000	0.033	23.946	0.000
RIDGE	0.004	0.016	0.037	-0.029	0.242	0.062	0.840	0.216	3.881	0.000	0.047	17.773	0.000
ENET	0.004	0.015	0.054	-0.027	0.261	0.063	0.905	0.217	4.172	0.000	0.040	22.576	0.000
All Instruments													
OLS	0.001	0.004	0.097	-0.017	0.160	0.067	0.554	0.231	2.395	0.017	0.065	8.465	0.000
3PRF	0.003	0.013	0.001	-0.086	0.193	0.062	0.670	0.215	3.113	0.002	0.047	14.340	0.000
PLS	0.004	0.014	0.122	0.087	0.264	0.068	0.916	0.234	3.915	0.000	0.063	14.545	0.000
PCR	0.005	0.019	0.136	0.104	0.238	0.067	0.825	0.233	3.539	0.000	0.054	15.290	0.000
LASSO	0.004	0.014	-0.019	-0.044	0.271	0.063	0.940	0.219	4.296	0.000	0.038	24.768	0.000
RIDGE	0.004	0.015	0.034	-0.016	0.248	0.067	0.858	0.233	3.674	0.000	0.054	15.940	0.000
ENET	0.003	0.013	-0.011	-0.035	0.255	0.063	0.883	0.218	4.052	0.000	0.040	21.857	0.000

B: UMV

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.002	0.010	0.055	-0.110	0.218	0.062	0.757	0.216	3.508	0.001	0.046	16.626	0.000
3PRF	0.003	0.011	0.144	0.000	0.258	0.063	0.893	0.217	4.122	0.000	0.046	19.347	0.000
PLS	0.004	0.017	0.139	0.111	0.246	0.062	0.853	0.216	3.941	0.000	0.054	15.805	0.000
PCR	0.005	0.016	0.167	0.083	0.299	0.063	1.035	0.218	4.749	0.000	0.052	19.905	0.000
LASSO	0.002	0.009	0.041	-0.071	0.187	0.062	0.648	0.215	3.011	0.003	0.049	13.293	0.000
RIDGE	0.002	0.011	0.077	-0.040	0.224	0.062	0.774	0.216	3.587	0.000	0.048	16.120	0.000
ENET	0.002	0.009	0.029	-0.074	0.260	0.063	0.901	0.217	4.157	0.000	0.044	20.464	0.000
FRED-MD													
OLS	0.001	0.004	0.116	0.043	0.158	0.062	0.546	0.215	2.547	0.011	0.057	9.636	0.000
3PRF	0.002	0.012	0.135	-0.021	0.197	0.062	0.681	0.215	3.163	0.002	0.060	11.418	0.000
PLS	0.003	0.016	0.150	0.076	0.198	0.062	0.687	0.215	3.189	0.002	0.058	11.751	0.000
PCR	0.004	0.017	0.118	0.055	0.255	0.063	0.882	0.217	4.070	0.000	0.059	14.887	0.000
LASSO	0.002	0.010	0.098	0.017	0.201	0.062	0.697	0.215	3.239	0.001	0.046	15.132	0.000
RIDGE	0.003	0.014	0.142	-0.021	0.230	0.062	0.797	0.216	3.688	0.000	0.054	14.801	0.000
ENET	0.003	0.012	0.138	0.047	0.261	0.063	0.905	0.217	4.173	0.000	0.044	20.408	0.000
All Instruments													
OLS	0.001	0.004	0.125	-0.007	0.168	0.067	0.583	0.232	2.519	0.012	0.070	8.374	0.000
3PRF	0.003	0.012	0.167	-0.027	0.213	0.062	0.738	0.216	3.421	0.001	0.059	12.574	0.000
PLS	0.004	0.014	0.104	0.032	0.263	0.068	0.911	0.234	3.897	0.000	0.060	15.097	0.000
PCR	0.004	0.017	0.133	0.053	0.237	0.067	0.820	0.233	3.517	0.001	0.069	11.858	0.000
LASSO	0.002	0.009	0.069	-0.026	0.228	0.063	0.790	0.218	3.628	0.000	0.043	18.550	0.000
RIDGE	0.003	0.014	0.147	-0.021	0.229	0.067	0.794	0.233	3.407	0.001	0.060	13.331	0.000
ENET	0.002	0.009	0.038	-0.045	0.234	0.063	0.812	0.217	3.736	0.000	0.043	19.023	0.000

(Continued)

Table C.3: Results for the 6 Portfolios Formed on Size and Book-to-Market *(Continued)*

C: MVATE

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.004	0.022	0.059	-0.025	0.203	0.062	0.703	0.215	3.266	0.001	0.050	14.175	0.000
3PRF	0.004	0.020	0.161	0.034	0.223	0.062	0.773	0.216	3.583	0.000	0.052	14.856	0.000
PLS	0.007	0.029	0.169	-0.011	0.247	0.062	0.857	0.216	3.958	0.000	0.061	14.037	0.000
PCR	0.009	0.031	0.144	0.034	0.300	0.063	1.039	0.218	4.768	0.000	0.048	21.704	0.000
LASSO	0.004	0.025	0.027	-0.010	0.166	0.062	0.575	0.215	2.678	0.008	0.046	12.624	0.000
RIDGE	0.005	0.029	0.132	-0.022	0.174	0.062	0.604	0.215	2.811	0.005	0.060	10.138	0.000
ENET	0.006	0.024	0.033	-0.098	0.236	0.062	0.816	0.216	3.774	0.000	0.044	18.609	0.000
FRED-MD													
OLS	0.001	0.014	-0.175	-0.006	0.037	0.062	0.128	0.213	0.599	0.550	0.035	3.648	0.000
3PRF	0.005	0.027	-0.019	-0.064	0.188	0.062	0.652	0.215	3.031	0.003	0.048	13.619	0.000
PLS	0.006	0.032	0.138	0.037	0.201	0.062	0.695	0.215	3.226	0.001	0.051	13.732	0.000
PCR	0.009	0.033	0.131	0.105	0.275	0.063	0.951	0.217	4.378	0.000	0.043	22.089	0.000
LASSO	0.007	0.030	-0.039	-0.053	0.226	0.062	0.783	0.216	3.625	0.000	0.032	24.708	0.000
RIDGE	0.007	0.030	0.036	-0.023	0.246	0.062	0.850	0.216	3.930	0.000	0.047	18.052	0.000
ENET	0.008	0.029	0.050	-0.018	0.261	0.063	0.905	0.217	4.173	0.000	0.039	23.380	0.000
All Instruments													
OLS	0.000	0.018	-0.130	-0.057	-0.002	0.066	-0.008	0.230	-0.034	0.973	0.042	-0.184	0.854
3PRF	0.005	0.025	-0.001	-0.087	0.191	0.062	0.661	0.215	3.072	0.002	0.048	13.744	0.000
PLS	0.007	0.028	0.132	0.095	0.265	0.068	0.917	0.234	3.921	0.000	0.064	14.399	0.000
PCR	0.009	0.037	0.143	0.106	0.240	0.067	0.831	0.233	3.565	0.000	0.054	15.490	0.000
LASSO	0.007	0.026	-0.019	-0.038	0.274	0.063	0.950	0.219	4.339	0.000	0.037	25.655	0.000
RIDGE	0.007	0.029	0.044	-0.016	0.247	0.067	0.856	0.233	3.669	0.000	0.054	15.731	0.000
ENET	0.007	0.026	-0.026	-0.033	0.266	0.063	0.921	0.219	4.206	0.000	0.039	23.441	0.000

Table C.3 reports some summary statistics of the OOS (Jan-1996 to Dec-2017) for the 6 portfolios formed on size and book-to-market by estimator and three different set of conditioning information (Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios (UMV). Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). The table reports the monthly returns ($\hat{\mu}$), monthly returns ($\hat{\sigma}$), first and second autocorrelation ($\hat{\rho}_1$ and $\hat{\rho}_2$) of the monthly returns. The Sharpe ratios (SR) are presented monthly and annualized, as well as the standard error (se). For the sake of completeness, it is also reported the t-statistic (t), the standard p-value (p), as well as the HAC standard error ($se(HAC)$), the t-statistic obtained using HAC ($t(HAC)$) and its p-value ($p(HAC)$).

Table C.4: Results for the 25 Portfolios Formed on Size and Book-to-Market

A: CMV

Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	Monthly SR	Monthly se	Annual SR	Annual se	t	p-val	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.002	0.007	-0.108	0.124	0.307	0.063	1.062	0.218	4.868	0.000	0.038	27.935	0.000
3PRF	0.002	0.007	0.005	0.044	0.327	0.063	1.132	0.219	5.171	0.000	0.044	25.789	0.000
PLS	0.003	0.010	0.017	0.046	0.307	0.063	1.063	0.218	4.870	0.000	0.046	23.139	0.000
PCR	0.004	0.011	0.092	-0.042	0.383	0.064	1.325	0.221	6.000	0.000	0.054	24.516	0.000
LASSO	0.002	0.006	0.016	0.167	0.276	0.063	0.957	0.217	4.407	0.000	0.047	20.181	0.000
RIDGE	0.003	0.008	-0.130	0.147	0.313	0.063	1.083	0.218	4.961	0.000	0.039	27.911	0.000
ENET	0.002	0.006	-0.123	0.155	0.315	0.063	1.092	0.218	4.998	0.000	0.036	30.736	0.000
FRED-MD													
OLS	0.000	0.002	-0.061	-0.031	0.161	0.062	0.559	0.215	2.605	0.010	0.045	12.313	0.000
3PRF	0.002	0.007	0.096	-0.006	0.282	0.063	0.978	0.217	4.497	0.000	0.052	18.951	0.000
PLS	0.003	0.009	0.026	0.057	0.296	0.063	1.024	0.218	4.701	0.000	0.043	23.716	0.000
PCR	0.004	0.011	0.079	0.048	0.371	0.064	1.286	0.220	5.834	0.000	0.052	24.585	0.000
LASSO	0.002	0.008	-0.071	-0.076	0.215	0.062	0.743	0.216	3.448	0.001	0.044	17.027	0.000
RIDGE	0.003	0.011	0.053	-0.014	0.274	0.063	0.949	0.217	4.370	0.000	0.049	19.392	0.000
ENET	0.002	0.008	-0.025	-0.053	0.280	0.063	0.970	0.217	4.465	0.000	0.048	20.323	0.000
All Instruments													
OLS	0.000	0.002	-0.005	0.009	0.201	0.067	0.696	0.232	2.996	0.003	0.053	13.137	0.000
3PRF	0.002	0.007	0.076	-0.005	0.290	0.063	1.006	0.218	4.621	0.000	0.050	19.946	0.000
PLS	0.003	0.009	0.080	0.050	0.331	0.068	1.145	0.236	4.849	0.000	0.053	21.547	0.000
PCR	0.004	0.013	0.007	0.029	0.307	0.068	1.063	0.235	4.518	0.000	0.055	19.254	0.000
LASSO	0.002	0.008	-0.098	0.018	0.231	0.063	0.799	0.220	3.637	0.000	0.040	20.170	0.000
RIDGE	0.003	0.010	0.006	0.017	0.305	0.068	1.057	0.235	4.493	0.000	0.054	19.399	0.000
ENET	0.002	0.008	-0.118	-0.018	0.302	0.064	1.045	0.222	4.701	0.000	0.040	25.901	0.000

B: UMV

Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	Monthly SR	Monthly se	Annual SR	Annual se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.002	0.006	-0.078	0.107	0.293	0.063	1.017	0.218	4.668	0.000	0.037	27.165	0.000
3PRF	0.002	0.007	0.033	-0.005	0.330	0.063	1.143	0.219	5.219	0.000	0.042	27.225	0.000
PLS	0.003	0.011	0.099	0.074	0.322	0.063	1.114	0.219	5.094	0.000	0.041	27.417	0.000
PCR	0.004	0.011	0.121	0.024	0.386	0.064	1.337	0.221	6.049	0.000	0.050	26.591	0.000
LASSO	0.001	0.005	0.016	0.102	0.279	0.063	0.968	0.217	4.453	0.000	0.046	20.973	0.000
RIDGE	0.002	0.007	-0.075	0.042	0.311	0.063	1.076	0.218	4.928	0.000	0.040	26.583	0.000
ENET	0.002	0.005	-0.066	0.088	0.316	0.063	1.095	0.218	5.014	0.000	0.037	29.458	0.000
FRED-MD													
OLS	0.000	0.002	-0.058	-0.023	0.166	0.062	0.576	0.215	2.683	0.008	0.045	12.696	0.000
3PRF	0.002	0.007	0.113	-0.001	0.262	0.063	0.907	0.217	4.182	0.000	0.059	15.475	0.000
PLS	0.003	0.009	0.089	0.057	0.299	0.063	1.036	0.218	4.756	0.000	0.043	24.224	0.000
PCR	0.004	0.011	0.072	0.042	0.361	0.064	1.252	0.220	5.687	0.000	0.051	24.710	0.000
LASSO	0.002	0.007	-0.062	-0.106	0.209	0.062	0.724	0.216	3.357	0.001	0.048	15.114	0.000
RIDGE	0.003	0.010	0.135	-0.047	0.277	0.063	0.960	0.217	4.417	0.000	0.051	18.862	0.000
ENET	0.002	0.007	-0.017	-0.123	0.267	0.063	0.924	0.217	4.259	0.000	0.052	17.747	0.000
All Instruments													
OLS	0.000	0.002	0.001	0.011	0.204	0.067	0.706	0.232	3.039	0.003	0.054	13.132	0.000
3PRF	0.002	0.007	0.105	-0.007	0.278	0.063	0.964	0.217	4.436	0.000	0.054	17.895	0.000
PLS	0.003	0.009	0.115	-0.014	0.303	0.068	1.051	0.235	4.469	0.000	0.053	19.951	0.000
PCR	0.004	0.012	0.046	0.019	0.293	0.068	1.015	0.235	4.322	0.000	0.062	16.363	0.000
LASSO	0.001	0.006	-0.070	0.014	0.230	0.063	0.797	0.220	3.628	0.000	0.043	18.443	0.000
RIDGE	0.003	0.010	0.101	-0.020	0.306	0.068	1.061	0.235	4.511	0.000	0.057	18.572	0.000
ENET	0.002	0.006	-0.066	-0.057	0.280	0.064	0.970	0.222	4.376	0.000	0.047	20.478	0.000

(Continued)

Table C.4: Results for the 25 Portfolios Formed on Size and Book-to-Market *(Continued)*

C: MVATE

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.003	0.013	-0.127	0.142	0.270	0.063	0.934	0.217	4.302	0.000	0.037	25.109	0.000
3PRF	0.004	0.014	-0.019	0.027	0.307	0.063	1.064	0.218	4.876	0.000	0.043	25.024	0.000
PLS	0.006	0.019	-0.011	0.071	0.300	0.063	1.039	0.218	4.766	0.000	0.045	23.133	0.000
PCR	0.008	0.020	0.098	-0.040	0.380	0.064	1.318	0.221	5.967	0.000	0.052	25.274	0.000
LASSO	0.003	0.012	0.020	0.177	0.278	0.063	0.962	0.217	4.427	0.000	0.045	21.145	0.000
RIDGE	0.005	0.015	-0.121	0.169	0.299	0.063	1.035	0.218	4.750	0.000	0.038	27.334	0.000
ENET	0.003	0.013	-0.128	0.199	0.259	0.063	0.898	0.217	4.143	0.000	0.049	18.277	0.000
FRED-MD													
OLS	0.000	0.039	-0.010	0.045	0.006	0.062	0.020	0.213	0.096	0.924	0.043	0.472	0.637
3PRF	0.004	0.015	0.142	0.018	0.261	0.063	0.904	0.217	4.171	0.000	0.050	17.999	0.000
PLS	0.005	0.019	0.041	0.012	0.285	0.063	0.989	0.218	4.546	0.000	0.045	21.889	0.000
PCR	0.008	0.022	0.082	0.058	0.374	0.064	1.295	0.221	5.872	0.000	0.051	25.189	0.000
LASSO	0.003	0.016	-0.074	-0.074	0.212	0.062	0.735	0.216	3.410	0.001	0.043	16.970	0.000
RIDGE	0.006	0.020	0.069	-0.001	0.275	0.063	0.951	0.217	4.380	0.000	0.049	19.403	0.000
ENET	0.004	0.016	-0.011	-0.089	0.286	0.063	0.990	0.218	4.553	0.000	0.045	21.977	0.000
All Instruments													
OLS	-0.001	0.025	-0.007	0.035	-0.054	0.066	-0.186	0.230	-0.810	0.419	0.046	-4.040	0.000
3PRF	0.004	0.014	0.105	0.013	0.254	0.063	0.880	0.217	4.063	0.000	0.050	17.726	0.000
PLS	0.005	0.017	0.068	0.063	0.314	0.068	1.089	0.236	4.621	0.000	0.053	20.460	0.000
PCR	0.008	0.024	0.009	0.031	0.310	0.068	1.074	0.235	4.562	0.000	0.055	19.574	0.000
LASSO	0.004	0.015	-0.100	0.019	0.232	0.063	0.802	0.220	3.649	0.000	0.039	20.450	0.000
RIDGE	0.006	0.020	0.040	0.024	0.299	0.068	1.035	0.235	4.405	0.000	0.054	18.999	0.000
ENET	0.005	0.015	-0.097	-0.106	0.302	0.064	1.045	0.223	4.681	0.000	0.040	26.200	0.000

Table C.4 reports some summary statistics of the OOS (Jan-1996 to Dec-2017) for the 25 portfolios formed on size and book-to-market by estimator and three different set of conditioning information (Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios (UMV). Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). The table reports the monthly returns ($\hat{\mu}$), monthly returns ($\hat{\sigma}$), first and second autocorrelation ($\hat{\rho}_1$ and $\hat{\rho}_2$) of the monthly returns. The Sharpe ratios (SR) are presented monthly and annualized, as well as the standard error (se). For the sake of completeness, it is also reported the t-statistic (t), the standard p-value (p), as well as the HAC standard error ($se(HAC)$), the t-statistic obtained using HAC ($t(HAC)$) and its p-value ($p(HAC)$).

Table C.5: Results for the 100 Portfolios Formed on Size and Book-to-Market

A: CMV

Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	Monthly SR	Monthly se	Annual SR	Annual se	t	p-val	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.001	0.009	-0.175	0.106	0.076	0.062	0.262	0.214	1.227	0.221	0.072	3.635	0.000
3PRF	0.001	0.012	-0.208	0.065	0.117	0.062	0.404	0.214	1.890	0.060	0.095	4.269	0.000
PLS	0.002	0.010	-0.002	0.009	0.245	0.062	0.848	0.216	3.920	0.000	0.151	5.624	0.000
PCR	0.003	0.009	0.104	0.090	0.303	0.063	1.048	0.218	4.808	0.000	0.127	8.270	0.000
LASSO	0.001	0.009	-0.051	-0.002	0.087	0.062	0.302	0.214	1.415	0.158	0.096	3.140	0.002
RIDGE	0.001	0.013	-0.027	0.010	0.060	0.062	0.207	0.213	0.970	0.333	0.082	2.512	0.013
ENET	0.001	0.009	0.047	0.022	0.097	0.062	0.336	0.214	1.573	0.117	0.101	3.314	0.001
FRED-MD													
OLS	0.000	0.001	0.185	0.005	0.128	0.062	0.443	0.214	2.068	0.040	0.065	6.830	0.000
3PRF	0.002	0.007	-0.255	0.134	0.210	0.062	0.728	0.216	3.376	0.001	0.042	17.141	0.000
PLS	0.002	0.007	0.081	0.283	0.350	0.063	1.211	0.220	5.513	0.000	0.060	20.049	0.000
PCR	0.002	0.011	0.072	0.081	0.203	0.062	0.703	0.215	3.266	0.001	0.144	4.898	0.000
LASSO	0.001	0.003	0.049	0.115	0.312	0.063	1.079	0.218	4.943	0.000	0.041	26.408	0.000
RIDGE	0.001	0.005	-0.027	0.082	0.274	0.063	0.950	0.217	4.375	0.000	0.062	15.315	0.000
ENET	0.001	0.004	-0.027	0.139	0.307	0.063	1.064	0.218	4.878	0.000	0.037	28.613	0.000
All Instruments													
OLS	0.000	0.001	0.211	-0.001	0.118	0.067	0.409	0.231	1.773	0.078	0.082	4.972	0.000
3PRF	0.002	0.007	-0.188	0.220	0.240	0.062	0.832	0.216	3.846	0.000	0.049	16.828	0.000
PLS	0.002	0.006	0.106	0.049	0.309	0.068	1.072	0.235	4.554	0.000	0.074	14.532	0.000
PCR	0.003	0.010	0.119	0.031	0.269	0.068	0.932	0.234	3.983	0.000	0.154	6.069	0.000
LASSO	0.001	0.004	0.179	0.223	0.323	0.068	1.120	0.236	4.747	0.000	0.042	26.651	0.000
RIDGE	0.001	0.005	-0.042	0.147	0.243	0.067	0.842	0.233	3.608	0.000	0.038	21.963	0.000
ENET	0.001	0.004	0.212	0.261	0.359	0.068	1.243	0.237	5.241	0.000	0.040	31.397	0.000

B: UMV

Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	Monthly SR	Monthly se	Annual SR	Annual se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.001	0.009	-0.175	0.102	0.072	0.062	0.248	0.213	1.164	0.245	0.072	3.471	0.001
3PRF	0.001	0.010	-0.165	0.086	0.133	0.062	0.462	0.214	2.157	0.032	0.104	4.445	0.000
PLS	0.002	0.010	0.002	0.017	0.241	0.062	0.836	0.216	3.867	0.000	0.154	5.439	0.000
PCR	0.003	0.009	0.132	0.109	0.315	0.063	1.090	0.218	4.992	0.000	0.110	9.935	0.000
LASSO	0.001	0.008	-0.026	0.014	0.100	0.062	0.346	0.214	1.617	0.107	0.104	3.327	0.001
RIDGE	0.001	0.011	-0.010	0.036	0.083	0.062	0.289	0.214	1.351	0.178	0.095	3.031	0.003
ENET	0.001	0.008	0.058	0.040	0.103	0.062	0.357	0.214	1.669	0.096	0.104	3.444	0.001
FRED-MD													
OLS	0.000	0.001	0.186	0.006	0.128	0.062	0.442	0.214	2.064	0.040	0.065	6.802	0.000
3PRF	0.001	0.008	-0.266	0.128	0.197	0.062	0.684	0.215	3.177	0.002	0.039	17.320	0.000
PLS	0.002	0.007	0.065	0.297	0.346	0.063	1.199	0.220	5.463	0.000	0.056	21.277	0.000
PCR	0.002	0.010	0.070	0.082	0.209	0.062	0.723	0.216	3.356	0.001	0.140	5.185	0.000
LASSO	0.001	0.003	0.055	0.123	0.305	0.063	1.057	0.218	4.846	0.000	0.040	26.500	0.000
RIDGE	0.001	0.005	-0.033	0.087	0.265	0.063	0.918	0.217	4.232	0.000	0.059	15.576	0.000
ENET	0.001	0.004	-0.037	0.141	0.301	0.063	1.044	0.218	4.789	0.000	0.036	28.759	0.000
All Instruments													
OLS	0.000	0.001	0.211	-0.001	0.118	0.067	0.409	0.231	1.771	0.078	0.082	4.959	0.000
3PRF	0.002	0.007	-0.194	0.210	0.228	0.062	0.791	0.216	3.663	0.000	0.047	16.654	0.000
PLS	0.002	0.005	0.134	0.062	0.315	0.068	1.092	0.236	4.634	0.000	0.066	16.579	0.000
PCR	0.002	0.009	0.131	0.007	0.272	0.068	0.944	0.234	4.031	0.000	0.136	6.925	0.000
LASSO	0.001	0.004	0.184	0.234	0.315	0.068	1.090	0.236	4.626	0.000	0.042	25.906	0.000
RIDGE	0.001	0.005	-0.021	0.165	0.247	0.067	0.857	0.233	3.670	0.000	0.037	23.256	0.000
ENET	0.001	0.004	0.222	0.280	0.352	0.068	1.218	0.237	5.142	0.000	0.040	30.241	0.000

(Continued)

Table C.5: Results for the 100 Portfolios Formed on Size and Book-to-Market *(Continued)*

C: MVATE

					Monthly		Annual						
Estimator	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}_1$	$\hat{\rho}_2$	SR	se	SR	se	t	p	$se(HAC)$	$t(HAC)$	$p(HAC)$
Goyal													
OLS	0.000	0.024	-0.094	0.078	0.003	0.062	0.010	0.213	0.047	0.962	0.042	0.239	0.811
3PRF	0.002	0.026	-0.175	0.064	0.059	0.062	0.204	0.213	0.955	0.341	0.060	3.374	0.001
PLS	0.005	0.017	0.042	0.035	0.279	0.063	0.965	0.217	4.440	0.000	0.101	9.541	0.000
PCR	0.005	0.017	0.103	0.108	0.312	0.063	1.082	0.218	4.957	0.000	0.098	11.065	0.000
LASSO	0.001	0.017	-0.016	0.048	0.084	0.062	0.292	0.214	1.365	0.173	0.089	3.263	0.001
RIDGE	0.001	0.026	-0.010	0.060	0.044	0.062	0.154	0.213	0.720	0.472	0.070	2.197	0.029
ENET	0.001	0.021	0.014	0.026	0.043	0.062	0.150	0.213	0.703	0.483	0.067	2.231	0.027
FRED-MD													
OLS	0.003	0.069	-0.136	0.022	0.047	0.062	0.163	0.213	0.763	0.446	0.025	6.613	0.000
3PRF	0.003	0.023	-0.220	0.009	0.135	0.062	0.469	0.214	2.190	0.029	0.015	31.078	0.000
PLS	0.004	0.019	0.094	0.210	0.230	0.062	0.796	0.216	3.683	0.000	0.101	7.858	0.000
PCR	0.004	0.022	0.082	0.084	0.198	0.062	0.687	0.215	3.189	0.002	0.139	4.931	0.000
LASSO	0.002	0.008	0.074	0.113	0.240	0.062	0.833	0.216	3.851	0.000	0.037	22.748	0.000
RIDGE	0.003	0.015	-0.034	0.082	0.188	0.062	0.652	0.215	3.030	0.003	0.029	22.333	0.000
ENET	0.001	0.012	0.192	0.065	0.129	0.062	0.448	0.214	2.092	0.037	0.069	6.465	0.000
All Instruments													
OLS	-0.002	0.085	-0.103	-0.020	-0.026	0.066	-0.091	0.230	-0.394	0.694	0.040	-2.284	0.023
3PRF	0.003	0.020	-0.187	0.084	0.145	0.062	0.502	0.214	2.344	0.020	0.026	19.497	0.000
PLS	0.003	0.013	0.110	-0.064	0.263	0.068	0.912	0.234	3.898	0.000	0.046	19.874	0.000
PCR	0.005	0.020	0.125	0.031	0.271	0.068	0.937	0.234	4.003	0.000	0.140	6.698	0.000
LASSO	0.003	0.010	0.117	0.200	0.253	0.067	0.875	0.234	3.748	0.000	0.036	24.412	0.000
RIDGE	0.002	0.015	-0.049	0.083	0.117	0.067	0.407	0.231	1.762	0.079	0.036	11.442	0.000
ENET	0.002	0.012	-0.021	-0.093	0.179	0.067	0.620	0.232	2.675	0.008	0.060	10.313	0.000

Table C.5 reports some summary statistics of the OOS (Jan-1996 to Dec-2017) for the 100 portfolios formed on size and book-to-market by estimator and three different set of conditioning information (Goyal, FRED-MD and "All Instruments"). Each panel shows the test of pairs of Sharpe ratios for three different framework to build efficient portfolios. Panel A reports conditionally mean-variance (CMV) efficient portfolios. Panel B reports unconditionally mean-variance efficient portfolios (UMV). Panel C presents the mean-variance tracking error (MVATE) portfolios. We split the results depending on conditioning information set used. Goyal variables comprises: *b/m, dfr, dfy, infl, ltr, lty, ntis, svar, tms* and *tbl* . The FRED-MD is a large dataset containing 128 macroeconomic and financial variables. Finally, "All Instruments" is the combination of Goyal and FRED-MD datasets with the Economic Policy Uncertainty (EPU) index and the Financial Stress Indicator (FSI). The table reports the monthly returns ($\hat{\mu}$), monthly returns ($\hat{\sigma}$), first and second autocorrelation ($\hat{\rho}_1$ and $\hat{\rho}_2$) of the monthly returns. The Sharpe ratios (SR) are presented monthly and annualized, as well as the standard error (se). For the sake of completeness, it is also reported the t-statistic (t), the standard p-value (p), as well as the HAC standard error ($se(HAC)$), the t-statistic obtained using HAC ($t(HAC)$) and its p-value ($p(HAC)$).