Statistica Sinica **25** (2015), 759-786
doi:http://dx.doi.org/10.5705/ss.2012.283

HYBRID-GARCH: A GENERIC CLASS OF MODELS FOR VOLATILITY PREDICTIONS USING HIGH FREQUENCY DATA

Xilong Chen, Eric Ghysels and Fangfang Wang

SAS Institute Inc., University of North Carolina at Chapel Hill and University of Illinois at Chicago

Abstract: We propose a general GARCH framework that allows one to predict volatility using returns sampled at a higher frequency than the prediction horizon. We call the class of models **H**igh Frequenc**Y** Data-**B**ased P**R**oject**I**on-**D**riven GARCH, or HYBRID-GARCH models, as volatility dynamics are driven by what we call HYBRID processes. The HYBRID processes can involve data sampled at any frequency. We study the theoretical properties as well as statistical inference. An application reports the superior out-of-sample forecasting performance of the new class of models, including the time of the recent financial crisis.

Key words and phrases: Filtering, GARCH jump diffusion, HYBRID process, realized measure, temporal aggregation, weak GARCH.

1. Introduction

Multi-period volatility forecasts feature prominently in asset pricing, portfolio allocation, risk-management, and most other areas of finance where long-horizon measures of risk are necessary. Such forecasts can be constructed in three quite different ways. The first approach is to estimate a horizon-specific model of the volatility, such as a weekly or monthly GARCH, that can then be used to form direct predictions of volatility over the next week, month, etc. The second approach is to estimate a daily model and then iterate forward the daily forecasts to obtain weekly or monthly predictions. The forecasting literature refers to the first approach as "direct" and the second as "iterated". A third method is the mixed-data sampling (MIDAS) approach introduced by Ghysels, Santa-Clara, and Valkanov (2005, 2006). A MIDAS model uses, for example, daily squared returns to produce directly multi-period volatility forecasts and can be viewed as a middle ground between the direct and the iterated approaches. The MIDAS volatility literature (see Ghysels and Valkanov (2012)) has mostly focused on regressions-based models. It is the purpose of this paper to introduce ideas similar to MIDAS models in GARCH-type models. The advantages of this approach is that one focuses directly on multi-period forecasts, as in the direct approach, while one preserves the use of high-frequency data.

We propose a unifying framework, based on a generic GARCH-type model, that addresses the issue of volatility forecasting involving forecast horizons of a different frequency than the information set. Hence, we propose a class of GARCH models that can handle volatility forecasts over the next five business days and use past (intra-)daily data, or tomorrow's expected volatility while using intra-daily returns. We call the class of models **H**igh Frequenc**Y** Data-**B**ased P**R**oject**I**on-**D**riven GARCH models as the GARCH dynamics are driven by what we call HYBRID processes. HYBRID-GARCH models - by their very nature - relate to many topics discussed in the extensive literature on volatility forecasting. These topics include - but are not limited to - iterated versus direct forecasting, temporal aggregation, weak versus semi-strong GARCH, as well as various estimation procedures. Since there are quite a few papers written on these topics it is obviously hard to cite a comprehensive list here. Nevertheless, it is worth noting that we study three broad classes of HYBRID processes: (1) parameter-free processes that are purely data-driven, (2) structural HYBRIDs in which one assumes an underlying data generating process (DGP) or some dynamic structure for the high frequency data, and (3) HYBRID filter processes.

To motivate the class of models, it is worth recalling that a key ingredient of conditional volatility models is that more weight is attached to the most recent returns (i.e. information). In the case of the original ARCH model (see e.g. Engle (1982)) that means the most recent (daily) squared returns have more weight when predicting future (daily) conditional volatility. How does this apply to intra-daily financial data? The foundation of so-called realized volatility (RV) modeling is the theory of continuous time semi-martingale stochastic processes, more specifically stochastic volatility continuous time jump-diffusions. While intra-daily data are used to construct RV, prediction models put more weight on recent (daily) RV but, despite the use of intra-daily data, do not differentiate among intra-daily returns. If volatility is a persistent process, it would be natural to weight intra-daily data differently, as pointed out by Malliavin and Mancino (2005). The arguments also apply to lower frequency volatility prediction models, such as (total) weekly volatility. Here the choice is between a GARCH model - using past weekly returns, de facto putting equal weight to the daily returns within the week - and a GARCH model for weekly forecasts using daily returns. It is the latter that is novel, and an example of the class of models is introduced in the paper. Compared to Malliavin and Mancino (2005), we go beyond linear projections, albeit in a discrete time setting. Our models do have a connection with continuous time models as well when we restrict our attention to linear projections.

The paper is structured as follows. Section 2 provides an overview of the models, and the various classes of HYBRID processes involved. Section 3 is devoted to the statistical properties of the HYBRID GARCH model, defining the various parameter estimators and studying their asymptotic properties. In Section 4 we study continuous time limiting arguments and examine how the HYBRID process can be structurally linked to continuous time diffusions. We also examine the role played by jumps. The small sample behavior of various estimators is examined in Section 5 via simulation. An empirical application in Section 6 reports the superior out-of-sample forecasting performance of the new class of models, including during the time of the recent financial crisis. Section 7 concludes the paper. Technical details are collected in Chen, Ghysels, and Wang (2014).

2. HYBRID Processes

The paper considers models for the volatility prediction of a log-price process $p \doteq (p_s)_{-\infty<s<\infty}$, which is a continuous-time semi-martingale defined on some probability space $(\Omega, \mathcal{F}, P)$. Suppose the process is observed at a frequency $m \in \mathbb{N}^*$, with returns $r_{k/m} = p_{k/m} - p_{(k-1)/m}$ observed for $k \in \mathbb{Z}$, where the unit interval corresponds to one day with m observations per day. The realized variance on day t, RV_t, is defined as $RV_t \doteq \sum_{j=0}^{m-1} r_{t-j/m}^2$. The volatility object of interest is

$$\sigma_{t+1|t}^2 \doteq P_l(RV_{t+1}|\mathcal{I}_t), \quad t \in \mathbb{Z}, \tag{2.1}$$

the orthogonal projection of the realized variance RV_{t+1} onto the information up to day t, $\mathcal{I}_t$.

To mimic the volatility dynamics, we consider a GARCH-type recursion

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma H_t(\phi), \quad t \in \mathbb{Z}, \tag{2.2}$$

where $\alpha > 0, 0 < \beta < 1, \gamma > 0$, and $H_t(\phi)$ is some functional of the observed returns $\vec{r}_t = (r_{t-1+1/m}, r_{t-1+2/m}, \ldots, r_{t-1/m}, r_t)^\top$ that may or may not depend on the parameter $\phi \in \mathbb{R}^d$ $(d \geq 1)$. The process $V_{t+1|t}$ at (2.2) is referred to a **H**igh Frequenc**Y** Data-**B**ased P**R**oject**I**on-**D**riven (HYBRID) GARCH process, and H_t is called a *HYBRID* process.

Remark 1. The HYBRID process H_t can be purely data-driven and not depend on parameters. The obvious case would be a simple squared return process such that $V_{t+1|t}$ has the typical GARCH(1,1) dynamic.

In this paper, we intend to study volatility forecast under minimal assumptions on the log-price process. The structure of $\sigma_{t+1|t}^2$ is not tractable if the price or return process is unspecified. Therefore, to facilitate our analysis, we have an assumption on the discretely sampled return process $\{r_{k/m}, k \in \mathbb{Z}\}$.

Assumption 1. $\{r_{k/m}, k \in \mathbb{Z}\}$ *are non-degenerate and linearly independent with* $Er^4_{k/m} < \infty$.

Let $\mathcal{L}_t$ be the closed span of $\{1, r_{t-k/m},\ r^2_{t-k/m};\ k = 0, 1, 2\ldots\}$, and $\mathcal{F}^d_t \doteq \sigma(r_{t-k/m}, k = 0, 1, 2\ldots)$ the sigma field generated by the discretely observed return process. We consider two specifications for the information set $\mathcal{I}_t$.

Assumption 2 (Scenario 1)**.** $\mathcal{I}_t = \mathcal{L}_t$ *for* $t \in \mathbb{Z}$*, and* $P_l(r_s|\mathcal{L}_{s-1/m}) = 0$*, where* $s = k/m$ *and* $k \in \mathbb{Z}$.

Assumption 3 (Scenario 2)**.** $\mathcal{I}_t = \mathcal{F}^d_t$ *for* $t \in \mathbb{Z}$*, and* $P_l(r_s|\mathcal{F}^d_{s-1/m}) = 0$ *with* $s = k/m$ *and* $k \in \mathbb{Z}$.

In Scenario 1, $\sigma^2_{t+1|t}$ is viewed as the best linear predictor. Scenario 2 defines a more general situation, where $\sigma^2_{t+1|t}$ is a conditional variance. In particular, under Assumption 3, the prediction equations imply that $E(r_s|\mathcal{F}^d_{s-1/m}) = 0$ and $E(RV_{t+1}|\mathcal{F}^d_t) = E(R^2_{t+1}|\mathcal{F}^d_t) = \sigma^2_{t+1|t}$, where $R_{t+1} = \sum_{j=0}^{m-1} r_{t+1-j/m}$.

The distance between $V_{t+1|t}$ and $\sigma^2_{t+1|t}$ is determined by the structure of the HYBRID process as well as unknown parameters. Write $V_{t+1|t}$ as $V_{t+1|t}(\theta)$ where θ collects all the distinct parameters in the HYBRID GARCH model (2.2). We discuss the model $\sigma^2_{t+1|t} = V_{t+1|t}(\theta_0)$, where θ_0 can be understood as the 'true' parameter. This is done for simplicity and one could think of generalizations where $V_{t+1|t}(\theta_0)$ is an approximation of $\sigma^2_{t+1|t}$. Moreover, we have an assumption that guarantees that the HYBRID process is non-negative and satisfies measurability and identifiability when it comes to parameter estimation.

Assumption 4. *Let* $t \in \mathbb{Z}$*,* $H_t \in \mathcal{I}_t$*. For a parameterized* H_t *and* $\phi = (\phi_1, \ldots, \phi_d)^\top$*, define* $H(\phi, \vec{r}_t) \doteq H_t(\phi)$*. There exists a connected set* $\Phi \subset \mathbb{R}^d$ *such that, for* $j \in \{1, \ldots, d\}$, (1) *the second-order partial derivative* $\partial^2 H_t(\phi)/\partial\phi_j^2$ *exists and is continuous on* Φ*;* (2) $H(\phi, \cdot)$*,* $\partial H(\phi, \cdot)/\partial\phi_j$*, and* $\partial^2 H(\phi, \cdot)/\partial\phi_j^2$ *are* $\mathcal{B}(R^m)/\mathcal{B}(R)$ *measurable;* (3) 1*,* $H_t(\phi)$*, and* $\partial H_t(\phi)/\partial\phi_1, \ldots, \partial H_t(\phi)/\partial\phi_d$ *are linearly independent.*

A detailed discussion of Assumption 4 is provided in the companion document Chen, Ghysels, and Wang (2014). The HYBRID process that satisfies Assumption 4 is also referred to as a *HYBRID filtering process.*

3. Estimation of the HYBRID GARCH Model

The objective is to find an optimal θ which minimizes the distance between $V_{t+1|t}(\theta)$ and $\sigma^2_{t+1|t}$. We work exclusively with returns sampled at fixed frequency without referring to an explicit data generating process and make the assumption that

Assumption 5. $\{r_{k/m}, k \in \mathbb{Z}\}$ *is strictly stationary and ergodic.*

We take the returns as sampled from the underlying continuous-time log price process. A discussion on the continuous time limits is deferred to Section 4.

Throughout this section, we assume that the HYBRID process is parameterized, and the parameter ϕ is not related to α, β, γ appearing in (2.2). Hence $\theta = (\alpha, \beta, \gamma, \phi)^\top \in \mathbb{R}^{3+d}$. The discussion that follows can easily carry over to the situation where ϕ is functional of α or β or γ. Let $\Theta = \{\theta = (\alpha, \beta, \gamma, \phi)^\top : \alpha > 0, 0 < \beta < 1, \gamma > 0, \phi \in \Phi\}$ where Φ is a connected set in Assumption 4. We suppose that $\mathcal{C}$ is a convex compact subset of Θ such that the true parameter $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \phi_0)^\top$ is an interior point of $\mathcal{C}$.

3.1. Estimation under Scenario 1

Given the observations $\{r_{1/m}, \ldots, r_1, \ldots, r_{T-1+1/m}, \ldots, r_T\}$, a natural estimator of θ_0 in Scenario 1 is the minimizer of $T^{-1}\sum_{t=1}^T \left(RV_t - \tilde{V}_t(\theta)\right)^2$ where $\tilde{V}_t$ is defined recursively by

$$\tilde{V}_t(\theta) = \alpha + \beta\tilde{V}_{t-1}(\theta) + \gamma H_{t-1}(\phi), t \geq 1 \quad \text{and} \quad \tilde{V}_0 = \tilde{v}, \tag{3.1}$$

where $\tilde{v}$ is any arbitrary deterministic value. The minimizer exists due to Jennrich (1969) and Gallant and White (1988), and is denoted by $\tilde{\theta}_T^{mdrv}$, *minimum-distance RV-based estimator.* Let $\varepsilon_t(\theta) = RV_t - V_{t|t-1}(\theta)$.

Proposition 1 (Scenario 1)**.** *If Assumptions* 1, 2, 5, *hold,* $\theta_0 = \arg\min_{\theta\in\mathcal{C}} E\varepsilon_t(\theta)^2$.

Proof. see Section S3.1 of Chen, Ghysels, and Wang (2014).

The true parameter θ_0 is identifiably unique – see Gallant and White (1988) for the definition of identifiable uniqueness. Proposition 1 allows us to establish the consistency of $\tilde{\theta}_T^{mdrv}$. The asymptotic normality needs another assumption.

Assumption 6. $\{r_{k/m}, k \in \mathbb{Z}\}$ *is strictly stationary and α-mixing with the mixing coefficient $\alpha(j)$ satisfying $\sum_{j=0}^\infty \alpha(j)^{v_2/(2+v_2)} < \infty$ for some $v_2 > 0$, and $Er_{k/m}^{4(2+v_2)} < \infty$.*

Denote by ∇ the vector differential operator w.r.t θ so that ∇f is the gradient (column vector) of scaler function f: $\nabla f = (\partial_1 f, \ldots, \partial_{d+3} f)^\top$ where ∂_k is the partial derivative w.r.t. the k^{th} parameter in $\theta = (\alpha, \beta, \gamma, \phi)$.

Theorem 1 (Scenario 1)**.** *Under Assumptions* 1, 2, 4, *and* 5*.*

(1) $\tilde{\theta}_T^{mdrv}$ *is identifiably unique and is a strongly consistent estimator of* θ_0*.*

(2) *Under the further Assumption* 6, $\lim_{T\to\infty}$ *var* $\left((1/\sqrt{T})\sum_{t=1}^T \varepsilon_t \nabla\varepsilon_t(\theta_0)\right)$ *exists and is finite, denoted by* Ω^{mdrv}*.*

(3) *If* Ω^{mdrv} *is positive definite,* $\sqrt{T}(\tilde{\theta}_T^{mdrv} - \theta_0) \Rightarrow N(0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1})$, *where* $0 < \Sigma^{md} = E\nabla V_{t|t-1}(\theta_0)(\nabla V_{t|t-1}(\theta_0))' < \infty$.

Proof. see Section S3.2 of Chen, Ghysels, and Wang (2014).

The existence of Ω^{mdrv} and the asymptotic normality follows from the fact that $\varepsilon_t \partial_k \varepsilon_t$ is *near-epoch dependent* on the return process, and is therefore a *mixingale* when the return process is α-mixing (see Lemma S3.5 of Chen, Ghysels, and Wang (2014)). It should be noted that the size of α-mixing is $-(v_2+2)/v_2$ (see Assumption 6). It is weaker than the size required in Theorem 5.7 of Gallant and White (1988), and of Goncalves and White (2004) as well (see page 24 of Gallant and White (1988) for the definition of size).

3.2. Estimation under Scenario 2

We consider situations where the HYBRID GARCH model produces conditional variance prediction. In such a case we are at liberty to consider both minimum distance estimator as in Section 3.1, and the quasi-maximum likelihood estimator that is standard in the GARCH literature. Consider

$$\tilde{\theta}_T^{mdrv} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left(RV_t - \tilde{V}_t(\theta) \right)^2, \tag{3.2}$$

$$\tilde{\theta}_T^{mdr2} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left(R_t^2 - \tilde{V}_t(\theta) \right)^2, \tag{3.3}$$

$$\tilde{\theta}_T^{lhr2} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left(\log \tilde{V}_t(\theta) + \frac{R_t^2}{\tilde{V}_t(\theta)} \right), \tag{3.4}$$

$$\tilde{\theta}_T^{lhrv} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left(\log \tilde{V}_t(\theta) + \frac{RV_t}{\tilde{V}_t(\theta)} \right), \tag{3.5}$$

where $\tilde{V}_t$ is defined in (3.1). We also consider the estimator derived using the Multiplicative Error Model, which shares some similarities with the likelihood-RV-based estimator $\tilde{\theta}_T^{lhrv}$.

To make sure $\tilde{V}_t(\theta) \in L^2$ and its second order partial derivatives are well defined, we need an assumptions on H_t.

Assumption 7. *For* $\vec{r}_t = (r_{t-1+1/m}, r_{t-1+2/m}, \ldots, r_{t-1/m}, r_t)^T$ *satisfying Assumption* 1,

(1) $E\sup_{\phi \in \overline{\Phi^0}} H(\phi, \vec{r}_t)$, $E\sup_{\phi \in \overline{\Phi^0}} |\frac{\partial H(\phi, \vec{r}_t)}{\partial \phi_i}|$, *and* $E\sup_{\phi \in \overline{\Phi^0}} |\frac{\partial^2 H(\phi, \vec{r}_t)}{\partial \phi_i \partial \phi_j}|$ *are finite, for* $i, j \in \{1, \ldots, d\}$.

(2) $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^2$, $E(\sup_{\phi\in\overline{\Phi^0}} |\frac{\partial H(\phi,\vec{r}_t)}{\partial\phi_i}|)^2$, *and* $E(\sup_{\phi\in\overline{\Phi^0}} |\frac{\partial^2 H(\phi,\vec{r}_t)}{\partial\phi_i\partial\phi_j}|)^2$ *are finite, for* $i, j \in \{1,\ldots,d\}$.

(3) $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^4$ *and* $E(\sup_{\phi\in\overline{\Phi^0}} |\frac{\partial H(\phi,\vec{r}_t)}{\partial\phi_i}|)^4$ *are finite, for* $i \in \{1,\ldots,d\}$.

(4) $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^{2(2+v_2)}$ *and* $E(\sup_{\phi\in\overline{\Phi^0}} |\frac{\partial H(\phi,\vec{r}_t)}{\partial\phi_i}|)^{2(2+v_2)}$ *are finite, for* $i \in \{1,\ldots,d\}$, *where* v_2 *is defined in Assumption* 6.

Remark 2. We do not need Assumption 7 in Scenario 1, because the HYBRID process H_t in Scenario 1 is a weighted sum of 1, the intermediate returns and squared returns from day $t-1$ to t (see Assumption 4).

3.2.1. Minimum distance and quasi-likelihood estimators

We start by extending Proposition 1 to the case of Scenario 2. Let $e_t(\theta) = R_t^2 - V_{t|t-1}(\theta)$.

Proposition 2 (Scenario 2). *Under Assumptions* 1, 3, 4, 5, *and* 7(2), $\theta_0 = \arg\min_{\theta\in\mathcal{C}} E\left(\varepsilon_t(\theta)\right)^2 = \arg\min_{\theta\in\mathcal{C}} E\left(e_t(\theta)\right)^2$.

Here $\varepsilon_t(\theta_0)\partial_i\varepsilon_t(\theta_0)$ and $e_t(\theta_0)\partial_i e_t(\theta_0)$ are martingale difference sequences. Ω^{mdrv} defined in Theorem 1 becomes $E[(RV_t - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)']$. Let $\Omega^{mdr2} = E[(R_t^2 - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)']$. Both Ω^{mdrv} and Ω^{mdr2} are finite and positive definite under suitable regularity conditions. We therefore have the following regarding $\tilde{\theta}_T^{mdrv}$ and $\tilde{\theta}_T^{mdr2}$ defined in (3.2) and (3.3).

Theorem 2 (Scenario 2). *Under Assumptions* 1, 3, 4, 5, *and* 7(2),

(1) $\tilde{\theta}_T^{mdrv}$, $\tilde{\theta}_T^{mdr2}$ *are identifiably unique and they converge to* θ_0 *a.s.*

(2) *If further* $Er^8 < \infty$ *and Assumption* 7(3) *holds,* $\sqrt{T}(\tilde{\theta}_T^{mdrv} - \theta_0)$ *converges in distribution to* $N(0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1})$ *and* $\sqrt{T}(\tilde{\theta}_T^{mdr2} - \theta_0)$ *converges in distribution to* $N(0, (\Sigma^{md})^{-1}\Omega^{mdr2}(\Sigma^{md})^{-1})$.

The proof is an application of the Martingale Central Limit Theorem and appears in Section S3.3 of the companion document Chen, Ghysels, and Wang (2014).

Proposition 3 (Scenario 2). *Suppose* $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^2 < \infty$. *Under Assumptions* 1, 3, 4, *and* 5, $\theta_0 = \arg\min_{\theta\in\mathcal{C}} E\left(\log V_{t|t-1}(\theta) + RV_t/V_{t|t-1}(\theta)\right) = \arg\min_{\theta\in\mathcal{C}} E\left(\log V_{t|t-1}(\theta) + R_t^2/V_{t|t-1}(\theta)\right)$.

Proof. see Section S3.1 of Chen, Ghysels, and Wang (2014).

Therefore θ_0 can be estimated by $\tilde{\theta}_T^{lhrv}$ in (3.4) or $\tilde{\theta}_T^{lhr2}$ in (3.5). Let

$$\Sigma^{lh} = E\left(V_{t|t-1}^{-2}(\theta_0)\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'\right),$$

$$\Omega^{lhr2} = E\left(V_{t|t-1}^{-4}(\theta_0)(R_t^2 - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'\right),$$

and

$$\Omega^{lhrv} = E\left(V_{t|t-1}^{-4}(\theta_0)(RV_t - V_{t|t-1}(\theta_0))^2 \nabla V_{t|t-1}(\theta_0) \nabla V_{t|t-1}(\theta_0)'\right).$$

Theorem 3 (Scenario 2). *Suppose* $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^2 < \infty$ *and Assumptions* 1, 3, 4, *and* 5 *hold.*

(1) $\tilde{\theta}_T^{lhrv}$, $\tilde{\theta}_T^{lhr2}$ *are identifiably unique and they converge to* θ_0 *a.s.*

(2) *If* $E(r^{4+v}) < \infty$ *for some* $v > 0$, *and for* $i, j \in \{1, \ldots, d\}$,

$$\left|\frac{\partial H(\phi,\vec{x})/\partial\phi_i}{H(\phi,\vec{x})}\right| \le g(\phi), \quad \left|\frac{\partial^2 H(\phi,\vec{x})/\partial\phi_i\partial\phi_j}{H(\phi,\vec{x})}\right| \le g(\phi) \quad \forall\vec{x}\in\mathbb{R}^m, \phi\in\Phi, \tag{3.6}$$

where g *is real-valued and continuous in* ϕ, *then* $\sqrt{T}(\tilde{\theta}_T^{lhrv} - \theta_0)$ *converges in distribution to* $N\left(0, (\Sigma^{lh})^{-1}\Omega^{lhrv}(\Sigma^{lh})^{-1}\right)$, *and* $\sqrt{T}(\tilde{\theta}_T^{lhr2} - \theta_0)$ *converges in distribution to* $N\left(0, (\Sigma^{lh})^{-1}\Omega^{lhr2}(\Sigma^{lh})^{-1}\right)$.

The proof is an application of the Martingale Central Limit Theorem and appears in Section S3.4 of Chen, Ghysels, and Wang (2014).

Remark 3. The likelihood estimation considered here is slightly different from what is discussed in the literature. First of all, $\sigma^2_{t|t-1}$ is studied in $L^2(\Omega,\mathcal{F},P)$ instead of $L^1(\Omega,\mathcal{F},P)$. Secondly, the objective function appearing in (3.4) is not the joint quasi-log-likelihood function (modulo a constant) of $\{R_1,\ldots,R_T\}$. Instead of conditioning on $R_1,\ldots,R_{t-1}$, $\log\tilde{V}_t(\theta) + R_t^2/\tilde{V}_t(\theta)$ is conditional quasi-log-likelihood w.r.t. a finer set, the sigma field generated by the high frequency returns up to day $t-1$.

Remark 4. The discussion can be extended to strictly periodically stationary and periodically ergodic time series as well; the proofs only require $\vec{r}_t$ to be strictly stationary ergodic.

Among the four estimators $\tilde{\theta}_T^{mdr2}$, $\tilde{\theta}_T^{mdrv}$, $\tilde{\theta}_T^{lhr2}$ and $\tilde{\theta}_T^{lhrv}$, the likelihood-based estimators are superior to the minimum-distance ones in terms of moment conditions. However, it is hard to compare the efficiency between the R^2-based and the RV-based estimators, $\tilde{\theta}_T^{mdr2}$ vs. $\tilde{\theta}_T^{mdrv}$, and $\tilde{\theta}_T^{lhr2}$ vs. $\tilde{\theta}_T^{lhrv}$, because the sign of $E\left[(R_t^2 - V_{t|t-1}(\theta_0))^2 - (RV_t - V_{t|t-1}(\theta_0))^2|\mathcal{F}_{t-1}^d\right] = E(R_t^4 - RV_t^2|\mathcal{F}_{t-1}^d)$ is unclear for an arbitrary return process. We consider a special case.

Corollary 1. *Suppose the DGP is a semi-strong GARCH*$(1,1)$ *and* $E(r_s^3|\mathcal{F}_{s-1/m}^d) \ge 0$. *Then* $E(R_t^4 - RV_t^2|\mathcal{F}_{t-1}^d) > 0$. *Under the assumptions in Theorems* 2 *and* 3, $\tilde{\theta}_T^{mdrv}$ *(or* $\tilde{\theta}_T^{lhrv}$*) has a smaller asymptotic variance than* $\tilde{\theta}_T^{mdr2}$ *(or* $\tilde{\theta}_T^{lhr2}$*).*

Remark 5. The observation in Corollary 1 does not provide a helpful guide when it comes to practical implementation, since conditional skewness varies significantly over time (see, for instance, Harvey and Siddique (1999), Harvey and Siddique (2000), and Ghysels, Plazzi, and Valkanov (2011), among others). The squared daily return R_t^2 is an extremely noisy estimator of ex-post volatility compared to the realized variance RV_t. The latter results in a more accurate ex-post volatility forecast.

Remark 6. Theorems 1, 2, and 3 discuss long-span asymptotics: we fix the length between observations and let the time span increase. It is natural to tie the discretely sampled returns to the underlying continuous-time log price process, and to examine the behavior of the estimators when the length between observations decreases. Let $(p_s)_{-\infty<s<\infty}$ be a Brownian semimartingale: $dp_s = \sigma_s dW_s$, with volatility σ_s strictly stationary and α-mixing of size $-(2+v_2)/v_2$, and $E(\sigma_s^{4(2+v_2)}) < \infty$. Under Scenario 1, the discretely sampled returns satisfy Assumption 6 due to the Burkholder-Davis-Gundy inequality and the definition of α-mixing. Note that θ_0 might depend on m. We assume that θ_0 is interior to Θ uniformly in m, and Σ^{md} and Ω^{mdrv} are positive definite uniformly in m. An application of Theorem 5.1 of Gallant and White (1988) yields that $\left(\Omega^{mdrv}\right)^{-1/2}\Sigma^{md}\sqrt{T}(\tilde{\theta}_T^{mdrv} - \theta_0) \Rightarrow N(0, I_{d+3})$, as $m, T \to \infty$, where I_{d+3} is an identity matrix.

Under Scenario 2, suppose that the volatility σ_t is strictly stationary ergodic, and integrable of order 8. If Σ^{md}, Ω^{mdrv}, Ω^{mdr2}, Σ^{lh}, Ω^{lhrv}, and Ω^{lhr2} are positive definite uniformly in m, we have $\left(\Omega^{mdrv}\right)^{-1/2}\Sigma^{md}\sqrt{T}(\tilde{\theta}_T^{mdrv} - \theta_0)$, $\left(\Omega^{mdr2}\right)^{-1/2}\Sigma^{md}\sqrt{T}(\tilde{\theta}_T^{mdr2} - \theta_0)$, $\left(\Omega^{lhrv}\right)^{-1/2}\Sigma^{lh}\sqrt{T}(\tilde{\theta}_T^{lhrv} - \theta_0)$, and $\left(\Omega^{lhr2}\right)^{-1/2}\Sigma^{lh}\sqrt{T}(\tilde{\theta}_T^{lhr2} - \theta_0)$ are asymptotically $N(0, I_{d+3})$. When $m \to \infty$, $RV_t = \int_{t-1}^t \sigma_s^2 ds + o_p(1)$, but $R_t^2 = \int_{t-1}^t \sigma_s^2 ds + 2\int_{t-1}^t \int_{t-1}^s \sigma_{s'} dW_{s'} \sigma_s dW_s$. This also explains why the RV-based estimators are desirable.

3.2.2. Multiplicative error models

Inspired by the Multiplicative Error Model of Engle (2002) and the subsequent work by Engle and Gallo (2006), Lanne (2006), Cipollini, Engle, and Gallo (2006), we consider the model $RV_{t+1} = \sigma_{t+1|t}^2 \eta_{t+1}$, $t \in \mathbb{Z}$, where η_{t+1} is i.i.d. with mean 1, and $\sigma_{t+1|t}^2$ is the conditional expectation of RV_{t+1} up to time t, $\sigma_{t+1|t}^2 = V_{t+1|t}(\theta_0)$. One can estimate θ_0 using the marginal distribution of η_t, with the estimator denoted by $\tilde{\theta}_T^{mem}$.

The marginal distribution is commonly chosen as unit exponential (see Engle (2002)), Gamma distribution as suggested in Engle and Gallo (2006), or a mixture of two Gamma distributions (see Lanne (2006)). Suppose that the error term is Gamma distributed. The conditional density of RV_{t+1} is $f(RV_{t+1}|\mathcal{F}_t^d) = \Gamma(g)^{-1} g^g RV_{t+1}^{g-1} (\sigma_{t+1|t}^2)^{-g} \exp(-gRV_{t+1}/\sigma_{t+1|t}^2)$, and the parameter space is $\Theta \times$

$\{g > 0\}$. Since estimators of θ_0 and g are asymptotically independent (see the discussion in Engle and Gallo (2006) and Cipollini, Engle, and Gallo (2006)), the *point* estimator $\tilde{\theta}_T^{mem}$ is then the same as $\tilde{\theta}_T^{lhrv}$. It follows from the proof of Theorem 3 that $\sqrt{T}(\tilde{\theta}_T^{mem} - \theta_0)$ converges to $N\left(0, (g\Sigma^{lh})^{-1}\right)$ in distribution.

If the existence of an appropriate parametric marginal density cannot be verified, quasi-likelihood estimation is used, which yields the same asymptotic result as $\tilde{\theta}_T^{lhrv}$. But since the innovations are independent, Ω^{lhrv} in Theorem 3 is $E(RV_t^2/V_{t|t-1}^2(\theta_0) - 1)\Sigma^{lh}$, and hence the moment condition $Er_{k/m}^{4+v} < \infty$ is unnecessary. The asymptotic variance-covariance matrix is $(E\eta_t^2 - 1)(\Sigma^{lh})^{-1}$.

4. Continuous Time Limits

With $\sigma_{t+1|t}^2$ as the volatility prediction based on the discretely sampled returns, it is natural to link it with the continuous-time log-price process. Denote by $[p,p]_t$ the quadratic variation of the process (p_s) over $[0,t]$, and $\mathcal{F}_t^c \doteq \sigma(p_s, s \le t)$ the sigma field generated by the continuous-time log-price process. Then for $t \in \mathbb{Z}$, $E_t([p,p]_{t+1} - [p,p]_t) \doteq E([p,p]_{t+1} - [p,p]_t|\mathcal{F}_t^c)$ is the predictable daily increment of quadratic variation of the log price process. In this section, we discuss a special case in which we can explicitly link $E_t([p,p]_{t+1} - [p,p]_t)$, $\sigma_{t+1|t}^2$ and $V_{t+1|t}$. This allows us to examine how the HYBRID process $\{H_t(\phi), t \in \mathbb{Z}\}$ is structurally linked to $(p_s)_{-\infty<s<\infty}$. Moreover, we characterize how the presence of jumps has an impact on the HYBRID process.

Inspired by Drost and Werker (1996), we consider a continuous-time GARCH model as the data generating process:

$$\begin{aligned} dp_t &= \sigma_t dL_t, \\ d\sigma_t^2 &= \theta(\omega - \sigma_t^2)dt + \sqrt{2\lambda\theta}\sigma_t^2 dB_t, \\ L_t &= \sqrt{1-\eta}W_t + \sqrt{\eta}N_t, \end{aligned} \tag{4.1}$$

where $\theta > 0$, $\omega > 0$, $\lambda \in (0,1)$, $\eta \in [0,1]$. B_t and W_t are standard Brownian motions. N_t is a compound Poisson process with jump measure J_N and Lévy measure $\nu(dy) = \zeta f(dy)$ where f is the Normal density with mean 0 and variance $1/\zeta$. Moreover, B_t, W_t, and N_t are independent of each other.

The discretely sampled return $r_s = p_s - p_{s-1/m}$, where $s = k/m$ and $k \in \mathbb{Z}$, follows a weak GARCH(1,1)

$$P_l(r_{s+1/m}|\mathcal{L}_s) = 0, \quad \sigma_{s+1/m|s}^2 = a + b\sigma_{s|s-1/m}^2 + cr_s^2, \tag{4.2}$$

where $\sigma_{s+1/m|s}^2$ is the orthogonal projection of $r_{s+1/m}^2$ onto $\mathcal{L}_s$. The coefficients (a,b,c) are functionals of the structural parameters in (4.1). The relationship between (a,b,c,κ) and $(\theta,\omega,\lambda,v_L^*)$ can be found in Drost and Werker (1996,

p. 42), where κ is the kurtosis of r_s and $v_L^* = EL_1^4 - 3 = 3\eta^2/\zeta$. With some algebra, we have $\sigma^2_{s+k/m|s} = a\left(1-(b+c)^{k-1}\right)/(1-(b+c)) + (b+c)^{k-1}\sigma^2_{s+1/m|s}$ for $k \in \mathbb{Z}^+$. Consequently, the total volatility over the period $(t, t+1]$ for $t \in \mathbb{Z}$, $\sum_{k=1}^{m} \sigma^2_{t+k/m|t} = P_l(RV_{t+1}|\mathcal{L}_t) = \sigma^2_{t+1|t}$, can be characterized by

$$\sigma^2_{t+1|t} = \alpha_m + \beta_m \sigma^2_{t|t-1} + \gamma_m \sum_{j=0}^{m-1} \beta_m^{j/m} r^2_{t-j/m}, \tag{4.3}$$

where $\alpha_m = a((1-b^m)/(1-b))((m(1-b)-cd_m)/(1-(b+c)))$, $\beta_m = b^m$, $\gamma_m = cd_m$, and $d_m = (1-(b+c)^m)/(1-(b+c))$. Equation (4.3) has the form of (2.2) with $H_t = \sum_{j=0}^{m-1} \beta_m^{j/m} r^2_{t-j/m}$, and $\sigma^2_{t+1|t}$ coincides with the HYBRID GARCH process $V_{t+1|t}$.

Remark 7. When the return series follows a weak GARCH(1,1) as in (4.2), the HYBRID process is structurally linked to the return process, and hence H_t is referred to as a structural HYBRID process. The structural HYBRID specification allows the parameters evaluated under different sampling frequencies to be linked to each other explicitly:

$$\alpha_m = \alpha_1 \tfrac{1-\beta_1^m}{1-\beta_1} \tfrac{m(1-\beta_1)-\gamma_1 d_m}{1-(\beta_1+\gamma_1)}, \quad \beta_m = \beta_1^m, \quad \gamma_m = \gamma_1 d_m, \tag{4.4}$$

where $d_m = (1-(\beta_1+\gamma_1)^m)/(1-(\beta_1+\gamma_1))$, and $(\alpha_1, \beta_1, \gamma_1) = (a, b, c)$. A direct implication of (4.4) is that one can use parameter estimates from say a daily model with for example 5-min returns, to formulate a weekly or lower frequency model with the same 5-min returns.

We turn our attention to the predictable increment of quadratic variation. Note that $[p,p]_t = (1-\eta)\int_0^t \sigma_s^2 ds + \eta \int_0^t \int_{-\infty}^{\infty} \sigma_s^2 y^2 J_N(ds, dy)$. We start with examining how $V_{t+1|t}$ (or $\sigma^2_{t+1|t}$) relates to the prediction $E_t([p,p]_{t+1} - [p,p]_t)$. We write $V_{t+1|t}$ as $V^{(m)}_{t+1|t}$ to emphasize the role of sampling frequency m. On the one hand, using a continuous time filtration, the forecast of the daily increment of the quadratic variation is

$$E_t([p,p]_{t+1} - [p,p]_t) = \omega\left(1 - \theta^{-1}(1-e^{-\theta})\right) + \theta^{-1}(1-e^{-\theta})\sigma_t^2, \quad t \in \mathbb{Z}. \tag{4.5}$$

On the other hand, the forecast using $\mathcal{L}_t$ yields the HYBRID GARCH equation (4.3). In view of the relation between $(\theta, \omega, \lambda, v_L^*)$ and (a, b, c, κ) stated in Drost and Werker (1996), we have the following.

Proposition 4. *When the Lévy measure associated with the jump process features excess kurtosis, $v_L^* > 0$, then*

$$\lim_{m\to\infty} \alpha_m = \omega\left(1 - e^{-\theta(1+\phi)}\right)\left(1 - \frac{\phi}{1+\phi}\theta^{-1}(1-e^{-\theta})\right),$$
$$\lim_{m\to\infty} \beta_m = e^{-\theta(1+\phi)}, \quad \lim_{m\to\infty} \gamma_m = (1-e^{-\theta})\phi,$$
$$\lim_{m\to\infty} \sum_{j=0}^{m-1} \beta_m^{j/m} r_{t-j/m}^2 = \int_{(t-1,t]} e^{-\theta(1+\phi)(t-s)} d[p,p]_s \quad \textit{in probability,} \tag{4.6}$$

where $\phi = \sqrt{1 + 2\lambda/(\theta v_L^*)} - 1$. When there are no jumps in the price process, $v_L^* = 0$, we have $\lim_{m\to\infty} \alpha_m = \omega\left(1 - \theta^{-1}(1-e^{-\theta})\right)$, $\lim_{m\to\infty} \beta_m = 0$, and $\lim_{m\to\infty} \gamma_m/\sqrt{m} = \sqrt{\lambda/\theta}(1-e^{-\theta})$. Moreover $\sqrt{m}\sum_{j=0}^{m-1} \beta_m^{j/m} r_{t-j/m}^2$ converges to $(\theta\lambda)^{-1/2}\sigma_t^2$ in L^2.

Proof. see Section S3.5 of Chen, Ghysels, and Wang (2014).

Though $RV_{t+1} = \sum_{j=0}^{m-1} r_{t+1-j/m}^2$ is a consistent estimator of $[p,p]_{t+1} - [p,p]_t$, the HYBRID GARCH process $V_{t+1|t}^{(m)}$ may not consistently estimate $E_t([p,p]_{t+1} - [p,p]_t)$. From Proposition 4, $V_{t+1|t}^{(m)}$ consistently estimates $E_t([p,p]_{t+1} - [p,p]_t)$ only when there are no jumps in the price process.

Corollary 2. *Given a continuous time GARCH (4.1) as the DGP, the process $\{V_{t+1|t}^{(m)}, t \in \mathbb{Z}\}_{m\geq 1}$ at (4.3) converges to $\{E_t([p,p]_{t+1} - [p,p]_t), t \in \mathbb{Z}\}$ uniformly on compacts in probability if and only if there are no jumps in the price process.*

Proof. see Section S3.5 of Chen, Ghysels, and Wang (2014).

Remark 8. The realized multipower variation cannot be incorporated into our framework; the underlying information set is $\mathcal{L}_t$, the closed span of $\{1, r_{t-k/m}, r_{t-k/m}^2;\ k = 0, 1, 2 \ldots\}$, and the realized multipower variation is made up of products of the successive returns.

Without jumps, the HYBRID GARCH process still involves intra-period weighted returns: (4.3) has intra-period weights that are powers of β_m. It follows from the proof of Proposition 4 that what drives the HYBRID process (as $m \to \infty$) is the instantaneous volatility σ_t^2, not the integrated process estimated by the RV. The instantaneous volatility σ_t^2 can be consistently estimated by the same intra-period weighted sum $mc\sum_{j=0}^{m-1} b^j r_{t-j/m}^2$. Put differently, we can view the HYBRID process as a spot volatility estimator that shares some features with other data-driven spot volatility estimators considered by Foster and Nelson (1996), Zhang (2001), Andreou and Ghysels (2002), Fan, Fan, and Jiang

(2007), Fan and Wang (2008), Mykland and Zhang (2008), Zhao and Wu (2008), Malliavin and Mancino (2005), among others.

One could think of continuous time limits in the general HYBRID filtering context and potentially link them to $E_t([p,p]_{t+1} - [p,p]_t)$. The discussion relies on the approach of Drost and Werker (1996) using exact discretization limits, which is compatible with structural HYBRID processes. The continuous time limits of the HYBRID GARCH process also pertain to the joint long-span and infill asymptotics of minimum distance and quasi-likelihood estimators. We leave the broader question of diffusion limits, as in Nelson (1992), Nelson and Foster (1995), among others, and HYBRID filtering processes for future research.

5. Simulation Study

We consider a finite sample simulation study in this section as it will become clear that asymptotic analysis is not sufficient to appraise which estimators are the most attractive for empirical work.

We consider two data generating processes. The first is a discrete-time GARCH process, *strong* GARCH(1,1):

$$\begin{aligned} r_{s+1/m} &= \sqrt{v_{s+1/m|s}}\varepsilon_{s+1/m}, \quad \varepsilon_{s+1/m} \overset{i.i.d.}{\sim} N(0,1), \\ v_{s+1/m|s} &= a + bv_{s|s-1/m} + cr_s^2, \end{aligned} \tag{5.1}$$

where $s = k/m$ and $k \in \mathbb{Z}$. The second is a GARCH diffusion process: Model (4.1) with $\eta = 0$. The discretely-sampled high frequency return $r_s = p_s - p_{s-1/m}$, where $s = k/m$ and $k \in \mathbb{Z}$, is therefore a *weak* GARCH(1,1):

$$\begin{aligned} P_l(r_{s+1/m}|\mathcal{L}_s) = 0, \quad P_l(r^2_{s+1/m}|\mathcal{L}_s) = \sigma^2_{s+1/m|s}, \\ \sigma^2_{s+1/m|s} = a + b\sigma^2_{s|s-1/m} + cr_s^2, \end{aligned} \tag{5.2}$$

where $a = m^{-1}\omega(1 - e^{-m^{-1}\theta})$, $c = e^{-m^{-1}\theta} - b$, and $|b| < 1$ is the solution to $b/(1+b^2) = (\rho e^{-\theta/m} - 1)/(\rho(1 + e^{-2\theta/m}) - 2)$ with $\rho = [4(e^{-\theta/m} - 1 + \theta/m) + 2\theta/m(1+\theta/m(1-\lambda)/\lambda)]/(1-e^{-2\theta/m})$. We then construct the HYBRID GARCH process (4.3) based on model (5.1) and model (5.2) respectively.

The values of the parameters in model (5.1) were $a = 2.8E-06, b = 0.9770, c = 0.0225$, and are taken from Meddahi and Renault (2004). It is easy to check that r_s has finite 8^{th} moment (see Bollerslev (1986)) and it satisfies Assumption 5. For the GARCH diffusion process, we considered $\theta = 0.0350$, $\omega = 0.6365$, $\lambda = 0.2962$ (values taken from Andersen, Bollerslev, and Lange (1999)). The values of α_m, β_m and γ_m are reported in Table 1.

In the simulation experiment, we considered 1,000 replications of sample path (4.3), each having the first 1,000 observations burn-in and consisting of 500

Table 1. Summary of structural parameters in simulation study.

	a	b	c	α_m	β_m	γ_m
			Strong GARCH (1,1)			
$m = 5$	2.8E-06	0.9770	0.0225	0.0001	0.8902	0.1124
$m = 78$	2.8E-06	0.9770	0.0225	0.0147	0.1628	1.7216
$m = 288$	2.8E-06	0.9770	0.0225	0.1429	0.0012	6.0365
			Weak GARCH (1,1)			
$m = 24$	3.86e-05	0.9794	0.0192	0.0216	0.6065	0.4523
$m = 144$	1.07e-06	0.9915	0.0082	0.0204	0.2945	1.1619
$m = 288$	2.69e-07	0.9940	0.0059	0.0195	0.1776	1.6590

and 1,000 observations left in the estimation sample. For the continuous-time case, we used Euler discretization to simulate the diffusion process: take one day as a reference measure, and simulate 24 hours of trading with $dt = 1/86{,}400$.

The estimators considered were: $\tilde{\theta}_T^{mdrv}$, defined at (3.2), the companion estimator $\tilde{\theta}_T^{mdr2}$ replacing RV by R^2, the (quasi-)likelihood-based estimators $\tilde{\theta}_T^{lhr2}$ defined at (3.4), and $\tilde{\theta}_T^{lhrv}$ at (3.5). The study also included the MEM method described in Section 3.4. Note that the estimators of θ_0 and g are asymptotically independent, and thus $\tilde{\theta}_T^{mem}$ is *asymptotically* the same as $\tilde{\theta}_T^{lhrv}$.

The simulation results for strong GARCH(1,1) are reported in Table 2. The results for weak GARCH are similar and therefore appear in the companion document Chen, Ghysels, and Wang (2014, Table 1). To streamline the discussion we refer to the estimators $\tilde{\theta}_T^{mdrv}$, $\tilde{\theta}_T^{mdr2}$, $\tilde{\theta}_T^{lhrv}$, $\tilde{\theta}_T^{lhr2}$, and $\tilde{\theta}_T^{mem}$ as respectively MDRV, MDR2, LHRV, LHR2, and MEM. The numbers in parenthesis in Table 2 are MSE for LHR2, and relative MSE (with respect to LHR2) for LHRV, MDR2, MDRV, and MEM. For g (in the MEM estimator), we only report the sample variance.

The results in Table 2 are quite easy to summarize. The bold-faced entries between parentheses indicate the best estimator for the various parameters. The estimator that appears to have the best finite sample properties is LHRV. It is typically vastly better than the estimators based on R^2, either minimum distance or likelihood-based. Compared to the LHR2 estimator, we also see that MDRV - which uses also RV but via a minimum distance criterion - is also less efficient, except when $m = 5$ and $T = 1{,}000$. The MEM estimator, which is asymptotically equivalent to LHRV, is occasionally in small samples the most efficient for one parameter in particular, namely α_m. This means that the most efficient estimation of the unconditional mean of the volatility dynamic process can be achieved with the MEM principle which directly estimates the volatility process.

Overall, the simulation results suggest that $\tilde{\theta}_T^{lhrv}$ and $\tilde{\theta}_T^{mem}$ have the most desirable finite sample properties.

Table 2. **Small sample property of various estimators, Strong GARCH.** The table displays estimation of α_m, b, γ_m (and g for the MEM estimation procedure) of a high frequency data GARCH(1,1) appearing in equation (5.1) with sample size 500 (Panel I:III) and sample size 1,000 (Panel IV:VI), where the true values of α_m, b, γ_m are shown in the first line of each panel. The estimators considered are: *mdrv*, defined in (3.2), and the companion estimator *mdr*2, replacing *RV* by R^2, as well as (quasi-)likelihood-based estimators *lhr*2, defined in (3.4), and *lhrv*, defined in (3.5). The table also includes the *mem* method described in subsection 3.4. The numbers in the parenthesis are MSE for *lhr*2, relative MSE (with respect to *lhr*2) for *lhrv*, *mdr*2, *mdrv*, *mem*. For g, we only report sample variance.

	α_m		b		γ_m		g
Panel I: $m = 5$, $T = 500$							
True Value	0.000070		0.977000		0.112388		
LHR2	0.000259	(0.000000)	0.972401	(0.000194)	0.128465	(0.004700)	
LHRV	0.000261	(1.015169)	0.974313	(**0.271144**)	0.119444	(**0.258291**)	
MDR2	0.000454	(9.371349)	0.966299	(11.194958)	0.145875	(5.802282)	
MDRV	0.000328	(6.150788)	0.975828	(0.502932)	0.108347	(0.278161)	
MEM	0.000052	(**0.154507**)	0.976053	(0.782214)	0.117574	(0.790375)	2.344735 (0.022762)
Panel II: $m = 78$, $T = 500$							
True Value	0.014749		0.977000		1.721640		
LHR2	0.022270	(0.000429)	0.971857	(0.000324)	2.012838	(1.733246)	
LHRV	0.016544	(**0.053291**)	0.976488	(**0.018644**)	1.743047	(**0.018305**)	
MDR2	0.052818	(9.689061)	0.944327	(28.476856)	3.505934	(16.788438)	
MDRV	0.029753	(1.037355)	0.975510	(0.319931)	1.742003	(0.327187)	
MEM	0.004353	(0.430669)	0.979426	(0.049645)	1.575400	(0.043334)	9.716956 (6.113690)
Panel III: $m = 288$, $T = 500$							
True Value	0.142879		0.977000		6.036455		
LHR2	0.154505	(0.008603)	0.970251	(0.000693)	7.462548	(38.821448)	
LHRV	0.147024	(**0.062688**)	0.976511	(**0.033043**)	6.133488	(**0.046095**)	
MDR2	0.381905	(22.197008)	0.891170	(63.448471)	20.803390	(38.496335)	
MDRV	0.222810	(2.186104)	0.970839	(1.738098)	7.032169	(1.141032)	
MEM	0.131788	(0.289360)	0.976838	(0.129298)	6.109518	(0.185204)	7.711094 (2.311111)
Panel IV: $m = 5$, $T = 1,000$							
True Value	0.000070		0.977000		0.112388		
LHR2	0.000245	(0.000000)	0.972167	(0.000110)	0.129344	(0.002517)	
LHRV	0.000245	(0.967746)	0.972823	(0.487674)	0.126027	(0.416324)	
MDR2	0.000313	(3.687974)	0.971769	(4.640177)	0.128260	(3.806918)	
MDRV	0.000252	(0.877527)	0.976031	(**0.274003**)	0.109991	(**0.265061**)	
MEM	0.000039	(**0.212108**)	0.976308	(0.955344)	0.116601	(1.014513)	2.347128(0.011373)
Panel V: $m = 78$, $T = 1,000$							
True Value	0.014749		0.977000		1.721640		
LHR2	0.018895	(0.000220)	0.974736	(0.000127)	1.830847	(0.648751)	
LHRV	0.015719	(**0.054137**)	0.976620	(**0.023895**)	1.742325	(**0.024776**)	
MDR2	0.049303	(16.035449)	0.948179	(72.655723)	3.356842	(50.957657)	
MDRV	0.027384	(1.388565)	0.975913	(0.425466)	1.735309	(0.445478)	
MEM	0.003780	(0.858029)	0.979336	(0.105341)	1.585187	(0.090727)	9.866582 (5.116359)
Panel VI: $m = 288$, $T = 1,000$							
True Value	0.142879		0.977000		6.036455		
LHR2	0.144974	(0.004208)	0.974875	(0.000156)	6.480715	(10.762565)	
LHRV	0.145066	(**0.065397**)	0.976639	(**0.187942**)	6.117979	(**0.226982**)	
MDR2	0.344683	(31.382805)	0.912688	(207.120933)	17.660329	(107.729227)	
MDRV	0.210306	(2.883476)	0.972164	(4.841902)	6.853770	(3.094757)	
MEM	0.126214	(0.591231)	0.977462	(0.273088)	5.971204	(0.299844)	7.479886 (3.057745)

6. HYBRID GARCH Models in Practice

The proposed class of HYBRID GARCH processes allows us to investigate some intriguing empirical modeling strategies. We pose a number of practical questions pertaining to model specification.

- To predict daily volatility, are we better off estimating intra-daily weighting schemes, despite the additional parameters involved, compared to using realized volatility and related data-driven HYBRID processes?
- How should we handle asymmetries? Can we simply rely on sign-sensitive aggregates such as semi-variances, or should we rather estimate intra-daily news impact curves as suggested by Chen and Ghysels (2011), albeit at a cost of additional parameters?
- When we are interested in weekly horizon forecasts, should we keep using intra-daily data with their own weighting scheme, or should we rely on simple daily realized volatilities? Hence, is the right sampling frequency of returns intra-daily? Or, can we get by with daily aggregates?
- Considering the recent financial crisis, should we choose a totally different modeling strategy, or do model specifications successful prior to the crisis remain?

To address these questions we need to formulate appropriate HYBRID filtering processes that feature weighting schemes, news impact functions etc. We start with the specification:

$$H(\phi, \vec{r}_t) = \sum_{j=0}^{m-1} \Psi_j(\phi) r_{t-j/m}^2, \quad \sum_{j=0}^{m-1} \Psi_j(\phi) = 1, \tag{6.1}$$

where the weights $(\Psi_0(\phi), \Psi_1(\phi), \Psi_2(\phi), \ldots, \Psi_{m-1}(\phi))^\top$ are determined by the low-dimensional functional specification used by Chen and Ghysels (2011), and inspired by the MIDAS regression format of Ghysels, Santa-Clara, and Valkanov (2006), Ghysels, Sinko, and Valkanov (2006), and Ghysels, Rubia, and Valkanov (2009). The weighting schemes can also handle intra-daily seasonal patterns - a topic discussed in further detail by Chen, Ghysels, and Wang (2011).

Deviating from the linear projection paradigm, the HYBRID filtering structure allows us also to consider HYBRID GARCH models that feature intra-daily news impact curves, similar to the framework of Chen and Ghysels (2011), except that the latter uses a MIDAS regression format. The HYBRID processes we consider are

$$H(\phi, \vec{r}_t) = \sum_{j=0}^{m-1} \Psi_j(\phi) NIC(\phi, r_{t-j/m}), \quad \sum_{j=0}^{m-1} \Psi_j(\phi) = 1, \tag{6.2}$$

where $NIC(\phi, \cdot)$ stands for a high frequency data news impact curve. The parameter vector ϕ contains parameters pertaining to the weights as well as the news impact curve. Regarding the specification of the latter, we consider

$$NIC(\phi, r) = br^2 + \delta r^2 \mathbf{1}_{r<0}, \tag{6.3}$$

with b and δ the parameters that are in ϕ. The HYBRID process constructed using news impact curve (6.3) is in Scenario 2. A detailed discussion about the weights and the parameter space is provided in Section S2 of Chen, Ghysels, and Wang (2014).

As an application, we consider a comparison of daily and weekly volatility forecasts using models that belong to the HYBRID GARCH family. The models are shown in Tables 5, 6, and 7.

We used the S&P 500 Futures 5-minute returns from April 21st, 1982 to March 31st, 2013. To estimate the models and perform out-of-sample evaluations, we used a rolling window sampling scheme that moves forward monthly. There are 229 rolling windows and each window contains 120-months in-sample data and 24-month set-aside for out-of-sample appraisals. To understand the impact of the subprime mortgage crisis on modeling strategies, we also considered two subsets: a before-crisis sample that contained 160 rolling windows from April 21st, 1982 to July, 2007; an including-crisis sample that contained 69 rolling windows from August, 1996 until March, 2013.

All models were estimated using MEM and LHRV, inspired by our Monte Carlo simulation findings. The out-of-sample forecast performance of models using the same estimation method and across different estimation methods were evaluated via the Giacomini and White (2006) test, henceforth denoted by GW, which can be viewed as a generalization, or a conditional version of the Diebold and Mariano (1995) and West (1996) tests. Another appeal of using the GW test is that it can handle non-nested models, which is the case in our application. In fact, Giacomini and White (2006) stress the difference between what they call forecasting methods versus forecasting models. Loosely speaking, forecasting methods are the combination of estimation sample, model specification, and prediction sample. In our application this is most relevant, as we not only compare models involving high frequency data directly, or daily aggregate measures, but we also include ARCH-type models involving daily returns, the original models that were used in the literature on asymmetries in volatility. The loss function we used in the GW test is QLike, which has desirable properties and is robust to measurement error noises in volatility (see Patton (2011)). The QLike loss function is defined as $L(h_t, RV_t) = \log h_t + RV_t/h_t - (\log RV_t + 1)$.

We constructed a performance score for each model based on the GW tests' p-values and ratios,

$$s_A = \frac{\sum_{B \neq A} \mathbf{1}_{p_{AB}<=\alpha} r_{AB} + 0.5(1 - \mathbf{1}_{p_{AB}<=\alpha})}{(n(n-1)/2)}, \tag{6.4}$$

where s_A is the score of model A; p_{AB} is the p-value of GW test comparing models A and B; α is the significance level and set as 10% in the paper; r_{AB} is the ratio of model A being predicted as better choice than model B; n is the number of models. The performance score is normalized such that the summation across all models is one.

We start with a comparisons of conditional predictive ability for each model between the two methods. We skip the details, as the results are easy to summarize. The results consistently show that the choice of estimation method, LHRV or MEM, is not important, with a few exceptions where LHRV is preferred. Hence, to compare the forecast performance of models, we only consider the LHRV estimation method.

The ranks of forecast performance based on GW tests are shown in Tables 3 and 4 for daily models and weekly models, respectively. Examining the results in Table 3, we note the following for daily model comparisons. GARCH and TGARCH models using daily returns are dominated by models using intra-daily data. Symmetric models are always dominated by their asymmetric counterparts, which implies that asymmetry does matter despite the extra parameters required to estimate the news impact curves. Focusing on the first two columns of Table 3 for full-sample results, we find that for symmetric models, RV GARCH model is dominated by all other HYBRID or slope constrained HYBRID models; for asymmetric models, SemiRV GARCH model is also always less preferred than any HYBRID or slope constrained HYBRID asymmetric model. These results support two key findings for daily forecasting of volatility: (i) asymmetries matter and (ii) the weighting scheme does matter. Including the samples which cover the recent financial crisis does have some impact on modeling strategy. Focusing on the last two columns of Table 3, we find that SemiRV (RV) GARCH model is not always dominated by HYBRID models in the asymmetric (symmetric) model category. However, whether or not we consider to include the crisis, the best model always belongs to the HYBRID TGARCH class, with or without slope and other parameter constraints.

Table 3. Rank of Models Based on Giacomini-White Tests for Daily Models.

Rank 1 is the best forecast model.

	1982.4 to 2013.3		1982.4 to 2007.7		1996.8 to 2013.3	
Rank	Name	Score	Name	Score	Name	Score
1	FC1 HYBRID SC TGARCH	0.1130	HYBRID TGARCH	0.1117	FC0 HYBRID SC TGARCH	0.1078
2	HYBRID TGARCH	0.1075	FC1 HYBRID SC TGARCH	0.1088	HYBRID SC TGARCH	0.1074
3	FC0 HYBRID SC TGARCH	0.1055	FC1 HYBRID TGARCH	0.1009	FC1 HYBRID SC TGARCH	0.1016
4	FC1 HYBRID TGARCH	0.0998	FC0 HYBRID SC TGARCH	0.1004	SemiRV GARCH	0.0933
5	HYBRID SC TGARCH	0.0951	FC0 HYBRID TGARCH	0.0934	HYBRID TGARCH	0.0897
6	FC0 HYBRID TGARCH	0.0945	HYBRID SC TGARCH	0.0879	FC0 HYBRID TGARCH	0.0806
7	SemiRV GARCH	0.0809	SemiRV GARCH	0.0835	FC1 HYBRID TGARCH	0.0755
8	HYBRID GARCH	0.0616	HYBRID GARCH	0.0716	FC0 HYBRID SC GARCH	0.0728
9	FC0 HYBRID SC GARCH	0.0517	FC1 HYBRID SC GARCH	0.0494	RV GARCH	0.0685
10	FC1 HYBRID SC GARCH	0.0486	FC1 HYBRID GARCH	0.0473	HYBRID SC GARCH	0.0482
11	FC1 HYBRID GARCH	0.0392	FC0 HYBRID SC GARCH	0.0400	HYBRID GARCH	0.0469
12	FC0 HYBRID GARCH	0.0375	FC0 HYBRID GARCH	0.0386	FC0 HYBRID GARCH	0.0391
13	HYBRID SC GARCH	0.0284	HYBRID SC GARCH	0.0302	FC1 HYBRID SC GARCH	0.0388
14	RV GARCH	0.0255	RV GARCH	0.0238	FC1 HYBRID GARCH	0.0214
15	TGARCH	0.0114	TGARCH	0.0125	TGARCH	0.0083
16	GARCH	0.0000	GARCH	0.0000	GARCH	0.0000

Table 4. Rank of models based on Giacomini-White tests for weekly models. Rank 1 is the best forecast model.

	1982.4 to 2013.3		1982.4 to 2007.7		1996.8 to 2013.3	
Rank	Name	Score	Name	Score	Name	Score
1	FC0 HYBRID SC TGARCH	0.06359	FC0 HYBRID SC TGARCH	0.05923	FC0 HYBRID SC TGARCH	0.06337
2	FC1 HYBRID SC TGARCH	0.06098	SemiRV GARCH	0.05874	FC1 HYBRID SC TGARCH	0.06222
3	FC0 HYBRID SC GARCH	0.05648	FC1 HYBRID SC TGARCH	0.05764	FC0 HYBRID TGARCH	0.05867
4	FC1 HYBRID TGARCH	0.05620	FC0 HYBRID SC GARCH	0.05644	FC1 HYBRID TGARCH	0.05687
5	SemiRV GARCH D	0.05618	FC1 HYBRID TGARCH	0.05460	SemiRV GARCH D	0.05392
6	SemiRV GARCH	0.05402	FC1 HYBRID SC GARCH	0.05458	FC0 HYBRID SC GARCH	0.05372
7	FC0 HYBRID TGARCH	0.05392	SemiRV GARCH D	0.05412	HYBRID TGARCH	0.05127
8	FC1 HYBRID SC GARCH	0.05367	FC0 HYBRID TGARCH	0.05193	HYBRID SC GARCH	0.05022
9	FC0 HYBRID GARCH	0.04841	RV GARCH	0.05183	FC1 HYBRID SC GARCH	0.04878
10	HYBRID SC GARCH	0.04470	FC0 HYBRID GARCH	0.05144	FC0 HYBRID GARCH	0.04823
11	RV GARCH	0.04373	HYBRID SC GARCH	0.04363	FC1 HYBRID GARCH	0.04798
12	HYBRID TGARCH	0.04359	FC1 HYBRID GARCH	0.04293	SemiRV GARCH	0.04598
13	FC1 HYBRID GARCH	0.04258	HYBRID GARCH	0.04224	HYBRID GARCH	0.04468
14	HYBRID GARCH	0.04217	HYBRID TGARCH	0.04140	RV GARCH	0.03673
15	FC1 HYBRID SC TGARCH D	0.02823	RV GARCH D	0.02781	FC0 HYBRID SC TGARCH D	0.03263
16	FC0 HYBRID SC TGARCH D	0.02745	FC0 HYBRID SC GARCH D	0.02703	FC1 HYBRID SC TGARCH D	0.03043
17	FC0 HYBRID SC GARCH D	0.02734	FC0 HYBRID SC TGARCH D	0.02701	FC0 HYBRID TGARCH D	0.02784
18	FC1 HYBRID SC GARCH D	0.02694	FC1 HYBRID SC GARCH D	0.02622	FC0 HYBRID SC GARCH D	0.02494
19	RV GARCH D	0.02677	FC1 HYBRID SC TGARCH D	0.02615	FC1 HYBRID TGARCH D	0.02399
20	FC0 HYBRID TGARCH D	0.02515	FC0 HYBRID GARCH D	0.02309	FC1 HYBRID SC GARCH D	0.02169
21	FC0 HYBRID GARCH D	0.02270	FC0 HYBRID TGARCH D	0.02282	FC0 HYBRID GARCH D	0.01964
22	FC1 HYBRID TGARCH D	0.01967	HYBRID SC TGARCH	0.02279	HYBRID TGARCH D	0.01699
23	FC1 HYBRID GARCH D	0.01858	FC1 HYBRID GARCH D	0.01925	RV GARCH D	0.01689
24	HYBRID SC TGARCH	0.01579	FC1 HYBRID TGARCH D	0.01560	FC1 HYBRID GARCH D	0.01504
25	HYBRID TGARCH D	0.01201	HYBRID SC GARCH D	0.01281	HYBRID GARCH D	0.01479
26	HYBRID GARCH D	0.01063	HYBRID TGARCH D	0.01127	HYBRID SC TGARCH	0.01029
27	HYBRID SC GARCH D	0.00969	HYBRID GARCH D	0.00903	TGARCH	0.00970
28	TGARCH	0.00416	GARCH	0.00402	GARCH	0.00620
29	GARCH	0.00320	TGARCH	0.00357	HYBRID SC GARCH D	0.00410
30	HYBRID SC TGARCH D	0.00145	HYBRID SC TGARCH D	0.00079	HYBRID SC TGARCH D	0.00220

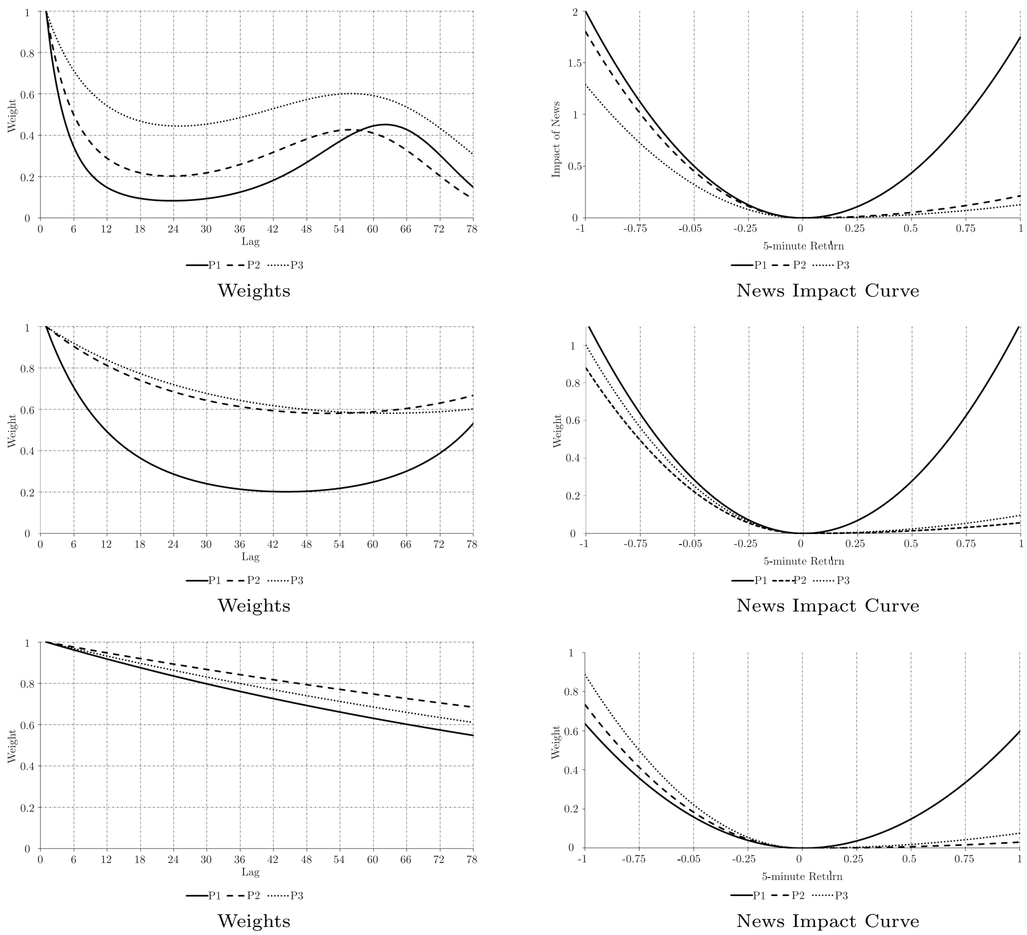

Figure 1. Best Daily Models in Three Subsamples - Panels (a), (c), and (e) display the weights of the best daily models for three samples: FC1 HYBRID SC TGARCH model for the full sample, HYBRID TGARCH model for the before-crisis sample, and FC0 HYBRID SC TGARCH model for the including-crisis sample. The three representative subsamples of our rolling sample scheme are: P1 is for period May 1982 to April 1992; P2 for period September 1991 to August 2001; P3 for period April 2001 to March 2011. Panels (b), (d), and (f), display the news impact curve of the same models in the same samples.

To further document these findings, Figure 1 panels (a), (c), and (e) display the weights of the best daily models for three samples: FC1 HYBRID SC TGARCH model for the full sample, HYBRID TGARCH model for the before-crisis sample, and FC0 HYBRID SC TGARCH model for the including-crisis sample. The three representative subsamples of our rolling sample scheme are:

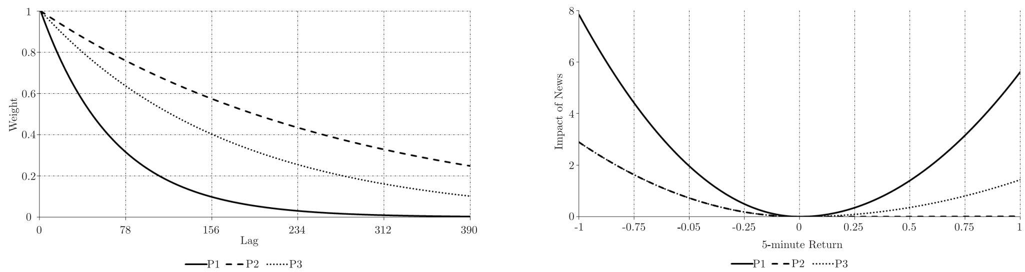

Figure 2. Best Weekly Model in Three Subsamples. The three representative subsamples of our rolling sample scheme are: P1 is for period May 1982 to April 1992; P2 for period September 1991 to August 2001; P3 for period April 2001 to March 2011.

P1 is for period May 1982 to April 1992; P2 for period September 1991 to August 2001; P3 for period April 2001 to March 2011. It appears from the figure that the weighting scheme patterns are fairly stable across subsamples. In Figure 1 panels (b), (d), and (f), we display the news impact curve of the same models in the same samples. We see again a stable pattern across subsamples and a pattern that is distinctly asymmetric and features larger impact of bad news - as commonly documented in the literature.

For weekly models, we report the rank of forecast performance in Table 4. As the forecast horizon increases from one day to one week, the importance of weighting schemes and the asymmetry effect decreases, although they still do matter. Moreover, for the sample which includes the crisis, we note that the SemiRV GARCH model using daily returns shows better performance than the SemiRV GARCH model using 5-minute returns. However, the best forecast model is FC0 HYBRID SC TGARCH model in all three sets of samples. Figure 2 shows the weighting schemes and news impact curves in three representative subsamples of our rolling sample scheme: P1 is for period May 1982 to April 1992; P2 for period September 1991 to August 2001; P3 for period April 2001 to March 2011. We see again a stable weighting scheme pattern and an asymmetric effect, similar as those in daily models.

Table 5. Summary of Model Specifications (A)

$r_{s/m}$ is the high-frequency return and m is the frequency in one day, $r_{s/m}$ is the ith high-frequency return at day t if $s = t-1+i/m, i = 1, \ldots, m$. The return at period t, daily, weekly, and bi-weekly, is denoted as $R_t \equiv \sum_{i=1}^{mh} r_{ht-(i-1)/m}$, where h is the number of days in one period. We use R_t^D to denote daily return at day t. The realized volatility at period t is $RV_t = \sum_{i=1}^{mh} r^2_{ht-(i-1)/m}$, and the realized semivariance is $SemiRV_t = \sum_{i=1}^{mh} \mathbf{1}_{r_{ht-(i-1)/m}<0} r^2_{ht-(i-1)/m}$. All "*** D" models, models No. (11) to (14), (19) to (22), and (27) to (30), are only applied to weekly or bi-weekly data.

(1)	GARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma R_t^2$
(2)	TGARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma R_t^2 + d\mathbf{1}_{R_t<0} R_t^2$
(3)	RV GARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma RV_t$
(4)	SemiRV GARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma RV_t + \delta SemiRV_t$
(5)	RV GARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{h} {R^d_{ht-(j-1)}}^2$
(6)	SemiRV GARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{h} (1 + \delta \mathbf{1}_{R^d_{ht-(j-1)}<0}) {R^d_{ht-(j-1)}}^2$
(7)	HYBRID GARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) r^2_{t-(j-1)/m}$
(8)	HYBRID TGARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) (1 + \delta \mathbf{1}_{r_{t-(j-1)/m}<0}) r^2_{t-(j-1)/m}$
(9)	HYBRID SC GARCH	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{mh} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) V_{t\|t-1}$ $+ \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) r^2_{t-(j-1)/m}$
(10)	HYBRID SC TGARCH	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{mh} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) V_{t\|t-1}$ $+\gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) (1 + \delta \mathbf{1}_{r_{t-(j-1)/m}<0}) r^2_{t-(j-1)/m}$

Table 6. Summary of Model Specifications (B)

(11)	HYBRID GARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2)\right) {R^d_{ht-(j-1)}}^2$
(12)	HYBRID TGARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2)\right)(1 + \delta \mathbf{1}_{R^d_{ht-(j-1)}<0}) {R^d_{ht-(j-1)}}^2$
(13)	HYBRID SC GARCH D	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{h} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2)\right) V_{t\|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2)\right) {R^d_{ht-(j-1)}}^2$
(14)	HYBRID SC TGARCH D	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{h} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2)\right) V_{t\|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2)\right)(1 + \delta \mathbf{1}_{R^d_{ht-(j-1)}<0}) {R^d_{ht-(j-1)}}^2$
(15)	FC1 HYBRID GARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh)\right) r^2_{t-(j-1)/m}$
(16)	FC1 HYBRID TGARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh)\right)(1 + \delta \mathbf{1}_{r_{t-(j-1)/m}<0}) r^2_{t-(j-1)/m}$
(17)	FC1 HYBRID SC GARCH	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{mh} (\phi_0 + \phi_1 i/mh)\right) V_{t\|t-1} + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh)\right) r^2_{t-(j-1)/m}$
(18)	FC1 HYBRID SC TGARCH	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{mh} (\phi_0 + \phi_1 i/mh)\right) V_{t\|t-1} + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh)\right)(1 + \delta \mathbf{1}_{r_{t-(j-1)/m}<0}) r^2_{t-(j-1)/m}$
(19)	FC1 HYBRID GARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h)\right) {R^d_{ht-(j-1)}}^2$
(20)	FC1 HYBRID TGARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h)\right)(1 + \delta \mathbf{1}_{R^d_{ht-(j-1)}<0}) {R^d_{ht-(j-1)}}^2$

Table 7. Summary of Model Specifications (C)

(21)	FC1 HYBRID SC GARCH D	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{h}(\phi_0 + \phi_1 i/h)\right)V_{t\|t-1} + \gamma\sum_{j=1}^{h}\exp\left(\sum_{i=1}^{j-1}(\phi_0 + \phi_1 i/h)\right){R^d_{ht-(j-1)}}^2$
(22)	FC1 HYBRID SC TGARCH D	$V_{t+1\|t} = \alpha + \exp\left(\sum_{i=1}^{h}(\phi_0 + \phi_1 i/h)\right)V_{t\|t-1} + \gamma\sum_{j=1}^{h}\exp\left(\sum_{i=1}^{j-1}(\phi_0 + \phi_1 i/h)\right)(1 + \delta\mathbf{1}_{R^d_{ht-(j-1)}<0}){R^d_{ht-(j-1)}}^2$
(23)	FC0 HYBRID GARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma\sum_{j=1}^{mh}\exp((j-1)\phi_0)r^2_{t-(j-1)/m}$
(24)	FC0 HYBRID TGARCH	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma\sum_{j=1}^{mh}\exp((j-1)\phi_0)(1 + \delta\mathbf{1}_{r_{t-(j-1)/m}<0})r^2_{t-(j-1)/m}$
(25)	FC0 HYBRID SC GARCH	$V_{t+1\|t} = \alpha + \exp(mh\phi_0)V_{t\|t-1} + \gamma\sum_{j=1}^{mh}\exp((j-1)\phi_0)r^2_{t-(j-1)/m}$
(26)	FC0 HYBRID SC TGARCH	$V_{t+1\|t} = \alpha + \exp(mh\phi_0)V_{t\|t-1} + \gamma\sum_{j=1}^{mh}\exp((j-1)\phi_0)(1 + \delta\mathbf{1}_{r_{t-(j-1)/m}<0})r^2_{t-(j-1)/m}$
(27)	FC0 HYBRID GARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \gamma\sum_{j=1}^{h}\exp((j-1)\phi_0){R^d_{ht-(j-1)}}^2$
(28)	FC0 HYBRID TGARCH D	$V_{t+1\|t} = \alpha + \beta V_{t\|t-1} + \sum_{j=1}^{h}\exp((j-1)\phi_0)(\gamma + \delta\mathbf{1}_{R^d_{ht-(j-1)}<0}){R^d_{ht-(j-1)}}^2$
(29)	FC0 HYBRID SC GARCH D	$V_{t+1\|t} = \alpha + \exp(h\phi_0)V_{t\|t-1} + \gamma\sum_{j=1}^{h}\exp((j-1)\phi_0){R^d_{ht-(j-1)}}^2$
(30)	FC0 HYBRID SC TGARCH D	$V_{t+1\|t} = \alpha + \exp(h\phi_0)V_{t\|t-1} + \sum_{j=1}^{h}\exp((j-1)\phi_0)(\gamma + \delta\mathbf{1}_{R^d_{ht-(j-1)}<0}){R^d_{ht-(j-1)}}^2$

7. Discussion

We introduced the versatile class of HYBRID GARCH models involving data sampled at any higher frequency. We studied the theoretical properties as well as statistical inference of this new. An empirical application reports the superior out-of-sample forecasting performance of the new class of models, including during the recent financial crisis. In particular, to predict daily volatility, we find that GARCH and TGARCH models using daily returns are dominated by models using intra-daily data. Symmetric models are always dominated by their asymmetric counterparts, which implies that asymmetry does matter despite the extra parameters required to estimate the news impact curves. Overall, we have two key findings for daily forecasting of volatility: asymmetries matter, and the weighting scheme does matter. These are the salient features of our new class of models. Including the samples which cover the recent financial crisis does have some impact on modeling strategy, but the main conclusions remain. Similar results hold for a weekly volatility forecast horizon. The new class of models also features appealing theoretical properties. We show that in the absence of jumps and a continuous time GARCH diffusion DGP, we obtain unbiased predictions of the increment of quadratic variation. We leave the broader question of diffusion limits - as in Nelson (1992), Nelson and Foster (1995), among others - and HYBRID filtering processes for future research.

Acknowledgement

We thank Rob Engle, Per Mykland, Eric Renault, Neil Shephard, and George Tauchen for insightful comments. An early version of the paper was presented at the Stevanovich Center - CREATES conference *Financial Econometrics and Statistics: Current Themes and New Directions,* Skagen, Denmark 4-6 June 2009 under the title High Frequency GARCH Models. We also thank participants for comments at the Oxford-Mann Institute and the following conferences: 2010 FERM in Taipei, 2010 NBER-NSF Time Series conference, Duke University, 2010 New Researchers in Statistics and Probability in Vancouver, 2010 SoFiE conference in Melbourne, and the 2010 Quantitative Methods in Business Applications in Beijing. The second author benefited from funding from a Marie Curie FP7-PEOPLE-2010-IIF grant.

References

Andersen, T. G., Bollerslev, T. and Lange, S. (1999). Forecasting financial market volatility: sample frequency vis-à-vis forecast horizon. *J. Empirical Finance* **6**, 457-477.

Andreou, E. and E. Ghysels (2002). Rolling sample volatility estimators: some new theoretical, simulation and empirical results. *J. Business and Economic Statist.* **20**, 363-376.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **31**, 307-327.

Chen, X. and Ghysels, E. (2011). News - good or bad - and its impact on predictions over multiple horizons. *Review of Financial Studies* **24**, 46-81.

Chen, X., Ghysels, E. and Wang, F. (2011). Hybrid garch models and intra-daily return periodicity. *J. Time Series Econom.* **3**, Article 11.

Chen, X., Chysels, E. and F. Wang, F. (2014). Hybrid-garch: A generic class of models for volatility predictions using high frequency data — technical appendix. Technical report.

Cipollini, F., Engle, R. and Gallo, G. (2006). Vector multiplicative error models: representation and inference. Wording Paper, University of Florence.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *J. Business and Economic Statist.* **13**, 253-263.

Drost, F. C. and B. J. Werker (1996). Closing the gap: Continuous time garch modeling. *J. Econom.* **74**, 31-57.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* **50**, 987-1008.

Engle, R. (2002). New frontiers for ARCH models. *J. Appl. Econom.* **17**, 425-446.

Engle, R. and Gallo, G. (2006). A multiple indicators model for volatility using intra-daily data. *J. Econom.* **131**, 3-27.

Fan, J., Fan, Y. and Jiang, J. (2007). Dynamic integration of time-and state-domain methods for volatility estimation. *J. Amer. Statist. Assoc.* **102**, 618-631.

Fan, J. and Y. Wang (2008). Spot volatility estimation for high-frequency data. *Statist. and its Interface*, 279-288.

Foster, D. P. and Nelson, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64**, 139-174.

Gallant, A. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell Oxford.

Ghysels, E., Plazzi, A. and Valkanov, R. (2011). Conditional skewness of stock market returns in developed and emerging markets and its economic fundamentals. *Swiss Finance Institute Research Paper* (11-06).

Ghysels, E., Rubia, A. and Valkanov, R. (2009). Multi-period forecasts of volatility: direct, iterated, and mixed-data approaches. Working paper, UNC.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005). There is a risk-return tradeoff after all. *J. Financial Econom.* **76**, 509-548.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131**, 59–95.

Ghysels, E., Sinko, A. and Valkanov, R. (2006). MIDAS regressions: further results and new directions. *Econom. Rev.* **26**, 53-90.

Ghysels, E. and R. Valkanov (2012). Forecasting volatility with MIDAS. In *Handbook of Volatility Models and Their Applications* (Edited by L. Bauwens, C. Hafner, and S. Laurent), 383-401. Wiley.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* **74**, 1545-1578.

Goncalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *J. Econom.* **119**, 199-219.

Harvey, C. R. and Siddique, A. (1999). Autoregressive conditional skewness. *J. Financial and Quantitative Anal.* **34**.

Harvey, C. R. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *J. Finance* **55**, 1263-1295.

Jennrich, R. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.

Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *J. Financial Econom.* **4**, 594.

Malliavin, P. and Mancino, M. (2005). A Fourier transform method for nonparametric estimation of multivariate volatility. *Ann. Statist.* **37**, 1983-2010.

Meddahi, N. and Renault, E. (2004). Temporal aggregation of volatility models. *J. Econom.* **119**, 355-379.

Mykland, P. A. and Zhang, L. (2008). Inference for volatility-type objects and implications for hedging. *Statist. and Its Interface* **1**, 255-278.

Nelson, D. (1992). Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *J. Econom.* **60**, 61-90.

Nelson, D. and Foster, D. (1995). Filtering and forecasting with misspecified ARCH models II: Getting the right variance with the wrong model. *J. Econom.* **67**, 303-335.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *J. Econom.* **160**, 246-256.

West, K. (1996). Asymptotic inference about predictive ability. *Econometrica* **64**, 1067-1084.

Zhang, L. (2001). From martingales to ANOVA: Implied and realized volatility. The University of Chicago: Ph.D. dissertation, Department of Statistics.

Zhao, Z. and Wu, W. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36**, 1854-1878.

ETS, SAS Institute Inc., Cary, NC 27513, USA.

E-mail: Xilong.Chen@sas.com

Department of Finance, Kenan-Flagler Business School, and Department of Economics, University of North Carolina at Chapel Hill, NC 27599, USA.

E-mail: eghysels@unc.edu

Department of Information and Decision Sciences, University of Illinois at Chicago, NC 27599, USA.

E-mail: ffwang@uic.edu

(Received September 2012; accepted May 2014)