Journal of Econometrics 231 (2022) 348–360

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

From zero to hero: Realized partial (co)variances

Tim Bollerslev [a,*], Marcelo C. Medeiros [b], Andrew J. Patton [a], Rogier Quaedvlieg [c]

[a] Department of Economics, Duke University, United States
[b] Department of Economics, PUC-Rio, Brazil
[c] Department of Business Economics, Erasmus University Rotterdam, Netherlands

ARTICLE INFO

Article history:
Received 3 July 2020
Received in revised form 9 April 2021
Accepted 22 April 2021
Available online 9 November 2021

JEL classification:
C22
C51
C53
C58

Keywords:
High-frequency data
Realized variation
Volatility forecasting

ABSTRACT

This paper proposes a generalization of the class of realized semivariance and semi-covariance measures introduced by Barndorff-Nielsen et al. (2010) and Bollerslev et al. (2020a) to allow for a finer decomposition of realized (co)variances. The new "realized partial (co)variances" allow for multiple thresholds with various locations, rather than the single fixed threshold of zero used in semi (co)variances. We adopt methods from machine learning to choose the thresholds to maximize the out-of-sample forecast performance of time series models based on realized partial (co)variances. We find that in low dimensional settings it is hard, but not impossible, to improve upon the simple fixed threshold of zero. In large dimensions, however, the zero threshold embedded in realized semi covariances emerges as a robust choice.

1. Introduction

We are extremely grateful for the opportunity to contribute to this special issue in honor of Francis X. Diebold. Frank's wide-ranging seminal contributions have importantly shaped the research frontier in time series and financial econometrics over the past four decades.

As a case in point, Frank was among a small group of researchers in the 1980s to first embrace the now widely-accepted empirical fact that financial market volatility is both time-varying and highly predictable (Diebold, 1986). He was also among the first to explicitly recognize that the variation in volatilities are strongly connected across assets and markets, and this connectedness has important implications not only for forecasting, but also for financial asset pricing and the practice of risk management. The factor ARCH class of models first proposed in Diebold and Nerlove (1989), in particular, provides an especially convenient framework for succinctly characterizing these joint dynamic dependencies, and related parametric GARCH and stochastic volatility type models remained the norm for modeling and forecasting time-varying variances and covariances up until the early 2000s (see, e.g., the survey in Andersen et al. (2006)).

Meanwhile, motivated by the increased availability of reliable high-frequency data for a host of financial assets, Andersen et al. (2003) proposed the use of traditional ARMA type procedures for directly modeling the dynamic dependencies

* Correspondence to: Department of Economics, Duke University, 213 Social Sciences Building, Box 90097, Durham, NC 27708-0097, United States.
E-mail addresses: boller@duke.edu (T. Bollerslev), mcm@econ.puc-rio.br (M.C. Medeiros), andrew.patton@duke.edu (A.J. Patton), quaedvlieg@ese.eur.nl (R. Quaedvlieg).

https://doi.org/10.1016/j.jeconom.2021.04.013
0304-4076/

in daily and lower frequency so-called realized volatilities constructed from higher frequency intraday returns (see also Andersen and Bollerslev (1998), Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) for additional motivation and theoretical justification). This idea spurred somewhat of a paradigm shift, and realized volatilities are now widely-used in both academia and financial practice, with the "heterogenous autoregressive" (HAR) model of Corsi (2009) having emerged as the most commonly-used specification (see also Andersen et al. (2007) for one of the earliest uses of the HAR model).

The "classic" realized volatility measure is computed simply as the sum of squared high frequency returns, and thus does not contain any information from the signs of these returns. Motivated by the downside risk measures sometimes employed in asset pricing (dating back to Roy (1952) and Markowitz (1959)), Barndorff-Nielsen et al. (2010) first proposed decomposing the total realized variation into separate up and down semivariance measures defined by the summation of the squared positive and negative high-frequency returns.

Underscoring the practical usefulness of this decomposition from a volatility forecasting perspective, Patton and Sheppard (2015) find that embedding the realized semivariances within the HAR modeling framework results in significantly improved volatility forecasts for aggregate equity index returns, albeit only modest gains for individual equities. Bollerslev et al. (2020a) extend this idea to a multivariate setting, and show how realized covariance matrices may be similarly decomposed into four realized semicovariance components based on the signs of the underlying high-frequency returns. They also go on to empirically demonstrate how this four-way decomposition may be used in the construction of improved covariance matrix and portfolio variance forecasts.

Set against this background, we propose an extension of the realized semivariance and semicovariance measures previously considered in the literature. Semi-variances and covariances stratify returns by their sign; that is, using a "threshold" of zero. Our work in this paper is directly inspired by a July 19th, 2018, post on Frank's *No Hesitation* blog.[1] After attending the 2018 NBER Summer Institute, Frank shared his thoughts on the intersection between two of the presentations: Gu et al. (2020) on the use of machine-learning methods to asses the risk premiums in financial markets, and the aforementioned (Bollerslev et al., 2020a) paper on realized semicovariances. Contrasting the agnostic nature of the machine learning methods used in the former paper with the economically-motivated zero threshold employed in the second paper, Frank concluded his post with the following musings:

"*So here's a fascinating issue to explore: Hit the Bollerslev et al. realized covariance data with machine learning (in particular, tree methods like random forests) and see what happens. Does it "discover" the Bollerslev et al. result? If not, why not, and what does it discover? Does it improve upon Bollerslev et al.?*"

This special issue in honor of Frank's many contributions, including those in volatility forecasting, seems to us to be the perfect opportunity to answer Frank's challenge.

In so doing, we firstly propose a new class of *partial* (co)variance measures, which allow for tresholds other than zero, and for multiple thresholds. We use a data-driven approach to determine the optimal number and location of the thresholds, based on the out-of-sample forecasting performance of a model using partial covariances with a given set of thresholds. Our algorithm for determining the thresholds is inspired by tree methods, with the objective of dividing the support of the joint return distribution into sub-regions to improve the quality of the forecasts. However, our approach differs importantly from off-the-shelf tree methods in the sense that the branching occurs at a different frequency from the dynamic modeling: the thresholds apply to the high-frequency intraday returns used in constructing the realized partial covariance measures, while performance is assessed using the resulting daily volatility forecasts.

Our empirical analysis begins with a detailed investigation of the optimal choice of thresholds in univariate volatility forecasting models. We use high frequency data on a set of highly liquid Dow Jones Industrial Average (DJIA) individual stocks, as well as the S&P 500 aggregate market index. We find that when restricting partial variances to a single threshold, zero does indeed emerge as the hero: models based on semivariances tend to do as well or better than more flexible partial variance-based models. However, allowing for two or more thresholds leads to significant forecast improvements. The locations of the optimal thresholds in the two-threshold case are typically in the left and right tails of the distribution of the high-frequency returns, suggesting that optimally chosen partial variances do not simplify to semivariances.

We next look at multivariate volatility forecasting, with dimensions ranging from two to fifty. To accommodate the increased dimension, we rely on a much wider set of S&P 500 constituent stocks. As in the univariate case, when the partial covariances are restricted to a single threshold, zero emerges as the optimal choice. When allowing for multiple tresholds, we find that an "enhanced" semicovariance model emerges as the winner: this model has one threshold at zero, as in semicovariances, and another flexibly-chosen threshold, which tends to be in the left tail of the intraday return distribution. Applying partial covariances to returns with jumps removed, we find that a semicovariance-based model re-emerges as the preferred specification.

The plan for the rest of the paper is as follows. Section 2 introduces partial variances and covariances, and relates them to semivariances and covariances. The univariate and multivariate volatility forecasting results are presented in

[1] https://fxdiebold.blogspot.com/2018/07/machine-learning-volatility-and.html.

Sections 3 and 4, respectively. Section 5 concludes. The online Supplemental Appendix contains additional details and empirical results.

2. Partial variances and covariances

The realized semivariances of Barndorff-Nielsen et al. (2010) decompose realized variance into two separate parts associated with positive and negative high-frequency returns. These measures naturally link to up- and down-side variances, and the long history of these types of measures in finance (e.g., Hogan and Warren (1972, 1974), Fishburn (1977) and Ang et al. (2006)). Bollerslev et al. (2020a) extend the univariate semivariance measures to the multivariate context with the notion of realized semicovariances.

To help fix ideas, let $r_{t,k,i}$ denote the return over the kth intradaily time-interval on day t for asset i. Denote the $N \times 1$ vector of returns, over equally-spaced intra-daily intervals, for the set of N assets by $\mathbf{r}_{t,k} \equiv \left[r_{t,k,1}, \ldots, r_{t,k,N}\right]'$. Further, let $n(x) \equiv \min\{x, 0\}$ and $p(x) \equiv \max\{0, x\}$ denote the element-wise positive and negative elements of the real vector x, so that $\mathbf{r}_{t,k} = n(\mathbf{r}_{t,k}) + p(\mathbf{r}_{t,k})$. The three $N \times N$ daily realized semicovariance matrices are then defined by:

$$\begin{aligned} \mathbf{N}_t &\equiv \sum_{k=1}^{m} n(\mathbf{r}_{t,k})\, n(\mathbf{r}_{t,k})', \\ \mathbf{M}_t &\equiv \sum_{k=1}^{m} n(\mathbf{r}_{t,k})\, p(\mathbf{r}_{t,k})' + p(\mathbf{r}_{t,k})\, n(\mathbf{r}_{t,k})', \\ \mathbf{P}_t &\equiv \sum_{k=1}^{m} p(\mathbf{r}_{t,k})\, p(\mathbf{r}_{t,k})' \ , \end{aligned} \tag{1}$$

where m denotes the total number of intradaily return observations available per day. It follows readily that,

$$\mathbf{RCOV}_t \equiv \sum_{k=1}^{m} \mathbf{r}_{t,k}\mathbf{r}_{t,k}' = \mathbf{N}_t + \mathbf{M}_t + \mathbf{P}_t. \tag{2}$$

Thus, the semicovariances provide an exact, additive decomposition of the conventional daily realized covariance matrix into "negative", "positive", and "mixed" semicovariance matrices. Since $\mathbf{RCOV}_t$, $\mathbf{N}_t$ and $\mathbf{P}_t$ are all defined as sums of vector outer-products, these matrices are all positive semidefinite. On the other hand, since the diagonal elements of $\mathbf{M}_t$ are zero by definition, this matrix is necessarily indefinite.[2] Note also that for a single asset ($N = 1$), $\mathbf{P}_t$ and $\mathbf{N}_t$ are the positive and negative (scalar) realized semivariance measures, while $\mathbf{M}_t \equiv 0$.

The threshold at zero that underlies the semivariation measures may seem somewhat arbitrary from a statistical perspective. However, from an economic perspective the choice of zero is naturally motivated by notions of loss aversion and prospect theory (Kahneman and Tversky, 1979), which have been extensively studied in the empirical finance literature, see, e.g., Ang et al. (2006) and Bollerslev et al. (2020c) and the many references therein. Correspondingly, econometric volatility forecasts based on zero-threshold semi(co)variation type measures have also been found to outperform the forecasts from otherwise comparable forecasting models that do not exploit such measures, see, e.g., Kroner and Ng (1998), Cappiello et al. (2006), Patton and Sheppard (2015) and Bollerslev et al. (2020b).

We generalize semi(co)variances to *partial* (co)variances by considering the use of multiple thresholds, rather than a single threshold at zero. In particular, we use the fact that

$$\mathbf{r}_{t,k} = \sum_{g=1}^{G} f_g(\mathbf{r}_{t,k}) \tag{3}$$

represents an exact decomposition of the kth intradaily return vector $\mathbf{r}_{t,k}$ into G components based on the partition functions f_g:

$$f_g(\boldsymbol{x}) = \boldsymbol{x} \circ \mathbf{1}\{\mathbf{c}_g < \boldsymbol{x} \leq \mathbf{c}_{g+1}\}, \tag{4}$$

i where the tresholds satisfy $\mathbf{c}_1 = -\infty$, $\mathbf{c}_{G+1} = \infty$, and $\mathbf{c}_{g-1} \leq \mathbf{c}_g \ \forall\, g$. Analogous to the realized semicovariances presented in Eq. (1), realized partial covariances are then defined by:

$$\boldsymbol{PCOV}_t^{(g,g')} \equiv \sum_{k=1}^{m} f_g(\mathbf{r}_{t,k}) f_{g'}(\mathbf{r}_{t,k})' \tag{5}$$

[2] Since the two matrices that define $\mathbf{M}_t$ are each other's transpose, we combine them into a single "mixed" component. In situations where the ordering of the assets matter, these two terms could be treated differently resulting in two different "mixed" components.

Combining Eqs. (2), (3) and (5) we easily see that realized partial covariances provide a finer decomposition of the realized covariance matrix:

$$\begin{aligned} \boldsymbol{RCOV}_t &\equiv \sum_{k=1}^{m} \mathbf{r}_{t,k}\mathbf{r}_{t,k}' \\ &= \sum_{k=1}^{m} f_1(\mathbf{r}_{t,k})f_1(\mathbf{r}_{t,k})' + f_1(\mathbf{r}_{t,k})f_2(\mathbf{r}_{t,k})' + \cdots + f_G(\mathbf{r}_{t,k})f_G(\mathbf{r}_{t,k})' \\ &= \sum_{g=1}^{G}\sum_{g'=1}^{G} \boldsymbol{PCOV}_t^{(g,g')} \end{aligned} \tag{6}$$

When the ordering of the assets is arbitrary, as in our empirical analysis below, it may again make sense to combine the "mixed" partial covariances (where $g \neq g'$), leading to a decomposition of $\boldsymbol{RCOV}$ into $G(G+1)/2$ partial covariance matrices rather than G^2 such matrices. This, of course, is completely analogous to the definition of $\boldsymbol{M}$ in Eq. (1).

Note that when the cross-sectional dimension is one, i.e., we only consider a single asset, the $G(G-1)/2$ "mixed" partial covariance terms reduce to zeros and we have only G realized partial variances:

$$RV_t \;=\; \sum_{k=1}^{m} f_1(r_{t,k})^2 + \cdots + f_G(r_{t,k})^2 \;=\; \sum_{g=1}^{G} PV_t^{(g)} \tag{7}$$

In both the univariate and multivariate cases, when $G = 2$ and $f_1(\mathbf{x}) = n(\mathbf{x})$ and $f_2(\mathbf{x}) = p(\mathbf{x})$, realized partial (co)variances reduce to realized semi (co)variances. When $G = 1$, realized partial (co)variances trivially reduce to the standard realized (co)variances.

The freedom to choose the number and location of the thresholds in the realized partial covariances affords a great deal of flexibility. Inspired by the widespread adoption of machine learning techniques in many areas of empirical economics, in our empirical work below we rely on algorithms rooted in regression trees to determine the number and locations of the thresholds. A key difference, however, between off-the-shelf tree methods as in Breiman et al. (1984) and the approach that we use here is that our approach operates on two different time scales, i.e., the forecasting model is estimated at the daily frequency, while the thresholds are determined at the high-frequency intraday level. By contrast, traditional regression tree models are estimated with data sampled at a single frequency. Moreover, the forecasting models estimated here are all linear in the parameters, while off-the-shelf tree-based methods are inherently non-linear. Modern regression-based tree methods, such as random forests, also typically combine a large number of different trees to construct the forecasts, rendering the interpretability of the branches difficult, while we purposely rely on a single "tree" thereby affording easier interpretation of the results.

3. Optimal partial variances for volatility forecasting

We begin our empirical investigations with an analysis of the optimal choice of threshold in univariate volatility forecasting models. This analysis helps guide our analysis in the next section of larger dimensional partial covariance matrices.

All of the empirical results in this section are based on high-frequency returns for the S&P 500 SPY ETF and the Dow Jones Industrial Average constituent stocks. The sample period starts on April 21, 1997, one thousand trading days before the April 9, 2001 decimalization of the NASDAQ, and ends on December 31, 2013, for a total of 4202 observations. The raw tick-by-tick data was obtained from the TAQ database. To help guard against complications associated with market microstructure "noise", we follow the advice of Liu et al. (2015) and rely on a 5-min sampling frequency in the construction of the daily realized measures.[3] The same data and 5-min sampling frequency was also previously used in Bollerslev et al. (2016), and we refer to that article for additional justification and details about our "cleaning" of the data.

To accommodate any large persistent movements in return volatility, we rely time-varying, asset-specific quantiles of the volatility-standardized return distributions as thresholds when actually implementing the realized partial variation measures. Specifically, for each asset, we obtain the quantiles as

$$c_{t,g} = RV_t^{1/2} \cdot \hat{Q}\left(\frac{r_{t,k}}{RV_t^{1/2}}\;;\; q_g\right), \text{ for } g = 2, 3, \ldots, G, \tag{8}$$

where $\hat{Q}\left(z\;;\;q_g\right)$ returns the sample q_g quantile of z.

Since the median of almost all high-frequency return distributions is nearly indistinguishable from zero, when $q_g = 0.5$ partial (co)variances effectively correspond to semi(co)variances. The use of other and/or additional quantiles will, of

[3] Diebold and Strasser (2013) discusses various underlying economic mechanisms that may possibly account for this "noise".

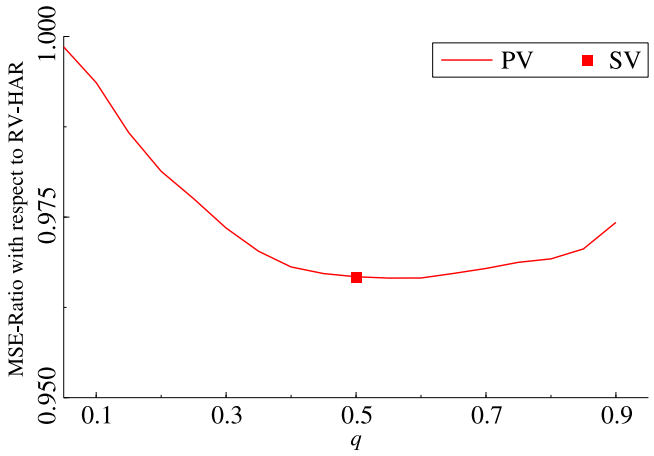

Fig. 1. PV(2)-HAR forecasting performance. *Note:* The figure shows the out-of-sample MSE of the PV(2)-HAR model for the S&P 500 SPY ETF relative to the MSE for the RV-HAR model as a function of the quantile threshold used in the definition of the partial variances.

course, result in distinctly different partial variation measures emphasizing other parts of the time-varying joint return distributions.

The Heterogeneous Autoregressive (HAR) model of Corsi (2009) has arguably emerged as the benchmark model for realized volatility-based forecasting, see, e.g., Bollerslev et al. (2018) and the many references therein. The basic RV-HAR model relies on the lagged daily, weekly and monthly realized volatilities to linearly forecast the future volatility:

$$RV_t = \phi_0 + \phi_1 RV_{t-1} + \phi_2 RV_{t-1|t-5} + \phi_3 RV_{t-1|t-22} + \epsilon_t, \tag{9}$$

where $RV_{t-1|t-h} = h^{-1} \sum_{i=1}^{h} RV_{t-i}$. An important extension of the RV-HAR model, due to Patton and Sheppard (2015), decomposes the lagged daily realized variance into the two semivariances.[4] We term this the SV-HAR model:

$$RV_t = \phi_0 + \phi_1^+ SV_{t-1}^+ + \phi_1^- SV_{t-1}^- + \phi_2 RV_{t-1|t-5} + \phi_3 RV_{t-1|t-22} + \epsilon_t, \tag{10}$$

where $SV_t^+ = \sum_{k=1}^{m} p(r_{t,k})^2$ and $SV_t^- = \sum_{k=1}^{m} n(r_{t,k})^2$. In empirical applications of the SV-HAR model, the SV_t^- typically appears relatively more important than the SV_t^+ term.

In a direct extension of the SV-HAR model, we define the class of PV(G)-HAR models based on the realized partial variances:

$$RV_t = \phi_0 + \sum_{g=1}^{G} \phi_1^g PV_{t-1}^{(g)} + \phi_2 RV_{t-1|t-5} + \phi_3 RV_{t-1|t-22} + \epsilon_t. \tag{11}$$

In parallel to the SV-HAR model in (10) we only decompose the daily lag to allow for more clear-cut comparisons concerning the choice of threshold(s) and the forecast gains from freely choosing the thresholds.

We begin our empirical analyses with a simple question: Do the forecasts from any PV(2)-HAR model beat the forecasts from the semivariance-based SV-HAR model? The PV(2)-HAR model, of course, collapses to the SV-HAR model when the threshold is fixed at zero, corresponding approximately to choosing the median as the partial variance threshold. Fig. 1 plots the out-of-sample MSEs for the one-day-ahead forecasts for the S&P 500 portfolio, based on daily-updated rolling sample estimation windows of 1000 observation, as a function of the quantile used in the construction of the partial variances. As the figure shows, the lowest MSE is obtained close to the median threshold, corresponding to the SV-HAR model. Hence, for a PV(G)-HAR type model to convincingly outperform the SV-HAR model, it must involve more than one threshold ($G > 2$), or a dynamically varying choice of threshold(s).

To further investigate this, we next extend the models underlying Fig. 1 to accommodate time-varying threshold(s). As above, we rely on rolling estimation windows of 1000 observations for forecasting the one-day-ahead volatility. We allow the grid of quantiles to range from 0.05 to 0.95 in increments of 0.05. We select the threshold, or combination of thresholds, as the quantiles that minimize the average of the MSE based on 5-fold cross-validation (CV).[5] More specifically, for each model, we estimate the model based on four of the five sub-samples and compute the MSE for the fifth subsample.

[4] The weekly and monthly realized variances can, of course, be decomponsed into their positive and negative semivariance components in a similar manner. However, Patton and Sheppard (2015) report the largest gains came from decomposing only the daily lag.

[5] k-fold CV methods are widely used for model selection in statistics and machine learning (Arlot and Celisse, 2010; Cerqueira et al., 2020). Bergmeir et al. (2018) provide compelling evidence that CV can also be effectively used to evaluate time-series models. Additional details on the specific CV procedure that we use here are provided in the Supplemental Appendix S1.

Table 1
Univariate models: Unconditional forecasting performance.

	RV	SV	PV(2)	PV(3)	PV(4)	PV(G*)
Panel A: S&P500						
MSE	2.5186	2.4345	2.4388	2.3923	2.8534	2.3923
p-val. dm_{RV}		0.164	0.176	0.049	0.674	0.050
p-val. dm_{SV}			0.821	0.011	0.715	0.011
p-val. MCS	0.404	0.132	0.198	1.000	0.587	0.892
QLIKE	0.1387	0.1361	0.1360	0.1347	0.2017	0.1348
p-val. dm_{RV}		0.005	0.003	0.001	0.994	0.001
p-val. dm_{SV}			0.345	0.072	0.996	0.079
p-val. MCS	0.041	0.396	0.396	1.000	0.082	0.753
Panel B: Individual stocks						
$\overline{MSE}$	14.982	14.886	14.888	14.421	15.600	14.421
#sig. dm_{RV}		10	13	24	0	24
#sig. dm_{SV}			4	20	0	20
#sig. MCS	11	11	9	27	2	27
$\overline{QLIKE}$	0.1547	0.1532	0.1529	0.1494	0.1828	0.1494
#sig. dm_{RV}		18	19	26	0	26
#sig. dm_{SV}			5	21	0	21
#sig. MCS	3	7	7	27	0	27

Note: The table reports the forecasting performance of the different models. The top panel shows the results for the S&P 500. The bottom panel reports the average loss and 5% rejection frequencies of the Diebold–Mariano tests for each of the individual stocks. The one-sided tests between PV-HAR against RV-HAR and SV-HAR are denoted by dm_{RV} and dm_{SV}, respectively. MCS denotes the *p*-value of that model being in the Model Confidence Set, or the number of times that model is in the 80% Model Confidence Set. PV(G*) dynamically chooses the best among the RV-HAR, SV-HAR and HAR-PV(G) models with 2, 3 or 4 thresholds.

We then choose the quantile(s) that minimize the average MSE over the five different sub-samples. We consider four versions of the PV(G)-HAR model: three using a fixed number of thresholds (G=1,2,3,4), and a version (denoted G*) in which we allow for a varying number of thresholds determined through CV. The latter model thus uses between zero and three thresholds, such that it nests all the fixed threshold variants, as well as the basic RV-HAR and SV-HAR models.

Table 1 summarizes the forecasting results for the S&P 500 portfolio and the results obtained across all of the individual DJIA stocks.[6] In addition to the PV(G)-HAR models, we also report the results for the benchmark RV-HAR and SV-HAR models. First consider the model performance in terms of MSE and QLIKE.[7] For both the S&P500 and the individual stocks, the PV(G)-HAR models with 2, 3 or a variable number of thresholds (G*) outperform the basic HAR model. Compared to the SV-HAR however, the PV(2)-HAR tends to perform slightly worse, while the PV(3)-HAR and PV(G*)-HAR offer clear improvements. The PV(4)-HAR performs extremely poorly, both for the S&P500 and the individual stocks, indicating over-fitting and too fine a partition of the realized variance.

To formally assess the performance of the models, each of the panels presents results of Diebold and Mariano (1995) tests between the RV-HAR, the SV-HAR and the remaining models. For the S&P500, we present the p-values, while for the 28 individual DJIA stocks, we present the number of rejections at the 5% level. While the SV-HAR clearly improves over RV-HAR based on the numerical losses, statistical significance is often lacking. The improvements is not significant for the S&P500 and 18 out of 28 stocks, based on MSE, while the results based on QLIKE are stronger, being significant for the S&P500 and 18 stocks. On the other hand, the PV(3) and PV(G*) models significantly outperform the RV-HAR in almost all cases. They also tend to significantly improve over the SV-HAR. This is further corroborated by the final row of each panel, which presents p-values and frequency of inclusion of the 80% Model Confidence Set (MCS) of Hansen et al. (2011). For the S&P500, the PV(3) and PV(G*) are both quite convincingly included in the MCS, as evidenced by their high p-values. Across the individual stocks, we also observe that these same two models are almost always included in the MCS. By comparison, the RV-HAR and SV-HAR models are included in the MCS for less than half of the firms.[8]

The forecasting performance of the PV-HAR models discussed above demonstrated two main points. First, improving over SV-HAR is difficult. When restricted to a single threshold, consistently beating the zero threshold is very challenging. Rather, to get close to the performance of the SV-HAR, we need two thresholds, with further improvements available by dynamically selecting the thresholds using cross-validation. To help understand these findings, it is instructive to consider the models that were actually selected by the PV(G*)-HAR approach. Table 2 presents summary statistics on the number and locations of the quantiles selected. The first set of columns show that for the vast majority of stocks and most of

6 Parameter estimates for the different models are summarized in Supplemental Appendix S3

7 QLIKE loss is computed as $QLIKE_t \equiv RV_t/h_t - \log(RV_t/h_t) - 1$, where h_t denotes the forecast.

8 As a robustness check, we repeated the entire analysis minimizing QLIKE in the CV step, resulting in very similar findings and forecasting performance. These results are reported in Supplemental Appendix S2.

Table 2
Univariate models: Quantile selection.

	PV(G*) Model Selection				Unc. Threshold Selection			Cond. Threshold Selection		
	SV	PV(2)	PV(3)	PV(4)	PV(2)	PV(3)		PV(2)	PV(3)	
					$\bar{q}_2$	$\bar{q}_2$	$\bar{q}_3$	$\bar{q}_2$	$\bar{q}_2$	$\bar{q}_3$
S&P 500										
	0.000	0.001	0.999	0.000	0.78	0.14	0.70	0.67	0.14	0.70
Stocks										
5% Quantile	0.000	0.000	0.980	0.000	0.43	0.06	0.69	0.73	0.06	0.69
Mean	0.001	0.004	0.995	0.000	0.64	0.09	0.75	0.86	0.08	0.75
Median	0.000	0.000	0.998	0.000	0.64	0.08	0.76	0.95	0.08	0.76
95% Quantile	0.007	0.019	1.000	0.000	0.85	0.15	0.85	0.95	0.14	0.85

Note: The table reports the quantile selection of the PV(G) models. The first set of columns show how often each number of thresholds are selected in the PV(G*) model. The second set of columns summarizes the thresholds for the PV(2) and PV(3) models. The final set of columns show the same descriptives, conditional on the specific model being selected by the PV(G*) model.

the time, the model is simply the PV(3) model. However, for a few of the stocks and certain time periods, only a single thresholds is selected. The second set of columns provide summary statistics on the time-series averages of the selected threshold, as well as their distributions across stocks. For the single threshold model, the threshold is typically above the median, with an average of around 0.80 for the market index, and 0.66 for the individual stocks. When allowing for two thresholds, as in the PV(3) model, the high thresholds are typically accompanied by an additional threshold in the lower tail of the distribution, ranging roughly between the 0.05 and 0.15 quantiles. The final set of columns report the average quantile, conditional on PV(G*) selecting the single or double threshold model. As these results show, for the individual stocks the single threshold model is almost exclusive chosen during times when the optimally selected threshold is relatively high.

The time-series averages of the selected quantiles reported in Table 2 only tell part of the story. Going one step further, Fig. 2 plots the cross-sectional averages and distributions over time. The top row presents the results for the S&P 500 SPY ETF, while the bottom row presents the results averaged across the individual stocks. In the single threshold case, the quantiles selected for the market index are typically fairly high. However, they are closer to the median around times of financial difficulties, starting with the recession in the early 2000 s and subsequently during the financial crisis of 2008. Interestingly, the value of the cross-validated threshold for the market already dropped in early 2007, preceding the financial crisis and the subsequent dramatic rise in the overall level of volatility. The results for the PV(3) model (with two thresholds) shown in the top-right panel reveal similar dynamics, with both the upper and lower thresholds converging to the median during the financial crisis. Compared to the results for the S&P 500, the values of the selected quantiles averaged across the DJIA stocks appear quite stable through time. In the single threshold setting, the average quantile is typically around 0.65. When allowing for two thresholds, a very low quantile combined with a fairly high quantile emerge as the best choice.

The selection of two thresholds, one high and one low, is reminiscent of the HAR-J model of Andersen et al. (2007), which decomposes the realized variance into separate continuous- and jump-variation components. However, whereas the HAR-J model is based on the separately estimated continuous and jump components, where the latter combines the variation stemming from both positive and negative jumps, our empirical results suggest that the direction of the jumps actually matters, and that it is primarily the variation stemming from especially large negative jumps that should be "filtered out" and treated differently in the construction of the forecasts.[9]

Meanwhile, the challenge posed by Frank in his *No Hesitation* blog relates explicitly to the forecast improvements afforded by the use of the semicovariance measures reported in Bollerslev et al. (2020a). Hence, to directly address Frank's questions, we next extend the above univariate analyses to a multivariate context.

4. Multivariate partial covariance models

To allow for the construction of large dimensional portfolios, we expand our previous sample of DJIA stocks to all stocks that were part of the S&P 500 index during the 1993–2014 sample period. Removing stocks with fewer than 2,000 daily observations, this results in a total of 749 unique firms. We consider equally portfolios comprised of up to 50 stocks, with all of the results reported below based on 100 such randomly selected portfolios of a given size. Since the use of more extreme thresholds may easily result in many zero-valued realized partial covariances, we reduce our grid search in the multivariate setting to thresholds ranging between the {0.1,0.2, …,0.9} volatility-standardized quantiles.

[9] The Supplemental Appendix S6 contains a more thorough analysis of the forecast performance of the PV(G)-HAR model based on the Conditional Superior Predictive Ability (CSPA) test of Li et al. (2021). As that analysis also shows, the improved performance can only be partly attributed to the model isolating jumps.

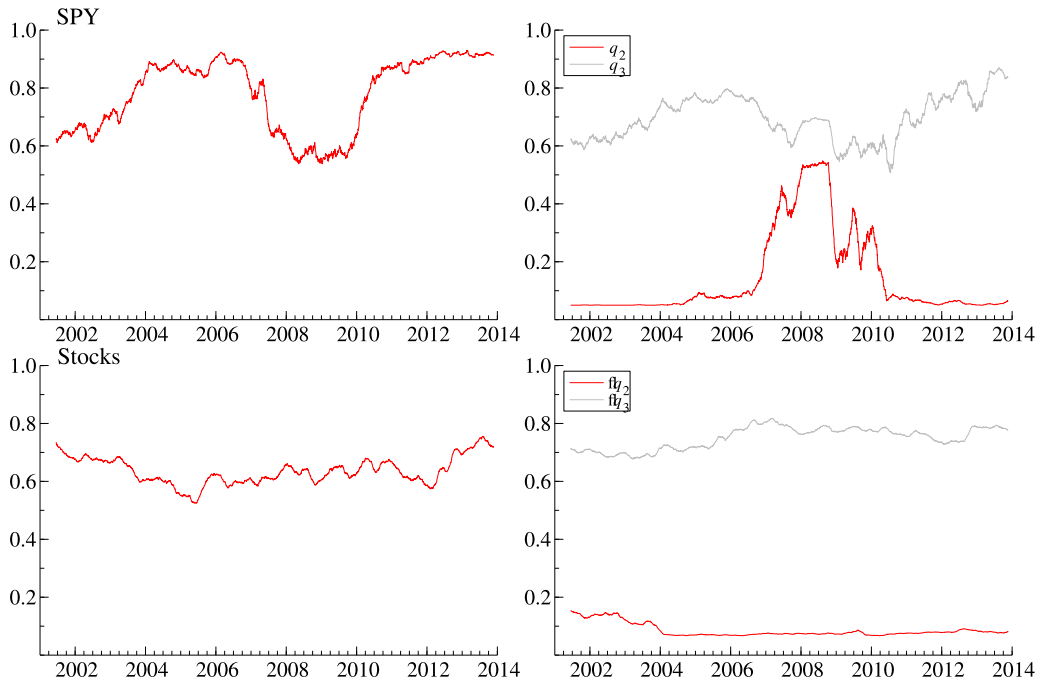

Fig. 2. Univariate models: The time-series of selected threshold quantiles. *Note:* The top row presents the selected quantiles in the PV(2)- and PV(3)-HAR models for the S&P 500 SPY ETF realized variance series. The bottom row shows the average selected threshold across the DJIA stocks for the same two models. The time-series are smoothed using a $[t-25:t+25]$-moving average.

The complications associated with market microstructure "noise" is further compounded in the multivariate setting due to asynchronous prices, and the so-called Epps effect. To address this issue, for the stoicks analyzed here, which includes some smaller-cap and less liquid stocks, Bollerslev et al. (2020a) advocated the use of a coarse 15-min sampling frequency in the calculation of their realized semicovariance measures. While this sparse sampling effectively alleviates the biases stemming from market microstructure "noise", it obviously also reduces the accuracy of the estimates compared to the estimates that would obtain with more finely sampled returns in the absence of any "noise". The literature has proposed a number of alternative procedures to more efficiently address these issues, including the kernel-based approach (Barndorff-Nielsen et al., 2011), pre-averaging (Christensen et al., 2010), and subsampling (Zhang et al., 2005). Out of these, the multivariate realized kernel has arguably emerged as the most popular "noise" robust approach for realized covariance matrix estimation. Accordingly, we rely on realized kernel-based estimates constructed from 1-min returns, denoted $\boldsymbol{MK}_t$ below, as our object of interest and multivariate measure appearing on the "left-hand-side" in the various forecasting models.

The smoothing out of the "noise" in the high-frequency returns that underlie the realized kernel estimates is at odds with the intuition and motivation for the high-frequency return-based decompositions of the covariance matrix that underlie the definitions of the semicovariance and new partial covariance measures. Hence, to maintain interpretability of the new realized measures that we use as explanatory variables on "right-hand-side" in our forecasting models, we instead rely on a subsampling scheme, in which we subsample 15-min returns at the 1-min frequency: the 15-min returns used in determining the thresholds for the partial covariance being effectively void of the detrimental impact of microstructure "noise", while averaging over the 15 different subsampled estimates increase the precision.

More formally, let m denote the total number of high-frequency returns available per day at the highest sampling frequency. Define $M \equiv m/S$, where S refers to the accumulation factor, or number of subgrids (for simplicity assume that M is an integer). Let $\mathbf{r}^s_{t,k}$ denote the sequence of $k = 1, \ldots, M-1$, lower-frequency intraday returns originating at the sth subgrid; i.e., $\mathbf{r}^s_{t,k} = \sum_{i=1}^{S} \mathbf{r}_{t,kS+s+i}$. The subsampled realized partial covariances are then simply defined by:

$$\widetilde{\boldsymbol{PCOV}}_t^{(g,g')} \equiv \frac{1}{S}\sum_{s=1}^{S}\left[\frac{M}{M-1}\sum_{k=1}^{M-1} f_g(\mathbf{r}^s_{t,k}) f_{g'}(\mathbf{r}^s_{t,k})'\right], \tag{12}$$

where the partion functions f_g are applied grid-by-grid. Effectively, this is simply the average of the previously defined realized partial covariances estimated on each of the S overlapping time grids. For notational convenience, we will drop the explicit use of tilde in the sequel, and simply refer to the subsampled estimates by $\boldsymbol{PCOV}_t$. Similarly, we will refer to the subsampled versions of the realized covariance and semicovariance matrices obtanied as special cases of (12) by $\boldsymbol{RCOV}_t$ and $\boldsymbol{SCOV}_t$ for short.

Table 3
Multivariate models: PCOV(G*)-HAR model selection.

N	1	2	5	10	20	50
RCOV	0	0	0	0	0	0
SCOV	6	4	0	0	0	0
PCOV(2)	0	0	0	0	0	0
SCOV$^+$	30	36	38	80	95	100
PCOV(3)	64	60	62	20	5	0

Note: The table reports the frequency, out of 100 random portfolios, with which various threshold combinations are selected, as a function of the portfolio dimension N. RCOV implies no threshold, SCOV is the single zero threshold, PCOV(2) indicates a single non-zero threshold, SCOV$^+$ refers to two thresholds, one of which is zero, while PCOV(3) denotes two thresholds, none of which equal zero.

Mirroring our univariate forecast analyses, we take the basic multivariate HAR model applied to the *vech* of the covariance matrices as our benchmark model,

$$\begin{aligned}\text{vech}(\boldsymbol{MK}_t) = \boldsymbol{\phi}_0 &+ \phi_1 \text{vech}(\boldsymbol{RCOV}_{t-1}) \\ &+ \phi_2 \text{vech}(\boldsymbol{RCOV}_{t-1|t-5}) + \phi_3 \text{vech}(\boldsymbol{RCOV}_{t-1|t-22}) + \epsilon_t,\end{aligned} \tag{13}$$

where $\boldsymbol{\phi}_0$ is an $N(N+1)/2-$vector, and the remaining coefficients are assumed to be scalar. We will refer to this as the RCOV-HAR model. The SCOV-HAR model extends the benchmark model to allow for different influences of the three daily semicovariance matrices:[10]

$$\begin{aligned}\text{vech}(\boldsymbol{MK}_t) = \boldsymbol{\phi}_0 &+ \phi_1^N \text{vech}(\boldsymbol{N}_{t-1}) + \phi_1^M \text{vech}(\boldsymbol{M}_{t-1}) + \phi_1^P \text{vech}(\boldsymbol{P}_{t-1}) \\ &+ \phi_2 \text{vech}(\boldsymbol{RCOV}_{t-1|t-5}) + \phi_3 \text{vech}(\boldsymbol{RCOV}_{t-1|t-22}) + \epsilon_t.\end{aligned} \tag{14}$$

Finally, in a parallel to the univariate models defined in (11), we consider new multivariate PCOV(G)-HAR partial covariance-based models. Guided by the univariate results, we allow for up to three partitions, resulting in a maximum of nine partial covariances, regardless of the cross-sectional dimension. Given that the ordering of our assets is arbitrary, we combine the "mixed" partial covariances, leaving us with a maximum of six partial covariances:

$$\begin{aligned}\text{vech}(\boldsymbol{MK}_t) = \boldsymbol{\phi_0} &+ \sum_{g=1}^{G} \sum_{g'=g}^{G} \phi_1^{(g,g')} \text{vech}(\overline{\boldsymbol{PCOV}}_{t-1}^{(g,g')}) \\ &+ \phi_2 \text{vech}(\boldsymbol{RCOV}_{t-1|t-5}) + \phi_3 \text{vech}(\boldsymbol{RCOV}_{t-1|t-22}) + \epsilon_t,\end{aligned} \tag{15}$$

where

$$\overline{\boldsymbol{PCOV}}_t^{(g,g')} \equiv \begin{cases} \boldsymbol{PCOV}_t^{(g,g)} & \text{for } g = g' \\ \boldsymbol{PCOV}_t^{(g,g')} + \boldsymbol{PCOV}_t^{(g',g)}, & \text{for } g \neq g'. \end{cases} \tag{16}$$

We rely on the same 5-fold CV procedure to optimally select the number and values of the time-varying thresholds. We do so by optimizing the multivariate MSE, or Frobenius norm, defined as:

$$MSE_t \equiv \text{Tr}((\boldsymbol{MK}_t - \boldsymbol{H}_t)'(\boldsymbol{MK}_t - \boldsymbol{H}_t))/N^2,$$

where $\boldsymbol{H}_t$ denotes the model forecast.

Table 3 provides a summary of the different models that were selected for the different portfolio dimensions N. Specifically, the columns titled RCOV and SCOV show the number of times out of 100 for which the CV procedure selected the model in (13) or (15), respectively. The row labelled SCOV$^+$ refers to a model that adds one additional threshold to the SCOV model (recall the SCOV model already contains a threshold at zero), while PCOV(2) and PCOV(3) denote one- or two-threshold models, where all of the thresholds differ from zero. As in the univariate case, the preferred number of thresholds is generally two: one of PCOV(3) or SCOV$^+$ is selected for between 94% and 100% of the portfolios across all values of N. Interestingly, the basic RCOV model is never selected, while the SCOV model is only selected for 6% of the randomly selected individual stocks ($N = 1$) and never for any of portfolios involving 5 or more stocks. Taken as a whole, the SCOV$^+$ model, which involves zero plus one additional flexibly-chosen threshold, emerges as the clear winner, especially for larger dimensions ($N \geq 10$) where it is nearly always the preferred model.[11]

The emergence of the SCOV$^+$ model as the preferred specification is further underscored by Fig. 3, which reports the distributions of the selected quantiles for the PCOV models across the 100 random portfolios for different portfolio sizes

[10] This model differs slightly from the specific model recommended in Bollerslev et al. (2020a), which involves a complete HAR-structure on $\boldsymbol{N}_t$, along with the monthly lag of $\boldsymbol{M}_t$. However, it more closely resembles the SV-HAR models analyzed in the previous section. It also provides a more natural benchmark against which to assess the general PCOV(G)-HAR partial covariance class of models.

[11] The actual parameter estimates for the SCOV$^+$ models, reported in Supplemental Appendix S3, further reveal that the joint most-negative partial covariance component typically stands out as the most important.

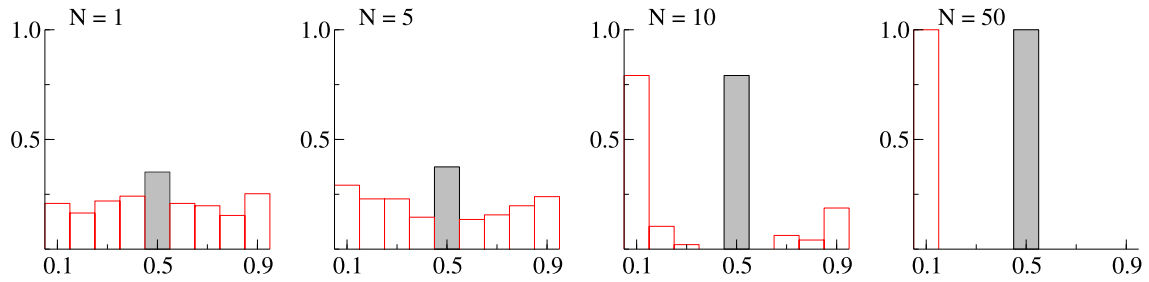

Fig. 3. Multivariate models: Quantile selection. *Note:* The figures shows the frequency with which the different quantiles are selected as the optimal threshold(s). Each column corresponds to a different cross-sectional dimension N. The shaded bar at 0.5 denotes the semicovariance threshold.

Table 4
Multivariate models: Forecasting results.

N	MSE			QLIKE		
	RCOV	SCOV	PCOV(G*)	RCOV	SCOV	PCOV(G*)
1	22.943	22.590	21.429	0.178	0.168	0.155
	(0.80)	(0.85)	(0.83)	(0.65)	(0.72)	(0.72)
2	15.987	15.619	15.350	0.425	0.465	0.439
	(0.77)	(0.83)	(0.93)	(0.44)	(0.33)	(0.81)
5	7.314	7.121	6.987	1.293	1.242	1.248
	(0.37)	(0.49)	(0.99)	(0.22)	(0.16)	(0.97)
10	4.881	4.686	4.552	3.414	3.349	2.760
	(0.13)	(0.32)	(0.98)	(0.21)	(0.21)	(0.98)
20	3.702	3.574	3.461	9.541	9.275	8.726
	(0.00)	(0.25)	(1.00)	(0.00)	(0.00)	(1.00)
50	2.896	2.837	2.821	36.478	35.824	34.454
	(0.00)	(0.00)	(1.00)	(0.00)	(0.00)	(1.00)

Note: The table provides the MSE and QLIKE for the three HAR-based models, for various cross-sectional dimensions N. The reported number is the average loss over 100 random portfolios, while the number in parentheses presents the fraction of portfolios for which the model is included in the 80% Model Confidence Set.

N. As in the univariate case, the two thresholds typically selected are zero and a low threshold. In fact, as Table 3 shows, for $N = 50$ that is the *only* model and set of thresholds selected.

The previous results all pertain to in-sample model selection and threshold estimation. We next consider out-of-sample multivariate forecasting. In parallel to the univariate analysis, we rely on rolling estimation windows of 1000 days for recursively forecasting the covariance matrices one-day-ahead.[12] Because of the high computational costs associated with the calculation of the multivariate forecasts, we select the thresholds for the PCOV models only once, using the initial training sample of the first 2000 days. Note, this differs from the univariate analysis, where we allowed for optimally chosen thresholds in each of the rolling estimation windows.

Table 4 reports the average MSE and QLIKE loss for the different models, together with their prevalence in the corresponding 80% MCSs.[13] While the average losses for the SCOV and PCOV models are typically lower than the losses for the benchmark RCOV model, the MCS does not appear to be able to convincingly discriminate between the models for portfolios of size $N \leq 10$. However, consistent with the earlier findings, the reductions in loss obtained by the threshold models, and the partial covariation-based models in particular, appear to be increasing in the cross-sectional dimension N. In fact, the RCOV model is never included in any of the MCSs for the $N \geq 20$ dimensional forecasts. Recalling the earlier results discussed in Table 3 and Fig. 3, $N = 20$ also coincides with the dimension for which the PCOV(G*)-HAR model consistently selects the SCOV$^+$ model with a low second threshold. Thus, while a variety of stock-specific thresholds appear to be optimal for the smaller dimensional portfolios, the choice of zero, possibly augmented with a low additional threshold, emerges as a "robust" best choice in large dimensions.[14]

Our findings that the PCOV(G*) models tend to select at least one threshold in the tails, suggests that the success of the partial covariance-based forecasting models may be related to "jumps". In order to further investigate this interplay between the optimal choice of threshold(s) and jumps, we consider a collection of alternative models, where in the spirit

[12] Supplemental Appendix S5 contains complementary results for forecasting the variance of portfolios using the partial covariances of the individual stocks included in the portfolios.

[13] Multivariate QLIKE loss is computed as $QLIKE_t \equiv \text{Tr}(\boldsymbol{H}_t^{-1}\boldsymbol{M}\boldsymbol{K}_t) - \log|\boldsymbol{H}_t^{-1}\boldsymbol{M}\boldsymbol{K}_t| - N$, where $\boldsymbol{H}_t$ denotes the covariance matrix forecast.

[14] Supplemental Appendix S4 reports the results from forecasting the covariance matrices of stocks within the same industry. Consistent with the idea of stock-specific thresholds, the relative performance of the PCOV(G*) models appears marginally better in that context, and the selected thresholds also differ from the SCOV$^+$ thresholds, suggesting that there is indeed some commonality in the choice of thresholds for related stocks.

Table 5
Multivariate models: PCOV(G*)†-HAR model selection.

N	1	2	5	10	20	50
RCOV†	0	0	0	0	0	0
SCOV†	6	4	10	15	21	32
PCOV(2)†	0	0	0	0	0	0
SCOV+†	45	33	57	78	79	68
PCOV(3)†	49	63	33	7	0	0

Note: The table reports the frequency, out of 100 random portfolios, with which various threshold combinations are selected, as a function of the portfolio dimension *N*, for the HAR models using continuous (partial) covariance on the right-hand side. RCOV† implies no threshold, SCOV† is the single zero threshold, PCOV(2)† indicates a single non-zero threshold, SCOV†+ refers to two thresholds, one of which is zero, while PCOV(3)† denotes two thresholds, none of which equal zero.

Table 6
Multivariate models: Forecasting results continuous (Partial) covariances.

N	MSE			QLIKE		
	RCOV†	SCOV†	PCOV(G*)†	RCOV†	SCOV†	PCOV(G*)†
1	24.431	23.975	23.915	0.184	0.194	0.178
	(0.94)	(0.86)	(0.89)	(0.68)	(0.57)	(0.61)
2	14.994	14.751	14.480	0.434	0.434	0.411
	(0.83)	(0.88)	(0.93)	(0.58)	(0.25)	(0.77)
5	7.494	7.225	7.119	1.447	1.390	1.210
	(0.53)	(0.64)	(0.96)	(0.40)	(0.18)	(0.90)
10	5.029	4.867	4.800	3.478	3.292	3.207
	(0.36)	(0.53)	(0.98)	(0.17)	(0.36)	(0.92)
20	3.959	3.809	3.743	9.905	9.400	9.165
	(0.09)	(0.34)	(1.00)	(0.09)	(0.39)	(1.00)
50	2.970	2.946	2.938	36.796	36.404	35.861
	(0.10)	(0.42)	(1.00)	(0.00)	(0.12)	(1.00)

Note: The table provides the MSE and QLIKE for the three HAR-based models, for various cross-sectional dimensions *N*, using jump-robust right-hand side variables. The reported number is the average loss over 100 random portfolios, while the number in brackets presents the fraction of portfolios for which the model is included in the 80% Model Confidence Set.

of the C-HAR model of Andersen et al. (2007), we replace the different realized measures used in predicting the future realized covariance matrices with the same measures constructed from high-frequency returns void of jumps. Following the commonly used approach in the literature, we rely on dynamic thresholds of three times the trailing scaled bipower variation, further adjusted for the strong intraday periodicity in volatility, to identify the jumps (see Bollerslev and Todorov (2011) for additional details). We refer to these "continuous" realized variation measures and versions of the different models by the † symbol, as in, e.g., PCOV†. As above, we also make use of subsampling techinques to further enhance the efficiency of all the estimates.

Looking at the resulting models selected by the 5-fold CV criteria reported in Table 5, the basic SCOV† model is selected much more frequently than the SCOV model based on the conventional measures that do not filter out jumps reported in Table 3.[15] Meanwhile, the same convergence towards the SCOV+† model evident for the conventional measures remains true, although the removal of the variation stemming from jumps does negate the necessity of an additional threshold for around 20%–30% of the randomly selected larger portfolios.

Further elaborating on this, Table 6 shows the corresponding forecasting results, analogous to Table 4 where the jumps were not removed in the calculation of the different measures.

Taken together, the results show that it is the *form* of the selected model that changes when jumps are excluded, not the predictive fit. In particular, when estimating the partial covariances with high-frequency returns void of jumps, the simpler semicovariance-based models are preferable, while when jumps are not removed, a partial covariance-based model with a threshold that effectively separates the negative jumps improves forecast performance.

5. Conclusion: Is zero optimal?

Motivated by a question posed by Frank Diebold in a 2018 blog post, this paper proposes a generalization of the class of realized semi (co)variance measures introduced by Barndorff-Nielsen et al. (2010) and Bollerslev et al. (2020a) to allow

[15] Recall that the results for the conventional measures in Table 3 showed that the combination of the zero threshold with one additional (typically negative) threshold was the preferred choice in the majority of cases, especially for larger dimensions.

for a finer decomposition of realized (co)variances. The new realized partial (co)variances allow for multiple thresholds, rather than the single fixed threshold of zero used in semi (co)variances. The number and values of the thresholds used to construct the partial (co)variances can be freely chosen by the researcher. In this paper, we adopt methods from machine learning to optimally choose the threshold(s) so as to maximize the out-of-sample forecast performance of simple time series forecasting models based on the realized partial (co)variances.

Overall we find that it is hard, but not impossible, to improve upon the simple fixed threshold of zero embedded in realized semi (co)variances. In particular, when restricting attention to partial (co)variances with a *single* threshold, the zero threshold cannot be consistently beaten. However, when multiple thresholds are allowed for, a univariate model based on partial variances with two optimally-chosen thresholds significantly improve upon existing semivariance-based models. In our multivariate analyses, an "enhanced" semicovariance-based model, with one threshold at zero and another flexibly-chosen threshold, typically in the left tail, emerges as the winner. Meanwhile, applying the same partial covariance models to variation measures constructed from high-frequency returns void of jumps, a semicovariance-based model re-emerges as the preferred specification.

Acknowledgments

We would like to thank Serena Ng (the Editor), the associate editor, and two anonymous referees for their thoughtful comments and suggestions, as well as Frank Diebold for his inspiring challenge. Patton and Quaedvlieg gratefully acknowledge support from, respectively, Australian Research Council Discovery Project 180104120 and Netherlands Organisation for Scientific Research Grant 451-17-009.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2021.04.013.

References

Andersen, T.G., Bollerlsev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39 (4), 885–905.
Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Volatility and correlation forecasting. In: Elliott, G., Granger, C.W., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier, pp. 777–878, Ch. 15.
Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Rev. Econ. Stat. 89 (4), 701–720.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Amer. Statist. Assoc. 96, 42–55.
Ang, A., Chen, J., Xing, Y., 2006. Downside risk. Rev. Financ. Stud. 19 (4), 1191–1239.
Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statist. Surv. 4, 40–79.
Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econometrics 162 (2), 149–169.
Barndorff-Nielsen, O.E., Kinnebrock, S., Shephard, N., 2010. Measuring downside risk: Realised semivariance. In: Bollerslev, T., Russell, J.R., Watson, M.W. (Eds.), Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle. Oxford University Press, Oxford, pp. 117–136.
Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Series B, 64, 253–280.
Bergmeir, C., Hyndman, R.J., Koo, B., 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Comput. Statist. Data Anal. 120, 70–83.
Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: Modeling and managing volatility. Rev. Financ. Stud. 31 (7), 2729–2773.
Bollerslev, T., Li, J., Patton, A.J., Quaedvlieg, R., 2020a. Realized semicovariances. Econometrica 88 (4), 1515–1551.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. J. Econometrics 192 (1), 1–18.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2020b. Multivariate leverage effects and realized semicovariance garch models. J. Econometrics 217 (2), 411–430.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2020c. Realized semibetas: Signs of things to come. J. Financial Econom. forthcoming.
Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. J. Finance 66 (6), 2165–2211.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. CRC Press.
Cappiello, L., Engle, R., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. J. Financ. Econom. 4 (4), 537–572.
Cerqueira, V., Torgo, L., Mozetič, I., 2020. A survey of cross-validation procedures for model selection. Mach. Learn. 109, 1997–2028.
Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. J. Econometrics 159 (1), 116–133.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econom. 7 (2), 174–196.
Diebold, F.X., 1986. Modeling the persistence of conditional variances: A comment. Econometric Rev. 5 (1), 51–56.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econom. Statist. 13 (3), 253–263.
Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. J. Appl. Econometrics 4 (1), 1–21.
Diebold, F.X., Strasser, G.H., 2013. On the correlation structure of microstructure noise: A financial economic approach. Rev. Econom. Stud. 80 (4), 1304–1337.
Fishburn, P.C., 1977. Mean-risk analysis with risk associated with below-target returns. Amer. Econ. Rev. 67 (2), 116–126.
Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Hogan, W.W., Warren, J.M., 1972. Computation of the efficient boundary in the E-S portfolio selection model. J. Financ. Quant. Anal. 7 (4), 1881–1896.

Hogan, W.W., Warren, J.M., 1974. Toward the development of an equilibrium capital-market model based on semivariance. J. Financ. Quant. Anal. 9 (1), 1–11.
Kahneman, D., Tversky, A., 1979. Prospect theory. Econometrica 47 (2), 263–291.
Kroner, K.F., Ng, V.K., 1998. Modeling asymmetric comovements of asset returns. Rev. Financ. Stud. 11 (4), 817–844.
Li, J., Liao, Z., Quaedvlieg, R., 2021. Conditional superior predictive ability. Rev. Econom. Stud. forthcoming.
Liu, L.Y., Patton, A.J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple assets. J. Econometrics 187, 293–311.
Markowitz, H.M., 1959. Portfolio Selection. Wiley.
Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Rev. Econ. Stat. 97 (3), 683–697.
Roy, A.D., 1952. Safety first and the holding of assets. Econometrica 431–449.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394–1411.