Supplemental Appendix to:

From Zero to Hero: Realized Partial (Co)Variances

Tim Bollerslev, Marcelo C. Medeiros, Andrew J. Patton, Rogier Quaedvlieg

This version: February 15, 2021

S1 Cross-validation procedure

In this appendix we formally describe the cross-validation (CV) procedure that we rely on in choosing the thresholds. All of our forecasting models can be written as:

$$RV_t = \boldsymbol{x}_{t-1}(q)'\boldsymbol{\phi} + \epsilon_t,$$

where $\boldsymbol{x}_{t-1}(q)$ includes the constant and the additional regressors. Note that the regressors, and thus the forecasts, depend on the chosen quantile q. Let $\mathcal{Q}$ be the set potential values of q. Furthermore, let $\mathsf{V} \subseteq \{1, \ldots, T\}$ denote the indices of the observations in the validation set, and $\mathsf{T} \subseteq \{1, ..., T\}$ denote the indices of the observations in the training set. Often, but not always, $\mathsf{T} := \mathsf{V}^c$. Let $\widehat{\boldsymbol{\theta}}_{\mathsf{T}}(q)$ be the parameter estimate based on the training data T using the quantile $q \in \Lambda \times \mathcal{Q}$. For each q let

$$CV(q, \mathsf{V}) = \sum_{t \in \mathsf{V}} \left[RV_t - \boldsymbol{x}_{t-1}' \widehat{\boldsymbol{\theta}}_{\mathsf{T}}(q) \right]^2,$$

denote the corresponding prediction error over the validation set V. Let $\mathsf{V} = \{\mathsf{V}_1, \ldots, \mathsf{V}_B\}$ be a user specified collection of validation sets (with corresponding training sets $\{\mathsf{T}_1, \ldots, \mathsf{T}_B\}$). The cross validation error for the q is then calculated as

$$CV(q) = \sum_{i=1}^{B} CV(q, \mathsf{V}_i),$$

resulting in the quantile choice,

$$\widehat{q} \in \arg\min_{q \in \mathcal{Q}} CV(q).$$

The final parameter estimate is then found as $\widehat{\boldsymbol{\beta}}(\widehat{q})$ based on all observations $1, \ldots, T$.

Turning to the choice of B and the corresponding $\{\mathsf{V}_1, \ldots, \mathsf{V}_B\}$, there are two standard

ways to split the sample: *exhaustive* and *non-exhaustive* CV. For the exhaustive class of CV methods, leave-v-out CV is the one most commonly used. The idea is to use v observations as the validation set, using the remaining observations to estimate the parameters of the model. This is done for all possible ways of choosing v observations out of T. Thus, $B = \binom{T}{v}$ with each V_i having cardinality v in the above general setting. $\mathsf{T}_i = \mathsf{V}_i^c,\ i = 1, ..., B$. A popular choice is to set $v = 1$, such that $B = T$ and $\mathsf{V_i} = \{i\}$. This is known as the *leave-one-out CV*.

Exhaustive CV is computationally very intensive as it performs cross-validation over many sample splits. A remedy to the computational intensity is to use non-exhaustive CV methods. Among these, *B-fold CV* is the most popular. In B-fold CV, the sample $\{RV_t, \boldsymbol{x}_{t-1}(q)\}_{t=1}^T$ is (randomly) partitioned into B subsamples with "approximately" the same number of observations. Thus, $\mathsf{V}_1, \ldots, \mathsf{V}_B$ have roughly the same cardinality. In fact, one typically chooses each V_i to have T/B observations (up to rounding) such that $\max_{1 \leq i \leq j \leq B} \big| |\mathsf{V_i}| - |\mathsf{V}_j| \big| \leq 1$. In many implementations the B validation groups are disjoint, such that each observation belongs to one group only. Again, $\mathsf{T}_i = \mathsf{V}_i^c,\ i = 1, ..., B$. Setting B equal to 5 or 10 are typical choices. We rely B=5 in the paper.

S2 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 S.1 reports the forecasting performance of the different models, directly mirroring the MSE-based results in Table 1 in the main part of the paper. 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 S.1: Univariate Models: Unconditional Forecasting performance based on QLIKE-CV

	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 and Mariano (1995) 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 (Hansen et al. (2011)), 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.

S3 The relative importance of partial (co)variances

We present in-sample estimation results for the various models based on the SPY below in Table S.1. Since the average 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 of the main part of the paper.

We next assess the relative importance of the multivariate partial covariances. 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

Table S.1: Univariate Models: Parameter Estimates

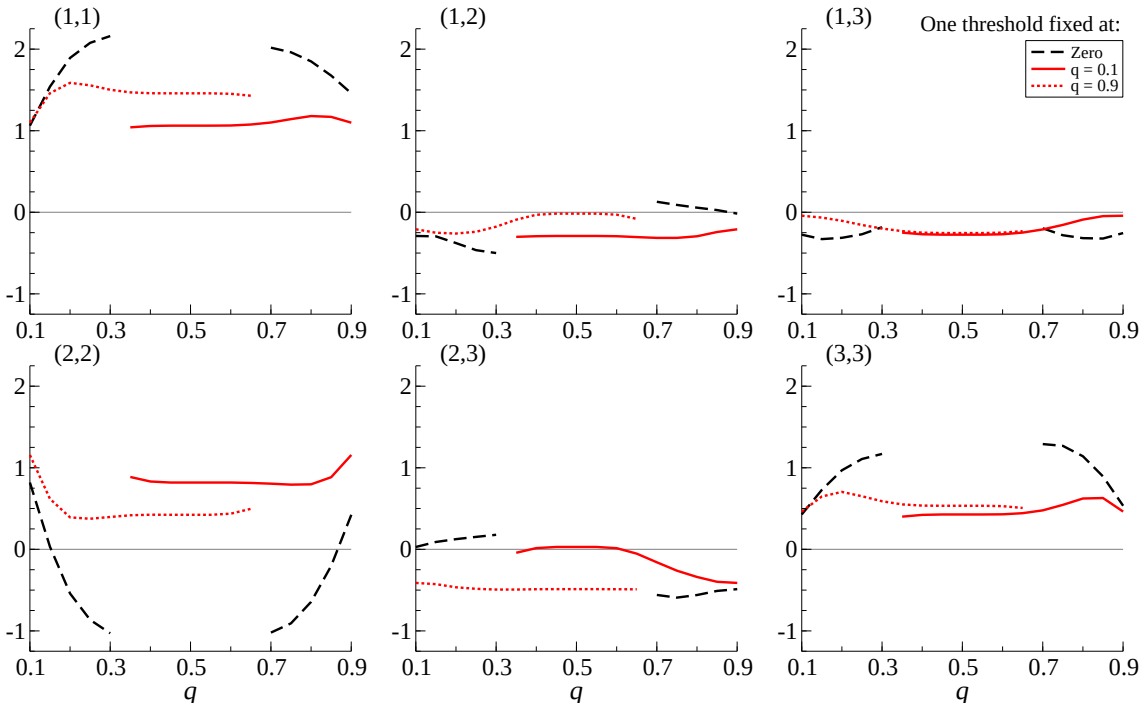
	RV	SV	PV(2)	PV(3)	PV(4)
Selected Quantiles			0.95	0.05	0.15
				0.75	0.50
					0.85
RV_{t-1}	0.444				
	(0.021)				
SV^{-}_{t-1}		0.384			
		(0.017)			
SV^{+}_{t-1}		0.040			
		(0.020)			
$PV^{(1)}_{t-1}$			0.545	0.045	0.159
			(0.022)	(0.021)	(0.024)
$PV^{(2)}_{t-1}$			-0.066	0.417	-0.026
			(0.018)	(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.

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 S.1.

While the figures summarize a large number of PCOV(3)-HAR models, let us start with a

Figure S.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.

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 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.

S4 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.[1]

We 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 S.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 'Energy' firms, where the standard SCOV model is actually selected.

[1] 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).

Table S.1: Industry forecasting performance

		PCOV	MSE			QLIKE		
	#Stocks	Thresholds	RCOV	SCOV	PCOV(G)	RCOV	SCOV	PCOV(G)
NoDur	9	0.1, 0.9	1.75	1.70	1.67**	3.10	3.27	2.84**
Durbl	2	0.1	7.93	7.81	7.67*	0.38	0.38	0.36*
Manuf	22	SCOV, 0.1	4.46	4.11	4.00**	10.86	11.35	10.14**
Enrgy	8	SCOV	7.35	6.00	6.00	2.23	2.21	2.21
HiTec	20	SCOV, 0.9	4.92	4.80	4.65**	9.16	8.54	8.74
Telcm	3	SCOV, 0.1	3.12	3.11	3.07**	4.18	4.18	4.13*
Shops	15	SCOV, 0.1	4.24	4.12	3.99**	5.58	5.62	5.22**
Hlth	10	SCOV, 0.1	3.23	3.05	2.83**	4.92	5.12	4.80**
Utils	11	SCOV, 0.1	2.94	2.94	2.71**	4.15	4.12	3.75**
Finance	13	0.1, 0.9	8.74	8.68	8.44**	6.59	6.75	6.54**
Others	8	0.1, 0.9	10.55	10.55	10.29*	2.49	2.20	1.97**

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.

S5 Forecasting portfolio variance

In addition to forecasting the covariance of a portfolio, the realized partial covariances may 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, $r_{t,k}^{p} = \boldsymbol{w}'\mathbf{r}_{t,k}$, may then be decomposed into realized partial covariances as:

$$RV_t^p \equiv \boldsymbol{w}'\boldsymbol{RCOV}_t\boldsymbol{w} = \sum_{g=1}^{G}\sum_{g'=g}^{G}\boldsymbol{w}'\overline{\boldsymbol{PCOV}}_t^{(g,g')}\boldsymbol{w} \equiv \sum_{g=1}^{G}\sum_{g'=g}^{G}PV_t^{p(g,g')} \tag{S5.1}$$

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 based on all of the portfolio partial variances, as defined in equation (S5.1) above.[2] Since the partial variances represent a complete decomposition of the portfolio variance, this extended representation may be rewritten as:

$$\begin{aligned} RV_t^p &= \phi_0 + \sum_{g=1}^{G}\sum_{g'=g}^{G}\phi_1^{(g,g')}PV_t^{p(g,g')} + \epsilon_t \\ &= \phi_0 + \underbrace{\left(\sum_{g=1}^{G}\sum_{g'=g}^{G}\phi_1^{(g,g')}\frac{PV_{t-1}^{p(g,g')}}{RV_{t-1}^{p}}\right)}_{\phi_{1,t}} RV_{t-1}^p + \epsilon_t, \end{aligned} \tag{S5.2}$$

[2] Of course, since we use Multivariate Kernel estimates on the left-hand side and sub-sampled partial covariance estimates on the right-hand side, the model is not strictly autoregressive.

Figure S.1: Portfolio Variance Models: Implied time-varying parameter

Note: The figure plots the implied time-varying 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.

thus affording a simple time-varying parameter interpretation of the more general model.[3] 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 3 in the main part of the paper. The first combines SCOV with the 0.1

[3]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.

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 S.1 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 0.410, the three other models have average daily coefficients between 0.545 and 0.565, 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 generally 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.

Next, we presents forecasting results for the portfolio variance. We use the three HAR-based models (RV^p, SV^p, PV^p), paralleling those in Section 4 of the main paper. A summary of the results is reported in Table S.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 (Hansen et al. (2011)) 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 S.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	6.986	6.682	6.575	0.128	0.129	0.128
	(0.63)	(0.82)	(0.96)	(0.57)	(0.44)	(0.74)
5	2.404	2.223	2.113	0.106	0.110	0.103
	(0.84)	(0.89)	(1.00)	(0.34)	(0.30)	(0.87)
10	1.895	1.644	1.540	0.118	0.112	0.113
	(0.86)	(0.92)	(1.00)	(0.55)	(0.82)	(0.75)
20	1.445	1.225	1.197	0.124	0.116	0.114
	(0.35)	(1.00)	(1.00)	(0.25)	(0.84)	(0.92)
50	1.353	1.105	1.086	0.124	0.121	0.124
	(0.12)	(1.00)	(1.00)	(0.78)	(1.00)	(0.52)

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.

S6 Conditional Superior Predictive Ability of the Partial Covariance Models

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

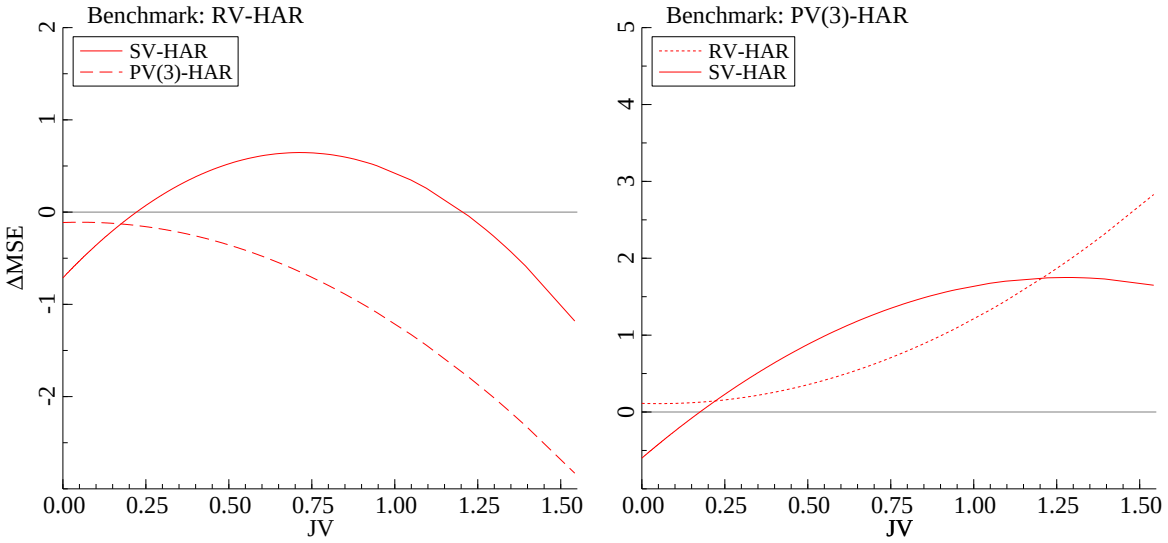
$$H_0 : \mathbb{E}[L_{j,t} - L_{0,t}|X_t = x] \geq 0 \; \forall \; x \in \mathcal{X}, 1 \leq j \leq J,$$

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 in the main part of the paper. 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.[4]

Figure S.2 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

[4]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 S.2: 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.

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 S.2. 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 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,

Table S.2: 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.

and the PV(G)-HAR models surviving the test for all but six of the series.

References

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.

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

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

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.