From Zero to Hero: Realized Partial (Co)Variances

Tim Bollerslev
Department of Economics
Duke University

Marcelo C. Medeiros
Department of Economics
PUC-Rio

Andrew J. Patton
Department of Economics
Duke University

Rogier Quaedvlieg
Department of Business Economics
Erasmus School of Economics

July 2, 2020

Abstract

This paper proposes a generalization of the class of realized semivariance and semicovariance measures introduced by Barndorff-Nielsen, Kinnebrock and Shephard (2010) and Bollerslev, Li, Patton and Quaedvlieg (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.

Keywords: High-frequency data; realized variation; volatility forecasting.

JEL codes: C22, C51, C53, C58.

Corresponding author: Tim Bollerslev, Department of Economics, Durham, NC 27707, USA; boller@duke.edu.

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 procedures remained the norm for modeling and forecasting time-varying variances and covariances up until the early 2000s (see, e.g., the survey in Andersen, Bollerslev, Christoffersen and Diebold (2006)).

Meanwhile, motivated by the increased availability of reliable high-frequency data for a host of financial assets, Andersen, Bollerlsev, Diebold and Labys (2003) proposed the use of traditional ARMA type procedures for directly modeling the dynamic dependencies in daily and lower frequency so-called realized volatilities constructed from higher frequency intraday returns (see also Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (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, Bollerslev and Diebold (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, as exemplified by the up and downside betas of Ang, Chen and Xing (2006), Barndorff-Nielsen et al. (2010) 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, as well as 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 19^{th}, 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, Kelly and Xiu (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:

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

"*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 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 optimal 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 a wider set of 749 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 Sections 3 and 4, respectively. Section 5 concludes. The Appendix contains additional details and 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., Markowitz (1959), Mao (1970), Hogan and Warren (1972, 1974) and Fishburn (1977)). 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 k^{th} 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 $\mathbf{r}_{t,k} \equiv [r_{t,k,1}, ..., r_{t,k,N}]'$. 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 $\boldsymbol{r}_{t,k} = n(\boldsymbol{r}_{t,k}) + p(\boldsymbol{r}_{t,k})$. 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 semivariation measures may seem somewhat arbitrary from a statistical perspective. However, from an economic perspective the choice of zero is rooted in loss aversion and prospect theory (Kahneman and Tversky, 1980), which have been extensively studied in the empirical finance literature, see, e.g., Ang et al. (2006) and Bollerslev, Patton and Quaedvlieg (2020c) and the many references therein. Volatility forecasts based on zero-threshold semi(co)variation type measures have also previously 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, Engle and Sheppard (2006), Patton and Sheppard

[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.

(2015) and Bollerslev, Patton and Quaedvlieg (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

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

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

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

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 equation (1), realized partial covariances are defined by:

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

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

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

When the ordering of the assets is arbitrary, as in our empirical analysis below, it may 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 is analogous to the definition of $\boldsymbol{M}$ in equation (1).

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

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

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 automatically reduce to the familiar 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.

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 vast-dimensional partial covariance matrices.

All 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 4,202 observations. We rely on a 5-minute sampling frequency in the construction of the different daily realized variation measures.

The raw data was obtained from the TAQ database, and is the same as that analyzed in Bollerslev et al. (2020a); see that article for further details on the cleaning of the data and a thorough justification for our choice of 5-minute sampling.

When implementing the realized partial variances, in recognition of the large movements in return volatility over time, we use time-varying, asset-specific quantiles of the volatility-standardized return distributions as thresholds. Specifically:

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

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 course, result in distinctly different partial covariances 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, Hood, Huss and Pedersen (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 $RV_{t-1|t-k} = k^{-1} \sum_{i=1}^{k} RV_{t-i}$. An important extension of the RV-HAR model, due to Patton and Sheppard (2015), decomposes the lagged daily realized variance into the two semi-

variances.[3] 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:

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

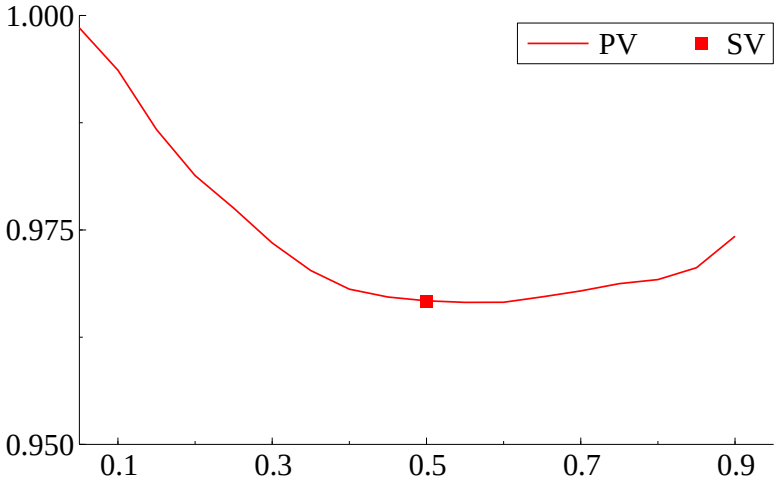
We decompose only the daily lag to most closely resemble the SV-HAR model, and to allow for more clear-cut comparisons concerning the choice of threshold(s).

We start 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 reduces to the SV-HAR when the threshold selected is zero, corresponding approximately to choosing the median as the partial variance threshold. Figure 1 plots the out-of-sample forecast MSEs for the S&P 500, based on rolling sample estimation windows of 1,000 observation, as a function of the quantile used in the construction of the partial variances. As the figure shows, the lowest MSE is obtained using the median as the threshold, corresponding to using the SV-HAR model. Hence for for a PV(G)-HAR type of model to outperform the SV-HAR model, it must involve more than one threshold ($G > 2$), or a dynamically varying choice of threshold(s).

We next extend the simple estimation underlying Figure 1 to accommodate time-varying threshold(s). We determine the thresholds using an exhaustive search combined with cross-

[3]The weekly and monthly realized variances can of course also be decomponsed into their positive and negative semivariance components, however Patton and Sheppard (2015) report that the largest gains came from decomposing only the daily lag.

Figure 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 relative to the MSE for the RV-HAR model as a function of the quantile threshold used in the definition of the partial variances.

validation (CV). The grid of quantiles ranges from 0.05 to 0.95 in increments of 5%. We select the threshold, or combination of thresholds, as the quantiles that minimize the average of the MSE over the CV folds. We use rolling windows of 1,000 observations and a 5-fold CV.[4] 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 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. In parallel to the estimation above, we rely on rolling windows of 1,000 observations for estimation, with the optimal quantile(s) selected by 5-fold CV.

Table 1 summarizes the results for the S&P 500 and across all of the individual stocks. In addition to the PV(G)-HAR models, we also report the results for the benchmark RV-HAR

[4]For a each model, we estimate the model based on four of the five sub-samples and compute the MSE for the fifth subsample. We choose the quantile that minimizes the average MSE over the five different sub-samples.

and SV-HAR models. First consider the model performance in terms of MSE and QLIKE.[5] For both the S&P500 and across the individual stocks, the PV(G)-HAR with 2, 3 or a variable number of regions (G^*) outperforms the basic HAR. 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, suggesting overfitting despite our use of cross-validation.

To formally compare the performance, 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 stocks, we present the number of rejections at the 5% level. While the SV-HAR clearly improves over RV-HAR, 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. The PV(3) and PV(G^*) models significantly outperform the RV-HAR in almost all cases. Moreover, they tend to significantly improve over the SV-HAR as well. 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, Lunde and Nason (2011). For the S&P500, the PV(3) and PV(G^*) are most convincingly parts of the model confidence set, as evidenced by their high p-values. Across the individual stocks, we also observe that these models are almost always included in the MCS, while the RV- and SV-HAR are included for less than half of the firms.[6]

The forecasting performance of the PV-HAR models demonstrated two major points. First, improving over SV-HAR is difficult. When restricted to a single threshold, consistently beating the zero threshold appears challenging. Rather, to get close to the SV-HAR, we need two thresholds, with further improvements available by dynamically selecting them using cross val-

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

[6]As a robustness check, we repeat the entire analysis minimizing QLIKE in the cross-validation step, resulting in overall very similar findings and forecasting performance. The results are reported in Appendix A.

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. 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.

idation. To help understand this result, 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 column shows that for the vast majority of time and stocks, the model is simply the PV(3) model, while, for a few stocks, the number of thresholds is reduced to one for some periods of time. The second column provides summary statistics on the

Table 2: Univariate Models: Quantile Selection

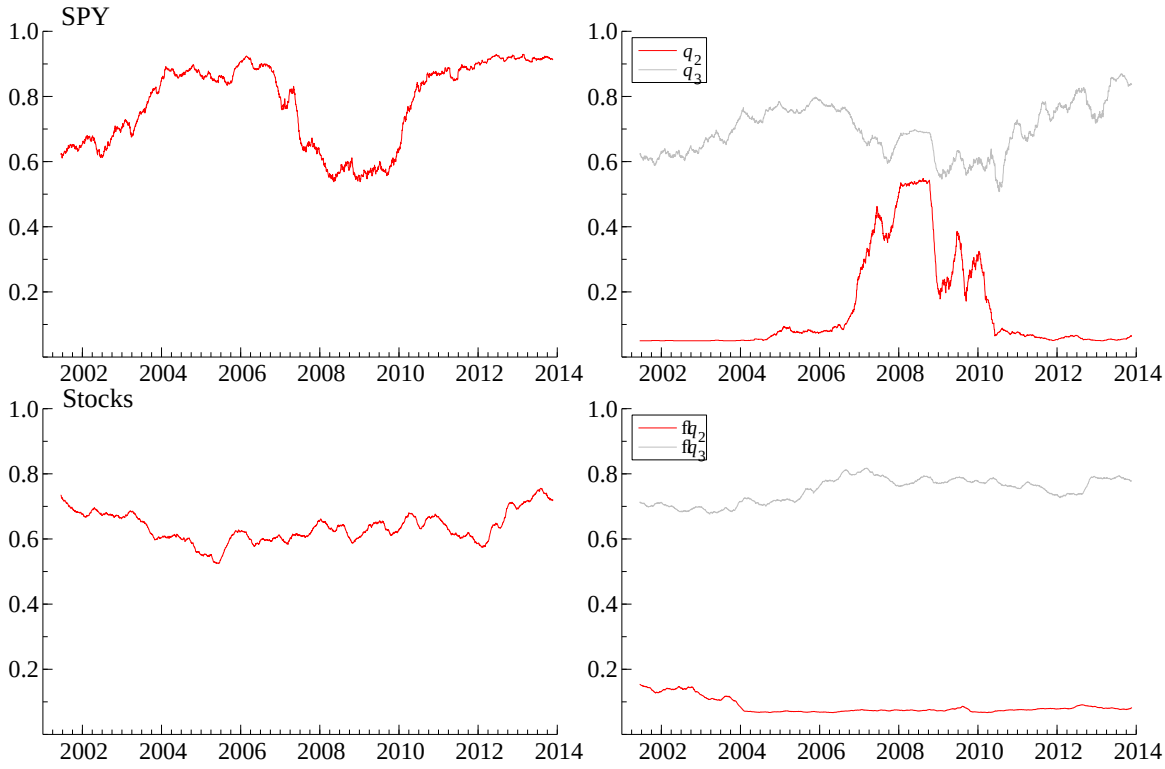
	PV(G*) Model Selection				Unc. Threshold Selection			Cond. Threshold Selection		
	SV	PV(2)	PV(3)	PV(4)	PV(2)	PV(3)		PV(2)	PV(3)	
					$\bar{q}_2$	$\bar{q}_2$	$\bar{q}_3$	$\bar{q}_2$	$\bar{q}_2$	$\bar{q}_3$
SPY										
	0.000	0.001	0.999	0.000	0.78	0.14	0.70	0.67	0.14	0.70
Stocks										
5% Quantile	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 column shows how often each number of thresholds is selected in the PV(G*) model. The second column shows descriptives of the thresholds for the PV(2) and PV(3) model. The final column shows these same descriptives, conditional on the being selected in the PV(G*) model.

time-series averages of the selected threshold, as well as its distribution 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, these 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 column shows the average quantile, conditional on PV(G*) selecting the single or double threshold model. The main difference is that the single threshold model appears to only be chosen in times where the threshold is very high for the individual stocks.

The time-series average of the selected quantiles only tells part of the story. Figure 2 plots the cross-sectional average and distribution over time. The top row presents the results for the SPY index, while the bottom row presents those for the individual stocks. In the single threshold case, the quantiles selected for the SPY are typically high, but are closer to the median around times of financial difficulties, starting with the recession in the early 2000s and subsequently during the financial crisis of 2008. Interestingly, the cross-validated threshold

Figure 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 SPY 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.

dropped in value already in early 2007, preceding the subsequent rise in volatility. The PV(3) model (with two thresholds) in the top-right panel shows similar dynamics where both the upper and lower threshold converge to the median during the financial crisis. When averaged across stocks, the quantiles appear are more stable: In the single threshold setting, the average quantile is around 0.65, but this number is the result of a clear bimodal distribution of very low, and very high quantiles, similar to what we pick up in the two threshold case.

To get a better understanding of the PV-HAR models' workings, we present in-sample estimation results for the various models based on the SPY below in Table 3. Since the average

Table 3: Univariate Models: Parameter Estimates

	RV	SV	PV(2)	PV(3)	PV(4)
Selected Quantiles			0.95	0.05	0.15
				0.75	0.50
					0.85
RV_{t-1}	0.444				
	(0.021)				
SV^{-}_{t-1}		0.384			
		(0.017)			
SV^{+}_{t-1}		0.040			
		(0.020)			
$PV^{(1)}_{t-1}$			0.545	0.045	0.159
			(0.022)	(0.021)	(0.024)
$PV^{(2)}_{t-1}$			-0.066	0.417	-0.026
			(0.018)	(0.026)	(0.027)
$PV^{(3)}_{t-1}$				-0.001	0.325
				(0.019)	(0.029)
$PV^{(4)}_{t-1}$					-0.008
					(0.023)
$RV_{t-1\|t-5}$	0.326	0.365	0.333	0.332	0.337
	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)
$RV_{t-1\|t-22}$	0.168	0.162	0.156	0.160	0.164
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)

Note: The table reports the full-sample parameter estimates for the S&P 500, with all of the right-hand-side explanatory variables standardized to have mean zero and unit variance. An intercept is included in all models but is not reported here in the interests of space. Heteroskedasticty robust standard errors are reported in parentheses.

values of the partial variances may differ greatly, we standardize all right-hand side variables to be mean zero, variance one, such that we can directly compare all the coefficients. We focus our discussion on the parameters for the daily lag, which is 0.44 for the base HAR model. For the SV-HAR we find the common result that the negative semivariance carries most of the information, with the positive semivariance's impact being roughly one-tenth of the positive semivariance. The three remaining columns provide the PV(G)-HAR estimates. We focus on

the SPY, but the selected quantiles are representative for those most commonly selected across the individual stocks as well. For the single-threshold partial-variance model, full-sample cross-validation finds a threshold of 0.9, selecting a term related to large positive returns, which carries an insignificant negative coefficient. The remaining variation obtains the largest positive coefficient across all models. The preferred PV(3)-HAR model, selects one low and one quantile. The variation in the middle of the distribution (captured by $PV_{t-1}^{(2)}$) appears to be the most important, with the variation stemming from large negative returns obtaining a secondary role, while the large positive returns are again insignificant. The three threshold model (PV(4)) selects the median and two tail quantiles. Unsurprisingly, the component associated with large negative returns has a high loading. More surprising is the fact that the above median partial variance is the other important component, while the below median partial variance is insignificant. In general, the model is highly impacted by small changes in the quantiles, in line with the poor out-of-sample performance previously documented in Table 1.

The selection of two thresholds, one high, and one low, is reminiscent of the HAR-J models of Andersen et al. (2007), who decompose realized variance into a continuous- and jump-variation component. They find that the continuous part is the predictable component, while the jump component, which is due to large and infrequent price jumps, leads to spuriously high RV forecasts when the right-hand side is not decomposed into the two components. The above results suggest that of primary importance is filtering out the large positive returns, rather than all of them.

In order to gain a deeper understanding of the performance of the PV(G)-HAR model and how it interplays with jumps, we apply the Conditional Superior Predictive Ability (CSPA) test of Li, Liao and Quaedvlieg (2020). The CSPA null hypothesis posits that

$$H_0 : \mathbb{E}[L_{j,t} - L_{0,t} | X_t = x] \geq 0 \; \forall \; x \in \mathcal{X}, 1 \leq j \leq J,$$

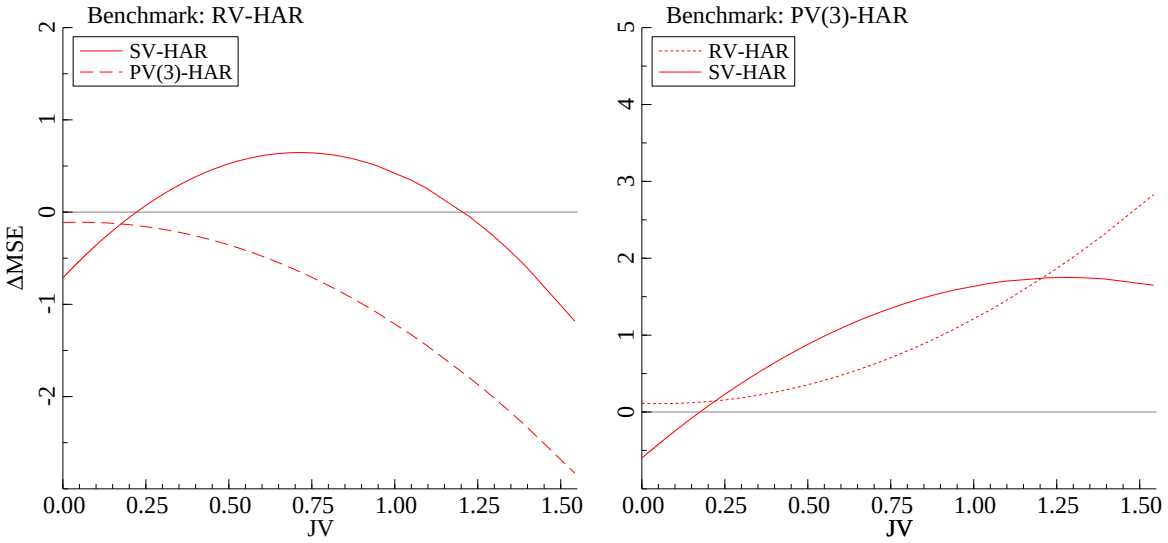
where $L_{0,t}$ is the time-t loss of the benchmark model, $L_{j,t}$ the loss of a set of alternative models, and X_t the conditioning variable. This test compliments the unconditional predictive performance comparisons presented in Table 1. A failure to reject the null indicates that the benchmark model is not dominated in *any* of the states dictated by the conditioning variable. This testing procedure also provides an estimate of the above conditional expectation function, which allows us to easily visualize the models' performance as function of the conditioning variable. We focus on lagged realized jump variation as a conditioning variable.[7]

Figure 3 provides the output of the CSPA test for the MSE loss of SPY forecasts. For simplicity we only consider the RV-, SV- and PV(3)-HAR models. The left panel uses HAR as the benchmark, and from the curves associated with the two competing models we see that the PV(3)-HAR model outperforms the HAR model for all values of jump variation, with greater outperformance when jump variation is higher. The SV-HAR model outperforms HAR only when jump variation is high or low; for intermediate values it underperforms HAR. A formal test of whether the minimum of the two curves is always positive fails to reject the null, indicating that the HAR model is not uniformly significantly outperformed by the two competing models. In the right panel we set the PV(3)-HAR model as the benchmark, and we observe that it underperforms SV-HAR for very low values of jump variation, but outperforms for all other values. Again, a formal test of whether the minimum of the two curves is always positive fails to reject the null, indicating that the PV(3)-HAR model is not uniformly significantly outperformed by the two competing models.

We apply the CSPA tests to all 28 individual stocks, and report the number of times the null is rejected in Table 4. We find that based using MSE loss, both RV and SV-HAR models are beaten by either the PV(3) or PV(G)-HAR models for about half the stocks, while these two PV models are only rejected with respect to the SV-HAR in one or two cases. The joint test against

[7]To identify jump returns, we employ the dynamic threshold of Mancini (2009) and Bollerslev and Todorov (2011a,b) based on three times the trailing scaled bipower variation, adjusted for intraday periodicity in the volatility. See Bollerslev and Todorov (2011b) for additional details.

Figure 3: Univariate Models: Conditional Superior Predictive Ability for SPY



Note: The figure shows the conditional expected loss differential functions, based on the CSPA test of Li et al. (2020), using jump variation as the conditioning variables. The left and right panels show results using RV-HAR and PV(3)-HAR as the benchmark model respectively.

Table 4: Univariate Models: Conditional Superior Predictive Ability

	MSE				QLIKE			
	RV	SV	PV(3)	PV(G*)	RV	SV	PV(3)	PV(G*)
Panel A: One-versus-one CSPA tests against competing models								
RV		3	0	0		6	3	2
SV	5		2	1	19		5	5
PV(2)	11	3	0	0	17	9	6	3
PV(3)	13	14		0	27	25		0
PV(4)	0	0	0	0	3	2	0	0
PV(G)	12	13	0		27	25	0	
Panel B: One-versus-all CSPA tests against competing models								
	16	14	3	2	28	28	6	6

Note: This table reports the rejection frequencies of the CSPA test using lagged jump-variation as the conditioning variable. Each entry represents the number of stocks for which the column-model fails the CSPA null against the row model.

all competing models is only rejected for two or three stocks. Rejections based on QLIKE loss are more frequent, with both RV and SV-HAR CSPA hypothesis being rejected for all stocks, and the PV(G)-HAR models surviving the test for all but six of the series.

4 Vast-dimensional partial covariance models

We next extend the above univariate analyses to a multivariate context. Mirroring the univariate forecast comparisons, we take the basic HAR model applied to the *vech* of the realized covariance matrices as our benchmark model:

$$\begin{aligned} \text{vech}(\boldsymbol{RCOV}_t) = \boldsymbol{\phi}_0 &+ \phi_1 \text{vech}(\boldsymbol{RCOV}_{t-1}) \\ &+ \phi_2 \text{vech}(\boldsymbol{RCOV}_{t-1|t-5}) + \phi_3 \text{vech}(\boldsymbol{RCOV}_{t-1|t-22}) + \epsilon_t \end{aligned} \tag{12}$$

where $\boldsymbol{\phi}_0$ is an $N(N+1)/2-$vector, and the remaining coefficients are 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:[8]

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

Finally, we consider the PCOV(G)-HAR partial covariance class of models. Motivated by the results in the previous section, 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

[8]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.

of six partial covariances:

$$\begin{aligned}\text{vech}(\boldsymbol{RCOV}_t) = \boldsymbol{\phi_0} \;&+\; \sum_{g=1}^{G}\sum_{g'=g}^{G}\phi_1^{(g,g')}\text{vech}(\overline{\boldsymbol{PCOV}}_t^{(g,g')}) \\ &+\; \phi_2\text{vech}(\boldsymbol{RCOV}_{t-1|t-5}) + \phi_3\text{vech}(\boldsymbol{RCOV}_{t-1|t-22}) + \epsilon_t\end{aligned} \tag{14}$$

where

$$\overline{\boldsymbol{PCOV}}_t^{(g,g')} \equiv \begin{cases} \boldsymbol{PCOV}_t^{(g,g)} & \text{for } g = g' \\ \boldsymbol{PCOV}_t^{(g,g')} + \boldsymbol{PCOV}_t^{(g',g)}, & \text{for } g \neq g' \end{cases} \tag{15}$$

We consider portfolios ranging in dimension from 2 to 50 assets, with all of the results reported below based on 100 randomly selected portfolios of a given dimension. To facilitate the calculation, we rely on a wider dataset which expands 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 stocks. As most of these stocks are less frequently traded than the DJIA stocks, we resort to a coarser 15-minutes sampling frequency in the construction of the realized measures.[9] In order to avoid too many zero-valued partial covariances, we also further reduce our grid search to thresholds ranging between the $\{0.1,0.2,...,0.9\}$ quantiles.

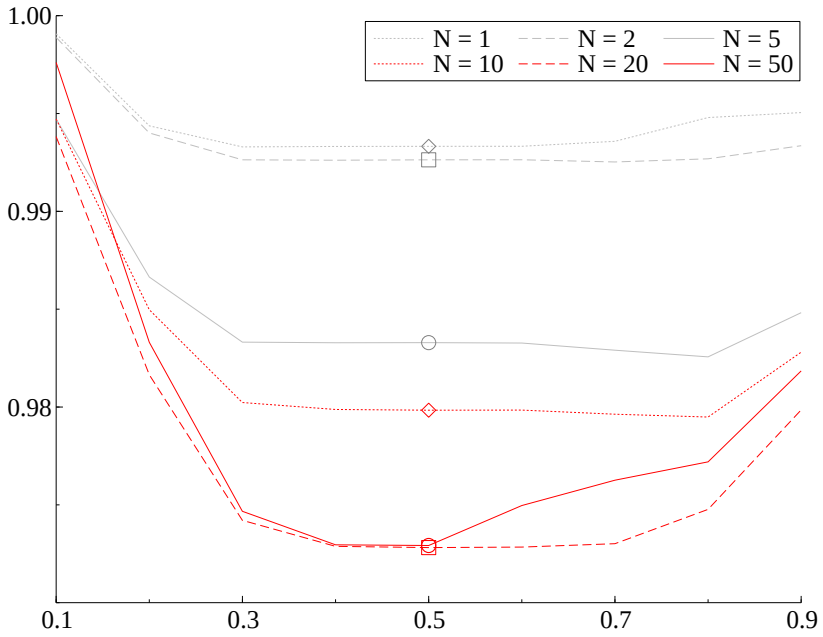
In parallel to the univariate analysis in the previous section, we begin by asking whether any single threshold is able to beat the zero threshold. We focus on the multivariate MSE, or Frobenius norm, defined as:

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

where $\boldsymbol{H}_t$ denotes the forecast. Figure 4 shows the resulting MSEs averaged over time and across the 100 random portfolios for different single-threshold PCOV models and a variety

[9]This same dataset also underlies the multivariate analysis in Bollerslev et al. (2020a), and we refer to that study for a more detailed discussion of the data.

Figure 4: Multivariate Models: PCOV(2)-HAR Forecasting Performance



Note: The figure shows the out-of-sample MSE of the PCOV(2)-HAR model, averaged over 100 random portfolios of size N, relative to the MSE for the RCOV-HAR model, as a function of the quantile threshold used in the definition of the partial covariances. The symbols are located at the corresponding loss of the SCOV-HAR model.

of cross-sectional dimensions. To ease interpretation we again report the ratios of the MSEs relative to the MSE for the benchmark RCOV model. One conclusion from Figure 4 echoes that from Figure 1 for the S&P 500, and the results for each of the individual stocks reported in Table 1, namely that the fixed zero threshold is optimal or close to optimal unconditionally. Consistent with the findings pertaining to semi covariances and the zero threshold in Bollerslev et al. (2020a), Figure 4 also reveals that the gains from using partial covariances are increasing in the dimension of the covariance matrices to be forecast more generally, as the curves shift nearly monotonically downwards as N increases.

To allow for multiple thresholds and the choice of thresholds to vary over time, we rely on the same 5-fold CV procedure to optimally select the thresholds. Table 5 provides a summary of

Table 5: Multivariate Models: PCOV(G*)-HAR Model Selection

N	1	2	5	10	20	50
RCOV	0	0	0	0	0	0
SCOV	0	2	3	3	0	0
PCOV(2)	8	9	15	5	0	0
SCOV$^+$	49	20	48	84	100	100
PCOV(3)	43	69	40	13	0	0

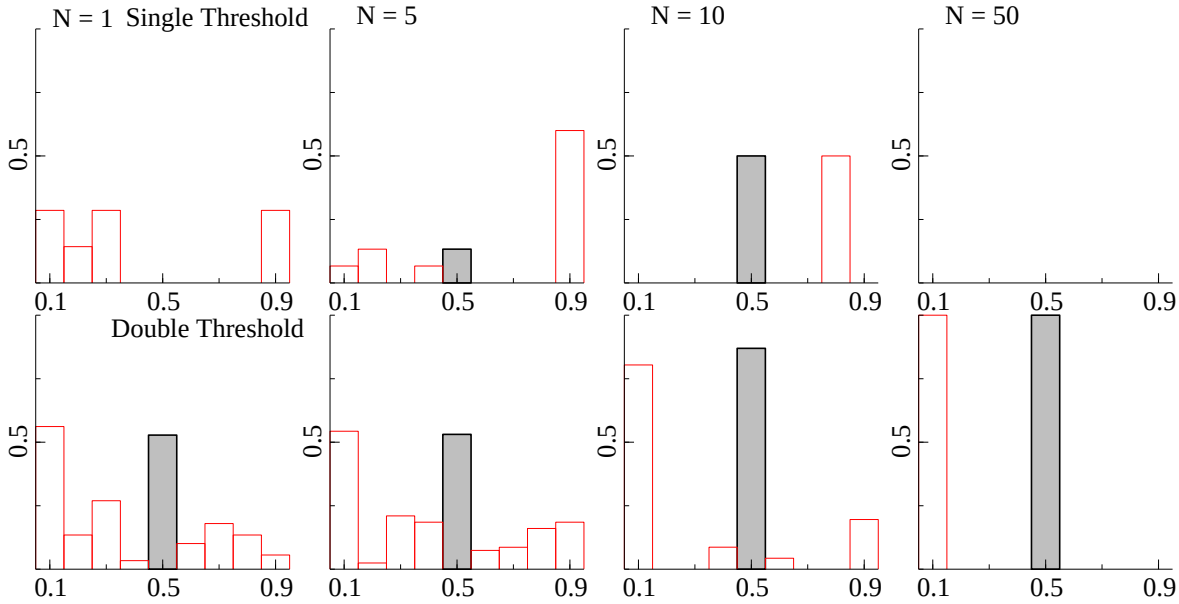
Note: The table reports the frequency, out of 100 random portfolios, with which various threshold combinations are selected, as a function of the portfolio dimension N. RCOV implies no threshold, SCOV is the single zero threshold, PCOV(2) indicates a single non-zero threshold, SCOV$^+$ refers to two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

the different models that were selected over the full sample 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 (12) or (14), "SCOV$^+$" refers to a model that adds one additional threshold to the SCOV model (which already contains a threshold at zero), while "PCOV(2)" and "PCOV(3)" denote one- or two-threshold models where all of the thresholds can differ from zero. As in the univariate setting, the preferred number of thresholds is generally two: one of PCOV(3) or SCOV$^+$ is selected for between 88% and 100% of the portfolios across all values of N. The basic RCOV model is never selected and the original SCOV model for no more than 3% of portfolios. It is noteworthy that the SCOV$^+$ model, which involves zero plus one additional flexibly-chosen threshold, emerges as the clear winner, especially for larger dimensions ($N \geq 20$) where it is always the preferred model.[10]

The emergence of the SCOV$^+$ model as the preferred model is further underscored by Figure 5, which reports the distributions of the selected quantiles for the PCOV models across the 100 random portfolios for different portfolio sizes N. As in the univariate results, when a single threshold is selected, the location of the threshold is typically in the tails. Meanwhile, when

[10]Looking at the parameter estimates for the SCOV$^+$ models, the joint most-negative partial covariance component typically stands out as the most important. Additional details are reported in Appendix B.

Figure 5: Multivariate Models: Quantile Selection



Note: The figures shows the frequency with which the different quantiles are selected as the optimal threshold(s), conditional on selecting a model with one threshold (top row), or two thresholds (bottom row). Each column corresponds to a different cross-sectional dimension N. The shaded bar at 0.5 denotes the semicovariance threshold.

two thresholds are selected, they are typically zero and a low threshold. In fact, as Table 5 shows, for $N = 50$ that is the *only* model and set of thresholds selected. This finding explains the empty top-right panel in the figure; the single threshold model is never chosen for $N = 50$.

4.1 Forecasting RCOV

The results in the previous section pertained to in-sample model selection and threshold estimation. We turn next to out-of-sample multivariate forecasting. As in our univariate analysis, we use rolling windows of length 1,000 to construct the forecasts. Due to 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 2,000 observations.

In addition to the multivariate MSE, or Frobenius norm, defined above, we also consider

a multivariate QLIKE loss function. Since the dimension of some of the portfolios exceed the number of intraday observations ($m = 26$), the ex-post realized covariance matrix are not necessarily positive definite. Hence, we adapt the conventional multivariate QLIKE to accommodate rank deficiency (see also Bollerslev et al. (2020a)):

$$QLIKE_t \equiv \text{Tr}(\boldsymbol{H}_t^{-1}\boldsymbol{RCOV}_t) - \log|\boldsymbol{H}_t^{-1}\widetilde{\boldsymbol{RCOV}}_t| - N,$$

where $\boldsymbol{H}_t$ refers to the forecast, and $\widetilde{\boldsymbol{RCOV}}_t$ denotes a shrunk version of $\boldsymbol{RCOV}_t$ that ensures positive definiteness.[11] This modification only affects the normalization of the QLIKE loss function, not the ranking of the models.

Table 6 reports the average MSE and QLIKE loss for the different models, together with their prevalence in the 80% MCS. Consistent with the earlier findings, for both of the loss functions, the loss reductions obtained by the threshold-based models, as well as their significance, is increasing in the cross-sectional dimension N. In particular, while the losses for the SCOV and PCOV models are generally lower than the losses for the benchmark RCOV model, the RCOV model is still included in the MCS for many of the portfolios of size $N \leq 10$. Meanwhile, the QLIKE loss always rejects the RCOV model for the $N \geq 20$ dimensional forecasts. $N = 20$ also coincides with the dimension for which the PCOV(G*)-HAR consistently selects the SCOV$^+$ with a low second threshold. Intuitively, for small dimensions, a variety of different thresholds are typically selected. However, these optimal thresholds are unlikely to coincide across the different sets of stocks and portfolios. Instead, the choice of zero, possibly augmented with a low additional threshold, emerges as a "robust" choice in large dimensions.[12]

[11] The shrinkage is formally defined by $\widetilde{C}_t \equiv (1-\lambda_t)\, C_t^* + \lambda_t Diag\{C_t^*\}$, where $\lambda_t = \arg\min(\lambda \in \{0, 0.1, ..., 1\} : |\widetilde{C}_t| > 0)$ and $C_t^* \equiv I_d \odot \max\left\{\widehat{C}_t, 0.001 \cdot J_d\right\} + (J_d - I_d) \odot \widehat{C}_t$.

[12] To further corroborate this conjecture, Appendix C reports the results where we consider the forecasts of portfolios comprised of stocks within the same industry. Overall, the relative performance of the PCOV(G*) models is marginally improved, and the selected thresholds also differ from the SCOV$^+$ thresholds, suggesting that there is some commonality in deviations from the optimal zero threshold for related stocks.

Table 6: Multivariate Models: Forecasting Results

	MSE			QLIKE		
N	RCOV	SCOV	PCOV(G*)	RCOV	SCOV	PCOV(G*)
1	21.273	21.270	21.216	0.227	0.227	0.228
	(0.97)	(0.89)	(0.98)	(0.75)	(0.89)	(0.81)
2	14.825	14.709	14.716	0.557	0.561	0.560
	(0.86)	(0.94)	(0.96)	(0.83)	(0.34)	(0.59)
5	8.622	8.473	8.408	2.066	2.075	2.080
	(0.47)	(0.84)	(0.99)	(0.71)	(0.19)	(0.58)
10	6.378	6.241	6.180	6.289	6.281	6.300
	(0.13)	(0.82)	(0.98)	(0.42)	(0.71)	(0.61)
20	5.634	5.501	5.426	22.131	22.011	21.988
	(0.00)	(0.62)	(1.00)	(0.00)	(0.97)	(1.00)
50	5.001	4.890	4.802	171.353	170.220	170.241
	(0.13)	(0.46)	(1.00)	(0.00)	(1.00)	(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.

Table 6 reveals that the PCOV model is almost always included in the model confidence set (MCS). To put the model to a more difficult test, we again apply the conditional CSPA test of Li et al. (2020), using the lagged VIX as the conditioning variable.[13] The overall rejection frequencies reported in Table 7 are higher than for the unconditional MCS-based tests, reflective of the greater power available in this environment. The RCOV-HAR model is rejected for between one- and two-thirds of the 100 random portfolios, signifying that in certain situations, either the SCOV-HAR and/or the PCOV-HAR models provide better forecasts than the basic RCOV model. The results also show that the PCOV models are rejected less frequently overall than the SCOV models.

[13]The jump variation measures used in the univariate CSPA tests in Section 3 do not readily extend to a multivariate context. The VIX, meanwhile, provides a simple scalar time series that is obviously related to market wide jump activity.

Table 7: Multivariate Models: Conditional Superior Predictive Ability Tests

	MSE			QLIKE		
N	RCOV	SCOV	PCOV(G*)	RCOV	SCOV	PCOV(G*)
1	18	11	9	33	20	15
2	35	26	8	30	56	39
5	65	16	9	42	75	39
10	78	20	8	78	58	32
20	34	8	6	48	4	6
50	52	12	4	64	5	2

Note: The table provides the number (out of 100) of portfolios for which the joint CSPA hypothesis is rejected (at the 95% significance level) in favor of the alternative that either of the two competing models outperforms the model for some value of the lagged VIX.

4.2 Forecasting portfolio variances

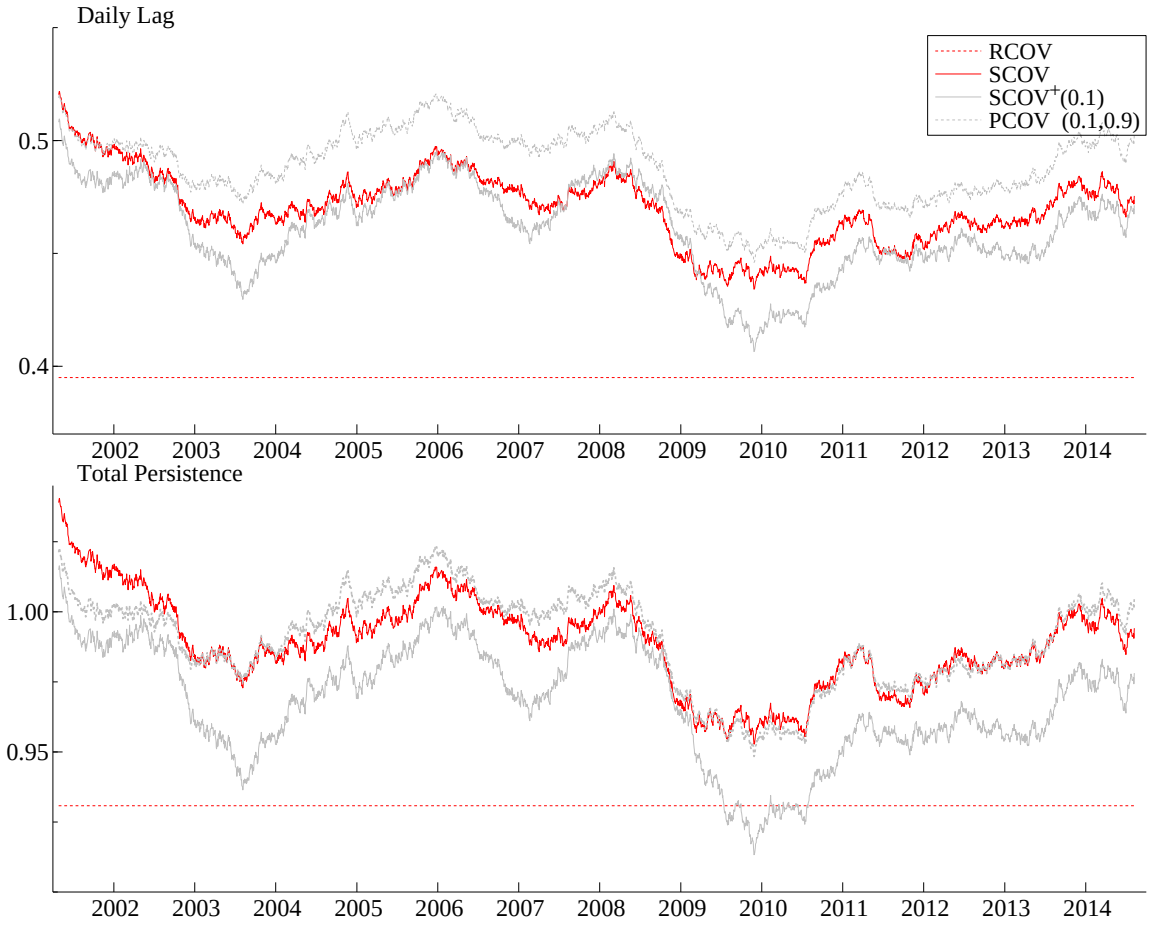
In addition to forecasting the covariance of a portfolio, the realized partial covariances may also be used in forecasting the variance of a portfolio. Let $\boldsymbol{w}$ denote the set of portfolio weights. The realized variance of the portfolio returns, $r_{t,k}^{p} = \boldsymbol{w}'\boldsymbol{r}_{t,k}$, may then be decomposed into realized partial covariances as:

$$RV_t^p \equiv \boldsymbol{w}^{'}\boldsymbol{RCOV}_t\boldsymbol{w} = \sum_{g=1}^{G}\sum_{g'=g}^{G}\boldsymbol{w}^{'}\overline{\boldsymbol{PCOV}}_t^{(g,g')}\boldsymbol{w} \equiv \sum_{g=1}^{G}\sum_{g'=g}^{G}PV_t^{p(g,g')} \tag{16}$$

where $PV_t^{p(g,g')}$ refers to the (scalar) realized partial variance of the portfolio formed using weight vector $\boldsymbol{w}$. The $G(G+1)/2$ scalar portfolio partial variances, associated with each of the partial *co*variances, are obviously distinct from any quantity that may be computed directly from the portfolio returns. As such, this representation of the portfolio variance in terms of the portfolio partial variances allows for an additional information on where the previously documented multivariate forecast improvements for the PCOV models stem from.

For simplicity of exposition, consider an AR(1) forecasting model for the portfolio variance

Figure 6: Portfolio Variance Models: Implied autoregressive parameter



Note: The figure plots the implied time autoregressive parameter for predicting the variance of random portfolios of dimension $N = 10$. The figure shows the average across 100 random portfolios, further smoothed by a $[t - 100 : t + 100]$-moving average. The top panel shows the daily lag, while the bottom panel provides the sum of the daily, weekly and monthly coefficients. The models are estimated on the full out-of-sample period.

based on all of the portfolio partial variances, as defined in equation (16) above. Since the partial variances represent a complete decomposition of the portfolio variance, this extended

AR(1) representation may be rewritten as:

$$\begin{aligned} RV_t^p &= \phi_0 + \sum_{g=1}^{G}\sum_{g'=g}^{G} \phi_1^{(g,g')} PV_t^{p(g,g')} + \epsilon_t \\ &= \phi_0 + \underbrace{\left(\sum_{g=1}^{G}\sum_{g'=g}^{G} \phi_1^{(g,g')} \frac{PV_{t-1}^{p(g,g')}}{RV_{t-1}^p} \right)}_{\phi_{1,t}} RV_{t-1}^p + \epsilon_t, \end{aligned} \tag{17}$$

thus affording a simple time-varying AR(1) parameter interpretation of the more general model.[14] Correspondingly, the implied time series estimates for $\phi_{1,t}$, augmented with a measure of the long-run persistence obtained by adding the weekly and monthly HAR coefficients, provide a simple visual for understanding the differences between the various HAR models. We consider four models: the RCOV-HAR, SCOV-HAR, and two different PCOV-HAR models motivated by the results in Figure 5. The first combines SCOV with the 0.1 quantile (denoted "SCOV$^+$") and the second employs the 0.1 and 0.9 quantiles. We estimate each of the models for 100 randomly selected equally weighted portfolios of size $N = 10$ (i.e., $\boldsymbol{w} = \boldsymbol{\iota}_N/10$).

The top panel in Figure 6 plots the average parameter estimates for each of the different models, and shows that all of the SCOV and PCOV models put higher weight on the daily lag than the RCOV model. While the RCOV model has a coefficient of 0.395, the three other models have average daily coefficients between 0.484 and 0.504, with the PCOV(0.1,0.9) model having the highest value. The bottom panel shows that the total persistence of the PCOV models also fairly closely mirrors that of the SCOV model, with the exception of the SCOV$^+$(0.1) model, which is slightly less persistent. Overall, however, the semi- and partial covariance-based models clearly allow for faster incorporation of new information and more persistent longer-run dynamic dependence than the traditional RV-HAR forecasting models.[15]

[14] This interpretation naturally extends to the HAR model, with the weekly and monthly lags added, although in all models considered in this paper those coefficients are constant.

[15] Additional results for these models are presented in Appendix D.

4.3 The role of jumps

The previous sections showed that for small dimensions, the partial covariance HAR models tend to select two thresholds, at least one of which is in the tails. For vast-dimensional models, the PCOV models always select the zero threshold and a low quantile. These results suggest that the success of the partial covariances may be related to jumps. This section further investigates that intersection. In doing so, we consider a collection of alternative models, where, in the spirit of the C-HAR model of Andersen et al. (2007) the explanatory variables are based on continuous variation only, whilst keeping the (total) realized covariance as the dependent variable. To obtain the continuous measures, we again rely on the dynamic threshold of three times the trailing scaled bipower variation, further adjusted for the strong intraday periodicity in volatility. We refer to these "continuous variation" measures and versions of the models by $\boldsymbol{PCOV}_t^{\dagger}$, and similarly for the other measures and models.

We start with a discussion of the cross-validation-based quantile selection for the models. Recall that the analogous results for the total covariation-based models presented in Table 5 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. Table 8 shows that, compared to the base scenario, SCOV† is selected more often. Meanwhile, the same convergence towards the SCOV$^{\dagger+}$ model evident in the base case remains true, although the removal of the variation stemming from jumps does negate the necessity of an additional threshold for 20% of the portfolios.

Table 9 shows the corresponding forecasting results, analogous to Table 6, where the jump and continuous variation were not separated. Overall, the average loss values in Table 9 are very close to the corresponding values in Table 6, with 80% of the results across the different models and dimensions differing by less than 1%.[16] Taken together, the results in Tables 8

[16]The models are estimated on the same (random) portfolios, so the average loss values are directly comparable.

Table 8: Multivariate Models: PCOV(G^*)†-HAR Model Selection

N	1	2	5	10	20	50
RCOV†	0	0	0	0	0	0
SCOV†	3	2	6	7	15	20
PCOV(2)†	0	0	0	0	0	0
SCOV$^{\dagger+}$	43	32	45	84	85	80
PCOV(3)†	54	66	49	9	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$^{\dagger+}$ refers to two thresholds, one of which is zero, while PCOV(3)† denotes two thresholds, none of which equal zero.

and 9, show that it is the *form* of the selected model that changes when jumps are excluded, not the predictive fit. Specifically, when applying partial covariances to non-jump returns, the simpler semicovariance-based models are preferable, while when jumps are retained, a partial covariance-based model with a threshold that effectively separates the 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 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. The number and location of the thresholds used to construct partial (co)variances can be freely chosen by the researcher. In this paper, we adopt methods from machine learning to choose the threshold(s) so as to maximize the out-of-sample forecast performance of time series models based on the

Table 9: Multivariate Models: Forecasting Results Continuous (Partial) Covariances

	MSE			QLIKE		
N	RCOV	SCOV	PCOV(G*)	RCOV	SCOV	PCOV(G*)
1	21.324	21.488	21.341	0.227	0.228	0.227
	(0.98)	(0.83)	(0.91)	(0.88)	(0.77)	(0.84)
2	14.843	14.758	14.740	0.558	0.564	0.561
	(0.73)	(0.83)	(1.00)	(0.93)	(0.48)	(0.38)
5	8.613	8.513	8.474	2.070	2.185	2.075
	(0.54)	(0.82)	(0.99)	(0.71)	(0.55)	(0.49)
10	6.369	6.172	6.235	6.311	6.290	6.279
	(0.15)	(1.00)	(0.75)	(0.13)	(0.59)	(0.88)
20	5.624	5.398	5.492	22.241	21.902	22.011
	(0.05)	(1.00)	(0.62)	(0.00)	(1.00)	(0.32)
50	4.993	4.760	4.874	172.410	169.993	170.192
	(0.10)	(1.00)	(0.30)	(0.00)	(1.00)	(0.46)

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.

resulting 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, this is not the case when multiple thresholds are allowed for. In our univariate analyses, a model based on partial variances with two optimally-chosen thresholds significantly improve upon semivariance-based models. In our multivariate analyses, an "enhanced" semicovariance 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.

References

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

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 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. chapter 15, pp. 777–878.

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

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

Ang, A., Chen, J., Xing, Y., 2006. Downside risk. Review of Financial Studies 19, 1191–1239.

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. Journal of the Royal Statistical Society Series B, 64, 253–280.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H., 2018. Risk everywhere: Modeling and managing volatility. Review of Financial Studies 31, 2729–2773.

Bollerslev, T., Li, J., Patton, A.J., Quaedvlieg, R., 2020a. Realized semicovariances. Econometrica .

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2020b. Multivariate leverage effects and realized semicovariance garch models. Journal of Econometrics 217, 411–430.

Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2020c. Realized semibetas: Signs of things to come. Available at SSRN 3528276 .

Bollerslev, T., Todorov, V., 2011a. Estimation of jump tails. Econometrica 79, 1727–1783.

Bollerslev, T., Todorov, V., 2011b. Tails, fears, and risk premia. Journal of Finance 66, 2165–2211.

Cappiello, L., Engle, R., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics 4, 537–572.

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

Diebold, F.X., 1986. Modeling the persistence of conditional variances: A comment. Econometric Reviews 5, 51–56.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. Journal of Applied Econometrics 4, 1–21.

Fishburn, P.C., 1977. Mean-risk analysis with risk associated with below-target returns. American Economic Review 67, 116–126.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hogan, W.W., Warren, J.M., 1972. Computation of the efficient boundary in the E-S portfolio selection model. Journal of Financial and Quantitative Analysis 7, 1881–1896.

Hogan, W.W., Warren, J.M., 1974. Toward the development of an equilibrium capital-market model based on semivariance. Journal of Financial and Quantitative Analysis 9, 1–11.

Kahneman, D., Tversky, A., 1980. Prospect theory. Econometrica 12.

Kroner, K.F., Ng, V.K., 1998. Modeling asymmetric comovements of asset returns. Review of Financial Studies 11, 817–844.

Li, J., Liao, Z., Quaedvlieg, R., 2020. Conditional superior predictive ability. Working Paper .

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

Mao, J.C.T., 1970. Survey of capital budgeting: Theory and practice. Journal of Finance 25, 349–360.

Markowitz, H.M., 1959. Portfolio Selection. Wiley.

Patton, A.J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Appendix A QLIKE based cross-validation

In this appendix we report the forecasting results for univariate models when we select the number and location of thresholds by minimizing cross-validated QLIKE loss, rather than MSE loss. Table A.1 reports the forecasting performance of the different models, directly mirroring the MSE-based results in Table 1. The relative performance of the models does not substantially change, but the PV(G)-HAR forecasts tend to be slightly worse, even when considering QLIKE loss.

Table A.1: Univariate Models: Unconditional Forecasting performance

	RV	SV	PV(2)	PV(3)	PV(G*)
Panel A: S&P500					
MSE	2.5186	2.4345	2.4359	2.4059	2.3927
p-val. dm_{RV}		0.164	0.166	0.109	0.101
p-val. dm_{SV}			0.669	0.013	0.025
p-val. MCS	0.298	0.108	0.108	0.298	1.000
QLIKE	0.1387	0.1361	0.1362	0.1346	0.1349
p-val. dm_{RV}		0.005	0.007	0.000	0.001
p-val. dm_{SV}			0.594	0.040	0.080
p-val. MCS	0.008	0.178	0.178	1.000	0.229
Panel B: Individual Stocks					
$\overline{MSE}$	14.982	14.886	14.888	14.416	14.420
#sig. dm_{RV}		10	11	23	22
#sig. dm_{SV}			2	22	22
#sig. MCS	10	11	12	27	22
$\overline{QLIKE}$	0.1547	0.1532	0.1530	0.1497	0.1499
#sig. dm_{RV}		18	19	23	23
#sig. dm_{SV}			7	21	20
#sig. MCS	5	6	7	26	20

Note: The table reports the forecasting performance of the different models, where the PV(G) thresholds are obtained by cross-validation minimizing QLIKE, rather than MSE. 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. 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.

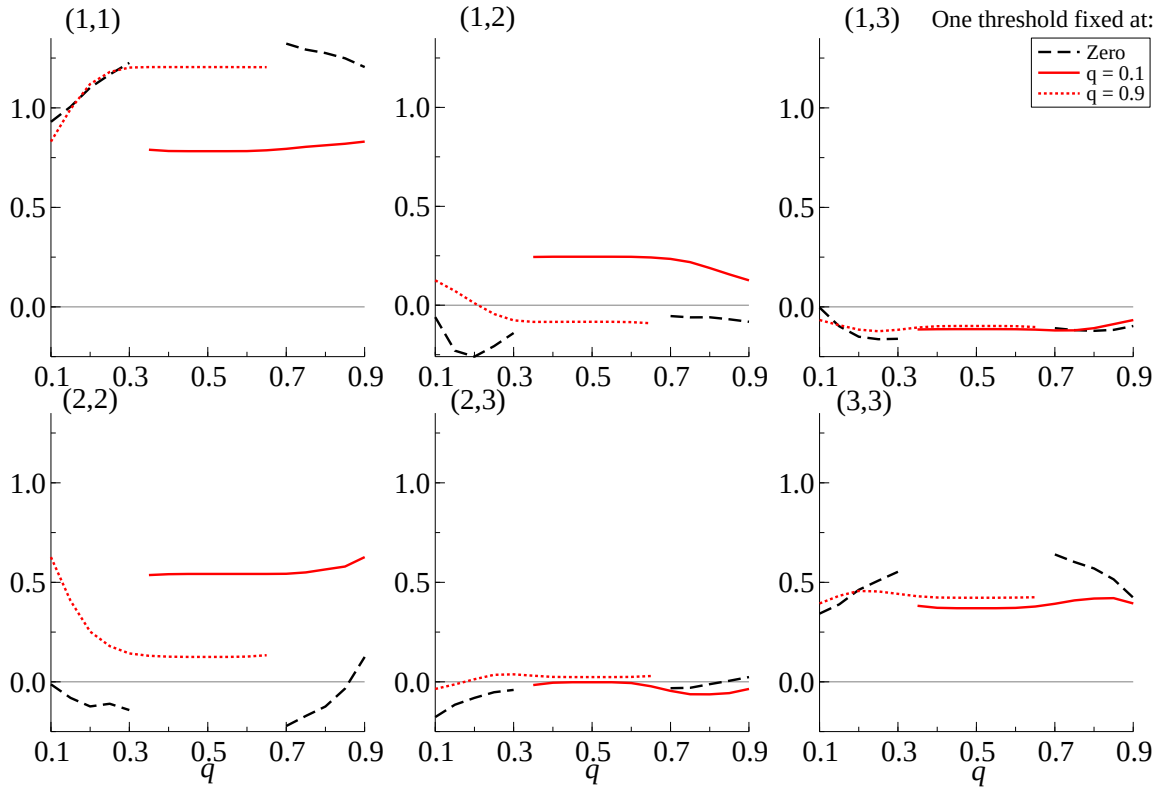
Appendix B The relative importance of partial covariances

In this appendix we report more detailed results on the relative importance of the various partial covariance matrices. We consider random portfolios of cross-sectional dimension $N = 10$, and fix one of two thresholds to either the zero threshold, or the utmost left, and right quantile, $q = 0.1$ or $q = 0.9$. We vary the second threshold across the range of quantiles, sufficiently far from the first to avoid ill-behaved, near empty, partial covariances. In order for the coefficients to be comparable across quantiles, we first standardize all right-hand side variables to be mean zero and variance one. The resulting average parameter estimates for each of the six partial covariance matrices are presented in Figure B.1.

While the figures summarize a large number of PCOV(3)-HAR models, let us start with a general overview, and than delve into some special cases. First, it is clear that the 'concordant' partial covariances play a central role. $\boldsymbol{PCOV}_t^{(1,1)}$, which is related to the most negative returns, generally has the highest standardized coefficient, followed by the positive component $\boldsymbol{PCOV}_t^{(3,3)}$, and finally the variation stemming from returns in the center $\boldsymbol{PCOV}_t^{(2,2)}$. The mixed components typically have negative coefficients.

A number of highlights emerge from the figure. First, consider the dotted lines, where one threshold is fixed at $q_g = 0.9$. This is the model that was found to be important in the univariate case. Here, we see that as the second quantile converges towards the left-tail, the coefficient on the (3,3) element remains mostly constant, while the coefficient on (2,2) increases and that of (1,1) decreases, signifying that the variation coming from large negative returns has moderate predictive content, while the variation stemming from below median returns less far into the tail plays an important role. A similar, but less pronounced version, holds for the reverse case, based on the solid line, where we keep one quantile fixed at 0.1. The center (2,2) component becomes more important when the second quantile goes deeper into the tail. The dashed line, where the first threshold is fixed at zero again corroborates this story. The predictive content is

Figure B.1: Multivariate Models: Parameter Estimates.



Note: We plot average parameter estimates over 100 random portfolios of size $N = 10$, for the PCOV(3)-HAR model, as a function of quantile selection. The right-hand side variables were first standardized to mean zero, variance one. For each of the three line-types, the first threshold is fixed, while the x-axis of the graphs denotes the second quantile. Each panel refers to one of the six partial covariances.

in large, but not extremely large returns. When the second quantile goes deeper into the tails, the (1,1) and (3,3) importance diminishes, while the center (2,2) increases in importance. The most prevalent combination of thresholds, SCOV$^+$ with the second threshold at 0.1, correspond to the point where the dashed line meets the y-axis. Interestingly, the middle component receives close to zero loading. The left-tail component, while typically small in magnitude plays an important role, indicating that the majority of the predictability comes from variation related to joint positive returns.

Appendix C Forecasting industry covariance matrices

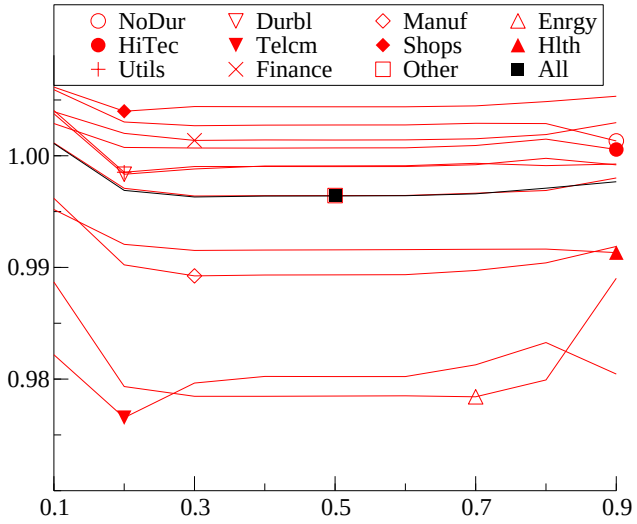
The main paper showed that the PCOV(G*)-HAR model converges to the SCOV$^+$ model when the dimension increases. For small dimensions, we select the complete range of thresholds, while for $N = 50$, we always select the same model. This suggests that there is variability in the optimal threshold across stocks, which are unlikely to coincide in any given random portfolio, resulting in the 'robust' zero threshhold. To corroborate this conjecture, we form portfolios of similar stocks based on Kenneth French's Industry classification.[17]

Figure C.1 provides the MSE loss ratios relative to the RCOV-HAR model averaged across the various industries. Averaged over all stocks, the zero threshold minimizes MSE, while within each industry, a non-zero fixed threshold would have resulted in lower loss. Conversely, the curve for the remainder 'Other' firms closely follows the unconditional curve, and also minimizes at the median. It is difficult to attach meaning to the exact minimizing quantile, but it does suggest that when considering portfolios comprised of similar stocks it may be possible to find a non-zero threshold that significantly beats the zero threshold.

We next consider a forecasting exercise where we use covariance matrices for stocks within a single industry. To ensure we have a sufficiently long sample period, we consider a subset of stocks that have traded continuously over the sample period, resulting in 121 different stocks. We then form industry portfolios, ranging in size from 2 to 22. The resulting forecasting performance is provided in Table C.1. The table shows more convincing forecasting improvements relative to the scenario of random portfolios. Based on MSE, the Diebold-Mariano test of equal forecasting performance for the SCOV and PCOV(G) model is rejected in favor of the PCOV(G) for all portfolios except the remaining set of 'other' firms. Even for the larger portfolio, we do not always select the $SCOV^+$ with a low extra threshold. Moreover, one of the selected thresholds is typically very close to the ex-post optimal shown in Figure C.1.

[17] We use the 10-industry classification provided in Kenneth French's Data Library, and add a separate class for Financial firms (SIC codes 6000-6800).

Figure C.1: Industry MSE Loss Ratios as a function of PCOV(2) threshold



Note: The figure shows the out-of-sample MSE of the PV(2)-HAR model, by industry, relative to the MSE for the RV-HAR model, as a function of the quantile threshold used in the definition of the partial variances. The symbol denotes quantile minimizing the MSE.

Table C.1: Industry forecasting performance

		PCOV	MSE			QLIKE		
	#Stocks	Thresholds	RCOV	SCOV	PCOV(G)	RCOV	SCOV	PCOV(G)
NoDur	9	0.2, 0.9	2.75	2.68	2.62**	5.65	5.64	5.64
Durbl	2	0.1	10.95	10.91	10.85*	0.58	0.58	0.57*
Manuf	22	0.3, 0.9	6.95	6.69	6.58**	27.61	27.33	27.35
Enrgy	8	SCOV, 0.8	9.64	9.06	8.89**	4.38	4.42	4.40*
HiTec	20	SCOV, 0.9	7.32	7.23	7.12**	21.43	21.42	21.40*
Telcm	3	SCOV, 0.1	4.85	4.80	4.72**	9.12	9.07	9.06*
Shops	15	SCOV, 0.1	7.24	7.16	7.15*	12.60	12.55	12.54*
Hlth	10	0.1, 0.9	3.37	3.28	3.22**	7.72	7.72	7.70**
Utils	11	SCOV, 0.1	3.64	3.79	3.70**	8.94	9.01	8.99
Finance	13	0.1, 0.7	11.98	11.87	11.75*	11.51	11.58	11.52**
Others	8	0.2, 0.9	13.89	13.73	13.73	4.30	4.29	4.29

Note: The table shows the average loss of industry covariance matrix forecasts. ** and * denote significance of a Diebold and Mariano (1995) test of PCOV(G)-HAR against SCOV-HAR at the 1% and 5% significance level respectively.

Appendix D Forecasting portfolio variance

This appendix presents forecasting results for the portfolio variance defined and discussed in Section 4.2. We use the three HAR-based models (RV^p, SV^p, PV^p), paralleling those in Section 4. A summary of the results is reported in Table D.1. The MSE of the $PV^p(G^*)$ model is consistently lower than that of the other two models. The improvements over the SV^p HAR are diminishing in N, while the improvements over RV^p are increasing in N. The Model Confidence Set fails to really separate the two decomposed models however. The results based on QLIKE are more mixed, and the decomposed variance models fail to consistently improve over the RV-HAR model.

Table D.1: Multivariate Models: Portfolio Variance Forecasting Results

	MSE			QLIKE		
N	RV^p	SV^p	$PV^p(G^*)$	RV^p	SV^p	$PV^p(G^*)$
2	7.596	7.457	7.150	0.182	0.181	0.184
	(0.88)	(0.92)	(0.98)	(0.70)	(0.89)	(0.44)
5	3.188	2.994	2.962	0.153	0.190	0.166
	(0.81)	(0.93)	(1.00)	(0.84)	(0.66)	(0.47)
10	2.305	2.121	2.116	0.150	0.164	0.184
	(0.65)	(0.84)	(1.00)	(0.98)	(0.34)	(0.21)
20	1.962	1.780	1.772	0.154	0.157	0.155
	(0.42)	(1.00)	(1.00)	(1.00)	(0.09)	(0.44)
50	1.876	1.678	1.676	0.160	0.164	0.163
	(0.19)	(1.00)	(0.94)	(1.00)	(0.12)	(0.30)

Note: The table provides the MSE and QLIKE for the three HAR-based models for the portfolio variance, for various cross-sectional dimensions N. 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.