HYBRID-GARCH
A Generic Class of Models for Volatility Predictions using Mixed Frequency Data*

Xilong Chen[†] **Eric Ghysels**[‡] **Fangfang Wang**[§]

First Draft: April 2009

This Draft: April 10, 2011

Abstract

We propose a general GARCH framework that allows the 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 the volatility dynamics are driven by what we call HYBRID processes. The HYBRID processes can involve data sampled at any frequency.

Keywords: HYBRID process, weak GARCH, GARCH jump diffusion, realized measure, temporal aggregation, filtering, misspecification

---

[*] We like to 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 like to 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.

[†] ETS, SAS Institute Inc., Cary, NC 27513. Email: `Xilong.Chen@sas.com`

[‡] Department of Finance, Kenan-Flagler Business School, and Department of Economics, University of North Carolina at Chapel Hill. Email: `eghysels@unc.edu`

[§] Department of Information and Decision Sciences, University of Illinois at Chicago. Email: `ffwang@uic.edu`

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 fundamentally different ways. The first approach is to estimate a horizon-specific model of the volatility, such as a weekly or monthly GARCH which can then be used to form direct predictions of volatility over the next week, month, etc. The second approach is to estimate a daily autoregressive volatility forecasting 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 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 **?**) 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 are quite straightforward: one focuses directly on multi-period forecasts, like in the direct approach, but one preserves the use of high frequency data. Neither the direct nor the iterated approach features such advantages combined.

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 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 large literature on volatility forecasting. These topics include - but are not limited to - iterated versus direct forecasting (as already noted), temporal aggregation, weak versus semi-strong GARCH, as well as various estimation procedures. Since there are quite a few papers written on these topics already it is 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 where one assumes an underlying data generating process (DGP) or some dynamic structure for the high frequency data and finally (3) HYBRID filter processes. The first class of processes - those that involve parameter-free HYBRID processes - relate to a flurry of recent papers, including Engle and Gallo (2006), de Vilder and Visser (2008), Visser (2011), Shephard and Sheppard (2010), Hansen, Huang, and Shek (2010) suggesting the use of (daily) realized volatilities, high-low range or realized kernels or generic realized measures.

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 high-frequency - that is 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 weights 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 recently by Malliavin and Mancino (2005).[1] 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 we introduce 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.

As far as empirical specifications go, we obtain some powerful findings that deviate substantially from the existing literature. GARCH and TGARCH models with daily data are dominated at all forecast horizons by models involving intra-daily models. Models featuring intra-daily asymmetries (news impact curves applied to intra-daily returns) dominate symmetric models up to weekly horizons, while the reverse is true for longer horizons, a finding also reported in Chen and Ghysels (2011). Models using daily realized volatility (or semi-variance versions) are less preferred than HYBRID

---

[1] The paper was brought to our attention by George Tauchen after we wrote a first draft of our paper and presented it at the FERM meeting in Taipei, June 2010.

involving intra-daily weighting scheme even for longer horizons.

The paper is structured as follows. Section 2 provides a general overview of the models, their estimation and the various classes of HYBRID processes involved. The section is deliberately non-technical. It is followed by a technical section diverging further on the various model specifications in section 3. Section 4 is devoted to the statistical properties of the HYBRID GARCH model when it is correctly specified, and defining the various parameter estimators and studying both their asymptotic and small sample properties (via simulation). Section 5 covers the empirical model specifications, with empirical findings appearing in Section 6. Section 7 concludes the paper. Some technical details are collected in an appendix, whereas all the proofs appear in a companion document Chen, Ghysels, and Wang (2010).

2 HYBRID Processes and Estimators: Overview

This section provides a general overview of the models, their estimation and the various classes of HYBRID processes involved. We deliberately compromise on technical details in order to provide the reader with a general overview of the paper's contributions. The subsequent sections of the paper will address in detail all the technicalities and provide the main results.

The volatility dynamics of a generic HYBRID GARCH model is as follows:

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma H_t \tag{2.1}$$

where H_t will be called a HYBRID process. When H_t is simply a daily squared return we have the volatility dynamics of a standard daily GARCH(1,1), or H_t a weekly squared return those of a standard weekly GARCH(1,1). However, what would happen if we want to attribute an individual weight to each of the five days in a week? In this case we might consider a process $H_t \equiv \sum_{j=0}^{4} \omega_j r_{t-j/5}^2$, where we use the notation $r_{t-j/5}$ to indicate intra-period returns - the daily observations of week t in this case.[2] This is an example of a parameter-driven HYBRID process $H_t \equiv H(\phi, \vec{r}_t)$ where $\vec{r}_t = (r_{t-1+1/m}, r_{t-1+2/m}, \ldots, r_{t-1/m}, r_t)^T$ is $\mathbb{R}^m-$valued random vector (in this case $m = 5$). In addition, the weights $(\omega_j(\phi), j = 0, \ldots, m-1)$ are governed by a low-dimensional parameter vector ϕ. One can think of at least two possibilities: (1) the weights are treated as additional parameters and estimated as such (with m small

---

[2] When days spill over into the previous week, we assume $r_{t-j/m} \equiv r_{t-1-(j-5)/m}$.

this is possible, but not as m gets large), or (2) anchor the weights ω_j to an underlying daily GARCH(1,1) in which case the parameters α, β and γ and the weights in ϕ are jointly related to the assumed daily DGP.

Parameter-free HYBRID Processes

The HYBRID process H_t may 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) dynamics. More recently, however, other purely data-driven examples of what we call generic HYBRID processes have been suggested. For example Engle and Gallo (2006), de Vilder and Visser (2008), Visser (2011), Shephard and Sheppard (2010), Hansen, Huang, and Shek (2010) suggest the use of (daily) realized volatilities, high-low range or realized kernels or generic realized measures. It is important to note that typically parameter-free HYBRID processes do not differentiate intra-period returns, i.e. an equal weighting scheme is supposed - although some kernel-weighting or pre-averaging may take place to eliminate so-called micro-structure noise. It should also be noted that typically an extra equation is added, namely consider the case of (daily) realized volatility RV_t :

$$\begin{aligned} V_{t+1|t} &= \alpha + \beta V_{t|t-1} + \gamma RV_t \\ RV_t &= a + bRV_{t-1} + e_t \end{aligned} \tag{2.2}$$

where

$$RV_t \equiv \sum_{j=1}^{m} r^2_{t-(j-1)/m}$$

is the realized volatility computed as the sum of intra-daily squared returns. Intra-daily returns are characterized as $r_{t-(j-1)/m}$ for $j = 1, \ldots, m$, where the latter is the number of intra-daily observations. The extra equation in (2.2) is added for the purpose of multi-period ahead forecasting. On this topic, the system of equations in (2.2) de facto results in multiple step predictions involving HYBRID processes of the type (with an abuse of notation - for a multiple horizon h with corresponding low frequency t) $H_t \equiv \sum_{j=0}^{h} \omega_j RV_{t-j/h}$, where the weights ω_j relate to γ and b in (2.2). What sets HYBRID processes apart is that we refrain from adding an additional equation, but rather let the weighting scheme that determines H_t handle the mapping between forecast horizon and the frequency of conditioning information - high frequency returns. Changing either the forecast horizon or the sampling frequency will result in different HYBRID processes.

While these may not be obviously related to each other, there is one case where they are, and this is discussed next.

Structural HYBRID Processes

Suppose we consider a daily weak GARCH(1,1), as defined by Drost and Nijman (1993), then the implied weekly prediction, using past daily returns is:

$$V_{t+1|t} = \alpha_m + \beta_m V_{t|t-1} + \gamma_m \sum_{j=0}^{m-1} \beta_m^{j/m} r_{t-j/m}^2, \quad t \in \mathbb{Z} \tag{2.3}$$

with $m = 5$, and where α_m, β_m and γ_m depend on the daily GARCH(1,1) parameters α_1, β_1 and γ_1 (see more details later in equation (3.4)). Note that all the parameters are driven by the daily parameters. Therefore, while the HYBRID process is parameter-driven it is in principle an integral part of the volatility dynamics and $H(\phi, \vec{r}_t)$ in (2.1) does not involve stand-alone parameters ϕ. This will have consequences when we elaborate on the estimation of HYBRID GARCH models.

HYBRID Filtering Processes

Here the HYBRID process $H(\phi, \vec{r}_t)$ in (2.1) involves parameters that are *not* explicitly related to α, β and γ appearing in (2.1). One can view this as a GARCH model driven by a filtered high frequency process, where the filter weights - (hyper-)parameterized by ϕ - are estimated jointly with the volatility dynamics parameters. The choice of parameterizations is inspired by Chen and Ghysels (2011). The commonly used specifications are exponential, beta, linear, hyperbolic, and geometric weights. This approach has implications too as far as estimation is concerned. Unlike the structural HYBRID case, we now can consider likelihood-based methods, although the regularity conditions required are novel and more involved as those of the usual QMLE approach to GARCH estimation.

Asymmetries - Nonlinearities

Is the response of future volatility to past return symmetric? It was noted earlier that Malliavin and Mancino (2005) advocated the use of linear intra-period weighting in the context of continuous time diffusions. Such weighting schemes are inherently symmetric. Perhaps that should not be the case. In particular, Chen and Ghysels

(2011) examine whether the sign and magnitude of intra-daily returns have impact on expected volatility the next day or over longer future horizons. They revisit the concept of news impact curves introduced by Engle and Ng (1993). Overall, they find that moderately good (intra-daily) news reduces volatility (the next day), while both very good news (unusual high intra-daily positive returns) and bad news (negative returns) increase volatility, with the latter having a more severe impact.

So far we have done the same as Malliavin and Mancino (2005) in terms of the formulation of HYBRID processes in the context of discrete time GARCH dynamics. At this stage, we start to deviate from the linear projection paradigm and continue the logic of GARCH modeling. Namely, we consider HYBRID GARCH models that feature intra-daily news impact curves - similar to the framework of Chen and Ghysels (2011), except that the latter use a MIDAS regression format. The HYBRID processes we consider are of the following type:

$$H(\phi, \vec{r}_t) = \sum_{j=0}^{m-1} \Psi_j(\phi) NIC(\phi, r_{t-j/m}), \quad \sum_{j=0}^{m-1} \Psi_j(\phi) = 1 \tag{2.4}$$

where $NIC(\phi, \cdot)$ stands for a high frequency data news impact curve. The parameter vector ϕ determines the weights as well as the parameters that determine the news impact curve. Regarding the specification of the latter, we consider the parametric news impact curves studied in Chen and Ghysels (2011), namely:

$$NIC(\phi, r) = \gamma(r - \delta)^2 \tag{2.5}$$

$$NIC(\phi, r) = \gamma r^2 + \delta r^2 \mathbf{1}_{r<0} \tag{2.6}$$

with γ and δ the parameters that are in the parameter vector ϕ. In principle we could also consider more general specifications - notably involving non-linearities. We refrain from doing so, although the theoretical analysis in this paper will cover non-linear HYBRID processes as well.

Estimation Procedures

Various estimation procedures will be considered - some tailored to specific cases of HYBRID processes. Let us first collect all the parameters of the model appearing in (2.1) in a parameter vector called $\theta \in \Theta$, with the (pseudo-) true parameter being denoted θ_0. One has to keep in mind that specific cases - notably involving structural HYBRID processes - may involve constraints across the parameters in (2.1) or the filtering weights of the HYBRID process may also be hyper-parameterized, so that the dimension of θ (denoted as d) depends on the specific circumstances considered. For this generic setting we have the following estimator: $\tilde{\theta}_T^{mdrv} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} (RV_t - V_{t|t-1}(\theta))^2$ where $\mathcal{C}$ is a convex compact subset of Θ such that θ_0 is in the interior of $\mathcal{C}$. This minimum distance estimator involves observations about *RV*, realized volatility or possibly a realized measure that corrects for microstructure effects etc. This estimator applies to volatility models involving all possible HYBRID processes, including structural ones for which a weak GARCH assumption is required. Note that this means that $V_{t|t-1}(\theta)$ in the above estimator is based on a best *linear predictor,* not the conditional variance - a technical issue that will be discussed in the next section.

A companion estimation procedure involves a single squared return process, namely, $\tilde{\theta}_T^{mdr2} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} (R_t^2 - V_{t|t-1}(\theta))^2$. It has a likelihood-based version: $\tilde{\theta}_T^{lhr2} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left(\log V_{t|t-1}(\theta) + R_t^2 / V_{t|t-1}(\theta) \right)$, which requires far more stringent in terms of regularity conditions, notably because $V_{t|t-1}(\theta)$ is a conditional variance, and in fact does not apply to all types of HYBRID processes - in particular structural ones. The estimator $\tilde{\theta}_T^{lhr2}$ is reminiscent of QMLE estimators for semi-strong GARCH models - yet the mixed data frequencies add an extra layer of complexity discussed later in the paper. One can again replace daily squared returns by, say *RV* and consider the following estimator: $\tilde{\theta}_T^{lhrv} = \arg\min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left(\log V_{t|t-1}(\theta) + RV_t / V_{t|t-1}(\theta) \right)$. The choice of R^2 versus RV in $\tilde{\theta}_T^{mdr2}$ versus $\tilde{\theta}_T^{mdrv}$ and $\tilde{\theta}_T^{lhr2}$ versus $\tilde{\theta}_T^{lhrv}$ has efficiency implications that will be discussed as well.

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 also consider the following model

$$RV_{t+1} = \sigma_{t+1|t}^2 \eta_{t+1} \tag{2.7}$$

where η_{t+1} is independent and identically distributed with mean 1, and $\sigma_{t+1|t}^2$ is the conditional expectation of RV_{t+1} given information at time t (in the next section we will explore in more detail the relation between $\sigma_{t+1|t}^2$ and RV_{t+1}). Suppose the cumulative distribution function of η is F. The choice of F could be a unit exponential (see Engle (2002)), or a Gamma distribution as suggested in Engle and Gallo (2006), or a mixture of two gamma distributions of Lanne (2006). The resulting class of estimators will be denoted by $\tilde{\theta}_T^{mem}$.

3 HYBRID Processes

We will start with some notation. Suppose the underlying probability space is $(\Omega, \mathcal{F}, P)$. Let $\|X\|_p = (E|X|^p)^{1/p}$ for $X \in L^p(\Omega, \mathcal{F}, P)$ and $p < \infty$. $\|A\| = \sqrt{tr(A^T A)}$ for $A \in \mathbb{R}^{n \times n}$ or $A \in \mathbb{R}^{n \times 1}$ and $n \geq 1$. For $X \in L^2(\Omega, \mathcal{F}, P)$ and $\mathcal{I}$ a closed subspace of $L^2(\Omega, \mathcal{F}, P)$, $P_l(X|\mathcal{I})$ indicates the *orthogonal* projection of X onto $\mathcal{I}$. We write $A > 0$ if A is a positive definite matrix, and $A \geq 0$ if A is positive semi-definite. If A is finite element by element, then we write $A < \infty$. Moreover, to emphasize the role of ϕ in the process $H(\phi, \vec{r}_t)$ we will use the notation $H_t(\phi) \equiv H(\phi, \vec{r}_t)$. The notation related to derivatives is shown at the beginning of Appendix A.

3.1 Three processes of interest

Suppose the underlying (log) price process p_s is a semimartingale defined on $(\Omega, \mathcal{F}, P)$. r_s is the return sampled at frequency m, i.e., $r_s = p_s - p_{s-1/m}$. We are interested in the next-period volatility forecast based on the available discretely-sampled returns, denoted by $\sigma^2_{t+1|t}$. It is defined as the *orthogonal* projection of $RV_{t+1} = \sum_{j=0}^{m-1} r^2_{t+1-j/m}$ onto $\mathcal{I}_t$, which is a closed subspace of $L^2(\Omega, \mathcal{F}, P)$ and represents the information up to time t. In other words, $\sigma^2_{t+1|t} = P_l(RV_{t+1}|\mathcal{I}_t)$. We therefore implicitly assume that $RV_t \in L^2(\Omega, \mathcal{F}, P)$, or the return has finite fourth moment.

Denote by $[p, p]_s$ the quadratic variation of $\{p_s\}$. The predicted increment in quadratic variation is expressed as $E_t([p, p]_{t+1} - [p, p]_t) \equiv E([p, p]_{t+1} - [p, p]_t|\sigma(p_s, s \leq t))$. We make a distinction between three objects: (1) $V_{t+1|t}$, (2) $\sigma^2_{t+1|t}$, and (3) $E_t([p, p]_{t+1} - [p, p]_t)$. The latter two are population quantities, while (1) pertains to the specification of the HYBRID GARCH model. Because the first two are formulated using the available (high frequency) returns, they are not necessarily linked with an explicit continuous-time/discrete-time DGP.

Various model specifications can be considered for the HYBRID GARCH, but this is not our concern at this point. Instead, we are interested in the relation between $\sigma^2_{t+1|t}$ and $V_{t+1|t}$. Moreover, we will also examine how predicted increments in quadratic variation relate to the HYBRID GARCH model-based predictions $V_{t+1|t}$. These relationships can only be well understood when (1) we impose a structure on the underlying returns and (2) we explicitly link the HYBRID GARCH to the DGP or at least some dynamic structure for high frequency returns.

At the outset it should also be noted that $V_{t+1|t}$ inherits the properties of the HYBRID process H_t and vice versa. Hence, we will interchangeably talk about features of HYBRID process H_t and features of $V_{t+1|t}$.

3.2 HYBRID Filtering Processes

Suppose that the available returns are sampled at frequency m, and they are expressed as r_s where s is of the form $t+k/m$ for some $t \in \mathbb{Z}$ and $k = 0, 1, 2 \ldots, m-1$. In general the structure of $\sigma^2_{t+1|t}$ is not tractable due to the ignorance of r_s and $\mathcal{I}_t$. Therefore to justify the approximation of $\sigma^2_{t+1|t}$ with $V_{t+1|t}$, we consider two specifications for $\sigma^2_{t+1|t}$, and call them Scenarios 1 and 2 appearing in respectively Assumptions 3.1 and 3.2. The former views $\sigma^2_{t+1|t}$ as the best linear predictor. Alternatively, we also consider a more general situation in Scenario 2, where $\sigma^2_{t+1|t}$ is a conditional variance. Suppose the return process $\{r_s\}$ satisfy Assumption A.1.[3] Denote by $\mathcal{F}^{t+m}_{t-m}$ the sigma field generated by the high frequency returns from $t-m$ to $t+m$, i.e., $\sigma(r_s, t-m-1+1/m \leq s \leq t+m)$.

Assumption 3.1 (Scenario 1). *$\mathcal{I}_t = \mathcal{L}_t$, the closed span of $\{1, r_{t-k/m}, r^2_{t-k/m}; k = 0, 1, 2 \ldots\}$ and $P_l(r_s|\mathcal{L}_{s-1/m}) = 0$. Therefore $\sigma^2_{t+1|t}$ is the best linear predictor.*

Assumption 3.2 (Scenario 2). *$\mathcal{I}_t = \mathcal{F}^t_{-\infty}$, the sigma field generated by the high frequency returns up to time t and $P_l(r_s|\mathcal{F}^{s-1/m}_{-\infty}) = 0$. The prediction equations therefore indicate that $E(r_s|\mathcal{F}^{s-1/m}_{-\infty}) = 0$ and $E(RV_{t+1}|\mathcal{F}^t_{-\infty}) = E(R^2_{t+1}|\mathcal{F}^t_{-\infty}) = \sigma^2_{t+1|t}$, where $R_{t+1} = \sum_{j=0}^{m-1} r_{t+1-j/m}$.*

Since we use $V_{t+1|t}$ driven by $H(\phi, \vec{r}_t)$ to mimic the dynamics of $\sigma^2_{t+1|t}$, H_t is required to satisfy Assumption A.4 in both scenarios. Particularly in Scenario 1, H_t is also assumed to be a weighted sum of 1, the intermediate returns and squared returns from period $t-1$ to t. Assumption A.4 essentially guarantees that the HYBRID process is non-negative and satisfies measurability and identifiability when it comes to parameter estimation via of extremum estimators. A necessary condition is that the dimension of ϕ is not larger than m. A more detailed discussion on Assumption A.4 is available in Chen, Ghysels, and Wang (2010). The HYBRID processes that satisfy Assumption A.4 are also referred to as *HYBRID filtering processes.*

In both scenarios we can hyper-parameterize the filter weights, namely:

$$H_t(\phi) = \sum_{j=0}^{m-1} \Psi_j(\phi) r^2_{t-j/m}, \quad \sum_{j=0}^{m-1} \Psi_j(\phi) = 1 \tag{3.1}$$

---

[3] For convenience we collected all the regularity conditions in Appendix A.

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

When we consider HYBRID processes driven by news impact curves as in equation (2.4) we should note that the HYBRID process with $NIC(\phi_2, \cdot)$ given by (2.5) is specified under Scenario 1, while the HYBRID process constructed using news impact curve (2.6) is a special case of Scenario 2. To ensure the HYBRID process (2.4) meets Assumption A.4, the weight functions need to satisfy some additional conditions that appear in Assumption A.6.

3.3 Structural HYBRID Processes

When the underlying high frequency returns $\{r_s\}$ follow a weak GARCH(1,1) of Drost and Nijman (1993), the dynamics of $\sigma^2_{t+1|t}$ can be fully specified. In other words, with proper parameterization, the HYBRID GARCH process $V_{t+1|t}$ and $\sigma^2_{t+1|t}$ coincide. The HYBRID processes are therefore called *structural HYBRID processes*.

Suppose that $r_{s+1/m}$ is orthogonal to $\mathcal{L}_s$, i.e., $P_l(r_{s+1/m}|\mathcal{L}_s) = 0$, and $\sigma^2_{s+1/m|s}$ defined as the orthogonal projection of $r^2_{s+1/m}$ onto $\mathcal{L}_s$ satisfies

$$\sigma^2_{s+1/m|s} = a + b\sigma^2_{s|s-1/m} + cr^2_s. \tag{3.2}$$

With some algebra one obtains $\sigma^2_{s+k/m|s} = a\left(1-(b+c)^{k-1}\right)/(1-(b+c)) + (b+c)^{k-1}\sigma^2_{s+1/m|s}$ for $k \in \mathbb{Z}^+$. Consequently, the total volatility over the period $(t, t+1]$, denoted by $V_{t+1|t} \equiv \sum_{k=1}^m \sigma^2_{t+k/m|t} \equiv \sigma^2_{t+1|t}$, can be characterized by the following GARCH-type of equation:

$$V_{t+1|t} = \alpha_m + \beta_m V_{t|t-1} + \gamma_m \sum_{j=0}^{m-1} \beta_m^{j/m} r^2_{t-j/m} \tag{3.3}$$

where

$$\alpha_m = a\frac{1-b^m}{1-b}\frac{m(1-b)-cd_m}{1-(b+c)}, \quad \beta_m = b^m, \quad \gamma_m = cd_m \tag{3.4}$$

and $d_m = (1-(b+c)^m)/(1-(b+c))$. Clearly, (3.3) is of the form (2.1) with $H_t =$

$\sum_{j=0}^{m-1} \beta_m^{j/m} r^2_{t-j/m}$, and H_t is referred to as *structural HYBRID process*. The distinct difference between structural HYBRID and HYBRID filtering processes is that the underlying return process follows a weak GARCH(1,1) for the structural HYBRID. The HYBRID filtering process can be viewed as an extension of the structural HYBRID process by allowing for more flexible return dynamics.

It is worth noting that the structural HYBRID model allows the parameters evaluated under different sampling frequencies to be linked to each other explicitly, as is evident from (3.4). A direct implication of the relationship (3.4) 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.

3.4 Diffusions, Jumps and HYBRID Processes

The HYBRID processes discussed so far pertained to discretely sampled returns at different frequencies. We turn our attention now to HYBRID processes structurally linked to continuous time processes. In addition, we also examine the possible structural HYBRID interpretation of the purely RV-driven HYBRID GARCH process, i.e., $H(\phi, \vec{r}_t) = RV_t$ (see the first equation in (2.2)). Moreover, we will characterize how the presence of jumps will have an impact on the discrete-time HYBRID process.

Inspired by Drost and Werker (1996), we consider a continuous-time GARCH model as the DGP, namely,

$$\begin{aligned} dp_t &= \sigma_t dL_t \\ d\sigma_t^2 &= \theta(\omega - \sigma_t^2)dt + \sqrt{2\lambda\theta}\sigma_t^2 dB_t \\ L_t &= \sqrt{1-\eta}W_t + \sqrt{\eta}N_t \end{aligned} \tag{3.5}$$

with $\theta > 0$, $\omega > 0$, $\lambda \in (0,1)$, and $\eta \in [0,1]$. 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. Note that $EL_t = EL_t^3 = 0$, $EL_t^2 = t$, and $EL_t^4 = 3t^2 + 3t\eta^2/\zeta$. The discretely sampled returns $r_s = p_s - p_{s-1/m}$ from (3.5) follows a weak GARCH(1,1) appearing in equation (3.2) as discussed in Drost and Werker (1996). The relation between $(\theta, \omega, \lambda, v_L^*)$ in (3.5) and (a, b, c, k) in (3.2) is stated in Drost and Werker (1996), where $v_L^* \equiv EL_1^4 - 3 = 3\eta^2/\zeta$ and k is the kurtosis of r_s.[4]

We turn our attention now to some prediction formulas associated with this

---

[4] Note that (a, b, c, k) and r_s depend on m.

framework. Note that the Quadratic Variation (QV) of p is $[p,p]_t = (1-\eta)\int_0^t \sigma_s^2 ds + \eta \int_0^t \int_{-\infty}^{\infty} \sigma_s^2 y^2 J_N(ds, dy)$. We start with examining how $V_{t+1|t} \equiv \sigma_{t+1|t}^2$ viewed as linear projection onto $\mathcal{L}_t$ relates to prediction of $E_t([p,p]_{t+1} - [p,p]_t)$. In this subsection, we write $V_{t+1|t}$ as $V_{t+1|t}^{(m)}$ to emphasize the role of sampling frequency m.

On the one hand, using a continuous time filtration, the forecast of the increment of QV is as follows:

$$\begin{aligned} E_t([p,p]_{t+1} - [p,p]_t) &= (1-\eta)\int_t^{t+1} E_t(\sigma_s^2) ds + \eta \int_t^{t+1} \int_{-\infty}^{\infty} E_t(\sigma_s^2) y^2 \zeta f(dy) ds \\ &= \omega\left(1 - \theta^{-1}(1-e^{-\theta})\right) + \theta^{-1}(1-e^{-\theta})\sigma_t^2. \end{aligned} \tag{3.6}$$

On the other hand, the forecast using $\mathcal{L}_t$ yields the HYBRID GARCH equation appearing in equation (3.3). What we will show is that, although $RV_{t+1} = \sum_{j=0}^{m-1} r_{t+1-j/m}^2$ 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 $E_t([p,p]_{t+1} - [p,p]_t)$. Using the relation between $(\theta, \omega, \lambda, v_L^*)$ and (a, b, c, k) stated in Drost and Werker (1996), we have the following proposition:

Proposition 3.1. *When the Lévy measure associated with the jump process features excess kurtosis, i.e.* $v_L^* > 0$,

$$\lim_{m\to\infty} \alpha_m = \omega\left(1 - e^{-\theta(1+\phi)}\right)\left(1 - \frac{\phi}{1+\phi}\theta^{-1}(1-e^{-\theta})\right) \tag{3.7}$$

$$\lim_{m\to\infty} \beta_m = e^{-\theta(1+\phi)} \tag{3.8}$$

$$\lim_{m\to\infty} \gamma_m = (1-e^{-\theta})\phi \tag{3.9}$$

$$\lim_{m\to\infty} \sum_{j=0}^{m-1} \beta_m^{j/m} r_{t-j/m}^2 = \int_{(t-1,t]} e^{-\theta(1+\phi)(t-s)} d[p,p]_s \quad \textit{in probability} \tag{3.10}$$

where $\phi = \sqrt{1 + 2\lambda/(\theta v_L^*)} - 1$. *In contrast, when there are no jumps in the price process, i.e.* $v_L^* = 0$,

$$\lim_{m\to\infty} \alpha_m = \omega\left(1 - \theta^{-1}(1-e^{-\theta})\right), \quad \lim_{m\to\infty} \beta_m = 0, \quad \lim_{m\to\infty} \frac{\gamma_m}{\sqrt{m}} = \sqrt{\lambda/\theta}(1-e^{-\theta}).$$

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.

The proof - as well as all subsequent ones - are omitted here; they appear in Chen, Ghysels, and Wang (2010). Comparing equations (3.3) and (3.6), we note from

Proposition 3.1 that $V^{(m)}_{t+1|t}$ is a consistent estimate of $E_t([p,p]_{t+1} - [p,p]_t)$ when there are no jumps in the price process. In contrast, when jumps are present $V^{(m)}_{t+1|t}$ does not provide a consistent estimate of $E_t([p,p]_{t+1} - [p,p]_t)$ because the limit of $V^{(m)}_{t+1|t}$ involves a whole sample path of volatility up to time t. This is stated formally in the following corollary:

Corollary 3.1. *Given a continuous time GARCH (3.5) as the DGP, the process* $\{V^{(m)}_{t+1|t}, t\}_{m\geq 1}$ *defined by equation (3.3) converges to* $\{E_t([p,p]_{t+1} - [p,p]_t), t\}$ *uniformly on compact sets in probability if and only if there are no jumps in the price process.*

Note that without jumps, the HYBRID GARCH process still involves intra-period weighted returns, namely equation (3.3) has intra-period weights that are powers of β_m. Furthermore, it follows from the proof of Proposition 3.1 that what drives the HYBRID process as $m \to \infty$, is the instantaneous volatility σ_t^2, not the integrated process estimated by RV. The instantaneous volatility σ_t^2 can be consistently estimated by that very same intra-period weighted sum $mc\sum_{j=0}^{m-1} b^j r^2_{t-j/m}$. Put differently, we can view the HYBRID process as a spot volatility estimator which 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.

To conclude it should also be noted that one could think of continuous time limits in the HYBRID filtering context and potentially link them to $E_t([p,p]_{t+1} - [p,p]_t)$. In the above discussions we relied on the approach of Drost and Werker (1996) using exact discretization limits - which is compatible with structural HYBRID processes. 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.

4 Estimation

We study the statistical properties of the HYBRID GARCH model-based parameter estimation in this section. The structural HYBRID can be viewed as a special case of the HYBRID filtering processes. Therefore, we will focus on the latter and the results derived for the HYBRID filtering processes will carry over to the structural HYBRID accordingly. We will work exclusively with returns sampled at fixed frequency without referring to an explicit DGP and make the assumption that the returns are strictly stationary and ergodic (formally stated as Assumption A.2 in Appendix A). Throughout this section we will assume there are no model specification errors (the analysis of potentially misspecified models is covered in Chen, Ghysels, and Wang (2010)).

4.1 Estimation under Scenario 1

We use $V_{t|t-1}(\theta)$ solved from equation (2.1) to approximate $\sigma^2_{t|t-1}$. The distance between the two time series $V(\theta) \equiv (V_{t|t-1}(\theta), t \in \mathbb{N})$ and $\sigma^2 \equiv \{\sigma^2_{t|t-1}, t \in \mathbb{N}\}$ is defined as $d(V, \sigma^2) = \sum_{t=1}^{\infty} 2^{-t} \min(\|V_{t|t-1} - \sigma^2_{t|t-1}\|_2, 1)$. Note that $\|RV_t - V_{t|t-1}(\theta)\|_2 - \|RV_t - \sigma^2_{t|t-1}\|_2 = \|V_{t|t-1}(\theta) - \sigma^2_{t|t-1}\|_2$ due to Assumption A.1. $0 \leq d(RV, V(\theta)) - d(RV, \sigma^2) \leq d(V(\theta), \sigma^2)$ where $RV = (RV_t, t \in \mathbb{N})$. $d(V(\theta), \sigma^2)$ hence measures ignorance of the true dynamics. $d(V(\theta), \sigma^2)$ (for a suitable choice of θ) is 0 when equation (2.1) *correctly* describes the dynamics of $\sigma^2_{t|t-1}$.

Given H that satisfies Assumption A.4, $d(V(\theta), \sigma^2)$ has a minimum over $\mathcal{C}$, say at $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \phi_0) \in \mathcal{C}$, where $\mathcal{C}$ is a convex compact subset of the interior of $\Theta \equiv \{(\alpha, \beta, \gamma, \phi) : \alpha > 0, 0 < \beta < 1, \gamma > 0, \phi \in \Phi\}$ and Φ is a connected set which collects all the possible values of ϕ's such that $H_t(\phi)$ meets Assumption A.4. Assume $d(V(\theta_0), \sigma^2) = 0$. A natural estimator of θ_0 under Scenario 1, given $\{r_{1/m}, \ldots, r_1, \ldots, r_{T-1+1/m}, \ldots, r_T\}$, is the minimizer of $\min_{\theta \in \mathcal{C}} T^{-1} \sum_{t=1}^{T} \left(RV_t - \tilde{V}_t(\theta)\right)^2$ and $\tilde{V}_t$ is defined recursively by

$$\tilde{V}_t(\theta) = \alpha + \beta \tilde{V}_{t-1}(\theta) + \gamma H_{t-1}(\phi), t \geq 1 \quad \text{and} \quad \tilde{V}_0 = \tilde{v} \tag{4.1}$$

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*. Note that θ_0 is identifiably unique.[5] Namely, Chen, Ghysels, and Wang (2010) show that: letting $\varepsilon_t(\theta) = RV_t - V_{t|t-1}(\theta)$,

Proposition 4.1 (Scenario 1)**.** *Suppose that Assumptions A.1 and A.2 hold.* $\theta_0 = \arg\min_{\theta \in \mathcal{C}} E\varepsilon_t(\theta)^2$.

We therefore have the following:

Theorem 4.1 (Scenario 1)**.** *Under Assumptions A.1 and A.2,*
(1) $\tilde{\theta}_T^{mdrv}$ *is identifiably unique and it is a strongly consistent estimator of* θ_0.

---

[5] See Gallant and White (1988) for the definition of identifiable uniqueness.

(2) Additionally under Assumption A.3 $\lim_{T\to\infty}$ *var* $\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\nabla\varepsilon_t(\theta_0)\right)$ *exists and is finite, denoted by* Ω^{mdrv}.

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

The existence of Ω^{mdrv} and the asymptotic normality follow from the fact that $\varepsilon_t\partial_k\varepsilon_t$ is *near epoch dependent* (NED) on the underlying return process under suitable moment conditions, and is therefore a *mixingale* when the return process is $\alpha-$mixing (see Assumption A.3). It should also be noted that the size of $\alpha-$mixing is $-v_2/(v_2+2)$ (see Assumption A.3). It is weaker than the size required in Theorem 5.7 of Gallant and White (1988), i.e., $-2v_2/(v_2+2)$, and Goncalves and White (2004) as well which requires a size of $-\delta v_2/(v_2+2)$ for some $\delta > 2$, although in a different context.[6]

4.2 Estimation under Scenario 2

We now consider situations where the HYBRID GARCH model produces conditional variance predictions. In such a case we are at liberty to consider both minimum distance estimators as in Section 4.1, and quasi-maximum likelihood estimators that are standard in the GARCH literature. More specifically, let us consider the following estimators:

$$\tilde{\theta}_T^{mdrv} = \arg\min_{\theta\in\mathcal{C}} \frac{1}{T}\sum_{t=1}^{T}\left(RV_t - \tilde{V}_t(\theta)\right)^2 \tag{4.2}$$

$$\tilde{\theta}_T^{mdr2} = \arg\min_{\theta\in\mathcal{C}} \frac{1}{T}\sum_{t=1}^{T}\left(R_t^2 - \tilde{V}_t(\theta)\right)^2 \tag{4.3}$$

$$\tilde{\theta}_T^{lhr2} = \arg\min_{\theta\in\mathcal{C}} \frac{1}{T}\sum_{t=1}^{T}\left(\log\tilde{V}_t(\theta) + \frac{R_t^2}{\tilde{V}_t(\theta)}\right) \tag{4.4}$$

$$\tilde{\theta}_T^{lhrv} = \arg\min_{\theta\in\mathcal{C}} \frac{1}{T}\sum_{t=1}^{T}\left(\log\tilde{V}_t(\theta) + \frac{RV_t}{\tilde{V}_t(\theta)}\right) \tag{4.5}$$

with $\tilde{V}_t$ defined in (4.1). We also consider estimator derived using Multiplicative Error Model which shares some similarities with the likelihood-RV-based estimator $\tilde{\theta}_T^{lhrv}$.

---

[6] See page 24 of Gallant and White (1988) for the definition of size.

4.2.1 Minimum Distance and Quasi-Likelihood Estimators

We start by extending Proposition 4.1 to the case of Scenario 2: letting $e_t(\theta) = R_t^2 - V_{t|t-1}(\theta)$,

Proposition 4.2. *Suppose that Assumptions A.1, A.2, and A.5(2) hold. Under Scenario 2,* $\theta_0 = \arg\min_{\theta\in\mathcal{C}} E\left(\varepsilon_t(\theta)\right)^2 = \arg\min_{\theta\in\mathcal{C}} E\left(e_t(\theta)\right)^2$.

Note that $\varepsilon_t(\theta_0)\partial_i\varepsilon_t(\theta_0)$ and $e_t(\theta_0)\partial_i e_t(\theta_0)$ are martingale difference sequences. Ω^{mdrv} defined in Theorem 4.1 becomes $E[(RV_t - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)']$. Define Ω^{mdr2} as $E[(R_t^2 - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)']$. Both Ω^{mdrv} and Ω^{mdr2} are finite and positive definite under suitable regularity conditions. Consider $\tilde\theta_T^{mdrv}$ and $\tilde\theta_T^{mdr2}$ defined in (4.2) and (4.3).

Theorem 4.2 (Scenario 2)**.** *Under Assumptions A.1, A.2, and A.5(2),*
(1) $\tilde\theta_T^{mdrv}$, $\tilde\theta_T^{mdr2}$ *are identifiably unique and they converge to* θ_0 *a.s.*
(2) Further assume that $Er^8 < \infty$ *and Assumption A.5(3) holds,* $\sqrt{T}(\tilde\theta_T^{mdrv} - \theta_0) \Longrightarrow N(0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1})$ *and* $\sqrt{T}(\tilde\theta_T^{mdr2} - \theta_0) \Longrightarrow N(0, (\Sigma^{md})^{-1}\Omega^{mdr2}(\Sigma^{md})^{-1})$.

Moreover, we have the following result that ties the QMLE estimators:

Proposition 4.3 (Scenario 2)**.** *Suppose* $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^2 < \infty$. *Under Assumptions A.1 and A.2,* $\theta_0 = \arg\min_{\theta\in\mathcal{C}} E\left(\log V_{t|t-1}(\theta) + RV_t/V_{t|t-1}(\theta)\right) = \arg\min_{\theta\in\mathcal{C}} E\left(\log V_{t|t-1}(\theta) + R_t^2/V_{t|t-1}(\theta)\right)$.

Therefore θ_0 can also be estimated by $\tilde\theta_T^{lhrv}$ in (4.4) or $\tilde\theta_T^{lhr2}$ in (4.5). Define

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

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

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

Theorem 4.3 (Scenario 2)**.** *Suppose* $E(\sup_{\phi\in\overline{\Phi^0}} H(\phi,\vec{r}_t))^2 < \infty$ *and Assumptions A.1, and A.2 hold.*
(1) $\tilde\theta_T^{lhrv}$, $\tilde\theta_T^{lhr2}$ *are identifiably unique and they converge to* θ_0 *a.s.*
(2) Suppose $E(r^{4+v}) < \infty$ *for some* $v > 0$, *and the following holds*

$$|\partial_\phi H(\phi,\vec{x})/H(\phi,\vec{x})| \le g(\phi), \quad |\partial_\phi^2 H(\phi,\vec{x})/H(\phi,\vec{x})| \le g(\phi) \quad \forall\vec{x}\in\mathbb{R}^m, \phi\in\Phi \tag{4.6}$$

for some real-valued function g which is continuous in ϕ. $\sqrt{T}(\tilde{\theta}_T^{lhrv} - \theta_0) \Rightarrow N\left(0, (\Sigma^{lh})^{-1}\Omega^{lhrv}(\Sigma^{lh})^{-1}\right)$*, and* $\sqrt{T}(\tilde{\theta}_T^{lhr2} - \theta_0) \Rightarrow N\left(0, (\Sigma^{lh})^{-1}\Omega^{lhr2}(\Sigma^{lh})^{-1}\right)$*.*

The likelihood estimation considered here is slightly different from what is discussed in the literature. First of all, $\sigma_{t|t-1}^2$ is studied in $L^2(\Omega, \mathcal{F}, P)$ instead of $L^1(\Omega, \mathcal{F}, P)$. Secondly, the objective function appearing in (4.4) is not the joint quasi-log-likelihood function (modulo a constant) of $\{R_1, R_2, R_3, \ldots, R_T\}$. Instead of conditioning on R_1, R_2, $\ldots$, 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 time $t-1$.

It should also be noted that the discussion can be extended to strictly periodically stationary and periodically ergodic time series as well.[7] This is because the proofs only require $\vec{r}_t$ to be strictly stationary ergodic.

Note that we have four estimators of θ_0: $\tilde{\theta}_T^{mdr2}$, $\tilde{\theta}_T^{mdrv}$, $\tilde{\theta}_T^{lhr2}$ 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 R^2-based estimator and RV-based estimator (i.e., $\tilde{\theta}_T^{mdr2}$ v.s. $\tilde{\theta}_T^{mdrv}$, and $\tilde{\theta}_T^{lhr2}$ v.s. $\tilde{\theta}_T^{lhrv}$), because the sign of $E_{t-1}\left[(R_t^2 - V_{t|t-1}(\theta_0))^2 - (RV_t - V_{t|t-1}(\theta_0))^2\right] = E_{t-1}(R_t^4 - RV_t^2)$ is unclear for an arbitrary return process. Next we consider a special case.

Corollary 4.1. *Suppose the DGP is a semi-strong GARCH(1,1) and* $E(r_s^3|F_{s-1/m}) = 0$*. Then* $E_{t-1}(R_t^4 - RV_t^2) > 0$*. Under the assumptions in Theorems 4.2 and 4.3,* $\tilde{\theta}_T^{mdrv}$ *(or* $\tilde{\theta}_T^{lhrv}$*) has a smaller asymptotic variance than* $\tilde{\theta}_T^{mdr2}$ *(or* $\tilde{\theta}_T^{lhr2}$*).*

4.2.2 Multiplicative Error Models

We turn our attention now to the Multiplicative Error Model of Engle (2002) appearing in (2.7). Denote the solution to equation (2.1) as $V_{t+1|t}(\theta)$. Suppose that $\theta_0 \in \Theta$ is such that $\sigma_{t+1|t}^2 = V_{t+1|t}(\theta_0)$. Therefore, we can use the marginal distribution of η appearing in (2.7) to formulate the estimation of θ_0. The estimator is then denoted by $\tilde{\theta}_T^{mem}$.

Suppose that the appropriate probability distribution for the error term is a Gamma distribution. In other words, the conditional density of RV_{t+1} is $f(RV_{t+1}|\mathcal{F}_{-\infty}^t) = \Gamma(g)^{-1}g^g RV_{t+1}^{g-1}(\sigma_{t+1|t}^2)^{-g}\exp(-gRV_{t+1}/\sigma_{t+1|t}^2)$. Hence $E(RV_{t+1}|\mathcal{F}_{-\infty}^t) = \sigma_{t+1|t}^2$, and $Var(RV_{t+1}|\mathcal{F}_{-\infty}^t) = \sigma_{t+1|t}^4/g$. The parameter space becomes $\Theta \times \{g > 0\}$. As pointed out by Engle and Gallo (2006) and Cipollini, Engle, and Gallo (2006), the estimation of θ_0 and g are asymptotically independent. The *point* estimation $\tilde{\theta}_T^{mem}$ of θ_0 is then

---

[7] See Aknouche and Bibi (2009) for the definition.

same as $\tilde{\theta}_T^{lhrv}$. It follows from the proof of Theorem 4.3 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 density can not be verified, one can consider quasi-likelihood estimation which will yield the same asymptotic result as $\tilde{\theta}_T^{lhrv}$. But since the innovation η is independent, Ω^{lhrv} in Theorem 4.3 becomes $E(RV_t^2/V_{t|t-1}^2(\theta_0) - 1)\Sigma^{lh}$ and hence we do not need the moment condition $Er^{4+v} < \infty$ to establish the asymptotic normality of $\tilde{\theta}_T^{mem}$. Its asymptotic variance-covariance matrix becomes $(E\eta_t^2 - 1)(\Sigma^{lh})^{-1}$.

We conclude this section with a finite sample simulation study as it will become clear that asymptotic analysis is not sufficient to appraise which estimators are the most attractive for empirical work. The details of the simulation design appear in Appendix B.

The results are quite easy to summarize and therefore not reported in detail (see Chen, Ghysels, and Wang (2010) for the details). The estimator that appears to have the best finite sample properties is LHRV. The LHRV estimator is typically vastly better than the estimators based on R^2, either minimum distance or likelihood-based. Compared to the LHR2 estimator, we also find that MDRV - which uses also RV but via a minimum distance criterion - is also less efficient, except in one case $m = 5$ and $T = 1000$. It should also be noted that 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 estimates directly the volatility process. The simulation results indicate that $\tilde{\theta}_T^{lhrv}$ and $\tilde{\theta}_T^{mem}$ appear to have the most desirable finite sample properties - the reason why we select these estimators for our empirical analysis.

5 Model Specifications and Evaluations

The class of HYBRID GARCH processes we introduce allows us to address quite a few intriguing empirical modeling strategies. For example:

- 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, also 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, or should we reply on simple daily realized volatilities? Hence, is the right sampling frequency of returns intra-daily? Or, can we come by with daily aggregates?
- The same question applies to longer horizons, such as bi-weekly volatility forecasts, which are most relevant for value-at-risk calculations.

To address these questions we introduce various model specifications that feature weighting schemes, news impact functions etc. We will focus on cases involving intra-daily returns in a first subsection. The next subsection covers weekly and bi-weekly predictions.

5.1 Daily Volatility Forecasts

We start with data-driven processes followed by parametric specifications for the HYBRID processes. Recall that:

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma H_t$$

In the sequel we will provide various specifications for the H_t process.

Data-driven HYBRID processes

A standard GARCH(1,1) model can be viewed as a special case of HYBRID-GARCH with the HYBRID process:

$$H_t \equiv r_t^2 \tag{5.1}$$

where the return in this case is the daily return. Using intra-daily data, we can also consider RV and SemiRV GARCH models, respectively with:

$$H_t \equiv RV_t \tag{5.2}$$

Likewise, we can account for asymmetries by computing semi-variances, as suggested by Barndorff-Nielsen, Kinnebrock, and Shephard (2008)

$$SemiRV_t \equiv \sum_{j=1}^{m} r_{t-(j-1)/m}^2 \mathbf{1}_{r_{t-(j-1)/m}^2 < 0} \tag{5.3}$$

yielding the SemiRV HYBRID process:

$$H_t \equiv RV_t + \delta SemiRV_t \tag{5.4}$$

The appeal of these HYBRID processes is that they involve only a small set of parameters. A standard GARCH(1,1) involves three parameters. The RV GARCH has the same number of parameters, but uses intra-daily information via realized volatility. The SemiRV GARCH adds one extra slope parameter δ that potentially captures asymmetries of negative returns.

Last but not least, we also considered so called realized volatility measures that are corrected for microstructure noise. Microstructure noise may mask the true price variation. This has prompted a substantial literature on how to correct measures of quadratic variation based on intra-daily data - which we will denote as RV_t^*. See for example Aït-Sahalia, Mykland, and Zhang (2005), Bandi and Russell (2006), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006) and Hansen and Lunde (2006), and references therein. We therefore also considered data-driven HYBRID processes that correct for the presence of intra-daily microstructure noise. One may use sub-sampling (see Zhang, Mykland, and Ait-Sahalia (2005) and Aït-Sahalia, Mykland, and Zhang (2011)), or the realized kernel (see Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011)), or pre-averaging (see Jacod, Li, Mykland, Podolskij, and Vetter (2009)) to construct RV_t^*. In our empirical applications, we will use the realized kernel method. More specifically, we will use the Oxford-Man Institute's *realised library* data described in Shephard and Sheppard (2010).[8] We refrain from reporting all the details of the empirical results for two reasons: (1) to save space and (2) since our findings are similar to Patton and Sheppard (2009) - namely that the use of realized measures corrected for microstructure noise yields results that are typically similar (or often worse) to the forecast performance using RV.

---

[8] See `http://realized.oxford-man.ox.ac.uk/` for further detail.

News Impact Curves and HYBRID processes

The computation of semi-variances brings us to the subject of asymmetric volatility models. In particular, it has been observed that 'good news' and 'bad news' have different impact for the prediction of future volatility. To capture asymmetries, Engle and Ng (1993) introduced the notion of a *news impact curve*, both as an object of economic interest and a diagnostic tool for volatility modeling. One commonly used specification was advocated by Glosten, Jagannathan, and Runkle (1993), which involves the following HYBRID process that pertains to the so called TGARCH or threshold GARCH model:

$$H_t \equiv r_t^2 + \delta \mathbf{1}_{r_t<0} r_t^2 \tag{5.5}$$

Inspired by the above specification Chen and Ghysels (2011) extended the notion of news impact curve applicable to a mixture of high and low frequency returns. Their approach involves MIDAS regressions with intra-daily news impact curves. Overall, they find that moderately good (intra-daily) news reduces volatility (the next day), while both very good news (unusual high intra-daily positive returns) and bad news (negative returns) increase volatility, with the latter having a more severe impact. Chen and Ghysels (2011) also find that asymmetries disappear over longer horizons. Moreover, they find that models featuring asymmetries dominate in terms of out-of-sample forecasting performance, especially during the 2007-2008 financial crisis.

We will introduce HYBRID processes involving intra-daily news impact curves - similar to Chen and Ghysels (2011) - but within the context of HYBRID-GARCH. Before we do, we need to discuss how to handle weighting schemes for intra-daily data.

How to Better Use High-Frequency Financial Data

The main empirical results of the paper pertain to whether we can make better use of intra-daily returns. The first approach involves a HYBRID process that allows for unequal weighting of intra-daily returns. The parametric specification of the HYBRID process is as follows:

$$H(\phi, \vec{r}_t) \equiv H_t \equiv \sum_{j=1}^{m} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/m + \phi_2 (i/m)^2)\right) r_{t-(j-1)/m}^2 \tag{5.6}$$

where the use of the (quadratic) Exponential Almon MIDAS filter, with $\phi = (\phi_0, \phi_1, \phi_2)$, is inspired by a similar approach proposed in Ghysels, Santa-Clara, and Valkanov

(2005).[9]

Note that when $\phi_0 = \phi_1 = \phi_2 = 0$, we have the equal weighting scheme of the RV specification. Estimating the parameters amounts to allowing for different weighting schemes. Compared to the RV-driven HYBRID process, we have at most three extra parameters to estimate in order to retrieve more information from the intra-daily return series. The extra parameters come at a cost that may cause poor out-of-sample forecast performance. This is an empirical question of course that we will address in the next section.

While parsimony is a concern, we do want to discuss first specifications involving more parameters than we have handled so far. Indeed, following the theme of asymmetries discussed earlier, we also consider HYBRID processes involving both intra-daily weighting schemes as well as news impact curves, yielding:

$$H_t \equiv \sum_{j=1}^{m} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/m + \phi_2 (i/m)^2)\right)(1 + \delta \mathbf{1}_{r_{t-(j-1)/m}<0}) r^2_{t-(j-1)/m} \tag{5.7}$$

Together with the parameters α, β, and γ in the generic HYBRID GARCH, this is the model specification with the largest parameter space. As it involves a flexible weighting scheme as well as an intra-daily news impact curve.

The concerns for parsimony prompts us to think about putting restrictions on the specification of the HYBRID process and/or putting restrictions on the parameters α and β that determine the volatility dynamics. We will make a distinction between parameter restriction on the parameters driving the HYBRID process and restrictions pertain to the parameters α and β. We will start with the latter.

Recall that with volatility being a persistent process, it is natural to weight intra-daily data differently. Recall that this is one of the motivations behind the class of HYBRID GARCH models. We can carry this logic a step further, as the persistence not only affects the intra-daily weighting scheme, but should also affect the parameter β in equation (2.1). The connection between the parameter vector (ϕ_0, ϕ_1, ϕ_2) and β relates to what we refer to earlier as structural HYBRID processes. One such restriction relates to what Chen, Ghysels, and Wang (2011) refer to as a Periodic HYBRID GARCH models because of its connection with the periodic GARCH(1,1) model of Bollerslev

---

[9] We opted for the Exponential Almon because it yields a convenient approach to testing restrictions, as opposed to the Beta polynomial proposed by Chen and Ghysels (2011) to handle intra-daily seasonality in the context of a MIDAS regression. Note also that the Exponential Almon lags in equation (5.6) are not normalized to add up to one.

and Ghysels (1996). The model can be written as follows:

$$
\begin{aligned}
V_{t+1|t} \quad &= \quad \alpha + \exp\left(\sum_{i=1}^{m} (\phi_0 + \phi_1 i/m + \phi_2 (i/m)^2)\right) V_{t|t-1} \qquad (5.8) \\
&\quad + \gamma \sum_{j=1}^{m} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/m + \phi_2 (i/m)^2)\right) r^2_{t-(j-1)/m}
\end{aligned}
$$

We will call this a HYBRID slope-constrained GARCH, or HYBRID SC GARCH as the parameter β is now replaced and linked to the parameters of the HYBRID process. Note that the above specification can also be enriched with an intra-daily news impact curve, yielding a HYBRID SC TGARCH.

Another set of restrictions pertains to the parametrization of the HYBRID process itself. In equation (5.6) we assumed a quadratic polynomial for the HYBRID process. We can set ϕ_2 equal to zero yielding an exponentially affine weighting scheme. We will refer to such a model as FC1 HYBRID GARCH, or FC1 HYBRID TGARCH, the latter involving news impact curves. These models can also feature slope constraints, such as for example FC1 HYBRID SC GARCH. Finally, we can further restrict the HYBRID process by imposing $\phi_2 = \phi_1 = 0$. This restriction on the HYBRID process, combined with its associated slope restriction amounts to assuming a high frequency data GARCH(1,1) data generating process (DGP) and is an example of a structural HYBRID GARCH with an underlying DGP. We will refer to this type of restriction as FC0.

5.2 Weekly and Bi-Weekly Volatility Forecasts

When we turn our attention to weekly and bi-weekly forecasts, we face interesting issues which we want to highlight in this subsection. In principle, these issues also apply to intra-daily returns and daily forecasts, but they are more acute for longer term forecasts. Before we do, we would like to note that all the models discussed in the previous subsection readily apply to multiple day horizons. For example the general HYBRID TGARCH with an h day forecast horizon can be written as:

$$
\begin{aligned}
V_{t+1|t} \quad &= \quad \alpha + \beta V_{t|t-1} \qquad (5.9) \\
&\quad + \gamma \sum_{j=1}^{mh} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/mh + \phi_2 (i/mh)^2)\right) (1 + \delta \mathbf{1}_{r_{t-(j-1)/m} < 0}) r^2_{t-(j-1)/m}
\end{aligned}
$$

where the above equation collapses to (5.7) when $h = 1$. Such extensions also apply to slope constrained GARCH models, as well as models involving constrained HYBRID processes.

Selecting the right sampling frequency

To motivate the discussion we start with a standard GARCH(1,1) model using weekly returns, namely:

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma (r_t^w)^2 \tag{5.10}$$

where we emphasize the fact that returns are sampled weekly with the superscript w. Likewise, we can think of a RV-GARCH(1,1) model:

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma RV_t^w \tag{5.11}$$

with the same convention for the notation. We could, however, also consider a HYBRID GARCH with daily returns to replace equation (5.10):

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2) \right) (r_{ht-(j-1)}^d)^2 \tag{5.12}$$

where the superscript d refers to daily returns and $h = 5$ for the weekly/daily sampling frequency combination. We will denote the above model as HYBRID GARCH D. Likewise, we can also apply the same logic to RV-driven GARCH models, namely replacing equation (5.11) with:

$$V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma \sum_{j=1}^{h} \exp\left(\sum_{i=1}^{j-1} (\phi_0 + \phi_1 i/h + \phi_2 (i/h)^2) \right) RV_{ht-(j-1)}^d \tag{5.13}$$

Obviously, as before, we are facing a trade-off between parameter proliferation and the level of aggregation of returns and/or RV. In each of these cases, we can constraint the HYBRID process parameters, or we can impose a slope restriction. Likewise, we can also impose news impact curves when we consider (daily) returns. The generic specification in equation (5.9) uses all high frequency data directly and involves a general HYBRID process with intra-daily news impact curves. The key question therefore, is whether to use intra-daily data, or daily returns or daily RV, instead of weekly returns or weekly RV.

It should also be noted that the selection of the sampling frequency pertains to the issue of considering an iterated forecast for the next five days versus a direct forecast based on a coarser data set of past weekly returns. Relatively little work has been done on the comparison between direct and iterated forecasts in the context of volatility. One exception is Ghysels, Rubia, and Valkanov (2009) who consider a regression-based approach and compare several approaches of producing multi-period ahead forecasts of volatility –iterated, direct, and mixed-data sampling (MIDAS) regressions. The latter is shown to be the best and is similar to the HYBRID GARCH class of models proposed here. While the MIDAS approach dominates, it is not clear whether we should use equation (5.9) applicable to all high frequency data directly, or daily aggregates such as RV^d, or daily returns, etc. These questions pertain to the selection of the proper sampling frequency in a mixed sampling frequency setting.

5.3 Data and Model Evaluation

We use the S&P 500 Futures 5-minute returns from April 1st, 1982 to December 31st, 2008. To estimate the models and perform out-of-sample evaluations, we use a rolling window sample which is moved forward monthly. There are 178 rolling windows and each window contains 120-months in-sample data and 24-month set aside for out-of-sample appraisals. We consider three horizons: one-day, one-week, and two-weeks ahead forecasts. All models, with specifications listed in Table 1 through 3, are estimated using two methods: 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 are evaluated via the Giacomini and White (2006) test, henceforth denoted 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 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 will not only compare models involving high frequency data directly, or daily aggregate measures, but we will also include ARCH-type models involving daily returns - i.e. the original models that were used in the literature on asymmetries in volatility.

The loss function we use 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 follows:

$$L(h_t, RV_t) = \log h_t + \frac{RV_t}{h_t} - (\log RV_t + 1) \tag{5.14}$$

Finally, we construct a score for each model based on the GW tests' p-values and ratios:

$$s_A = \frac{\sum_{B \neq A} \mathbf{1}_{p_{AB} <= \alpha} r_{AB} + 0.5(1 - \mathbf{1}_{p_{AB} <= \alpha})}{(n(n-1)/2)} \tag{5.15}$$

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 score is normalized such that the summation of scores of all models is one.[10]

6 Empirical Findings

The comparison of Conditional Predictive Ability (CPA) of each model between different methods are shown in Table 4 for daily, weekly, and bi-weekly data. The results consistently show that the choice of estimation method - MEM or LHRV - is not important, except that for a few exceptions, LHRV is preferred. Hence, to compare the forecast performance of models, we only need to consider the LHRV method. The p-values of GW tests are shown in Table 5 for daily models. The details for the weekly and bi-weekly data, and the statistics of GW test are omitted in order to save space. We also report the ratio of choosing model A (the model No. shown in the row header) against model B (the model No. shown in the column header) according to the decision rule discussed in Giacomini and White (2006). When the p-value of GW test is below the significance level of interest, for example, 10%, the ratio is an indicator of the relative forecast performance of the two models.

Examining the results in Table 5, we note the following for daily model comparisons:

1. GARCH and TGARCH models using daily returns, Model (1) and (2), are dominated by all other models, i.e. models using intra-daily data.
2. Symmetric models are always dominated by their asymmetric counterparts, which implies that asymmetry does matter, despite the required extra parameters.

---

[10] Other methods for comparing large classes of models exist, see e.g. Hansen, Lunde, and Nason (2011) and references therein. These tests are unconditional, however, whereas we prefer to rely on conditional tests.

3. Focusing on symmetric models, the RV GARCH model, Model (3), is dominated by all other HYBRID or slope constrained HYBRID models. Focusing on asymmetric models, the SemiRV GARCH model, Model (4), is also always less preferred than any HYBRID or slope constrained HYBRID asymmetric models.

These results support two key findings for daily forecasting of volatility: (1) asymmetries matter and (2) the weighting scheme also does matter.

To further document these findings we turn our attention to Figure 1 where we display the weigths of the best daily model, i.e. the HYBRID TGARCH model, in three representative subsamples of our rolling sample scheme: P1 is for period December, 1986 to November, 1996; P2 for period August, 1989 to July, 1999; P3 for period April, 1993 to March, 2003.[11] It appears from the figure that the weighting schemes are fairly stable across subsamples. In Figure 3 we display the news impact curve of the best daily model in the same three subsamples. We see again a stable pattern across subsamples and a pattern that is distinctly asymmetric and feature larger impact of bad news - as commonly documented in the literature.

Next we turn to the weekly models, where we find that:

1. Same as daily models, GARCH and TGARCH models are dominated by all other models.

2. Symmetric models are dominated by their asymmetric counterparts in most cases, which implies that asymmetric effects still matter at weekly horizons.

3. The models using daily returns (having D in their acronym), are always dominated by their counterparts using 5-minute returns. In fact, almost all models based on 5-minute returns show better performance than any model based on daily returns, which implies that using 5-minute returns, instead of daily returns, may result in better forecast performance for weekly models.

4. RV GARCH models perform worse than most HYBRID or slope constrained HYBRID models. The same conclusion applies to SemiRV GARCH models. Hence, the weight scheme still matters for weekly horizons.

Figures 2 and 4 display respectively the weighting scheme and news impact curves of the best weekly model, i.e. FC0 HYBRID SC TGARCH model (see again below), in the aforementioned three subsamples. Arguably, the weighting schemes and impact

---

[11] We will discuss later how the best model is selected - the results appear in Table 6.

curves vary more across subsamples. Nevertheless all display the two key features supporting our findings: (1) the weighting scheme matters and (2) asymmetries do too.

Finally we turn our attention to bi-weekly models:

1. GARCH and TGARCH models with daily data are again dominated by other models.
2. As to the asymmetric effect, now the conclusion is different: many asymmetric models are dominated by their symmetric counterparts. It implies that asymmetric effects have less importance as the horizon increases, a finding also reported in Chen and Ghysels (2011).
3. The models using daily returns are still less preferred to their counterparts using 5-minute returns.
4. RV and SemiRV GARCH models are still less preferred than most HYBRID or slope constrained HYBRID models, which means that the weight scheme still matters for bi-weekly horizon.

Figure 5 covers news impact curve (we skip the weighting scheme) of the best bi-weekly model, i.e. FC1 HYBRID SC GARCH model (see again below), in the same three subsamples. We clearly see that the asymmetries have disappeared and the news impact curves are stable across subsamples as well.

The score of a model is a good indicator of model's forecast performance. We rank all models according to their score in Table 6 which provides an easy summary. As shown in Table 6, the top 6 models are always HYBRID or slope constrained HYBRID models, and the last two models are always GARCH and TGARCH models.

7 Conclusions

We proposed a general unifying framework that allows the use of different frequency returns to model conditional heteroskedasticity. We call the class of models HYBRID-GARCH models, as the volatility dynamics are driven by what we call HYBRID processes. The topics addressed in this paper have many applications, given the wide use of multi-period volatility forecasts in risk and portfolio analysis.

The main conclusions, based on our analysis are fairly simple:

1. to use intra-daily weighting schemes other than purely aggregation like RV and SemiRV;
2. to use 5-minute returns *directly* rather than daily returns or daily aggregates like RV even in long horizon forecast;
3. to include asymmetric effect at short horizons but not for long horizons.

These results have implications that go against most of the current practice of (1) using direct forecasting methods for long horizon forecasts, (2) using daily aggregates of intra-daily data, and (3) using daily returns or realized measures for long horizon forecasts.

References

AÏT-SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2005): "How often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

——— (2011): "Ultra high frequency volatility estimation with dependent microstructure noise," *Journal of Econometrics*, 160(1), 160–175.

AKNOUCHE, A., AND A. BIBI (2009): "Quasi-maximum likelihood estimation of periodic GARCH and periodic ARMA-GARCH processes," *Journal of Time Series Analysis*, 30(1), 19–46.

ANDERSEN, T. G., T. BOLLERSLEV, AND S. LANGE (1999): "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance*, 6(5), 457–477.

ANDREOU, E., AND E. GHYSELS (2002): "Rolling sample volatility estimators: some new theoretical, simulation and empirical results," *Journal of Business and Economic Statistics*, 20, 363–376.

BANDI, F. M., AND J. R. RUSSELL (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79, 655–692.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2006): "Regular and Modified Kernel-Based Estimators of Integrated Variance:

The Case with Independent Noise," Discussion paper, Department of Mathematical Sciences, University of Aarhus.

——— (2008): "Designing realized kernels to measure the ex postvariation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

——— (2011): "Subsampling realised kernels," *Journal of Econometrics*, 160(1), 204–219.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard (2008): "Measuring downside risk realised semivariance," Discussion Paper, Oxford.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, 31(3), 307–327.

Bollerslev, T., and E. Ghysels (1996): "On Periodic Autoregressive Conditional Heteroskedasticity," *Journal of Business and Economic Statistics*, 14, 139–151.

Chen, X., and E. Ghysels (2011): "News - good or bad - and its impact on predictions over multiple horizons," .

Chen, X., E. Ghysels, and F. Wang (2010): "HYBRID-GARCH: A Generic Class of Models for Volatility Predictions using High Frequency Data," Paper available at `www.unc.edu/~eghysels`.

——— (2011): "HYBRID GARCH Models and Intra-Daily Return Periodicity," *Journal of Time Series Econometrics*, 3(1), Article 11.

Cipollini, F., R. Engle, and G. Gallo (2006): "Vector multiplicative error models: representation and inference," Working Paper 12690, National Bureau of Economic Research.

de Vilder, R., and M. Visser (2008): "Ranking and Combining Volatility Proxies for Garch and Stochastic Volatility Models," MPRA paper 11001.

Diebold, F., and R. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Drost, F. C., and T. E. Nijman (1993): "Temporal aggregation of GARCH processes," *Econometrica*, 61(4), 909–927.

Drost, F. C., and B. J. Werker (1996): "Closing the Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31–57.

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1008.

——— (2002): "New frontiers for ARCH models," *Journal of Applied Econometrics*, 17(5), 425–446.

Engle, R. F., and G. M. Gallo (2006): "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131(1–2), 3–27.

Engle, R. F., and V. Ng (1993): "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

Fan, J., Y. Fan, and J. Jiang (2007): "Dynamic integration of time-and state-domain methods for volatility estimation," *Journal of the American Statistical Association*, 102, 618–631.

Fan, J., and Y. Wang (2008): "Spot volatility estimation for high-frequency data," *Statistics and its Interface*, pp. 279–288.

Foster, D. P., and D. B. Nelson (1996): "Continuous record asymptotics for rolling sample variance estimators," *Econometrica*, 64(1), 139–174.

Gallant, A., and H. White (1988): *A unified theory of estimation and inference for nonlinear dynamic models*. Basil Blackwell Oxford.

Ghysels, E., A. Rubia, and R. Valkanov (2009): "Multi-Period Forecasts of Volatility: Direct, Iterated, and Mixed-Data Approaches," Working paper, UNC.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005): "There is a risk-return tradeoff after all," *Journal of Financial Economics*, 76, 509–548.

——— (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.

Ghysels, E., A. Sinko, and R. Valkanov (2006): "MIDAS Regressions: Further Results and New Directions," *Econometric Reviews*, 26, 53 – 90.

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

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): “On the relation between the expected value and the volatility of the nominal excess return on stocks,” *Journal of Finance*, 48(5), 1779–1801.

Goncalves, S., and H. White (2004): “Maximum likelihood and the bootstrap for nonlinear dynamic models,” *Journal of Econometrics*, 119(1), 199–219.

Hansen, P. R., Z. A. Huang, and H. H. Shek (2010): “Realized GARCH: A Complete Model of Returns and Realized Measures of Volatility,” Working paper, Stanford University.

Hansen, P. R., and A. Lunde (2006): “Realized Variance and Market Microstructure Noise,” *Journal of Business and Economic Statistics*, 24(2), 127–161.

Hansen, P. R., A. Lunde, and J. M. Nason (2011): “Model confidence sets for forecasting models,” *Econometrica*, 79(2), 453–497.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): “Microstructure noise in the continuous case: the pre-averaging approach,” *Stochastic Processes and their Applications*, 119(7), 2249–2276.

Jennrich, R. (1969): “Asymptotic properties of non-linear least squares estimators,” *The Annals of Mathematical Statistics*, 40(2), 633–643.

Lanne, M. (2006): “A mixture multiplicative error model for realized volatility,” *Journal of Financial Econometrics*, 4(4), 594.

Malliavin, P., and M. Mancino (2005): “A Fourier transform method for nonparametric estimation of multivariate volatility,” *Annals of Statistics*, 37, 1983–2010.

Meddahi, N., and E. Renault (2004): “Temporal aggregation of volatility models,” *Journal of Econometrics*, 119(2), 355–379.

Mykland, P. A., and L. Zhang (2008): “Inference for volatility-type objects and implications for hedging,” *Statistics 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,” *Journal of Econometrics*, 60, 61–90.

NELSON, D., AND D. FOSTER (1995): “Filtering and forecasting with misspecified ARCH models II: Getting the right variance with the wrong model,” *Journal of Econometrics*, 67, 303–335.

PATTON, A. (2011): “Volatility forecast evaluation and comparison using imperfect volatility proxies,” *Journal of Econometrics*, 160(1), 246–256.

PATTON, A., AND K. SHEPPARD (2009): “Optimal combinations of realised volatility estimators,” *International Journal of Forecasting*, 25(2), 218–238.

SHEPHARD, N., AND K. SHEPPARD (2010): “Realising the future: forecasting with high frequency based volatility (HEAVY) models,” *Journal of Applied Econometrics*, 23(2), 197–231.

VISSER, M. P. (2011): “GARCH parameter estimation using high-frequency data,” *Journal of Financial Econometrics*, 9(1), 162–197.

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.

ZHANG, L., P. MYKLAND, AND Y. AIT-SAHALIA (2005): “A tale of two time scales: Determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association*, 100(472), 1394–1411.

ZHAO, Z., AND W. WU (2008): “Confidence bands in nonparametric time series regression,” *Annals of Statistics*, 36, 1854–1878.

Technical Appendices

A Regularity Conditions

The purpose of this appendix is to collect the regularity conditions used in the paper. In what follows, we use ∇ to denote the vector differential operator (w.r.t. θ) so that ∇f is the gradient (column vector) of scaler function f, and $Hess(f)$ the Hessian matrix of f, i.e., $ent_{i,j}Hess(f) = \partial_i \partial_j f$ where ∂_k denotes the partial derivative w.r.t. the k^{th} parameter in $\theta = (\alpha, \beta, \gamma, \phi)$. When ϕ is a vector, ∂_ϕ refers to the partial derivative w.r.t. *each component* of ϕ, say ϕ_i, and ∂_ϕ^2 is treated as $\partial_{\phi_i}\partial_{\phi_j}$. ∇_ϕ is a vector differential operator w.r.t. ϕ when ϕ is a vector.

Assumption A.1. *$r_s^2 \in L^2(\Omega, \mathcal{F}, P)$ for some probability space $(\Omega, \mathcal{F}, P)$, and $P_l(RV_{t+1}|\mathcal{I}_t) = \sigma^2_{t+1|t}$, where $RV_{t+1} = \sum_{j=0}^{m-1} r^2_{t+1-j/m}$, $\mathcal{I}_t$ is a closed subspace of $L^2(\Omega, \mathcal{F}, P)$ and it consists of the information on the high frequency returns up to time t. r_s's are non-degenerate, and linearly independent.*

Assumption A.2. *$\{r_s\}$ is strictly stationary and ergodic.*

Assumption A.3. *$\{r_s\}$ is strictly stationary and strong mixing. The mixing coefficient $\alpha(k)$ satisfies $\sum_{k=0}^{\infty} \alpha(k)^{v_2/(2+v_2)} < \infty$ for some $v_2 > 0$. And $Er_s^{4(2+v_2)} < \infty$.*

Assumption A.4. *Suppose that Φ is a connected set and H is a mapping from $\Phi \times \mathbb{R}^m$ to $\mathbb{R}^+$ such that (1) $H(\cdot, \vec{x}) \in C^2$ for $\vec{x} \in \mathcal{R}^m$; (2) $H(\phi, \cdot)$, $\partial_\phi H(\phi, \cdot)$, $\partial_\phi^2 H(\phi, \cdot)$ are $\mathcal{B}(R^m)/\mathcal{B}(R)$ measurable for $\phi \in \Phi$; (3) for r_t satisfying Assumption A.1,* 1 *and $H(\phi, \vec{r}_t)$ and (each component of) $\partial_\phi(H(\phi, \vec{r}_t))$ are linearly independent for all $\phi \in \Phi$.*

Assumption A.5 (Moment conditions on H)**.** *For r_t satisfying Assumption A.1,*

(1) $E\sup_{\phi \in \overline{\Phi^0}} H(\phi, \vec{r}_t)$, $E\sup_{\phi \in \overline{\Phi^0}} |\partial_\phi H(\phi, \vec{r}_t)|$ and $E\sup_{\phi \in \overline{\Phi^0}} |\partial_\phi^2 H(\phi, \vec{r}_t)|$ are finite.

(2) $E(\sup_{\phi \in \overline{\Phi^0}} H(\phi, \vec{r}_t))^2$, $E(\sup_{\phi \in \overline{\Phi^0}} |\partial_\phi H(\phi, \vec{r}_t)|)^2$ and $E(\sup_{\phi \in \overline{\Phi^0}} |\partial_\phi^2 H(\phi, \vec{r}_t)|)^2$ are finite.

(3) $E(\sup_{\phi \in \overline{\Phi^0}} H(\phi, \vec{r}_t))^4$ and $E(\sup_{\phi \in \overline{\Phi^0}} |\partial_\phi H(\phi, \vec{r}_t)|)^4$ are finite.

(4) $E(\sup_{\phi \in \overline{\Phi^0}} H(\phi, \vec{r}_t))^{2(2+v_2)}$ and $E(\sup_{\phi \in \overline{\Phi^0}} |\partial_\phi H(\phi, \vec{r}_t)|)^{2(2+v_2)}$ are finite, where v_2 is defined in Assumption A.3.

Assumption A.6. *The rank of the matrix $(\nabla_{\phi_1}\Psi_0 \quad \nabla_{\phi_1}\Psi_1 \quad \ldots \quad \nabla_{\phi_1}\Psi_{m-1})$ is same as the dimension of ϕ_1.*

Assumption A.7. *(1) θ_* is identifiably unique in $\mathcal{C}$. (2) $\theta_* \in \mathcal{C}^0$. (3) The determinant of $E[Hess\left(\varepsilon_t^2\right)(\theta_*)]$ (or $E[Hess\left(e_t^2\right)(\theta_*)]$) is positive.*

Assumption A.8. *(1) θ_{**} is identifiably unique in $\mathcal{C}$. (2) $\theta_{**} \in \mathcal{C}^0$. (3) The determinant of $E[Hess(l_t)(\theta_{**})]$ (or $E[Hess(h_t)(\theta_{**})]$) is positive.*

B Simulation study

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

$$r_{s+1/m} = \sqrt{v_{s+1/m|s}}\varepsilon_{s+1/m}, v_{s+1/m|s} = a + bv_{s|s-1/m} + cr_s^2, \varepsilon_{s+1/m} \overset{iid}{\sim} N(0,1). \tag{B.1}$$

The second is a GARCH diffusion process, i.e., Model (3.5) with $\eta = 0$. The discretely-sampled high frequency return $r_s{=}p_s - p_{s-1/m}$ is therefore a *weak* GARCH(1,1):

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

with $a = \omega(1-e^{-\theta/m})/m$, $c = e^{-\theta/m} - b$, and $|b| < 1$ is the solution to $b/(1+b^2) = (\rho e^{-\theta/m} - 1)/(\rho(1+e^{-2\theta/m})-2)$ where $\rho = [4(e^{-\theta/m}-1+\theta/m)+2\theta/m(1+\theta/m(1-\lambda)/\lambda)]/(1-e^{-2\theta/m})$. We then construct the HYBRID GARCH process (3.3) based on either model (B.1) or model (B.2).

The values of parameters in model (B.1) are $a = 2.8E-06, b = 0.9770, c = 0.0225$ which 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 A.2. For the GARCH diffusion process, we consider $\theta = 0.0350$, $\omega = 0.6365$, $\lambda = 0.2962$ which are based on the daily Deutschemark - US dollar exchange rate from October 1, 1987 to September 30, 1992 (See Andersen, Bollerslev, and Lange (1999)). The values of α_m, β_m and γ_m are reported in table below with m = 5, 78, 288 for model (B.1) and with m = 24, 144, 288 for model (B.2).

	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

The estimators considered are: $\tilde{\theta}_T^{mdrv}$, defined in (4.2), and the companion estimator $\tilde{\theta}_T^{mdr2}$, replacing RV by R^2, as well as (quasi-)likelihood-based estimators $\tilde{\theta}_T^{lhr2}$ defined in (4.4), and $\tilde{\theta}_T^{lhrv}$, defined in (4.5). Finally, the simulation study also includes the MEM method described in Section 4.2.2. Recall that Engle and Gallo (2006) and Cipollini, Engle, and Gallo (2006) noted that the estimation of θ_0 and g are asymptotically independent, and thus $\tilde{\theta}_T^{mem}$ is *asymptotically* the same as $\tilde{\theta}_T^{lhrv}$. The purpose of this section is to examine differences in small sample behavior.

In the simulation experiment, we consider 1000 replications of sample path (3.3), each having the first 1000 observations burn-in and consisting of 500 and 1000 observations left in the sample. For the continuous case, we use Euler discretization to simulate the diffusion process: take one day as a reference measure, and simulate 24 hours of trading with $dt = 1/86400$.

Table 1: Summary of Model Specifications (A)

Let $r_{s/m}$ denote the high-frequency return and m is the frequency in one day, i.e. $r_{s/m}$ is the ith high-frequency return at day j if $s = j-1+i/m, i = 1, ..., m$. The return at period t of interest, namely daily, weekly, and bi-weekly return, 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, i.e. $h = 1$ for daily return, 5 for weekly return, and 10 for bi-weekly return. Especially, 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, namely the twelve models with model 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 2: Summary of Model Specifications (B)

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

Table 3: Summary of Model Specifications (C)

(21)	FC1 HYBRID SC GARCH D	$V_{t+1\|t} = \alpha + \exp\Big(\sum_{i=1}^{h}(\phi_0 + \phi_1 i/h)\Big)V_{t\|t-1}$ $+ \gamma \sum_{j=1}^{h} \exp\Big(\sum_{i=1}^{j-1}(\phi_0 + \phi_1 i/h)\Big){r^d_{ht-(j-1)}}^2$
(22)	FC1 HYBRID SC TGARCH D	$V_{t+1\|t} = \alpha + \exp\Big(\sum_{i=1}^{h}(\phi_0 + \phi_1 i/h)\Big)V_{t\|t-1}$ $+\gamma \sum_{j=1}^{h} \exp\Big(\sum_{i=1}^{j-1}(\phi_0 + \phi_1 i/h)\Big)(1 + \delta \mathbf{1}_{r^d_{ht-(j-1)}<0}){r^d_{ht-(j-1)}}^2$
(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	$V_{t+1\|t} = \alpha + \exp(h\phi_0)V_{t\|t-1} + \sum_{j=1}^{h} \exp((j-1)\phi_0)(\gamma + \delta \mathbf{1}_{r^d_{ht-(j-1)}<0}){r^d_{ht-(j-1)}}^2$

Table 4: GW Test for comparing daily models between methods

Each model is estimated through two methods: MEM and LHRV. The GW tests show that the conditional predictive ability of each model might be different due to its estimation methods. The column names "stat" stands for "Statistics" and "ratio" for "the ratio of choosing model estimated through method A other than that through method B according to the decision rule". A ratio is only meaningful when the corresponding p-value is smaller than the given significance level, for example, 10%.

	Daily Models			Weekly Models			Bi-Weekly Models		
Method A	LHRV			LHRV			LHRV		
Method B	MEM			MEM			MEM		
MODEL	stat	p-value	ratio	stat	p-value	ratio	stat	p-value	ratio
(1)	1.18	0.55	0.0	0.54	0.76	1.0	2.00	0.37	1.0
(2)	0.70	0.70	0.1	1.10	0.58	1.0	0.23	0.89	0.0
(3)	1.04	0.59	1.0	0.27	0.87	0.0	2.09	0.35	0.0
(4)	0.28	0.87	0.0	3.69	0.16	1.0	1.71	0.43	0.0
(5)				1.16	0.56	0.0	0.83	0.66	1.0
(6)				2.58	0.27	0.0	1.29	0.53	1.0
(7)	2.98	0.23	0.9	5.56	0.06	1.0	0.92	0.63	0.9
(8)	1.28	0.53	0.1	6.83	0.03	1.0	5.16	0.08	1.0
(9)	0.33	0.85	0.0	2.58	0.27	0.9	1.89	0.39	0.1
(10)	2.12	0.35	0.0	5.31	0.07	1.0	1.53	0.46	1.0
(11)				1.17	0.56	0.1	1.39	0.50	0.1
(12)				4.03	0.13	0.0	1.18	0.55	0.9
(13)				0.66	0.72	0.0	1.93	0.38	1.0
(14)				3.85	0.15	0.0	3.37	0.19	1.0
(15)	4.99	0.08	1.0	3.76	0.15	1.0	4.21	0.12	1.0
(16)	1.99	0.37	0.0	2.68	0.26	0.1	3.08	0.21	1.0
(17)	0.95	0.62	0.9	1.63	0.44	0.1	2.02	0.36	1.0
(18)	2.38	0.30	1.0	0.92	0.63	0.2	2.16	0.34	0.0
(19)				1.89	0.39	1.0	1.78	0.41	1.0
(20)				2.07	0.35	1.0	0.39	0.82	0.0
(21)				6.42	0.04	0.0	2.27	0.32	1.0
(22)				5.42	0.07	0.0	2.40	0.30	0.0
(23)	0.68	0.71	1.0	0.11	0.95	1.0	2.96	0.23	0.0
(24)	2.46	0.29	0.0	1.03	0.60	0.0	1.76	0.42	0.0
(25)	2.00	0.37	0.1	3.99	0.14	0.9	3.05	0.22	1.0
(26)	0.17	0.92	0.1	3.52	0.17	0.0	2.03	0.36	1.0
(27)				0.79	0.67	0.9	1.36	0.51	1.0
(28)				2.63	0.27	0.9	0.81	0.67	1.0
(29)				1.13	0.57	0.0	2.05	0.36	0.0
(30)				4.31	0.12	1.0	2.73	0.25	1.0

Table 5: GW tests for daily models estimated with LHRV method

All daily models are estimated via LHRV method. Each cell shows the p-value of GW test and in the bracket is the ratio of choosing model A other than model B according to the decision rule.

Model B	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)
Model A								
(1)	[]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]
(2)	0.00[1.0]	[]	0.00[0.1]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.1]	0.00[0.0]
(3)	0.00[1.0]	0.00[0.9]	[]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.3]	0.00[0.0]
(4)	0.00[1.0]	0.00[1.0]	0.00[1.0]	[]	0.00[0.7]	0.00[0.1]	0.00[1.0]	0.00[0.2]
(7)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.3]	[]	0.00[0.0]	0.00[1.0]	0.00[0.2]
(8)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.9]	0.00[1.0]	[]	0.00[1.0]	0.00[0.8]
(9)	0.00[1.0]	0.00[0.9]	0.00[0.7]	0.00[0.0]	0.00[0.0]	0.00[0.0]	[]	0.00[0.0]
(10)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.8]	0.00[0.8]	0.00[0.2]	0.00[1.0]	[]
(15)	0.00[1.0]	0.00[0.9]	0.00[0.9]	0.00[0.0]	0.00[0.1]	0.00[0.0]	0.04[0.8]	0.00[0.0]
(16)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.9]	0.00[0.9]	0.00[0.2]	0.00[1.0]	0.01[0.8]
(17)	0.00[1.0]	0.00[0.9]	0.00[0.9]	0.00[0.0]	0.00[0.1]	0.00[0.0]	0.00[1.0]	0.00[0.0]
(18)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.9]	0.00[1.0]	0.00[0.4]	0.00[1.0]	0.00[0.9]
(23)	0.00[1.0]	0.00[0.9]	0.00[0.8]	0.00[0.0]	0.00[0.1]	0.00[0.0]	0.00[0.7]	0.00[0.0]
(24)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.6]	0.00[0.9]	0.00[0.3]	0.00[1.0]	0.00[0.6]
(25)	0.00[1.0]	0.00[0.9]	0.00[0.9]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.8]	0.00[0.0]
(26)	0.00[1.0]	0.00[1.0]	0.00[1.0]	0.00[0.9]	0.00[0.9]	0.00[0.2]	0.00[1.0]	0.00[0.8]
Score	0.000	0.013	0.023	0.081	0.070	**0.113**	0.028	0.089

Model B	(15)	(16)	(17)	(18)	(23)	(24)	(25)	(26)
Model A								
(1)	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.0]
(2)	0.00[0.1]	0.00[0.0]	0.00[0.1]	0.00[0.0]	0.00[0.1]	0.00[0.0]	0.00[0.1]	0.00[0.0]
(3)	0.00[0.1]	0.00[0.0]	0.00[0.1]	0.00[0.0]	0.00[0.2]	0.00[0.0]	0.00[0.1]	0.00[0.0]
(4)	0.00[1.0]	0.00[0.1]	0.00[1.0]	0.00[0.1]	0.00[1.0]	0.00[0.4]	0.00[1.0]	0.00[0.1]
(7)	0.00[0.9]	0.00[0.1]	0.00[0.9]	0.00[0.0]	0.00[0.9]	0.00[0.1]	0.00[1.0]	0.00[0.1]
(8)	0.00[1.0]	0.00[0.8]	0.00[1.0]	0.00[0.6]	0.00[1.0]	0.00[0.7]	0.00[1.0]	0.00[0.8]
(9)	0.04[0.2]	0.00[0.0]	0.00[0.0]	0.00[0.0]	0.00[0.3]	0.00[0.0]	0.00[0.2]	0.00[0.0]
(10)	0.00[1.0]	0.01[0.2]	0.00[1.0]	0.00[0.1]	0.00[1.0]	0.00[0.4]	0.00[1.0]	0.00[0.2]
(15)	[]	0.00[0.0]	0.06[0.2]	0.00[0.0]	0.00[0.7]	0.00[0.0]	0.06[0.5]	0.00[0.0]
(16)	0.00[1.0]	[]	0.00[1.0]	0.00[0.3]	0.00[1.0]	0.00[0.8]	0.00[1.0]	0.09[0.4]
(17)	0.06[0.8]	0.00[0.0]	[]	0.00[0.0]	0.00[0.7]	0.00[0.0]	0.00[0.6]	0.00[0.0]
(18)	0.00[1.0]	0.00[0.7]	0.00[1.0]	[]	0.00[1.0]	0.00[0.7]	0.00[1.0]	0.00[0.8]
(23)	0.00[0.3]	0.00[0.0]	0.00[0.3]	0.00[0.0]	[]	0.00[0.0]	0.00[0.4]	0.00[0.0]
(24)	0.00[1.0]	0.00[0.2]	0.00[1.0]	0.00[0.3]	0.00[1.0]	[]	0.00[1.0]	0.00[0.4]
(25)	0.06[0.5]	0.00[0.0]	0.00[0.4]	0.00[0.0]	0.00[0.6]	0.00[0.0]	[]	0.00[0.0]
(26)	0.00[1.0]	0.09[0.6]	0.00[1.0]	0.00[0.2]	0.00[1.0]	0.00[0.6]	0.00[1.0]	[]
Score	0.043	0.102	0.050	**0.111**	0.039	0.093	0.043	**0.103**

Table 6: Rank of models according to scores

Score is a good indicator of how much the model is better or worse than the others. The higher the score of a model is, the better of the predictive ability the model might be. The table lists the rank of models for different horizons according to their scores.

	Daily Models		Weekly Models		Bi-weekly Models	
Rank	Name	Score	Name	Score	Name	Score
1	HYBRID TGARCH	0.113	FC0 HYBRID SC TGARCH	0.059	FC1 HYBRID SC GARCH	0.053
2	FC1 HYBRID SC TGARCH	0.111	FC1 HYBRID SC TGARCH	0.058	FC0 HYBRID GARCH	0.052
3	FC0 HYBRID SC TGARCH	0.103	FC0 HYBRID TGARCH	0.057	FC0 HYBRID SC GARCH	0.052
4	FC1 HYBRID TGARCH	0.102	FC1 HYBRID TGARCH	0.056	FC0 HYBRID TGARCH	0.050
5	FC0 HYBRID TGARCH	0.093	HYBRID SC TGARCH	0.055	HYBRID GARCH	0.049
6	HYBRID SC TGARCH	0.089	FC0 HYBRID GARCH	0.054	FC1 HYBRID TGARCH	0.049
7	SemiRV GARCH	0.081	FC1 HYBRID SC GARCH	0.052	FC0 HYBRID SC TGARCH	0.049
8	HYBRID GARCH	0.070	SemiRV GARCH	0.051	FC1 HYBRID GARCH	0.048
9	FC1 HYBRID SC GARCH	0.050	FC0 HYBRID SC GARCH	0.051	FC1 HYBRID SC TGARCH	0.048
10	FC1 HYBRID GARCH	0.043	SemiRV GARCH D	0.050	HYBRID TGARCH	0.047
11	FC0 HYBRID SC GARCH	0.043	FC1 HYBRID GARCH	0.047	HYBRID SC TGARCH	0.044
12	FC0 HYBRID GARCH	0.039	HYBRID TGARCH	0.044	SemiRV GARCH D	0.043
13	HYBRID SC GARCH	0.028	RV GARCH	0.043	HYBRID SC GARCH	0.043
14	RV GARCH	0.023	HYBRID SC GARCH	0.043	RV GARCH	0.042
15	TGARCH	0.013	HYBRID GARCH	0.037	SemiRV GARCH	0.040
16	GARCH	0.000	FC0 HYBRID SC TGARCH D	0.026	FC0 HYBRID SC GARCH D	0.034
17			RV GARCH D	0.025	FC1 HYBRID SC GARCH D	0.026
18			FC0 HYBRID SC GARCH D	0.025	FC0 HYBRID SC TGARCH D	0.025
19			HYBRID TGARCH D	0.020	RV GARCH D	0.023
20			FC0 HYBRID GARCH D	0.019	FC1 HYBRID SC TGARCH D	0.021
21			FC0 HYBRID TGARCH D	0.019	FC0 HYBRID TGARCH D	0.021
22			FC1 HYBRID GARCH D	0.017	HYBRID SC TGARCH D	0.020
23			FC1 HYBRID TGARCH D	0.016	FC1 HYBRID TGARCH D	0.020
24			FC1 HYBRID SC GARCH D	0.016	FC0 HYBRID GARCH D	0.020
25			HYBRID GARCH D	0.015	HYBRID TGARCH D	0.019
26			FC1 HYBRID SC TGARCH D	0.015	FC1 HYBRID GARCH D	0.019
27			HYBRID SC GARCH D	0.014	HYBRID SC GARCH D	0.018
28			HYBRID SC TGARCH D	0.013	HYBRID GARCH D	0.016
29			GARCH	0.002	GARCH	0.005
30			TGARCH	0.002	TGARCH	0.004

Figure 1: Weights of Best Daily Model in Three Subsamples

The figure shows the weigths of the best daily model, i.e. HYBRID TGARCH model, in three subsamples. P1 is for period December, 1986 to November, 1996; P2 for period August, 1989 to July, 1999; P3 for period April, 1993 to March, 2003.

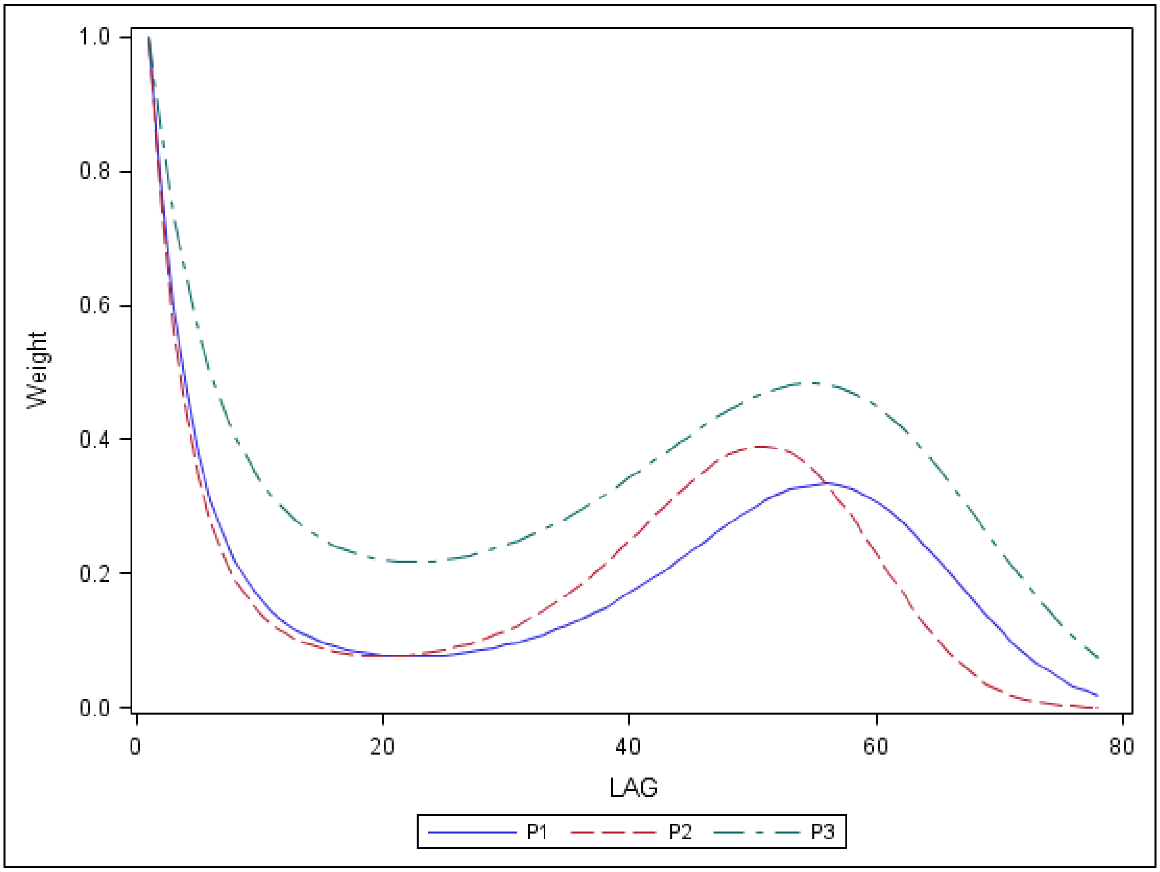

Figure 2: Weights of Best Weekly Model in Three Subsamples

The figure shows the weigths of the best weekly model, i.e. FC0 HYBRID SC TGARCH model, in three subsamples. P1 is for period December, 1986 to November, 1996; P2 for period August, 1989 to July, 1999; P3 for period April, 1993 to March, 2003.

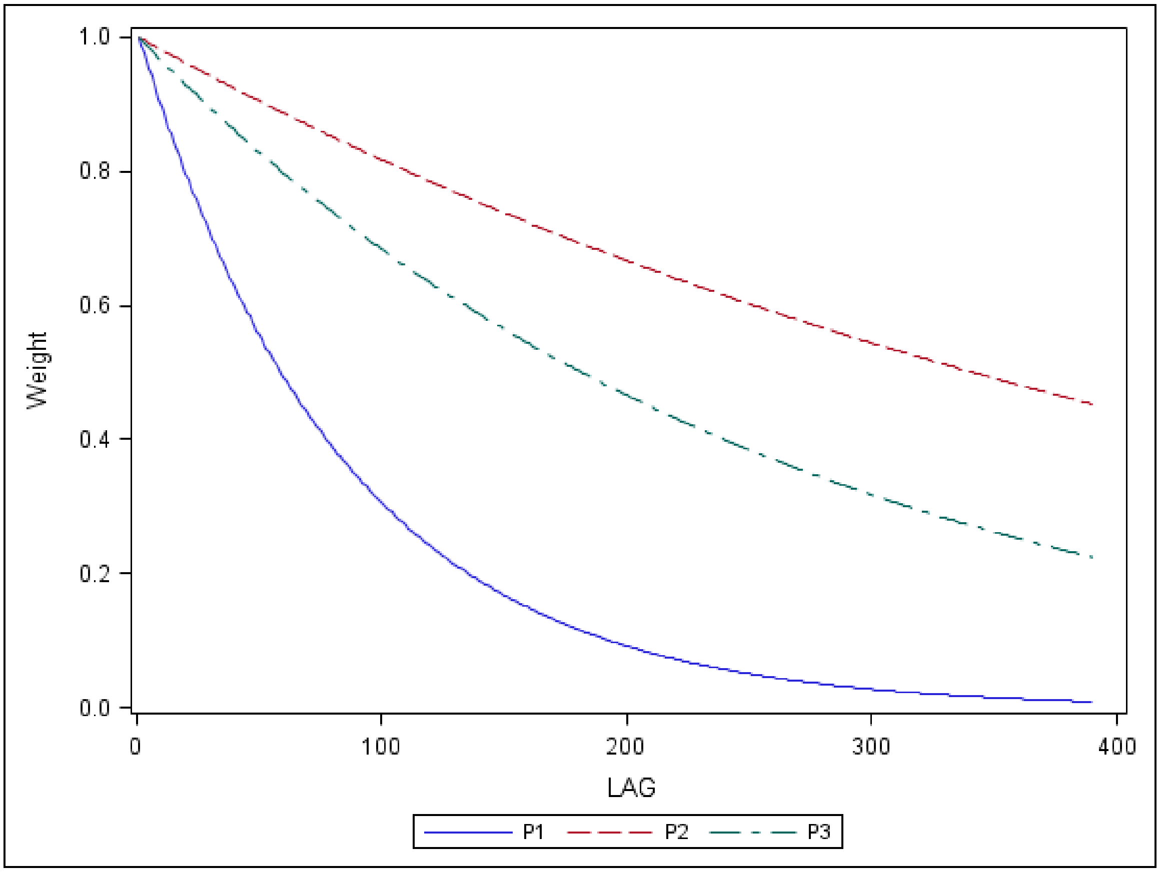

Figure 3: News Impact Curve of Best Daily Model in Three Subsamples

The figure shows the news impact curve of the best daily model, i.e. HYBRID TGARCH model, in three subsamples. P1 is for period December, 1986 to November, 1996; P2 for period August, 1989 to July, 1999; P3 for period April, 1993 to March, 2003.

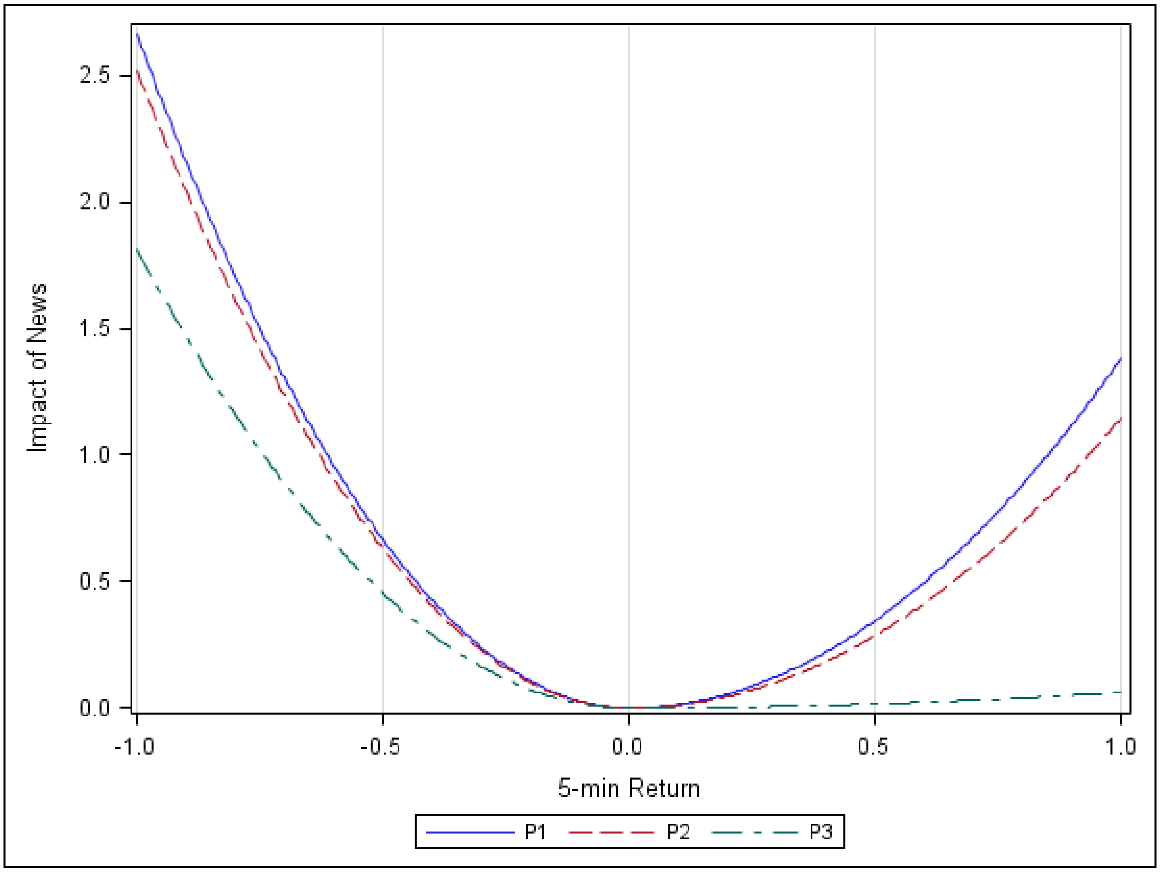

Figure 4: News Impact Curve of Best Weekly Model in Three Subsamples
The figure shows the news impact curve of the best weekly model, i.e. FC0 HYBRID SC TGARCH model, in three subsamples. P1 is for period December, 1986 to November, 1996; P2 for period August, 1989 to July, 1999; P3 for period April, 1993 to March, 2003.

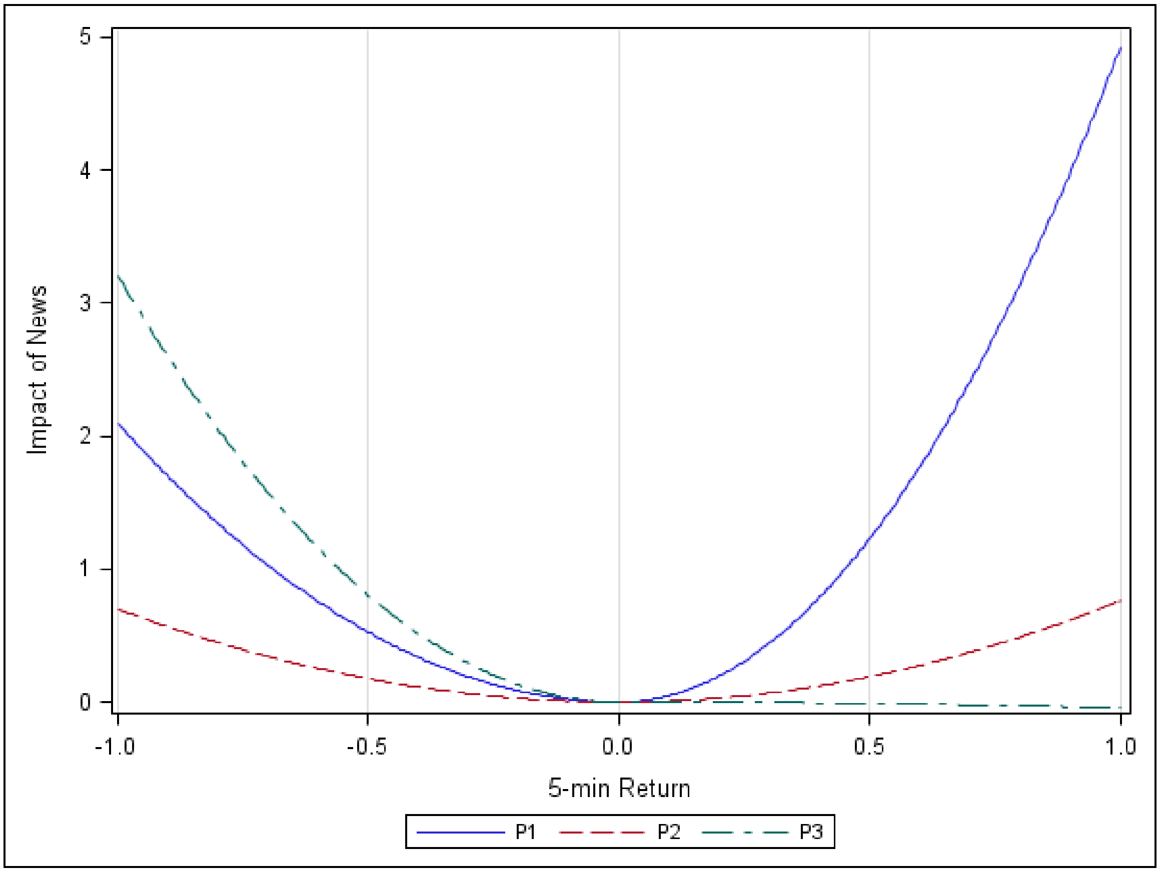

Figure 5: News Impact Curve of Best Bi-weekly Model in Three Subsamples

The figure shows the news impact curve of the best bi-weekly model, i.e. FC1 HYBRID SC GARCH model, in three subsamples. P1 is for period December, 1986 to November, 1996; P2 for period August, 1989 to July, 1999; P3 for period April, 1993 to March, 2003.