



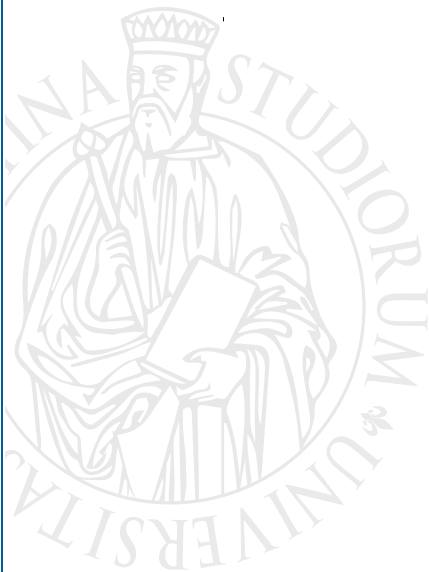
UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

**DISIA**

DIPARTIMENTO DI STATISTICA,  
INFORMATICA, APPLICAZIONI  
"GIUSEPPE PARENTI"

## **Realized Volatility Forecasting: Robustness to Measurement Errors**

Fabrizio Cipollini, Giampiero M. Gallo,  
Edoardo Otranto



**DISIA WORKING PAPER  
2019/04**

© Copyright is held by the author(s).

# Realized Volatility Forecasting: Robustness to Measurement Errors

Fabrizio Cipollini

DiSIA, Università di Firenze, Italy

e-mail: [fabrizio.cipollini@unifi.it](mailto:fabrizio.cipollini@unifi.it)

Giampiero M. Gallo\*

Corte dei Conti and NYU in Florence, Italy

e-mail: [giampiero.gallo@nyu.edu](mailto:giampiero.gallo@nyu.edu)

Edoardo Otranto

Università di Messina, Italy

e-mail: [eotranto@unime.it](mailto:eotranto@unime.it)

## Abstract

In this paper, we reconsider the issue of measurement errors affecting the estimates of a dynamic model for the conditional expectation of realized variance arguing that heteroskedasticity of such errors may be adequately represented with a multiplicative error model. Empirically we show that the significance of quarticity/quadratic terms capturing attenuation bias is very important within an HAR model, but is greatly diminished within an AMEM, and more so when regime specific dynamics account for a faster mean reversion when volatility is high. Model Confidence Sets confirm such robustness both in- and out-of-sample.

**Keywords:** Realized volatility, Forecasting, Measurement errors, HAR, AMEM, Markov switching, Volatility of volatility

**JEL codes:** C22, C51, C53, C58

---

\*Address correspondence to: Giampiero M. Gallo, Corte dei conti, Sezione regionale di controllo per la Lombardia, Via Marina 5, 20121 Milan, Italy, Email: [giampiero.gallo@nyu.edu](mailto:giampiero.gallo@nyu.edu). The views expressed in the article are those of the authors and do not involve the responsibility of the Corte dei conti. The authors thank Fulvio Corsi.

# 1 Introduction

Financial market volatility exhibits persistence, be it implicit in option prices (as in VIX Whaley, 2009), estimated through ultra-high frequency data (UHF $\bar{D}$  – any of the several flavors of realized variance and daily range, cf. Andersen *et al.*, 2006; Alizadeh *et al.*, 2002), or derived as the conditional (one-step ahead) variance of returns in a GARCH-type model (cf. Teräsvirta, 2009). While in the GARCH modeling, measurement and prediction coincide, when volatility measurement is UHF $\bar{D}$ -based, a separate modeling strategy has to be set up, invariably grounded on some sort of an autoregressive scheme. The choice is by now abundant: in general, models aim at capturing empirical regularities in the series, mainly persistence, a slowly moving *average* level of volatility, jumps, the presence of regimes, and possible long memory features.

Whatever the method used to derive it, realized volatility is but an estimator of integrated volatility, and, in spite of it being consistent, the measurement error involved has an impact on the modeling phase. This point is forcefully pursued by Bollerslev *et al.* (2016) (henceforth, BPQ) who illustrate the theoretical consequences of modeling and forecasting realized variance in the presence of additive errors-in-variables, shown to be proportional to the sqrt-integrated quarticity (HARQ). They suggest to remedy an attenuation bias in the Heterogeneous Autoregressive model (HAR, Corsi, 2009) by inserting an interaction term between the lagged realized variance and the (square root of) realized quarticity. A negative sign of the related coefficient ensures that the higher the measurement error, the lower the persistence of the past into the present.<sup>1</sup>

In an empirical perspective, we argue that the sqrt-realized quarticity is strongly correlated with the realized variance (in our panel of 29 assets, the median is 0.938). This has two consequences: measurement errors can be heuristically assumed proportional to the integrated variance and, as such, have a multiplicative representation; next to the HARQ extension to the HAR, substituting the interaction term between the sqrt-realized quarticity and the realized variance with a *curvature* term, i.e. the square of the realized variance, provides a model (HAR2) with a very similar performance.

Modeling the error heteroskedasticity mitigates the attenuation bias: we show a drastic reduction in the statistical significance of the extra terms (either interaction or curvature) both in a HAR-GARCH (Corsi *et al.*, 2008) and in an Asymmetric Multiplicative Error Model (AMEM, Engle, 2002).

While correcting for the presence of measurement errors, both HARQ and HAR2 empirically improve upon a base HAR implying a faster rate of mean reversion in the presence of high volatility. This suggests a model which takes volatility regimes into consideration may mitigate measurement error effects even more. Maheu and McCurdy (2002) use a Markov Switching ARMAX model where regime dependency is found both in the conditional mean and the conditional variance of the realized volatility.<sup>2</sup>

---

<sup>1</sup>Recently, Buccheri and Corsi (2019) address the issue of the relationship between the log of realized variance and the log of integrated variance, resorting to a Kalman filter technique for prediction purposes.

<sup>2</sup>More in general, similar paths have been followed, by Brownlees and Gallo (2010) or Brownlees and Gallo (2011) for splines, and Gallo and Otranto (2015) for Smooth Transition (ST-AMEM) MEMs. Further refinements in HAR present in the literature consider jumps (Andersen *et al.*, 2007); smooth transition dynamics (McAleer and Medeiros, 2008); the possibility of combining HAR and MEMs with Markov Switching

We find that combining both multiplicative error structure and a regime-specific dynamics in a MS-AMEM (Gallo and Otranto, 2015) reduces the relevance of the interaction/curvature terms even more. These results are strengthened by the analysis of prediction results both in- and out-of-sample which show that, while HAR-GARCH is a competitive model, AMEM, but more so, MS-AMEM provide better realized variance forecasts than HAR. This better out-of-sample performance is not a trivial result, also in view of the Hansen (2010) discussion about the trade-off between prediction ability and complexity of the model.

The structure of the paper is as follows: in Section 2 we report the high correlation between the sqrt-realized quarticity and realized variance and we heuristically justify a multiplicative measurement error structure for the relationship between realized and integrated variance. In Section 3 we document that augmenting the HAR with an interaction term involving (sqrt-) quarticity or with a curvature term (the square of realized variance) provides very similar results, confirming the proportionality of the measurement errors to the level of volatility. In Section 4 we propose to refine the specification of the model involving heteroskedasticity, asymmetric behavior in the presence of negative past returns, and regimes: we present three classes of models, the HAR-GARCH, the AMEM, and the MS-AMEM, which variously address these issues. A simulation study (Section 5) shows that when data are generated with multiplicative errors with or without regimes but without extra terms, estimation by HAR generally gives significant interaction or curvature terms, while for the HAR-GARCH it does not. Section 6 provides evidence both in- (6.1) and out-of-sample (6.2) across all models. Concluding remarks follow.

## 2 Multiplicative Measurement Errors in Realized Variance

As noted by BPQ, within a standard framework for the evolution of prices as a diffusion process in continuous time, the Integrated Variance ( $IV_t$ ) is not directly observable but can be measured by one of the several versions of realized variance ( $RV_t$ ) using ultra-high frequency data (Andersen and Benzoni, 2009) with an empirical intra-daily interval  $\Delta$  for sampling returns (e.g. 5 minutes). This implies the measurement error relationship (cf. Barndorff-Nielsen and Shephard (2002) for definitions and details)

$$RV_t = IV_t + \eta_t \tag{1}$$

where, conditional on the Integrated Quarticity  $IQ_t$ ,  $(\eta_t|IQ_t)$  has a distribution with mean zero and variance equal to  $2\Delta IQ_t$ , or, which is the same, the  $\eta_t$ 's are conditionally heteroskedastic with a dependence on the level of  $IQ_t$ . Since Integrated Quarticity and Integrated Variance are both functions of the spot volatility, one can say that the measurement error is dependent on the integrated variance itself, i.e. the measurement errors are more relevant, the higher the variance of the returns.

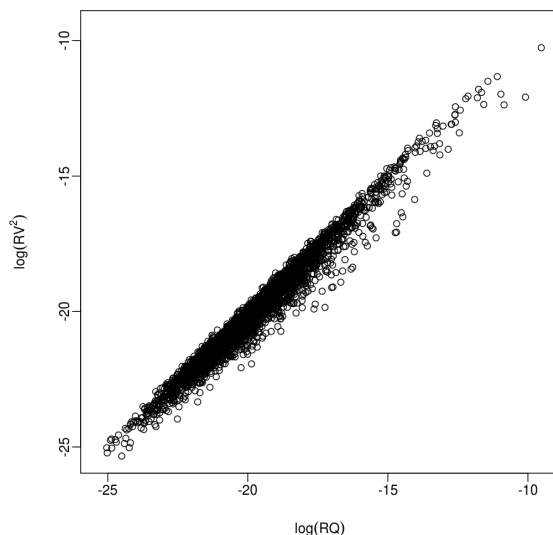
For practical purposes,  $IQ_t$  can be consistently estimated by *Realized Quarticity*  $RQ_t$ ,

---

(Gallo and Otranto, 2015).

which is built from the fourth power of intraday returns in  $\Delta$  intervals and, as such, intuitively related to the squared realized variance. As an empirical evidence of this claim, a scatterplot of the realized quarticity and the squared realized variance in reference to the SPY ETF (Figure 1, in log–log scale to enhance readability) shows that the two series are really close to each other: the Pearson correlation in log–log is 0.983, in levels is 0.909 and between the square roots thereof is 0.954.

Figure 1: SPY over the period Jan. 3, 2003 to Dec. 31, 2015. Scatterplot of Realized Quarticity versus Squared Realized Variance (calculated at 5–minutes intervals) – log–log scale to enhance readability (Pearson correlation=0.983).



This association is further confirmed in Table 1, where (across all 29 tickers in Appendix A) we summarize the five statistics on the distribution of: a. in the first column, the Pearson correlations between  $\ln RQ$  and  $\ln RV^2$ ; b. in the second column, of the Spearman correlations between  $RQ$  and  $RV^2$ ; c. in the third column, the Pearson correlations between  $RQ$  and  $RV^2$ ; d. in the last column, the Pearson correlations between  $RQ^{1/2}$  and  $RV$ . All statistics point to very high values, evidencing that, from a practical point of view,  $RQ$  proves to be proportional to  $RV^2$ , or, which will be relevant in what follows, the square root of  $RQ$  proportional to  $RV$ .

Table 1: Five–number summary of the distributions of calculated Pearson ( $\rho_P$ ) and Spearman ( $\rho_S$ ) correlations across the 29 series, for various pairs of functions of realized quarticity and realized variance.

	$\rho_P(\ln RQ, \ln RV^2)$	$\rho_S(RQ, RV^2)$	$\rho_P(RQ, RV^2)$	$\rho_P(RQ^{1/2}, RV)$
min	0.965	0.959	0.619	0.846
Q1	0.968	0.962	0.797	0.926
median	0.972	0.966	0.864	0.938
Q3	0.974	0.969	0.943	0.954
max	0.985	0.980	0.986	0.975

The empirical evidence shown in the last column of Table 1 suggests that, for the respective theoretical quantities,  $IQ^{1/2}$  can be heuristically assumed proportional to  $IV$ , i.e.  $IQ^{1/2} \approx \delta IV$ , with  $\delta$  very close to 1. As a consequence, an alternative representation of the measurement error relationship in Equation (1) is possible:

$$RV_t = IV_t + \eta_t = IV_t + \sqrt{2\Delta}IQ_t^{1/2}z_t \approx IV_t + \sqrt{2\Delta}\delta IV_t z_t = \underbrace{IV_t(1 + \sqrt{2\Delta}\delta z_t)}_{\varepsilon_t}, \quad (2)$$

where  $z_t$  is an iid error term with mean zero and unit variance. This convenient manipulation provides an expression which highlights the proportional (i.e. heteroskedastic) nature of the measurement errors<sup>3</sup>.

### 3 The HAR Approach Extended

When modeling  $IV_t$  on the basis of observable past values of  $RV_t$ , BPQ discuss the attenuation bias in the first order autoregressive coefficient induced by measurement errors in the HAR framework (Corsi, 2009). BPQ suggest an augmented HAR, labeled HARQ, where a component dependent on the square root of the quarticity is aimed at capturing the measurement error:

$$RV_t = \omega + \underbrace{(\alpha_D + \alpha_E RQ_{t-1}^{1/2})}_{\alpha_{1,t-1}} RV_{t-1} + \alpha_W \overline{RV}_{t-(2:5)} + \alpha_M \overline{RV}_{t-(6:22)} + \varepsilon_t. \quad (3)$$

$\overline{RV}_{t-(h:k)}$  is the mean of  $t-h$  to  $t-k$  lagged  $RV$  values.<sup>4</sup>  $\alpha_E$  is assumed to be non-positive, so that  $\alpha_{1,t-1} \leq \alpha_D$  and non-increasing in  $RQ_{t-1}^{1/2}$ . An interpretation of how this model works in practice<sup>5</sup> is that the information in the past  $RV_{t-1}$  is purposely given less weight in days where the  $RQ$  is high, a feature which is related to the heteroskedasticity in the measurement errors.

The heart of this model lies on the extra term  $RQ_{t-1}^{1/2}RV_{t-1}$  with a coefficient  $\alpha_E$  expected to be negative. This means that the reaction to the lagged realized variance is smaller (less persistence) in the presence of an interaction term between (the square root of) the realized quarticity and the realized variance.

The high correlation between  $RQ^{1/2}$  and  $RV$  motivates the HAR2 alternative specification

$$RV_t = \omega + \underbrace{(\alpha_D + \alpha_E RV_{t-1})}_{\alpha_{1,t-1}} RV_{t-1} + \alpha_W \overline{RV}_{t-(2:5)} + \alpha_M \overline{RV}_{t-(6:22)} + \varepsilon_t \quad (4)$$

Comparison between (3) and (4) shows that the empirical behavior of the two models would

<sup>3</sup>Multiplicative measurement errors are discussed by Carroll *et al.* (2006) in a different context sharing the similarity that the errors are proportional to the latent variable.

<sup>4</sup>For better interpretation purposes, we prefer to avoid overlaps among regressors, and hence lag 1 is not included in the weekly average, and lags 1 to 5 are not included in the monthly one.

<sup>5</sup>In their empirical results obtained on data for the S&P500 and 27 constituents of the Dow Jones Index, BPQ find evidence of negative significant  $\alpha_E$ , and of an improved forecasting performance of the HARQ over HAR.

be close to one another. Even though both models correct for the presence of measurement errors in the base HAR,

- the Q specification highlights such a presence with the interpretation that lagged volatility observations with high quarticity are more affected by measurement errors and should receive less weight in forecasting;
- the 2 specification highlights the idea that volatility dynamics is linkable to market behavior: bursts of volatility are shortly lived, hence, in the presence of high lagged volatility, the model would hasten the mean reversion by paying less attention, so-to-speak, to those observations.

We examine 28 components of the DJ30 Industrial Index plus the SPY ETF <sup>6</sup> over the period Jan. 3, 2003 to Dec. 31, 2015 (3273 observations total). We build the series of realized variances from TAQ data, cleaned according to the Brownlees and Gallo (2006) procedure. The first 2015 observations (up to Dec. 31, 2010) serve for the in-sample analysis, while we use the rest for an out-of-sample rolling analysis: we estimate on a 8-year period (starting from 2003-2010) and we produce one-step ahead predictions for the following year. As a leading example we will work on the SPY ETF which mimics the behavior of the Standard and Poor's 500.

In Table 2 we report the estimates of the base HAR, the HARQ and the HAR2 for three different choices of sampling intervals in the construction of the realized variance of SPY ETF, as a robustness check on the sensitivity of the curvature effect to the construction of the series. The results show the significance of the  $\alpha_E$  parameters across the Q and 2 specifications, confirming the need to augment the base HAR model. An improvement in Mean Square Error (MSE) is noticeable and substantially equivalent across the two augmented specifications, accompanied by the disappearance of significance of the monthly coefficient  $\alpha_M$ . An opposite behavior is shown by the Quasi-Likelihood (QL) statistics, which increases in the augmented specifications. The results are consistent across sampling interval choices.

The observational equivalence between the two models HARQ and HAR2, and the characteristic behavior relative to the HAR model are shown in Figure 2<sup>7</sup> for the October 6–17, 2008 highly turbulent period. The left-hand side panel shows the behavior of the time-varying coefficient  $\alpha_{1,t-1}$  (including the constant level for HAR), which clearly drops in the presence of the burst of volatility on October 10 (a Friday) which reflects on the model predictions for Oct. 13; the right-hand side panel shows how the fitted behavior is modified accordingly in the augmented models, avoiding the HAR overprediction.

When the analysis is repeated across the panel of DJ30 components we get similar results, in the sense that (including the S&P500) we get 25 out of 29 cases in which  $\alpha_E$  is significant (at 1% significance level) in the HARQ model and 26 out of 29 in the

---

<sup>6</sup>Cf. Table 8 in the Appendix for the list of tickers of realized variances available over the entire sample period. Two components currently present in the index were excluded, Travelers and Visa, because of incomplete data.

<sup>7</sup>We mirror Figure 2 (center and right panels) of Bollerslev *et al.* (2016).

Table 2: Estimates of the HAR, HARQ and HAR2 on the realized variance of the Standard & Poor’s SPY ETF. Estimation period: Jan. 3, 2003 – Dec. 31, 2010. The estimates are repeated across the intra-daily sampling periods to construct the realized variance series. Newey–West HAC standard errors in parenthesis. The rows  $MSE\%$  and  $QL\%$  show the corresponding error measures in percentage of the value reported by the base HAR.

	1 minute			5 minutes			10 minutes		
	Q	2		Q	2		Q	2	
$\omega$	0.274 (0.146)	-0.173 (0.137)	-0.227 (0.175)	0.228 (0.106)	-0.167 (0.123)	-0.231 (0.176)	0.213 (0.091)	-0.228 (0.167)	-0.228 (0.167)
$\alpha_D$	0.285 (0.119)	0.931 (0.130)	0.895 (0.158)	0.331 (0.092)	0.856 (0.106)	0.843 (0.149)	0.367 (0.081)	0.851 (0.157)	0.879 (0.175)
$\alpha_E$		-0.350 (0.066)	-0.456 (0.089)		-0.382 (0.051)	-0.527 (0.114)		-0.447 (0.123)	-0.522 (0.148)
$\alpha_W$	0.470 (0.161)	0.273 (0.096)	0.315 (0.098)	0.443 (0.148)	0.284 (0.089)	0.314 (0.093)	0.398 (0.132)	0.271 (0.054)	0.256 (0.065)
$\alpha_M$	0.154 (0.084)	-0.060 (0.118)	-0.051 (0.123)	0.144 (0.062)	0.002 (0.092)	0.015 (0.111)	0.154 (0.058)	0.047 (0.052)	0.037 (0.066)
$MSE$	54.8	46.2	46.6	42.1	36.8	37.8	38.8	35.1	35.0
$QL$	0.159	0.182	0.223	0.171	0.190	0.219	0.176	0.335	0.444
$MSE\%$	100.0	84.3	85.0	100.0	87.4	89.8	100.0	90.5	90.2
$QL\%$	100.0	114.5	140.3	100.0	111.1	128.1	100.0	190.3	252.3

HAR2 specification for the five-minute sampling interval.<sup>8</sup> One may note that when the realized variance is made available but not the underlying ultra-high frequency data a HAR2 model may be a good patch for the HAR; cases in point are the well-known Realized Library (Heber *et al.*, 2009), or alternatives such as the range derived measures (Garman and Klass, 1980).

Focusing on model diagnostics, though, we find that all three specifications have substantial residual correlation. Still summarizing across the 29 variances, we find significance of the Ljung–Box joint test statistic at lag 5 for residuals, absolute residuals and squared residuals in a number of cases reported in Table 3. This finding is consistent with the findings by Corsi *et al.* (2008) about the presence of clustering in the residuals of HAR.

## 4 Heteroskedasticity, Asymmetry and Regimes

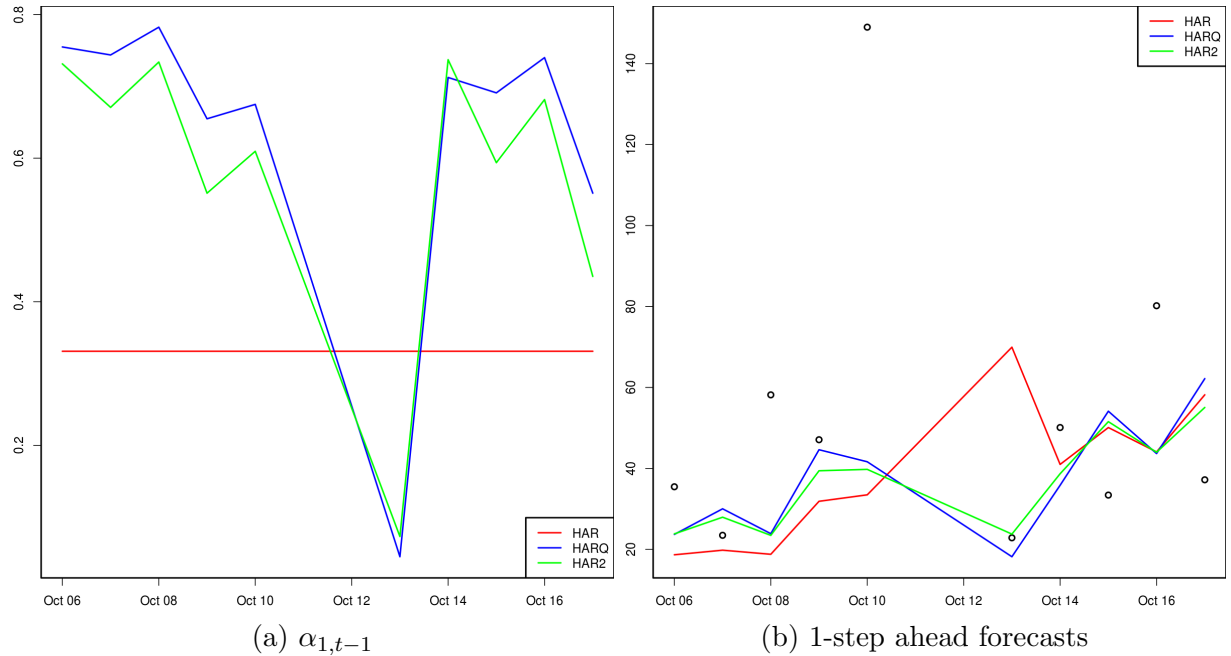
### 4.1 The HAR-GARCH

The issue of volatility of volatility playing a role in modeling volatility was recognized in the HAR family by Corsi *et al.* (2008) who refined the features of the HAR model adding a GARCH structure on the innovation term: such a HAR-GARCH model has the advantage of accounting for clustering in the residuals, which is indeed capturing the time-varying

<sup>8</sup>Results change only slightly at different sampling frequency, being 29 and 28, respectively, in the one-minute case, and 27, respectively, 24 in the ten-minute case.



Figure 2: SPY: Examples of estimated  $\alpha_{1,t-1}$  for different HAR models (left panel) and corresponding actual values (open dots) and volatility forecasts (right panel). October 2008, cf. Fig 2 in BPQ.



features of volatility of volatility. In the context of the discussion here, considering a general specification which can be augmented for interaction/curvature terms, such a model can be written as

$$RV_t = \omega + (\alpha_D + \alpha_E X_{t-1})RV_{t-1} + \alpha_W \overline{RV}_{t-(2:5)} + \alpha_M \overline{RV}_{t-(6:22)} + \sqrt{h_t} z_t, \quad (5)$$

where  $\varepsilon_t \equiv \sqrt{h_t} u_t$  has the usual structure for the innovation under GARCH where  $h_t = E(\varepsilon_t^2 | \mathcal{I}_{t-1})$  provided that  $z_t \sim (0, 1)$ , and  $h_t$  has the usual conditionally autoregressive (with variants) specification. The base model corresponds to  $X_{t-1} = 0$ , while the HARQ-GARCH has  $X_{t-1} = RQ_{t-1}^{1/2}$  and the HAR2-GARCH has  $X_{t-1} = RV_{t-1}$ , the latter two being the augmented models which allow for a time-varying coefficient on  $RV_{t-1}$ , as before.

In Table 4, we report results for the HAR-GARCH model. After accounting for heteroskedasticity in the residuals, the striking result is the statistical insignificance of the interaction/curvature terms, confirmed by a very small gain in terms of loss functions. Comparing the simpler versions of either model, we notice a higher relevance given to the daily term, and smaller coefficients for the weekly and monthly terms. We keep that in mind when interpreting the relative performance both in- and out-of-sample between the two models.

Table 3: Number of assets for which the Ljung–Box statistics at lag 5 turns out to be statistically significant at 1% significance level. Estimation period: Jan. 3, 2003 – Dec. 31, 2010. The estimates are repeated across the intra–daily sampling frequency at which the realized variance series is built.  $LB_5$  is calculated on residuals,  $LBa_5$  on absolute residuals, and  $LB2_5$  on squared residuals.

	1 minute			5 minutes			10 minutes		
	HAR	HARQ	HAR2	HAR	HARQ	HAR2	HAR	HARQ	HAR2
$LB_5$	27	27	26	27	26	26	26	26	25
$LBa_5$	28	28	28	29	29	29	29	29	29
$LB2_5$	20	17	16	27	22	21	24	22	21

## 4.2 The AMEM

Equation 2 shows that a model with multiplicative errors captures the heteroskedastic nature of the measurement errors in the realized volatility. A suitable approach to modeling  $RV_t$  in a forecasting perspective (i.e. conditioning on the information set  $\mathcal{I}_{t-1}$  at  $t-1$ , as before) is the MEM(Engle, 2002). A quite general formulation for  $RV_t$  is

$$RV_t = \mu_t \varepsilon_t \quad \mu_t = \mu(\mathcal{I}_{t-1}; \theta) \quad \varepsilon_t \stackrel{i.i.d.}{\sim} D^+(1, \sigma^2).$$

Here,  $\varepsilon_t$  is an i.i.d. innovation term following a generic distribution over a non–negative support with unit mean and variance  $\sigma^2$ . The model has the important feature (in this context) that the dependent variable is conditionally heteroskedastic:

$$E(RV_t | \mathcal{I}_{t-1}) = \mu_t \quad V(RV_t | \mathcal{I}_{t-1}) = \sigma^2 \mu_t^2$$

In what follows, we will adopt an Asymmetric MEM (AMEM) specification for  $\mu_t$  (Engle and Gallo, 2006), where past values of  $RV_{t-1}$  have a different impact on the current  $RV_t$  according to the sign of past returns on that asset. Following Engle and Gallo (2006) we adopt a Gamma distribution for  $\varepsilon_t$  dependent on just one parameter  $a$ , as implied by the unit expectation constraint. We have

$$\begin{aligned} \mu_t &= \omega + \beta_1 \mu_{t-1} + (\alpha_1 + \alpha_E X_{t-1}) RV_{t-1} + \gamma_1 RV_{t-1}^{(-)} \\ \varepsilon_t | \mathcal{I}_{t-1} &\sim \text{Gamma}(a, 1/a) \end{aligned}$$

Again, the base AMEM model corresponds to  $X_{t-1} = 0$ , setting  $X_{t-1} = RQ_{t-1}^{1/2}$  delivers the AMEMQ, while  $X_{t-1} = RV_{t-1}$  corresponds to the AMEM2, with the latter two having the interaction/curvature terms, respectively.

## 4.3 The MS-AMEM

In all augmented models the main feature is a faster mean reversion behavior: thus an additional question arises as to whether the mean reversion occurs to a constant level or to a time–varying level. To that end, among the many possible specifications which allow

Table 4: Coefficient estimates on SPY realized variance for the HAR-GARCH models. The rows  $MSE\%$  and  $QL\%$  show the corresponding error measures in percentage of the value reported by the base HAR-GARCH. Sample period: Jan. 3, 2003–Dec. 31, 2010.

	HAR-GARCH		
		Q	2
$\omega$	0.100 (0.034)	0.096 (0.034)	0.099 (0.034)
$\alpha_D$	0.439 (0.036)	0.444 (0.038)	0.439 (0.037)
$\alpha_E$		-0.067 (0.086)	-0.006 (0.187)
$\alpha_W$	0.363 (0.030)	0.361 (0.030)	0.363 (0.030)
$\alpha_M$	0.103 (0.037)	0.103 (0.036)	0.103 (0.036)
Const	0.039 (0.011)	0.029 (0.011)	0.029 (0.011)
ARCH	0.236 (0.030)	0.234 (0.030)	0.236 (0.030)
GARCH	0.763 (0.004)	0.765 (0.004)	0.763 (0.004)
$MSE$	42.5	40.5	42.4
$QL$	0.165	0.164	0.165
$MSE\%$	100.0	95.3	99.8
$QL\%$	100.0	99.4	100.0

for a low frequency component which slowly evolves through time (cf., for example, Engle and Rangel, 2008; Brownlees and Gallo, 2010; Barigozzi *et al.*, 2014; Gallo and Otranto, 2017), we choose the Markov Switching AMEM (cf. Gallo and Otranto, 2015, for details) again allowing for all three versions, the base MS-AMEM, and the augmented ones, MS-AMEMQ and MS-AMEM2.

$$\begin{aligned}\mu_{t,s_t} &= \omega_{s_t} + \beta_{s_t}\mu_{t-1,s_{t-1}} + (\alpha_{s_t} + \alpha_E X_{t-1})RV_{t-1} + \gamma_{s_t}RV_{t-1}^{(-)} \\ \varepsilon_t | s_t, \mathcal{I}_{t-1} &\sim \text{Gamma}(a_{s_t}, 1/a_{s_t})\end{aligned}$$

where  $s_t \in \{1, 2, 3\}$ <sup>9</sup> and  $P(s_t = j | s_{t-1} = i) = p_{ij}$  ( $\sum_i p_{ij} = 1$ ), and  $X_t$  is defined as before.

The estimation results from the AMEM, AMEMQ, AMEM2 and MS-AMEM (with three states) are reported in Table 5: to be noticed the insignificance of the interaction and curvature effects (coefficient  $\alpha_E$ ), leading to a substantial equivalence of the three models, while the asymmetric effects play an important role from a statistical point of view. For the MS-AMEMQ and MS-AMEM2, for this sample period and this asset, the estimation converged to  $\hat{\alpha}_E$  being numerically equal to zero,<sup>10</sup> hence we do not report the estimation results in the table. In terms of MSE, AMEMQ shows some improvement over the base model, despite the non significant  $\alpha_E$ ; MS-AMEM shows better performance than the other models for both loss functions.

## 5 A Simulation Study

To further the discussion about the behavior of these models, we want to see how the interaction/curvature effects behave in the presence of heteroskedasticity and slow moving low frequency component. We have performed a simple simulation exercise in which the data are generated by an AMEM or by an MS-AMEM.<sup>11</sup> In view of the BPQ results on measurement errors, the AMEM and the MS-AMEM are not the true models, but in practical applications they result in good diagnostics for the estimated residuals and so they lend themselves to be good benchmarks.

We test how many times the  $\hat{\alpha}_E$  coefficient turns out to be significantly negative at two different significance levels, using both OLS standard errors and HAC (Newey and West, 1987) in estimated HAR2 models for different sample periods  $T$ .

The results, reported in Table 6, show fairly clearly that in the presence of heteroskedasticity, and even more so when also regimes are introduced, one is very likely to get a significant  $\hat{\alpha}_E$  coefficient, even using HAC robust standard errors. The results on the importance of a proper inclusion of heteroskedasticity are confirmed when considering the results obtained estimating with HAR-GARCH (bottom panel of the same table). Also in this case,

<sup>9</sup>We adopt three states as they proved suitable to capture the features of the data for similar sample periods (cf. the results in Gallo and Otranto, 2015).

<sup>10</sup>This is not always true across assets and estimation periods, hence we can still evaluate the performance of the augmented models in terms of in- and out-of-sample prediction.

<sup>11</sup>The coefficients were set at values of the estimated for SPY – cf. Table 5. We performed 1000 replications for each combination of  $T$  and DGP. The results are similar for the HARQ model and are not reported here.

Table 5: Coefficient estimates for SPY realized variance within the AMEM and MS-AMEM groups. The rows  $MSE\%$  and  $QL\%$  show the corresponding error measures in percentage of the value reported by the base AMEM. Sample period: Jan. 3, 2003–Dec. 31, 2010.

	AMEM	AMEMQ	AMEM2	MS(3)-AMEM
$\omega$	0.177 (0.008)	0.176 (0.008)	0.177 (0.008)	0.138 (0.009)
$k_3$				0.945 (0.093)
$\alpha_1$	0.482 (0.033)	0.485 (0.033)	0.482 (0.033)	0.155 (0.018)
$\alpha_3$				0.005 (0.016)
$\alpha_E$		-0.059 (0.082)	-0.0002 (0.004)	
$\beta_1$	0.332 (0.032)	0.330 (0.032)	0.332 (0.032)	0.562 (0.025)
$\beta_2$				0.884 (0.019)
$\beta_3$				0.875 (0.013)
$\gamma_1$	0.189 (0.031)	0.190 (0.032)	0.189 (0.031)	0.128 (0.021)
$\gamma_2$				0.116 (0.019)
$\gamma_3$				0.114 (0.027)
$p_{11}$				0.938 (0.019)
$p_{12}$				0.049 (0.018)
$p_{21}$				0.059 (0.022)
$p_{22}$				0.919 (0.020)
$p_{31}$				0.071 (0.006)
$p_{32}$				0.078 (0.016)
$a_1$	3.777 (0.232)	3.777 (0.232)	3.777 (0.231)	5.809 (0.331)
$a_2$				7.213 (0.406)
$a_3$				2.020 (0.476)
$MSE$	45.1	41.7	45.1	39.0
$QL$	0.155	0.154	0.155	0.103
$MSE\%$	100.0	92.5	100.0	86.5
$QL\%$	100.0	99.4	100.0	66.5

Table 6: Simulation results. Percentage of times the  $\hat{\alpha}_E$  coefficient turns out to be significantly negative at different significance levels in an HAR2 model (top panel with OLS and HAC standard errors) and in an HAR2-GARCH model (bottom panel with ML and Sandwich based standard errors). Data simulated under AMEM and MS-AMEM DGPs (with estimated SPY coefficients – cf. Table 5). 1000 replications each.

Model	DGP	$T$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
HAR2	AMEM		OLS		HAC	
		1000	83.0	78.8	63.9	45.8
		2000	85.8	81.9	63.8	47.6
	3000	86.0	82.9	64.3	48.0	
	MS-AMEM	1000	98.5	97.3	90.5	80.5
		2000	99.1	98.4	92.2	82.2
3000		99.2	98.6	92.9	84.1	
HAR2-GARCH	AMEM		ML		Sandwich	
		1000	15.5	4.6	12.7	3.7
		2000	17.0	6.2	14.2	4.6
	3000	15.3	6.4	13.2	5.1	
	MS-AMEM	1000	39.2	25.3	23.6	9.9
		2000	45.5	29.8	27.7	10.6
3000		51.6	33.6	31.1	13.3	

we distinguish between calculations of standard errors according to the inverse of the Hessian (ML), or robust (sandwich estimator). The rates of failures to reject are much lower, less so when the data are generated according to the MS-AMEM, signifying in both cases that the presence of regimes in a multiplicatively heteroskedastic model may be mistaken for the significance of curvature (or interaction) terms.

## 6 Comparing Model Performances

### 6.1 In-sample

When brought to the evaluation of their empirical performance, the models presented above which somehow correct for heteroskedasticity are estimated on all our tickers (cf. the Appendix for a list), with results synthesized in Table 7: out of the 29 assets considered, we report the number of statistically significant (one-sided, 1% as before)  $\hat{\alpha}_E$  coefficients entering the equations for the realized variances, as well as the times the Ljung Box diagnostics (at lag 5) on residuals (respectively in levels, in absolute values and squared) are significant at 1%. The behavior across models is fairly similar, in that we notice a much lower presence of significant  $\hat{\alpha}_E$  and an overall better performance in terms of residual diagnostics relative to the base HAR specifications discussed before.<sup>12</sup>

<sup>12</sup>One could think that the presence of significant interaction/curvature terms may actually signal some remaining mis-specification in the models, in a way reminiscent of a White (1980)-type test equation for

Table 7: Summary of the estimation results for the HAR-GARCH, AMEM and MS-AMEM in their base and augmented versions. In-sample period: Jan. 3, 2003–Dec. 31., 2010. Number of assets for which the estimated  $\alpha_E$  coefficients are significant (one-sided test, 1% significance) across the 29 asset realized variances, and for which the 5-lag Ljung–Box statistics on residuals ( $LB_5$ ), absolute residuals ( $LBa_5$ ), and squared residuals ( $LB2_5$ ) are statistically significant at 1%.

	HAR- GARCH	HARQ- GARCH	HAR2- GARCH	AMEM	AMEMQ	AMEM2	MS- AMEM	MS- AMEMQ	MS- AMEM2
$\alpha_E$		8	3		7	2		3	0
$LB_5$	7	7	9	2	2	2	9	7	7
$LBa_5$	5	3	4	3	3	3	2	3	3
$LB2_5$	0	0	0	0	0	0	0	0	0

The models can be compared in reference to the Model Confidence Set (MCS) procedure of Hansen *et al.* (2003) and Hansen *et al.* (2011), on the basis of the Mean Squared Error (MSE) and the Quasi-Likelihood (QL), which are consistent loss functions in the sense of Patton (2011). The test statistic used to calculate the MCS is the semi-quadratic  $T_{SQ}$  (less conservative with respect to the range statistic Clements *et al.*, 2009):

$$T_{SQ} = \sum_{i \neq j \in \mathcal{M}} \frac{\bar{l}_{ij}^2}{\widehat{var}(\bar{l}_{ij})} \quad (6)$$

where  $\bar{l}_{ij}$  is the mean of the loss differentials between model  $i$  and model  $j$  belonging to the set of models  $\mathcal{M}$ , and its variance is obtained from the bootstrap procedure (10,000 replications) described in Hansen *et al.* (2003).

In Figure 3 we report the graphs by asset showing the values of the MSE for each model, visually enclosing the set of models with the best in-sample prediction performance according to the MCS procedure; note that scaling the loss differential by its variance may lead at times to models with a lower loss function not being included in the MCS.

For most assets the MS class of models shows a lower MSE (excluding just two tickers, Goldman Sachs – GS, where the HARQ model is the best one, and Intel – INTC, where the HAR and HAR–GARCH families have a better performance);<sup>13</sup> this is expected, given its richer parameterization. As far as the MCS is concerned, for all tickers does the MS class enter in the MCS: 14 out of 29 times without competitors, and, with just one exception (Apple – AAPL), the MS model without interaction/curvature term has the same performance as with. By contrast, for 11 assets all the models belong to the best MCS.

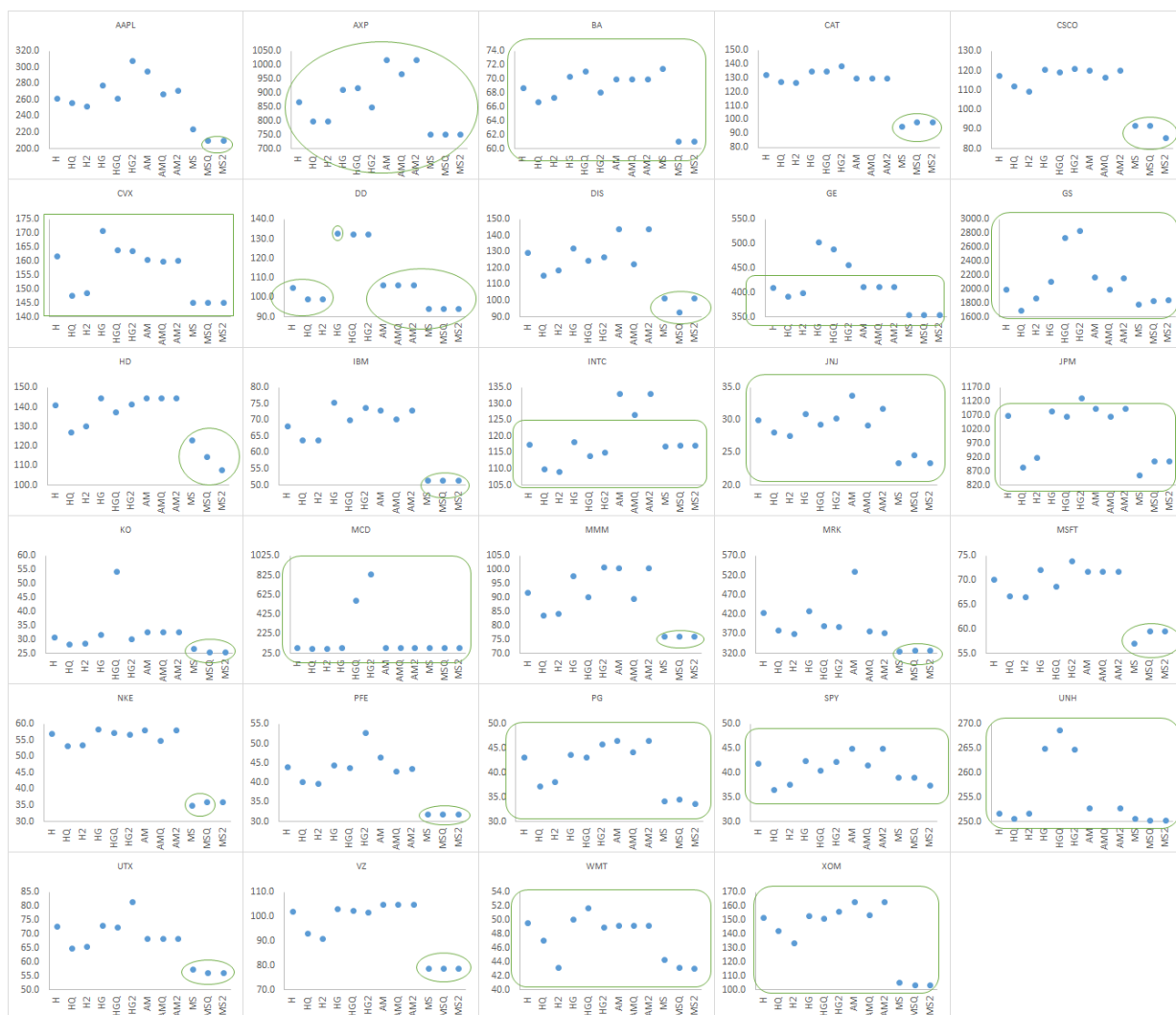
A similar graph, Figure 4, reports the results for the QL loss function showing a more clear-cut behavior in the performance of the models: at least one member of the MS–AMEM family enters the MCS, with the base model entering the best set for 14 assets,

---

model diagnostics.

<sup>13</sup>Being based on LS-type estimation, it is not surprising to find a good MSE performance of HAR (and also of HAR-GARCH).

Figure 3: In-sample MSE loss function for 12 models and 29 assets. The enclosures contain, for each asset, the models belonging to the MCS at 5% significance level. Models are represented by short acronyms in this order: H for HAR, HQ for HARQ, H2 for HAR2, HG for HAR-GARCH, HGQ for HARQ-GARCH, HG2 for HAR2-GARCH, AM for AMEM, AMQ for AMEMQ, AM2 for AMEM2, MS for MS-AMEM, MSQ for MS-AMEMQ, MS2 for MS-AMEM2.





just the two models with interaction/curvature term together for 6 assets, with the Q model by itself in 4 cases and the 2 model in 5 cases.

Overall these results seem to suggest that the effects of a measurement error in the realized variance show up in-sample even with more sophisticated models, although an explicitly heteroskedastic model (allowing for a time-varying local average level of volatility) is able to mitigate such effects.

## 6.2 Out-of-sample

The model comparison is performed similarly for the out-of-sample forecasting capability. As mentioned before, we estimate models starting from 2003-2010 and produce one-step ahead forecasts for the following year; we then move the sample period ahead by one year (i.e. 2004-2011) and we forecast the year 2012, repeating the procedure until we estimate over 2007-2014 and we forecast over 2015.

Working with these forecast errors, analogously to what we presented in the in-sample case, we build MSE and QL loss functions, reported in Figure 5, respectively, Figure 6, which give a somewhat different outlook on the nature of the results. The higher variability of the results for the MSE is confirmed also in this exercise. For the MSE, although a few MCS include all models, all of them include the three MS models: we have 18 instances in which they are the only ones in the set, 1 (DD) in which the base MS is in the set by itself, 1 (MCD) in which only the MS models with additional terms significantly improve the set; moreover, 22 MCS's exclude the HAR models. The picture is sharper for QL in Figure 6 since no HAR models enter the MCS, with the loss function values of the three MS models being consistently the lowest: in 20 out of 29 cases all three appear in the MCS, in the remaining nine, 2 involve the base MS model by itself (IBM and UTX), 4 the MS model with the Q term (BA, CAT, GS and MCD), 2 the MS model with the 2 term (MRK and UNH), and 1 the MS models with either additional term, but with values of the loss functions no better than the base model. Moreover, within the HAR family higher loss functions are present for the HARQ and the HAR2 than the HAR in 13 cases for the MSE (see, e.g. GS, JPM and VZ). These findings are more apparent in terms of QL with 18 cases where the introduction of the interaction/curvature term worsens the out-of-sample performance of the model (see, e.g., JPM, UTZ, and VZ) and in most others does not improve substantially the base HAR model.

The introduction of heteroskedasticity in HAR, or adopting AMEM models with or without MS dynamics produces similar values of MSE and QL within each model family, with the latter class models always showing their dominance in both exercises, in some cases with non significant differences with respect to other models when the loss function is MSE. The cases of MRK and XOM in Figure 5 are interesting: in this case the presence of measurement error terms improve the forecasting performance, but a MS-AMEM without them has the same performance. In Figure 6 only the MS family enters in the best set; in 20 cases the best set contains all three models, whereas only in 7 cases there is an improvement due to the measurement error terms.

It may be important to stress that the superiority of MS models is not a trivial result in out-of-sample terms and cannot be attributed to overparameterization: the better per-

Figure 4: In-sample QL loss function for 12 models and 29 assets. The enclosures contain, for each asset, the models belonging to the MCS at 5% significance level. Models are represented by short acronyms in this order: H for HAR, HQ for HARQ, H2 for HAR2, HG for HAR-GARCH, HGQ for HARQ-GARCH, HG2 for HAR2-GARCH, AM for AMEM, AMQ for AMEMQ, AM2 for AMEM2, MS for MS-AMEM, MSQ for MS-AMEMQ, MS2 for MS-AMEM2.

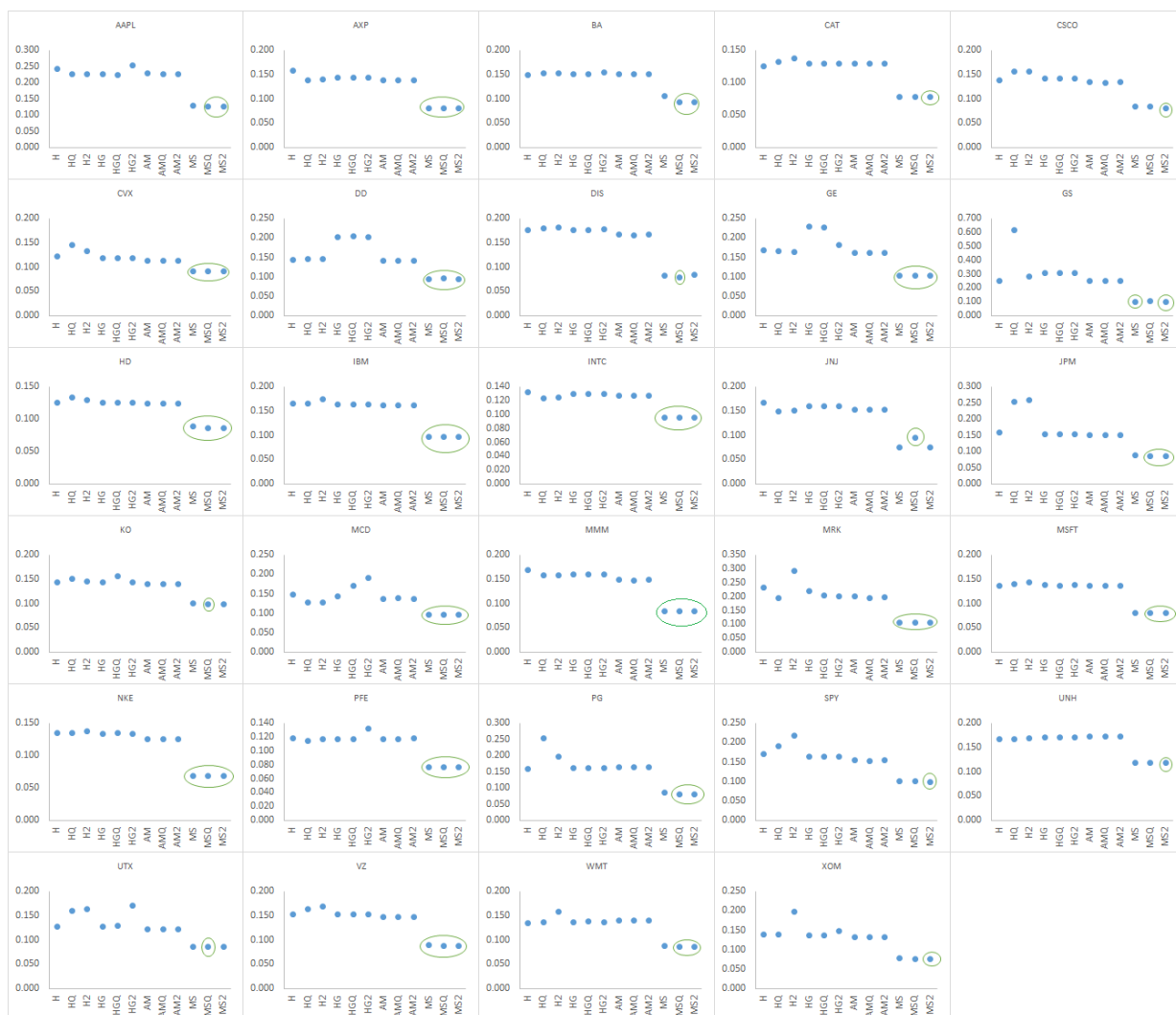
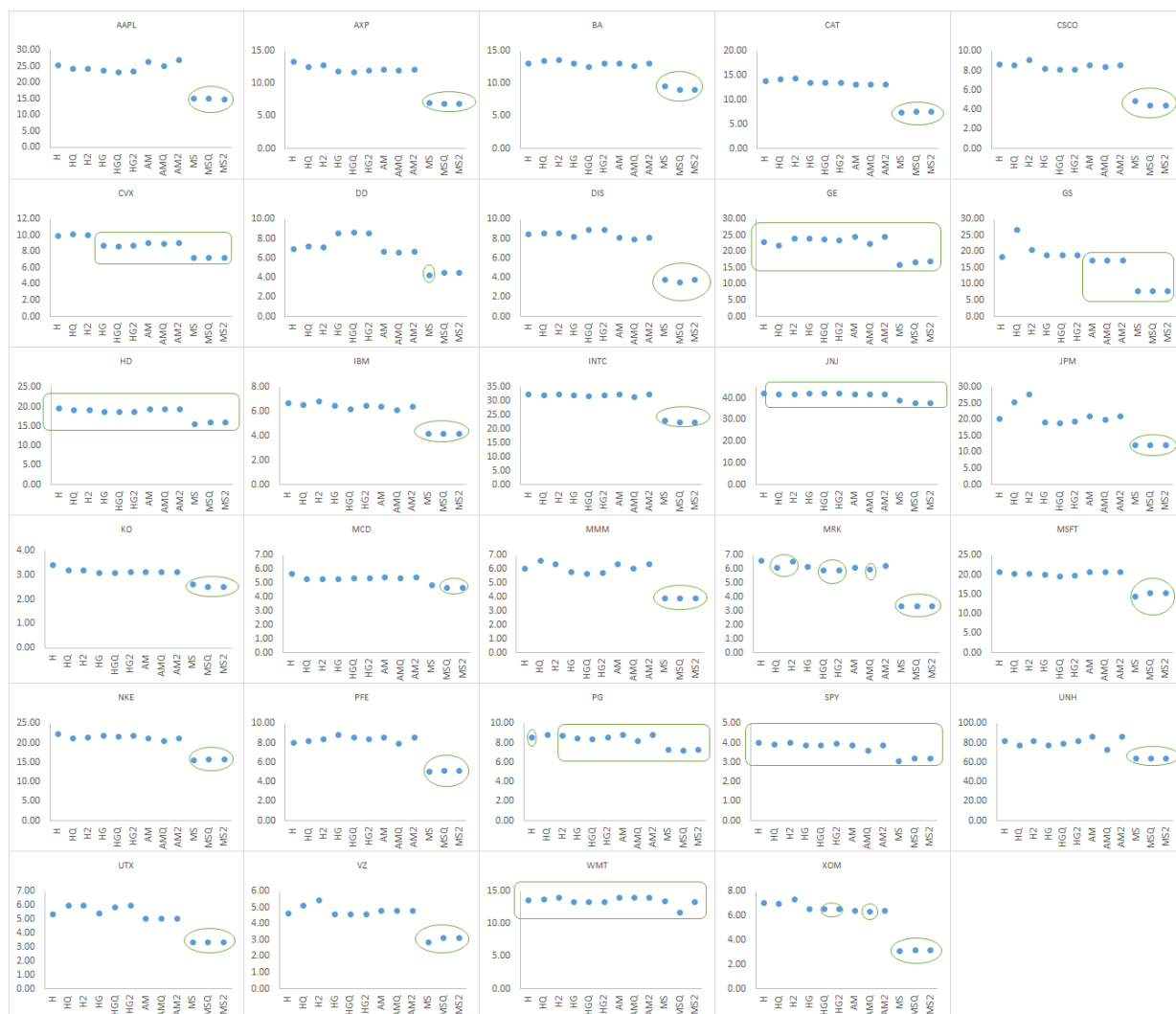


Figure 5: Out-of-sample MSE loss function for 12 models and 29 assets. The enclosures contain, for each asset, the models belonging to the MCS at 5% significance level. Models are represented by short acronyms in this order: H for HAR, HQ for HARQ, H2 for HAR2, HG for HAR-GARCH, HGQ for HARQ-GARCH, HG2 for HAR2-GARCH, AM for AMEM, AMQ for AMEMQ, AM2 for AMEM2, MS for MS-AMEM, MSQ for MS-AMEMQ, MS2 for MS-AMEM2.



formance of simple (perhaps linear) models with respect to more sophisticated models is an empirical regularity, with also a theoretical justification (Hansen, 2010).

## 7 Concluding Remarks

Dynamic modeling of realized variance behavior is important to forecast conditional return variance, taking full advantage of the better properties of the ultra-high frequency-based estimates of variance. Since these are estimates they are affected by measurement errors, as long recognized in the realized variance literature recalled earlier. The variance estimates are less accurate when quarticity is high. In this paper, we stressed that the (square root of) quarticity has a high correlation with realized variance itself: this means that realized variance estimates are less accurate the higher the variance itself, or that measurement errors are heteroskedastic.

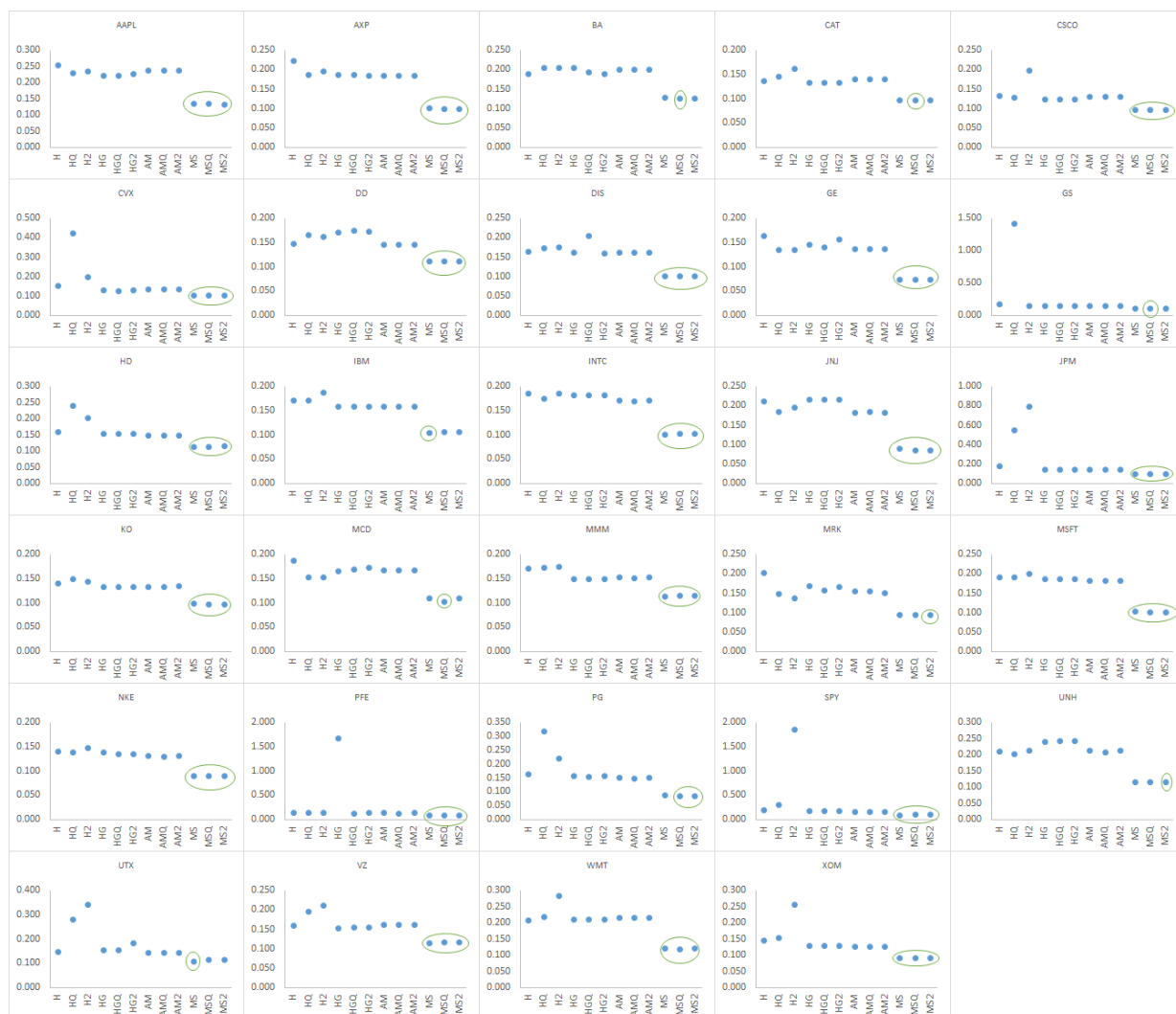
The important point made by Bollerslev *et al.* (2016) is that when predicting variance based on an HAR model, a term can be inserted to capture such an estimation bias in the coefficient of the lagged realized variance. This interaction effect (i.e. a stronger mean reversion when quarticity is high) is shown in a variety of instances to be statistically significant: the practical implication is that the augmented model pays smaller attention, so to speak, to lagged values of realized variance when quarticity is high.

Starting from the consideration that quarticity and realized variance (under a number of transformations or correlation measures) are strongly correlated, the model may be augmented by allowing the dynamics of volatility to depend on a curvature effect, i.e. on the square of the variance. Interaction and curvature effects prove to be observationally equivalent, in that they deliver very similar results within the HAR model, both in terms of numerical values and in terms of statistical significance across a representative sample of US market assets (the S&P500 plus 28 components of the Dow Jones index). As a consequence, we argue that in the measurement relationship between realized variance and the integrated variance, the measurement error may be represented as proportional to the latter, and, as such, heteroskedasticity is a factor to be made explicit.

We have explored several directions, all pointing to the importance of adopting a heteroskedastic model which gives different importance to observations: the HAR-GARCH model of Corsi *et al.* (2008) where HAR is enriched by a GARCH-type structure to account for the volatility of volatility (related to quarticity), the AMEM model of Engle and Gallo (2006), where a multiplicative error structure makes the variance of realized variance proportional to the square of its expectation. In a further extension of the AMEM, the MS-AMEM of Gallo and Otranto (2015), dynamics is regulated by regime-specific coefficients (both the speed of mean reversion, and the level of the mean to which the model reverts change). We find that making the realized variance sensitive to the sign of past returns adds an important driver of the dynamics.

In all these models, the statistical significance of the interaction/curvature terms is greatly reduced, and more strongly so for the MS-AMEM, providing evidence that accounting for heteroskedasticity provides robustness to measurement errors. This result is confirmed by a simulation study where generating observations considering multiplicative

Figure 6: Out-of-sample QL loss function for 12 models and 29 assets. The enclosures contain, for each asset, the models belonging to the MCS at 5% significance level. Models are represented by short acronyms in this order: H for HAR, HQ for HARQ, H2 for HAR2, HG for HAR-GARCH, HGQ for HARQ-GARCH, HG2 for HAR2-GARCH, AM for AMEM, AMQ for AMEMQ, AM2 for AMEM2, MS for MS-AMEM, MSQ for MS-AMEMQ, MS2 for MS-AMEM2.



errors, and/or with the AMEM and the MS-AMEM (without interaction/curvature terms) and when HAR is estimated with either of those terms, it turns up being statistically significant.

The analysis we performed both in- and out-of-sample shows that the MS-AMEM is by far the best performing model, with a good second-best performance of the HAR-GARCH model (even over AMEM). The augmented versions of the models show no sign of improved performance in these cases, while such better performance is confirmed for HARQ and HAR2 over HAR.

The documented presence of measurement errors in estimating realized variance is unquestionable, but the fact that such errors are proportional to the variance itself can be turned to the advantage of an improved specification for its conditional expectation. Since, in general, we should pay less attention to high values of variance when modeling its dynamics for prediction purposes, heteroskedastic models, such as HAR-GARCH and AMEM are equipped to do so, the latter from the outset. However, the issue of the presence of a low frequency volatility component in the data is unrelated to measurement errors and needs to be adequately and separately addressed. In some cases, moreover, quarticity may not be available, as it happens when daily realized variances are made available without the underlying intra daily data, or when range measures are an alternative choice instead.

All analysis here is conducted for the realized variance. In the HAR model this is a natural choice because of the additivity of the variance when aggregating at the weekly and monthly levels. An alternative, not analyzed here, is to model volatility directly, rather than variance as in Cipollini *et al.* (2013) and Gallo and Otranto (2015), in view of the fact that the volatility has less severe peaks in correspondence to volatility bursts. As a matter of fact, Cipollini *et al.* (2013) show that there is a definite improvement in the forecasting performance of the (multivariate) modeling of realized volatility over realized variance. Finally, the HARQ model has been recently extended to the multivariate case by Bollerslev *et al.* (2018). The challenge would be to develop a similar exercise, dealing with multivariate MEM (Cipollini *et al.*, 2013; Taylor and Xu, 2017) with and without the presence of Markov Switching dynamics in the time-varying covariance matrix (see, for example, Otranto, 2010).

## References

- Alizadeh, S., Brandt, M. W. and Diebold, F. X. (2002) Range-based estimation of stochastic volatility models, *The Journal of Finance*, **57**, 1047–1091.
- Andersen, T. G. and Benzoni, L. (2009) Realized volatility, in *Handbook of Financial Time Series* (Eds.) T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch, Springer Verlag.
- Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006) Volatility and correlation forecasting, in *Handbook of Economic Forecasting* (Eds.) G. Elliott, C. W. J. Granger and A. Timmermann, North Holland.
- Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007) Roughing it up: Including jump

- components in the measurement, modeling and forecasting of return volatility, *Review of Economics and Statistics*, **89**, 701–720.
- Barigozzi, M., Brownlees, C., Gallo, G. and Veredas, D. (2014) Disentangling systematic and idiosyncratic dynamics in panels of volatility measures, *Journal of Econometrics*, **182**, 364–384.
- Barndorff-Nielsen, O. and 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., Patton, A. J. and Quaadvlieg, R. (2016) Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics*, **192**, 1–18.
- Bollerslev, T., Patton, A. J. and Quaadvlieg, R. (2018) Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions, *Journal of Econometrics*, **207**, 71–91.
- Brownlees, C. T. and Gallo, G. M. (2006) Financial econometric analysis at ultra-high frequency: Data handling concerns, *Computational Statistics and Data Analysis*, **51**, 2232–2245.
- Brownlees, C. T. and Gallo, G. M. (2010) Comparison of volatility measures: a risk management perspective, *Journal of Financial Econometrics*, **8**, 29–56.
- Brownlees, C. T. and Gallo, G. M. (2011) Shrinkage estimation of semiparametric multiplicative error models, *International Journal of Forecasting*, **27**, 365–378.
- Buccheri, G. and Corsi, F. (2019) Hark the shark: Realized volatility modelling with measurement errors and nonlinear dependencies, Tech. Rep. SSRN 3089929, University of Venice Ca’ Foscari.
- Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006) *Measurement Error in Nonlinear Models*, Chapman and Hall/CRC, 2 edn.
- Cipollini, F., Engle, R. F. and Gallo, G. M. (2013) Semiparametric vector MEM, *Journal of Applied Econometrics*, **28**, 1067–1086.
- Clements, A., Doolan, M., Hurn, S. and Becker, R. (2009) On the efficacy of techniques for evaluating multivariate volatility forecasts, in *NCER Working Paper Series*, National Centre for Econometric Research, vol. 41.
- Corsi, F. (2009) A simple approximate long-memory model of realized volatility, *Journal of Financial Econometrics*, **7**, 174–196.
- Corsi, F., Mittnik, S., Pigorsch, C. and Pigorsch, U. (2008) The volatility of realized volatility, *Econometric Reviews*, **27**, 46–78.

- Engle, R. F. (2002) New frontiers for ARCH models, *Journal of Applied Econometrics*, **17**, 425–446.
- Engle, R. F. and Gallo, G. M. (2006) A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics*, **131**, 3–27.
- Engle, R. F. and Rangel, J. G. (2008) The spline-GARCH model for low frequency volatility and its global macroeconomic causes, *Review of Financial Studies*, **21**, 1187–1222.
- Gallo, G. M. and Otranto, E. (2015) Forecasting realized volatility with changing average levels, *International Journal of Forecasting*, **31**, 620–634.
- Gallo, G. M. and Otranto, E. (2017) Combining sharp and smooth transitions in volatility dynamics: a fuzzy regime approach, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, forthcoming.
- Garman, M. B. and Klass, M. J. (1980) On the estimation of security price volatilities from historical data, *The Journal of Business*, **53**, 67–78.
- Hansen, P. R. (2010) A winner’s curse for econometric models: on the joint distribution of in-sample fit and out-of-sample fit and its implications for model selection, Tech. rep., Stanford University.
- Hansen, P. R., Lunde, A. and Nason, J. M. (2003) Choosing the best volatility models: The model confidence set approach\*, *Oxford Bulletin of Economics and Statistics*, **65**, 839–861.
- Hansen, P. R., Lunde, A. and Nason, J. M. (2011) The model confidence set, *Econometrica*, **79**, 453–497.
- Heber, G., Lunde, A., Shephard, N. and Sheppard, K. (2009) OMI’ s realised library, version 0.1, Tech. rep., Oxford-Man Institute, University of Oxford.
- Maheu, J. M. and McCurdy, T. H. (2002) Nonlinear features of realized FX volatility, *Review of Economics and Statistics*, **84**, 668–681.
- McAleer, M. and Medeiros, M. (2008) A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries., *Journal of Econometrics*, **147**, 104–119.
- Newey, W. and West, K. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703–08.
- Otranto, E. (2010) Asset allocation using flexible dynamic correlation models with regime switching, *Quantitative Finance*, **10**, 325–338.
- Patton, A. J. (2011) Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics*, **160**, 246 – 256, realized Volatility.



- Taylor, N. and Xu, Y. (2017) The logarithmic vector multiplicative error model: an application to high frequency NYSE stock data, *Quantitative Finance*, **17**, 1021–1035.
- Teräsvirta, T. (2009) An introduction to univariate GARCH models, in *Handbook of Financial Time Series* (Eds.) T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch, Springer Verlag, chap. 1, pp. 17–42.
- Whaley, R. E. (2009) Understanding the VIX, *The Journal of Portfolio Management*, **35**, 98–105.
- White, H. (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, **48**, 817–38.

## Appendix – List of tickers

Table 8: List of tickers of the realized variance series

AAPL	Apple
AXP	American Express
BA	Boeing
CAT	Caterpillar
CSCO	Cisco
CVX	Chevron
DD	DuPont
DIS	Disney
GE	General Electric
GS	Goldman Sachs
HD	Home Depot
IBM	IBM
INTC	Intel
JNJ	Johnson and Johnson
JPM	J.P. Morgan Chase
KO	Coca Cola
MCD	McDonald's
MMM	3M
MRK	Merck
MSFT	Microsoft
NKE	Nike
PFE	Pfizer
PG	Procter and Gamble
SPY	SPY ETF
UNH	United Health
UTX	United Technologies
VZ	Verizon
WMT	Walmart
XOM	Exxon Mobil

