HYBRID-GARCH: A Generic Class of Models for Volatility Predictions using High Frequency Data

Xilong Chen† Eric Ghysels‡ Fangfang Wang§

First Draft: April 2009
This Draft: February 14, 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 High Frequency Data-Based Projector-Driven 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. Long horizon volatility 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, monthly, or quarterly GARCH which can then be used to form direct predictions of volatility over the next week, month, or quarter. The second approach is to estimate a daily autoregressive volatility forecasting model and then iterate over the daily forecasts for the necessary number of periods to obtain weekly, monthly, or quarterly predictions of the volatility. 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 say daily squared returns to produce directly multi-period volatility forecasts and can be viewed as a middle ground between the direct and the iterated approaches. The MIDAS volatility literature (see Ghysels and Valkanov (2011)) 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 on 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 feature 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 High Frequency Data-Based Projection-Driven GARCH models as the GARCH dynamics are driven by what we call HYBRID processes. 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 high frequency data structure and finally (3) HYBRID filter 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. However, 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 DGP 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 (2008), Shephard and Sheppard (2009), 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).\footnote{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.}

Besides introducing the new generic class, the paper makes a number of other contributions. First, we provide a jump diffusion framework that underlies the class of HYBRID-GARCH models and show how the presence of jumps affects its parameters.
in population. These results extend earlier work of Drost and Werker (1996). Second, we introduce a variety of new estimators for GARCH-type models that extend various existing estimators, ranging from the QMLE of Bollerslev and Wooldridge (1992) and the MEM procedures of Engle and Gallo (2006), Lanne (2006). Third, one appealing feature of the class of HYBRID-GARCH models is that the sampling frequency of returns becomes a *choice variable*. This is important in empirical applications.

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). We extend the results to potentially misspecified HYBRID GARCH models in Section 5. Section 6 concludes the paper. All technical details are collected in an appendix.

## 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:

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

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 \( H_t \equiv H(\phi, \vec{r}_t) \)

\(^2\)When days spill over into the previous week, we assume \( r_{t-j/m} \equiv r_{t-1-(j-5)/m} \).
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 $\phi$. One can think of at least two possibilities: (1) the weights are treated as additional parameters and estimated as such (with $m$ small this is possible, but not as $m$ gets large), or (2) anchor the weights $\omega_j$ to an underlying daily GARCH(1,1) in which case the parameters $\alpha$, $\beta$ and $\gamma$ and the weights in $\phi$ are jointly related to the assumed daily DGP. The discussion so far implicitly relates to many issues we discuss next.

### 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 (2008), Shephard and Sheppard (2009), 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 (see also later). It should also be noted that typically an extra equation is added, namely consider the case of (daily) realized volatility $RV_t$:

\[
\begin{align*}
V_{t+1|t} &= \alpha + \beta V_{t|t-1} + \gamma RV_t \\
RV_t &= a + b RV_{t-1} + e_t
\end{align*}
\] (2.2)

where the extra equation is added for the purpose of multi-period ahead forecasting. On this topic, it is worth noting that 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 $t$) $H_t \equiv \sum_{j=0}^{K} \omega_j RV_{t-j/K}$, where the weights $\omega_j$ relate to $\gamma$ 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.

\textit{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:

\begin{equation}
V_{t+1|t} = \alpha_m + \beta_m V_{t|t-1} + \gamma_m \sum_{j=0}^{m-1} \beta_i^{j/m} r_{t-j/m}^2, \quad t \in \mathbb{Z} 
\end{equation}

with \( m = 5 \), and where \( \alpha_m, \beta_m \) and \( \gamma_m \) depend on the daily GARCH(1,1) parameters \( \alpha_1, \beta_1 \) and \( \gamma_1 \) (see more details later in equation (3.6)). 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 \( \phi \). This will have consequences when we elaborate on the estimation of HYBRID GARCH models. Indeed, the context of temporal aggregation precludes us from using, say standard QMLE methods, a topic that will be discussed later.

\textit{HYBRID Filtering Processes}

Here the HYBRID process \( H(\phi, \vec{r}_t) \) in (2.1) involves parameters that are not explicitly related to \( \alpha, \beta \) and \( \gamma \) 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 \( \phi \) are estimated jointly with the volatility dynamics parameters. The choice of the parameterizations is inspired by Chen and Ghysels (2009). 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 for instance in Bollerslev and Wooldridge (1992).

\textit{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 (2009) 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 (2009), except that the latter use a MIDAS regression format. The HYBRID processes we consider are of the following type:

$$H_t(\phi) = \sum_{j=0}^{m-1} \Psi_j(\phi_1) NIC(\phi_2, r_{t-j/m}), \quad \sum_{j=0}^{m-1} \Psi_j(\phi_1) = 1 \quad (2.4)$$

where $NIC(\phi_2, \cdot)$ stands for a high frequency data news impact curve. The parameter vector $\phi$ combines both $\phi_1$ and $\phi_2$, the latter being 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 (2009), namely:

$$NIC(\phi_2, r) = b(r - c)^2 \quad (2.5)$$
$$NIC(\phi_2, r) = br^21_{r \geq 0} + cr^21_{r < 0} \quad (2.6)$$

with $b$ and $c$ the parameters that are in the sub-vector $\phi_2$.

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 $\theta_0$. 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 $\theta$ (denoted as $d$) depends on the specific circumstances considered. For this generic setting we have the following estimators:

$$\hat{\theta}_{T}^{mdr} = \arg \min_{\theta \in C} \frac{1}{T} \sum_{t=1}^{T} (RV_t - V_{t|t-1}(\theta))^2$$

where $C$ is a convex compact subset of $\Theta$ such that $\theta_0$ is in the interior of $C$. 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:

$$\hat{\theta}_{T}^{mdr2} = \arg \min_{\theta \in C} \frac{1}{T} \sum_{t=1}^{T} (R^2_t - V_{t|t-1}(\theta))^2$$

The above estimator has a likelihood-based version, namely:

$$\hat{\theta}_{T}^{thr2} = \arg \min_{\theta \in C} \frac{1}{T} \sum_{t=1}^{T} \left( \log V_{t|t-1}(\theta) + \frac{R^2_t}{V_{t|t-1}(\theta)} \right)$$

requiring 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 $\hat{\theta}_{T}^{mdr2}$ 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:

$$\hat{\theta}^{lhrv}_T = \arg\min_{\theta \in C} \frac{1}{T} \sum_{t=1}^{T} \left( \log V_{t|t-1}(\theta) + \frac{RV_t}{V_{t|t-1}(\theta)} \right)$$

The choice of $R^2$ versus $RV$ in $\hat{\theta}_T^{mdr2}$ versus $\hat{\theta}_T^{mdrv}$ and $\hat{\theta}_T^{lhr2}$ versus $\hat{\theta}_T^{lhrv}$ 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

$$RV_{t+1} = \sigma^2_{t+1|t} \eta_{t+1}$$

where $\eta_{t+1}$ is independent and identically distributed with mean 1 and $\sigma^2_{t+1|t}$ 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 $\sigma^2_{t+1|t}$ and $RV_{t+1}$). Suppose the cumulative distribution function of $\eta$ 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 $L^2(\Omega, \mathcal{F}, P)$, $P(X|\mathcal{I})$ 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 entrywise, then we write $A < \infty$. Moreover, to emphasize the role of $\phi$ in the process $H(\phi, \tilde{r}_t)$ we will use the notation $H(\phi, \tilde{r}_t) \equiv H_t(\phi)$.

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 $RV_{t+1} = \sum_{j=0}^{m-1} r_{t+1-j/m}$ 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, $\sigma^2_{t+1|t} = P_t(RV_{t+1}|\mathcal{I}_t)$. Therefore we 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 in terms of 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 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., $\sigma(r_s, t-m-1+1/m \leq s \leq t+m)$.

\(^3\)For convenience we collected all the regularity conditions in Appendix A.
Assumption 3.1 (Scenario 1). \( \mathcal{I}_t = \mathcal{L}_t \), the closed span of \( \{1, r_{t-k/m}, r_{t-k/m}^2; k = 0, 1, 2 \ldots \} \) and \( P_t(r_s|\mathcal{L}_{s-1/m}) = 0 \). Therefore \( \sigma^2_{t+1|t} \) is the best linear predictor.

Assumption 3.2 (Scenario 2). \( \mathcal{I}_t = \mathcal{F}_{t-\infty}^t \), the sigma field generated by the high frequency returns up to time \( t \) and \( P_t(r_s|\mathcal{F}_{s-1/m}^{-\infty}) = 0 \). The prediction equations therefore indicate that

\[
E(r_s|\mathcal{F}_{s-1/m}^{-\infty}) = 0, \quad E(RV_{t+1}|\mathcal{F}_{t}^{t}) = E(R^2_{t+1}|\mathcal{F}_{t}^{t}) = \sigma^2_{t+1|t} \quad (3.1)
\]

where \( R_{t+1} \equiv \sum_{j=0}^{m-1} r_{t+1-j/m} \).

Since we use \( V_{t+1|t} \) driven by \( H(\phi, \tilde{r}_t) \) to mimic the dynamics of \( \sigma^2_{t+1|t} \), \( H_t \) is required to satisfy Assumption A.4, and under 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 extremum estimators. A necessary condition is that the dimension of \( \phi \) is not larger than \( m \). A more detailed discussion on Assumption A.4 is provided in Appendix A.1. 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:

\[
H_t(\phi) = \sum_{j=0}^{m-1} \Psi_j(\phi)r_{t-j/m}^2, \quad \sum_{j=0}^{m-1} \Psi_j(\phi) = 1 \quad (3.2)
\]

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 (2009) which were inspired by MIDAS regression format of Ghysels, Santa-Clar, 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 (2010).

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 since \( r_{t-j/m}1_{r_{t-j/m}<0} \) and \( r_{t-j/m}1_{r_{t-j/m}\geq0} \) are

---

4 See Brockwell and Davis (1991).
not contained in the linear projection. 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 \( \{r_s\} \) satisfy 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 \( L_s \), i.e., \( P_l(r_{s+1/m}|L_s) \) = 0, for \( s \), which is the orthogonal projection of \( r^2_{s+1/m} \) onto \( L_s \) satisfies

\[
\sigma^2_{s+1/m|s} = a + b\sigma^2_{s-1/m} + cr^2_s.
\] (3.3)

With some algebra one obtains:

\[
\sigma^2_{s+k/m|s} = a \frac{1 - (b + c)^{k-1}}{1 - (b + c)} + (b + c)^{k-1}\sigma^2_{s+1/m|s}
\] (3.4)

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:

\[
V_{t+1|t} = \alpha_m + \beta_m V_{t|t-1} + \gamma_m \sum_{j=0}^{m-1} \beta^{j/m}_m r_{t-j/m}
\] (3.5)

where

\[
\alpha_m = \alpha_1 \frac{1-\beta^m_m}{1-\beta_1} \frac{m(1-\beta_1) - \gamma d_m}{1-(\beta_1 + \gamma_1)}
\]

\[
\beta_m = \beta^m_1
\]

\[
\gamma_m = \gamma_1 d_m
\] (3.6)

and \( d_m = (1 - (\beta_1 + \gamma_1)^m)/(1 - (\beta_1 + \gamma_1)) \). Clearly, (3.5) is of the form (2.1) with \( H_t = \sum_{j=0}^{m-1} \beta^{j/m}_m r_{t-j/m}^2 \), and \( H_t \) is referred to as structural HYBRID process.

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 (3.6). A direct implication of the relationship (3.6) 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.

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 processes. The HYBRID filtering process can be viewed as an extension of the structural HYBRID process by allowing for more flexible return dynamics.

### 3.4 Diffusions, Jumps and HYBRID Processes

So far the HYBRID processes pertained to discretely sampled returns at different frequencies. We turn our attention 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., $H(\phi, \bar{r}_t) = RV_t$ (see the first equation in (2.2)). Moreover, we will also 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,

\[
\begin{align*}
    dp_t &= \sigma_t dL_t \\
    d\sigma_t^2 &= \theta(\omega - \sigma_t^2)dt + \sqrt{2\lambda\theta}\sigma_t^2 dB_t \\
    L_t &= \sqrt{1-\eta}W_t + \sqrt{\eta}N_t
\end{align*}
\]  

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 density of a normal distribution with mean 0 and variance $1/\zeta$. Moreover, $B_t$, $W_t$ and $N_t$ are independent of each other. Note that $EL_t = EL_t^2 = 0$, $EL_t^2 = t$, and $EL_t^4 = 3t^2 + 3t\eta^2/\zeta$. The discretely sampled returns $r_s = p_s - p_{s-1/m}$ from (3.7) follows a weak GARCH(1,1) appearing in equation (3.3) as discussed in Drost and Werker (1996). The relation between $(\theta, \omega, \lambda, v_L^*)$ in (3.7) and $(a, b, c, k)$ in (3.3) 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 $r_s$.

We turn our attention now to some prediction formulas associated with this framework. Note first 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 $L_t$ relate to prediction of $E_t([p, p]_{t+1} - [p, p]_t)$. In this subsection,

---

$^5$Note that $(a, b, c, k)$ and $r_s$ depend on $m$. 

12
we write $V_{t+1|t}^{(m)}$ 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:

$$
E_t([p, p]_{t+1} - [p, p]_t) = (1 - \eta) \int_t^{t+1} E_t(\sigma^2_s)ds + \eta \int_t^{t+1} \int_{-\infty}^{\infty} E_t(\sigma^2_s) y^2 \zeta f(dy)ds \\
= \omega \left(1 - \theta^{-1}(1 - e^{-\theta})\right) + \theta^{-1}(1 - e^{-\theta}) \sigma_t^2. \quad (3.8)
$$

On the other hand, the forecast using $\mathcal{L}_t$ yields the HYBRID GARCH equation appearing in equation (3.5).

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 $[p, p]_{t+1} - [p, p]_t$, it does not imply that the HYBRID GARCH process $V_{t+1|t}^{(m)}$ is a consistent estimator of $E_t([p, p]_{t+1} - [p, p]_t)$ as well. Based on 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. $v^*_L > 0$,

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

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 $\sqrt{m} \sum_{j=0}^{m-1} \beta_j/m \tau_{t-j/m}^2$ converges to $(\theta \lambda)^{-1/2} \sigma_t^2$ in $L^2$.

Proof: see Appendix B.2.

When we compare equations (3.5) and (3.8) we note from Proposition 3.1 that $V_{t+1|t}^{(m)}$ 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 $V_{t+1|t}^{(m)}$ does not provide a consistent estimate of $E_t([p, p]_{t+1} - [p, p]_t)$ because the limit of $V_{t+1|t}^{(m)}$ 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 (3.7) as the DGP, the process \( \{V_{t+1|t}^{(m)}, t\}_{m \geq 1} \) defined by equation (3.5) converges to \( \{E_t([p, p]_{t+1} - [p, p]_t), t\} \) uniformly on compacts in probability if and only if there are no jumps in the price process.\(^6\)

Proof: see Appendix B.2.

Note that without jumps, the HYBRID GARCH process still involves intra-period weighted returns, namely equation (3.5) has intra-period weights that are powers of $\beta_m$. Furthermore, it follows from the proof of Proposition 3.1 that what drives the HYBRID process as $m \to \infty$, is the instantaneous volatility $\sigma_t^2$, not the integrated process estimated by RV. The instantaneous volatility $\sigma_t^2$ can be consistently estimated by that very same intra-period weighted sum $mc \sum_{j=0}^{m-1} b^j r_{t-j/m}^2$. Put differently, we can view the HYBRID process as a spot volatility estimator 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 processes 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 processes accordingly. In the rest of the paper, we will work exclusively with the returns sampled at fixed frequency without referring to an explicit DGP and make

\(^6\)See Protter (2004) for the definition.
the assumption that the returns are strictly stationary and ergodic (formally stated as Assumption A.2 in Appendix A).

A natural approach to identify parameters and hence estimate the HYBRID GARCH model is to make use of the projection equation (A.1), which yields the minimum-distance estimation and it is applied to both scenarios. For Scenario 2, we also consider the likelihood estimation and the estimation via Multiplicative Error Models. So the purpose of this section is to provide the asymptotic analysis of various estimators. In addition, we also include 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.

4.1 Estimation under Scenario 1

We use the HYBRID GARCH process $V_{t|t-1}(\theta)$ solved from equation (2.1) to approximate $\sigma^2_{t|t-1}$. It is necessary to define the distance between the two time series, i.e., $V(\theta) \equiv (V_{t|t-1}(\theta), t \in \mathbb{N})$ and $\sigma^2 \equiv \{\sigma^2_{t|t-1}, t \in \mathbb{N}\}$. Consider the following metric on a countable product of $L^2(\Omega, \mathcal{F}, P)$:

$$d(X, Y) = \sum_{t=1}^{\infty} 2^{-t} \min(\|X_t - Y_t\|_2, 1),$$

where $X = (X_t, t \geq 1)$ and $Y = (Y_t, t \geq 1)$. $d(X, Y) = 0$ if and only if $\|X_t - Y_t\|_2 = 0$ for any $t$. Note that $\|RV_t - V_{t|t-1}(\theta)\|_2 - \|RV_t - \sigma^2_{t|t-1}\|_2 = \|V_{t|t-1}(\theta) - \sigma^2_{t|t-1}\|_2$ for all possible $\theta$'s due to Assumption A.1. Therefore $0 \leq d(RV, V(\theta)) - d(RV, \sigma^2) \leq d(V(\theta), \sigma^2)$ hence measures our ignorance of the true dynamics. $d(V(\theta), \sigma^2)$ is 0 when equation (2.1) correctly describes the dynamics of $\sigma^2_{t|t-1}$. This is true for the structural HYBRID, where one is able to find a suitable $\theta$ so that $d(V(\theta), \sigma^2) = 0$. In this section we will focus on the situation where the HYBRID GARCH can be correctly specified and leave the topic of mis-specification to Section 5.

For $H$ satisfying Assumption A.4, we assume there exists $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \phi_0)$ such that $d(V(\theta_0), \sigma^2) = 0$, or equivalently,

$$V_{t|t-1}(\theta_0) = \sigma^2_{t|t-1} \quad \forall t \in \mathbb{N}. \quad (4.1)$$

The parameters satisfy $\alpha_0 > 0, 0 < \beta_0 < 1, \gamma_0 > 0, \phi_0 \in \Phi$ where $\Phi$ is a connected set that collects all the possible values of $\phi$'s such that $H_t(\phi)$ meets Assumption A.4. The parameter space is then defined as $\Theta = \{ (\alpha, \beta, \gamma, \phi) : \alpha > 0, 0 < \beta < 1, \gamma > 0, \phi \in \Phi \}$.

Theorem 4.1. Suppose that Assumptions A.1 and A.2 hold. Under Scenario 1 (i.e.,
Assumption 3.1),
\[ \theta_0 = \arg \min_{\theta \in \mathcal{C}} E \left( R V_t - V_{t|t-1}(\theta) \right)^2 \]  
(4.2)

where \( \mathcal{C} \) is a convex compact subset of \( \Theta \) satisfying \( \theta_0 \in \mathcal{C}^0 \subset \mathcal{C} \subset \Theta^0 \).

Proof: see Appendix B.3

Theorem 4.1 ensures the ‘true’ parameter \( \theta_0 \) is identifiably unique.\(^7\) Hence a natural estimator for \( \theta_0 \) is the minimizer of

\[ \min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left( R V_t - \tilde{V}_t(\theta) \right)^2 \]  
(4.3)

based on observations \( \{ r_{1/m}, r_{2/m}, \ldots, r_1, \ldots, r_{T-1+1/m}, \ldots, r_T \} \), where \( \tilde{V}_t \) is defined recursively by

\[ \tilde{V}_t(\theta) = \alpha + \beta \tilde{V}_{t-1}(\theta) + \gamma H_{t-1}(\phi), t \geq 1 \quad \text{and} \quad \tilde{V}_0 = \tilde{v} \]  
(4.4)

and \( \tilde{v} \) is any arbitrary deterministic value. The minimizer of (4.3) exists due to Jennrich (1969) and Gallant and White (1988) and is denoted by \( \hat{\theta}_{mdrv}^T \), minimum-distance RV-based estimator. The following theorem indicates that \( \hat{\theta}_{mdrv}^T \) is identifiably unique and it will converge to \( \theta_0 \) a.s.

**Theorem 4.2.** Suppose that Assumptions A.1 and A.2 hold. Under Scenario 1 \( \hat{\theta}_{mdrv}^T \) is identifiably unique and it is a strongly consistent estimator of \( \theta_0 \).

Proof: see Appendix B.4

Now we turn our attention to the asymptotic distribution of \( \hat{\theta}_{mdrv}^T \). Define \( \varepsilon_t(\theta) = R V_t - V_{t|t-1}(\theta) \). Note that under suitable moment condition \( \varepsilon_t \partial_k \varepsilon_t \) is near epoch dependent (NED) on the underlying return process (see Lemma B.5), and is therefore a mixingale when the return process is \( \alpha \)-mixing (see Assumption A.3). More precisely, we have \( \| E(\varepsilon_t \partial_k \varepsilon_t | \mathcal{F}_m^{t-m} ) \|_2 \leq \psi_m C_0 \) where \( \psi_m = \rho^{m/2} + \alpha([m/2])^{v_2/(2+2v_2)} \) for some constants \( 0 < \rho < 1 \) and \( C_0 > 0 \).\(^8\) Therefore the asymptotic normality of \( \hat{\theta}_{mdrv}^T \) can be derived using mixingale central limit theorem. First of all, we need to show that the asymptotic variance-covariance matrix exists.

**Proposition 4.1.** For any \( k = 1, 2, \ldots, d \), (the dimension of \( \theta \)) under Assumptions A.3 and 3.1, \( \lim_{T \to \infty} \text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t(\theta_0) \partial_k \varepsilon_t(\theta_0) \right) \) exists and is finite.

\(^7\)See Gallant and White (1988) for the definition of identifiable uniqueness.

\(^8\)This is derived using Theorem 3.1 of McLeish (1975).
Proof: see Appendix B.4

Use $\Omega^{mdrv}$ to denote the limiting variance-covariance matrix

$$\lim_{T \to \infty} \text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t(\theta_0) \nabla \varepsilon_t(\theta_0) \right).$$

When $\Omega^{mdrv}$ is positive definite, $\tilde{\theta}_T^{mdrv}$ will be asymptotically normal which is stated in the following theorem:

**Theorem 4.3.** Suppose Assumptions A.1, A.3 hold and $\Omega^{mdrv} > 0$. Under Scenario 1

$$\sqrt{T}(\tilde{\theta}_T^{mdrv} - \theta_0) \Rightarrow N(0, (\Sigma^{md})^{-1} \Omega^{mdrv} (\Sigma^{md})^{-1}),$$

where $0 < \Sigma^{md} = E \nabla V_{t|t-1}(\theta_0) (\nabla V_{t|t-1}(\theta_0))^\prime < \infty$.

Proof: see Appendix B.4

The proof is an application of mixingale central limit theorem of Davidson (1992). It is worth noting 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.9

### 4.2 Estimation under Scenario 2

In this section we 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 the previous section, and quasi-maximum likelihood estimators that are standard in the GARCH literature (see e.g. Bollerslev and Wooldridge (1992)). More specifically, let us consider the following estimators:

$$\tilde{\theta}_T^{mdrv} = \arg \min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left( RV_t - \tilde{V}_t(\theta) \right)^2$$  \hspace{1cm} (4.5)

$$\tilde{\theta}_T^{mdrv}_2 = \arg \min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left( R^2_t - \tilde{V}_t(\theta) \right)^2$$  \hspace{1cm} (4.6)

$$\tilde{\theta}_T^{hvr2} = \arg \min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left( \log \tilde{V}_t(\theta) + \frac{R^2_t}{\tilde{V}_t(\theta)} \right)$$  \hspace{1cm} (4.7)

$$\tilde{\theta}_T^{hvr} = \arg \min_{\theta \in \mathcal{C}} \frac{1}{T} \sum_{t=1}^{T} \left( \log \tilde{V}_t(\theta) + \frac{RV_t}{\tilde{V}_t(\theta)} \right)$$  \hspace{1cm} (4.8)

---

9See page 24 of Gallant and White (1988) for the definition of size.
with $\tilde{V}_t$ defined in (4.4). We also consider the estimator derived using Multiplicative Error Model which shares some similarities with the likelihood-RV-based estimator, $\tilde{\theta}^{thrv}_T$.

### 4.2.1 Minimum Distance and Quasi-Likelihood Estimators

We start by extending Theorem 4.1 to the case of Scenario 2:

**Theorem 4.4.** Suppose that Assumptions A.1, A.2, and A.5(2) hold. Under Scenario 2 (i.e. Assumption 3.2):

$$\theta_0 = \arg\min_{\theta \in \mathcal{C}} E \left( RV_t - V_{t|t-1}(\theta) \right)^2 = \arg\min_{\theta \in \mathcal{C}} E \left( R_t^2 - V_{t|t-1}(\theta) \right)^2.$$  \hspace{1cm} (4.9)

**Proof:** see Appendix B.3

**Theorem 4.5.** Suppose that Assumptions A.1, A.2, and A.5(2) hold. Under Scenario 2, $\tilde{\theta}^{mdrv}_T$ and $\tilde{\theta}^{mdr}_2$ are identifiably unique and they converge to $\theta_0$ a.s. Further assume that $Er^8 < \infty$ and Assumption A.5(3) holds, then

$$\sqrt{T}(\tilde{\theta}^{mdrv}_T - \theta_0) \rightarrow N \left( 0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1} \right)$$

$$\sqrt{T}(\tilde{\theta}^{mdr}_2 - \theta_0) \rightarrow N \left( 0, (\Sigma^{md})^{-1}\Omega^{mdr_2}(\Sigma^{md})^{-1} \right)$$

where

$$0 < \Sigma^{md} = E[\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'] < \infty, \hspace{1cm} (4.10)$$

$$0 < \Omega^{mdrv} = E[(RV_t - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'] < \infty, \hspace{1cm} (4.11)$$

$$0 < \Omega^{mdr_2} = E[(R_t^2 - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'] < \infty \hspace{1cm} (4.12)$$

**Proof:** see Appendix B.5

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

**Theorem 4.6.** Suppose $E(\sup_{\phi \in \Phi} H(\phi, \tilde{r}_t))^2 < \infty$ and Assumptions A.1 and A.2 hold. Under Scenario 2 (i.e. Assumption 3.2):

$$\theta_0 = \arg\min_{\theta \in \mathcal{C}} E \left( \log V_{t|t-1}(\theta) + \frac{R_t^2}{V_{t|t-1}(\theta)} \right) = \arg\min_{\theta \in \mathcal{C}} E \left( \log V_{t|t-1}(\theta) + \frac{RV_t}{V_{t|t-1}(\theta)} \right).$$  \hspace{1cm} (4.13)

**Proof:** see Appendix B.3

18
Theorem 4.7. Suppose $E(\sup_{\phi \in \Phi} H(\phi, \tilde{r}_i))^2 < \infty$ and Assumptions A.1 and A.2 hold. Under Scenario 2, $\hat{\theta}^{thrv}_T$ and $\tilde{\theta}^{thrv}_T$ are identifiably unique and they converge to $\theta_0$ a.s. Further assume $E(r^{4+v}) < \infty$ for some $v > 0$ and inequality (B.1) holds,

$$\sqrt{T}(\hat{\theta}^{thrv}_T - \theta_0) \Longrightarrow N\left(0, (\Sigma^{lh})^{-1} \Omega^{thr}(\Sigma^{lh})^{-1}\right)$$

$$\sqrt{T}(\tilde{\theta}^{thrv}_T - \theta_0) \Longrightarrow N\left(0, (\Sigma^{lh})^{-1} \Omega^{thr}(\Sigma^{lh})^{-1}\right)$$

where

$$0 < \Sigma^{lh} = E\left(V_{t|t-1}^{-2}(\theta_0)\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'\right) < \infty,$$  \hspace{1cm} (4.14)

$$0 < \Omega^{thr} = E\left(V_{t|t-1}^{-4}(\theta_0)(R^2_i - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'\right) < \infty$$  \hspace{1cm} (4.15)

$$0 < \Omega^{thr} = E\left(V_{t|t-1}^{-4}(\theta_0)(RV_i - V_{t|t-1}(\theta_0))^2\nabla V_{t|t-1}(\theta_0)\nabla V_{t|t-1}(\theta_0)'\right) < \infty.$$  \hspace{1cm} (4.16)

Proof: see Appendix B.6

The likelihood estimation introduced in this paper is slightly different from the one discussed in the literature. First of all, $\sigma^2_{t|t-1}$ is studied in $L^2(\Omega, \mathcal{F}, P)$ instead of $L^1(\Omega, \mathcal{F}, P)$, therefore we need the assumption $E(\sup_{\phi \in \Phi} H(\phi, \tilde{r}_i))^2 < \infty$ in Theorems 4.6 and 4.7. However, the properties derived for likelihood estimators can carry over to $L^1(\Omega, \mathcal{F}, P)$ by simply assuming $E \sup_{\phi \in \Phi} H(\phi, \tilde{r}_i) < \infty$.\(^{10}\) Secondly, $\hat{\theta}^{thrv}_T$ is different from the traditional quasi-MLE in that the objective function appearing in (4.7) 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 $R_1, R_2, \ldots, R_{t-1}$, $\tilde{V}_t(\theta) + R^2_t/\tilde{V}_t(\theta)$ 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 properties stated in Theorems 4.5 and 4.7 can apply to periodic time series by replacing Assumption A.2 with an assumption that the return process is strictly periodically stationary and periodically ergodic.\(^{11}\) This is because proofs of Theorems 4.5 and 4.7 only require $\tilde{r}_t$ to be strictly stationary ergodic.

Under scenario 2, we have four estimators: $\hat{\theta}^{mdrv}_T, \hat{\theta}^{mdrv}_T, \hat{\theta}^{thrv}_T$ and $\tilde{\theta}^{thrv}_T$. The likelihood-based estimators are superior to the minimum-distance ones in that they require weaker moment conditions to establish asymptotic normality. However, it is not straightforward to compare the efficiency between $R^2$-based estimator and $RV$-based estimator (i.e., $\hat{\theta}^{mdrv}_T$ v.s. $\hat{\theta}^{mdrv}_T$, and $\hat{\theta}^{thrv}_T$ v.s. $\tilde{\theta}^{thrv}_T$), because the sign

\(^{10}\)The lemmas appearing in Appendix B.6 are stated under this assumption.

\(^{11}\)See Aknouche and Bibi (2009) for the definition.
of \(E_{t-1}\left[(R_t^2 - V_{t|t-1}(\theta)) - (RV_t - V_{t|t-1}(\theta))^2\right] = E_{t-1}(R_t^4 - RV_t^2)\) 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 Theorem 4.5 and Theorem 4.7, \(\hat{\theta}_T^{mdr}\) (or \(\hat{\theta}_T^{lhrv}\)) has a smaller asymptotic variance than \(\tilde{\theta}_T^{mdr^2}\) (or \(\tilde{\theta}_T^{lhr^2}\)).

**Proof:** see Appendix B.7

### 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 \(\eta\) appearing in (2.7) to formulate the estimation of \(\theta_0\). The estimator is then denoted by \(\hat{\theta}_T^{mem}\).

Suppose that the appropriate probability distribution for the error term is a Gamma distribution. The conditional density of \(RV_{t+1}\) is

\[
f(RV_{t+1}|F_{t-\infty}) = \frac{1}{\Gamma(g)} g^g \frac{RV_{t+1}^{g-1}}{(\sigma_{t+1|t}^2)^g} \exp\left(-g RV_{t+1}/\sigma_{t+1|t}^2\right).
\]

Hence \(E(RV_{t+1}|F_{t-\infty})=\sigma_{t+1|t}^2\), and \(Var(RV_{t+1}|F_{t-\infty})=\sigma_{t+1|t}^4/g\). 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 \(\theta_0\) and \(g\) are asymptotically independent, and the point estimation \(\hat{\theta}_T^{mem}\) of \(\theta_0\) is same as \(\hat{\theta}_T^{lhrv}\). Hence we have

\[
\sqrt{T}(\hat{\theta}_T^{mem} - \theta_0) \longrightarrow N\left(0, (g \Sigma^h)^{-1}\right),
\]

which follows from the proof of \(\hat{\theta}_T^{lhrv}\) directly.

If the existence of an appropriate parametric density function can not be verified, one can consider quasi-likelihood estimation which will yield the same asymptotic result as \(\hat{\theta}_T^{lhrv}\) (see Theorem 4.7). Because the innovation \(\eta\) is independent, \(\Omega^{lhrv}\) in Theorem 4.7 becomes \(E(RV_t^2/V_{t|t-1}(\theta_0) - 1)\Sigma^h\) and hence we do not need the moment condition \(Er^{4+v} < \infty\) to establish the asymptotic normality of \(\hat{\theta}_T^{mem}\). Its asymptotic variance-covariance matrix becomes \((E\eta_t^2 - 1)(\Sigma^h)^{-1}\).
4.3 Simulation study

We conclude the section with a simulation study which helps us better understand the small-sample properties of the estimators discussed in this section. We first describe the design, followed by a discussion of the simulation results.

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

\[ r_{s+1/m} = \sqrt{v_{s+1/m}} \varepsilon_{s+1/m}, \quad v_{s+1/m} = a + bv_{s|s-1/m} + cr_s^2, \quad \varepsilon_{s+1/m} \sim \mathcal{N}(0,1), \] (4.17)

The values of parameters in model (4.17) are reported in table 4.17, while in model (4.18) their values are fixed at \( \omega = 0.0350, \lambda = 0.2962 \). The discretely-sampled high frequency return \( r-s=p_s-p_{s-1/m} \) is therefore a weak GARCH(1,1) as below:

\[ \sigma^2_{s+1/m|s} = a + b\sigma^2_{s|s-1/m} + cr^2_s, \sigma^2_{s+1/m|s} = P_t(v^2_{s+1/m}|\mathcal{L}_s), P_t(r_{s+1/m}|\mathcal{L}_s) = 0 \] (4.18)

with \( a = \omega(1 - e^{-\theta/m})/m, \ c = e^{-\theta/m} - b, \) and \( |b| < 1 \) is the solution to

\[ \frac{b}{1+b^2} = \frac{\rho e^{-\theta/m} - 1}{\rho(1 + e^{-2\theta/m}) - 2}, \] where \( \rho = \frac{4(e^{-\theta/m} - 1 + \theta/m) + 2\theta/m(1 + \theta/m(1 - \lambda)/\lambda)}{1 - e^{-2\theta/m}}. \]

Based on either Model (4.17) or Model (4.18), we construct the HYBRID GARCH process \( V_{t+1|t} = \alpha_m + \beta_m V_{t|t-1} + \gamma_m \sum_{j=0}^{m-1} \beta^m_j r^2_{t-j/m} \) with \( \alpha_m = a \frac{a^{1-b_m} m(1-b-m) - cd(m)}{1-(b+c)}, \beta_m = b^m, \gamma_m = cd(m) \) and \( d(m) = (1 - (b + c)^m)/(1 - (b + c)). \)

The values of parameters in model (4.17) are \( a = 2.8E - 06, b = 0.9770, c = 0.0225 \) which are reported in table 1 with \( m = 5, 78 \) and \( 288 \) for model (4.17) and \( m = 24, 144, 288 \) for model (4.18). Note that \( a, b, c \) are fixed across different values of \( m \) in model (4.17), while in model (4.18) their values are different for different \( m \).

The estimators considered are: \( \hat{\theta}_T^{mdr}, \) the minimizer of (4.3), and the companion estimator \( \hat{\theta}_T^{mdr^2} \), replacing \( RV \) by \( R^2 \), as well as (quasi-)likelihood-based estimators
\(\tilde{\theta}_T^{\text{drv}}\) defined in (4.7), and \(\tilde{\theta}_T^{\text{lhrv}}\), defined in (4.8). 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 \(\theta_0\) and \(g\) are asymptotically independent, and thus \(\tilde{\theta}_T^{\text{mem}}\) is asymptotically the same as \(\tilde{\theta}_T^{\text{lhrv}}\). The purpose of this section is to examine differences in small sample behavior. To streamline the discussion we will refer to the estimators \(\tilde{\theta}_T^{\text{mdrv}}\), \(\tilde{\theta}_T^{\text{mdr}2}\), \(\tilde{\theta}_T^{\text{lhrv}}\), \(\tilde{\theta}_T^{\text{lhr}2}\) and \(\tilde{\theta}_T^{\text{mem}}\) as respectively MDRV, MDR2, LHRV, LHR2 and MEM. The numbers in parenthesis in Table 2 and Table 3 are MSE for LHR2, relative MSE (with respect to LHR2) for LHRV, MDR2, MDRV, and MEM. For \(g\) (in the MEM estimator), we only report the sample variance.

In the simulation experiment, we consider 1000 replications of sample path (3.5), 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 \(dt = 1/86400\). The simulation results for strong GARCH(1,1) and weak GARCH(1,1) are reported in Table 2 and Table 3 respectively.

The results in Table 2 are quite easy to summarize. The bold-faced entries between parentheses indicate the best estimator for the various parameters. The estimator that appears to have the best finite sample properties is LHRV. 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 see 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 \(\alpha_m\). 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.

Table 3 tells a similar story. LHRV dominates almost all the other estimators, especially the \(R^2\)-based ones. However when \(m = 144\), the data tends to choose MDRV for the estimation of \(\alpha_m\) and MEM for \(\gamma_m\). For \(\beta_m\), both MEM and LHRV perform better than the others.

The simulation results inspired us to use LHRV and MEM as the principal estimators in the empirical analysis.
5 Misspecified HYBRID GARCH processes

The distance between \( V(\theta) \equiv (V_{t-1}(\theta), t \in \mathbb{N}) \) and \( \sigma^2 \equiv (\sigma^2_{t-1}, t \in \mathbb{N}) \), denoted by \( d(V(\theta), \sigma^2) \), describes our ignorance about the true dynamics. In Section 4, we assumed \( d(V(\theta), \sigma^2) = 0 \) for a suitable choice of \( \theta \). This is not always the case especially when the DGP is unknown. Consider the situation where \( V_{t+1|t} = \alpha + \beta V_{t|t-1} + \gamma RV_t \) is estimated instead of a linearly weighted high frequency return HYBRID process, as in equation (3.2). When the underlying DGP is unknown, it is hard to tell which one is the true description of \( \sigma^2_{t|t-1} \). Most likely, both of them are misspecified. Another example pertains to ignored asymmetries. Suppose the news impact curve is as in (2.5) with \( c \neq 0 \). What happens to the estimation of HYBRID GARCH model involving a linear weighting scheme with squared returns in that case?

In this section we study the potential misspecification of HYBRID GARCH processes and how this affects model estimation. The discussion follows the work of Domowitz and White (1982) and Bates and White (1985). But slightly different from their work, we focus on a situation where the optimand depends on all previous observations which are realized by mixing processes. As before, the high frequency returns are assumed to satisfy Assumption A.2 throughout this section.

Estimation under Scenario 1

Given \( H \) that satisfies Assumption A.4, \( d(V(\theta), \sigma^2) \) has a minimum over \( C \), say at \( \theta_* \), where \( C \) is a convex compact subset of \( \Theta^0 \). Assume \( d(V(\theta_*), \sigma^2) > 0 \) which implies \( \sigma^2_{t|t-1} \neq V_{t|t-1}(\theta_*) \) for any \( t \). A natural estimator of \( \theta_* \) under Scenario 1 is \( \hat{\theta}_T^{mdrv} \) determined by equation (4.3). Without further restriction \( \hat{\theta}_T^{mdrv} \) could be divergent or converge to something different than \( \theta_* \). We need therefore stronger conditions in order to have \( \hat{\theta}_T^{mdrv} \) behave nicely. This is studied in the following theorem. Let \( \varepsilon_t(\theta) = RV_t - V_{t|t-1}(\theta) \).

**Theorem 5.1** (Scenario 1). Under Assumptions A.1, A.2 and A.7(1), we have
1. \( \hat{\theta}_T^{mdrv} \) is identifiably unique and it is a strongly consistent estimator of \( \theta_* \).
2. Under additional Assumptions A.3 and A.7(2, 3), \( E[\text{Hess}(\varepsilon_t^2)(\theta_*)] \), denoted by \( 2\Sigma^{md} \), is positive definite and finite; \( \lim_{T \to \infty} \text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \nabla \varepsilon_t(\theta_*) \right) \) exists and is finite, denoted by \( \Omega^{mdrv} \).
3. If \( \Omega^{mdrv} > 0 \), \( \sqrt{T}(\hat{\theta}_T^{mdrv} - \theta_*) \Rightarrow N(0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1}) \).

Proof: see Appendix B.8

23
Estimation under Scenario 2

Given that \( \inf_{\theta \in C} \| \sigma_{t|t-1} - V_{t|t-1}(\theta) \|_2 = \| \sigma_{t|t-1} - V_{t|t-1}(\theta_*) \|_2 > 0 \), the natural estimator of \( \theta_* \) as in Scenario 1 is \( \hat{\theta}^{mdrv}_T \) defined in (4.5) or \( \hat{\theta}^{mdr^2}_T \) defined in (4.6). This is because \( \| RV_t - V_{t|t-1}(\theta) \|_2 = \| \sigma_{t|t-1} - RV_t \|_2 + \| \sigma_{t|t-1} - V_{t|t-1}(\theta) \|_2 \), and \( R^2_t - V_{t|t-1}(\theta) \) is minimized at \( \theta_* \) if and only if \( R^2_t - V_{t|t-1}(\theta) \) is. The properties of \( \hat{\theta}^{mdrv}_T \) and \( \hat{\theta}^{mdr^2}_T \) are studied in the following theorem. Let \( \varepsilon_t(\theta) = RV_t - V_{t|t-1}(\theta) \) and \( e_t(\theta) = R^2_t - V_{t|t-1}(\theta) \).

**Theorem 5.2 (Scenario 2).** Under Assumptions A.1, A.2, A.5(2) and A.7(1),

1. \( \hat{\theta}^{mdrv}_T, \hat{\theta}^{mdr^2}_T \) are identifiably unique and they converge to \( \theta_* \) a.s.
2. Under additional Assumptions A.3 and A.7(2, 3), \( E[\text{Hess}(e_t^2)(\theta_*)] = E[\text{Hess}(\varepsilon_t^2)(\theta_*)] \) is positive definite and is finite, denoted by \( 2\Sigma^{md} \); And \( \lim_{T \to \infty} \text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t(\theta_*) \nabla \varepsilon_t(\theta_*) \right) \) and \( \lim_{T \to \infty} \text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_t(\theta_*) \nabla e_t(\theta_*) \right) \) exist and are finite, denoted by \( \Omega^{mdrv} \) and \( \Omega^{mdr^2} \) respectively.
3. If \( \Omega^{mdrv} > 0 \), \( \sqrt{T}(\hat{\theta}^{mdrv}_T - \theta_*) \xrightarrow{a.s.} N(0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1}) \),
4. If \( \Omega^{mdr^2} > 0 \), \( \sqrt{T}(\hat{\theta}^{mdr^2}_T - \theta_*) \xrightarrow{a.s.} N(0, (\Sigma^{md})^{-1}\Omega^{mdr^2}(\Sigma^{md})^{-1}) \).

Note that \( \{\varepsilon_t(\theta_*) \partial_t \varepsilon_t(\theta_*) , t \in \mathbb{Z} \} \) is not a martingale difference sequence, nor is \( \{e_t(\theta_*) \partial_t e_t(\theta_*) , t \in \mathbb{Z} \} \). When the HYBRID GARCH model is misspecified, the extra probabilistic structure introduced in Scenario 2 does not provide further insight into the asymptotic properties of estimators, compared with Scenario 1. The proofs are therefore similar to those of Theorem 5.1, and will be skipped.

As with the likelihood estimation for the correctly-specified HYBRID GARCH, we consider minimizing the distance between \( \sigma_{t|t-1}^2 \) and \( V_{t|t-1}(\theta) \) by minimizing \( E(\log V_{t|t-1}(\theta) + \sigma_{t|t-1}^2/V_{t|t-1}(\theta)) \) inspired by Kullback-Leibler distance. Since it is not symmetric, the minimizer, denoted by \( \theta_{**} \), may not coincide with \( \theta_* \). Define \( l_t(\theta) = \log V_{t|t-1}(\theta) + RV_t/V_{t|t-1}(\theta) \) and \( h_t(\theta) = \log V_{t|t-1}(\theta) + R^2_t/V_{t|t-1}(\theta) \). One can therefore estimate \( \theta_{**} \) using \( \tilde{\theta}^{hrv}_T \) defined in (4.7) or \( \tilde{\theta}^{hrr^2}_T \) defined in (4.8) in that \( El_t(\theta) = Eh_t(\theta) = E(\log V_{t|t-1}(\theta) + \sigma_{t|t-1}^2/V_{t|t-1}(\theta)) \).

**Theorem 5.3 (Scenario 2).** Suppose \( E(\sup_{\phi \in \Theta} H(\phi, \bar{\eta}^l_t))^2 < \infty \) and Assumptions A.1, A.2 and A.8(1) hold.

1. \( \tilde{\theta}^{hrv}_T, \tilde{\theta}^{hrr^2}_T \) are identifiably unique and they converge to \( \theta_{**} \) a.s.
2. Suppose inequality (B.1) holds. Under additional Assumptions A.8(2, 3), \( E\text{Hess}(l_t)(\theta_{**}) = E\text{Hess}(h_t)(\theta_{**}) \), denoted by \( \Sigma^t \). It is positive definite and finite.
Additionally, under Assumption A.3, \( \lim_{T \to \infty} \text{Var}(1/\sqrt{T} \sum_{t=1}^{T} \nabla l_t(\theta_{**})) \) and \( \lim_{T \to \infty} \text{Var}(1/\sqrt{T} \sum_{t=1}^{T} \nabla h_t(\theta_{**})) \) exist and are finite, denoted by \( \Omega_{lhrv} \) and \( \Omega_{lhr2} \) respectively.

If \( \Omega_{lhrv} > 0 \), \( \sqrt{T} (\tilde{\theta}_{lhrv} T - \theta_{**}) \Rightarrow N(0, (\Sigma^{lh})^{-1} \Omega_{lhrv}(\Sigma^{lh})^{-1}) \).

If \( \Omega_{lhr2} > 0 \), \( \sqrt{T} (\tilde{\theta}_{lhr2} T - \theta_{**}) \Rightarrow N(0, (\Sigma^{lh})^{-1} \Omega_{lhr2}(\Sigma^{lh})^{-1}) \).

Proof: see Appendix B.9

The proofs of Theorem 5.3(3-5) rely on the fact that \( \nabla l_t(\theta_{**}) \) and \( \nabla h_t(\theta_{**}) \) are NED on the underlying and hence are mixingales under Assumption A.3. This is studied in Lemma B.6(3). With inequality (B.1) held, the moment condition required to establish the near epoch dependence property of \( \nabla l_t(\theta) \) or \( \nabla h_t(\theta) \) is weaker than that of \( \varepsilon_t(\theta) \partial_t \varepsilon_t(\theta) \) or \( e_t(\theta) \partial_t e_t(\theta) \) (see Lemmas B.5 and B.6). And when \( H_t(\phi) \) is a linear filter, inequality (B.1) is always satisfied. Besides, although we require \( E_r^{4(2+v_2)} < \infty \) in Theorem 5.3(3-5), this condition can be weakened as \( E_r^{2(2+v_2)u} \) for some \( u > 1 \) due to Lemma B.6 and equation B.10.

The ‘likelihood’ function that depends on the entire history of the weakly dependent observations when the model is misspecified is not widely studied in the literature. Huber (1967) and White (1982) studied the likelihood estimation of misspecified models when the underlying DGP is i.i.d. Domowitz and White (1982) and Bates and White (1985) extended the study to dependent data but the optimand depends on a finite number of recent lagged observations, or m-dependent. The extension to allow dependence on all previous observations was addressed by Gallant and White (1988) but the mixing condition is stronger than what is considered here.

It should also be noted that although the criterion \( E(\log V_{t|t-1}(\theta) + \sigma_{t|t-1}^{2}/V_{t|t-1}(\theta)) \) is justified using Kullback-Leibler distance, we do not need to specify the true and approximating (conditional) density functions. Instead we work with second conditional moment solely and treat the first conditional moment to be 0.

6 Discussion

The topic addressed in this paper has many applications as it covers multi-period forecasts of stock market return volatilities. 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

\[ \text{Note that they are no longer martingales.} \]
volatility dynamics are driven by what we call HYBRID processes. Besides the many potential empirical applications (see for instance Chen, Ghysels, and Wang (2010)), there are also many outstanding theoretical issues that warrant further research. One such topic is that of 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. 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. How does this affect the estimation of HYBRID GARCH models? Namely, suppose the contaminated high frequency returns are denoted by $r^*$, how does $H(\phi, \vec{r}_t)$ compare to $H(\phi, \vec{r}_t^*)$? There are some answers one can provide. Namely, HYBRID-GARCH could be applied to pre-averaged data, see for example Jacod, Li, Mykland, Podolskij, and Vetter (2009), as a way to handle the presence of microstructure noise. One could also consider block-sampling schemes, as suggested in Mykland and Zhang (2009). We leave these questions for further research.

References


Technical Appendices

A Regularity Conditions

The purpose of this appendix is to collect the regularity conditions used in the paper. We provide a listing of the conditions which are used at various places in the main body. In what follows, we use $\nabla$ to denote the vector differential operator (w.r.t $\theta$) so that $\nabla f$ is the gradient (column vector) of scalar function $f,$ and $\text{Hess}(f)$ the Hessian matrix of $f,$ i.e., $\text{ent}_{t,j}\text{Hess}(f) = \partial_i \partial_j f$ where $\partial_k$ denotes the partial derivative w.r.t. the $k$th parameter in $\theta = (\alpha, \beta, \gamma, \phi).$ When $\phi$ is a vector, $\partial_\phi$ refers to the partial derivative w.r.t. each component of $\phi,$ say $\phi_i,$ and $\partial_\phi^2$ is treated as $\partial_{\phi_i} \partial_{\phi_j},$ $\nabla_\phi$ is a vector differential operator w.r.t. $\phi$ when $\phi$ is a vector.

Assumption A.1. Let $r_t^2 \in L^2(\Omega, \mathcal{F}, P)$ for some probability space $(\Omega, \mathcal{F}, P),$ and

$$P_t(RV_{t+1}|\mathcal{I}_t) = \sigma_{t+1}^2$$

(A.1)

where $RV_{t+1} = \sum_{j=0}^{m-1} r_{t+1-j/m},$ $\mathcal{I}_t$ 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 $E r_s^{4(2+v_2)} < \infty.$

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

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

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

Assumption A.6. The rank of the following matrix

$$\left( \begin{array}{cccc} \nabla_{\phi_1} \Psi_0 & \nabla_{\phi_1} \Psi_1 & \ldots & \nabla_{\phi_1} \Psi_{m-1} \end{array} \right)$$

(A.2)

is same as the dimension of $\phi_1.$
Assumption A.7. (1) \( \theta_* \) is identifiably unique in \( \mathcal{C} \). (2) \( \theta_* \in C^0 \). (3) The determinant of \( E[Hess(x_i^2)(\theta_*)] \) (or \( E[Hess(x_i^2)(\theta_*)] \)) is positive.

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

A.1 More details on Assumption A.4

Assumption A.4 essentially guarantees that the HYBRID process is non-negative and measurable, and satisfies identifiability if it is parameterized. Conditions (1) and (2) are very standard. Here we give more explanations about condition (3) which is also related to the choice of \( \Phi \).

**Example 1.** Consider \( H_t(\phi) = r_t^2 + \phi^2 r_{t-1/2}^2 \) with \( \Phi = \mathbb{R} \). \( H_t \) does not meet condition (3). To see this, suppose there exists \( c = (c_1, c_2, c_3) \in \mathbb{R}^3 \) such that \( c_1 + c_2 H_t + c_3 \partial_x H_t = 0 \), which is equivalent to \( c_1 + c_2 r_t^2 + \phi^2 c_3 r_{t-1/2}^2 = 0 \). Clearly, \( c_1 = c_2 = 0 \) but \( c_3 \) may not be 0 unless \( \phi \neq 0 \). Hence condition (3) is not satisfied. But the discussion indicates that \( H_t(\phi) \) with \( \Phi \) being either \((0, \infty)\) or \((-\infty, 0)\), or a connected subset of either of them will meet condition (3).

**Example 2.** Consider the HYBRID process constructed by a MIDAS component with exponential Almon lag polynomial:

\[
H_t(\phi) = \sum_{j=0}^{m-1} (\tilde{\gamma} + b_j(\eta)) r_{t-j/m}^2
\]

(A.3)

and

\[
b_j(\eta) = \exp\{\eta(j/m) + \eta_2(j/m)^2\} \left( \sum_{k=0}^{m-1} \exp\{\eta(k/m) + \eta_2(k/m)^2\} \right)^{-1}, \quad \tilde{\gamma} > 0, \eta_1, \eta_2 \in \mathbb{R}, \phi = (\tilde{\gamma}, \eta_1, \eta_2)^T.
\]

For easy discussion we let \( m = 5 \). We consider several situations:

(1) Consider \( \tilde{\gamma} > 0 \) and \( \eta_1, \eta_2 \neq 0 \). \( \partial H_t/\partial \tilde{\gamma} = \sum_{j=0}^{m-1} r_{t-j/m}^2 \), \( \partial H_t/\partial \eta_1 = \sum_{j=0}^{m-1} (\partial b_j/\partial \eta_1) r_{t-j/m}^2 \), and \( \partial H_t/\partial \eta_2 = \sum_{j=0}^{m-1} (\partial b_j/\partial \eta_2) r_{t-j/m}^2 \). For \( c = (c_1, c_2, c_3, c_4, c_5) \in \mathbb{R}^5 \), suppose that \( c_1 + c_2 H_t + c_3 \partial H_t/\partial \tilde{\gamma} + c_4 \partial H_t/\partial \eta_1 + c_5 \partial H_t/\partial \eta_2 = 0 \), which is equivalent to \( c_1 + \sum_{j=0}^{m-1} [c_2 (\tilde{\gamma} + b_j) + c_3 + c_4 (\partial b_j/\partial \eta_1) + c_5 (\partial b_j/\partial \eta_2)] r_{t-j/m}^2 = 0 \). Hence \( c_1 = 0 \), and \( c_2 (\tilde{\gamma} + b_j) + c_3 + c_4 (\partial b_j/\partial \eta_1) + c_5 (\partial b_j/\partial \eta_2) = 0 \) for \( j = 0, 1, \ldots, 4 \).

Note that \( \sum_{j=0}^{m-1} b_j(\eta) = 1 \). We have \( \sum_{j=0}^{m-1} \partial b_j/\partial \eta_1 = 0 \) and \( \sum_{j=0}^{m-1} \partial b_j/\partial \eta_2 = 0 \). Therefore \( c_2 (m \tilde{\gamma} + 1) + mc_3 = 0 \), and \( c_2 (b_j - 1/m) + c_4 (\partial b_j/\partial \eta_1) + c_5 (\partial b_j/\partial \eta_2) = 0 \) (\( \forall j \)), or equivalently

\[
\begin{pmatrix}
 b_0 - 1/5 & \partial b_0/\partial \eta_1 & \partial b_0/\partial \eta_2 \\
 b_1 - 1/5 & \partial b_1/\partial \eta_1 & \partial b_1/\partial \eta_2 \\
 b_2 - 1/5 & \partial b_2/\partial \eta_1 & \partial b_2/\partial \eta_2 \\
 b_3 - 1/5 & \partial b_3/\partial \eta_1 & \partial b_3/\partial \eta_2 \\
 b_4 - 1/5 & \partial b_4/\partial \eta_1 & \partial b_4/\partial \eta_2 \\
\end{pmatrix}
\begin{pmatrix}
 c_2 \\
 c_4 \\
 c_5 \\
\end{pmatrix}
= \begin{pmatrix}
 0 \\
 0 \\
 0 \\
 0 \\
\end{pmatrix}.
\]

(A.4)

Note that

\[
\frac{\partial b_j(\eta)}{\partial \eta_1} = b_j(\eta) \left( \frac{j}{m} - \sum_{k=0}^{m-1} k b_k(\eta) \right), \quad \frac{\partial b_j(\eta)}{\partial \eta_2} = b_j(\eta) \left( \frac{j^2}{m^2} - \sum_{k=0}^{m-1} \frac{k^2}{m^2} b_k(\eta) \right).
\]

32
The rank of the coefficient matrix in (A.4) is 3. We have \( c_2 = c_4 = c_5 = 0 \), and hence \( c_3 = 0 \) as well.

It follows that 1, \( H_t \), and each component of \( \partial \phi(H_t) \) are linearly independent.

(2) When \( \bar{\gamma} > 0 \), and either \( \eta_1 \neq 0, \eta_2 = 0 \) or \( \eta_1 = 0, \eta_2 \neq 0 \), \( H_t \), and each component of \( \partial \phi(H_t) \) are linearly independent. The proof is similar to (1).

(3) When \( \bar{\gamma} > 0 \) and \( \eta_1, \eta_2 = 0 \), \( H_t(\phi) = \sum_{j=0}^{m-1} (\bar{\gamma} + 1/m)r_{t-j/m}^2 \).

For \( c = (c_1, c_2, c_3) \in \mathbb{R}^3 \), \( c_1 + c_2 H_t + c_3 \partial H_t / \partial \bar{\gamma} = 0 \) is equivalent to \( c_1 + \sum_{j=0}^{m-1} (c_2(\bar{\gamma} + 1/m) + c_3)r_{t-j/m}^2 = 0 \), which implies \( c_1 = 0 \), and \( c_2(\bar{\gamma} + 1/m) + c_3 = 0 \). Because \( c_2 \) and \( c_3 \) may not be zero at the same time, 1, \( H_t \), and each component of \( \partial \phi(H_t) \) are linearly dependent.

The discussion indicates that when \( \Phi \) is a connected subset of \{ \( (\bar{\gamma}, \eta_1, \eta_2) : \bar{\gamma} > 0, \eta_1^2 + \eta_2^2 \neq 0 \} \), \( H_t(\phi) \) satisfies condition (3).

**Example 3.** Consider the HYBRID process driven by NIC as in equation (2.4), i.e.,

\[
H_t(\phi) = \sum_{j=0}^{m-1} \Psi_j(\phi_1) NIC(\phi_2, r_{t-j/m}), \quad \sum_{j=0}^{m-1} \Psi_j(\phi_1) = 1
\]

where \( \phi = (\phi_1, \phi_2) \), and the weights \( (\Psi_0(\phi_1), \Psi_1(\phi_1), \ldots, \Psi_{m-1}(\phi_1))^T \) are determined by a low-dimensional functional specification. In this example, we focus on how to choose weights and \( \Phi \) with NIC specified in (2.5) and (2.6) to meet condition (3). The degenerate case that \( \phi_1 \equiv 0 \) and \( \phi_2 \equiv 0 \) is excluded in the discussion.

(1) Consider \( NIC(\phi_2, r) = br^2 1_{r \geq 0} + cr^2 1_{r < 0} \) \( (b \neq 0, c \neq 0) \). For \( c = (c_1, c_2, c_3, c_4, c_5) \), \( c_1 + c_2 H_t + c_3 \partial H_t / \partial b + c_4 \partial H_t / \partial c + c_5 \nabla \phi_t H_t = 0 \) is equivalent to

\[
c_1 + \sum_{j=0}^{m-1} [c_2 \Psi_j b + c_3 \Psi_j + c_5 \nabla \phi_t \Psi_j] 1_{r_{t-j/m} \geq 0} r_{t-j/m}^2 + \sum_{j=0}^{m-1} [c_2 \Psi_j c + c_4 \Psi_j + c_5 \nabla \phi_t \Psi_j] 1_{r_{t-j/m} < 0} r_{t-j/m}^2 = 0
\]

1_{r_{t-j/m} \geq 0} and 1_{r_{t-j/m} < 0} are linearly independent of each other. We have \( c_1 = 0, c_2 \Psi_j b + c_3 \Psi_j + c_5 \nabla \phi_t \Psi_j = 0 \) for \( j = 0, 1, \ldots, m-1 \). Note that \( \sum_{j=0}^{m-1} \Psi_j(\eta) = 1 \). It implies that \( \nabla \phi_t \Psi_j = 0 \), and thus \( c_2 b + c_3 = 0, c_2 c + c_4 = 0, \) and \( c_5 \nabla \phi_t \Psi_j = 0 \) (\( \forall j \)). Note that \( c_5 \) is 0 if the weights satisfy Assumption A.6, but \( c_2, c_3, c_4 \) may be non-zero. Therefore 1, \( H_t \), and each component of \( \partial \phi(H_t) \) are linearly dependent.

In order to have \( H_t(\phi) \) meet condition (3), one could consider \( NIC(\phi_2, r) = r^2 1_{r \geq 0} + r^2 1_{r < 0} \) or \( NIC(\phi_2, r) = br^2 1_{r \geq 0} + r^2 1_{r < 0} \) and have the weights satisfy Assumption A.6.

(2) \( H_t(\phi) \) with \( NIC(\phi_2, r) = b(r - c)^2 \) \( (b > 0, c \neq 0) \) does not meet condition (3) either. Proof is similar to (1). However, \( H_t(\phi) \) with \( NIC(\phi_2, r) = (r - c)^2 \) and weights satisfying Assumption A.6 could satisfy condition (3).

**B** Proofs

**B.1** Preliminary results

We first present some useful results in this section which will facilitate the proofs of the theorems appearing in the paper. The lemmas below are stated under Assumptions A.1 and A.4.
Lemma B.1. Under Assumptions A.2 and A.5(1), \( \partial_t V_{t-1}(\theta) \), \( \partial_i \partial_j V_{t-1}(\theta) \) are strictly stationary ergodic for \( \theta \in \mathcal{C} \) and \( i,j \in \{1,2,\ldots,d\} \). Moreover under the additional Assumption A.5(2), \( E(\sup_{\theta \in \mathcal{C}} V_{t-1}(\theta))^2 \), \( E(\sup_{\theta \in \mathcal{C}} |\partial_t V_{t-1}(\theta)|)^2 \), and \( E(\sup_{\theta \in \mathcal{C}} |\partial_i \partial_j V_{t-1}(\theta)|)^2 \) are bounded.

Proof: Note that \( H_t, \partial_t H_t, \partial_i \partial_j H_t \) are strictly stationary ergodic. \( V_{t-1}(\theta) = \gamma + \sum_{k=0}^{\infty} \beta^k H_{t-1-k}(\phi) \) a.s for \( \theta \in \mathcal{C} \). And it is easy to check that \( \sum_{k=0}^{\infty} \partial_t (\gamma^k \beta^k) \) and \( \sum_{k=0}^{\infty} \partial_i \partial_j (\gamma^k \beta^k) \) are absolutely summable uniformly on \( \mathcal{C} \), which implies that \( \partial_t V_{t-1} = \partial_t (\alpha/(1 - \beta)) + \sum_{k=0}^{\infty} \partial_i \partial_j (\gamma^k \beta^k H_{t-1-k}(\phi)) \) a.s. and hence they are strictly stationary ergodic.

Since \( \mathcal{C} \subset \mathbb{R}^d \) is bounded, one can always find constants (say) \( c_1 > 0, c_2 > 0, \) and \( 0 < c_3 < 1 \) such that \( V_{t-1}(\theta) \leq c_1 + c_2 \sum_{k=0}^{\infty} c_3^k \sup_{\phi \in \mathcal{C}} H_{t-1-k}(\phi) \). Note that \( \sum_{k=0}^{\infty} c_3^k (\sup_{\phi \in \mathcal{C}} H_{t-1-k}(\phi))^2 \leq \infty \) a.s. due to Assumption A.5(2). We have \( V_{t-1}(\theta)^2 \leq 2c_1^2 + \frac{2c_2^2}{1-c_3} \sum_{k=0}^{\infty} c_3^k (\sup_{\phi \in \mathcal{C}} H_{t-1-k}(\phi))^2 \) a.s. due to the Cauchy-Schwarz inequality and hence \( E(\sup_{\theta \in \mathcal{C}} V_{t-1}(\theta))^2 \) is \( O(1) \). Similarly \( E(\sup_{\theta \in \mathcal{C}} |\partial_t V_{t-1}(\theta)|)^2 \) and \( E(\sup_{\theta \in \mathcal{C}} |\partial_i \partial_j V_{t-1}(\theta)|)^2 \) are \( O(1) \).

Lemma B.2. Fix \( \theta \in \mathcal{C} \). If \( p^T \nabla V_{t-1}(\theta) = 0 \) a.s. for any \( t \in \mathbb{Z} \), then \( p = 0 \).

Proof: Let \( p = (p_1, p_2, p_3, p_4) \in \mathbb{R}^d \), where \( p_4 \) is of the same dimension as \( \phi \). Note that \( \nabla V_{t+1|t}(\theta) = \nabla \alpha + (\nabla \beta) V_{t-1}(\theta) + \nabla (\gamma H_t(\phi)) \). \( p^T \nabla V_{t-1}(\theta) = 0 \) a.s. implies

\[
p_1 + p_2 V_{t-1}(\theta) + p_3 H_t(\phi) + p_4^T \nabla \phi H_t(\phi) = 0 \quad \text{a.s.}
\]

Since \( p_3 H_t(\phi) + p_4^T \nabla \phi H_t(\phi) \in \mathcal{I}_t, p_2 = 0 \) and hence \( p_1 + p_3 H_t(\phi) + p_4^T \nabla \phi H_t(\phi) = 0 \) a.s. Assumption A.4 implies that \( p_1 = p_3 = 0 \) and \( p_4 = 0 \) (since \( \gamma > 0 \)).

Lemma B.3. For \( \theta \in \mathcal{C} \), \( V_{t-1}(\theta) = V_{t-1}(\theta_0) \) a.s. \( \forall t \in \mathbb{Z} \) if and only if \( \theta = \theta_0 \).

Proof: Sufficiency is apparent. We need to check necessity. If \( V_{t-1}(\theta) = V_{t-1}(\theta_0) \) a.s. \( \forall t \in \mathbb{Z} \),

\[
0 = \alpha - \alpha_0 + (\beta - \beta_0) V_{t-1}(\theta_0) + (\gamma H(\phi, \vec{r}_t) - \gamma_0 H(\phi_0, \vec{r}_t)), \quad \text{a.s.}
\]

Since \( V_{t-1}(\theta_0) \in \mathcal{I}_{t-1} \) and \( \gamma H(\phi, \vec{r}_t) - \gamma_0 H(\phi_0, \vec{r}_t) \in \mathcal{I}_t \), we have \( \beta = \beta_0 \) and hence \( (\alpha - \alpha_0) + (\gamma H(\phi, \vec{r}_t) - \gamma_0 H(\phi_0, \vec{r}_t)) = 0 \) a.s. Note that \( \gamma H(\phi, \vec{r}_t) - \gamma_0 H(\phi_0, \vec{r}_t) = H(\vec{\phi}, \vec{r}_t)(\gamma - \gamma_0) + \gamma(\phi - \phi_0)^T \nabla \phi H(\vec{\phi}, \vec{r}_t) \) where \( (\gamma, \vec{\phi}) \) is between \( (\gamma, \phi) \) and \( (\gamma_0, \phi_0) \) and it may depend on \( t \). Assumption A.4 indicates that \( \alpha = \alpha_0, \gamma = \gamma_0 \) and \( \phi = \phi_0 \). In other words \( \theta = \theta_0 \).

Lemma B.4. Suppose that \( E(\sup_{\phi \in \mathcal{C}} H(\phi, \vec{r}_t))^\delta < \infty \) for some \( \delta > 0 \), and

\[
|\partial_\phi H(\phi, \vec{x})/H(\phi, \vec{x})| \leq g(\phi), \quad |\partial^2_\phi^2 H(\phi, \vec{x})/H(\phi, \vec{x})| \leq g(\phi) \quad \forall \phi \in \mathbb{R}^m, \phi \in \Phi \quad (B.1)
\]

where \( g \) is real-valued and continuous in \( \phi \). Then we have

\[
E(\sup_{\theta \in \mathcal{C}} |\partial_t V_{t-1}(\theta)/V_{t-1}(\theta)|)^v < \infty, \quad E(\sup_{\theta \in \mathcal{C}} |\partial_t \partial_j V_{t-1}(\theta)/V_{t-1}(\theta)|)^v < \infty \quad \forall v > 0. \quad (B.2)
\]

Proof: \( |\partial_\phi H(\phi, \vec{x})/H(\phi, \vec{x})| \) and \( |\partial^2_\phi^2 H(\phi, \vec{x})/H(\phi, \vec{x})| \) are bounded on \( \mathcal{C} \), and suppose the upper bound is \( M_1 > 0 \). Note that \( |\partial_t (\alpha/(1 - \beta))| \leq (1/\alpha + 1/(1 - \beta)) \alpha/(1 - \beta) \) and \( |\partial_t (\gamma \beta^k)| \leq (1/\gamma + k/\beta) \gamma \beta^k \).

34
And (B.1) implies that \( |\partial_t v_{i|t-1}| \leq |\partial_t (\alpha/(1 - \beta))| + \sum_{k=0}^{\infty} C(k)\gamma^k H_{t-1-k}(\phi) \) where \( C(k) = M_1 + 1/\gamma + k/\beta \). Therefore, on \( \mathcal{C} \)

\[
\left| \frac{\partial_t V_{i|t-1}}{V_{i|t-1}}(\theta) \right| \leq (1/\alpha + 1/(1 - \beta)) + C(N) + (1 - \beta)/\alpha \sum_{k>N} C(k)\gamma^k H_{t-1-k}(\phi), \quad \text{for } N \in \mathbb{N}.
\]

Because one can always find constants \( M_2 > 0 \) and \( 0 < \rho_* < 1 \) such that \((1 - \beta)/\alpha C(k)\gamma^k \leq M_2\rho_*^k\), \(1/\alpha + 1/(1 - \beta) < M_2 \) and \( C(N) \leq M_2 N \) on \( \mathcal{C} \), we have for \( \theta \in \mathcal{C} \)

\[
\left| \frac{\partial_t V_{i|t-1}}{V_{i|t-1}}(\theta) \right| \leq M_2 + M_2 N + M_2 \sum_{k>N} \rho_*^k H_{t-1-k}(\phi).
\]

The rest of the discussion is similar to the proof of Lemma 5.2 of Berkes, Horváth, and Kokoszka (2003), and hence we have \( E\sup_{\theta \in \mathcal{C}} |\partial_t V_{i|t-1}/V_{i|t-1}|^v < \infty \) for any \( v > 0 \).

The second inequality in (B.2) follows from a similar argument.

**Lemma B.5.** Suppose that \( Er_s^k < \infty \), \( r_s \) is strictly stationary, and Assumption A.5(3) is true. Define \( \varepsilon_t(\theta) = RV_t - V_{i|t-1}(\theta) \). Fix \( k = 1, \ldots, d \) and \( \theta \in \mathcal{C} \). Then \( \|\varepsilon_t\|_2 < \infty \) and \( \sup_{\theta} \|\varepsilon_t\partial_k\varepsilon_t - E(\varepsilon_t\partial_k\varepsilon_t|F_{t-m}^i)\|_2 \leq C\rho^m \) for some constants \( C > 0 \) and \( 0 < \rho < 1 \). Therefore \( \{\varepsilon_t\partial_k\varepsilon_t, t \in \mathbb{Z}\} \) is near epoch dependent on \( \{\tilde{r}_t\} \). This is also true when \( RV_t \) is replaced with \( R_t^2 \).

**Proof:** Define \( Z_t = \varepsilon_t\partial_k\varepsilon_t \). Note that \( E\sup_{\theta \in \mathcal{C}} |\partial_t V_{i|t-1}(\theta)| < \infty \) and \( E\sup_{\theta \in \mathcal{C}} |(\partial_t V_{i|t-1}(\theta))|^4 < \infty \), which follows from an argument similar to the proof of Lemma B.1. We have \( \|Z_t\|_2 \leq \|\varepsilon_t\|_4 \|\partial_k\varepsilon_t\|_4 < \infty \).

Since \( \varepsilon_t(\theta) = RV_t - \frac{\alpha}{1 - \beta} - \gamma \sum_{j=0}^{\infty} \beta^j H_{t-1-j}(\phi) \), it can be written as \( \varepsilon_t(\theta) = \sum_{j=0}^{\infty} c_j(\theta)\tilde{H}_{t-j}(\phi) \) where \( c_0(\theta) = 1, \tilde{H}_t(\theta) = RV_t - \alpha/(1 - \beta) \), and \( c_j(\theta) = -\gamma\beta^{j-1}, \tilde{H}_{t-j}(\theta) = H_{t-j}(\phi) \) for \( j \geq 1 \). Hence

\[
Z_t = \sum_{i,j=0}^{\infty} c_i\tilde{H}_{t-i}\partial_k(c_j\tilde{H}_{t-j}) = \left( \sum_{0 \leq i,j \leq m} + \sum_{0 \leq i,m,j \geq 0} c_i\tilde{H}_{t-i}\partial_k(c_j\tilde{H}_{t-j}) \right) = Z_t^{(m)} + \xi_t^{(m)} + \eta_t^{(m)}.
\]

Note that \( \|\xi_t^{(m)}\|_2 \leq \sum_{0 \leq i,m,j \geq 0} \|c_i\partial_k(c_j)(\tilde{H}_{t-i}\tilde{H}_{t-j})\|_2 + \|c_i\partial_k(c_j)\tilde{H}_{t-i}\|_2 \leq M_2\rho^m \) for some constants \( C > 0 \) and \( 0 < \rho < 1 \). Similarly, \( \|\eta_t^{(m)}\|_2 \leq 2M^2B_1/(1 - \rho)\rho^m + 1 \). Note that \( \|Z_t - E(Z_t|F_{t-m}^i)\|_2 \leq \|Z_t - Z_t^m\|_2 \). Therefore \( \sup_t \|Z_t - E(Z_t|F_{t-m}^i)\|_2 \leq C\rho^m \) for some constants \( C > 0 \) and \( 0 < \rho < 1 \).

**Lemma B.6.** Define \( l_t(\theta) = \log V_{i|t-1} + RV_t/V_{i|t-1}(\theta) \). Suppose that \( r_s \) is strictly stationary. Then

1. \( E\sup_{\theta \in \mathcal{C}} |l_t(\theta)| < \infty \) if \( E\sup_{\theta \in \mathcal{C}} H(\phi, \tilde{r}_t) < \infty \).

2. \( E\sup_{\theta \in \mathcal{C}} |l_t(\theta)| < \infty \) and inequality (B.1) holds. Then \( E\sup_{\theta \in \mathcal{C}} |\partial_t l_t(\theta)| \) \( < \infty \) and \( E\sup_{\theta \in \mathcal{C}} |\partial_t l_t(\theta)| \) \( < \infty \) if additionally assume that \( Er_s^{4+v} < \infty \) for some \( v > 0 \), then \( E(\sup_{\theta \in \mathcal{C}} |\partial_t l_t(\theta)|)^2 < \infty \).

3. \( E\sup_{\theta \in \mathcal{C}} |l_t^{4+v} < \infty \) for some \( u > 0 \) and inequality (B.1) holds. For each \( i = 1, 2, \ldots, d \) and \( \theta \in \mathcal{C} \), \( \{\partial_t l_t, t \in \mathbb{Z}\} \) is near epoch dependent on \( \{\tilde{r}_t\} \) and particularly \( \sup_t \|\partial_t l_t - E(\partial_t l_t|F_{t-m}^i)\|_2 \leq C\rho^m \) for some constants \( C > 0 \) and \( 0 < \rho < 1 \).

This is also true when \( RV_t \) is replaced with \( R_t^2 \).
Proof: (1) Note that $\log \alpha \leq l_t(\theta) \leq \log V_{t\mid t-1}(\theta) + RV_t/\alpha$. Hence $|l_t(\theta)| \leq \max\{|\log \alpha|, |V_{t\mid t-1}(\theta) + RV_t/\alpha|\}$. Since $E\sup_{\theta \in \mathcal{C}} V_{t\mid t-1}(\theta) < \infty$, which follows from an argument similar to the proof of Lemma B.1, we have $E\sup_{\theta \in \mathcal{C}} |l_t(\theta)| < \infty$.

(2) Note that $\partial_t l_t = (1 - RV_t/V_{t\mid t-1})\partial_t V_{t\mid t-1}/V_{t\mid t-1} \text{ and } \partial_t^2 l_t = (1 - RV_t/V_{t\mid t-1})(\partial_t \partial_t V_{t\mid t-1}/V_{t\mid t-1} + (2RV_t/V_{t\mid t-1} - 1)(\partial_t V_{t\mid t-1}/V_{t\mid t-1})(\partial_t V_{t\mid t-1}/V_{t\mid t-1})$. We have, due to Lemma B.4,

$$E\sup_{\theta \in \mathcal{C}} |\partial_t^2 l_t(\theta)| \leq E\sup_{\theta \in \mathcal{C}} (1 + RV_t/\alpha)^2 E\sup_{\theta \in \mathcal{C}} \partial_t V_{t\mid t-1}/V_{t\mid t-1}^2 < \infty \quad E\sup_{\theta \in \mathcal{C}} \partial_t l_t(\theta)|^2 \leq E\sup_{\theta \in \mathcal{C}} (1 + RV_t/\alpha)^2 E\sup_{\theta \in \mathcal{C}} (4 + v)/(2 + v) \quad E\sup_{\theta \in \mathcal{C}} |\partial_t \partial_t l_t(\theta)| \leq E\sup_{\theta \in \mathcal{C}} (1 + RV_t/\alpha)^2 E\sup_{\theta \in \mathcal{C}} (\partial_t V_{t\mid t-1}/V_{t\mid t-1})^2$$

(3) Let $X_t = (\varepsilon_t, Y_t, V_{t\mid t-1})^T$, $X_t^{(m)} = (\varepsilon_t^{(m)}, Y_t^{(m)}, V_{t\mid t-1}^{(m)})^T$ where $Y_t = \partial_t \varepsilon_t$, $V_{t\mid t-1}^{(m)} = \alpha/(1 - \beta) + \gamma \sum_{k=0}^{m-1} \beta^k H_{t-1-k}(\phi)$, $\varepsilon_t^{(m)} = RV_t - V_{t\mid t-1}^{(m)}$, $Y_t^{(m)} = -\partial_t V_{t\mid t-1}$. Define function $b : \mathbb{R}^3 \rightarrow \mathbb{R}$ as $b(x_1, x_2, x_3) = x_1 x_2 / x_3^2$. Thus $b(X_t) = \partial_t l_t$. The proof of near epoch dependence is an application of Theorem 4.2 of Gallant and White (1988) with $d(X_t, X_t^{(m)}) = |\varepsilon_t - \varepsilon_t^{(m)}| + |Y_t - Y_t^{(m)}| + |V_{t\mid t-1} - V_{t\mid t-1}^{(m)}|$. Note that $|b(X_t) - b(X_t^{(m)})| \leq B(X_t, X_t^{(m)})d(X_t, X_t^{(m)})$ a.s. where

$$B(X_t, X_t^{(m)}) = |Y_t/V_{t\mid t-1}^2 + |\varepsilon_t^{(m)}|/V_{t\mid t-1}^2 + |Y_t^{(m)}|Y_t^{(m)}/(V_{t\mid t-1}^{(m)}V_{t\mid t-1})/(V_{t\mid t-1}^{(m)}V_{t\mid t-1} + 1/V_{t\mid t-1}^{(m)})$$

Because $|Y_t/V_{t\mid t-1}|, |(Y_t - Y_t^{(m)})/V_{t\mid t-1}|, |Y_t^{(m)}|/V_{t\mid t-1}^2$ are finite for any $\delta > 0$ due to Lemma B.4 and there exist $u_1, u_2 > 0$ such that $2 + u_2/(2 + u_2) = (1 + u_1)^2$, we have

$$\left| \frac{\varepsilon_t^{(m)}}{V_{t\mid t-1}^{(m)}} \right| \leq \left| \frac{RV_t/\alpha + 1}{V_{t\mid t-1}^{(m)}} \right| < \infty \quad \left| \frac{Y_t^{(m)} - V_{t\mid t-1}^{(m)}}{V_{t\mid t-1}^{(m)}} \right| < \infty \quad \left| \frac{Y_t/(V_{t\mid t-1}^2 + |\varepsilon_t^{(m)}|/V_{t\mid t-1}^2 + |Y_t^{(m)}|/V_{t\mid t-1}^2)}{d(X_t, X_t^{(m)})/V_{t\mid t-1}^2} \right| < \infty$$

Therefore $B(X_t, X_t^{(m)})d(X_t, X_t^{(m)}) \leq \infty$, and $B(X_t, X_t^{(m)}) < \infty$ as well.

Note also that $d(X_t, X_t^{(m)}) \leq |Y_t - Y_t^{(m)}|/V_{t\mid t-1}^2 + 2|V_{t\mid t-1}^2 - V_{t\mid t-1}^{(m)}|/2$, and $|V_{t\mid t-1}^2 - V_{t\mid t-1}^{(m)}| \leq \gamma \sum_{j=m} H_{t-1-j}(\phi)$ and $|Y_t - Y_t^{(m)}| \leq \gamma \sum_{j=m} \gamma \beta^j |H_{t-1-j}(\phi)| + |\partial_t \gamma \beta^j |H_{t-1-j}(\phi)|/2$. It follows from the proof of Lemma B.5 that $d(X_t, X_t^{(m)}) \leq C_1 \rho_t^m$ for some constants $C_1 > 0$ and $0 < \rho_t < 1$, and hence $\partial_t l_t, t \in \mathbb{Z}$ is near epoch dependent on $\{\varepsilon_t\}$. In particular, since

$$\left| \partial_t l_t - E(\partial_t l_t|\mathcal{F}_{t-m}^t) \right| \leq 2\left| B(X_t, X_t^{(m)})d(X_t, X_t^{(m)}) \right|^{2 + u_2}/2 + 2\left| B(X_t, X_t^{(m)})d(X_t, X_t^{(m)}) \right|^{n_2}/2 + 2\left| d(X_t, X_t^{(m)}) \right|^{n_2}/2$$

due to Lemma 4.1 of Gallant and White (1988), $\sup_t \left| \partial_t l_t - E(\partial_t l_t|\mathcal{F}_{t-m}^t) \right| \leq C\rho^m$ for some constants $C > 0$ and $0 < \rho < 1$. ■
B.2 Proofs of Proposition 3.1 and Corollary 3.1

Proof of Proposition 3.1: Drost and Werker (1996) shows that \((a, b, c)\) relates to \((\theta, \omega, \lambda, v_L^*)\) as follows: letting \(h = 1/m\),

\[
a = \omega(1 - e^{-\theta h})h, \quad c = e^{-\theta h} - b
\]

and \(|b| < 1\) is the solution to

\[
\frac{b}{1 + b^2} = \frac{\rho e^{-\theta h} - 1}{\rho(1 + e^{-2\theta h}) - 2},
\]

where

\[
\rho = \frac{4(e^{-\theta h} - 1 + \theta h) + 2\theta h(1 + (v/2 + \theta h)(1 - \lambda)/\lambda)}{1 - e^{-2\theta h}}, \quad v = \frac{\theta v_L^*}{1 - \lambda}.
\]

Note that

\[
\rho = 1 + h\theta(1 + 1/\lambda) + \theta^2h^2/\lambda + v(1 + h\theta + \theta^2h^2/3) + o(h^2)
\]

where \(v = (v/2)(1 - \lambda)/\lambda = \theta v_L^*/(2\lambda)\). Therefore

1. when \(v_L^* > 0, b = 1 - h\theta(1 + \phi) + o(h)\) and \(c = e^{-\theta h} - b = h\theta\phi + o(h)\) where \(\phi = 1 + 1/\theta - 1 = 1\). It implies that, as \(m \to \infty\),

\[
\beta_m = b^m \to e^{-\theta(1 + \phi)}, \quad \frac{c}{1 - b} = \frac{e^{-\theta h} - b}{1 - b} \to \frac{\phi}{1 + \phi}, \quad \frac{d_m}{m} = \frac{1 - (b + c)^m}{m(1 - b - c)} \to \theta^{-1}(1 - e^{-\theta})
\]

\[
\alpha_m = ma(1 - b^m) \to \omega (1 - e^{-\theta(1 + \phi)}) \left(1 - \frac{\phi}{1 + \phi}\theta^{-1}(1 - e^{-\theta})\right)
\]

\[
\gamma_m = cd_m \to (1 - e^{-\theta})\phi.
\]

Next we need to show \(\lim_{m \to \infty} \sum_{j=0}^{m-1} \beta_j^{j/m} r_{t-j/m}^2 = \int_{t-1,t} e^{-\theta(1+\phi)(t-s)} d[p,p], s\) in probability. Note that \(\sum_{i=1}^{m} e^{-\theta(1+\phi)(t-i)} r_{ti}^2\) converges to \(\int_{t-1,t} e^{-\theta(1+\phi)(t-s)} d[p,p], s\) in probability where \(t_i = t - 1 + i/m\).\(^{13}\)

And for any \(\epsilon > 0\),

\[
P\left(\left|\sum_{j=0}^{m-1} \beta_j^{j/m} r_{t-j/m}^2 - \sum_{i=1}^{m} e^{-\theta(1+\phi)(t-i-1)} r_{ti}^2\right| > \epsilon\right) \leq \frac{\omega}{\epsilon} \left(\sum_{j=0}^{m-1} (j/m) \log(\beta_m) + \theta(1 + \phi)(j/m + 1/m)\right)
\]

\[
\leq \frac{\omega}{\epsilon} \left(\log(\beta_m) + \theta(1 + \phi)/2 + \theta(1 + \phi)/m\right).
\]

Therefore \(\lim_{m \to \infty} \sum_{j=0}^{m-1} \beta_j^{j/m} r_{t-j/m}^2 = \int_{t-1,t} e^{-\theta(1+\phi)(t-s)} d[p,p], s\) in probability.

2. when \(v_L^* = 0, b = 1 - \sqrt{h\theta} + o(h^{1/2})\) and \(c = \sqrt{h\theta} + o(h^{1/2})\). It implies that, as \(m \to \infty\),

\[
\beta_m = b^m \to 0, \quad \frac{c}{1 - b} \to 1, \quad \frac{d_m}{m} \to \theta^{-1}(1 - e^{-\theta})
\]

\(^{13}\)See Protter (2004).
\[
\alpha_m = \frac{ma(1 - b^m)}{1 - (b + c)} \left(1 - \frac{cd_m}{m(1 - b)}\right) \rightarrow \omega \left(1 - \theta^{-1}(1 - e^{-\theta})\right)
\]
\[
\gamma_m = \sqrt{\lambda/\theta}(1 - e^{-\theta})
\]

Next we show that \(\sqrt{m} \sum_{j=0}^{m-1} \beta_{jm} r_{t-j/m}^2\) converges to \((\theta \lambda)^{-1/2} \sigma_t^2\) in \(L^2\), which is equivalent to show that \(\lim_{m \to \infty} mc \sum_{j=0}^{m-1} b^j r_{t-j/m} = \sigma_t^2\) in \(L^2\). Let \(\widetilde{RV}_t = \sum_{j=0}^{m-1} b^j r_{t-j/m}^2\).

\[
E \left[ mc \sum_{j=0}^{m-1} b^j r_{t-j/m}^2 - \sigma_t^2 \right]^2 = \frac{\omega^2}{1 - \lambda} + m^2 \omega^2 E(\widetilde{RV}_t^2) - 2mcE(\widetilde{RV}_t \sigma_t^2)
\]

Note that

\[
E(\widetilde{RV}_t^2) = \sum_{j=0}^{m-1} b^{2j} E(r_{t-j/m}^4) + 2 \sum_{j<i} b^{i+j} E(r_{t-i/m}^2 r_{t-j/m}^2)
\]
\[
= \sum_{j=0}^{m-1} b^{2j}(kh_j^2 \omega^2) + 2 \sum_{j<i} b^{i+j} \left[ h_j^2 \omega^2 + \frac{\omega^2 \lambda e^{h \theta} (1 - e^{-h \theta})^2}{1 - \lambda} e^{-(i-j)\theta/m}\right]
\]
\[
E(\widetilde{RV}_t \sigma_t^2) = \sum_{j=0}^{m-1} b^j E(r_{t-j/m} \sigma_t) = \sum_{j=0}^{m-1} b^j E \left( \int_{t-j/m-1/m}^{t-j/m} \sigma_t \sigma_u dL_u \right)^2
\]
\[
= \sum_{j=0}^{m-1} b^j \int_{t-j/m-1/m}^{t-j/m} E(\sigma_t^2 \sigma_u^2) du = \sum_{j=0}^{m-1} b^j \omega^2 \omega^2 \int_{t-j/m-1/m}^{t-j/m} (1 + \frac{\lambda e^{-(t-u)\theta}}{1 - \lambda}) du
\]
\[
= \frac{\omega^2}{m} \left[ \frac{1 - b^m}{1 - b} + \frac{m \lambda}{1 - \lambda} \theta^{-1}(1 - e^{-\theta/m}) \frac{1 - b^m e^{-\theta}}{1 - be^{-\theta/m}}\right]
\]

Therefore,

\[
E \left[ mc \sum_{j=0}^{m-1} b^j r_{t-j/m}^2 - \sigma_t^2 \right]^2
\]
\[
= \frac{\omega^2}{1 - \lambda} + k \omega^2 \frac{e^2 (1 - b^{2m})}{1 - b^2} + 2 \omega^2 \sum_{j<i} b^i + 2m^2 \omega^2 \frac{\omega^2 \lambda e^{h \theta} (1 - e^{-h \theta})^2}{1 - \lambda} \sum_{j<i} b^i e^{-(i-j)\theta/m}
\]
\[
- 2k \omega^2 \left[ \frac{1 - b^m}{1 - b} + \frac{m \lambda}{1 - \lambda} \theta^{-1}(1 - e^{-\theta/m}) \frac{1 - b^m e^{-\theta}}{1 - be^{-\theta/m}}\right]
\]
\[
= \frac{\omega^2}{1 - \lambda} + T_1 + T_2 + T_3 - T_4
\]

Note that

\[
T_1 = k \omega^2 \frac{e^2 (1 - b^{2m})}{1 - b^2} = \left( \frac{3}{1 - \lambda} + o(1) \right) \omega^2 \frac{h \theta \lambda + o(h)}{2 \sqrt{h}} (1 - (b^m)^2) \rightarrow 0
\]
\[
T_2 = 2 \omega^2 \sum_{j<i} b^i = 2 \omega^2 e^2 \frac{(b - b^m)(1 - b^m)}{(1 - b)^2 (1 + b)} \rightarrow \omega^2
\]
\[ T_4 = 2\omega^2 \left[ \frac{1 - b^m}{1 - b} + \frac{m\lambda}{1 - \lambda} \theta^{-1}(1 - e^{-\theta/m}) \frac{1 - b^m e^{-\theta}}{1 - be^{-\theta/m}} \right] \]
\[ = 2\omega^2 \left[ \frac{e}{1 - b}(1 - b^m) + \frac{\theta^{-1}\lambda}{1 - \lambda} m\sqrt{h} (\sqrt{\theta} + o(1)) \frac{h\theta + o(h)}{\sqrt{h}} \right] \rightarrow 2\omega^2 \frac{1}{1-\lambda} \]
\[ T_3 = 2m^2\omega^2 \frac{\omega \theta^2}{1 - \lambda} \frac{e^{h\theta}(1 - e^{-h\theta})^2}{e^{h\theta}} \sum_{j<i} b^{i+j} e^{-(i-j)\theta/m} \]
\[ = \frac{2\omega^2 \lambda}{(1 - \lambda)\theta^2} m^2 e^{h\theta}(1 - e^{-h\theta})^2 \left( \frac{b^{1-m} - b^m e^{-\theta}}{1 - be^{-\theta/m}} - \frac{b^2 - b^m}{1 - b^2} \right) \]
\[ = \frac{2\omega^2 \lambda}{(1 - \lambda)\theta^2} \frac{1}{h^2} \left( \frac{1}{\sqrt{\theta} + o(1)} \right)^2 \frac{h(1 + h\theta + o(h))(h\theta + o(h))}{\sqrt{h}(\sqrt{\theta} + o(1))} \]
\[ \times \left( \frac{1 - \sqrt{h}(\sqrt{\theta} + o(1))}{\sqrt{h}(\sqrt{\theta} + o(1))} - \frac{1 - 2\sqrt{h}(\sqrt{\theta} + o(1))}{2\sqrt{h}(\sqrt{\theta} + o(1))} \right) \rightarrow \frac{\omega^2 \lambda}{1 - \lambda}. \]

Therefore \( mc \sum_{j=0}^{m-1} b^{(m-j)r^2/t-j/m} \) converges to \( \sigma_t^2 \) in \( L^2 \).

**Proof of Corollary 3.1:** Sufficiency follows from the fact that for each \( s > 0, \lim_{m \to \infty} P(\sup_{t \leq s} |V_t^{(m)}|) \leq 0. \) Therefore \( V_t \) converges to \( V_t = a.s. \) to \( \omega(1 - e^{-\theta(1+\phi)})/(1+\phi) \) in probability, which however contradicts the results in Proposition 3.1.

**B.3 Proofs of Theorems 4.1, 4.4 and 4.6**

**Proof of Theorem 4.1:** Note that \( ||RV_t - V_{t-1}(\theta)||_2 = ||RV_t - \sigma_t^2||_2 + ||V_{t-1}(\theta) - \sigma_t^2||_2 \) for all possible \( \theta \)'s. When the model is correctly specified, \( \min_{\theta \in \Theta} ||RV_t - V_{t-1}(\theta)||_2 = ||RV_t - \sigma_t^2||_2 \) and \( ||RV_t - V_{t-1}(\theta_0)||_2 = 0. \) To show that \( \theta_0 \) is unique, suppose there exists \( \theta_1 \in \Theta \) such that \( ||RV_t - V_{t-1}(\theta_1)||_2 = \min_{\theta \in \Theta} ||RV_t - V_{t-1}(\theta)||_2. \) It implies \( ||V_{t-1}(\theta_1) - \sigma_t^2||_2 = 0, \) and \( \theta_1 = \theta_0 \) in the sense of Lemma B.3. Therefore \( \theta_1 = \theta_0 \), which follows from Lemma B.3.

**Proof of Theorem 4.4:** Similar to the proof of Theorem 4.1.

**Proof of Theorem 4.6:** It suffices to justify the first equality. The proof for the second is same. Define \( l_t(\theta) = \log V_{t-1}(\theta) + R_t^2/V_{t-1}(\theta) \). \( E \sup_{\theta \in \Theta} |l_t(\theta)| < \infty \) due to Lemma B.6. Under Assumption 3.2 \( E(l_t(\theta) - l_t(\theta_0)) = E \left( \frac{V_{t-1}(\theta_0)}{V_{t-1}(\theta)} - 1 - \log \frac{V_{t-1}(\theta)}{V_{t-1}(\theta_0)} \right) \geq 0. \) The equality holds if and only if \( V_{t-1}(\theta) = V_{t-1}(\theta_0), \) or \( \theta = \theta_0 \) which is implied by Lemma B.3. Therefore \( El_t(\theta) \) is uniquely minimized at \( \theta_0. \)

**B.4 Proof of Theorems 4.2 and 4.3, and Proposition 4.1**

Let \( \varepsilon_t(\theta) = RV_t - V_t(\theta), \hat{\varepsilon}_t(\theta) = RV_t - \hat{V}_t(\theta), O_T(\theta) = 1/T \sum_{t=1}^{T} \varepsilon_t^2(\theta), \hat{O}_T(\theta) = 1/T \sum_{t=1}^{T} \hat{\varepsilon}_t^2(\theta). \) The proof is started with \( \hat{\theta}_T^{\text{driv}} = \arg \min_{\theta \in \Theta} O_T(\theta). \)

**Lemma B.7.** Under Assumptions 3.1 and A.2, \( \hat{\theta}_T^{\text{driv}} \) is identifiably unique and converges to \( \theta_0 \) a.s.
Proof: Note that \( \varepsilon_t(\theta) \) is strictly stationary ergodic and \( E \sup_{\theta \in C}(\varepsilon_t(\theta))^2 < \infty \) (see Lemma B.1). \( O_T(\theta) \) converges to \( E(\varepsilon^2_t(\theta)) \) a.s. uniformly on \( C \) due to uniform SLLN (see Rao (1962)). And Theorem 4.1 implies that \( \theta_0 \) is identifiable unique. Therefore the results follow from Lemma A.1 of Goncalves and White (2004) and Theorem 3.3 of Gallant and White (1988).

Lemma B.8. Under Assumptions 3.1 and A.2, \( \lim_{T \to \infty} \sup_{\theta \in C} |O_T(\theta) - \bar{O}_T(\theta)| \overset{a.s.}{=} 0. \)

Proof: Note that there exists \( \kappa > 1 \) such that \( \lim_{t \to \infty} \kappa^t \sup_{\theta \in C} |V_t| = 0 \) according to Theorem 3.1 of Bougerol (1993) or Theorem 2.8 of Straumann and Mikosch (2006). In other words, \( \forall \delta > 0, \exists T_0 > 0 \) such that \( \kappa^t \sup_{\theta \in C} |\varepsilon_t(\theta) - \tilde{\varepsilon}_t(\theta)| < \delta \) for \( t > T_0 \). Hence \( \sup_{\theta \in C} |\tilde{\varepsilon}_t(\theta)| \leq 2\delta \kappa^{-t} \sup_{\theta \in C} |\varepsilon_t(\theta)| + \delta^2 \kappa^{-2t} \) when \( t > T_0 \). Since under Assumption A.2 \( E \sup_{\theta \in C} |\varepsilon_t(\theta)| \) is bounded away from 0, \( E \log \sup_{\theta \in C} |\varepsilon_t(\theta)| \) is finite as well. Considering Lemma 2.1 of Straumann and Mikosch (2006), we have \( \lim_{t \to \infty} \sup_{\theta \in C} |\tilde{\varepsilon}_t(\theta) - \tilde{\varepsilon}_t^2(\theta)| = 0 \) a.s., and hence \( \lim_{T \to \infty} \sup_{\theta \in C} |O_T(\theta) - \bar{O}_T(\theta)| \leq \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in C} |\tilde{\varepsilon}_t^2(\theta) - \tilde{\varepsilon}_t(\theta)| = 0 \) a.s.

Proof of Theorem 4.2: Lemma B.8 implies that \( \bar{O}_T(\theta) \) converges to \( E\tilde{\varepsilon}_t^2(\theta) \) a.s. uniformly on \( C \). Similar to the proof of Lemma B.7, we have \( \bar{\theta}_T^{mdv} \) is identifiably unique and it converges to \( \theta_0 \) a.s.

Proof of Proposition 4.1: Define \( Z_t = \varepsilon_t(\theta_0) \) for \( \theta_0 \) and for \( \theta_0 \). Then \( EZ_\theta = 0 \). Lemma B.5 implies that \( \{Z_t\} \) is near epoch dependent on \( \{r_t\} \) and \( \sup_{\theta \in C} \|Z_t - E(Z_t|\mathcal{F}_{t-1})\|_2 \leq C \rho^m \) for some constants \( C > 0 \) and \( 0 < \rho < 1 \). Define \( \Omega_T = var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t\right) \). Note that \( \Omega_T = \gamma(0) + 2 \sum_{k=1}^{T-1} (1 - k/T) \gamma(k) \) where \( \gamma(k) = cov(Z_k, Z_0) \). And for \( k > 0 \)

\[
|\gamma(2k)| = |E(Z_t Z_{t-2k})| \leq |E[(Z_t - E(Z_t|\mathcal{F}_{t-k}))Z_{t-2k}] + E[E(Z_t|\mathcal{F}_{t-k})Z_{t-2k}]| \\
\leq \|Z_t - E(Z_t|\mathcal{F}_{t-k})\|_2 \|Z_{t-2k}\|_2 + 2\|Z_t\|_2 \alpha(k) \bar{\alpha}(k)^{1/2} \alpha(k)^{1/2} (2 + \bar{\alpha}(k)) \leq C \rho^k \|Z_t\|_2 + 2\|Z_t\|_2 \alpha(k) \bar{\alpha}(k)^{1/2} (2 + \bar{\alpha}(k)) \tag{B.4}
\]

The second inequality follows from Davydov (1968). Therefore \( \sum_{k=0}^{\infty} |\gamma(k)| < \infty \) under assumption A.3 and thus \( \lim_{T \to \infty} \Omega_T \) exists and is finite.

The proof of Theorem 4.3 needs the following lemmas:

Lemma B.9. Under Assumptions 3.1 and A.2,

\[
\lim_{T \to \infty} \limsup_{N \to \infty} \sup_{\theta \in B(\theta_0, 1/N) \cap C} \|Hess(O_T)(\theta) - 2\Sigma^{md}\| = 0 \quad \text{a.s.} \tag{B.5}
\]

where \( B(\theta_0, 1/N) = \{\theta \in \mathbb{R}^d : |\theta - \theta_0| < 1/N\} \) and \( 0 < \Sigma^{md} = E \nabla V_{t,-1}(\theta_0)(\nabla V_{t,-1}(\theta_0))' \). \( \Sigma^{md} \) is not empty for sufficiently large \( N \). \( H \) under Scenario 1 meets condition A.4 automatically. Note that \( Hess(O_T) = \frac{1}{T} \sum_{t=1}^{T} Hess(\varepsilon_t) \), and \( \partial_i \partial_j \varepsilon_t^2(\theta_0) = 2\varepsilon_i \varepsilon_j (\partial_i \varepsilon_t(\theta_0) \partial_j \varepsilon_t(\theta_0)) \), and \( \partial_i \partial_j \varepsilon_t(\theta_0) = 2E \partial_i \varepsilon_t(\theta_0) \partial_j \varepsilon_t(\theta_0) \) due to \( \partial_i \varepsilon_t(\theta) \in I_{t-1} \). Hence \( Hess(\varepsilon_t^2)(\theta_0) = 2\Sigma^{md} \). Clearly, \( \Sigma^{md} \geq 0 \) and O(1). Suppose that there exists \( T \) such that \( p \nabla V_{t,-1}(\theta_0)(\nabla \varepsilon_t(\theta_0))' = 0 \), which is equivalent to \( p \nabla V_{t,-1}(\theta_0) = 0 \) a.s. for all \( t \). Lemma B.2 implies \( p \equiv 0 \) and hence \( \Sigma^{md} > 0 \).

\(^{14}\text{Suppose the dimension of the parameter space is } d.\)
Note that $\sup_{\theta \in \Theta(\theta_0, 1/N) \cap C} \left\| \text{Hess}(O_T)(\theta) - 2\Sigma^{md} \right\| \leq \sup_{\theta \in C} \left\| \text{Hess}(O_T)(\theta) - E\text{Hess}(\varepsilon_1^2)(\theta) \right\| + E \sup_{\theta \in \Theta(\theta_0, 1/N) \cap C} \left\| \text{Hess}(\varepsilon_1^2)(\theta) - \text{Hess}(\varepsilon_1^2)(\theta_0) \right\|$, and $E \sup_{\theta \in C} \left\| \text{Hess}(\varepsilon_1^2)(\theta) \right\|$ is $O(1)$ uniformly in $t$ due to Lemma B.1. (B.5) follows from the dominated convergence theorem and uniform SLLN. □

**Lemma B.10.** Under Assumptions 3.1, A.3 and $\Omega^{mdrv} > 0$, $\sqrt{T}(\hat{\theta}^{mdrv} - \theta_0) \implies N(0, (\Sigma^{md})^{-1}\Omega^{mdrv}(\Sigma^{md})^{-1})$, where $\Sigma^{md} = E\nabla V_{i|t-1}(\theta_0)(\nabla V_{i|t-1}(\theta_0))'$.

**Proof:** Note that $-\nabla O_T(\theta_0) = \text{Hess}(O_T)(\hat{\theta}_T)(\hat{\theta}^{mdrv}_T - \theta_0)$ where $\hat{\theta}_T$ is between $\theta_0$ and $\hat{\theta}^{mdrv}_T$. Since $\theta_T$ converges to $\theta_0$ a.s., Lemma B.9 implies that $\text{Hess}(O_T)(\hat{\theta}_T)$ converges to $2\Sigma^{md}$ a.s. and $\text{Hess}(O_T)(\hat{\theta}_T)$ is invertible for sufficiently large $T$. Hence we have, for large $T$, $\sqrt{T}(\hat{\theta}^{mdrv}_T - \theta_0) = -(\text{Hess}(O_T)(\hat{\theta}_T))^{-1}\sqrt{T}\nabla O_T(\theta_0)$. The asymptotics result follows if $\sqrt{T}\nabla O_T(\theta_0)$ converges to $N(0, 4\Omega^{mdrv})$ in distribution. Therefore we just need to show that $\sqrt{T}\nabla O_T(\theta_0)$ converges to $N(0, 4\Omega^{mdrv}p)$ in distribution for any $p \in \mathbb{R}^d$ due to the Cramér-Wold device.

Note that $\sqrt{T}\nabla O_T(\theta_0) = \frac{2}{\sqrt{T}} \sum_{i=1}^T Z_i$, where $Z_i = \sum_{k=1}^d p_k Y_{k,t}$. Define $\Omega_T = \text{var}(\frac{1}{\sqrt{T}} \sum_{i=1}^T Z_i)$. $\Omega_T$ is $O(1)$ due to Proposition 4.1 and is uniformly positive definite because of $\Omega^{mdrv} > 0$, hence $\Omega_T^{-1}$ is $O(1)$. Consider $X_{Tt} = Z_t/\sqrt{T\Omega_T}$. $E(X_{Tt}) = 0$ and $\text{Var}(\sum_{t=1}^T X_{Tt}) = 1$. $\{X_{Tt}\}$ is near epoch dependent on $\{r_t\}$ of size 1 due to Lemma B.5 and $\{r_t\}$ is $\alpha$-mixing of size $v_2/(2 + v_2)$.

Note also that $\|Z_t\|_{2+v_2} < \infty$, and $T(1/\sqrt{T\Omega_T})^2$ is $O(1)$. An application of Theorem 3.6 of Davidson (1992) yields that $\sum_{t=1}^T X_{Tt}$ converges to $N(0,1)$ in distribution and hence $\sqrt{T}\nabla O_T(\theta_0)$ converges to $N(0, 4\Omega^{mdrv}p)$ in distribution. □

**Lemma B.11.** Under Assumptions 3.1 and A.2, $\lim_{T \to \infty} \sup_{\theta \in C} \sqrt{T}\|\nabla O_T(\theta) - \nabla O_T(\theta)\|_{\alpha*}^\alpha = 0$.

**Proof:** Note that for $t \geq 1$, $\hat{V}_t(\theta) = \alpha(1 - \beta^t) + \beta^t v + \sum_{k=0}^{t-1} \gamma^k \beta H_{t-k}(\phi)$, and $\hat{\partial}_i \hat{V}_t(\theta) = \partial_i \left( \alpha(1 - \beta^t) \right) + \partial_i (\beta^t v + \sum_{k=0}^{t-1} \partial_i (\gamma^k \beta H_{t-k}(\phi))$. Note also that $\hat{\partial}_i V_{i|t-1}(\theta) = \partial_i (\alpha/(1 - \beta)) + \sum_{k=0}^{\infty} \partial_i (\gamma^k \beta H_{i-1-k}(\phi))$ (see Lemma B.1). It is easy to check that both $\hat{\partial}_i V_{i|t-1}(\theta)$ and $\partial_i \hat{V}_t(\theta)$ satisfy

$$\hat{\partial}_i X_t = \partial_i \alpha + (\partial_i \beta) X_{t-1} + \beta (\partial_i X_{t-1}) + \partial_i (\gamma H_t(\phi)), \quad t \in \mathbb{Z}^+ \quad \text{(B.6)}$$

for each $i$. Since under Assumption A.2 the conditions of Proposition 6.1 of Straumann and Mikosch (2006) are met, then $\partial_i V_{i|t-1}(\theta)$ is the unique stationary ergodic solution to (B.6) and $\lim_{T \to \infty} \kappa_1^i \sup_{\theta \in C} |\partial_i V_{i|t-1}(\theta) - \hat{\partial}_i \hat{V}_t(\theta)| = 0$ for some $\kappa_1 > 1$, which implies $\lim_{T \to \infty} \kappa_1^i \sup_{\theta \in C} |\partial_i \varepsilon_i(\theta) - \hat{\partial}_i \hat{\varepsilon}_i(\theta)| = 0$. In other words, $\forall \delta > 0, \exists T_0 > 0$ such that $\sup_{\theta \in C} |\partial_i \varepsilon_i(\theta) - \hat{\partial}_i \hat{\varepsilon}_i(\theta)| < \kappa_1^i \delta$ and $\sup_{\theta \in C} |\varepsilon_i(\theta) - \hat{\varepsilon}_i(\theta)| < \kappa_\delta$ for $t > T_0$ (the second inequality is from (1)). Consequently,

$$\sqrt{T} \sup_{\theta \in C} |\varepsilon_i(\theta) - \hat{\varepsilon}_i(\theta)| \leq \sup_{\theta \in C} |\varepsilon_i(\theta) - \hat{\varepsilon}_i(\theta)| + \sqrt{T} \sup_{\theta \in C} |\varepsilon_i(\theta)| + \sqrt{T} \sup_{\theta \in C} |\varepsilon_i(\theta)|$$

for $t > T_0$. Note that $E \sup_{\theta \in C} |\partial_i \varepsilon_i(\theta)|$ and $E \sup_{\theta \in C} |\varepsilon_i(\theta)|$ are bounded away from 0. Same as the

15See page 24 of Gallant and White (1988) for the definition of size.
discuss in (1), we have \( \lim_{T \to \infty} \sqrt{T} \sup_{\theta \in \mathcal{C}} |\varepsilon_t(\theta)\partial_l\varepsilon_l(\theta) - \varepsilon_t(\theta)\partial_l\varepsilon_l(\theta)| = 0 \) a.s. Therefore,

\[
\lim_{T \to \infty} \sup_{\theta \in \Theta} \sqrt{T} |\partial_l O_T(\theta) - \partial_l \hat{O}_T(\theta)| \leq \lim_{T \to \infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} 2 \sup_{\theta \in \Theta} |\varepsilon_t(\theta)\partial_l\varepsilon_l(\theta) - \varepsilon_t(\theta)\partial_l\varepsilon_l(\theta)| \\
\leq \lim_{T \to \infty} 2 \sup_{\theta \in \Theta} |\varepsilon_t(\theta)\partial_l\varepsilon_l(\theta) - e_\varepsilon(\theta)\partial_l e_\varepsilon(\theta)| \stackrel{a.s.}{\longrightarrow} 0
\]

**Proof of Theorem 4.3:** Note that \( \nabla O_T(\hat{\theta}_T^m) - \nabla O_T(\hat{\theta}_T^e) = \text{Hess}(O_T)(\hat{\theta}_T^e)(\hat{\theta}_T^e - \hat{\theta}_T^m) \) where \( \hat{\theta}_T \in \mathcal{C} \) is between \( \hat{\theta}_T^m \) and \( \hat{\theta}_T^e \). On one hand, \( \sqrt{T}(\nabla O_T(\hat{\theta}_T^m) - \nabla O_T(\hat{\theta}_T^e)) = \sqrt{T}(\nabla O_T(\hat{\theta}_T^m) - \nabla O_T(\hat{\theta}_T^e)) \) converges to 0 a.s. due to Lemma B.11. On the other hand, since \( \lim_{T \to \infty} \hat{\theta}_T = \theta_0 \) a.s., \( \text{Hess}(O_T)(\hat{\theta}_T) \) converges to \( \Sigma^m \) a.s. due to Lemma B.9, and hence \( \text{Hess}(O_T)(\hat{\theta}_T) \) is positive definite for sufficiently large \( T \). Therefore we have

\[
\sqrt{T}(\hat{\theta}_T^m - \hat{\theta}_T^e) = (\text{Hess}(O_T)(\hat{\theta}_T))^{-1}(\nabla O_T(\hat{\theta}_T^m) - \nabla O_T(\hat{\theta}_T^e))
\]

converges to 0 in probability. An application of Slutsky’s theorem yields that \( \sqrt{T}(\hat{\theta}_T^m - \theta_0) \) converges to \( N(0,(\Sigma^m)^{-1}\Omega^m(\Sigma^m)^{-1}) \) in distribution.

**B.5 Proof of Theorem 4.5**

We will focus on the proof about \( \hat{\theta}_T^m \) and skip the proof regarding \( \hat{\theta}_T^e \) which will be same. Use the notations introduced in section B.4.

1. Note that lemmas B.7 and B.8 are still true under Scenario 2. Similar to the proof of Theorem 4.2, \( \hat{\theta}_T^m \) is identifiably unique and it converges to \( \theta_0 \) a.s.

2. Need to show that \( \sqrt{T}(\hat{\theta}_T^m - \theta_0) \implies N(0,(\Sigma^m)^{-1}\Omega^m(\Sigma^m)^{-1}) \).

Since Lemma B.9 still holds under Scenario 2, the first paragraph of the proof of Lemma B.10 is true under Scenario 2 as well. We only need to revise the proof in the second paragraph, i.e., the asymptotic normality of \( \sqrt{T}p^T\nabla O_T(\theta_0) \).

Note that \( Y_{l,t}(\theta_0) \equiv \varepsilon_t(\theta_0)\partial_l\varepsilon_l(\theta_0) \) is strictly stationary ergodic with finite second moment due to \( Ev^8 < \infty \) and Assumption A.5(3), following from an argument similar to the proof of Lemma B.1. \( Y_{l,t}(\theta_0) \) is a martingale difference sequence. Then \( \sqrt{T}p^T\nabla O_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} 2p^TY_l(\theta_0) \) converges to \( N(0,4p^T\Omega^m p) \) in distribution due to martingale central limit theorem where \( \Omega^m = EV_l(\theta_0)Y_l(\theta_0)^T = EV_l(\theta_0)Y_l(\theta_0) = E(V_l(\theta_0)V_l(\theta_0)^T) \). Note that \( p^T\Omega^m p > 0 \) if and only if \( p^T\nabla V_l(\theta_0) \neq 0 \) a.s. Hence \( \Omega^m \) is positive definite.

3. Note that \( \lim_{T \to \infty} \sqrt{T}(\hat{\theta}_T^m - \theta_0) = 0 \) in probability, which follows from an argument similar to the proof of Theorem 4.3 (since Lemmas B.8 and B.9 are true under Scenario 2). \( \sqrt{T}(\hat{\theta}_T^m - \theta_0) \) converges to \( N(0,(\Sigma^m)^{-1}\Omega^m(\Sigma^m)^{-1}) \) in distribution due to Slutsky’s theorem.

**B.6 Proof of Theorem 4.7**

We will give the proof regarding \( \hat{\theta}_T^h \), and skip the one for \( \hat{\theta}_T^l \), which will be similar. Define \( l_1(\theta) = \log V_l(\theta_0) + RV_l(\theta_0) \) and \( \hat{l}_1(\theta) = \log \hat{V}_l + RV_l(\theta_0) \). Let \( L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} l_1(\theta) \) and \( \hat{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \hat{l}_1(\theta) \). Suppose that \( \hat{\theta}_T^h \) is the solution to \( \min_{\theta \in \mathcal{C}} L_T(\theta) \).
Lemma B.12. Under Assumptions 3.2, A.2 and $E \sup_{\phi \in \Theta} H(\phi, \tilde{r}_T) < \infty$, $\hat{\theta}_T^{hrv}$ is identifiably unique and it converges to $\theta_0$ a.s.

Proof: Define $L(\theta) = \frac{1}{T} \sum_{t=1}^{T} E_t l_t(\theta) = E_{t} l_t(\theta)$ for $\theta \in \mathcal{C}$. $l_t$ is strictly stationary ergodic, and $E \sup_{\theta \in \mathcal{C}} |l_t(\theta)| < \infty$ (see Lemma B.6). $L_T(\theta)$ converges to $L(\theta)$ a.s. uniformly on $\mathcal{C}$ due to the uniform SLLN. Moreover $\theta_0$ is the unique minimizer of $L(\theta)$ (see Theorem 4.6). The results follow from Lemma A.1 of Goncalves and White (2004) and Theorem 3.3 of Gallant and White (1988).

Lemma B.13. Suppose inequality (B.1) holds. Let $B(\theta_0, 1/N) = \{ \theta \in \mathbb{R}^d : \|\theta - \theta_0\| < 1/N \}$. Under Assumptions 3.2, A.2 and $E \sup_{\phi \in \Theta} H(\phi, \tilde{r}_T) < \infty$,

$$\lim_{T \to \infty} \lim_{N \to \infty} \sup_{\theta \in B(\theta_0, 1/N) \cap \mathcal{C}} \|Hess(L_T)(\theta) - \Sigma^{th}\| = 0 \text{ a.s.}$$

(B.7)

where $0 < \Sigma^{th} = E \left( V_{t|t-1}^{-2}(\theta_0) \nabla V_{t|t-1}(\theta_0) \nabla V_{t|t-1}(\theta_0)' \right) < \infty$.

Proof: Since $\theta_0 \in \mathcal{C}$, $B(\theta_0, 1/N) \cap \mathcal{C}$ is not empty for sufficiently large $N$. Note that

$$\partial_i \partial_j L_T = \frac{1}{T} \sum_{t=1}^{T} \partial_i \partial_j l_t = \frac{1}{T} \sum_{t=1}^{T} \left( 1 - \frac{RV_t}{V_{t|t-1}} \right) \frac{\partial_i \partial_j V_{t|t-1}}{V_{t|t-1}} + \left( 2 \frac{RV_t}{V_{t|t-1}} - 1 \right) \frac{\partial_i V_{t|t-1} \partial_j V_{t|t-1}}{V_{t|t-1}^2}.$$

$\partial_i \partial_j l_t$ is strictly stationary ergodic. And $E \sup_{\theta \in \mathcal{C}} |\partial_i \partial_j l_t(\theta)| < \infty$ by Lemma B.6. $\Sigma^{th} > 0$ because $p' \nabla V_{t|t-1}(\theta_0) \neq 0$ a.s. for non-zero $p \in \mathbb{R}^d$ (see Lemma B.2).

Note that $\sup_{\theta \in B(\theta_0, 1/N) \cap \mathcal{C}} \|Hess(L_T)(\theta) - \Sigma^{th}\| \leq \sup_{\theta \in \mathcal{C}} \|Hess(L_T)(\theta) - E Hess(l_1)(\theta)\| + E \sup_{\theta \in B(\theta_0, 1/N) \cap \mathcal{C}} \|Hess(l_1)(\theta) - Hess(l_1)(\theta_0)\|$, and $E \sup_{\theta \in \mathcal{C}} \|Hess(l_1)(\theta)\|$ is $O(1)$ uniformly in $t$. Thus (B.7) is true due to the dominated convergence theorem and uniform SLLN.

Lemma B.14. Suppose that $Er^{A+u} < \infty$ for $u > 0$ and inequality (B.1) holds. Under Assumptions 3.2, A.2, and $E \sup_{\phi \in \Theta} H(\phi, \tilde{r}_T) < \infty$, $\sqrt{T} \nabla L_T(\theta_0) \Rightarrow N(0, \Omega^{hrv})$ where $\Omega^{hrv} = E \left( V_{t|t-1}^{-4}(\theta_0) (RV_t - V_{t|t-1}(\theta_0)) \nabla V_{t|t-1}(\theta_0) \nabla V_{t|t-1}(\theta_0) \right) > 0$.

Proof: Note that

$$\nabla L_T = \frac{1}{T} \sum_{t=1}^{T} \nabla l_t = \frac{1}{T} \sum_{t=1}^{T} \left( 1 - \frac{RV_t}{V_{t|t-1}} \right) \nabla V_{t|t-1}/V_{t|t-1}.$$

$\nabla l_t$ is strictly stationary ergodic. $E(\partial l_t(\theta_0))^2$ is finite due to Lemma B.6. And $E(RV_t|F_{t-1}) = V_{t|t-1}(\theta_0)$. Hence $\{\partial l_t(\theta_0), t \in \mathbb{Z}\}$ is a martingale difference sequence with finite second moment. Therefore, for nonzero $p \in \mathbb{R}^d$, $\sqrt{T} p' \nabla L_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} p' \nabla l_t(\theta_0)$ converges to $N(0, p' \Omega^{hrv} p)$ in distribution by the martingale central limit theorem, and $\Omega^{hrv}$ is positive definite because $p' \nabla V_{t|t-1}(\theta_0) \neq 0$ a.s. The result follows from the Cramer-Wold device.

Lemma B.15. Under assumptions 3.2 and A.2 and $E \sup_{\phi \in \Theta} H(\phi, \tilde{r}_T) < \infty$,

$$\lim_{T \to \infty} \sup_{\theta \in \mathcal{C}} |L_T(\theta) - \tilde{L}_T(\theta)| = 0 \text{ a.s.} \quad \text{(B.8)}$$
**Proof:** Note that $L_T(\theta) - \tilde{L}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} (l_t(\theta) - \tilde{l}_t(\theta))$. It suffices to show that $\lim_{t \to \infty} \sup_{\theta \in C} |l_t(\theta) - \tilde{l}_t(\theta)| = 0$ a.s. Since $|l_t(\theta) - \tilde{l}_t(\theta)| \leq |\log V_{t|t-1} - \log \tilde{V}_t| + \left| \frac{1}{V_{t|t-1}(\theta)} - \frac{1}{\tilde{V}_t(\theta)} \right| \leq \frac{R_V}{\alpha} |V_{t|t-1}(\theta) - \tilde{V}_t(\theta)|$ where the second inequality is due to the mean value theorem, we have $\sup_{\theta \in C} |l_t(\theta) - \tilde{l}_t(\theta)| \leq \frac{R_V}{\alpha} |V_{t|t-1}(\theta) - \tilde{V}_t(\theta)|$ for some $\alpha > 0$. Note that $E \log^{+} R_V < \infty$ and $\lim_{t \to \infty} \kappa^t \sup_{\theta \in C} |V_{t|t-1}(\theta) - \tilde{V}^t(\theta)|$ a.s. $= 0$ for some $\kappa > 1$ (see Lemma B.8). $\lim_{t \to \infty} \sup_{\theta \in C} |l_t(\theta) - \tilde{l}_t(\theta)| = 0$ a.s. by Lemma 2.1 of Straumann and Mikosch (2006).

**Lemma B.16.** Suppose that inequality (B.1) holds. Under assumptions 3.2 and A.2, and $E \sup_{\theta \in \Theta} H(\phi, \tilde{r}_t) < \infty$,

$$\lim_{T \to \infty} \frac{1}{T} \sup_{\theta \in C} \| \nabla L_T(\theta) - \nabla \tilde{L}_T(\theta) \| \overset{a.s.}{\leq} 0. \quad (B.9)$$

**Proof:** Since $\frac{1}{T} \sup_{\theta \in C} |\partial_i L_T(\theta) - \partial_i \tilde{L}_T(\theta)| \leq \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in C} |\partial_i l_t(\theta) - \partial_i \tilde{l}_t(\theta)|$ for each $i$, it suffices to show that $\lim_{t \to \infty} \frac{1}{T} \sup_{\theta \in C} |\partial_i l_t(\theta) - \partial_i \tilde{l}_t(\theta)| = 0$ a.s. Note that

$$\partial_i l_t - \partial_i \tilde{l}_t = \left(1 - \frac{R_V}{V_{t|t-1}}\right) \frac{\partial_i V_{t|t-1}}{V_{t|t-1}} - \left(1 - \frac{R_V}{\tilde{V}_t}\right) \frac{\partial_i \tilde{V}_t}{\tilde{V}_t}.$$ 

Applying the mean value theorem to $\partial_i l_t - \partial_i \tilde{l}_t$, we have

$$|\partial_i l_t - \partial_i \tilde{l}_t| \leq \left| \frac{\partial_i \tilde{V}_t - \partial_i V_{t|t-1}}{\alpha^2} \right| + \left| \frac{\partial_i V_{t|t-1}}{\alpha^2} \right| \left( \frac{2R_V}{\alpha} + 1 \right) |V_{t|t-1} - \tilde{V}_t| + \left(1 + \frac{R_V}{\alpha}\right) \frac{1}{\alpha} |\partial_i V_{t|t-1} - \partial_i \tilde{V}_t|$$

Note that

$$\sqrt{T} \left| \frac{\partial_i \tilde{V}_t - \partial_i V_{t|t-1}}{\alpha^2} \right| \left( \frac{2R_V}{\alpha} + 1 \right) |V_{t|t-1} - \tilde{V}_t| = \kappa^t |\partial_i \tilde{V}_t - \partial_i V_{t|t-1}| \kappa^t |V_{t|t-1} - \tilde{V}_t| \sqrt{T} \frac{2R_V}{\alpha} + 1 \right)$$

$$\sqrt{T} \left(1 + \frac{R_V}{\alpha}\right) \frac{1}{\alpha} |\partial_i V_{t|t-1} - \partial_i \tilde{V}_t| = \kappa^t |\partial_i V_{t|t-1} - \partial_i \tilde{V}_t| \sqrt{Tk^t} \left(1 + \frac{R_V}{\alpha}\right) \frac{1}{\alpha} \right.$$

Since $E \log^{+} R_V < \infty$, $E \log^{+} (\sup_{\theta \in C} |\partial_i V_{t|t-1}(\theta)|) \leq E \log^{+} (\sup_{\theta \in C} |\partial_i V_{t|t-1}(\theta)|) + E \log^{+} R_V < \infty$, and $\lim_{t \to \infty} \kappa^t \sup_{\theta \in C} |\partial_i V_{t|t-1} - \partial_i \tilde{V}_t| a.s. = 0$, $\lim_{t \to \infty} \kappa^t \sup_{\theta \in C} |V_{t|t-1} - \tilde{V}_t| a.s. = 0$ ($\kappa$ and $\kappa_1$ are defined in the proof of Lemma B.8), we have $\lim_{t \to \infty} \frac{1}{T} \sup_{\theta \in C} |\partial_i l_t(\theta) - \partial_i \tilde{l}_t(\theta)| = 0$ a.s.

**Proof of Theorem 4.7:** The results follow from an argument similar to the proofs of Theorems 4.2 and 4.3.

**B.7 Proof of Corollary 4.1**

For easy exposition, we define $r_{i,t} \equiv r_{i-1+i/m}$ for $i = 1, 2, \ldots, m$. Note that

$$E_{t-1}(R_i^2 - R_V) = 4 \sum_{k<l} E_{t-1}(r_{i,k}^2 r_{i,l}^2) + 4 \sum_{k<l} E_{t-1}(r_{i,k} r_{l,i}^2) + 4 \sum_{k<l} E_{t-1}(r_{i,k} r_{l,i}^3)$$
We have $E_{t-1}(r_{k,t}r_{l,t}r_{k,t}r_{l,t}) = 0$ and $E_{t-1}(r_{k,t}r_{l,t}^3) = 0$ under Semi-Strong GARCH assumption and $E(r_{k,t}^3|F_{s-1/m}) = 0$. Therefore $E_{t-1}(R_t^4 - RV_t^2) > 0$, which implies that $\text{ent}_{i,j}(\Sigma^{lh})^{-1}\Omega^{lh}\epsilon^{lh}(\Sigma^{lh})^{-1}) > \text{ent}_{i,j}(\Sigma^{lh})^{-1}\Omega^{lh}\epsilon^{lh}(\Sigma^{lh})^{-1})$, and $\text{ent}_{i,j}(\Sigma^{lh})^{-1}\Omega^{lh}\epsilon^{lh}(\Sigma^{lh})^{-1}) > \text{ent}_{i,j}(\Sigma^{lh})^{-1}\Omega^{lh}\epsilon^{lh}(\Sigma^{lh})^{-1})$.

\[ \text{B.8 Proof of Theorem 5.1} \]

(1) Similar to the proof of Theorem 4.2.

(2) The moment conditions allow us to differentiate under the integral sign, and together with Assumption A.7(1,2) we have $E(\varepsilon_t(\theta_\delta)\nabla\varepsilon_t(\theta_\delta))=0$. For any $\theta_n \neq \theta_* \in C^0$ and $|\theta_n - \theta_*| < 1/n$,

$$\|RV_t - V_{t|\delta}(\theta_n)\|_2 > \|RV_t - V_{t|\delta}(\theta_*)\|_2 \quad \Rightarrow \quad (\theta_n - \theta_*)\text{HESS}(\varepsilon_t^2)(\theta_n - \theta_*) > 0,$$

for some $\bar{\theta}_n$ between $\theta_n$ and $\theta_*$. Hence $\text{HESS}(\varepsilon_t^2)(\theta_0) > 0$. The eigenvalues of $\text{HESS}(\varepsilon_t^2)(\theta_0)$ are continuous in $\theta$. Note that $\bar{\theta}_n$ converges to $\theta_*$ and the eigenvalues of $\text{HESS}(\varepsilon_t^2)(\theta_0)$ are non-zero due to Assumption A.7(3). Hence they should be strictly positive, which implies that $\text{HESS}(\varepsilon_t^2)(\theta_0) > 0$. Similar to the proof of Proposition 4.1, we have $\lim_{T \to \infty} \text{var}(\frac{1}{T} \sum_{t=1}^T \varepsilon_t \nabla\varepsilon_t(\theta_0))$ exists and is finite.

(3) Similar to the proof of Theorem 4.3.

\[ \text{B.9 Proof of Theorem 5.3} \]

(1) Use the same argument as the proof of Lemma B.12 and Lemma B.15.

(2) $E\partial_\delta l_t = E\theta_j h_t = E \left(1 - \frac{\sigma_{v_{t|\delta}}^2}{v_{t|\delta}} \right) \partial_\delta \partial_j V_{t|\delta} + E \left(\frac{2\sigma_{v_{t|\delta}}^2}{v_{t|\delta}} - 1\right) \frac{\partial_\delta V_{t|\delta}}{v_{t|\delta}} \partial_j V_{t|\delta}$. The moment condition and Assumption A.7 guarantee that $E\nabla l_t(\theta_*) = \nabla E l_t(\theta_*) = 0$. Similar to the proof of Theorem 5.1(2), $\Sigma^{lh}$ is positive definite. $\Sigma^{lh} < \infty$ due to Lemma B.6.

(3) Consider $1/\sqrt{T} \sum_{t=1}^T p^T \nabla l_t(\theta_*)$ for $p \in \mathbb{R}^d$. Note that $Z_t \equiv p^T \nabla l_t(\theta_*)$ has mean 0. Take $u_1 = u_3 = \sqrt{2} - 1$ and $u_2 = v_2$. It follows from Lemma B.6(3) that $Z_t$ is near epoch dependent on $\bar{r}_t$ and satisfying $\sup_t \|Z_t - E(Z_t|\mathcal{F}_t\|_2 \leq C \rho^m$ for some constants $C > 0$ and $0 < \rho < 1$. Note that

$$\|Z_t\|_{2+v_2} \leq \|RV_t/\alpha + 1\|_{2(2+v_2)}\|p^T \nabla V_{t|\delta}/V_{t|\delta}\|_{2(2+v_2)} < \infty. \quad (B.10)$$

Similar to the proof of Proposition 4.1, we have $\lim_{T \to \infty} \text{Var}(1/\sqrt{T} \sum_{t=1}^T p^T \nabla l_t(\theta_*)$) exists and is finite. Similarly, $\lim_{T \to \infty} \text{Var}(1/\sqrt{T} \sum_{t=1}^T p^T \nabla h_t(\theta_*)$) exists and is finite.

(4) Note that equations B.7, B.8, and B.9 with $\theta_0$ replaced by $\theta_*$ are true under the assumptions of Theorem 5.3. And the proofs are similar to those of lemmas B.13, B.15, and B.16. Therefore we only need to show that $\sqrt{T} \nabla L_T(\theta_*) \Rightarrow N(0, \Omega^{lh})$ or $\sqrt{T} p^T \nabla L_T(\theta_*) \Rightarrow N(0, p^T \Omega^{lh}\epsilon^{lh})$ (for $p \in \mathbb{R}^d$) under the assumptions of Theorem 5.3, where $L_T(\theta) = \frac{1}{2} \sum_{t=1}^T l_t(\theta)$. Since $Z_t \equiv p^T \nabla l_t(\theta_*)$ has mean 0 and it is near epoch dependent on $\bar{r}_t$ due to lemma B.6, the proof will be similar to the argument in the second paragraph of the proof of lemma B.10, and hence is skipped. \[ \blacksquare \]
Table 1: Summary of structural parameters in simulation study
The table displays the values of parameters in model (4.17) and model (4.18) respectively with different values of $m$, and the associated structural parameters in the HYBRID GARCH model.

<table>
<thead>
<tr>
<th></th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>$\alpha_m$</th>
<th>$\beta_m$</th>
<th>$\gamma_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong GARCH (1,1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 5$</td>
<td>2.8E-06</td>
<td>0.9770</td>
<td>0.0225</td>
<td>0.0001</td>
<td>0.8902</td>
<td>0.1124</td>
</tr>
<tr>
<td>$m = 78$</td>
<td>2.8E-06</td>
<td>0.9770</td>
<td>0.0225</td>
<td>0.0147</td>
<td>0.1628</td>
<td>1.7216</td>
</tr>
<tr>
<td>$m = 288$</td>
<td>2.8E-06</td>
<td>0.9770</td>
<td>0.0225</td>
<td>0.1429</td>
<td>0.0012</td>
<td>6.0365</td>
</tr>
<tr>
<td>Weak GARCH (1,1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m = 24$</td>
<td>3.86E-05</td>
<td>0.9794</td>
<td>0.0192</td>
<td>0.0216</td>
<td>0.6065</td>
<td>0.4523</td>
</tr>
<tr>
<td>$m = 144$</td>
<td>1.07E-06</td>
<td>0.9915</td>
<td>0.0082</td>
<td>0.0204</td>
<td>0.2945</td>
<td>1.1619</td>
</tr>
<tr>
<td>$m = 288$</td>
<td>2.69E-07</td>
<td>0.9940</td>
<td>0.0059</td>
<td>0.0195</td>
<td>0.1776</td>
<td>1.6590</td>
</tr>
</tbody>
</table>
Table 2: HYBRID-GARCH Model: Small sample properties, Strong GARCH

The table displays estimation of $\alpha_m$, $\beta$, $\gamma_m$ (and $g$ for the MEM estimation procedure) of a high frequency data GARCH(1,1) appearing in equation (4.17) with sample size 500 (Panel I:III) and sample size 1000 (Panel IV:VI), where the true values of $\alpha_m$, $\beta$, $\gamma_m$ are shown in the first line of each panel. The estimators considered are: mdrv, the minimizer of (4.3), and the companion estimator mdr$^2$, replacing $RV$ by $R^2$, as well as (quasi-)likelihood-based estimators lhr$^2$, defined in (4.7), and lhrv, defined in (4.8). Finally, the table also includes the mem method described in subsection 4.2.2. The numbers in the parenthesis are MSE for lhr$^2$, relative MSE (with respect to lhr$^2$) for lhrv, mdr$^2$, mdrv, mem. For $g$, we only report sample variance.

$$
\begin{align*}
\alpha_m & & b & & \gamma_m & & g \\
\text{True Value} & 0.000070 & 0.977000 & 0.112388 & & & \\
\text{LHR2} & 0.000259 (0.000000) & 0.972401 (0.000194) & 0.128465 (0.004700) \quad & & \text{Panel I: } m = 5, T = 500 \\
\text{LHRV} & 0.000261 (1.015169) & 0.974313 (0.271144) & 0.119444 (0.258291) \quad & & \\
\text{MDR2} & 0.000454 (6.150788) & 0.966299 (11.194958) & 0.145875 (5.802282) \quad & & \\
\text{MDRV} & 0.000328 (9.371349) & 0.975828 (0.502932) & 0.108347 (0.278161) \quad & & \\
\text{MEM} & 0.000052 (0.154507) & 0.976053 (0.790375) & 0.117574 (2.344735) & 2.344735 (0.022762) \quad & \\
\end{align*}
$$

$$
\begin{align*}
\text{Panel II: } m = 78, T = 500 \\
\text{True Value} & 0.014749 & 0.977000 & 1.721640 \quad & & \\
\text{LHR2} & 0.022270 (0.000429) & 0.971857 (0.000324) & 2.012838 (1.733246) \quad & & \\
\text{LHRV} & 0.016544 (0.053291) & 0.976488 (0.018644) & 1.743047 (0.018305) \quad & & \\
\text{MDR2} & 0.052818 (9.689061) & 0.944327 (28.476856) & 3.505934 (16.788438) \quad & & \\
\text{MDRV} & 0.029753 (1.037355) & 0.975510 (0.319931) & 1.742003 (0.327187) \quad & & \\
\text{MEM} & 0.004353 (0.430669) & 0.979426 (0.043334) & 1.575400 (0.185204) & 9.716956 (6.113690) \quad & \\
\end{align*}
$$

$$
\begin{align*}
\text{Panel III: } m = 288, T = 500 \\
\text{True Value} & 0.142879 & 0.977000 & 6.036455 \quad & & \\
\text{LHR2} & 0.154505 (0.008603) & 0.970251 (0.000693) & 7.462548 (38.821448) \quad & & \\
\text{LHRV} & 0.147024 (0.062688) & 0.976511 (0.033043) & 6.133488 (0.046095) \quad & & \\
\text{MDR2} & 0.381905 (22.197008) & 0.891170 (63.48471) & 20.803390 (38.496335) \quad & & \\
\text{MDRV} & 0.22810 (2.186104) & 0.970839 (1.738098) & 7.032169 (1.141032) \quad & & \\
\text{MEM} & 0.131788 (0.289360) & 0.976838 (0.129298) & 6.109518 (0.185204) & 7.711094 (2.311111) \quad & \\
\end{align*}
$$

Table continued on next page ...
Panel IV: $m = 5, T = 1000$

<table>
<thead>
<tr>
<th>True Value</th>
<th>$\alpha_m$</th>
<th>$b$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHR2</td>
<td>0.000245 (0.000000)</td>
<td>0.972167 (0.000110)</td>
<td>0.129344 (0.002517)</td>
<td></td>
</tr>
<tr>
<td>LHRV</td>
<td>0.000245 (0.967746)</td>
<td>0.972823 (0.487674)</td>
<td>0.126027 (0.416324)</td>
<td></td>
</tr>
<tr>
<td>MDR2</td>
<td>0.000313 (3.687974)</td>
<td>0.971769 (4.640177)</td>
<td>0.128260 (3.806918)</td>
<td></td>
</tr>
<tr>
<td>MDRV</td>
<td>0.000252 (0.877527)</td>
<td>0.976031 (0.274003)</td>
<td>0.109991 (0.265061)</td>
<td></td>
</tr>
<tr>
<td>MEM</td>
<td>0.000039 (0.212108)</td>
<td>0.976308 (0.955344)</td>
<td>0.116601 (1.014513)</td>
<td>2.347128 (0.011373)</td>
</tr>
</tbody>
</table>

Panel V: $m = 78, T = 1000$

<table>
<thead>
<tr>
<th>True Value</th>
<th>$\alpha_m$</th>
<th>$b$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHR2</td>
<td>0.018895 (0.000220)</td>
<td>0.974736 (0.000127)</td>
<td>1.830847 (0.648751)</td>
<td></td>
</tr>
<tr>
<td>LHRV</td>
<td>0.015719 (0.054137)</td>
<td>0.976620 (0.023895)</td>
<td>1.742325 (0.024776)</td>
<td></td>
</tr>
<tr>
<td>MDR2</td>
<td>0.049303 (16.035449)</td>
<td>0.948179 (72.655723)</td>
<td>3.569842 (50.957657)</td>
<td></td>
</tr>
<tr>
<td>MDRV</td>
<td>0.027384 (1.388565)</td>
<td>0.975913 (0.445478)</td>
<td>1.735309 (0.445478)</td>
<td></td>
</tr>
<tr>
<td>MEM</td>
<td>0.003780 (0.858029)</td>
<td>0.979336 (0.105341)</td>
<td>1.585187 (0.090727)</td>
<td>9.866582 (5.116359)</td>
</tr>
</tbody>
</table>

Panel VI: $m = 288, T = 1000$

<table>
<thead>
<tr>
<th>True Value</th>
<th>$\alpha_m$</th>
<th>$b$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHR2</td>
<td>0.144974 (0.004208)</td>
<td>0.974875 (0.000156)</td>
<td>6.480715 (10.762565)</td>
<td></td>
</tr>
<tr>
<td>LHRV</td>
<td>0.145066 (0.065397)</td>
<td>0.976639 (0.187942)</td>
<td>6.117979 (0.226982)</td>
<td></td>
</tr>
<tr>
<td>MDR2</td>
<td>0.344683 (31.382805)</td>
<td>0.912668 (207.120933)</td>
<td>17.660329 (107.729227)</td>
<td></td>
</tr>
<tr>
<td>MDRV</td>
<td>0.210306 (2.883476)</td>
<td>0.972164 (4.841902)</td>
<td>6.853770 (3.094757)</td>
<td></td>
</tr>
<tr>
<td>MEM</td>
<td>0.126214 (0.591231)</td>
<td>0.977462 (0.273088)</td>
<td>5.971204 (0.299844)</td>
<td>7.479886 (3.057745)</td>
</tr>
</tbody>
</table>
Table 3: HYBRID-GARCH Model: Small sample property, GARCH Diffusion

The table displays estimation of $\alpha_m$, $\beta_m$, $\gamma_m$ (and $g$ for the MEM estimation procedure) of a GARCH diffusion process appearing in equation (3.7) ($\eta = 0$) with sample size 500 (Panel I:III) and sample size 1000 (Panel IV:VI), where the true values of $\alpha_m$, $\beta_m$, $\gamma_m$ are shown in the first line of each panel. The estimators considered are: $mdrv$, the minimizer of (4.3), and the companion estimator $mdr^2$, replacing $RV$ by $R^2$, as well as (quasi-)likelihood-based estimators $lhr^2$, defined in (4.7), and $lhrv$, defined in (4.8). Finally, the table also includes the $mem$ method described in subsection 4.2.2. The numbers in the parenthesis are MSE for $lhr^2$, relative MSE (with respect to $lhr^2$) for $lhrv, mdr^2, mdrv, mem$. For $g$, we only report sample variance.

<table>
<thead>
<tr>
<th>Panel</th>
<th>m</th>
<th>T</th>
<th>True Value</th>
<th>$\alpha_m$</th>
<th>$\beta_m$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>24</td>
<td>500</td>
<td>0.021560</td>
<td>0.606483</td>
<td>0.452303</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lhr2</td>
<td>0.028485</td>
<td>0.574957</td>
<td>0.519297</td>
<td>0.000239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lhrv</td>
<td>0.027866</td>
<td>0.592717</td>
<td>0.463234</td>
<td>0.085556</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mdr2</td>
<td>0.047512</td>
<td>0.554378</td>
<td>0.624201</td>
<td>2.350239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mdrv</td>
<td>0.029201</td>
<td>0.603333</td>
<td>0.444092</td>
<td>0.097035</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mem</td>
<td>0.003728</td>
<td>0.639632</td>
<td>0.439080</td>
<td>7.793987</td>
</tr>
<tr>
<td>II</td>
<td>144</td>
<td>500</td>
<td>0.020402</td>
<td>0.606483</td>
<td>0.452303</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lhr2</td>
<td>0.045201</td>
<td>0.283460</td>
<td>1.658002</td>
<td>0.003203</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lhrv</td>
<td>0.026274</td>
<td>0.285434</td>
<td>2.730282</td>
<td>0.043350</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mdr2</td>
<td>0.068484</td>
<td>0.377976</td>
<td>2.730282</td>
<td>1.846734</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mdrv</td>
<td>0.029991</td>
<td>0.289406</td>
<td>1.169336</td>
<td>0.012388</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mem</td>
<td>0.002005</td>
<td>0.308290</td>
<td>1.159722</td>
<td>26.320110</td>
</tr>
<tr>
<td>III</td>
<td>288</td>
<td>500</td>
<td>0.019472</td>
<td>0.606483</td>
<td>0.452303</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lhr2</td>
<td>0.043392</td>
<td>0.192863</td>
<td>3.269031</td>
<td>0.000315</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lhrv</td>
<td>0.023372</td>
<td>0.172963</td>
<td>1.650080</td>
<td>0.020106</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mdr2</td>
<td>0.072641</td>
<td>0.195808</td>
<td>6.074646</td>
<td>1.321687</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mdrv</td>
<td>0.028927</td>
<td>0.175222</td>
<td>1.667509</td>
<td>0.002766</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mem</td>
<td>0.020868</td>
<td>0.175481</td>
<td>1.935688</td>
<td>35.562272</td>
</tr>
</tbody>
</table>

Table continued on next page ...
Table 3 continued

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_m$</th>
<th>$\beta_m$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0.021560</td>
<td>0.606483</td>
<td>0.452303</td>
<td></td>
</tr>
<tr>
<td>lhr2</td>
<td>0.027721 (0.000070)</td>
<td>0.582303 (0.011110)</td>
<td>0.493834 (0.037197)</td>
<td></td>
</tr>
<tr>
<td>lhrv</td>
<td>0.027869 (0.808425)</td>
<td>0.590124 (0.101996)</td>
<td>0.465888 (0.064436)</td>
<td></td>
</tr>
<tr>
<td>mdr2</td>
<td>0.038240 (23.955470)</td>
<td>0.568033 (3.269306)</td>
<td>0.579447 (12.799811)</td>
<td></td>
</tr>
<tr>
<td>mdrv</td>
<td>0.026646 (1.013888)</td>
<td>0.603776 (0.184239)</td>
<td>0.447793 (0.134316)</td>
<td></td>
</tr>
<tr>
<td>mem</td>
<td>0.002638 (6.866083)</td>
<td>0.640964 (0.235021)</td>
<td>0.434746 (0.068354)</td>
<td>7.730200 (0.187919)</td>
</tr>
</tbody>
</table>

Panel IV: $m = 24$, $T = 1000$

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_m$</th>
<th>$\beta_m$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0.020402</td>
<td>0.294540</td>
<td>1.161865</td>
<td></td>
</tr>
<tr>
<td>lhr2</td>
<td>0.031422 (0.001167)</td>
<td>0.299256 (0.031683)</td>
<td>1.353009 (0.707128)</td>
<td></td>
</tr>
<tr>
<td>lhrv</td>
<td>0.023179 (0.112965)</td>
<td>0.290395 (0.029173)</td>
<td>1.171383 (0.012677)</td>
<td></td>
</tr>
<tr>
<td>mdr2</td>
<td>0.053584 (3.304682)</td>
<td>0.288688 (1.653277)</td>
<td>2.071579 (17.816284)</td>
<td></td>
</tr>
<tr>
<td>mdrv</td>
<td>0.026522 (0.078156)</td>
<td>0.291540 (0.056826)</td>
<td>1.162984 (0.028757)</td>
<td></td>
</tr>
<tr>
<td>mem</td>
<td>0.000591 (0.355572)</td>
<td>0.310927 (0.029184)</td>
<td>1.150915 (0.008941)</td>
<td>26.387767 (2.853065)</td>
</tr>
</tbody>
</table>

Panel V: $m = 144$, $T = 1000$

<table>
<thead>
<tr>
<th></th>
<th>$\alpha_m$</th>
<th>$\beta_m$</th>
<th>$\gamma_m$</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>0.019472</td>
<td>0.177589</td>
<td>1.659011</td>
<td></td>
</tr>
<tr>
<td>lhr2</td>
<td>0.034068 (0.001155)</td>
<td>0.197662 (0.028785)</td>
<td>2.132745 (3.979507)</td>
<td></td>
</tr>
<tr>
<td>lhrv</td>
<td>0.021192 (0.029121)</td>
<td>0.175443 (0.016039)</td>
<td>1.669002 (0.003373)</td>
<td></td>
</tr>
<tr>
<td>mdr2</td>
<td>0.058739 (4.439753)</td>
<td>0.200768 (1.623452)</td>
<td>4.015785 (20.04682)</td>
<td></td>
</tr>
<tr>
<td>mdrv</td>
<td>0.025429 (0.075973)</td>
<td>0.175460 (0.044330)</td>
<td>1.668098 (0.010440)</td>
<td></td>
</tr>
<tr>
<td>mem</td>
<td>0.014977 (0.189630)</td>
<td>0.177831 (0.064847)</td>
<td>1.960999 (1.213615)</td>
<td>35.716633 (92.418742)</td>
</tr>
</tbody>
</table>