*Journal of Financial Econometrics*, 2022, Vol. 20, No. 1, 160–186
doi: 10.1093/jjfinec/nbaa005
Advance Access Publication Date: 23 June 2020
Article

OXFORD



# Forecasting Equity Index Volatility by Measuring the Linkage among Component Stocks*

**Yue Qiu[1], Tian Xie[2], Jun Yu[3] and Qiankun Zhou[4]**

[1]Shanghai University of International Business and Economics, [2]Shanghai University of Finance and Economics, [3]Singapore Management University and [4]Louisiana State University

Address correspondence to Tian Xie, College of Business, Shanghai University of Finance and Economics, Shanghai, China, or e-mail: xietian001@hotmail.com

*We would like to thank the editor Silvia Gonçalves, the associated editor, and two referees for their valuable suggestions. Xie's research is supported by the Natural Science Foundation of China (71701175), the Chinese Ministry of Education Project of Humanities and Social Sciences (17YJC790174), and the Fundamental Research Funds for the Central Universities. Yu acknowledges financial support from the Lee Foundation. The usual caveat applies.

Received August 5, 2019; revised January 13, 2020; editorial decision March 25, 2020; accepted March 27, 2020

## Abstract

The linkage among the realized volatilities of component stocks is important when modeling and forecasting the relevant index volatility. In this article, the linkage is measured via an extended Common Correlated Effects (CCEs) approach under a panel heterogeneous autoregression model where unobserved common factors in errors are assumed. Consistency of the CCE estimator is obtained. The common factors are extracted using the principal component analysis. Empirical studies show that realized volatility models exploiting the linkage effects lead to significantly better out-of-sample forecast performance, for example, an up to 32% increase in the pseudo $R^2$. We also conduct various forecasting exercises on the linkage variables that compare conventional regression methods with popular machine learning techniques.

**Key words:** volatility forecasting, heterogeneous autoregression, common correlated effect, factor analysis, random forest

**JEL classification:** C31, C32, G12, G17

Volatility forecasting is central to financial institutions and market regulators. Portfolio managers tend to maximize returns when facing risk limits. With the development of realized variation based on high-frequency data, we are able to better measure financial market volatility. From then on, various volatility forecasting models have been put forward in the literature, such as the renowned fractionally integrated autoregressive moving average models used in Andersen et al. (2001) and the heterogeneous autoregressive (HAR) model proposed by Corsi (2009). No matter how complicated forms the above models can take, most of them rely on the asset-specific realized volatility histories. On the contrary, the comovement and spillover effect of risk across assets is well documented in the existing volatility literature and has been modeled by multivariate general autoregressive conditional heteroskedasticity (GARCH) and stochastic volatility (SV) models.[1]

This article exploits the linkages among the realized volatilities of component stocks to improve the corresponding stock index volatility forecasting. We propose a heterogeneous panel HAR (HARP) model assuming unobserved common factors to grasp the linkages. Our framework is based on the common correlated effect (CCE) estimator of Pesaran (2006) and allows us to extract unobserved common factors from the residuals of the econometric model. We demonstrate the consistency of the CCE estimator within our framework by applying Theorem 1 in Chudik and Pesaran (2015). Regarding the specification of cross-sectional unit regressors, we follow the realized semivariance models of Patton and Sheppard (2015). Another important step for our forecasting implementation is to model the dynamics of unobserved common factors. We conduct various forecasting exercises and compare regression methods with popular machine learning techniques.

We consider empirical applications to various equity indices, including the NASDAQ 100 exchange-traded fund (ETF), the Dow Jones Industrial Average (DJIA), the Dow Jones Transportation Average, and the Dow Jones Utility Average. Several novel findings are summarized as follows. First, the cross-sectional correlation of realized volatilities does exist for the considered component stocks. This finding holds irrespective of the underlying models.[2] Second, the in-sample results suggest that the role of unobserved common factors in explaining future volatility is nontrivial. They even carry partially the information contained in realized semivariances, especially the negative ones. Third and perhaps more importantly, we show that incorporating unobserved common factors into the HAR-type regressions leads to large and significant improvements in forecast accuracy.

It should be noted that there are other empirical applications of the CCE estimator in the literature. For example, the CCE approach allows Kapetanios and Pesaran (2004) to estimate asset return equations with both observed and unobserved common factors. Bernoth and Pick (2011) utilize the CCE framework to model the linkages between bank and insurance companies so as to improve forecasting the systemic risk. Chudik et al. (2017) develop tests for debt threshold effects in the context of dynamic heterogeneous panel data models, where the CCE estimator produces unbiased and consistent coefficient estimates for threshold variables. However, we are not aware of any application of the CCE estimator to the problem of volatility forecasting.

1 See Bauwens, Laurent, and Rombouts (2006) and Asai, McAleer, and Yu (2006) for more detailed reviews and discussions.

2 After some normalization of daily realized volatilities for a wide range of asset classes, Bollerslev et al. (2018) reach a conclusion closely related to ours. Their "normalized risk measures" exhibit almost identical unconditional distributions and similar highly persistent autocorrelation functions when comparing across assets and asset classes.

This article makes various contributions in several strands of literature. The first contribution is to extend the HAR model to a heterogeneous panel data model to exploit the linkages among the realized volatilities of component stocks. This idea is related to the volatility spillover literature under GARCH or SV framework (see, e.g., Engle, Ito, and Lin, 1990; King, Sentana, and Wadhwani, 1994), as we all exploit volatility information from other time series through some exogenous variables or presumptive common factors.[3] While the factor GARCH and SV models usually attribute the volatility comovement to asset common return factors and hence the consequent factor volatility, our formulation directly estimates common factors driving a panel of realized variances (RVs) by the CCE estimator. Moreover, our design explicitly deals with endogeneity concerns on model parameter estimates possibly caused by latent factors.

Our second contribution is to introduce the CCE approach to estimation of the panel HAR model and establish consistency of the CCE estimator. Third, we add to the literature on panels of realized volatilities.[4] Among this line of work, the nearest to ours are Bollerslev et al. (2018) and Cheng, Swanson, and Yang (2019), both of which augment the standard HAR regression with the factors extracted from panels of realized volatilities or returns.[5] In contrast to their setups, our linkage factors are extracted from the residuals of the CCE framework, a fairly dissimilar way of presuming the role of the common factors. We show that our model that exclusively exploits the linkage effects outside the asset's own histories can lead to increase in the out-of-sample $R^2$ by a wide margin, relative to volatility forecasting models without accommodating the linkage effects.

In the next section, we review a list of reference HAR-type models. Section 2 discusses the econometric approach and the forecasting procedure. Section 3 describes the data, which are analyzed in Section 4. Section 5 conducts some robustness checks. Finally, Section 6 concludes. An online supplement contains additional empirical and theoretical results.

3 The factor ARCH models were proposed by Diebold and Nerlove (1989) and Engle, Ng, and Rothschild (1990), and have been extended by Ng, Engle, and Rothschild (1992) and Bollerslev and Engle (1993) to model common persistence in conditional variances and covariances. The multivariate SV factor models were developed successively to become a more flexible alternative to GARCH-type models (see, e.g., King, Sentana, and Wadhwani, 1994; Pitt and Shephard, 1999). The origins of multivariate stochastic factor models were discussed in Shephard (2004), along with an intuitive argument of their basic features.

4 Early works of this literature are represented by Anderson and Vahid (2007), Gourieroux, Jasiak, and Sufana (2009), Bauer and Vorkink (2011), Bauwens, Storti, and Violante (2012), Hautsch, Kyj, and Oomen (2012), Halbleib and Voev (2014), and Asai and McAleer (2015). However, their feasibility to panel data of vast dimensions is not clear.

5 The above two papers treat volatility common factors in fairly differentiated ways. The framework of Cheng, Swanson, and Yang (2019) assumes unobservable common price factors governing panels of asset returns, which can be further used to generate volatility factors. Their factor estimation is based on a two-step shrinkage procedure in order to select the subset of assets, informative for extracting factors. Bollerslev et al. (2018) regard their common factors as the global risk factors influencing across asset classes, which are approximated by the average normalized realized volatilities over all asset classes.

## 1 A Review of Reference Models

Before moving into the panel-based HAR model, it is useful to review some HAR-type reference models, which act either as our basic specification for the CCE estimator or as comparison models in the subsequent exercise. Following Andersen and Bollerslev (1998), the $M$-sample daily RV at day $t$ can be calculated by summing the corresponding $M$ equally spaced intra-daily squared returns $r_{t,j}$. Here, the subscript $t$ indexes the day, and $j$ indicates the time interval within day $t$,

$$\mathrm{RV}_t \equiv \sum_{j=1}^{M} r_{t,j}^2, \quad \text{for } t = 1, 2, \ldots, T,\ j = 1, 2, \ldots, M, \tag{1}$$

where $r_{t,j} = p_{t,j} - p_{t,j-1}$ with $p_{t,j}$ being the log-price at time $(t, j)$.

To model realized variation, a series of HAR-type models were invented in the literature, see Corsi, Audrino, and Renò (2012) for a survey. The basic HAR model was introduced by Corsi (2009) and has gained great popularity because of its estimation simplicity and outstanding out-of-sample performance. The basic HAR model in Corsi (2009) postulates that the $h$-step-ahead daily $\mathrm{RV}_{t+h}$ can be modeled by

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_d \mathrm{RV}_t^{(1)} + \beta_w \mathrm{RV}_t^{(5)} + \beta_m \mathrm{RV}_t^{(22)} + \varsigma_{t+h}, \tag{2}$$

where the explanatory variables can take the general form of $\mathrm{RV}_t^{(l)}$. It is defined by

$$\mathrm{RV}_t^{(l)} \equiv l^{-1} \sum_{s=1}^{l} \mathrm{RV}_{t-s} \tag{3}$$

as the $l$ period averages of daily RV, the $\beta$s are the coefficients, and $\{\varsigma_t\}_t$ is the error term. Since each $\mathrm{RV}_t^{(l)}$ can be regarded as a volatility cascade, generated by the actions of distinct types of market participants trading at daily, weekly, or monthly frequencies (Müller et al., 1993), the lag structure in the HAR model is fixed at some lag index vector $\boldsymbol{l} = [1, 5, 22]$.

Andersen, Bollerslev, and Diebold (2007) extend the standard HAR model from two perspectives. First, they added the daily jump component $\mathrm{J}_t$ to Equation (2) to explicitly capture its impacts. The extended model is denoted as the HAR-J model,

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_d \mathrm{RV}_t^{(1)} + \beta_w \mathrm{RV}_t^{(5)} + \beta_m \mathrm{RV}_t^{(22)} + \beta^j \mathrm{J}_t + \varsigma_{t+h}, \tag{4}$$

where the empirical measurement of the squared jumps is $\mathrm{J}_t = \max(\mathrm{RV}_t - \mathrm{BPV}_t, 0)$, and the realized bipower variation (BPV) is defined as $\mathrm{BPV}_t \equiv (2/\pi)^{-1} \sum_{j=2}^{M} |r_{t,j-1}||r_{t,j}|$. Second, through a decomposition of RV by the $Z_{1,t}$ statistic in Huang and Tauchen (2005) into the continuous sample path and the jump components, they extend the HAR-J model by explicitly incorporating the above two types of volatility components. The $Z_{1,t}$ statistic distinguishes the "significant" jumps $\mathrm{CJ}_t$ from continuous sample path components $\mathrm{CSP}_t$:

$$\mathrm{CSP}_t \equiv \mathbb{I}(Z_t \le \Phi_\alpha) \cdot \mathrm{RV}_t + \mathbb{I}(Z_t \le \Phi_\alpha) \cdot \mathrm{BPV}_t,$$
$$\mathrm{CJ}_t \equiv \mathbb{I}(Z_t > \Phi_\alpha) \cdot \max(\mathrm{RV}_t - \mathrm{BPV}_t, 0),$$

where $Z_t$ is the ratio statistic defined in Huang and Tauchen (2005), and $\Phi_\alpha$ is the cumulative distribution function of a standard Gaussian distribution with $\alpha$ level of significance. The daily, weekly, and monthly average components of $\mathrm{CSP}_t$ and $\mathrm{CJ}_t$ are then constructed in the same manner as $\mathrm{RV}^{(l)}$ in Equation (3). The model specification for the continuous HAR-J, namely, HAR-CJ, is given by

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_d^c \mathrm{CSP}_t^{(1)} + \beta_w^c \mathrm{CSP}_t^{(5)} + \beta_m^c \mathrm{CSP}_t^{(22)} + \beta_d^j\ \mathrm{CJ}_t^{(1)} + \beta_w^j \mathrm{CJ}_t^{(5)} + \beta_m^j \mathrm{CJ}_t^{(22)} + \varsigma_{t+h}. \tag{5}$$

Note that compared with the HAR-J model, the HAR-CJ model explicitly controls for the weekly and monthly components of continuous jumps. Thus, the HAR-J model can be treated as a special and restrictive case of the HAR-CJ model for $\beta_d = \beta_d^c + \beta_d^j$, $\beta^j = \beta_d^j$, $\beta_w = \beta_w^c + \beta_w^j$, and $\beta_m = \beta_m^c + \beta_m^j$.

To capture the information from signed high-frequency variation, Patton and Sheppard (2015) developed a series of realized semivariance HAR (HAR-RS) models. The first one, HAR-RS-I model, completely decomposes the $\mathrm{RV}^{(1)}$ in Equation (2) into two asymmetric semivariances, $\mathrm{RS}_t^+$ and $\mathrm{RS}_t^-$,

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_d^+ \mathrm{RS}_t^+ + \beta_d^- \mathrm{RS}_t^- + \beta_w \mathrm{RV}_t^{(5)} + \beta_m \mathrm{RV}_t^{(22)} + \varsigma_{t+h}, \tag{6}$$

where $\mathrm{RS}_t^- = \sum_{j=1}^M r_{t,j}^2 \cdot \mathbb{I}\big(r_{t,j} < 0\big)$ and $\mathrm{RS}_t^+ = \sum_{j=1}^M r_{t,j}^2 \cdot \mathbb{I}\big(r_{t,j} > 0\big)$. To verify the actual effects of signed variations, they include an additional term capturing the leverage effect, $\mathrm{RV}_t^{(1)} \cdot \mathbb{I}(r_t < 0)$. The second model in Equation (7) is denoted as HAR-RS-II,

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_1 \mathrm{RV}_t^{(1)} \cdot \mathbb{I}(r_t < 0) + \beta_d^+ \mathrm{RS}_t^+ + \beta_d^- \mathrm{RS}_t^- + \beta_w \mathrm{RV}_t^{(5)} + \beta_m \mathrm{RV}_t^{(22)} + \varsigma_{t+h}. \tag{7}$$

The third and fourth models in Patton and Sheppard (2015), denoted as HAR-SJ-I (Equation 8) and HAR-SJ-II (Equation 9), respectively, examine the role that decomposing RVs into signed jump variations and BPV can play in forecasting volatility:

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_d^j \mathrm{SJ}_t + \beta_d^{bpv} \mathrm{BPV}_t + \beta_w \mathrm{RV}_t^{(5)} + \beta_m \mathrm{RV}_t^{(22)} + \varsigma_{t+h}, \tag{8}$$

$$\mathrm{RV}_{t+h} = \beta_0 + \beta_d^{j-} \mathrm{SJ}_t^- + \beta_d^{j+} \mathrm{SJ}_t^+ + \beta_d^{bpv} \mathrm{BPV}_t + \beta_w \mathrm{RV}_t^{(5)} + \beta_m \mathrm{RV}_t^{(22)} + \varsigma_{t+h}, \tag{9}$$

where $\mathrm{SJ}_t = \mathrm{RS}_t^+ - \mathrm{RS}_t^-$, $\mathrm{SJ}_t^+ = \mathrm{SJ}_t \cdot \mathbb{I}(\mathrm{SJ}_t > 0)$, and $\mathrm{SJ}_t^- = \mathrm{SJ}_t \cdot \mathbb{I}(\mathrm{SJ}_t < 0)$. The HAR-SJ-II model further extends the HAR-SJ-I model by distinguishing the effect of a positive jump variation from that of a negative jump variation.

Since $\mathrm{RV}_{t+h}$ is unobservable at time $t$, the models mentioned above must take $h$ periods of lags in order to estimate the coefficients in practice.

## 2 The Panel HAR Model

To control for possibly unobserved common effects in volatility across a class of assets, we construct a HARP model with error cross-sectional dependence, which is an extension of the framework of Chudik and Pesaran (2015). Let $y_{it}$ be the RV of the $i$-th individual asset at time $t$ for $i = 1, \ldots, N$; $t = 1, \ldots, T$. Suppose that $y_{it}$ can be described by the following heterogeneous dynamic panel data model as the $h$-period direct forecasting model,

$$y_{it} = \boldsymbol{\alpha}_{i,h}^\top \boldsymbol{d}_{t-h} + \sum_{l \in \mathcal{L}} \phi_{i,h}^{(l)} \bar{y}_{i,t-h}^{(l)} + \boldsymbol{\beta}_{i,h}^\top \boldsymbol{x}_{i,t-h} + u_{it}, \tag{10}$$

$$\bar{y}_{it-h}^{(l)} = l^{-1} \sum_{s=0}^{l-1} y_{i,t-h-s}, \tag{11}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where $\boldsymbol{d}_{t-h}$ is a $r \times 1$ vector of observed common effects, including deterministic such as intercepts or seasonal dummies, $\boldsymbol{x}_{i,t-h}$ is a $k \times 1$ vector of regressors specific to cross-sectional unit $i$ at time $t - h$, and $\boldsymbol{\alpha}_{i,h}$ and $\boldsymbol{\beta}_{i,h}$ are parameter vectors specific for the $h$-period forecasting model. $y_{i,t-h-s}$ represents the $s$-th lag of $y_{i,t-h}$, and $\bar{y}_{i,t-h}^{(l)}$ is the HAR component, which is the average of previous $l$ periods of $y_{i,t-h}$. $\phi_{i,h}^{(l)}$ is the coefficient for $\bar{y}_{i,t-h}^{(l)}$, $\mathcal{L}$ is the lag index vector of $l$ and we let $L = \max(\mathcal{L})$.[6] Furthermore, we assume that the RV of individual stocks is correlated beyond what can be explained by the observed determinants because the error term, $u_{it}$, comprises $m$ unobserved common factors,[7]

$$u_{it} = \gamma_{i,h}^{\top} \boldsymbol{f}_t + \epsilon_{it}, \tag{12}$$

where $\gamma_{i,h}$ is the $m \times 1$ vector of factor loadings, $\boldsymbol{f}_t$ is the $m \times 1$ vector of unobserved common factors that could themselves be serially correlated, and $\epsilon_{it}$ are the idiosyncratic errors assumed to be independently distributed of $(\boldsymbol{d}_{t-h}, \boldsymbol{x}_{i,t-h})$ and uncorrelated with the factors. Assume that $\boldsymbol{f}_t$ can be modeled by a VAR or by a more general relationship,[8]

$$\boldsymbol{f}_t = \Phi_{f,h} \boldsymbol{f}_{t-h} + \zeta_t. \tag{13}$$

Following Pesaran (2006) and Chudik and Pesaran (2015), the unobserved factors, $\boldsymbol{f}_t$, can be also correlated with $(\bar{y}_{i,t-h}^{(l)}, \boldsymbol{d}_{t-h}, \boldsymbol{x}_{i,t-h})$. To permit such a possibility, we assume a fairly general model for individual-specific regressors, $\boldsymbol{x}_{i,t-h}$,

$$\boldsymbol{x}_{i,t-h} = \Lambda_{i,h}^{\top} \boldsymbol{d}_{t-h} + \boldsymbol{\Pi}_{i,h} \boldsymbol{y}_{it-h,-L+1} + \boldsymbol{\Gamma}_{i,h}^{\top} \boldsymbol{f}_{t-h} + \boldsymbol{v}_{i,t-h}, \tag{14}$$

where $\boldsymbol{y}_{it-h,-L+1} = (y_{i,t-h}, \ldots, y_{i,t-h-L+1})^{\top}$, $\boldsymbol{\Lambda}_{i,h}$ and $\boldsymbol{\Gamma}_{i,h}$ are $r \times k$ and $m \times k$ matrix of factor loadings for observed and unobserved factors, respectively, $\boldsymbol{\Pi}_{i,h}$ is a $k \times L$ matrix of unknown coefficients, and $\boldsymbol{v}_{i,t-h}$ is assumed to follow a general linear covariance stationary process distributed independently of $\epsilon_{it}$. Equations (10)–(14) hitherto set out our panel HAR volatility forecasting model, named as the HARP model.[9]

## 2.1 The CCE Estimator

Clearly, forecasts of $y_{it}$ need consistent estimates of the parameters and unobserved common factors. Unfortunately, conventional panel estimators of Equation (10) yield inconsistent estimates of coefficients due to the correlation of regressors $(\boldsymbol{x}_{i,t-h}, \bar{y}_{i,t-h}^{(l)})$ and error terms $u_{it}$. To see this, since $\boldsymbol{f}_t$ is assumed to be serially correlated, the error term in Equation (10), $u_{it}$, is thus serially correlated through Equation (12), which further entails

6 Following the convention of the HAR-RV literature, we set $L = 22$ in our empirical exercise. In this literature, it is common to set $\mathcal{L} = [1, 5, 22]$ to claim that tomorrow's RV can be a sum of daily, weekly, and monthly averages of past RVs.

7 The recent paper by Bollerslev et al. (2018) interprets the common factors as combined economic forces from the investor sentiment, the variance risk premium and the news surprise variable.

8 The benchmark scheme of predicting $\hat{f}_{t+h}$ and its alternatives are further elaborated in Section 4.1.

9 Note that before we pin down the underlying specification of $x_{it}$, the HARP model can be quite flexible to accommodate other HAR-type specifications in the literature.

the correlation of $u_{it}$ and $y_{i,t-h-s}$. Since the HAR components $\bar{y}^{(l)}_{i,t-h}$ are linear functions of $y_{i,t-h-s}$, they are correlated with $u_{it}$. It is more apparent to witness the correlation between $\boldsymbol{x}_{i,t-h}$ and $u_{it}$, regarding the fact that they both contain serially correlated factors $\boldsymbol{f}_t$.

In this section, we address the issue of inconsistency and demonstrate how to estimate the slope coefficients $(\phi^{(1)}_{i,h}, \ldots, \phi^{(L)}_{i,h}, \boldsymbol{\beta}_{i,h})$ from Equation (10), employing the CCE estimator proposed by Pesaran (2006). We note that Equations (10) and (12) can be combined, and rewritten as

$$\begin{aligned} y_{it} &= \boldsymbol{\alpha}_{i,h}^\top \boldsymbol{d}_{t-h} + \sum_{l \in \mathcal{L}} \phi^{(l)}_{i,h} \bar{y}^{(l)}_{i,t-h} + \boldsymbol{\beta}_{i,h}^\top \boldsymbol{x}_{i,t-h} + \boldsymbol{\gamma}_{i,h}^\top \boldsymbol{f}_t + \epsilon_{it} \\ &= \boldsymbol{\alpha}_{i,h}^\top \boldsymbol{d}_{t-h} + \sum_{l=0}^{L-1} \psi_{ih,l} y_{i,t-h-l} + \boldsymbol{\beta}_{i,h}^\top \boldsymbol{x}_{i,t-h} + \boldsymbol{\gamma}_{i,h}^\top \boldsymbol{f}_t + \epsilon_{it} \\ &= \boldsymbol{\alpha}_{i,h}^\top \boldsymbol{d}_{t-h} + \boldsymbol{\psi}_{i,h}^\top \boldsymbol{y}_{it-h,-L+1} + \boldsymbol{\beta}_{i,h}^\top \boldsymbol{x}_{i,t-h} + \boldsymbol{\gamma}_{i,h}^\top \boldsymbol{f}_t + \epsilon_{it}, \end{aligned} \tag{15}$$

where $\boldsymbol{y}_{it-h,-L+1} = (y_{i,t-h}, \ldots, y_{i,t-h-L+1})^\top$ and $\boldsymbol{\psi}_{i,h} = \left(\psi_{ih,0}, \ldots, \psi_{ih,L-1}\right)^\top$ are given by

$$\begin{aligned} \psi_{ih,0} &= \phi^{(1)}_{i,h} + \frac{\phi^{(2)}_{i,h}}{2} + \cdots + \frac{\phi^{(L)}_{i,h}}{L}, \quad \psi_{ih,1} = \frac{\phi^{(2)}_{i,h}}{2} + \cdots + \frac{\phi^{(L)}_{i,h}}{L}, \\ &\vdots \\ \psi_{ih,L-2} &= \frac{\phi^{(L-1)}_{i,h}}{L-1} + \frac{\phi^{(L)}_{i,h}}{L}, \quad \psi_{ih,L-1} = \frac{\phi^{(L)}_{i,h}}{L}. \end{aligned}$$

As a result, we note that Equation (15) can be viewed as a panel restricted AR($L$) model with error cross-sectional dependence.

The parameters of interest in Equation (15) are $\boldsymbol{\psi}_i$ and $\boldsymbol{\beta}_i$ while $\boldsymbol{f}_t$ is unobserved. To estimate $\boldsymbol{\psi}_i$ and $\boldsymbol{\beta}_i$ in model (15), we can adopt the CCE) estimator proposed by Pesaran (2006), which is further extended by Chudik and Pesaran (2015) to the dynamic setting. To this end, let $z_{it} = (y_{it}, \boldsymbol{x}_{i,t-h}^\top)^\top$ and define the cross-sectional average of $z_{it}$ at period $t$ as $\bar{z}_t$.

The parameter of interests are $\boldsymbol{\theta}_{i,h} = (\boldsymbol{\psi}_{i,h}^\top, \boldsymbol{\beta}_{i,h}^\top)^\top$. Define

$$\Xi_i = \begin{pmatrix} \boldsymbol{y}_{ip_T,-L+1}^\top & \boldsymbol{x}_{i,p_T}^\top \\ \boldsymbol{y}_{ip_T+1,-L+1}^\top & \boldsymbol{x}_{i,p_T+1}^\top \\ \vdots & \vdots \\ \boldsymbol{y}_{iT-h,-L+1}^\top & \boldsymbol{x}_{i,T-h}^\top \end{pmatrix}, \bar{\mathrm{Q}} = \begin{pmatrix} \boldsymbol{d}_{p_T}^\top & \bar{\mathrm{z}}_{p_T+1}^\top & \bar{\mathrm{z}}_{p_T}^\top & \cdots & \bar{\mathrm{z}}_1^\top \\ \boldsymbol{d}_{p_T+1}^\top & \bar{\mathrm{z}}_{p_T+2}^\top & \bar{\mathrm{z}}_{p_T+1}^\top & \cdots & \bar{\mathrm{z}}_2^\top \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{d}_{T-h}^\top & \bar{\mathrm{z}}_{T-h+1}^\top & \bar{\mathrm{z}}_{T-h}^\top & \cdots & \bar{\mathrm{z}}_{T-h+1-p_T}^\top \end{pmatrix}, \tag{16}$$

where $p_T > L$ is a predetermined value. Let the projection matrix $\boldsymbol{M}_{\bar{\mathrm{Q}}} = \boldsymbol{I}_{T-p_T} - \bar{\mathrm{Q}}(\bar{\mathrm{Q}}^\top\bar{\mathrm{Q}})^+\bar{\mathrm{Q}}^\top$, where $\boldsymbol{I}_{T-p_T}$ is a $(T-p_T)\times(T-p_T)$ dimensional identity matrix, and $\boldsymbol{A}^+$ represents the Moore–Penrose generalized inverse of $\boldsymbol{A}$. The CCE estimator of $\boldsymbol{\theta}_{i,h}$ is given by

$$\hat{\theta}_{ih,CCE} = \left(\Xi_i^\top \boldsymbol{M}_{\bar{\mathrm{Q}}} \Xi_i\right)^{-1} \Xi_i^\top \boldsymbol{M}_{\bar{\mathrm{Q}}} \boldsymbol{y}_i, \tag{17}$$

where $\boldsymbol{y}_i = (y_{i,p_T+h}, y_{i,p_T+h+1}, \ldots, y_{i,T})^\top$.

Theorem 1 in Chudik and Pesaran (2015) shows the consistency of CCE estimator for the AR framework under certain assumptions. Since the HAR model can be regarded as a restricted AR model, the consistency of CCE estimator for the HARP model can be easily established by applying Theorem 1 in Chudik and Pesaran (2015). We elaborate this application in detail along with necessary assumptions in Online Appendix A. The Monte Carlo experiments are also conducted to investigate the general performance of the CCE estimator under the HARP framework in Online Appendix B.

## 2.2 Forecasting Realized Volatility

We are interested in the forecast of $y_{i,T+h}$ conditional on the information up to time $T$. Equation (15) may be rewritten for $y_{i,t+h}$ as the following $h$-period direct forecasting model[10]

$$y_{i,t+h} = \boldsymbol{\alpha}_{i,h}^{\top}\boldsymbol{d}_t + \boldsymbol{\psi}_{i,h}^{\top}\boldsymbol{y}_{it,-L+1} + \boldsymbol{\beta}_{i,h}^{\top}\boldsymbol{x}_{it} + \gamma_{i,h}^{\top}\boldsymbol{f}_{t+h} + \epsilon_{it+h}. \tag{18}$$

The forecast of $y_{i,T+h}$ contingent on the information up to time $T$ is therefore

$$\hat{y}_{i,T+h|T} = \hat{\boldsymbol{\alpha}}_{i,h}^{\top}\boldsymbol{d}_T + \hat{\boldsymbol{\psi}}_{i,h}^{\top}\boldsymbol{y}_{iT,-L+1} + \hat{\boldsymbol{\beta}}_{i,h}^{\top}\boldsymbol{x}_{iT} + \hat{\gamma}_{i,h}^{\top}\hat{\boldsymbol{f}}_{T+h|T}, \tag{19}$$

where $\hat{\boldsymbol{f}}_{T+h|T}$ is a forecast of $\boldsymbol{f}_{T+h}$. As argued above in Section 2.1, any conventional panel estimator of $\hat{\boldsymbol{\alpha}}_{i,h}$, $\hat{\boldsymbol{\psi}}_{i,h}^{\top}$, $\hat{\boldsymbol{\beta}}_{i,h}$ without the control of $\boldsymbol{f}_t$ is inconsistent.

Hence, to obtain the forecasts $\hat{y}_{i,T+h|T}$ from Equation (19), it requires consistent estimation of the corresponding parameters and forecasts of $\boldsymbol{f}_{t+h}$. We solve the above forecasting problem by employing a two-stage process: first, we obtain the initial consistent estimates of $(\boldsymbol{\psi}_{i,h}^{\top}, \boldsymbol{\beta}_{i,h}^{\top})^{\top}$ and unobserved common factors $\boldsymbol{f}_t$, utilizing the CCE estimator and principal components (PCs). The factor estimates from this procedure are consistent as well (Pesaran, 2006). For instance, given the consistent estimation of $\boldsymbol{\psi}_{i,h}$ and $\boldsymbol{\beta}_{i,h}$, we can obtain $g_{it} = \boldsymbol{\alpha}_{i,h}^{\top}\boldsymbol{d}_{t-h} + u_{it}$, as

$$\hat{g}_{it} = y_{it} - \hat{\boldsymbol{\psi}}_{i,h}^{\top}\boldsymbol{y}_{it-h,-L+1} - \hat{\beta}_{i,h}^{\top}\boldsymbol{x}_{it-h}. \tag{20}$$

After acquiring the residuals, $\hat{g}_{it}$, an estimate of $u_{it}$ is produced by integrating out the common observed factors, $\boldsymbol{d}_{t-h}$,

$$\hat{\mathrm{u}}_{\boldsymbol{i}} = \boldsymbol{M}_D\hat{\mathrm{g}}_{\boldsymbol{i}}, \tag{21}$$

where $\hat{\mathrm{u}}_i = \left(\hat{u}_{i,h+1}, \hat{u}_{i,h+2}, \ldots, \hat{u}_{i,T}\right)$ and $\boldsymbol{D} = \left(\boldsymbol{d}_1{}^{\top}, \boldsymbol{d}_2{}^{\top}, \ldots, \boldsymbol{d}_{\mathrm{T-h}}{}^{\top}\right)$. Bernoth and Pick (2011) pointed out that the orthogonality assumption of $\boldsymbol{d}_t$ to $\boldsymbol{f}_t$ is necessary to guarantee unbiasedness of the parameter estimates of the common factors, $\hat{\boldsymbol{\alpha}}_i$, although an absence of it will not bias the forecasts resulting from the above procedure.

With estimates of the residuals $\hat{u}_{it}$, we are able to extract the unobserved common factors, $\boldsymbol{f}_t$, from residuals of the $h$-step direct projection using the PC analysis. In our application, consistent estimation of $\boldsymbol{\psi}_{i,h}$ and $\boldsymbol{\beta}_{i,h}$ is guaranteed under any fixed number of unobserved factors, $m$.[11] However, since $\boldsymbol{f}_t$ and its factor loadings still matter for the forecasts by Equation (18), an estimate of $m$ seems essential. To resolve this, we apply the Bai and Ng's (2002) method to the residuals, $\hat{\mathrm{u}}_i$ given in Equation (21), which yields a choice of $m = 2$.[12]

10 Assuming certain vector autoregressive processes for $d_t$ and $f_t$, we can iterate on Equations (10) and (12), and then conduct recursive substitutions to yield the direct forecasting model. Bernoth and Pick (2009) provide a detailed mathematical induction on a similar question.

11 As suggested by Pesaran (2006), the number of unobserved factors, $m$, only becomes a practical issue if the focus of the analysis is on the factor loadings, for instance, the parameters of asset pricing factors.

12 We also test the forecasts for different values of $m$ and the results remain qualitatively similar.

In the second stage, the above estimates of $\boldsymbol{f}_t$ can then be plugged back to Equation (18) to estimate the parameters $\boldsymbol{\alpha}_{i,h}$, $\boldsymbol{\psi}_{i,h}$, $\boldsymbol{\beta}_{i,h}$ and $\boldsymbol{\gamma}_{i,h}$ by OLS. For the $h$-step prediction, we still produce direct forecasts of $\hat{\boldsymbol{f}}_{t+h}$ under certain hypothetical process of $\boldsymbol{f}_t$,[13] as the PC analysis only generates $\hat{\boldsymbol{f}}_t$ up to time $T$. It is noteworthy to mention that the two-stage procedure is iterated for each $h$ separately.

### 2.3 Assessment of the Forecasts

Forecast performance of $y_{it}$ is evaluated using the following criteria:

$$\text{MAFE}(h) = \frac{1}{V}\sum_{j=1}^{V} |e_{iT_j,h}|, \tag{22}$$

$$\text{MSFE}(h) = \frac{1}{V}\sum_{j=1}^{V} e^2_{iT_j,h}, \tag{23}$$

$$\text{SDFE}(h) = \sqrt{\frac{1}{V-1}\left(e_{iT_j,h} - \frac{1}{V}\sum_{j=1}^{V} e_{iT_j,h}\right)^2}, \tag{24}$$

where $e_{iT_j,h} = y_{iT_j,h} - \hat{y}_{iT_j,h}$ is the forecast error, $j = 1, 2, \ldots, V$, and $\hat{y}_{iT_j,h}$ is the $h$-day ahead forecast with information up to $T_j$, where $T_j$ stands for the last observation in each of the $V$ rolling windows. Another widely adopted method for evaluation is by means of the $R^2$-criterion of the Mincer–Zarnowitz regression,[14] given by

$$y_{iT_j,h} = a + b\hat{y}_{iT_j,h} + u_{T_j}, \text{for } j = 1, 2, \ldots, V, \tag{25}$$

Note that we choose the level-regression (25) over the log-regression, because Hansen and Lunde (2006) have argued that the $R^2$ from a regression of $\log(y_{iT_j,h})$ on a constant and $\log(\hat{y}_{iT_j,h})$ is unlikely to induce the same ranking of volatility models as the $R^2$ from the infeasible regressions (with the true volatility), unless a proportionate relationship exists between the estimated and true values of volatility.[15]

Based on the findings of Hansen and Lunde (2006) and Patton (2011), we also compute the expected values of Gaussian quasi-likelihood (QLIKE),

$$\text{QLIKE}(h) = \log \hat{y}_{iT_j,h} + \frac{y_{iT_j,h}}{\hat{y}_{iT_j,h}}, \text{for } j = 1, 2, \ldots, V, \tag{26}$$

The QLIKE function, along with the mean squared forecast error (MSFE) loss, has been demonstrated to be robust to noise in the proxy for volatility in Patton (2011). Moreover, Patton and Sheppard (2009) find that relative to the MSFE loss, the QLIKE loss has better power properties under the Diebold–Mariano test. In the last place, we complement the above results by running the unconditional Giacomini and White (2006) test for the mean absolute forecast error, in order to test the equal predictive ability of a pair of models.

13 Section 4.1 provides a detailed discussion on various forecasting schemes of $f_t$ and their implications.

14 Interested readers may refer to Mincer and Zarnowitz (1969) for more details.

15 Hansen and Lunde (2006) prove that if the proxy, $\tilde{\sigma}_t^2$, and the true volatility, $\sigma_t^2$, satisfy the equation $\tilde{\sigma}_t^2 = (1 - \upsilon_t)\sigma_t^2$ for some random variable, $\upsilon_t$, the ranking of volatility models remains unaffected by the measurement errors of $\tilde{\sigma}_t^2$.

## 3 Data Description

In the empirical exercise, we consider an application to the RV of the NASDAQ 100 Trust ETF tracking the NASDAQ 100 Index, with ticker symbol QQQ. To avoid the concerns of data mining, the information on unobserved common factors is extracted solely from a panel data of the NASDAQ 100 constituents.[16]

The data on the NASDAQ 100 ETF consists of high-frequency transaction prices from May 22, 2007 to October 20, 2017, which totals 2625 observations. The 104 constituents cover industry groups ranging from computer hardware and software, telecommunications, retail/wholesale trade, and biotechnology. Since the components of the NASDAQ 100 index are varying over time, we only include the stocks that always belong to the index during our sample period, in order to keep the panel balanced. In total, eighty-nine stocks are selected. A more detailed description of these stocks is given in Online Appendix D, including their ticker symbols, names, and Global Industry Classification sectors. The whole dataset and its relevant information are provided by Pitrading Inc., which base their source information from New York Stock Exchange's TAQ database. All the above data are obtained at one-minute increments. The intraday prices are then used to calculate daily RV measures by Equation (1).

To have an abundance of time series, we use the original data at one-minute intervals in our primary analysis. However, we are aware that a too high sampling frequency might cause microstructure bias to distort volatility estimates from its true daily variance. The previous work by Liu, Patton, and Sheppard (2015) offers some evidence on why five-minute RV is typically considered as the benchmark.[17] For the justification of using the one-minute data, we construct volatility signature plots for the NASDAQ 100 ETF (QQQ) and the eight representative stocks in Figure 1, to check if one-minute RV is an appropriate alternative to five-minute RV.[18] A description of the full company names and their weights is included in Table 1. The main pattern is that volatility signature plots for all the considered assets are flat, which is especially the case for QQQ. This indicates no apparent variations from one-minute sampling, relative to five-minute sampling, at least for our datasets.[19]

16 Our focus on the index volatility is reminiscent of the exercise in Bollersleva et al. (2019), where they exploit the information in the realized semicovariance matrix to improve the portfolio variance forecasts. In practice, the NASDAQ 100 Index ETF is usually regarded as one of the most important portfolios in the financial market.

17 Liu, Patton, and Sheppard (2015) conduct a comprehensive study for over 400 different realized measures, with a wide range of sampling frequencies, and they apply these to 11 years of daily data on thirty individual financial assets. Overall, they find it difficult to significantly outperform five-minute RV.

18 The volatility signature plot was first introduced by Andersen et al. (2000) to provide some guidance on the optimal sampling frequency.

19 Liu, Patton, and Sheppard (2015) conclude that as long as the assets are liquid enough, one-minute sampling is nevertheless sparse to avoid the problem of microstructure bias. For the robustness check, we also conduct a similar empirical analysis on five-minute sampling data in Section 5.2 and the HARP model still outperforms the risk models solely based on its own realized volatility components.

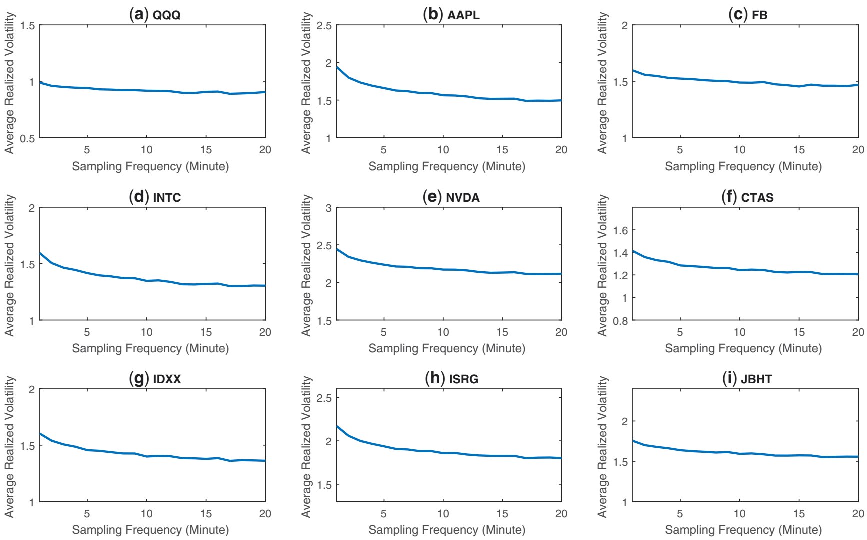


**Figure 1** Volatility signature plot for representative stocks and the NASDAQ 100 ETF. (**a–i**) Plot the average daily realized volatility (on the *y*-axis) for the NASDAQ 100 ETF, four highly liquid components (AAPL, FB, INTC, and NVDA) and four relatively illiquid components (CTAS, IDXX, ISRG, and JBHT) of the NASDAQ 100 Index over their corresponding sampling intervals (on the *x*-axis). The sampling intervals run from 1 minute to 20 minutes.

**Table 1** Descriptions of liquid and illiquid stocks

| Ticker | Company name | Weights in the NASDAQ 100 (%) |
|---|---|---|
| Liquid stocks | | |
| AAPL | Apple Inc. | 12.39 |
| FB | Facebook Inc. | 4.85 |
| INTC | Intel Corporation | 2.60 |
| NVDA | NVIDIA Corporation | 1.80 |
| Illiquid stocks | | |
| CTAS | Cintas Corporation | 0.27 |
| IDXX | IDEXX Laboratories Inc. | 0.25 |
| ISRG | Intuitive Surgical Inc. | 0.70 |
| JBHT | J.B. Hunt Transport Services Inc. | 0.16 |

We describe summary statistics of the RV series for the NASDAQ100 ETF in column 2 of Table 2. Due to the vast number of the NASDAQ 100 Index components, we only present the statistics of six representative stocks in columns 3–8 of Table 2. Table 2 documents the results of the sample mean, median, minimum, maximum, standard deviation, skewness, and kurtosis for the RV series over the full sample periods. Table 2 also reports the *p*-values[20] of the Jarque–Bera test for normality and those of the augmented Dickey–Fuller

20 In our exercises, we set the lower bound of the *p*-values of the Jarque–Bera and the ADF tests at 0.001. Values less than 0.001 are truncated at 0.001.

**Table 2** Summary statistics for the RV of the NASDAQ100 ETF and six representative stocks

| Statistic | QQQ | Tickers of representative stocks | | | | | |
|---|---|---|---|---|---|---|---|
| | | AAL | ALXN | DISCA | ISRG | QCOM | XLNX |
| Mean | 1.0536 | 17.0766 | 4.7820 | 3.5654 | 4.1525 | 2.3330 | 2.7815 |
| Median | 0.6119 | 7.2913 | 3.4835 | 2.0676 | 2.6538 | 1.4461 | 1.7512 |
| Maximum | 9.8061 | 669.3273 | 135.1100 | 583.1942 | 65.4187 | 34.2651 | 33.9595 |
| Minimum | 0.0687 | 0.5071 | 0.4013 | 0.4374 | 0.3030 | 0.1598 | 0.3030 |
| Standard deviation | 1.2467 | 30.9578 | 5.0603 | 13.5319 | 4.6312 | 2.7424 | 2.9660 |
| Skewness | 3.0783 | 7.8892 | 9.2852 | 35.4074 | 4.0143 | 4.2497 | 3.5105 |
| Kurtosis | 15.1628 | 117.0782 | 189.5792 | 1416.5222 | 29.7366 | 31.5784 | 21.4986 |
| Jarque–Bera | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| ADF | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |

*Notes:* The second column of Table 2 contains statistics for the RV of the NASDAQ100 ETF from May 22, 2007 to October 20, 2017, for a total of 2625 observations. The statistics of six component stocks of the NASDAQ 100 index are presented in columns 3–8. For JB and ADF test, statistics that are outside tabulated critical values, we report maximum (0.999) or minimum (0.001) $p$-values. All the statistics here are computed based on the data at one-minute increments.

(ADF) test for unit root. The null hypotheses of a normal distribution and a unit root are strongly rejected in all cases, whereas the other statistics disperse over a wide range.

### 3.1 Observed Common Factors

Note that all of the volatility models in our empirical exercise contain a set of observed macroeconomic factors $\boldsymbol{d}_t$. Inspired by recent research by Fernandes, Medeiros, and Scharth (2014), we include the following predictors in $\boldsymbol{d}_t$ contemporaneously: (i) the $j$-day continuously compounded return on the one-month crude oil futures contract for $j = 1, 5, 10, 22$, and 66 (oil $j$-day return); (ii) the first difference of the logarithm of the trade-weighted average of the foreign exchange value of the US dollar index against the Australian dollar, Canadian dollar, Swiss franc, euro, British pound sterling, Japanese yen, and Swedish krona (USD change); (iii) the excess yield of the Moody's seasoned Baa corporate bond over the Moody's seasoned AAA corporate bond (credit spread); (iv) the difference between the 10-year and 3-month Treasury constant maturity rates (term spread); and (v) the difference between the effective and target federal fund rates (FF deviation). While both oil prices (Oil) and term spread (TS) are concerned with various dimensions of the overall market conditions, USD change (USDI), and FF deviation (FED) are both linked to US macroeconomic states. Descriptive statistics for $\boldsymbol{d}_t$ are available in Online Appendix E.

## 4 Empirical Analysis

In this section, we conduct an empirical exercise to thoroughly examine both in-sample and out-of-sample performance of the HARP model. The result is in comparison with the performance of the autoregressive model (AR) and a battery of HAR-type models reviewed in Section 1. The rivalry models are listed as follows:

i. AR model: the simple autoregression model AR(22);
ii. HAR model: defined in (2);

iii. HAR-J model: defined in (4);
iv. HAR-CJ model: defined in (5);
v. HAR-RS-I model: defined in (6);
vi. HAR-RS-II model: defined in (7);
vii. HAR-SJ-I model: defined in (8);
viii. HAR-SJ-II model: defined in (9).

Bear in mind that all the considered volatility models explicitly control the effects of observed macroeconomic factors by $\boldsymbol{d}_t$. To justify the usage of the CCE estimator, we need to first verify the existence of the possible error cross-sectional dependence. In Table 3, we report the cross-section dependence (CD) test of Pesaran (2004, 2015) and their p-values, which are based on the average of pair-wise correlations of the residuals from various HAR-type regressions of individual stock volatility. For all regressions and forecast horizons, these residual terms show a considerable degree of cross-sectional dependence.[21] This implies that even after controlling for major predictors of volatility in our dataset, sizable CD still remains across component stocks, which supports the utilization of this information to improve the forecast accuracy.

The HARP model in our analysis is built upon the specification of HAR-RS-II, due to its sound out-of-sample performance documented in Patton and Sheppard (2015). This particularly implies that unit-specific regressors $\boldsymbol{x}_{it}$ are $(\mathrm{RV}_{it}^{(1)} \cdot \mathbb{I}(r_t < 0), \mathrm{RS}_{it}^{+}, \mathrm{RS}_{it}^{-})^{\top}$, while other regressors in Equation (10) are well defined above.[22] In the main experiment, we concentrate on forecasting the RV of the NASDAQ 100 ETF, where the information of unobserved common effects is extracted from the panel of realized volatilities of the NASDAQ 100 components. We want to see if the comovements of component stocks genuinely assist in predicting the index fund volatility.

For all of the exercises, we conduct an in-sample exercise with the full sample and a rolling window out-of-sample exercise. The window length is set at 1000. We also test other values of the window length and the results remain similar.[23] Each of the above candidate models is applied to the dataset, and a series of $h$ days ahead forecasts are obtained. Note that, to compute $h$-day ahead forecasts, we employ a direct forecasting approach in which we estimate $\mathrm{RV}_{t+h}$ in the above models.[24] We consider both short-horizon and long-horizon forecasts with $h = 1$, 5, 10, and 22. For assessing the out-of-sample performance, we calculate the five statistics in Section 2.3: (i) the MSFE; (ii) the standard deviation of forecast error (SDFE); (iii) the mean absolute forecast error (MAFE); (iv) the Mincer–Zarnowitz pseudo-$R^2$; and (v) the QLIKE function for each candidate model at each forecast horizon $h$.

21 Under the null of weak error cross-sectional dependence, the CD statistics are asymptotically distributed as $N(0, 1)$.

22 The lagged dependent variables $\bar{y}_{it}^{(l)}$ include $RV_{it}^{(5)}$ and $RV_{it}^{(22)}$, while $\mathrm{d}_t$ are observed macroeconomic factors, of which the details are explained in Section 3.1.

23 We test the results with the window length of 500 observations. Tables are provided in Online Appendix G.2.

24 This approach permits us to produce multi-step ahead forecasts without imposing any assumption about future realizations on the explanatory variables.

**Table 3** Results for CD test for RV of the NASDAQ 100 constituents

| Method | $h=1$ | | $h=5$ | | $h=10$ | | $h=22$ | |
|---|---|---|---|---|---|---|---|---|
| | CD | $p$-value | CD | $p$-value | CD | $p$-value | CD | $p$-value |
| RW | 5386.1615 | 0.0000 | 6661.1541 | 0.0000 | 6844.1628 | 0.0000 | 7426.2886 | 0.0000 |
| AR(22) | 4178.6508 | 0.0000 | 5563.3119 | 0.0000 | 6031.7421 | 0.0000 | 6814.7767 | 0.0000 |
| HAR | 4491.6446 | 0.0000 | 5719.2537 | 0.0000 | 6167.5320 | 0.0000 | 6957.3473 | 0.0000 |
| HAR-J | 4384.6859 | 0.0000 | 5655.8968 | 0.0000 | 6107.9819 | 0.0000 | 6895.6076 | 0.0000 |
| HAR-CJ | 4251.0679 | 0.0000 | 5474.3930 | 0.0000 | 5880.9295 | 0.0000 | 6596.3800 | 0.0000 |
| HAR-RS-I | 4386.1947 | 0.0000 | 5641.6365 | 0.0000 | 6118.8177 | 0.0000 | 6919.7391 | 0.0000 |
| HAR-RS-II | 4179.7576 | 0.0000 | 5530.1978 | 0.0000 | 6063.0118 | 0.0000 | 6884.2000 | 0.0000 |
| HAR-SJ-I | 4403.8828 | 0.0000 | 5646.2349 | 0.0000 | 6132.6288 | 0.0000 | 6926.7192 | 0.0000 |
| HAR-SJ-II | 4351.2087 | 0.0000 | 5611.6945 | 0.0000 | 6103.8054 | 0.0000 | 6900.0884 | 0.0000 |

*Notes:* CD is short for the cross-section dependence test statistic applied to the residuals of the asset-specific regression. RW indicates a random walk model. A large CD statistics indicates that the residuals of certain model estimation are correlated across individual component stocks.

### 4.1 The Scheme of Predicting $\hat{f}_{t+h}$

Before the study of various schemes for forecasting $\hat{f}_{t+h}$, an initial investigation on the statistical properties of $f_t$ seems necessary. We start by estimating $f_t$ based on the HARP specification and the CCE estimator, across various forecast horizons ($h = 1, 5, 10, 22$) separately. We report summary statistics of $f_t$ in Table 4.[25] As can be seen clearly, the null hypotheses of a normal distribution and a unit root process are both rejected for the two PCs ($f_1$ and $f_2$). Except for $f_2$ at $h=1$, the factors have highly persistent autocorrelation structures in other cases.

Due to the linear nature of the HARP specification and high persistence of $f_t$, we first contemplate three linear autoregressive models of forecasting $\hat{f}_{t+h}$: the AR model, the HAR model, the VAR model (Stock and Watson, 2002; Pesaran, Pick, and Allan, 2011),[26] where the number of lags is selected by the Bayes Information Criterion. As a complement to the above linear models, we also employ the random forest method, one of the popular machine learning tools, to capture the possibly nonlinear dynamics of $f_t$, since they are extracted from the noisy regression residuals.[27] For each of the above four methods, direct forecasts of $f_{t+h}$ based on the information up to time $t$ are performed, and a rolling window exercise is implemented to estimate the associated coefficients.

25 The sample autocorrelation functions of $f_t$ is reported in Figure A2 of Online Appendix F.

26 We also tried to apply the VARMA model to forecast $f_{t+h}$. However, the VARMA model requires the data to be highly stationary. Since we are using a rolling window exercise, this implies that $f_t$ needs to be highly stationary in each roll, which is usually not guaranteed in practice. Hence, we use the VAR model instead.

27 The inclusion of the random forest method provides a flexible way of accommodating the possible non-linearity of $f_t$. A detailed description of the random forest procedure is provided in Online Appendix C. The further investigation about how the random forest performs in the actual forecasting of RV is outside the scope of this article, which becomes the subject for future research.

**Table 4** Summary statistics for the unobserved common factors

| Statistic | $h=1$ | | $h=5$ | | $h=10$ | | $h=22$ | |
|---|---|---|---|---|---|---|---|---|
| | $f_1$ | $f_2$ | $f_1$ | $f_2$ | $f_1$ | $f_2$ | $f_1$ | $f_2$ |
| Mean | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Median | 0.2038 | 0.1382 | 0.2290 | 0.1189 | 0.2524 | −0.0393 | −0.2678 | 0.0697 |
| Maximum | 14.1083 | 7.2167 | 5.4090 | 20.6654 | 5.0550 | 12.5781 | 12.3710 | 15.2257 |
| Minimum | −14.1560 | −26.5783 | −18.9432 | −21.5940 | −15.4713 | −15.9108 | −6.7179 | −13.2345 |
| Standard deviation | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Skewness | −2.5078 | −11.2995 | −6.4684 | −0.8306 | −5.0666 | −2.0137 | 4.0513 | 0.7770 |
| Kurtosis | 56.8944 | 252.7725 | 88.6692 | 189.6395 | 54.0926 | 64.9030 | 35.8895 | 49.9837 |
| Jarque–Bera | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| ADF test | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |

*Notes:* Table 4 contains statistics for the unobserved common factors computed based on the CCE estimator and applied to RVs of the NASDAQ 100 constituents. For each forecast horizon ($h = 1, 5, 10, 22$), we extract two PCs. For JB and ADF tests, statistics that are outside tabulated critical values, we report maximum (0.999) or minimum (0.001) $p$-values.

With the HARP model and its benchmark specification (i.e., HAR-RS-II), the relative out-of-sample performance from the above four ways of forecasting $f_{t+h}$ is compared in Table 5. Looking across the columns, we see that the random forest method outperforms its three linear competitors for $h=5$ and 10. The Giacomini–White (GW) tests based on the mean absolute forecast errors, reported in Table 6, again corroborate this conclusion. In contrast, in the case of $h=1$ and 22, the performance of the random forest method is on a par with the other three methods. Based on the above outcomes, we believe that it is sensible to employ the random forest method as the benchmark forecasting scheme for $\hat{f}_{t+h}$ in the following exercises.

## 4.2 The In-Sample and Out-of-Sample Results for the NASDAQ 100 ETF

We begin our discussion by considering the in-sample results in Table 7, where the HARP model is compared with its baseline specification without unobserved common factors, the HAR-RS-II model. The race is evaluated by the in-sample predictive $R^2$'s. We find that, in the HARP model with $f_t$, the coefficients of $RS_t^-$ decrease substantially in magnitudes but remain significant for all horizons, and this effect is less obvious for the coefficients of $RS_t^+$. Akin to a substitution effect, the coefficients on all $f_t$ are relatively large and significant. Moreover, the HARP model explains 32% (at $h=1$) to 174% (at $h=22$) more of the variation in volatility than the HAR-RS-II model not containing $f_t$. The above finding implies that aside from additional information for explaining dependent variables, the unobserved common factors may partially capture the information contained in positive and negative realized semivariances.

Table 8 presents some descriptive results of the out-of-sample evaluation for forecasting 1, 5, 10, and 22 days ahead. In particular, we report the MSFE, SDFE, MAFE, QLIKE, and the pseudo $R^2$ from the rolling-window regressions for the HARP model as well as for the set of other candidate models. We find a consistent ranking of models across all forecast

**Table 5** The out-of-sample forecast comparison for different ways of forecasting $\hat{f}_{t+h}$

| Method | MSFE | QLIKE | MAFE | SDFE | Pseudo $R^2$ |
|---|---|---|---|---|---|
| Panel A: $h = 1$ | | | | | |
| $\text{HARP}_{\text{AR}}$ | 0.2190 | 0.1330 | 0.2753 | 0.4680 | 0.7767 |
| $\text{HARP}_{\text{HAR}}$ | 0.2160 | **0.1307** | 0.2722 | 0.4647 | 0.7799 |
| $\text{HARP}_{\text{VAR}}$ | 0.2159 | 0.1375 | 0.2775 | 0.4646 | 0.7800 |
| $\text{HARP}_{\text{RF}}$ | **0.2114** | 0.1339 | **0.2698** | **0.4598** | **0.7845** |
| Panel B: $h = 5$ | | | | | |
| $\text{HARP}_{\text{AR}}$ | 0.3666 | 0.2085 | 0.3758 | 0.6055 | 0.6272 |
| $\text{HARP}_{\text{HAR}}$ | 0.3622 | 0.2072 | 0.3724 | 0.6018 | 0.6316 |
| $\text{HARP}_{\text{VAR}}$ | 0.3564 | 0.1956 | 0.3604 | 0.5970 | 0.6375 |
| $\text{HARP}_{\text{RF}}$ | **0.3433** | **0.1931** | **0.3409** | **0.5859** | **0.6509** |
| Panel C: $h = 10$ | | | | | |
| $\text{HARP}_{\text{AR}}$ | 0.4075 | **0.2398** | 0.4024 | 0.6384 | 0.5862 |
| $\text{HARP}_{\text{HAR}}$ | 0.4075 | 0.2449 | 0.4059 | 0.6383 | 0.5862 |
| $\text{HARP}_{\text{VAR}}$ | 0.4150 | 0.2453 | 0.3790 | 0.6442 | 0.5785 |
| $\text{HARP}_{\text{RF}}$ | **0.4058** | 0.2654 | **0.3683** | **0.6370** | **0.5879** |
| Panel D: $h = 22$ | | | | | |
| $\text{HARP}_{\text{AR}}$ | 0.4670 | **0.2863** | 0.4595 | 0.6834 | 0.5261 |
| $\text{HARP}_{\text{HAR}}$ | **0.4629** | 0.3061 | 0.4614 | **0.6804** | **0.5303** |
| $\text{HARP}_{\text{VAR}}$ | 0.5193 | 0.3587 | 0.4410 | 0.7206 | 0.4731 |
| $\text{HARP}_{\text{RF}}$ | 0.4788 | 0.3748 | **0.4266** | 0.6919 | 0.5142 |

*Notes:* This table reports the out-of-sample results for predicting $h$-day future realized variation using the different models of forecasting $\hat{f}_{t+h}$. The candidate models are the HAR model ($\text{HARP}_{\text{HAR}}$), the AR($h$) model ($\text{HARP}_{\text{AR}}$), and the random forecast method ($\text{HARP}_{\text{RF}}$). The results are based on the transaction data of the NASDAQ 100 ETF spanning from May 22, 2007 to October 20, 2017 (a total of 2625 observations). We use a rolling window of 1000 observations to estimate the coefficients of the above models, and evaluate the out-of-sample forecast performance at four horizons ($h = 1$, 5, 10, and 22). Each panel in Table 5 corresponds to a specific forecast horizon, which varies from 1 day to 22 days. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

horizons: the AR(22) performs the worst, followed by models with high-frequency intraday data (i.e., the HAR, the HAR-J, and the HAR-CJ), while the more sophisticated RV models (i.e., the HAR-RS-I, the HAR-RS-II, the HAR-SJ-I, the HAR-SJ-II, and the HARP) perform better. The HARP model has the best performance in all cases of $h$. To further examine whether the outperformance is statistically significant, we perform the modified GW test in Table 9. The result shows that the outperformance of the HARP model is statistically significant at 5% for all forecast horizons.

It is perhaps more informative to focus on the performance comparison of the HAR, the HAR-RS-II, and the HARP models due to their nested specifications in sequence. The above comparison shows that relative to the basic HAR model, the HARP specification improves 26.2–37% more of the variation in future volatility. Even if we control the effect of semi-variance components in the HAR-RS-II model, the HARP can still explain 2.5% ($h = 1$) to 33% ($h = 22$) more of the variation in future volatility. When the horizon increases, the unobserved factors add more value to the forecast precision. A possible interpretation of this is rooted from the fact that $f_t$ is constructed from regression residuals. In the case of short

**Table 6** The GW test for the mean absolute forecast errors—different ways of forecasting $\hat{f}_{t+h}$

| Method | $HARP_{AR}$ | $HARP_{HAR}$ | $HARP_{VAR}$ | $HARP_{AR}$ | $HARP_{HAR}$ | $HARP_{VAR}$ |
|---|---|---|---|---|---|---|
| | | $h=1$ | | | $h=5$ | |
| $HARP_{AR}$ | – | – | – | – | – | – |
| $HARP_{HAR}$ | 0.0244 | – | – | 0.1274 | – | – |
| $HARP_{VAR}$ | 0.4173 | 0.0185 | – | 0.0097 | 0.0266 | |
| $HARP_{RF}$ | 0.0573 | 0.3722 | 0.4001 | **0.0000** | **0.0000** | **0.0000** |
| | | $h=10$ | | | $h=22$ | |
| $HARP_{AR}$ | – | – | – | – | – | – |
| $HARP_{HAR}$ | 0.3843 | – | – | 0.6001 | – | – |
| $HARP_{VAR}$ | 0.0088 | 0.0021 | – | 0.3418 | 0.2872 | – |
| $HARP_{RF}$ | **0.0003** | **0.0001** | **0.0141** | **0.0476** | **0.0325** | 0.1026 |

*Notes:* The modified GW test (Giacomini and White, 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally well as the *column method* (in horizontal headings) in terms of the absolute forecast error. ($h = 1, 5, 10, 22$ Table 6, Bold numbers indicate the null hypothesis can be rejected at 5% level of significance.

horizons, the information from major predictors of the HAR-RS-II model is plentiful so that there is not much remaining in the residuals. However, with the expansion of horizons, it is more likely to have extra information left unexplored in residuals.

## 5 Robustness Checks

This section presents three out-of-sample checks on the conclusions from the previous section, with the identical set of volatility models. The first is an application to the DJIA. The second is a test of the main results on the NASDAQ 100 ETF with a five-minute sampling interval. For space consideration, we only report the results for $h = 1$. Results for other forecast horizons can be found in Online Appendix G.1. The third considers estimating the unobserved factors alternatively based on the PC analysis and then use these factor estimates to augment the regression (18). Apart from the above trials, we also test our findings by other industrial indices in Online Appendix G.2, by a different window length in Online Appendix G.3, and by varying the sample period in Online Appendix G.4.

### 5.1 Evidence on the DJIA

The previous sections presented results for the NASDAQ 100 ETF. In this section, we report out-of-sample results for the DJIA index. Data at one-minute sampling interval is provided by Pitrading Incorporation and covers the same period as the main results. The DJIA is a weighted average index that includes the value of thirty large, publicly owned companies based in the United States. All forecasts are generated using rolling window regressions based on 1000 observations, and parameter estimates are updated daily.

The results are reported in Tables 10 and 11. We note that the HARP forecast is always the winner and significantly improves the out-of-sample forecast performance. Relative to the best semivariance-based specification, the HARP generates gains in the out-of-sample $R^2$ by 35.2%. The GW test in Table 11 implies that the improvement by HARP is

**Table 7** In-sample results for the benchmark specification with and without unobserved common factors

| | $h=1$ | | $h=5$ | | $h=10$ | | $h=22$ | |
|---|---|---|---|---|---|---|---|---|
| | $M$ | $M'$ | $M$ | $M'$ | $M$ | $M'$ | $M$ | $M'$ |
| Panel A: Feasible common factors | | | | | | | | |
| Constant | 0.3700 | 0.4505 | 0.4317 | 0.5096 | 0.4845 | 0.5260 | 0.6013 | 0.6428 |
| | (0.0413) | (0.0237) | (0.0508) | (0.0310) | (0.0565) | (0.0352) | (0.0629) | (0.0379) |
| OIL | 0.4373 | −0.2574 | −0.9595 | −0.9611 | 0.4554 | −0.1422 | −0.7933 | −1.0313 |
| | (0.6334) | (0.3622) | (0.7781) | (0.4736) | (0.8639) | (0.5377) | (0.9577) | (0.5770) |
| USD | −1.7878 | 1.2087 | −1.3487 | 0.7342 | −3.2909 | −1.2032 | 0.9706 | 2.3014 |
| | (3.0697) | (1.7551) | (3.7716) | (2.2960) | (4.1871) | (2.6061) | (4.6416) | (2.7970) |
| CS | −1.3083 | −0.8762 | −0.0156 | 0.7781 | 0.4170 | 1.1349 | 0.9932 | 1.6086 |
| | (0.6697) | (0.3829) | (0.8226) | (0.5009) | (0.9131) | (0.5683) | (1.0124) | (0.6104) |
| TS | −0.0590 | −0.0254 | −0.0419 | 0.0049 | −0.0298 | −0.0233 | −0.0303 | −0.0196 |
| | (0.0180) | (0.0103) | (0.0222) | (0.0135) | (0.0246) | (0.0154) | (0.0273) | (0.0165) |
| FFD | 1.7295 | 1.2395 | 1.0488 | 0.3062 | 1.5633 | 1.5272 | 1.7559 | 1.6254 |
| | (0.1547) | (0.0889) | (0.1901) | (0.1165) | (0.2112) | (0.1348) | (0.2344) | (0.1429) |
| Panel B: Cross-section-specific regressors | | | | | | | | |
| $RV_t^{(1)} \cdot \mathbb{I}(r_t < 0)$ | 0.0778 | 0.0798 | −0.1214 | −0.0585 | −0.0954 | −0.0634 | −0.0147 | 0.0068 |
| | (0.0333) | (0.0190) | (0.0409) | (0.0249) | (0.0453) | (0.0282) | (0.0502) | (0.0303) |
| $RS_t^+$ | 0.1129 | 0.1315 | 0.0926 | 0.0995 | 0.2715 | 0.1749 | 0.0196 | 0.0176 |
| | (0.0501) | (0.0287) | (0.0616) | (0.0375) | (0.0683) | (0.0426) | (0.0756) | (0.0457) |
| $RS_t^-$ | 0.2733 | 0.0957 | 0.1447 | 0.0673 | 0.1229 | 0.0459 | 0.1368 | 0.0731 |
| | (0.0163) | (0.0097) | (0.0201) | (0.0123) | (0.0223) | (0.0139) | (0.0247) | (0.0149) |
| $RV_t^{(5)}$ | 0.2379 | 0.1930 | 0.3287 | 0.1714 | 0.1177 | 0.0674 | 0.1744 | 0.0746 |
| | (0.0220) | (0.0126) | (0.0270) | (0.0166) | (0.0300) | (0.0187) | (0.0332) | (0.0201) |
| $RV_t^{(22)}$ | 0.2710 | 0.2490 | 0.2548 | 0.2466 | 0.3098 | 0.3770 | 0.2320 | 0.2923 |
| | (0.0165) | (0.0094) | (0.0203) | (0.0123) | (0.0225) | (0.0149) | (0.0249) | (0.0151) |
| Panel C: Estimated unobserved common factors | | | | | | | | |
| Common effect $f_1$ | | −0.4718 | | −0.6600 | | −0.7052 | | 0.8185 |
| | | (0.0086) | | (0.0108) | | (0.0124) | | (0.0132) |
| Common effect $f_2$ | | −0.4150 | | −0.3149 | | 0.3062 | | −0.2731 |
| | | (0.0081) | | (0.0109) | | (0.0133) | | (0.0136) |
| Panel D: Goodness of fit | | | | | | | | |
| $R^2$ | 0.6765 | 0.8944 | 0.5127 | 0.8196 | 0.4008 | 0.7681 | 0.2686 | 0.7347 |
| Adj. $R^2$ | 0.6753 | 0.8939 | 0.5108 | 0.8188 | 0.3985 | 0.7670 | 0.2658 | 0.7335 |

*Notes:* This table reports the in-sample results for predicting the $h$-day future realized volatility using the HAR-RS-II model and the HARP model. The estimation is based on the full sample data of the NASDAQ 100 ETF and considers a range of forecast horizons ($h = 1, 5, 10, 22$). Panel A reports the coefficient estimates of observed common factors and their standard errors (in parentheses). Panel B reports the coefficient estimates of unit-specific regressors and their standard errors (in brackets), while Panel C reports the relevant results for unobserved common factors. The bottom panel provides the in-sample predictive $R^2$ and adjusted $R^2$.

**Table 8** Out-of-sample forecast comparison of models for RV of the NASDAQ 100 ETF

| Method | MSFE | QLIKE | MAFE | SDFE | Pseudo $R^2$ |
|---|---|---|---|---|---|
| Panel A: $h=1$ | | | | | |
| AR(22) | 0.4428 | 0.1587 | 0.3282 | 0.6654 | 0.5487 |
| HAR | 0.4190 | 0.1569 | 0.3250 | 0.6473 | 0.5729 |
| HAR-J | 0.3783 | 0.1487 | 0.3114 | 0.6150 | 0.6145 |
| HAR-CJ | 0.3575 | 0.1491 | 0.3096 | 0.5979 | 0.6356 |
| HAR-RS-I | 0.2341 | 0.1371 | 0.2884 | 0.4839 | 0.7614 |
| HAR-RS-II | 0.2302 | 0.1349 | 0.2819 | 0.4798 | 0.7654 |
| HAR-SJ-I | 0.2498 | 0.1465 | 0.2927 | 0.4998 | 0.7454 |
| HAR-SJ-II | 0.2333 | 0.1382 | 0.2891 | 0.4830 | 0.7622 |
| HARP | **0.2114** | **0.1339** | **0.2698** | **0.4598** | **0.7845** |
| Panel B: $h=5$ | | | | | |
| AR(22) | 0.5623 | 0.2618 | 0.4247 | 0.7499 | 0.4280 |
| HAR | 0.4761 | 0.2547 | 0.4101 | 0.6900 | 0.5158 |
| HAR-J | 0.4768 | 0.2728 | 0.4052 | 0.6905 | 0.5151 |
| HAR-CJ | 0.4578 | 0.3255 | 0.3942 | 0.6766 | 0.5343 |
| HAR-RS-I | 0.4780 | 0.2548 | 0.3984 | 0.6914 | 0.5138 |
| HAR-RS-II | 0.4632 | 0.2612 | 0.3965 | 0.6806 | 0.5289 |
| HAR-SJ-I | 0.4874 | 0.2577 | 0.3999 | 0.6981 | 0.5043 |
| HAR-SJ-II | 0.4462 | 0.2443 | 0.3925 | 0.6680 | 0.5461 |
| HARP | **0.3433** | **0.1931** | **0.3409** | **0.5859** | **0.6509** |
| Panel C: $h=10$ | | | | | |
| AR(22) | 0.5782 | 0.3126 | 0.4460 | 0.7604 | 0.4128 |
| HAR | 0.5426 | 0.3322 | 0.4398 | 0.7366 | 0.4490 |
| HAR-J | 0.5223 | 0.9742 | 0.4368 | 0.7227 | 0.4696 |
| HAR-CJ | 0.5222 | 0.3965 | 0.4276 | 0.7226 | 0.4697 |
| HAR-RS-I | 0.5107 | 0.3472 | 0.4310 | 0.7146 | 0.4814 |
| HAR-RS-II | 0.5027 | 0.3552 | 0.4269 | 0.7090 | 0.4895 |
| HAR-SJ-I | 0.5056 | 0.3622 | 0.4308 | 0.7110 | 0.4866 |
| HAR-SJ-II | 0.9043 | 0.3170 | 0.4415 | 0.9509 | 0.0817 |
| HARP | **0.4058** | **0.2654** | **0.3683** | **0.6370** | **0.5879** |
| Panel D: $h=22$ | | | | | |
| AR(22) | 0.6179 | 0.3958 | 0.4941 | 0.7861 | 0.3730 |
| HAR | 0.6012 | 0.3952 | 0.4899 | 0.7754 | 0.3900 |
| HAR-J | 0.6112 | 0.3815 | 0.4885 | 0.7818 | 0.3799 |
| HAR-CJ | 0.6250 | 0.3982 | 0.4883 | 0.7906 | 0.3658 |
| HAR-RS-I | 0.6035 | 0.4550 | 0.4899 | 0.7768 | 0.3877 |
| HAR-RS-II | 0.6045 | 0.3888 | 0.4872 | 0.7775 | 0.3866 |
| HAR-SJ-I | 0.6017 | 0.4642 | 0.4898 | 0.7757 | 0.3894 |
| HAR-SJ-II | 0.6055 | 0.5216 | 0.4887 | 0.7781 | 0.3856 |
| HARP | **0.4788** | **0.3748** | **0.4266** | **0.6919** | **0.5142** |

*Notes:* This table reports the out-of-sample results for predicting $h$-day future realized variation using the different predictor variables and risk models. The results are based on the transaction data of the NASDAQ 100 ETF spanning from May 22, 2007 to October 20, 2017 (a total of 2625 observations). We use a rolling window of 1000 observations to estimate the coefficients of the models, and evaluate the out-of-sample forecast performance at four horizons ($h=1$, $h=5$, $h=10$, and $h=22$). Each panel in Table 8 corresponds to a specific forecast horizon, which ranges from 1 day to 22 days. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

**Table 9** The GW test for the mean absolute forecast errors—the NASDAQ 100 ETF

| Method | AR(22) | HAR | HAR-J | HAR-CJ | HAR-RS-I | HAR-RS-II | HAR-SJ-I | HAR-SJ-II |
|---|---|---|---|---|---|---|---|---|
| Panel A: $h = 1$ | | | | | | | | |
| AR(22) | – | – | – | – | – | – | – | – |
| HAR | 0.5936 | – | – | – | – | – | – | – |
| HAR-J | 0.0068 | 0.0000 | – | – | – | – | – | – |
| HAR-CJ | 0.0008 | 0.0000 | 0.4332 | – | – | – | – | – |
| HAR-RS-I | 0.0001 | 0.0001 | 0.0050 | 0.0058 | – | – | – | – |
| HAR-RS-II | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0304 | – | – | – |
| HAR-SJ-I | 0.0000 | 0.0000 | 0.0020 | 0.0028 | 0.0656 | 0.0000 | – | – |
| HAR-SJ-II | 0.0002 | 0.0002 | 0.0068 | 0.0086 | 0.3482 | 0.0190 | 0.1289 | – |
| HARP | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0060** | **0.0000** | **0.0000** |
| Panel B: $h = 5$ | | | | | | | | |
| AR(22) | – | – | – | – | – | – | – | – |
| HAR | 0.1284 | – | – | – | – | – | – | – |
| HAR-J | 0.0475 | 0.0062 | – | – | – | – | – | – |
| HAR-CJ | 0.0028 | 0.0002 | 0.0077 | – | – | – | – | – |
| HAR-RS-I | 0.0139 | 0.0003 | 0.0027 | 0.3819 | – | – | – | – |
| HAR-RS-II | 0.0093 | 0.0000 | 0.0003 | 0.5971 | 0.1716 | – | – | – |
| HAR-SJ-I | 0.0178 | 0.0030 | 0.0260 | 0.2597 | 0.0175 | 0.0714 | – | – |
| HAR-SJ-II | 0.0034 | 0.0000 | 0.0000 | 0.7023 | 0.0072 | 0.0176 | 0.0052 | – |
| HARP | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** |
| Panel C: $h = 10$ | | | | | | | | |
| AR(22) | – | – | – | – | – | – | – | – |
| HAR | 0.4620 | – | – | – | – | – | – | – |
| HAR-J | 0.3554 | 0.2268 | – | – | – | – | – | – |
| HAR-CJ | 0.0477 | 0.0042 | 0.0154 | – | – | – | – | – |
| HAR-RS-I | 0.1665 | 0.0236 | 0.0022 | 0.4104 | – | – | – | – |
| HAR-RS-II | 0.1279 | 0.0359 | 0.0165 | 0.9043 | 0.1295 | – | – | – |
| HAR-SJ-I | 0.1858 | 0.0536 | 0.0185 | 0.4746 | 0.8771 | 0.0495 | – | – |
| HAR-SJ-II | 0.6800 | 0.8733 | 0.7063 | 0.2821 | 0.4451 | 0.3663 | 0.4666 | – |
| HARP | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0001** |
| Panel D: $h = 22$ | | | | | | | | |
| AR(22) | – | – | – | – | – | – | – | – |
| HAR | 0.6077 | – | – | – | – | – | – | – |
| HAR-J | 0.4487 | 0.5539 | – | – | – | – | – | – |
| HAR-CJ | 0.5025 | 0.6427 | 0.9489 | – | – | – | – | – |
| HAR-RS-I | 0.5918 | 0.9796 | 0.4787 | 0.6039 | – | – | – | – |
| HAR-RS-II | 0.3305 | 0.3026 | 0.4892 | 0.7376 | 0.1795 | – | – | – |
| HAR-SJ-I | 0.5900 | 0.8755 | 0.6093 | 0.6654 | 0.8190 | 0.2828 | – | – |
| HAR-SJ-II | 0.4660 | 0.5700 | 0.8699 | 0.8888 | 0.3776 | 0.2854 | 0.5575 | – |
| HARP | **0.0075** | **0.0127** | **0.0136** | **0.0140** | **0.0124** | **0.0131** | **0.0126** | **0.0120** |

*Notes:* The modified GW test (Giacomini and White, 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Corresponding $p$-values for a number of forecasting horizons ($h = 1, 5, 10, 22$) are reported in Panels A–D of Table 9, respectively. Bold numbers indicate the null hypothesis can be rejected at 5% level of significance.

**Table 10** Out-of-sample forecast comparison of models on the DJIA ($h = 1$)

| Method | MSFE | QLIKE | MAFE | SDFE | Pseudo $R^2$ |
|---|---|---|---|---|---|
| AR(22) | 1.1207 | 0.2401 | 0.3852 | 1.0586 | 0.2990 |
| HAR | 0.9000 | 0.2177 | 0.3707 | 0.9487 | 0.4370 |
| HAR-J | 0.8720 | 0.2116 | 0.3648 | 0.9338 | 0.4545 |
| HAR-CJ | 0.9153 | 0.2154 | 0.3722 | 0.9567 | 0.4275 |
| HAR-RS-I | 0.9776 | 0.2063 | 0.3636 | 0.9887 | 0.3885 |
| HAR-RS-II | 0.8223 | 0.2049 | 0.3467 | 0.9068 | 0.4857 |
| HAR-SJ-I | 0.9325 | 0.2089 | 0.3662 | 0.9656 | 0.4168 |
| HAR-SJ-II | 1.1121 | 0.2036 | 0.3653 | 1.0546 | 0.3044 |
| HARP | **0.5487** | **0.1852** | **0.3005** | **0.7407** | **0.6568** |

*Notes:* This table reports the out-of-sample results for predicting one-day ahead future realized variation using the different predictor variables and risk models. The results are based on data of the DJIA spanning from May 22, 2007 to October 20, 2017 (a total of 2625 observations). We use a rolling window of 1000 observations to estimate the coefficients of the models, and evaluate the out-of-sample forecast performance. Bold numbers indicate the best performing model by each criterion.

**Table 11** The GW test for the mean absolute forecast errors-the DJIA ($h = 1$)

| Method | AR(22) | HAR | HAR-J | HAR-CJ | HAR-RS-I | HAR-RS-II | HAR-SJ-I | HAR-SJ-II |
|---|---|---|---|---|---|---|---|---|
| AR(22) | – | – | – | – | – | – | – | – |
| HAR | 0.0959 | – | – | – | – | – | – | – |
| HAR-J | 0.0239 | 0.0266 | – | – | – | – | – | – |
| HAR-CJ | 0.1254 | 0.3795 | 0.0016 | – | – | – | – | – |
| HAR-RS-I | 0.0129 | 0.0034 | 0.7456 | 0.0033 | – | – | – | – |
| HAR-RS-II | 0.0000 | 0.0000 | 0.0009 | 0.0000 | 0.0001 | – | – | – |
| HAR-SJ-I | 0.0316 | 0.1331 | 0.7198 | 0.0833 | 0.1478 | 0.0000 | – | – |
| HAR-SJ-II | 0.0324 | 0.1526 | 0.9165 | 0.0919 | 0.4069 | 0.0004 | 0.7530 | – |
| HARP | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** | **0.0000** |

*Notes:* The modified GW test (Giacomini and White, 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Bold numbers indicate the null hypothesis can be rejected at 5% level of significance.

significant at 1%. Overall, the above conclusions are in agreement with the results for the NASDAQ 100 ETF.

### 5.2 Variation in the Sampling Frequency

Here we examine the robustness of the HARP model to the intraday RVs constructed from a five-minute sampling frequency. This choice directly reflects the sampling frequency adopted in much of the existing realized volatility literature.[28] Summary statistics of the five-minute RVs of the NASDAQ 100 ETF are reported in Online Appendix G.5.

28 See also the theoretical comparisons of various volatility estimators in Andersen, Bollerslev, and Meddahi (2011) and Ghysels and Sinko (2011) from a forecasting perspective. Liu, Patton, and

**Table 12** Out-of-sample forecast comparison of models on the NASDAQ 100 ETF sampled at a five-minute frequency ($h = 1$)

| Method | MSFE | QLIKE | MAFE | SDFE | Pseudo $R^2$ |
|---|---|---|---|---|---|
| AR(22) | 1.1111 | 0.2327 | 0.4401 | 1.3125 | −0.8863 |
| HAR | 1.0463 | 0.2257 | 0.3824 | 1.0229 | −0.1457 |
| HAR-J | 0.7040 | 0.1963 | 0.3428 | 0.8390 | 0.2291 |
| HAR-CJ | 0.7178 | 0.1881 | 0.3293 | 0.8472 | 0.2139 |
| HAR-RS-I | 0.3090 | 0.1906 | 0.3269 | 0.5559 | 0.6616 |
| HAR-RS-II | 0.6451 | 0.1847 | 0.3255 | 0.8032 | 0.2936 |
| HAR-SJ-I | 0.2812 | 0.1909 | 0.3249 | 0.5303 | 0.6921 |
| HAR-SJ-II | 0.2641 | 0.1853 | 0.3125 | 0.5139 | 0.7108 |
| HARP | **0.2400** | **0.1757** | **0.2948** | **0.4899** | **0.7372** |

*Notes:* This table reports the out-of-sample results for predicting one-day future realized variation using the different predictor variables and risk models. The results are based on data of the Dow Jones Transportation Average spanning from May 22, 2007 to October 20, 2017 (a total of 2625 observations). We use a rolling window of 1000 observations to estimate the coefficients of the models and evaluate the out-of-sample forecast performance at $h = 1$. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

Comparing to statistics on one-minute intraday data in Table 2, we notice very minor changes. We then duplicate the rolling window regressions on the five-minute data in Tables 12 and 13. The losses for all the forecast criteria are systematically lower than those in the one-minute case. We also notice that the results are qualitatively the same as those based on the one-minute data. The experiment confirms that the robustness of HARP under different sampling frequencies.

## 5.3 Alternative Estimation of Unobserved Factors

Since our article is related to volatility spillover literature which always explores more information from other datasets, it is an interesting trial to compare the HARP model with the HAR models incorporating common factors extracted from other time series. Two alternative approaches are considered below. The first method (denoted by HARPCA1) conducts a direct estimation of unobserved common factors from panels of RVs of the NASDAQ 100 components, using the PC analysis. Following Kapetanios and Pesaran (2004), this is implemented in a two-stage procedure, where in the first-stage PCs for RVs of the NASDAQ 100 components are obtained as in Stock and Watson (2002), and in the second stage, the regression for the NASDAQ 100 ETF is estimated augmenting the observed regressors with estimated PCs. It can be seen that this approach immediately generates unobserved factors instead of approximating them initially by cross-section averages of the dependent variable and the observed regressors.[29]

Sheppard (2015) give a comprehensive empirical study of 400 volatility estimators across multiple assets.

29 A theoretical comparison is provided in Kapetanios and Pesaran (2004) on the small sample properties of the CCE method and the PC approach. After a series of Monte Carlo experiments, Kapetanios and Pesaran (2004) conclude that the PC augmented method does not perform as well as the CCE estimator and can be subject to substantial bias.

**Table 13** The GW test for the mean absolute forecast error—the NASDAQ 100 ETF at a five-minute sampling frequency ($h = 1$)

| Method | AR(22) | HAR | HAR-J | HAR-CJ | HAR-RS-I | HAR-RS-II | HAR-SJ-I | HAR-SJ-II |
|---|---|---|---|---|---|---|---|---|
| AR(22) | | | | | | | | |
| HAR | 0.0013 | | | | | | | |
| HAR-J | 0.0000 | 0.0000 | | | | | | |
| HAR-CJ | 0.0000 | 0.0000 | 0.0000 | | | | | |
| HAR-RS-I | 0.0000 | 0.0012 | 0.1849 | 0.8437 | | | | |
| HAR-RS-II | 0.0000 | 0.0000 | 0.0000 | 0.3558 | 0.9096 | | | |
| HAR-SJ-I | 0.0000 | 0.0048 | 0.2370 | 0.7770 | 0.5794 | 0.9649 | | |
| HAR-SJ-II | 0.0000 | 0.0008 | 0.0542 | 0.3025 | 0.0013 | 0.3993 | 0.0000 | |
| HARP | **0.0000** | **0.0000** | **0.0029** | **0.0379** | **0.0000** | **0.0033** | **0.0000** | **0.0000** |

*Notes:* The modified GW test (Giacomini and White, 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Bold numbers indicate the null hypothesis can be rejected at 5% level of significance.

The second choice implements the PC analysis of lagged RVs of the NASDAQ 100 ETF and thus augments the regression (18) with the information of its own lagged terms (denoted by HARPCA2). This is inspired by the work of Vortelinos (2017), which has already made the comparison between the HAR and the model augmented with PCs of the dependent variable's lagged terms. Vortelinos (2017) found that the HAR marginally outperforms the PC-based model.

The out-of-sample results are presented in Table 14, which shows that the HARP defeats the other two competitors under each of the evaluation criteria. Table 15 confirms the improvement is significant in all cases except for $h = 22$. It is not a surprise the HARPCA1 dominates the HARPCA2 in our sample, since the former is likely to access richer information from other time series rather than lagged RVs of the NASDAQ 100 ETF *per se*.

## 6 Conclusions

In this article, we argue that the linkages among component stock volatilities are important for forecasting the relevant index or index fund volatility. We develop a panel-based HAR model assuming unobserved common factors across cross-sectional units to capture the comovements in realized volatility. The framework configuration draws from the CCE-type estimators of Pesaran (2006) and Chudik and Pesaran (2015). It is shown that the CCE estimator is consistent. Monte Carlo studies confirm that the CCE estimator has sound finite sample performance.

We illustrate the relevance of the proposed HARP model by an empirical application to forecasting the realized volatility of the NASDAQ 100 ETF. The in-sample analysis discloses that the unobserved factors from the panel regression play an important role. They may capture information that is not contained already in the asset-specific realized volatility histories, such as the sentiment of the financial market, the news effect, or the varying risk premium.[30] Taking the unobserved factors into account can lead up to 174% increase in the

30 See, for example, Baker and Wurgler (2006) and Baker, Wurgler, and Yuan (2012) for the importance of investor sentiment in explaining the cross section of stock returns.

**Table 14** Out-of-sample comparison of the HARP and models with other factor estimates on the NASDAQ 100 ETF

| Method | MSFE | QLIKE | MAFE | SDFE | Pseudo $R^2$ |
|---|---|---|---|---|---|
| Panel A: $h = 1$ | | | | | |
| HARPCA1 | 0.2382 | 0.8017 | 0.2887 | 0.4881 | 0.7577 |
| HARPCA2 | 0.2517 | 0.1402 | 0.2892 | 0.5017 | 0.7435 |
| HARP | **0.2129** | **0.1307** | **0.2709** | **0.4614** | **0.7831** |
| Panel B: $h = 5$ | | | | | |
| HARPCA1 | 0.4310 | 1.3620 | 0.3681 | 0.6565 | 0.5616 |
| HARPCA2 | 0.4783 | 0.2661 | 0.4057 | 0.6916 | 0.5136 |
| HARP | **0.3458** | **0.2072** | **0.3409** | **0.5880** | **0.6483** |
| Panel C: $h = 10$ | | | | | |
| HARPCA1 | 0.4903 | 0.9658 | 0.3941 | 0.7002 | 0.5021 |
| HARPCA2 | 0.5243 | 0.3578 | 0.4339 | 0.7241 | 0.4675 |
| HARP | **0.4071** | **0.2449** | **0.3662** | **0.6380** | **0.5866** |
| Panel D: $h = 22$ | | | | | |
| HARPCA1 | 0.5650 | 1.1543 | 0.4390 | 0.7516 | 0.4267 |
| HARPCA2 | 0.7484 | 0.4043 | 0.4926 | 0.8651 | 0.2406 |
| HARP | **0.4835** | **0.3061** | **0.4258** | **0.6954** | **0.5093** |

*Notes:* The results are based on data of the NASDAQ 100 ETF spanning from May 22, 2007 to October 20, 2017 (a total of 2625 observations). HARPCA1 and HARPCA2 denote estimating the unobserved factors by the PC analysis of two different time series, respectively. We use a rolling window of 1000 observations to estimate the coefficients of the models, and evaluate the out-of-sample forecast performance. Bold numbers indicate the best performing model by each criterion at each forecast horizon.

**Table 15** The GW test for the mean absolute forecast errors—various spillover effect estimators

| Method | HARPCA1 | HARPCA2 | HARP | HARPCA1 | HARPCA2 | HARP |
|---|---|---|---|---|---|---|
| | | $h = 1$ | | | $h = 5$ | |
| HARPCA1 | – | – | – | – | – | – |
| HARPCA2 | 0.0000 | – | – | 0.0000 | – | – |
| HARP | **0.0002** | **0.0000** | – | **0.0005** | **0.0000** | |
| | | $h = 10$ | | | $h = 22$ | |
| HARPCA1 | – | – | – | – | – | – |
| HARPCA2 | 0.0000 | – | – | 0.0000 | – | – |
| HARP | **0.0261** | **0.0000** | – | 0.5612 | **0.0000** | – |

*Notes:* The modified GW test (Giacomini and White, 2006) is implemented to test the null hypothesis that the *row method* (in vertical headings) performs equally well as the *column method* (in horizontal headings) in terms of the absolute forecast error. Corresponding $p$-values for a number of forecasting horizons ($h = 1, 5, 10, 22$) are reported in Panels A–D of Table 6, respectively. Bold numbers indicate the null hypothesis can be rejected at 5% level of significance.

in-sample $R^2$. In the out-of-sample exercise, the HARP model that includes these latent common factors significantly outperforms benchmark models that do not. Moreover, the HARP model produces substantial predictive gains, as measured by 2.5–69.6% larger out-of-sample $R^2$ values, relative to those without common unobservable factors. Our findings are robust to different stock indices and alternative estimates of the unobserved common factors.

Our HARP estimator is not limited to the univariate forecasting exercises demonstrated in this article. It is possible that the HARP method can be further applied to the full realized covariance matrix. We leave these for future research.

## Supplemental Data

Supplemental data is available at *Journal of Financial Econometrics* online.

## References

Andersen, T. G., and T. Bollerslev. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39: 885–905.

Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2007. Roughing It up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics* 89: 701–720.

Andersen, T. G., T. Bollerslev, and N. Meddahi. 2011. Realized Volatility Forecasting and Market Microstructure Noise. *Journal of Econometrics* 160: 220–234.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2000. Great Realizations. *RISK* 13: 105–108.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96: 42–55.

Anderson, H. M., and F. Vahid. 2007. Forecasting the Volatility of Australian Stock Returns. *Journal of Business & Economic Statistics* 25: 76–90.

Asai, M., and M. McAleer. 2015. Forecasting Co-Volatilities via Factor Models with Asymmetry and Long Memory in Realized Covariance. *Journal of Econometrics* 189: 251–262.

Asai, M., M. McAleer, and J. Yu. 2006. Multivariate Stochastic Volatility: A Survey. *Econometric Reviews* 25: 145–175.

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

Baker, M., and J. Wurgler. 2006. Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance* 61: 1645–1680.

Baker, M., J. Wurgler, and Y. Yuan. 2012. Global, Local, and Contagious Investor Sentiment. *Journal of Financial Economics* 104: 272–287.

Bauer, G. H., and K. Vorkink. 2011. Forecasting Multivariate Realized Stock Market Volatility. *Journal of Econometrics* 160: 93–101.

Bauwens, L., G. Storti, and F. Violante. 2012. "Dynamic Conditional Correlation Models for Realized Covariance Matrices." CORE Discussion papers 2012060, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Bauwens, L., S. Laurent, and J. V. K. Rombouts. 2006. Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics* 21: 79–109.

Bernoth, K., and A. Pick. 2009. "Forecasting the Fragility of the Banking and Insurance Sectors." De Nederlandsche Bank. DNB Working paper 202.

Bernoth, K., and A. Pick. 2011. Forecasting the Fragility of the Banking and Insurance Sectors. *Journal of Banking & Finance* 35: 807–818.

Bollerslev, T., and R. F. Engle. 1993. Common Persistence in Conditional Variances. *Econometrica* 61: 167–186.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen. 2018. Risk Everywhere: Modeling and Managing Volatility. *The Review of Financial Studies* 31: 2729–2773.

Bollersleva, T., J. Li, J. A. Patton, and R. Quaedvlieg. 2019. "Realized Semicovariances." Discussion paper.

Cheng, M., N. R. Swanson, and X. Yang. 2019. "Forecasting Volatility Using Double Shrinkage Methods." Discussion paper.

Chudik, A., and M. H. Pesaran. 2015. Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors. *Journal of Econometrics* 188: 393–420.

Chudik, A., K. Mohaddes, M. H. Pesaran, and M. Raissi. 2017. Is There a Debt-Threshold Effect on Output Growth? *Review of Economics and Statistics* 99: 135–150.

Corsi, F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–196.

Corsi, F., F. Audrino, and R. Renò. 2012. "HAR Modeling for Realized Volatility Forecasting." In *Handbook of Volatility Models and Their Applications*, pp. 363–382. John Wiley & Sons, Inc.

Diebold, F. X., and M. Nerlove. 1989. The Dynamics of Exchange Rate Volatility: A Multivariate Latent Factor ARCH Model. *Journal of Applied Econometrics* 4: 1–21.

Engle, R. F., T. Ito, and W.-L. Lin. 1990. Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market. *Econometrica* 58: 525–542.

Engle, R. F., V. K. Ng, and M. Rothschild. 1990. Asset Pricing with a Factor-Arch Covariance Structure: Empirical Estimates for Treasury Bills. *Journal of Econometrics* 45: 213–237.

Fernandes, M., M. C. Medeiros, and M. Scharth. 2014. Modeling and Predicting the CBOE Market Volatility Index. *Journal of Banking & Finance* 40: 1–10.

Ghysels, E., and A. Sinko. 2011. Volatility Forecasting and Microstructure Noise. *Journal of Econometrics* 160: 257–271.

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

Gourieroux, C., J. Jasiak, and R. Sufana. 2009. The Wishart Autoregressive Process of Multivariate Stochastic Volatility. *Journal of Econometrics* 150: 167–181.

Halbleib, R., and V. Voev. 2014. Forecasting Covariance Matrices: A Mixed Approach. *Journal of Financial Econometrics* 14: 383–417.

Hansen, P. R., and A. Lunde. 2006. Consistent Ranking of Volatility Models. *Journal of Econometrics* 131: 97–121.

Hautsch, N., L. M. Kyj, and R. C. A. Oomen. 2012. A Blocking and Regularization Approach to High-Dimensional Realized Covariance Estimation. *Journal of Applied Econometrics* 27: 625–645.

Huang, X., and G. Tauchen. 2005. The Relative Contribution of Jumps to Total Price Variance. *Journal of Financial Econometrics* 3: 456–499.

Kapetanios, G., and M. H. Pesaran. 2004. "Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns." CESifo Discussion paper 1416.

King, M., E. Sentana, and S. Wadhwani. 1994. Volatility and Links between National Stock Markets. *Econometrica* 62: 901–933.

Liu, Y. L., J. A. Patton, and K. Sheppard. 2015. Does Anything Beat 5-Minute RV? A Comparison of Realized Measures across Multiple Asset Classes. *Journal of Econometrics* 187: 293–311.

Mincer, J., and V. Zarnowitz. 1969. “The Evaluation of Economic Forecasts.” In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pp. 3–46. National Bureau of Economic Research, Inc.

Müller, U. A., M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen, and J. Ward. 1993. “Fractals and Intrinsic Time - A Challenge to Econometricians.” Discussion paper.

Ng, V., R. F. Engle, and M. Rothschild. 1992. A Multi-Dynamic-Factor Model for Stock Returns. *Journal of Econometrics* 52: 245–266.

Patton, A. J. 2011. Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics* 160: 246–256.

Patton, A. J., and K. Sheppard. 2009. Optimal Combinations of Realised Volatility Estimators. *International Journal of Forecasting* 25: 218–238.

Patton, A. J., and K. Sheppard. 2015. Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility. *Review of Economics and Statistics* 97: 683–697.

Pesaran, M. H. 2004. “General Diagnostic Tests for Cross Section Dependence in Panels.” IZA Discussion paper 1240.

Pesaran, M. H. 2006. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica* 74: 967–1012.

Pesaran, M. H. 2015. Testing Weak Cross-Sectional Dependence in Large Panels. *Econometric Reviews* 34: 1089–1117.

Pesaran, M. H., A. Pick, and T. Allan. 2011. Variable Selection, Estimation and Inference for Multi-Period Forecasting Problems. *Journal of Econometrics* 164: 173–187.

Pitt, M. K., and N. Shephard. 1999. *Time Varying Covariances: A Factor Stochastic Volatility Approach*, pp. 547–570. Oxford: Oxford University Press.

Shephard, N. 2004. *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

Stock, J. H., and M. W. Watson. 2002. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics* 20: 147–162.

Vortelinos, D. I. 2017. Forecasting Realized Volatility: HAR against Principal Components Combining, Neural Networks and GARCH. *Research in International Business and Finance* 39: 824–839.