Forecasting Volatility Using Double Shrinkage Methods*

Mingmian Cheng[1], Norman R. Swanson[2] and Xiye Yang[2]
[1]Sun Yat-sen University and [2]Rutgers University

February 2020

Abstract

In this paper, we propose and evaluate a shrinkage based methodology that is designed to improve the accuracy of volatility forecasts. Our approach is based on a two-step procedure for extracting latent common volatility factors from a large dimensional and high-frequency dataset. In the first step, we apply either least absolute shrinkage operator (LASSO) or the elastic net (EN) shrinkage on estimated integrated volatilities, in order to select a subset of assets that are informative about the target asset. In the second step, we utilize (sparse) principal component analysis on the selected assets, in order to estimate latent return factors, which are in turn used to construct latent volatility factors. Our two-step method is found to yield more accurate volatility predictions than a variety of alternative models based on approaches such as direct application of (S)PCA and direct application of LASSO or EN shrinkage, when comparing *out-of-sample R^2s* and mean absolute forecasting errors, and when implementing predictive accuracy tests. Additionally model confidence sets are found to contain models solely based on our two-step approach. These forecasting gains are found to be robust to the use of original or log-scale realized volatility models, different data sampling frequencies, and different forecasting sub-periods.

Keywords: Forecasting, Latent Volatility Factor, Shrinkage, High-Frequency and High Dimensional Data.

JEL Classification: C22, C52, C53, C58.

1   Introduction

Accurate volatility estimation and prediction is crucial to successful risk management and asset allocation. In light of this fact, it is not surprising that the seminal contribution of Jacod (2018), originally published as a 1994 working paper, has spurred the development of a veritable arsenal of realized integrated volatility (IV) estimators. A very few of these include realized variance (RV) (Andersen et al. (2001)), jump robust RV based on multi-power variation and truncation (Barndorff-Nielsen and Shephard (2004), Mancini (2009), Corsi et al. (2010), Podolskij and Ziggel (2010)), and microstructure noise robust RV based on multi-scale variation and pre-averaging (Jacod et al. (2009) and Aït-Sahalia et al. (2011)). One important use of these sorts of IV estimators is in heterogeneous autoregressive (HAR) type forecasting models, as introduced in Corsi (2009), and built on by Andersen et al. (2007), Corsi et al. (2010), Duong and Swanson (2015), and Patton and Sheppard (2015), who augment the basic HAR model by the inclusion of a variety of jump variation related variables. Although very parsimonious, the HAR-type models analyzed in the above papers only assess whether information derived from the target asset is useful for IV prediction. In light of this, a number of authors examine whether there are useful sources of information other than the target asset itself, when predicting volatility in an HAR framework. For example, Audrino and Hu (2016) explore the importance of leverage and downside risk in an HAR forecasting framework, and Bollerslev et al. (2016) further improve volatility forecasting by allowing HAR type model coefficients to evolve according to the degree of measurement error. Other key papers from this nascent literature include the papers by Fernandes et al. (2014), who examine the relationship between the VIX, S&P500 volume, and S&P500 returns, and Audrino et al. (2019) who extend the HAR model by including sentiment, attention, and other economic variables.[1]

In this paper, we add to the volatility prediction literature by proposing and analyzing a two-step "double" shrinkage based methodology for extracting useful predictive information from a large dimensional and high-frequency asset returns dataset. The methodology that we develop hinges on the construction of latent IV factors, which are used to augment standard HAR prediction models. In the first step

---

[1]Evidently, there is a very deep and rich literature that uses HAR models for forecasting, and many of the important papers in the field are cited above. Other volatility forecasting models that are widely used in the financial econometrics literature include stochastic volatility (SV) models (e.g., see Meddahi (2001), Andersen et al. (2004), Andersen et al. (2011)), (G)ARCH-type models (e.g., see Andersen et al. (2003), Hansen and Lunde (2005), Brandt and Jones (2006)), and Mixed Data Sampling (MIDAS) models (e.g., see Ghysels et al. (2006), Ghysels and Sinko (2011)).

of our procedure, we apply either least absolute shrinkage operator (LASSO) or elastic net (EN) shrinkage on estimated integrated volatilities, in order to select a subset of assets that are informative about the target asset. In the second step, we apply either principal component analysis (PCA) or sparse principal component analysis (SPCA) on the selected assets (on a daily basis), in order to estimate latent return factors with flexible loadings that can vary from day to day. Then we use the estimated volatilities of these return factors to estimate latent IV factors. It is important to note that in the case where the second step involves using SPCA, we are essentially performing double shrinkage. In the first step, we shrink the set of IV estimators in order to "pare down" the original dataset, by removing those assets that are not relevant on the target asset's volatility. In the second step, we shrink the set of asset returns selected in the first step. One can immediately see that our second step indeed involves shrinkage by noting, as discussed in Zou et al. (2006), that SPCA can be interpreted as a variant of PCA, where PCA is first written in penalized regression form, and the coefficients from the penalized regression are treated using LASSO or EN shrinkage.[2] It is in this sense that our procedure involves "double shrinkage". Additionally, it is worth pointing out that the first step of our approach builds on methods developed in Bai and Ng (2008), in which "targeted predictors" are selected before the estimation of common factors; while the second step builds on the recent generalization of PCA to high-frequency data by Aït-Sahalia and Xiu (2017, 2019).[3]

One might argue that it is not necessary to use double shrinkage, as a simpler approach will likely yield similar forecasting performance. However, our results indicate that our double shrinkage method yields superior forecasting performance than many relatively simple variants of our approach, including the following alternatives: (i) the HAR model, which utilizes only "own-asset" information; (ii) an HAR model that includes latent factors estimated using PCA or SPCA on IV estimators for all stocks; (iii) an HAR model that includes IV predictors selected by conducting LASSO or EN shrinkage on IV estimates for all stocks; (iv) an HAR model, in which PCA or SPCA

---

is carried out using high-frequency returns of all stocks, and IV estimators that are constructed using the resultant principle components are included as predictors; (v) an HAR model that is specified using the two-step procedure discussed above, but fixes factor loadings; and (vi) an HAR model in which partial least squares (PLS) is used to select asset specific IV estimators for use in prediction.[4] Note that (iii) and (iv) involve specifying the HAR model using only the first and second steps of our two-step procedure, respectively. Namely, when comparing point *out-of-sample R^2* and mean absolute forecasting error (MAFE) criteria, our proposed double shrinkage procedure yields more accurate predictions than all of the above alternatives. In addition, application of the model confidence set (MCS) tests of Hansen et al. (2011) and the conditional predictive accuracy (GW) tests of Giacomini and White (2006) indicate that forecasts based on "double shrinkage" are significantly superior. For instance, confidence sets exclusively contain variants of our double shrinkage type HAR models.

Intuitively, the above results follow because our high-frequency and high dimensional dataset contains both relevant and irrelevant information on the target asset's volatility. Summarizing, our main task is to preserve relevant information while discarding irrelevant information. Now, there is no guarantee that use of PCA or SPCA will preserve all the relevant information for a particular target volatility. This is why solely using these dimension reduction methods on every asset yields inferior predictions, relative to those obtained using our double shrinkage approach. On the other hand, while LASSO and EN shrinkage can help select relevant assets, there is no guarantee that they can remove irrelevant information contained within particular selected assets (it would be a very strong assumption to claim that selected assets only contain relevant information). This is the role that subsequent application of PCA or SPCA plays.

Our empirical analysis points to a number of interesting findings, in addition to those mentioned above. First, we show that for virtually all forecasting scenarios, the forecasting gains associated with our proposed models range from approximately 6% to 9% (in terms of R^2 and MAFE). These findings are robust to different data frequencies (2.5-, 5-, and 10-minute), different functional forms (original or log-scale), and various different forecasting periods. Second, we find that implementing our procedure using SPCA in the second step sometimes yields prediction models that slightly dominate those associated with the use of PCA. For instance, when comparing predictions of SPY IV using factors constructed with 5-minute frequency data, factor-augmented models

---

[4]See Wold (1966).

with SPCA have 5%–9% larger *out-of-sample* R^2 values than those associated with the use of PCA. Finally, we find that *in-sample* fit improves when estimated IV factors are used as predictors, no matter how they are extracted.

The rest of the paper is organized as follows. Section 2 outlines the setup and modeling assumptions, and includes a brief discussion of some commonly used realized measures of integrated volatility. Section 3 discusses the forecasting framework, and briefly introduces LASSO, EN, PCA, and SPCA techniques. Section 4 introduces our experimental setup, includes a description of all forecasting models, and details the statistics used to analyze our experimental results. Section 5 briefly discusses the data used in the sequel, and Section 6 summarizes our key empirical findings. Finally, concluding remarks are gathered in Section 7.

2   Setup

Denote by X a d-dimensional log-price process. Following the high-frequency econometrics literature, assume that X follows an Itô-semimartingale defined on the filtered probability space $\left(\Omega, \mathbb{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}\right)$, and has the following representation:

$$
\begin{aligned}
X_t = X_0 &+ \int_0^t b_s ds + \int_0^t \sigma_s dW_s \\
&+ \int_0^t \int_{\{|x| \leq \epsilon\}} x(\mu - \nu)(ds, dx) + \int_0^t \int_{\{|x| \geq \epsilon\}} x\mu(ds, dx),
\end{aligned}
$$

where the instantaneous drift b_t and the spot volatility σ_t are both adapted, càdlàg (right continuous with left limit), and locally bounded. Additionally, W_t is a multidimensional standard Brownian motion, μ is a random jump measure with compensator ν, and $\epsilon > 0$ is an arbitrary threshold. For more details on Itô-semimartingales and continuous-time asset price modeling, see Aït-Sahalia and Jacod (2014), and the references cited therein.

Various realized measures have been developed to estimate latent volatility on a fixed interval $[0, T]$, using high-frequency intraday data. For instance, realized volatility, one of the most widely known measures, is given by:

$$
\text{RV}_t := \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i^n X^j \right)^2, \quad \forall t \in [0, T], \quad j = 1, ..., d, \tag{1}
$$

where $\lfloor \cdot \rfloor$ is the floor function and $\Delta_i^n X^j = X_{i\Delta_n}^j - X_{(i-1)\Delta_n}^j$ is the ith intraday return for j^{th} asset in X, with Δ_n defined as an equally-spaced sampling interval that shrinks

to zero. It is well-known that when asset prices are continuous on a fixed interval $[0, T]$, we have that:

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i^n X^j\right)^2 \overset{\mathbb{P}}{\longrightarrow} \int_0^t \left(\sigma_s^j\right)^2 ds, \quad \forall t \in [0, T], \quad j = 1, ..., d,$$

as $\Delta_n \to 0$, where σ_s^j is the spot volatility for j^{th} asset.

However, when asset prices are discontinuous on $[0, T]$, we instead have:

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i^n X^j\right)^2 \overset{\mathbb{P}}{\longrightarrow} \int_0^t \left(\sigma_s^j\right)^2 ds + \sum_{0 \leq s \leq t} \left(\Delta X_s^j\right)^2, \quad \forall t \in [0, T], \quad j = 1, ..., d,$$

where $\Delta X_s^j := X_s^j - X_{s-}^j \neq 0$, if and only if the jth asset, X^j, jumps at time s.

To separate integrated volatility from jump variation, one can use the threshold technique developed by Mancini (2001, 2009), and construct truncated realized volatility (TRV):

$$\text{TRV}_t := \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\Delta_i^n X^j\right)^2 \mathbf{1}_{\{|\Delta_i^n X^j| \leq c\Delta_n^\varpi\}} \overset{\mathbb{P}}{\longrightarrow} \int_0^t \left(\sigma_s^j\right)^2 ds, \tag{2}$$

for some $\varpi \in (0, 1/2)$, or use the multipower variation (MPV) estimator developed by Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen et al. (2006):

$$\text{MPV}_t := \Delta_n^{1-p^+/2} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor - k + 1} |\Delta_i^n X^j|^{p_1} \cdots |\Delta_{i+k-1}^n X^j|^{p_k} \overset{\mathbb{P}}{\longrightarrow} m_{p_1} \cdots m_{p_k} \int_0^t |\sigma_s^j|^{p^+} ds, \tag{3}$$

where $p_1, p_2, \ldots, p_k \geq 0$, $p^+ = p_1 + \cdots + p_k$ and $m_p = \mathbb{E}[|\mathcal{N}(0,1)|^p]$. One can also combine these two methods and use a truncated multipower variation estimator (see Corsi et al. (2010)). As a result, different components of the quadratic variation can be separately analyzed.

To facilitate the analysis of large dimensional datasets, we further assume that the continuous part of asset log-prices follows a continuous-time factor model on $[0, T]$. Namely, define $Y_t := X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s$ as the continuous part of X, and assume the following factor structure for Y_t:

$$Y_t = \Lambda_t F_t + Z_t, \tag{4}$$

where F_t is an r-dimensional $(r < d)$ unobservable common factor, Z_t is an idiosyncratic component, and Λ_t is a $d \times r$ factor loading matrix, each element of which is adapted

and càdlàg. Here, we specifically call F_t the common price factor, in order to distinguish it from the common volatility factor defined later. The common price factor, F_t, and the idiosyncratic component, Z_t, are assumed to follow continuous Itô-semimartingales, and are given by:

$$F_t = F_0 + \int_0^t h_s ds + \int_0^t \eta_s dB_s \tag{5}$$

and

$$Z_t = Z_0 + \int_0^t g_s ds + \int_0^t \gamma_s d\widetilde{B}_s,$$

respectively, where B_s and $\widetilde{B}_s$ are independent Brownian motions. All of the coefficient processes, h, η, g and γ are adapted to $(\mathcal{F}_t)_{t \geq 0}$ and have càdlàg paths, almost surely. The above factor model and general settings are also used by Aït-Sahalia and Xiu (2017).

3 Forecasting Models and Dimension Reduction Techniques

3.1 Forecasting Models

The proposed forecasting models and most of the simpler alternative models considered in this paper follow the structure of factor-augmented regressions that are widely exploited in the macroeconomic forecasting literature (e.g., see Stock and Watson (2002a,b, 2006), Bai and Ng (2006a, 2008), and the references cited therein). The model is:

$$y_{t+h} = \alpha' W_t + \beta' \Psi_t + \varepsilon_{t+h}, \tag{6}$$

where y_t is the daily integrated volatility of a target asset, h is the forecasting horizon, W_t is a set of observable variables, such as lags of y_t, and Ψ_t contains unobservable variables, such as the latent factors analyzed in this paper. In the context of volatility forecasting, we define the latent predictors, Ψ_t, based on the factor structure assumed in (4) and (5). Namely,

$$\Psi_t := \int_0^t \text{diag}(\Lambda_s \eta_s \eta_s' \Lambda_s') ds,$$

which is defined to be our so-called "common volatility factor" (also called an IV factor in the sequel). This nomenclature follows naturally, since η_s in the above integrand is the spot volatility of F_t (see (5)). As a result, Ψ_t is essentially defined as the integrated volatility of the common price factor F_t. Note that one cannot disentangle Λ from η unless certain identification conditions, such as $\eta \eta' = I_r$, are imposed. However, in the

context of forecasting, we don't have to disentangle these components from Ψ_t. This is because we are only interested in Ψ_t, which is the IV matrix of the r uncorrelated common factors, in our setup. In summary, it is worth stressing that, unlike many other applications of factor-augmented regression, we do not directly use weighted common factors, $\Lambda_t F_t$, that extracted from a large panel of observable data. Instead, we use estimated IVs of these common factors as predictors in our models (i.e. we use the Ψ_t).

Of note is that model (6) actually nests the large family of heterogeneous autoregressive (HAR) type models that are widely used in the literature. The original HAR model of Corsi (2009) is given by:

$$\text{RM}_{t+h} = \alpha_0 + \alpha_1 \text{RM}_t + \alpha_2 \text{RM}_{[t,t-4]} + \alpha_3 \text{RM}_{[t,t-21]} + \epsilon_{t+h}, \qquad (7)$$

where RM represents a realized measure of a target asset's integrated volatility, and $\text{RM}_{[t,t-p]}$ is the average of RM's, over the most recent $p+1$ days (i.e. $\text{RM}_{[t,t-p]} = \frac{1}{p+1} \sum_{i=0}^{p} \text{RM}_{t-i}$). As a result, the second and the third variables on the right hand side of (7) represent weekly and monthly RM averages, respectively. In practice, RV, TRV and MPV defined in (1), (2) and (3) are commonly used as the realized measures in the above HAR model. Let $W_t = [1 \ \text{RM}_t \ \text{RM}_{[t,t-4]} \ \text{RM}_{[t,t-21]}]'$ or $W_t = [1 \ \text{RM}_t \ \text{RM}_{[t,t-4]} \ \text{RM}_{[t,t-21]} \ Z_t]'$, where Z_t is a vector of predictors that may be added to the HAR model. It is evident that this family of HAR-type models is nested in model (6).

Heuristically, equation (6) combines two distinct sources of information for volatility prediction. The first source, W_t, follows an HAR structure, exploiting time series information on the target asset itself. The second source, Ψ_t, contains information extracted from a broader sources of information (i.e. from a large panel of variables other than the target asset).

A key contribution of this paper is that we propose a two-step shrinkage procedure for the extraction of predictors derived from the second source of information. As discussed above, this is done by applying LASSO or EN shrinkage to IV estimates of all assets in order to obtain a subset of assets whose IVs are relevant for predicting the target asset's IV. We then apply PCA or SPCA to this selected panel of high-frequency asset returns in order to estimate the common price factors, F_t, from which common IV factors, Ψ_t, are estimated. Finally, these IV factors are incorporated into equation (6), and IV forecasts are constructed. In the following two sub-sections, we briefly outline the techniques used in this two-step procedure (i.e. the LASSO, elastic net, PCA, and

SPCA).

3.2 LASSO and Elastic Net Shrinkage

In order to select stocks in the first step of our two-step procedure, we use either the LASSO (see Tibshirani (1996)) or the EN (see Zou and Hastie (2005)). Both techniques can be interpreted as regularized or penalized regression methods. Consider a regression of y_{t+h} on W_t and χ_t, where y_{t+h} and W_t are defined in (6), and χ_t is a vector of integrated volatilities on day t, for all assets in X_t. The LASSO estimator is the solution to the following problem:

$$\min_{\phi} \ \sum_t \left\{ \left\| y_{t+h} - \alpha' W_t - \sum_j \phi_j \chi_{j,t} \right\|^2 + \lambda \sum_j |\phi_j| \right\},$$

where the ϕ's are regression coefficients. Similarly, the EN estimator is a solution to the following problem:

$$\min_{\phi} \ \sum_t \left\{ \left\| y_{t+h} - \alpha' W_t - \sum_j \phi_j \chi_{j,t} \right\|^2 + \lambda \sum_j \left(\frac{(1-\theta)}{2} \phi_j^2 + \theta |\phi_j| \right) \right\}, \qquad (8)$$

where $\theta \in [0,1]$. Of note is that when $\theta = 1$, the EN is equivalent to the LASSO. Also, as θ shrinks toward 0, EN estimators approach those obtained via ridge regression. Furthermore, note that LASSO imposes an $\mathcal{L}_1$-norm penalty on coefficients in the model, while EN imposes a linear combination of $\mathcal{L}_1$-norm and $\mathcal{L}_2$-norm penalties on coefficients in the model. Finally, recall that it is the imposition of the $\mathcal{L}_1$-norm penalty that induces shrinkage to zero of some coefficients in the regression model; and it is variables with non-zero coefficients in the solution to these minimization problems that constitute the subset of variables selected in the first step of our procedure.

Tuning parameters for the implementation of the LASSO and EN are determined by hv-block cross validation (see Racine (2000)). As an example of the implementation of this technique, note that when EN is applied, we must estimate values for θ and λ in (8). That is, for a given dataset, one needs to pin down a pair of values among the matrix of all possible combinations of (θ, λ), such that the objective function in (8) generates minimum loss. In our experiments, we set possible θ's as 10 equally-spaced values between 0.1 and 1, while λ is chosen among 100 equally-spaced values between the smallest value required to retain all variables in our chosen subset, and the largest value required to drop all but one variable. Racine (2000)'s method is then used for

consistent cross-validation.

After the first step of our procedure, only assets with nonzero ϕ's are retained in our final set of selected target predictor assets, say $\widetilde{X}_t$. For two different target assets or for two different datasets with the same target asset, the selected pool of assets (i.e. $\widetilde{X}_t$) from which we construct the $\widehat{F}_t$'s and subsequently the $\widehat{\Psi}_t$'s can be substantially different. Intuitively, since the assets in $\widetilde{X}_t$ all have relatively large regression coefficients, they are potentially more informative about the target asset than other assets in X. Hence, the common volatility factors extracted using data from this selected pool of assets may be contaminated with less irrelevant information than would have been the case were the entire X dataset used in our analysis.

In closing this section, it is worth reiterating that shrinkage characterizes both steps of our procedure. As discussed in the introduction, in the first step, we shrink the set of all IV estimators, and select a subset of asset returns. In the second step, which is discussed next, PCA or SPCA is employed to estimate common factors from the subset of returns; and IV estimators are constructed using these factors. Of note is that when SPCA is utilized in the second step, we have a second layer of shrinkage. Namely, SPCA further shrinks the subset of assets extracted in the first step of our procedure, yielding factors that are linear combinations of fewer assets than those selected in the first step.

3.3 (Sparse) Principal Component Analysis

In order to carry out the second step of our procedure, we utilize PCA or SPCA, both of which are briefly discussed in this section. To start, consider the following covariance matrix estimator, defined on a fixed interval, $[0, T]$:

$$\widehat{\Sigma}_t = \frac{1}{t} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left\{ (\Delta_i^n X)(\Delta_i^n X)' \right\} \mathbf{1}_{\{\|\Delta_i^n X\| \leq c\Delta_n^\varpi\}}, \quad \forall t \in [0, T].$$

One carries out PCA by applying an eigenvalue-eigenvector decomposition to $\widehat{\Sigma}_t$, yielding r estimated eigenvalues, in descending order, say $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_r$, and corresponding estimated eigenvectors, $\widehat{\xi}_1, \widehat{\xi}_2, \cdots, \widehat{\xi}_r$. The first r principal components on the fixed interval are estimated as follows:

$$\widehat{\Delta_i^n F}_j = \widehat{\xi}_j' (\Delta_i^n X) \mathbf{1}_{\{\|\Delta_i^n X\| \leq c\Delta_n^\varpi\}}, \quad j = 1, \cdots, r.$$

With these estimated principal components, latent common volatility factors on day t can subsequently be estimated as follows:

$$\widehat{\Psi}_{j,t} = \frac{1}{t} \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \left(\widehat{\Delta_i^n F}_j \right)^2, \quad j = 1, \cdots, r.$$

Thus, for any $j = 1, \cdots, r$, we have $\widehat{\Psi}_{j,t} = \widehat{\xi}_j' \widehat{\Sigma}_t \widehat{\xi}_j = \widehat{\lambda}_j \widehat{\xi}_j' \widehat{\xi}_j$, which is equivalent to $\widehat{\lambda}_j$, if the eigenvector has unit-length.

Note that the above PCA procedure delivers the eigens (eigenvalues and eigenvectors) of the integrated volatility matrix. According to Aït-Sahalia and Xiu (2019), these eigens are different from the integrated eigens of the spot volatility matrix, when t does not shrink to zero. However, in finite samples, the time horizon t, which is one day in our empirical application, is relatively small compared to Δ_n, which we set equal to 2.5-, 5-, and 10-minute in our application. It is unpractical to further split our daily data into even smaller blocks. Hence, we do not distinguish eigens of integrated volatility versus integrated eigens of spot volatility in our empirical application, following the approach taken by Aït-Sahalia and Xiu (2017).

Note also that eigens are nonlinear functions of the spot volatility matrix. Hence, estimators of the integrated eigens have asymptotic biases (see Jacod and Rosenbaum (2013)). However, according to Aït-Sahalia and Xiu (2019), these bias terms are proportional to their associated eigens. Consequently, they share the same source of predictive power as eigens. Therefore, in terms of forecasting, there is no distortion of information. Hence, we don't remove the bias term in our empirical application.

In general, PCA yields nonzero factor loadings for (almost) all variables. Moreover, although the selected assets from the first step of our procedure all contain relevant information on the target asset volatility, they may have different signal-to-noise ratios. Consequently, if the loadings of low signal-to-noise ratio assets are non-zero, then may decrease overall the signal-to-noise ratio in the principal components. To avoid these drawbacks and to induce further parsimony, we also utilize SPCA in the second step of our procedure. This technique is closely related to PCA (see Jolliffe et al. (2003) and Zou et al. (2006)) and involves estimating "sparse" eigenvectors $\widehat{\widehat{\xi}}_j$, for $j = 1, 2, ..., r$, which are solutions to the following optimization problem:

$$\max_{\|\xi_j\|_2=1, \ \xi_j \perp \xi_1, \cdots, \xi_{j-1}} \xi_j' \widehat{\Sigma}_t \xi_j, \text{ subject to } \sum_{k=1}^{d} |\xi_{j,k}| \leq \delta, \tag{9}$$

where δ is a regularization parameter. Note that the constraint in (9) imposes an $\mathcal{L}_1$-norm penalty on the eigenvectors, and hence induces sparsity. The key feature of SPCA, thus, is that it yields sparse factor loadings, in the sense that loadings may be identically zero, a feature not feasible in the context of shrinkage on the $\mathcal{L}_2$-norm, such as that associated with ridge regression.

4   Experimental Setup

The forecasting models in our IV prediction experiments are all variants of equation (6). More specifically, we choose TRV as the realized measure of integrated volatility. The forecast horizon, h, is set to be one day. Thus, following an HAR structure, equation (6) can be written as:

$$\widehat{y}_{t+1} = \alpha'\widehat{W_t} + \beta'\widehat{\Psi}_t + \epsilon_{t+1}, \tag{10}$$

where $\widehat{y}_{t+1} = \text{TRV}_{t+1}$, $\widehat{W_t} = [1 \ \text{TRV}_t \ \text{TRV}_{[t,t-4]} \ \text{TRV}_{[t,t-21]}]'$, and the latent IV factors, $\widehat{\Psi}_t$, are constructed by implementing the two-step procedure discussed in Section 3. The vector $(\widehat{\alpha}', \widehat{\beta}')$ contains least squares estimates from the regression of $\widehat{y}_{t+1}$ on $[\widehat{W_t}', \ \widehat{\Psi}_t']$. In this context, an important issue is the determination of which IV factors to include in $\widehat{\Psi}_t$. We address this issue in a data-driven manner. Specifically, we set a relatively large value for r, say $r_{\max}$, in the second step of our procedure. Hence, there are $2^{r_{\max}}$ different combinations of the first $r_{\max}$ estimated IV factors under consideration (in our forecasting experiments, we set $r_{\max} = 6$, which yields 64 combinations). We then employ the hv-block cross validation method proposed by Racine (2000) to determine which combination to include in equation (10).

Our method is different from some widely used ones in the following ways. First, unlike Bai and Ng (2002), our method is "forecasting-oriented" instead of "fitting-oriented" in the sense that factors that are highly correlated with many variables in the dataset may still be dropped if they have little predictive power for the target variable. Second, unlike Stock and Watson (1998), it is possible that the combination of IV factors that are selected may not be the first r principal components corresponding to the largest r eigenvalues of the sample covariance matrix.

We examine four factor-augmented models based on our two-step procedure. They are referred to as LASSO-PCA-HFRet, EN-PCA-HFRet, LASSO-SPCA-HFRet, and EN-SPCA-HFRet, and are described in Table 1 under "Model Group" VII. In addition, several groups of alternative models are included in our experiments. (All of these

alternatives are also summarized in Table 1). Our first alternative model is the original HAR model (named HAR-TRV, as TRV is the IV measure used in our experiments), called model Group I in Table 1. This model is given as:

$$\widehat{y}_{t+1} = \alpha' \widehat{W}_t + \epsilon_{t+1}, \tag{11}$$

where $\widehat{y}_{t+1}$ and $\widehat{W}_t$ are defined in (10). Use of this benchmark model allows us to underscore the importance of additional sources of information for volatility forecasting, other than $\widehat{W}_t$. The rest of our alternative models utilize additional predictors and are nested by the following forecasting model:

$$\widehat{y}_{t+1} = \alpha' \widehat{W}_t + \phi' \widehat{\chi}_t + \epsilon_{t+1}, \tag{12}$$

where $\widehat{\chi}_t$ is a set of predictor variables.

Model Group II, which includes PCA-TRV and SPCA-TRV, is closest to the "diffusion index" model of Stock and Watson (1998, 2002a,b). Here, we apply (S)PCA to the daily volatilities of all assets in our dataset, and the resultant principal components are collected in $\widehat{\chi}_t$. In Group III models, denoted by LASSO-TRV and EN-TRV, we only implement the first step of our two-step procedure. Namely, we use either LASSO or EN shrinkage to select a subset of assets whose volatilities contain relevant information on the target asset's volatility. These volatilities are then collected in $\widehat{\chi}_t$. In Group IV models, denoted by PCA-HFRet and SPCA-HFRet, we only apply the second step of our procedure. Namely, we apply (S)PCA to high-frequency (intra-day) returns of all assets on a daily basis. Hence, we only assume that factor loadings are constant within each day, but allow the factor loadings to vary from day to day. We then estimate IV from these asset factors, which are collected in $\widehat{\chi}_t$. In Group V models, denoted by LASSO-PCA-TRV, EN-PCA-TRV, LASSO-SPCA-TRV and EN-SPCA-TRV, we employ the two-step procedure that we proposed. However, these models assume that factor loadings are constant over the entire sample period (this setting mimics the approach often used with discrete-time/low-frequency models) and do not take full advantage of the flexibility of the continuous-time model. Finally, our Group VI model, denoted by PLS-TRV, is similar in spirit to our main two-step procedure. This model employs the partial least squares (PLS) technique introduced by Wold (1966), which takes variable prediction into account when constructing a parsimonious set of factors. A growing literature has documented the good performance of PLS-type methods for

macroeconomic forecasting, when a large number of predictors are available (e.g., see Fuentes et al. (2015) and Groen and Kapetanios (2016)). In our experiments, PLS is applied using the target variable, $\widehat{y}_{t+1}$, and the volatilities of all assets in our dataset, in order to obtain a subset of volatilities which are collected in $\widehat{\chi}_t$. Summarizing, there are 16 forecasting models in our experiments, including 4 models associated with our proposed two-step procedure, and 12 alternative models (refer to Table 1 for details).

In the sequel, we carry out both *in-sample* regression analysis and *out-of-sample* forecast analysis, using a rolling-window estimation scheme. The length of the rolling window is 1259 days (i.e., approximately five years). For example, we first estimate all models using data from January 2, 2004 to December 31, 2008 (1259 trading days), and then construct one-day-ahead forecasts for January 2, 2009. Then, in order to forecast the volatilities on the next trading day (January 5, 2009), we re-estimate all models using data from January 5, 2004 to January 2, 2009 (again, 1259 trading days). We continue this procedure until we reach the end of our dataset. Finally, we obtain sequences of daily *out-of-sample* volatility forecasts for the sample period from January 2, 2009 to December 29, 2017, which contains 2265 trading days. All models are estimated using ordinary least squares.

In order to evaluate the forecasting performance of our models, we utilize three statistics, including:

(i) *In-sample* adjusted R^2;[5]

(ii) *Mean absolute forecasting error (MAFE)*;

(iii) *Out-of-sample* R^2 (see Campbell and Thompson (2008)), defined as:

$$Out\text{-}of\text{-}sample\ R^2 = 1 - \frac{\sum_{t=1}^{T}(y_t - \widehat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y}_t)^2},$$

where y_t is the ex-post value of volatility, $\bar{y}_t$ is the historical average of volatility, and $\widehat{y}_t$ is the forecast of volatility. Note that the *out-of-sample* R^2 values can be negative, indicating that the forecasting performance of the particular model is even worse than simply using historical averages. In addition, we carry out the model confidence set (MCS) tests of Hansen et al. (2011) for multiple model comparisons. Finally, we construct conditional predictive ability (GW) test statistics of the variety introduced in Giacomini and White (2006), for pair-wise model comparisons. In our implementation of both tests, we utilize the $\mathcal{L}_1$ loss function.

---

[5]We report the average *in-sample* adjusted R^2, across all rolling windows.

5   Data

We collect intraday observations on 380 constituents of the S&P 500 index and the SPDR S&P 500 ETF (SPY) from the TAQ database.[6] The full sample period is from January 2, 2004 to December 29, 2017, with 3524 trading days in total. Raw data are first filtered and cleaned, following Brownlees and Gallo (2006). To reduce the effect of microstructure noise, we aggregate the raw data into 2.5-, 5-, and 10-minute sampling frequencies for our forecasting experiments, following standard procedures described in Aït-Sahalia and Jacod (2012). Overnight returns are excluded from our analysis.

Each individual stock in our dataset is assigned to a sector, according to the Global Industry Classification Standard (GICS) code system. As shown in Figure 1, the 4 largest sectors are Consumer Discretionary (CD), Industrials (I), Information Technology (IT), and Health Care (HC). Approximately 53.4% individual stocks in our dataset belong to these four sectors, while the smallest sector, Telecommunication Services (TS), only contains 2.4% of stocks.

In our forecasting experiments, the target asset is SPY. For the sake of brevity, the main empirical findings presented in the sequel are based on data sampled at a 5-minute frequency. However, results based on robustness checks comparing different data sampling frequencies, different forecasting sub-periods, as well as log-scale volatility are also briefly reported.

6   Empirical Findings

6.1   Forecasting Performance

We begin by discussing the one-day-ahead predictive performance of the forecasting models outlined in Section 4, for the forecasting period from January 2, 2009 to December 29, 2017. As mentioned above, all empirical findings set daily SPY volatility as the target variable, and use data sampled at 5-minute frequency. A number of clear-cut findings emerge upon inspection of the results contained in Tables 2 and 3.

First, our two-step shrinkage procedure results in notable improvements in *out-of-sample* predictive accuracy, when comparing *out-of-sample* R^2 and MAFE values. For instance, in Table 2 we see that LASSO-SPCA-HFRet generates an approximately 7.89% increase in *out-of-sample* R^2 and an approximately 7.94% decrease in MAFE,

---

[6]Since the constituents of the S&P 500 index change over time, we only collect those that are always in the index from 2004 to 2017.

when compared to our benchmark HAR model (i.e., HAR-TRV). Additionally, the MCS test p-values reported in Table 2 confirm that factor-augmented models based on our proposed two-step procedure exhibit the best forecasting performance, among all models considered in this paper. [7] Indeed, three out of our four proposed models (from Group VII in Table 1) are the only models selected for the confidence set (these are the starred entries in the "MCS p-value column of the table), in a 90% model confidence set. Moreover, the p-values for all other models (with the exception of our fourth proposed model) are extremely small, indicating that the alternative models in Groups I - VI of Table 1 are predictively inferior.

Second, based on *out-of-sample* R^2 and MAFE values, the parsimonious benchmark HAR model (i.e., HAR-TRV) generally outperforms all other models, with the exception of the 4 aforementioned models in Group VII of Table 1. For example, inspection of Table 2 indicates that none of the alternative models in Groups II - VI have lower MAFE than HAR-TRV, except for our four double shrinkage models, which *all* have lower MAFE than HAR-TRV. Even worse, LASSO-TRV and EN-TRV actually generate negative *out-of-sample* R^2 values. This supports the claim that there is no guarantee that big data methods will yield superior predictive performance. Perhaps this is not surprising, given the extant empirical evidence suggesting how well the HAR model performs in prediction contexts. However, this finding does clearly indicate that the superior performance of our two-step procedure is a non-trivial outcome. Indeed, it appears that big data do contain useful information, and the trick is distinguishing between relevant and irrelevant information, when faced with high-frequency and high dimensional datasets.

Third, a less important, although still interesting finding emerges upon inspection of the *in-sample* results reported in the column denoted by "*IS-AdjR^2*" in Table 2. Here, we observe that all other models have better *in-sample* fit than the original HAR model. In particular, the increases in *in-sample* adjusted R^2 for various models range from 0.57% to 35.28%, when compared with the *in-sample* adjusted R^2 value of the HAR-TRV model. Thus, based solely on *in-sample* diagnostics, there appears to be substantial and broad gains associated with adding volatility factors to the benchmark HAR model, no matter how the factors are constructed. However, significant increases in ex ante predictive performance are found only for very few models, as discussed

---

[7]Note that, the way in which MCS p-values are constructed is different from that used for a traditional t-test. A model with a MCS p-value greater than the significance level indicates that its corresponding null hypothesis of equivalence is rejected and the model is eliminated from the sequential testing.

above. An extreme case is observed for LASSO-TRV and EN-TRV models. In these models, LASSO or EN shrinkage is performed to find stocks whose daily IVs are useful for predicting the target asset's IV, and those are used as predictors in the forecasting models. This procedure results in a remarkable improvement in *in-sample* fit. However, *out-of-sample* forecasting results are very poor, indicated by negative *out-of-sample* R^2 values, as well as over 50% larger MAFEs, compared with the HAR-TRV model. This finding constitutes strong evidence of an important difference between findings based on *in-* and *out-of-sample* experiments. It also indicates that LASSO and EN shrinkage may not be effective, when applied without recourse to further dimension reduction and shrinkage, such as that associated with the use of our two-step procedure.

We now discuss our fourth finding, which is based on examination of the results reported in Table 3, in which pair-wise GW test statistics are reported. Here, a significantly negative GW statistic indicates that the model in the corresponding row (the alternative model) has larger MAFE, when compared to the model in the corresponding column (one of our proposed four models from Group VII of Table 1). Consider the four models of Group VII by examining the last four rows of entries in the table. Evidently, there is little to choose between these four models, with the only exception that the EN-SPCA-HFRet model is dominated by the LASSO-SPCA-HFRet model, at the 10% significance level. Thus, all of our proised models appear to be adequate for volatility forecasting.

Fifth, examination of the statistics reported in the second and third rows of Table 3 further confirms that solely using the first step of our two-step procedure does not yield adequate forecasts. Indeed, GW test statistics for this Group of alternative models are approximately -100, indicating the worst forecasting performance among all alternative models. Moreover, the statistics contained the fourth to seventh rows of the table demonstrate that the widely used "diffusion index" type models (see Group II in Table 1) are also dominated by our proposed models. Our empirical findings therefore underscore the importance of "targeted" variable selection, such as that carried out using the LASSO or EN in the first step of our procedure. Still, inspection of the statistics reported in the eighth through twelfth rows in the table even targeted shrinkage is not enough, since our Group V and VI alternative models (i.e., LASSO-PCA-TRV, EN-PCA-TRV, LASSO-SPCA-TRV, EN-SPCA-TRV and PLS-TRV) are all significantly dominated by our proposed two-step shrinkage method, at the 1% level). This suggests that it is very important to take advantage of the continuous-time factor structures assumed in (4)

16

and (5), when performing multi-step targeted shrinkage. As discussed in Section 4, this structure allows us to specify our Group VII models in a flexibile manner, in the sense that factor loadings are only constant within a day, instead of being constant over the entire sample period, as is the case with the models of Group V in Table 1.

6.2 Selected Stocks and Latent Factors

Figures 2–4 and Table 4 summarize the results from each step of our proposed two-step shrinkage method. More specifically, Figure 2 shows the numbers of selected stocks in the first step of our procedure. Figure 3 presents the percentages of stocks (by sector) that comprise the variables selected in the first step of our procedure. Table 4 gives the most frequently selected stocks that are utilized in the construction of the latent IV factors in the second step of our procedure. Finally, a summary of the number of estimated IV factors used in our Group VII forecasting models are reported in Figure 4. A number of interesting findings based on these figures and tables are summarized below.

We begin by discussing the results from the use of our two different variable selection methods (i.e., the LASSO and EN), which are used in the first step of our procedure. First, as shown in Figure 2, the number of individual stocks selected by different variable selection methods is similar, over time. Not surprisingly, then, we see in Table 4, where the 50 most frequently selected individual stocks are reported (in descending order), 47 out of 50 stocks selected by the LASSO are also selected by the EN. Even the specific frequencies of selection for particular stock are similar across the two shrinkage operators, suggesting that there is little to choose between them in our context. However, the number of selected stocks may vary markedly, across different rolling windows, as evidenced in Figure 2. More interestingly, for periods when the number of stocks selected varies substantially, both variable shrinkage methods choose a relative large number of stocks. For instance, over the period of financial crisis (i.e., 2008 and 2009), the numbers of selected stocks usually ranges between approximately 100 and 150, and sometimes even surpasses 200. On the other hand, far fewer stocks are selected in other periods. This indicates that when the financial markets are in turmoil, individual stocks tend to be more correlated, as might be expected.

Second, Figure 3a shows that the industrial stocks tend to be selected most frequently, in the first step of our procedure. This is not surprising, given that the industrial sector is one of the largest among all sectors (see Figure 1). This is also

likely due to the fact that in our dataset there are less stocks in energy and financial sectors than in the industrial sector. However, when rescaled by the size of each sector, as depicted in Figure 3b, we find that the percentage of stocks in the energy sector is roughly 5% higher than that of industrial stocks, when selected using either the LASSO or the EN. Finally, recall that Table 4 contains the frequencies of selection associated with the most oft selected stocks. For example, the most frequently selected stock (i.e., *Williams Companies, Inc.* (WMB)), which is in the energy sector, is used in around 88% of the daily prediction models. Interestingly, stocks in the industrial, financial, health care and information technology sectors are also frequently selected, constituting around 14%, 20%, 18% and 12% of the 50 stocks listed in the table, respectively.

Third, turn to Figure 4, which depicts the average number of factors used in our four Group VII models, across the entire prediction period. Evidently (see Figure 4a), around 50% to 60% of the time only one factor is used in our forecasting models.[8] This suggests that our forecasting models are quite parsimonious. Turning to the rest of the plots in this figure, recall that, as discussed in Section 4, the way in which we determine which volatility factors are included in our models is more flexible than the approach used for factor-augmented models in the extant literature. This is reflected in Figures 4b – 4d. For example, inspection of Figure 4c indicates that, conditional on only one factor being used, our two-step shrinkage approach often chooses factors that are not associated with the largest eigenvalue. For instance, for approximately 12% to 23% of the periods, it is the third factor that is employed in the prediction model. Similarly, conditional on two factors being included, the combination of the second and third IV factors (see the bars indexed by "2&3" in Figure 4d) are the most commonly used predictors, with frequencies ranging from between approximately 18% and 26%, in our forecasting models, rather than the first two. Finally, we observe that the overall frequency at which each factor is incorporated in the forecasting models is stable, to some degree (see Figure 4b). These results again underscore the importance of targeted shrinkage.

---

[8]Note that the average number of factors used in the LASSO-PCA-HFRet, EN-PCA-HFRet, LASSO-SPCA-HFRet, and EN-SPCA-HFRet models is 1.591, 1.585, 1.565, and 1.769, respectively.

6.3 Robustness Checks

6.3.1 Log-transformation of IV Estimates

We replicated all of our experiments using logged data (i.e. logs of all TRV and latent factors were used in the 16 forecasting models discussed above). Table 5 reports results using this experimental setup that are comparable with those reported in Table 2.[9] Comparing the findings from these tables, we note the following. First, both *in-sample* and *out-of-sample* R^2 values for the various forecasting models increase and become much more stable, in the sense that R^2 values vary only within a relatively small range. For instance, *in-sample* R^2 values range between approximately 72% and 74%, and *out-of-sample* R^2 values range between approximately 73% and 77%. We still observe that all factor-augmented models achieve a higher *in-sample* R^2 values than the that associated with our benchmark HAR model. However, similar to the results reported in Section 6.1, only a very limited number of models have values that are greater than that of the benchmark HAR model. Second, and most importantly, the 90% model confidence set contains all of the four specifications associated with our proposed two-step shrinkage method, indicating that the superior performance of our proposed factor-augmented models is preserved when logs are taken. Finally, the forecasting performance of LASSO-TRV and EN-TRV is much more in line with the performance of our other alternative models under the log-transformation. This can be seen by comparing Tables 2 and 5. Figure 5 may offer an explanation for this empirical finding. Namely, we see in Figure 5 (compare with Figure 2) that the average number of stocks selected drops from approximately 50 to 30, resulting in substantially fewer predictors being included in LASSO-TRV and EN-TRV, under the log transformation.

6.3.2 Different Data Sampling Frequencies

We also carried out experiments in which we replicated our experiments using 2.5- and 10-minute data frequencies. Table 6 contains results from these experiments (for the sake of brevity, only MAFEs and the MCS test results are shown in table - additional results are available upon request). Again, we see that only the four specifications in Group VII yield lower MAFEs than that of the benchmark HAR model. Additionally, we again observe that the Group VII models are the only ones selected using the MCS confidence set test, at the 10% significance level, with the only exception being that,

---

[9]Although all experiments reported above were run in log-scale, for the sake of brevity, we reproduce only Table 2. Other results are available upon request from the authors.

when using 10-minute frequency data, the HAR-TRV model is also included in the 90% model confidence set. This suggests that while our models are still superior, the performance of the benchmark HAR model is approaching that of our models, as the sampling frequency decreases (note that the p-value for the benchmark model still remains approximately one half as large as those associated with our 4 proposed Group VII models, however). We conjecture the following explanation for this finding. A relatively low sampling frequency may result in less accurate estimates of integrated volatility, since jump variation components become more difficult to be excluded at lower frequencies. Additionally, the common factors estimated by PCA or SPCA may be more inaccurate when less data are utilized in their estimation, over a fixed interval. However, it should be stressed that are models still perform very well, at all three sampling frequencies considered in this paper.

6.3.3 Different Forecasting Sub-periods

Finally, we replicated our experiments using different forecasting periods. Specifically, we divided the full forecasting period from 01/02/2009 - 12/29/2017 into three equal sub-periods, including 01/02/2009 - 12/30/2011 (called period P1), 01/03/2012 - 12/31/2014 (called period P2), and 01/02/2015 - 12/29/2017 (called period P3). On average, there are 755 trading days over each period. MAFEs and MCS results are gathered in Table 7. Inspection of the results in this table reveal that for P1, the MAFEs for all models are generally much larger than those associated with the other two sub-periods. This is not surprising since the data used during this period are substantially more volatile than the rest periods. Still, our proposed factor-augmented models (i.e., those in Group VII) are the only ones that yield MAFEs below 0.01, and are the only ones included in the 90% model confidence set. As might be expected, during P2 (when financial markets were quite calm), MAFEs for all models decline sharply to around 25 - 30% of the levels seen during P1. In addition, the MAFEs during this period for all models vary within a relatively narrow range, from approximately 0.003 to 0.005. This observation suggests that it may be more challenging for a particular model to significantly outperform other models during P2. However, according to MCS test results reported in the table, only LASSO-PCA-HFRet, EN-PCA-HFRet and LASSO-SPCA-HFRet are contained in the 90% model confidence set, indicating yet again the superiority of our proposed shrinkage method. Finally, during P3, the MAFEs for various models are also contained within a small interval, but MCS test results are rather mixed. Indeed, only three

models (i.e., LASSO-TRV, EN-TRV, and LASSO-SPCA-TRV) are not included in the 90% model confidence set. Furthermore, the original HAR model achieves the lowest MAFE of 0.0046 and the highest MCS p-value of 1.0000. Therefore, our proposed factor-augmented models are not superior during P3 - instead, multiple other models are also useful.

7 Concluding Remarks

This paper investigates the forecasting benefits associate with using a new double shrinkage based specification methodology, in which latent IV factors are estimated using a combination of LASSO or ENt shrinkage, followed by factor estimation using (S)PCA). Our key finding is that we uncover substantial empirical evidence indicating that latent common volatility factors greatly improve the *out-of-sample* predictive accuracy of HAR models, as measured by both *out-of-sample* R^2 and mean absolute forecasting error (MAFE). Additionally model confidence sets almost always exculsively contain our proposed models, even though a wide variety of natural alternative specifications are included in our experiments. This result is robust to the use of original or log-scale data in model specification, as well as to the use of different data frequencies and different sample periods. This paper is meant as a starting point, as much remains to be done. For example, although our approach draws heavily on theoretical advances in the application of principal component analysis to high dimensional asset return made in Aït-Sahalia and Xiu (2017, 2019), it remains to ascertain whether our findings carry over to the use of other shrinkage methods, such as independent component analysis. It also remains to theoretically analyze the higher-order latent volatility factors that are extracted from first-order latent factors constructed using observed (asset) data. From an empirical perspective, it remains to assess whether the findings in this paper can be translated into profitable investment strategies, in real-time trading contexts, for example.

References

Aït-Sahalia, Y. and Jacod, J. 2012. Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature* 50:1007–1050.

Aït-Sahalia, Y. and Jacod, J. 2014. High-frequency financial econometrics. Princeton University Press.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. 2011. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* 160:160–175.

Aït-Sahalia, Y. and Xiu, D. 2017. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics* 201:384–399.

Aït-Sahalia, Y. and Xiu, D. 2019. Principal component analysis of high-frequency data. *Journal of the American Statistical Association* 114:287–303.

Andersen, T., Bollerslev, T., Diebold, F. X., and Labys, P. 2001. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96:42–55.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *Review of Economics and Statistics* 89:701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579–625.

Andersen, T. G., Bollerslev, T., and Meddahi, N. 2004. Analytical evaluation of volatility forecasts. *International Economic Review* 45:1079–1110.

Andersen, T. G., Bollerslev, T., and Meddahi, N. 2011. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* 160:220–234.

Audrino, F. and Hu, Y. 2016. Volatility forecasting: Downside risk, jumps and leverage effect. *Econometrics* 4:1–24.

AUDRINO, F., SIGRIST, F., AND BALLINARI, D. 2019. The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting, In Press* .

BAI, J. AND NG, S. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70:191–221.

BAI, J. AND NG, S. 2006a. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* 74:1133–1150.

BAI, J. AND NG, S. 2006b. Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics* 131:507–537.

BAI, J. AND NG, S. 2008. Forecasting economic time series using targeted predictors. *Journal of Econometrics* 146:304–317.

BARNDORFF-NIELSEN, O. E., GRAVERSEN, S. E., JACOD, J., AND SHEPHARD, N. 2006. Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22:677–719.

BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2:1–37.

BOLLERSLEV, T., PATTON, A. J., AND QUAEDVLIEG, R. 2016. Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics* 192:1–18.

BRANDT, M. W. AND JONES, C. S. 2006. Volatility forecasting with range-based egarch models. *Journal of Business & Economic Statistics* 24:470–486.

BROWNLEES, C. T. AND GALLO, G. M. 2006. Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis* 51:2232–2245.

CAMPBELL, J. Y. AND THOMPSON, S. B. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:1509–1531.

CARRASCO, M. AND ROSSI, B. 2016. In-sample inference and forecasting in misspecified factor models. *Journal of Business & Economic Statistics* 34:313–338.

CORRADI, V., DISTASO, W., AND FERNANDES, M. 2020. Testing for jump spillovers without testing for jumps. *Journal of the American Statistical Association* forthcoming.

CORRADI, V., DISTASO, W., AND SWANSON, N. R. 2009. Predictive density estimators for daily volatility based on the use of realized measures. *Journal of Econometrics* 150:119–138.

CORRADI, V., DISTASO, W., AND SWANSON, N. R. 2019. Predictive inference for integrated volatility. *Journal of the American Statistical Association* 114:1880–1893.

CORSI, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7:174–196.

CORSI, F., PIRINO, D., AND RENÒ, R. 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159:276–288.

DUONG, D. AND SWANSON, N. R. 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics* 187:606–621.

FAN, J., KE, Y., SUN, Q., AND ZHOU, W.-X. 2011. Farmtest: factor-adjusted robust multiple testing with approximate false discovery control. *Journal of the American Statistical Association* 106:1496–1512.

FERNANDES, M., MEDEIROS, M. C., AND SCHARTH, M. 2014. Modeling and predicting the cboe market volatility index. *Journal of Banking and Finance* 40:1–10.

FUENTES, J., PONCELA, P., AND RODRÍGUEZ, J. 2015. Sparse partial least squares in time series for macroeconomic forecasting. *Journal of Applied Econometrics* 30:576–595.

GHYSELS, E., SANTA-CLARA, P., AND VALKANOV, R. 2006. Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* 131:59–95.

GHYSELS, E. AND SINKO, A. 2011. Volatility forecasting and microstructure noise. *Journal of Econometrics* 160:257–271.

GIACOMINI, R. AND WHITE, H. 2006. Tests of conditional predictive ability. *Econometrica* 74:1545–1578.

GROEN, J. J. AND KAPETANIOS, G. 2016. Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics & Data Analysis* 100:221–239.

HANSEN, P. R. AND LUNDE, A. 2005. A forecast comparison of volatility models: Does anything beat a garch (1, 1)? *Journal of Applied Econometrics* 20:873–889.

HANSEN, P. R., LUNDE, A., AND NASON, J. M. 2011. The model confidence set. *Econometrica* 79:453–497.

JACOD, J. 2018. Limit of random measures associated with the increments of a brownian semimartingale. *Journal of Financial Econometrics* 16:526–569.

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M., AND VETTER, M. 2009. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications* 119:2249–2276.

JACOD, J. AND ROSENBAUM, M. 2013. Quarticity and other functionals of volatility: Efficient estimation. *The Annals of Statistics* 41:1462–1484.

JOLLIFFE, I. T., TRENDAFILOV, N. T., AND UDDIN, M. 2003. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics* 12:531–547.

MANCINI, C. 2001. Disentangling the jumps of the diffusion in a geometric jumping brownian motion. *Giornale dell'Istituto Italiano degli Attuari* LXIV:19–47.

MANCINI, C. 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36:270–296.

MEDDAHI, N. 2001. An eigenfunction approach for volatility modeling. *Université de Montréal. Département de Sciences Économiques.* .

PATTON, A. J. AND SHEPPARD, K. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97:683–697.

PODOLSKIJ, M. AND ZIGGEL, D. 2010. New tests for jumps in semimartingale models. *Statistical Inference for Stochastic Processes* 13:15–41.

RACINE, J. 2000. Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics* 99:39–61.

STOCK, J. H. AND WATSON, M. W. 1998. Diffusion indexes. *NBER Working Paper* 6702.

STOCK, J. H. AND WATSON, M. W. 2002a. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97:1167–1179.

STOCK, J. H. AND WATSON, M. W. 2002b. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20:147–162.

STOCK, J. H. AND WATSON, M. W. 2006. Forecasting with many predictors, pp. 515–554. *In* G. Elliott, C. Granger, and A. Timmermann (eds.), Handbook of Forecasting, vol. 1. North Holland.

SWANSON, N. R. AND XIONG, W. 2018. Big data analytics in economics: What have we learned so far, and where should we go from here? *Canadian Journal of Economics* 51:695–746.

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

WOLD, H. 1966. Estimation of principal components and related models by iterative least squares, pp. 391–420. *In* P. Krishnaiah (ed.), Multivariate Analysis. Academic Press.

ZOU, H. AND HASTIE, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67:301–320.

ZOU, H., HASTIE, T., AND TIBSHIRANI, R. 2006. Sparse principal component analysis. *Journal of Computational and Graphical Statistics* 15:265–286.

Table 1: Brief Descriptions of Forecasting Models*

	Model	Description
I	HAR-TRV	Only lagged volatilities of the target asset are used in model specification.
II	PCA-TRV	HAR augmented by principal components (PCs) constructed from volatility estimates.
	SPCA-TRV	HAR augmented by sparse principal components (SPCs) constructed from volatility estimates.
III	LASSO-TRV	HAR augmented by volatility estimates of assets selected using the least absolute shrinkage operator (LASSO).
	EN-TRV	HAR augmented by volatility estimates of assets selected using the elastic net (EN).
IV	PCA-HFRet	HAR augmented by volatilities of PCs constructed from returns of all assets.
	SPCA-HFRet	HAR augmented by volatilities of SPCs constructed from returns of all assets.
V	LASSO-PCA-TRV	HAR augmented by volatilities of PCs constructed from LASSO-selected asset returns, with constant factor loadings.
	EN-PCA-TRV	HAR augmented by volatilities of PCs constructed from EN-selected asset returns, with constant factor loadings.
	LASSO-SPCA-TRV	HAR augmented by volatilities of SPCs constructed from LASSO-selected asset returns, with constant factor loadings.
	EN-SPCA-TRV	HAR augmented by volatilities of SPCs constructed from EN-selected asset returns, with constant factor loadings.
VI	PLS-TRV	HAR augmented by volatilites selected using partial least squares (PLS).
VII	LASSO-PCA-HFRet	HAR augmented by volatilities of PCs constructed from LASSO-selected asset returns, with flexible factor loadings.
	EN-PCA-HFRet	HAR augmented by volatilities of PCs constructed from EN-selected asset returns, with flexible factor loadings.
	LASSO-SPCA-HFRet	HAR augmented by volatilities of SPCs constructed from LASSO-selected asset returns, with flexible factor loadings.
	EN-SPCA-HFRet	HAR augmented by volatilities of SPCs constructed from EN-selected asset returns, with flexible factor loadings.

* Notes: The HAR-TRV model uses no additional information other than the history of the target asset. The models in Group III (i.e., LASSO-TRV and EN-TRV) are based on the first step of our proposed two-step procedure. The models in Group IV (i.e. PCA-HFRet and SPCA-HFRet) are based on the second step of our proposed two-step procedure. The models in Group V (i.e. LASSO-PCA-TRV, EN-PCA-TRV, LASSO-SPCA-TRV and EP-SPCA-TRV), are the same as those based on our procedure, except that factor loadings are fixed. The models in Group are the models proposed in this paper (all other models are benchmark models against which our models are compared).

27

Table 2: Forecasting Results*

Forecasting Period: Jan. 02, 2009 – Dec. 29, 2017				
Forecasting Model	$IS\text{-}AdjR^2$	$OOS\text{-}R^2$	MAFE	MCS p-value
HAR-TRV	0.5604	0.5579	0.0063	0.0158
PCA-TRV	0.5641	0.5095	0.0067	0.0007
SPCA-TRV	0.5636	0.5182	0.0066	0.0010
LASSO-TRV	0.7578	-0.4111	0.0098	0.0000
EN-TRV	0.7581	-0.4301	0.0098	0.0000
PCA-HFRet	0.5662	0.5581	0.0065	0.0007
SPCA-HFRet	0.5707	0.5508	0.0066	0.0007
LASSO-PCA-TRV	0.5806	0.5292	0.0070	0.0001
EN-PCA-TRV	0.5822	0.5088	0.0069	0.0005
LASSO-SPCA-TRV	0.5847	0.5164	0.0070	0.0001
EN-SPCA-TRV	0.5834	0.5217	0.0069	0.0004
PLS-TRV	0.5791	0.4262	0.0071	0.0007
LASSO-PCA-HFRet	0.5998	0.5524	0.0059	0.4920*
EN-PCA-HFRet	0.6020	0.5691	0.0059	0.4920*
LASSO-SPCA-HFRet	0.5973	0.6019	0.0058	1.0000*
EN-SPCA-HFRet	0.6030	0.5590	0.0060	0.0313

*Notes: Entries in this table include *in-sample* adjusted R^2 (*IS-AdjR²*), *out-of-sample* R^2 (*OOS-R²*), mean absolute forecasting error (MAFE) and MCS p-values associated with the model confidence set test discussed in Section 4, for the target asset SPY. 90% model confidence set members are identified by one models with asterisk. All results are from forecasting experiments that utilize 5-minute frequency data. For complete details on various model specifications as well as evaluation criteria, refer to Section 4.

Table 3: Predictive Accuracy Test Result for Pair-Wise Model Comparisons*

Forecasting Model	LASSO-PCA-HFRet	EN-PCA-HFRet	LASSO-SPCA-HFRet	EN-SPCA-HFRet
HAR-TRV	-8.66** (0.013)	-8.44** (0.015)	-18.69*** (0.000)	-5.82* (0.055)
PCA-TRV	-34.15*** (0.000)	-30.57*** (0.000)	-37.36*** (0.000)	-27.41*** (0.000)
SPCA-TRV	-20.09*** (0.000)	-19.10*** (0.000)	-36.90*** (0.000)	-21.14*** (0.000)
LASSO-TRV	-100.94*** (0.000)	-116.01*** (0.000)	-100.94*** (0.000)	-104.88*** (0.000)
EN-TRV	-98.80*** (0.000)	-111.12*** (0.000)	-100.07*** (0.000)	-101.72*** (0.000)
PCA-HFRet	-18.93*** (0.000)	-17.82*** (0.000)	-34.65*** (0.000)	-13.29*** (0.000)
SPCA-HFRet	-21.35*** (0.000)	-19.79*** (0.000)	-37.97*** (0.000)	-15.72*** (0.000)
LASSO-PCA-TRV	-52.38*** (0.000)	-50.50*** (0.000)	-96.94*** (0.000)	-41.09*** (0.000)
EN-PCA-TRV	-45.06*** (0.000)	-45.86*** (0.000)	-82.12*** (0.000)	-40.46*** (0.000)
LASSO-SPCA-TRV	-60.94*** (0.000)	-60.34*** (0.000)	-105.19*** (0.000)	-48.62*** (0.000)
EN-SPCA-TRV	-35,67*** (0.000)	-36.96*** (0.000)	-61.75*** (0.000)	-31.60*** (0.000)
PLS-TRV	-42.02*** (0.000)	-38.92*** (0.000)	-60.28*** (0.000)	-41.67*** (0.000)
LASSO-PCA-HFRet	— —	-2.14 (0.343)	-1.09 (0.580)	2.61 (0.271)
EN-PCA-HFRet	2.14 (0.343)	— —	-0.76 (0.685)	4.41 (0.110)
LASSO-SPCA-HFRet	1.09 (0.580)	0.76 (0.685)	— —	4.97* (0.083)
EN-SPCA-HFRet	-2.61 (0.271)	-4.41 (0.110)	-4.97* (0.083)	— —

* Notes: Entries are GW test statistics and p-values (in parentheses). A negative test statistic indicates the forecasting model in the corresponding row generates a higher average loss, relative to the model in the corresponding column. ***, ** and * indicate rejections at 1%, 5% and 10% significance levels, respectively. For further details, refer to Section 4.

Table 4: Variable Selection Frequencies (Individual Stocks)*

	LASSO			Elastic Net	
Ticker	Sector	Frequency	Ticker	Sector	Frequency
WMB	E	0.872	WMB	E	0.886
LUV	I	0.634	LUV	I	0.633
DO	E	0.598	DO	E	0.603
CERN	HC	0.555	CERN	HC	0.557
SLM	F	0.547	DGX	HC	0.557
CSX	I	0.534	SLM	F	0.547
PRU	F	0.524	CSX	I	0.538
SRE	U	0.523	PRU	F	0.524
R	I	0.521	SRE	U	0.523
DGX	HC	0.509	R	I	0.521
MMC	F	0.428	AAPL	IT	0.466
AAPL	IT	0.427	MMC	F	0.435
ROK	I	0.426	ROK	I	0.430
AMD	IT	0.401	CNX	E	0.401
MCD	CD	0.398	MRK	HC	0.400
CNX	E	0.391	AMD	IT	0.397
CI	HC	0.385	FOSL	CD	0.389
MRK	HC	0.379	CI	HC	0.388
FOSL	CD	0.373	MCD	CD	0.387
SWKS	IT	0.362	COG	E	0.383
FITB	F	0.357	DE	I	0.358
TGT	CD	0.349	SWKS	IT	0.354
DE	I	0.345	FITB	F	0.354
LNC	F	0.343	TGT	CD	0.348
ADS	IT	0.336	LNC	F	0.343
VLO	E	0.329	VLO	E	0.334
FISV	IT	0.325	YUM	CD	0.332
KEY	F	0.315	ADS	IT	0.329
PVH	CD	0.311	FISV	IT	0.325
UHS	HC	0.306	PVH	CD	0.313
AFL	F	0.297	KEY	F	0.307
WYNN	CD	0.296	UHS	HC	0.307
AIV	RE	0.295	RCL	CD	0.302
CTXS	IT	0.294	AIV	RE	0.297
OMC	TS	0.294	WYNN	CD	0.297
UTX	I	0.294	CNC	HC	0.295
BK	F	0.293	UTX	I	0.295
CNC	HC	0.292	BAX	HC	0.294
ILMN	HC	0.291	CTXS	IT	0.294
ECL	M	0.290	BK	F	0.293
FL	CD	0.290	ECL	M	0.292
BAX	HC	0.290	CAH	HC	0.291
C	F	0.290	ILMN	HC	0.291
FFIV	IT	0.290	OMC	TS	0.291
JPM	F	0.289	C	F	0.289
CAH	HC	0.289	TIF	CD	0.288
GWW	I	0.289	AFL	F	0.288
IFF	M	0.288	FFIV	IT	0.288
TIF	CD	0.288	HSY	CS	0.287
HSY	CS	0.288	JPM	F	0.287

*Notes: Entries in the columns denoted by "Frequency" indicate the frequency at which a particular stock is selected in the first step by LASSO or EN, across all prediction periods. For brevity, we only list the 50 most frequently chosen stocks in this table. Stock tickers as well as the sectors to which they belong are also reported. See Sections 4 and 6.3 for further discussion.

Table 5: Forecasting Results for Experiments Replicated Using Log-scale Data*

	Forecasting Period: Jan. 02, 2009 – Dec. 29, 2017			
Forecasting Model	$IS\text{-}AdjR^2$	$OOS\text{-}R^2$	MAFE	MCS p-value
HAR-TRV	0.7225	0.7446	0.4382	0.0011
PCA-TRV	0.7244	0.7459	0.4371	0.0005
SPCA-TRV	0.7242	0.7455	0.4369	0.0033
LASSO-TRV	0.7335	0.7400	0.4418	0.0000
EN-TRV	0.7333	0.7396	0.4421	0.0000
PCA-HFRet	0.7236	0.7461	0.4371	0.0005
SPCA-HFRet	0.7237	0.7458	0.4373	0.0000
LASSO-PCA-TRV	0.7362	0.7421	0.4412	0.0000
EN-PCA-TRV	0.7362	0.7418	0.4418	0.0000
LASSO-SPCA-TRV	0.7347	0.7435	0.4387	0.0008
EN-SPCA-TRV	0.7347	0.7427	0.4395	0.0000
PLS-TRV	0.7378	0.7347	0.4429	0.0011
LASSO-PCA-HFRet	0.7324	0.7656	0.4108	1.0000*
EN-PCA-HFRet	0.7324	0.7655	0.4109	0.8260*
LASSO-SPCA-HFRet	0.7324	0.7643	0.4116	0.5945*
EN-SPCA-HFRet	0.7323	0.7647	0.4114	0.5945*

*Notes: See notes to Table 2. For further details, refer to Section 6.3.1.

Table 6: Forecasting Results with Different Sampling Frequencies*

Forecasting Period: Jan. 02, 2009 – Dec. 29, 2017				
Forecasting Model	MAFE		MCS p-value	
	2.5-min	10-min	2.5-min	10-min
HAR-TRV	0.0060	0.0066	0.0329	0.1925*
PCA-TRV	0.0066	0.0073	0.0014	0.0084
SPCA-TRV	0.0066	0.0074	0.0014	0.0088
LASSO-TRV	0.0092	0.0102	0.0000	0.0008
EN-TRV	0.0091	0.0102	0.0000	0.0014
PCA-HFRet	0.0064	0.0068	0.0014	0.0088
SPCA-HFRet	0.0063	0.0069	0.0014	0.0088
LASSO-PCA-TRV	0.0075	0.0077	0.0005	0.0014
EN-PCA-TRV	0.0070	0.0076	0.0005	0.0014
LASSO-SPCA-TRV	0.0075	0.0077	0.0005	0.0014
EN-SPCA-TRV	0.0070	0.0077	0.0005	0.0014
PLS-TRV	0.0069	0.0073	0.0014	0.0007
LASSO-PCA-HFRet	0.0057	0.0063	0.3871*	0.3717*
EN-PCA-HFRet	0.0056	0.0063	1.0000*	0.7114*
LASSO-SPCA-HFRet	0.0056	0.0063	0.8023*	0.3717*
EN-SPCA-HFRet	0.0056	0.0062	0.8023*	1.0000*

*Note: This table summarizes forecasting results based on experiments that utilize data sampled at 2.5- and 10- minute frequencies, respectively. For further details, refer to Table 2 and Section 6.3.2.

Table 7: Forecasting Results for Different Prediction Periods*

Forecasting Model	MAFE			MCS p-value		
	P1	P2	P3	P1	P2	P3
HAR-TRV	0.0106	0.0038	0.0046	0.0082	0.0008	1.0000*
PCA-TRV	0.0112	0.0037	0.0050	0.0037	0.0008	0.5682*
SPCA-TRV	0.0113	0.0037	0.0048	0.0037	0.0008	0.7262*
LASSO-TRV	0.0181	0.0048	0.0065	0.0000	0.0000	0.0004
EN-TRV	0.0183	0.0048	0.0064	0.0000	0.0000	0.0004
PCA-HFRet	0.0107	0.0038	0.0051	0.0040	0.0008	0.5565*
SPCA-HFRet	0.0108	0.0039	0.0051	0.0037	0.0008	0.5563*
LASSO-PCA-TRV	0.0117	0.0037	0.0054	0.0007	0.0008	0.1316*
EN-PCA-TRV	0.0117	0.0036	0.0054	0.0037	0.0008	0.1501*
LASSO-SPCA-TRV	0.0119	0.0038	0.0054	0.0007	0.0008	0.0846
EN-SPCA-TRV	0.0116	0.0038	0.0053	0.0037	0.0008	0.1501*
PLS-TRV	0.0123	0.0038	0.0052	0.0037	0.0008	0.5565*
LASSO-PCA-HFRet	0.0096	0.0030	0.0050	0.9768*	1.0000*	0.5682*
EN-PCA-HFRet	0.0096	0.0031	0.0049	0.9768*	0.4730*	0.7262*
LASSO-SPCA-HFRet	0.0096	0.0030	0.0047	1.0000*	0.6855*	0.7262*
EN-SPCA-HFRet	0.0097	0.0032	0.0050	0.3465*	0.0083	0.5682*

*Note: This table summarizes forecasting results based on experiments using different forecasting periods. Here, P1 denotes the period from 01/02/2009 to 12/30/2011, P2 denotes the period from 01/03/2012 to 12/31/2014, and P3 denotes the period from 01/02/2015 to 12/29/2017. For further details, refer to Table 2 and Section 6.3.3.
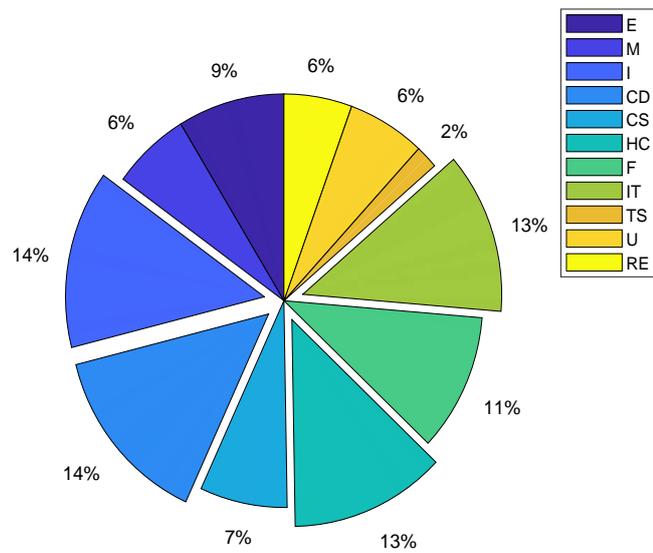
Figure 1: Percentage of Stocks in each Sector of the S&P 500 Stocks in Experimental Dataset*

*Notes: Following the Global Industry Classification Standard (GICS) coding system, the 380 constituents of the S&P 500 index in our dataset are classified into 11 sectors. The percentage of stocks in each sector is reported in this figure. The 11 sectors are Energy (E), Materials (M), Industrials (I), Consumer Discretionary (CD), Consumer Staples (CS), Health Care (HC), Financials (F), Information Technology (IT), Telecommunication Services (TS), Utilities (U), and Real Estate (RE). The 4 largest sectors (I, CD, HC, IT) are exploded in the pie chart for emphasis.
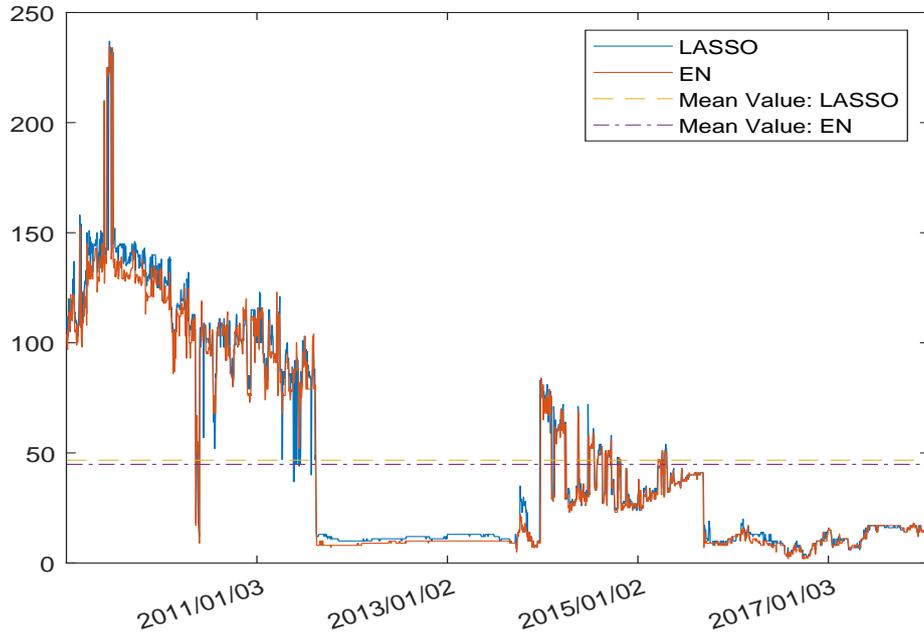
Figure 2: Number of Selected Stocks Using LASSO and EN Shrinkage*

*Notes: The number of individual stocks selected using the LASSO and EN for each rolling data window used in forecast model specification is plotted, over the prediction period. The dashed horizontal lines are the averages, over all rolling windows.


(a) Absolute Frequency

(b) Relative Frequency
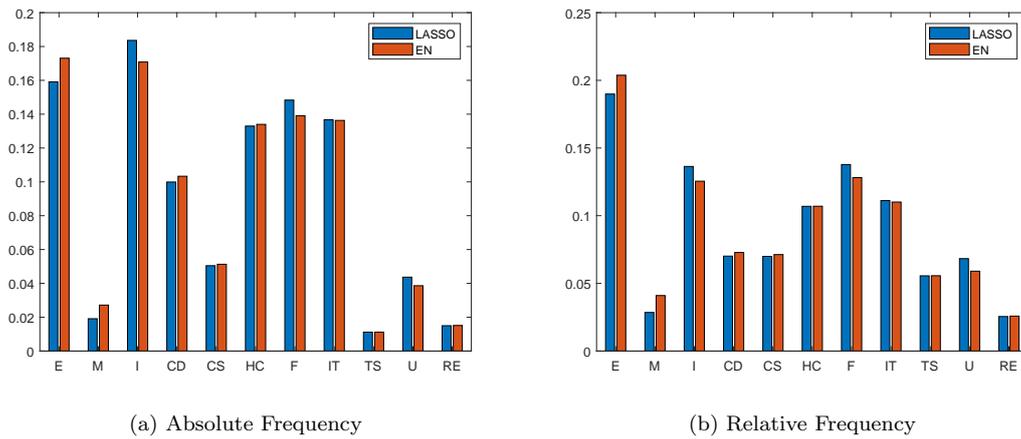
Figure 3: Variable Selection Results, by Sector*

*Notes: The distributions of selected individual stocks are plotted, by sector. Figure 3a reports average selection frequencies for select stocks. More specifically, for each rolling window, we calculate the ratio of the number of selected stocks in each sector to the total number of selected stocks, then take the average over all rolling windows. Figure 3b reports relative frequencies, which are calculated rescaling the average frequencies in Figure 3a by the size of each sector, as given in Figure 1.
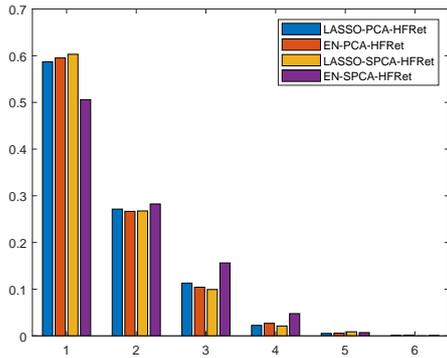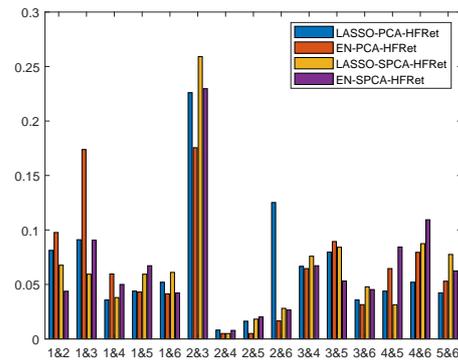
(a) Number of Factors
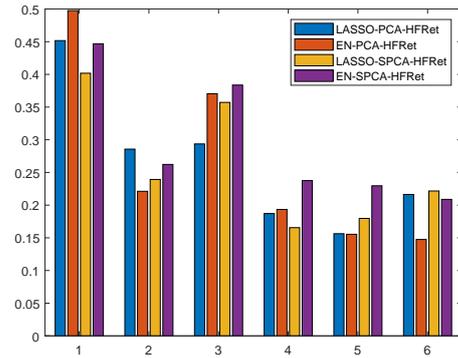
(b) Unconditional Frequency of Each Factor
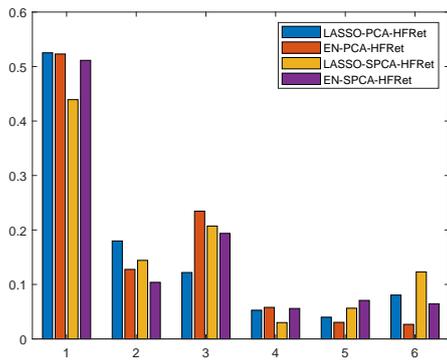
(c) Conditional Frequency (One Factor)

(d) Conditional Frequency (Two Factors)

Figure 4: Latent Common Volatility Factors*

*Notes: Figure 4a plots the average number of factors used for different specifications of our proposed two-step shrinkage method (i.e., the 4 models in Group VII in Table 1); Figure 4b plots the frequencies at which each factor is used, across all prediction periods; and Figures 4c and 4d report these frequencies, conditional on one factor and two factors being specified in the forecasting models, respectively. For further details, refer to Section 4.
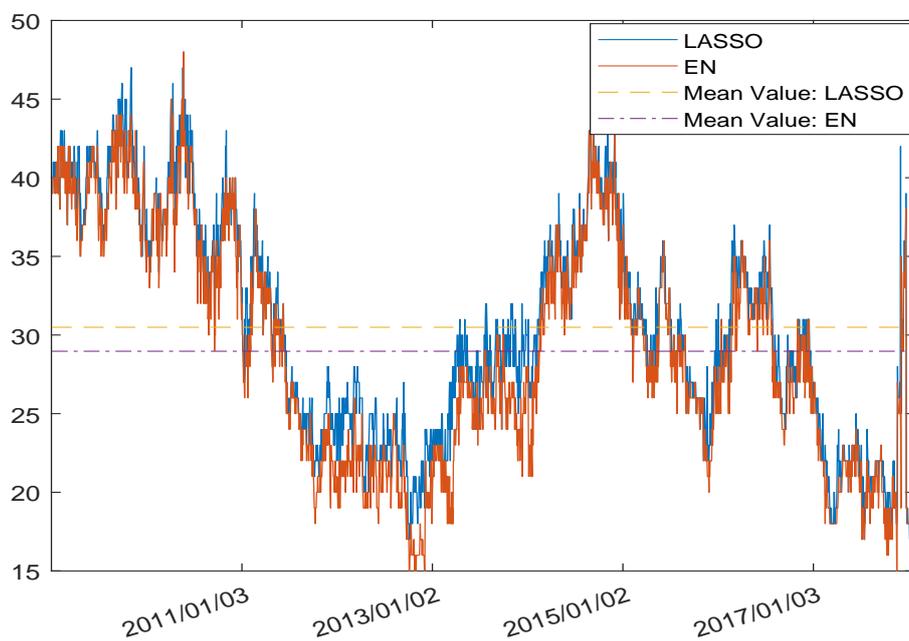
Figure 5: Number of Selected Stocks - Log-IV Estimates*

*Notes: As in Figure 2, the number of individual stocks selected using the LASSO and EN for each rolling data window used in forecast model specification is plotted, over the prediction period. However, in these plots, the prediction experiments utilize log-scale rather than original scale data. See Section 6.3.1 for further details.