Using skewness to estimate the semi-strong GARCH(1,1) model

Todd, Prono

Commodity Futures Trading Commission

11 November 2009
Using Skewness to Estimate the Semi-Strong GARCH(1,1) Model

Todd Prono
Commodity Futures Trading Commission

First Version: November 2009
This Version: May 2011

Abstract

IV estimators with an instrument vector composed only of past squared residuals, while applicable to the semi-strong ARCH(1) model, do not extend to the semi-strong GARCH(1,1) case because of underidentification. Augmenting the instrument vector with past residuals, however, renders traditional IV estimation feasible, if the residuals are skewed. The proposed estimators are much simpler to implement than efficient IV estimators, yet they retain improved finite sample performance over QMLE. Jackknife versions of these estimators deal with the issues caused by many (potentially weak) instruments. A Monte Carlo study is included, as is an empirical application involving foreign currency spot returns.

Keywords: GARCH, GMM, instrumental variables, continuous updating, many moments, robust estimation. JEL codes: C13, C22, C53.

1This paper replaces an earlier version entitled "Simple GMM Estimation of the Semi-Strong GARCH(1,1) Model." I wish to thank Stephen Kane, Celso Brunetti, Pat Fishe, seminar participants at the 2010 International Symposium on Forecasting, the 2010 Society for Computational Economics Conference on Computing in Economics and Finance, the Federal Reserve Board, Camp Econometrics VI, and the Commodity Futures Trading Commission for helpful comments and discussions. The views expressed herein are solely those of the author and do not reflect official positions of the Commodity Futures Trading Commission. In addition, the usual disclaimer applies.

2 Corresponding Author: Todd Prono, Commodity Futures Trading Commission, Office of the Chief Economist, 1155 21st, N.W., Washington, DC 20581. (202) 418-5460, tprono@cftc.gov.
1. Introduction

Despite a plethora of alternative volatility models intended to capture certain "stylized facts" of financial time series, the standard GARCH(1,1) model of Bollerslev (1986) remains the workhorse of conditional heteroskedasticity (CH) modeling in financial economics. By far, the most common estimator for this model is the quasi maximum likelihood estimator (QMLE). Properties of this estimator are well-studied. Weiss (1986) and Lumsdaine (1996) demonstrate that when applied to the strong GARCH(1,1) model, the QMLE is consistent and asymptotically normal (CAN). Bollerslev and Wooldridge (1992), Lee and Hansen (1994), and Escanciano (2009) generalize this result to the semi-strong GARCH(1,1) model. In this paper, I also consider estimation of the semi-strong GARCH(1,1) model, but I do so through the lens of generalized method of moments (GMM) estimators. I propose simple GMM estimators constructed from: (i) the covariances between past residuals and current squared residuals, (ii) the autocovariances between squared residuals. These estimators are asymptotically equivalent to instrumental variables (IV) estimators where the instrument vector is completely contained within the time $t - 1$ information set.

Weiss (1986), Rich, Raymond and Butler (1991), and Guo and Phillips (2001) discuss IV estimators for the ARCH(1) model that are based on the autocovariances between squared residuals. These estimators, however, do not extend to the GARCH(1,1) case because the autocovariances of squared residuals alone are insufficient for identifying the model. I show that the covariances between past residuals and current squared residuals are sufficient for identifying the GARCH(1,1) model, if the residuals are skewed, which differentiates my results from Baillie and Chung (2001) and Kristensen and Linton (2006), who both show that autocorrelations of squared residuals can be used to identify the GARCH(1,1) model. The key identifying assumption for the GMM estimators in this paper, therefore, is unconditional skewness in the residuals being modeled. Such a feature is common in many high frequency financial return series to which the GARCH(1,1) model is applied.

Bollerslev and Wooldridge (1992) recognize that the "results of Chamberlain (1982), Hansen (1982), White (1982), and Cragg (1983) can be extended to produce an instrumental variables estimator asymptotically more efficient than QMLE under nonnormality" (p. 5-6).
for the GARCH(1,1) model. Skoglund (2001) studies this result in detail for the strong GARCH(1,1) model. When applied to the semi-strong GARCH(1,1) model, however, this result necessitates the conditional variance function, its first derivative, as well as the third and fourth conditional moments to be included within the moment conditions. In contrast, the GMM estimators I propose require none of these features. Specifically, neither does the conditional variance function enter the moment conditions nor do the dynamics of the third and fourth moments need to be estimated. These omissions render my estimators simple. Such simplicity, of course, comes at the cost of diminished efficiency. However, even these simple estimators are shown to exhibit superior finite sample performance over QMLE.

The simple GMM estimators I propose are variance targeting estimators (VTE), since the unconditional variance is estimated in a preliminary first step and then plugged into the sample covariances and autocovariances used in a second step. These estimators are shown to be CAN under less restrictive moment existence criteria than in Weiss (1986), Rich, Raymond, and Butler (1991), Baillie and Chung (2001), and Kristensen and Linton (2006). Moreover, the first step variance estimate is shown to have no asymptotic effect on the second step ARCH and GARCH parameter estimates.

Since the proposed estimators are overidentified, the choice of a weighting matrix for the moment conditions is a material concern, especially for finite sample performance. Following Hansen (1982), the standard, optimal, choice for a weighting matrix involves the variance-covariance matrix of the functions comprising the moment conditions. However, since the estimators I propose define moment conditions in terms of the third and possibly the fourth moments, use of the variance-covariance matrix for these particular moment functions involves moment existence criteria up to at least the sixth and possibly the eighth moment. While not so strong as to exclude certain low ARCH, high GARCH processes encountered in empirical applications, such criteria are nevertheless quite strong, especially for certain financial data. Owing to this consideration, I propose a rank dependent correlation matrix as a robust analog to the variance-covariance matrix for use in the weighting matrix of simple GMM estimators for the semi-strong GARCH(1,1) model. This robust analog (i) requires no more than fourth moment existence for consistency, and (ii) provides superior finite sample performance over simple GMM estimators that utilize a non data dependent weighting
matrix like the identity matrix.

Because the proposed GMM estimators are IV estimators where the instrument vector is constructed from past residuals and past squared residuals, there are many potential instruments. From Newey and Windmeijer (2009), the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996) with an optimal weighting matrix is robust to the biases caused by many (potentially weak) instruments, as is the jackknife GMM estimator (JGMM). The finite sample properties of both of these estimators is investigated in the context of semi-strong GARCH(1,1) model estimation. In addition, I propose the jackknife CUE (JCUE) for cases where the optimal weighting matrix is unavailable out of a concern over the existence of higher moments, so the robust analog is used instead. Like the JGMM, the JCUE also removes the term responsible for the many (weak) moments bias from the objective function being minimized. In either the case of the JGMM or the JCUE, consistency is demonstrated without the need for considering the variance-covariance matrix of the moment functions. Doing so avoids the higher moment existence criteria requisite for the optimal CUE (OCUE), thus making the JGMM and the JCUE robust alternatives. Monte Carlo studies show both the OCUE and the JCUE to be more efficient than QMLE in finite samples. These efficiency gains relate to the number of instruments used in constructing the respective estimators.

2. The Model and Implications

For the sequence \( \{Y_t\}_{t \in \mathbb{Z}} \), let \( F_t \) be the associated \( \sigma \)-algebra where \( F_{t-1} \subseteq F_t \subseteq \cdots \subseteq F \). The first two conditional moments of \( Y_t \) are

\[
E [Y_t \mid F_{t-1}] = 0, \quad E [Y_t^2 \mid F_{t-1}] = h_t,
\]

where

\[
h_t = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 h_{t-1}.
\]

In what follows, \( \omega_0 \) denotes the true value, \( \omega \) any one of a set of possible values, and \( \hat{\omega} \) an estimate. Parallel definitions hold for all other parameter values. The model of (1)

\[3\] Throughout this paper, the OCUE refers to the CUE with an optimal weighting matrix.

4
and (2) describes a semi-strong GARCH(1,1) process according to Definition 2 of Drost and Nijman (1993). The more common strong GARCH(1,1) specification where $Y_t/h_t^{1/2}$ is iid and drawn from a known distribution nests as a special case. Consider the following additional assumptions.

**ASSUMPTION A1:** Let $\sigma_0^2 = \frac{\omega_0}{1 - (\alpha_0 + \beta_0)} > 0$, and define $\theta_0 = (\sigma_0^2, \alpha_0, \beta_0)$. $\theta_0 \in \Theta \subseteq \mathbb{R}^3$ is in the interior of $\Theta$, a compact parameter space. For any $\theta \in \Theta$, $\partial \leq \omega \leq W$, $\partial \leq \alpha \leq 1 - \partial$, $0 \leq \beta \leq 1 - \partial$, and $\alpha + \beta < 1$ for some constant $\partial > 0$, where $\partial$ and $W$ are given a priori.

The restrictions on $\theta$ ensure that $h_t$ is everywhere strictly positive. From Lumsdaine (1996), $\alpha$ is strictly positive because if $\alpha = 0$, then $h_t$ is completely deterministic, in which case $\omega_0$ and $\beta_0$ are not separately identified. Since $\beta \geq 0$, A1 nests the ARCH(1) model.

Under A1, $Y_t$ is covariance stationary with $E[Y_t^2] = \sigma_0^2$ following from Theorem 1 of Bollerslev (1986). In this case, the mean-adjusted form of (2) is

$$\tilde{h}_t = \alpha_0 \tilde{Y}_{t-1}^2 + \beta_0 \tilde{h}_{t-1},$$

(3)

where $\tilde{h}_t = h_t - \sigma_0^2$ and $\tilde{Y}_t^2 = Y_t^2 - \sigma_0^2$. An implication of (2) is that

$$\tilde{Y}_t^2 = \tilde{h}_t + W_t,$$

(4)

where $W_t$ is a martingale difference sequence (MDS) by construction, with $E[W_t | F_{t-1}] = 0$ and $E[W_t W_{t-k}] = 0 \forall \; k \geq 1$. Recursively substituting $\tilde{h}_{t-\tau}$ into (3) for $\tau \geq 1$ produces

$$\tilde{h}_t = \sum_{i=0}^{t-1} \alpha_0 \beta_0^i \tilde{Y}_{t-1-i}^2 + \beta_0^t \tilde{h}_0,$$

(5)

for some arbitrary constant $\tilde{h}_0$. Using (5) to solve (4) forward from $t = 1$ setting $\tilde{Y}_0^2 = 0$ produces

$$\tilde{Y}_t^2 = W_t + \alpha_0 \sum_{i=1}^{t-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} + \beta_0 (\alpha_0 + \beta_0)^{t-1} \tilde{h}_0,$$

(6)

which shows that the GARCH(1,1) model relates $\tilde{Y}_t^2$ to a weighted sum of current and past
innovations. A similar recursion is found for the ARCH(p) model in Guo and Phillips (2001).

**ASSUMPTION A2:** (i) $E[W_t Y_t] = \gamma_0 < \infty$, where $\gamma_0 \neq 0$. (ii) The sequence $\{W_t Y_t - \gamma_0\}$ is an $L^1$ mixingale as defined in Andrews (1988) and is uniformly integrable. (iii) The sequences $\{W_{t-l} Y_{t-k}\}$ where $k,l = 1,\ldots,K$ and $k \neq l$ are uniformly integrable.

From (1) and (2),

$$
E[Y_t^3] = E[\hat{Y}_t^2 Y_t] \\
= E[(\hat{h}_t + W_t) Y_t] \\
= E[W_t Y_t].
$$

Under A2(i), $Y_t$ is asymmetrically distributed with a stationary third moment. The process governing the conditional third moment of $Y_t$ is restricted by A2(ii). An $L^1$ mixingale exhibits weak temporal dependence in that the $m$-step-ahead forecast converges (in absolute expected value) to an unconditional mean of zero. This temporal dependence need not decay towards zero at any particular rate and includes certain autoregressive moving average (ARMA) and infinite order moving average (MA) processes. Given the functional form of (2), allowing the third moment to display similar dynamics seems natural. Moreover, Harvey and Siddique (1999) present empirical evidence from stock return data that the conditional third moment follows an ARMA-style process.

Uniform integrability allows a weak LLN to apply to $W_t Y_t - \gamma_0$ and $W_{t-l} Y_{t-k}$ (See Lemma 3 in the Appendix). A sufficient condition for this result is that the given sequence be $L^p$ bounded for some $p > 1$. According to Andrews (1988), however, "it is preferable to impose the uniform integrability assumption rather than an $L^p$ bounded assumption because the former allows for more heterogeneity in the higher order moments of the rv's" (p. 3). This statement guides the formulation of A2(ii) and A2(iii).

**LEMMA 1.** Let Assumptions A1 and A2(i) hold for the model of (1) and (2). Then

$$
E[\hat{Y}_t^2 Y_{t-1}] = \alpha_0 E[W_t Y_t],
$$

6
and
\[
E \left[ \hat{Y}_t^2 Y_{t-(k+1)} \right] = (\alpha_0 + \beta_0) E \left[ \hat{Y}_t^2 Y_{t-k} \right].
\] (9)

**Proof.** All proofs are stated in the Appendix. ■

Lemma 1 relates the covariance between \( Y_t^2 \) and \( Y_{t-k} \) to the third moment of \( Y_t \).\(^4\) Lemma 1 of Guo and Phillips (2001) establishes an analogous result for the ARCH(p) model. In contrast to Guo and Phillips, however, the Lemma presented here is central to identification by providing the moment condition in (8) that is only a function of the data and of \( \alpha_0 \). Separation of \( \alpha_0 \) from \( \beta_0 \) is the direct consequence of a nonzero third moment. Skewness in the distribution of \( Y_t \), therefore, is the key identifying assumption for the simple GMM estimators that I discuss.

Newey and Steigerwald (1997) explore the effects of skewness on the identification of CH models using the QMLE. This paper conducts a similar exploration for certain GMM estimators. Newey and Steigerwald show that given skewness, there exist conditions under which the standard QMLE for CH models is not identified. This paper, in contrast, develops simple GMM estimators that are not identified without such skewness.

**ASSUMPTION A3:** (i) \( E[W_t^2] = \lambda_0 < \infty \). (ii) The sequences \( \{W_t W_{t-k}\} \) are uniformly integrable. (iii) The sequence \( \{W_t^2 - \lambda_0\} \) is an \( L^1 \) mixingale and is uniformly integrable.

Suppose
\[
Y_t = h_t^{1/2} \epsilon_t,
\] (10)
where \( \epsilon_t \) is iid with a mean of zero and a unit variance. Then A3(i) is equivalent to assuming that
\[
(\kappa + 1) \alpha_0^2 + 2\alpha_0 \beta_0 + \beta_0^2 < 1, \quad \kappa = E \left[ \epsilon_t^4 \right] - 1,
\] (11)
which is the necessary and sufficient condition for establishing existence of the fourth moment of \( Y_t \) according to Theorem 1 of Zadrozny (2005).\(^5\) As a consequence, A3(i) strengthens A1 by requiring under the strong GARCH case of (10) that for any \( \theta \in \Theta \), \( (\kappa + 1) \alpha^2 + 2\alpha \beta + \beta^2 < 1 \).

---

\(^4\)See (24) in the Appendix.

\(^5\)If \( \epsilon_t \) is normally distributed, then this inequality follows from Theorem 2 of Bollerslev (1986) with \( \kappa = 2 \).
Of course, in the semi-strong GARCH case of (1) and (2), A3(i) also strengthens A1, but in an unknown way owing to possible dependence in the fourth moment of $\epsilon_t$. Finally, A3(i) also strengthens A2(i) by establishing the existence of the third moment of $Y_t$.

A3(ii)-(iii) permit a weak LLN to apply to the sample autocovariances of $Y_t^2$. A3(iii) assumes that the same general type of process governing the third moment (see A2(ii)) also governs the fourth. This assumption is supported empirically by the results of Hansen (1994).

**LEMMA 2.** *Given the model of (1) and (2), $Y_t^2$ is covariance stationary if and only if A1 and A3(i) hold. In this case,*

$$E\left[\hat{Y}_t^2\hat{Y}_{t-(k+1)}^2\right] = (\alpha_0 + \beta_0)E\left[\hat{Y}_t^2\hat{Y}_{t-(k)}^2\right].$$

(12)

Mark (1988), Bodurtha and Mark (1991), Rich, Raymond, and Butler (1991), as well as Guo and Phillips (2001) estimate ARCH models from the autocovariances of squared residuals. Baillie and Chung (2001) and Kristensen and Linton (2006) estimate the GARCH(1,1) model from the autocorrelations of squared residuals. For any of these cases, the squared residuals need to be covariance stationary. Lemma 2 provides necessary and sufficient conditions for this result and is closely related to Theorem 3 of Hafner (2003) (see also He and Teräsvirta 1999).

(12), like (9), provides moment conditions in terms of the parameters $\alpha_0$ and $\beta_0$. Under Lemma 2, however, there is no analog to (8). As a consequence, the autocovariances of squared residuals alone, while sufficient for identifying the ARCH(1) model, are generally seen as insufficient for identifying the GARCH(1,1) model (see 12 and 26).

Kristensen and Linton (2006) demonstrate how the autocorrelations of squared residuals can identify the GARCH(1,1) model. For this result, the first-order autocorrelation provides the function that separates the ARCH and GARCH parameters in an analogous way to (8). Unlike (8), however, separation of these two parameters results from the solution to a quadratic equation. Moreover, identification in Kristensen and Linton (2006) depends on the existence of the fourth moment (Lemma 1 requires only the third) and requires $\beta_0 > 0$.

---

(see, in contrast, A1).

3. Estimation

3.1. Notation

Partition the parameter vector $\theta$ into $(\lambda, \sigma^2)'$, where $\lambda = (\alpha, \beta)'$. For the sequence of observations $\{Y_t\}_{t=1}^T$ from a data vector $Y$, let $X_{t-2} = [Y_{t-2}, \ldots, Y_{t-k}]'$ and $Z_{t-2} = [Y_{t-2}^2 - \sigma^2, \ldots, Y_{t-k}^2 - \sigma^2]'$ for $2 \leq k \leq K$. Consider the following vector valued functions

$$g_{1, t}(Y; \lambda, \sigma^2) = (Y_t^2 - \sigma^2) Y_{t-1} - \alpha Y_t^3,$$

$$g_{2, t}(Y; \lambda, \sigma^2) = (Y_t^2 - \sigma^2) (X_{t-2} - (\alpha + \beta) X_{t-1}),$$

$$g_{3, t}(Y; \lambda, \sigma^2) = (Y_t^2 - \sigma^2) (Z_{t-2} - (\alpha + \beta) Z_{t-1}).$$
and the following definitions

\[
\begin{align*}
g_{i,t} (Y; \lambda, \sigma^2) &= g_{i,t} (\lambda, \sigma^2), \quad i = 1, 2, 3, \\
g_t (\lambda, \sigma^2) &= \left[ g_{i,t} (\lambda, \sigma^2) \right], \quad i = 1, \ldots, \max (i), \quad 2 \leq \max (i) \leq 3, \\
g_{m,t} (\lambda, \sigma^2) &= \text{mth element of } g_t (\lambda, \sigma^2), \\
\hat{g} (\lambda, \sigma^2) &= T (k)^{-1} \sum_{t=k+1}^{T} g_t (\lambda, \sigma^2), \quad \overline{g} (\lambda, \sigma^2) = E \left[ g_t (\lambda, \sigma^2) \right], \\
m_t (\sigma^2) &= Y_t^2 - \sigma^2, \quad \hat{m} (\sigma^2) = T^{-1} \sum_{t=1}^{T} Y_t^2 - \sigma^2, \\
\bar{g}_t (\lambda, \sigma^2) &= g_t (\lambda, \sigma^2) + S_{\sigma^2} (\lambda, \sigma^2) m_t (\sigma^2), \\
\bar{S}_\lambda (\lambda, \sigma^2) &= \frac{\partial \bar{g} (\lambda, \sigma^2)}{\partial \lambda}, \quad S_\lambda (\lambda, \sigma^2) = E \left[ \frac{\partial g_t (\lambda, \sigma^2)}{\partial \lambda} \right], \\
\bar{S}_{\sigma^2} (\lambda, \sigma^2) &= \frac{\partial \bar{g} (\lambda, \sigma^2)}{\partial \sigma^2}, \quad S_{\sigma^2} (\lambda, \sigma^2) = E \left[ \frac{\partial g_t (\lambda, \sigma^2)}{\partial \sigma^2} \right], \\
\Omega (\lambda, \sigma^2) &= \sum_{s=-(L-1)}^{s=(L-1)} E \left[ g_{t-s} (\lambda, \sigma^2) g_t (\lambda, \sigma^2)^\prime \right], \quad L \geq 1, \\
\bar{\Omega} (\lambda, \sigma^2) &= \sum_{s=-(L-1)}^{s=(L-1)} T (k)^{-1} \sum_{t=k+s+1}^{T} g_{t-s} (\lambda, \sigma^2) g_t (\lambda, \sigma^2)^\prime, \\
R \left[ g_{m,t} (\lambda, \sigma^2) \right] &= \text{rank of } g_{m,t} (\lambda, \sigma^2) \text{ in } g_{m,k+1} (\lambda, \sigma^2), \ldots, g_{m,T} (\lambda, \sigma^2), \\
\hat{\rho}_{m,n}^{(m,n)} (\lambda, \sigma^2) &= 1 - \frac{6}{T (k, s) (T (k, s)^2 - 1)} \sum_{t=k+s+1}^{T} \left( R \left[ g_{m,t} (\lambda, \sigma^2) \right] - R \left[ g_{n,t-s} (\lambda, \sigma^2) \right] \right)^2, \\
\hat{\Sigma} (\lambda, \sigma^2) &= \sum_{s=-(L-1)}^{s=(L-1)} \left[ \hat{\rho}_{m,n}^{(m,n)} (\lambda, \sigma^2) \right],
\end{align*}
\]

where \( m, n = 1, \ldots, 2k-1 \), \( T (k) = T - k \), and \( T (k, s) = T - k - s \).

### 3.2. CAN and Robust Estimators

Consider

\[
\hat{\lambda} = \arg \min_{\lambda \in \Lambda} \hat{g} (\lambda, \hat{\sigma}^2)^\prime M_T \hat{g} (\lambda, \hat{\sigma}^2), \quad (14)
\]

for some sequence of positive semi-definite \( M_T \), which is the familiar GMM estimator of Hansen (1982) with \( \hat{\sigma}^2 \) plugged-in from a preliminary first step. Given this plug-in feature, (14) is also a VTE similar to that studied in Engle and Mezrich (1996) as well as Francq, Horath, and Zakoian (2009). Since (14) minimizes a quadratic objective function, it is
also comparable to the minimum distance estimator (MDE) proposed by Baillie and Chung (2001).

If $M_T = M_T \left( \tilde{\lambda}, \tilde{\sigma}^2 \right)$, where $\tilde{\lambda}$ is a preliminary (and consistent) estimator of $\lambda_0$, then (14) is a two-step GMM estimator. If $M_T = M_T \left( \lambda, \tilde{\sigma}^2 \right)$, then (14) is a CUE. If $\max(i) = 2$, then sample covariances from Lemma 1 form the moment conditions in (14). Supplementing these moment conditions are sample autocovariances from Lemma 2, if $\max(i) = 3$.

To see the asymptotic equivalence of (14) to an IV estimator, redefine (4) as

$$\tilde{Y}^2_t = X'_{-1} \lambda_0 + W_t,$$

where $X_{-1} = \left( \tilde{Y}^2_{t-1}, \tilde{h}_{t-1} \right)'$. Next, let $Z_{-1} \in F_{t-1}$. Since $W_t$ is a MDS,

$$E \left[ Z_{-1} \left( \tilde{Y}^2_t - X'_{-1} \lambda_0 \right) \right] = 0,$$

which defines the population moment conditions for an infeasible IV estimator of $\tilde{h}_t$.

**PROPOSITION.** Let $Z_{-1} = \begin{bmatrix} Y_{t-1} \\ X_{t-2} \\ \tilde{Z}_{t-2} \end{bmatrix}$, where $\tilde{Z}_{t-2} = \left[ \tilde{Y}^2_{t-2}; \cdots; \tilde{Y}^2_{t-k} \right]'$ for $k \geq 2$. Then

$$E \left[ Z_{-1} \left( \tilde{Y}^2_t - X'_{-1} \lambda_0 \right) \right] = g \left( \lambda_0, \sigma^2_0 \right).$$

Given the consistency result of Theorem 1 below, this proposition establishes that (14) converges to the same probability limit as an infeasible IV estimator. Enabling this convergence is the fact that $Cov \left[ Y_i^2; Y_{i-k} \right] = Cov \left[ h_i; Y_{i-k} \right]$, and $Cov \left[ Y_i^2; Y_{i-k}^2 \right] = Cov \left[ h_i; Y_{i-k}^2 \right]$ for $k \geq 1$, since $W_t$ is a MDS, which allows for a restatement of (16) in terms of elements that are observed at time $t$. Of course, (14) is not linear in $\lambda_0$ because (16) is not linear in $\lambda_0$, owing to the dependence of $h_{t-1}$ on $\lambda_0$.

The Proposition uncovers an instrument vector that permits feasible estimation of (16). Notice that this instrument vector omits $\tilde{Y}^2_{t-1}$. If $\tilde{Y}^2_{t-1}$ is included as an instrument, then

---

7In this case, and throughout the ensuing discussions of potential IV estimators, infeasible references the fact that $h_{t-1}$ is not observed at time $t$. 

11
feasible estimation of (16) is no longer possible. To see this, append \( \tilde{Y}_{t-1}^2 \) to the end of \( Z_{-1} \) as
\[
\hat{Z}_{-1} = \begin{pmatrix} Z_{-1} \\ \tilde{Y}_{t-1}^2 \end{pmatrix},
\]
and then substitute \( \hat{Z}_{-1} \) for \( Z_{-1} \) in (16). The final row of \( E \left[ \hat{Z}_{-1} X_{-1} \lambda_0 \right] \) is
\[
\alpha_0 E \left[ \tilde{Y}_{t-1}^4 \right] + \beta_0 E \left[ \tilde{h}_{t-1} \tilde{Y}_{t-1}^2 \right].
\] (17)

Expanding the left term in (17) using (4) produces
\[
E \left[ \tilde{Y}_{t-1}^4 \right] = E \left[ (\tilde{h}_{t-1} + W_{t-1}) \tilde{Y}_{t-1}^2 \right] \\
= E \left[ \tilde{h}_{t-1} \tilde{Y}_{t-1}^2 \right] + E \left[ W_{t-1} \tilde{Y}_{t-1}^2 \right] \\
\neq E \left[ \tilde{h}_{t-1} \tilde{Y}_{t-1}^2 \right],
\]
in general, since \( E \left[ W_{t-1} \tilde{Y}_{t-1}^2 \right] \neq 0 \). As a consequence, (17) can only be simplified to
\[
(\alpha_0 + \beta_0) E \left[ \tilde{Y}_{t-1}^4 \right] - \beta_0 E \left[ W_{t} \tilde{Y}_{t}^2 \right],
\]
which preserves the explicit dependence of (16) on the conditional variance through the contemporaneous covariance between \( W_t \) and \( \tilde{Y}_{t}^2 \).

The move from \( Z_{-1} \) to \( \hat{Z}_{-1} \) represents a progression towards a more efficient IV estimator. The limit to this progression is the Efficient IV estimator analyzed by Skoglund (2001) for the strong GARCH(1,1) model. Generalizing this estimator to the semi-strong case produces
\[
\hat{\vartheta} = \arg \min_{\vartheta \in \Theta} \hat{f}(\vartheta)' \Lambda_T \hat{f}(\vartheta),
\]
where \( \vartheta = (\omega, \alpha, \beta)' \),

\[
\hat{f}(\vartheta) = T^{-1} \sum_{t=1}^{T} f_{t}(\vartheta) = T^{-1} \sum_{t=1}^{T} \left[ f_{i,t}(\vartheta) \right] \text{ for } i = 1, 2, 3,
\]

\[
f_{i,t}(\vartheta) = \frac{1}{\Delta_t} \left( \partial h_t \left( \partial \vartheta_i \right) \right) h_t^{1/2} \left[ \left( \frac{Y_t}{h_t^{1/2}} \right) E \left[ Y_t^3 \mid F_{t-1} \right] - h_t^{3/2} \left( \left( \frac{Y_t^2}{h_t} \right) - 1 \right) \right],
\]

\[
\Delta_t = h_t^3 \left( \frac{E \left[ Y_t^4 \mid F_{t-1} \right]}{h_t^2} - 1 \right) - E \left[ Y_t^3 \mid F_{t-1} \right]^2,
\]

\[
\Lambda_T = \left( T^{-1} \sum_{t} f_{t}(\vartheta)^{f_{t}(\vartheta)'} \right)^{-1}.
\]

The estimator \( \hat{\vartheta} \) depends explicitly on the conditional variance, its first derivative, and on both the third and fourth conditional moments of \( Y_t \). These higher conditional moments either have to be dealt with nonparametrically or assigned parametric forms. The former treatment involves some misspecification bias, since A2(ii) and A3(iii) are non Markovian. The latter treatment, by involving a set of nuisance parameters, requires preliminary estimators and suffers the usual logical inconsistency of requiring additional information from the higher conditional moments but not estimating the associated nuisance parameters simultaneously with the parameters governing the conditional variance (see Meddahi and Renault 1997).

As seen through the Proposition, \( \hat{\lambda} \), in contrast, while clearly dependent on the dynamics of \( h_t \), does not take the conditional variance as an explicit input. Moreover, as seen through Lemmas 1 and 2, \( \hat{\lambda} \) depends on the third and fourth moments of \( Y_t \) only unconditionally, meaning that \( \hat{\lambda} \) does not require estimation of higher moment dynamics beyond the second. The lack of explicit dependence within the moment functions of (14) on (i) the conditional variance and (ii) time-variation in the third and fourth moments renders \( \hat{\lambda} \) a simple estimator for the GARCH(1,1) model within the class of IV estimators discussed above.

Of course, simplicity, in this context, comes at a cost of sacrificed efficiency. \( \hat{\lambda} \) is an asymptotically less efficient estimator than is \( \hat{\vartheta} \). From Skoglund (2001), \( \hat{\vartheta} \) is strictly more efficient asymptotically than its QMLE counterpart if \( \left\{ \frac{Y_t}{h_t^{1/2}} \right\}_{t \in \mathbb{Z}} \) displays excess kurtosis relative to the normal distribution. A question studied in section 5 is the finite sample
efficiency of \( \hat{\lambda} \) relative to QMLE as a means of gauging the cost of estimator simplicity.

**THEOREM 1 (Consistency).** Consider the estimator in (14) for the model of (1) and (2). Let \( \hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2 \), and assume that \( M_T \overset{p}{\to} M_0 \), a positive semi-definite matrix and that \( M_0 \mathbf{g}(\lambda, \sigma_0^2) = 0 \) only if \( \lambda = \lambda_0 \). If \( \max (i) = 2 \), then \( \hat{\lambda} \overset{p}{\to} \lambda_0 \) given Assumptions A1–A2. If \( \max (i) = 3 \), then \( \hat{\lambda} \overset{p}{\to} \lambda_0 \) given Assumptions A1–A3.

Theorem 1 establishes weak consistency of a simple GMM estimator for semi-strong versions of the ARCH(1) and GARCH(1,1) models. When \( \max (i) = 2 \), third moment existence is necessary for this result. When \( \max (i) = 3 \), fourth moment existence becomes necessary, owing to the consideration of autocovariances between squared residuals. Theorem 4.4 of Weiss (1986), Rich et al. (1991), as well as Guo and Phillips (2001) require fourth moment existence for the consistency of their, respective, ARCH model estimators. Baillie and Chung (2001) and Kristensen and Linton (2006) require the same condition for autocorrelation-based estimators of the GARCH(1,1) model. Theorem 1, in contrast, relies on fourth moment existence only as a sufficient condition, provided that skewness is present. In this case, a necessary condition is third moment existence, which allows a relatively milder set of moment existence criteria to establish consistency.

When \( \beta_0 = 0 \), the solution to (14) is

\[
\hat{\alpha} = \left\{ \left( \sum_t \hat{U}_t \right) M_T \left( \sum_t \hat{U}_t \right) \right\}^{-1} \left( \sum_t \hat{U}_t \right) M_T \left( \sum_t \hat{V}_t \right), \tag{18}
\]

\[
\hat{U}_t = \begin{pmatrix}
Y_t^3 \\
(Y_t^2 - \hat{\sigma}^2) X_{t-1}
\end{pmatrix}, \quad \hat{V}_t = \begin{pmatrix}
(Y_t^2 - \hat{\sigma}^2) Y_{t-1} \\
(Y_t^2 - \hat{\sigma}^2) X_{t-2}
\end{pmatrix},
\]

if either \( M_T \) does not depend on \( \alpha \) or \( M_T = M_T (\hat{\alpha}, \hat{\sigma}^2) \). Given the Proposition, (18) is asymptotically equivalent to

\[
\hat{\alpha} = \left\{ \left( \sum_t \hat{Z}_{t-1} (Y_{t-1}^2 - \hat{\sigma}^2) \right) N_T \left( \sum_t \hat{Z}_{t-1} (Y_{t-1}^2 - \hat{\sigma}^2) \right) \right\}^{-1} \left( \sum_t \hat{Z}_{t-1} (Y_{t-1}^2 - \hat{\sigma}^2) \right) N_T \left( \sum_t \hat{Z}_{t-1} (Y_{t-1}^2 - \hat{\sigma}^2) \right)
\]

if \( N_T \overset{p}{\to} M_0 \), where \( \hat{\alpha} \) is a generalized IV estimator based on the population moment con-
conditions $E \left[ Z_{-1} \left( \tilde{Y}_t^2 - \alpha_0 \tilde{Y}_{t-1}^2 \right) \right] = 0$. In the special case of an ARCH(1) process, $Z_{-1}$ can be substituted for $Z_{-1}$ without affecting the feasibility of the IV estimator, given the result from (17). Such a substitution is asymptotically equivalent to appending the vector valued function

$$g_{4,t}(\alpha, \hat{\sigma}^2) = (Y_t^2 - \hat{\sigma}^2) \left( (Y_{t-1}^2 - \hat{\sigma}^2) - \alpha (Y_t^2 - \hat{\sigma}^2) \right)$$

(19) to $g_t(\alpha, \hat{\sigma}^2)$.

**THEOREM 2 (Asymptotic Normality).** Consider the estimator in (14) for the model of (1) and (2), letting $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2$. Assume (i) $M_T \rightarrow M_0$, a positive semi-definite matrix and that $M_0 g(\lambda, \sigma_0^2) = 0$ only if $\lambda = \lambda_0$; (ii) either Assumptions A1–A2 hold if $\max(i) = 2$, or Assumptions A1–A3 hold if $\max(i) = 3$; (iii) $S_{\lambda}(\lambda_0, \sigma_0^2)' M_0 \times S_{\lambda}(\lambda_0, \sigma_0^2)$ is nonsingular; (iv) $\sqrt{T(k)} \tilde{g}(\lambda_0, \sigma_0^2) \xrightarrow{d} N \left( 0, \Omega(\lambda_0, \sigma_0^2) \right)$.

Then

$$\sqrt{T(k)} \left( \hat{\lambda} - \lambda_0 \right) \xrightarrow{d} N \left( 0, H(\lambda_0, \sigma_0^2)^{-1} S_{\lambda}(\lambda_0, \sigma_0^2)' M_0 \Omega(\lambda_0, \sigma_0^2) M_0 S_{\lambda}(\lambda_0, \sigma_0^2) H(\lambda_0, \sigma_0^2)^{-1} \right),$$

where $H(\lambda_0, \sigma_0^2) = S_{\lambda}(\lambda_0, \sigma_0^2)' M_0 S_{\lambda}(\lambda_0, \sigma_0^2)$.

As a VTE, (14) is a two-step estimator, since the objective function is minimized conditional on a preliminary, or first-step, estimator $\hat{\sigma}^2$. In general, the variance of a first-step estimator impacts the variance of the second-step (see Newey and McFadden 1994). Under Theorem 2, this impact is seen through

$$\tilde{\Omega}(\lambda_0, \sigma_0^2) = \sum_{s=-(L-1)}^{s=(L-1)} E \left[ \tilde{g}_{t-s}(\lambda_0, \sigma_0^2) \tilde{g}_t(\lambda_0, \sigma_0^2)' \right],$$

which is the variance-covariance matrix of

$$\sqrt{T(k)} \tilde{g}(\lambda_0, \hat{\sigma}^2) = \sqrt{T(k)} \left\{ \tilde{g}(\lambda_0, \sigma_0^2) + S_{\sigma^2}(\lambda_0, \sigma_0^2) \tilde{m}(\sigma_0^2) \right\},$$

(20)

the term to which a Central Limit Theorem (CLT) is applied when deriving asymptotic normality. The second quantity on the right-hand-side of the equality in (20) sources the effect of $\hat{\sigma}^2$ on the asymptotic variance of $\hat{\lambda}$. Given Lemma 4 stated in the Appendix,
however, \( S_{\alpha^2} (\lambda_0, \sigma_0^2) = 0 \), which means that \( \tilde{g} (\lambda_0, \tilde{\sigma}^2) = \tilde{g} (\lambda_0, \sigma_0^2) \), \( \tilde{\Omega} (\lambda_0, \sigma_0^2) = \Omega (\lambda_0, \sigma_0^2) \), and, as a consequence, nothing is lost (asymptotically) by plugging \( \tilde{\sigma}^2 \) into (14) as opposed to \( \sigma_0^2 \). This result stands in contrast to the VTE studied by Francq, Horath, and Zakoian (2009), where the variance of \( \tilde{\sigma}^2 \) does, in fact, impact the variance of \( \hat{\lambda} \) asymptotically.

If \( g_t (\lambda_0, \sigma_0^2) \) is a MDS (the assumption made in Sections 5 and 6), then condition (iv) of Theorem 2 follows if \( E [ \| g_t (\lambda_0, \sigma_0^2) \|^2 ] < \infty \). Other CLTs for dependent data may also prove applicable, depending on the process for \( g_t (\lambda_0, \sigma_0^2) \). This process depends, in turn, on the processes governing \( W_t Y_t \) and \( W_t^2 \). The fact that temporal dependence in each of these, respective, sequences is only generally specified motivates condition (iv).

Theorem 4.4 of Weiss (1986) demonstrates the CAN property of an autocovariance-based estimator for the ARCH model if the eighth moment of residuals exists. Kristensen and Linton (2006) rely on this same condition in demonstrating their autocorrelation-based estimator to be CAN. Eighth moment existence is only a sufficient condition under Theorem 2. Provided that skewness is present, (14) is CAN given existence of the sixth moment.

Of course, the rather complicated asymptotic variance formula in Theorem 2 simplifies to the more familiar \( H (\lambda_0, \sigma_0^2)^{-1} \) if \( M_0 = \Omega (\lambda_0, \sigma_0^2)^{-1} \). From Hansen (1982), this choice of weighting matrix is optimal, since it minimizes the asymptotic variance of (14). Given this choice and provided that skewness is present, (14) can be expected to be more efficient asymptotically than the MDS estimator of Baillie and Chung (2001), since the former utilizes information from the third moment.

Rather than relying on asymptotic approximations (and the higher moment existence criteria those approximations entail), standard errors for (14) can be computed via the parametric bootstrap. Suppose that the data generating process for \( Y_t \) is characterized by (1), (2), and (10), where \( E [ \epsilon_t \mid F_{t-1} ] = 0 \), \( E [ \epsilon_t^2 \mid F_{t-1} ] = 1 \), and the higher moments of \( \epsilon_t \) follow \( L^{th} \) order Markov processes with a finite \( L \ll T \). Use (14) to obtain \( \hat{h}_t \). Let

\[ ^{8}\text{This result, perhaps, is not surprising given the Proposition and the demonstration in Wooldridge (1994) p. 2695-2696 that for an instrumental variable function defined in terms of some nuisance parameters, the limiting distribution of those nuisance parameters does not affect that of the parameters of interest if the nuisance parameters are consistently estimated.} \]

\[ ^{9}\text{The proof to Theorem 2 is based on the two-step GMM estimator. For the CUE, although the first order condition analogous to (31) contains an additional term, this term does not distort the limiting distribution. Pakes and Pollard (1989) discuss this result in detail as do Donald and Newey (2000).} \]
\( \hat{\varepsilon}_t = Y_t / \sqrt{h_t} \), and apply the nonoverlapping block bootstrap method of Carlstein (1986) to
these standardized residuals to obtain the bootstrap sample \( \hat{\varepsilon}_t^* \). Use these bootstrap residuals
to construct the series \( \hat{Y}_t^* = \sqrt{h_t^* \hat{\varepsilon}_t^*} \), where \( h_t^* \) depends on the parameter estimates from
the original data sample. Estimate the model of (1) and (2) on \( \hat{Y}_t^* \), making sure to center the
bootstrap moment conditions with the original parameter estimates as suggested in Hall and
Horowitz (1996). Repetition of this procedure permits the calculation of bootstrap standard
errors for \( \hat{\theta} \) that are robust to higher moment dynamics in \( \varepsilon_t \). This same procedure can
also be used to bootstrap the GMM objective function as discussed in Brown and Newey
(2002) for a non-parametric test of the overidentifying restrictions that speaks to the fit of
the GARCH(1,1) model to the given data under study.

### 3.3. The Weighting Matrix

The estimator in (14) requires specification of a weighting matrix. Use of the optimal
weighting matrix under Theorem 2 requires existence of, at least, the sixth moment and as
high as the eighth if autocovariances are also considered. Such an assumption may prove
overly restrictive, especially for certain financial data. A key question, therefore, is what
potential weighting matrices exist that economize on the number of higher moment exist-
ence criteria needed for consistency. One option, of course, is to use a non data dependent
weighting matrix like the identity matrix. Skoglund (2001), however, reports that the iden-
tity matrix used in the Efficient IV estimator for the strong GARCH(1,1) model results
in quite poor finite sample performance. This result is also found (though not reported)
in Monte Carlo studies of (14). Alternatively, one can consider using a robust analog to
\( \hat{\Omega}(\hat{\theta}) \) when constructing the weighting matrix. One such alternative is \( \hat{\Sigma}(\hat{\theta}) \). The matrix
\( \left[ p_{t,s}^{(m,n)}(\hat{\theta}) \right] \) is Spearman’s (1904) correlation matrix for the vector valued functions \( g_t(\hat{\theta}) \)
and \( g_{t-s}(\hat{\theta}) \). The matrix \( \hat{\Sigma}(\hat{\theta}) \), therefore, reflects rank dependent measures of contempora-
neous and lagged association between the sequences of vector valued functions that comprise
the moment conditions. The following lemma is useful for establishing consistency of \( \hat{\Sigma}(\hat{\theta}) \).

**Lemma 5.** Let \( a_{t,s}(\theta) = \{ R[g_{m,t}(\theta)] - R[g_{n,t-s}(\theta)] \}^2 \). For a \( \delta_t \to 0 \), define \( \Delta_{t,s}(\theta) = \sup_{\| \theta - \theta_0 \| \leq \delta_t} \| a_{t,s}(\theta) - a_{t,s}(\theta_0) \| \). Assume that \( E[\Delta_{t,s}(\theta)] < \infty \). Then for \( \hat{\theta} \overset{p}{\to} \theta_0 \),
\[ \hat{\rho}_{t,s}^{(m,n)}(\hat{\theta}) - \hat{\rho}_{t,s}^{(m,n)}(\theta_0) \xrightarrow{p} 0. \]

Consistency of \(\hat{\rho}_{t,s}^{(m,n)}(\hat{\theta})\) follows from Lemma 5 and selected results in Schmid and Schmidt (2007).\(^\text{10}\) Conditions for consistency involve the copula for \(g_{m,t}(\theta_0)\) and \(g_{n,t-s}(\theta_0)\) (specifically, existence and continuity of its partial derivatives), but do not explicitly impose higher moment existence criteria on either. It is in this sense, therefore, that \(\hat{\Sigma}(\hat{\theta})\) can be thought of as robust.

4. Many (Weak) Moments Bias Correction

For the estimator in (14), \(k\) (the number of lags, which corresponds to the number of instruments) needs to be specified. Standard GMM asymptotics point to efficiency gains from increasing \(k\). Work by Stock and Wright (2000), Newey and Smith (2004), Han and Phillips (2006), and Newey and Windmeijer (2009), however, discuss the biases of GMM estimators when the instrument vector is large, (possibly) inclusive of (many) weak instruments, and allowed to grow with the sample size. To see how these biases relate to \(k\), suppose that there exists a finite \(L\) such that \(E[g_t(\theta) \mid F_{t-L}]\) is constant.\(^\text{11}\) Let \(s^* = \{S : s \geq t + L \text{ or } s \leq t - L; s = 1, \ldots, T\}\). Then, the expectation of the GMM objective function \(\hat{g}(\theta)'M_T\hat{g}(\theta)\) for a nonrandom weighting matrix \(M_T\) is

\[
E[\hat{g}(\theta)'M_T\hat{g}(\theta)] = T(k)^{-2}E\left[\sum_{t \neq S} g_t(\theta)'M_Tg_s(\theta) + \sum_t g_t(\theta)'M_Tg_t(\theta)\right]
\]

\[
= T(k)^{-2}E\left[\sum_{t \neq S} g_t(\theta)'M_Tg_s(\theta) + \sum_{s=(L-1)}^{s=(L-1)} \sum_t g_t(\theta)'M_Tg_{t-s}(\theta)\right]
\]

\[
= \left(1 - \frac{L}{T(k)}\right)\hat{g}(\theta)'M_T\hat{g}(\theta) + T(k)^{-1}\sum_{s=(L-1)}^{s=(L-1)} E[ g_t(\theta)'M_Tg_{t-s}(\theta) ]
\]

\[
= \left(1 - \frac{L}{T(k)}\right)\hat{g}(\theta)'M_T\hat{g}(\theta) + T(k)^{-1}\text{tr}\left( M_T \sum_{s=(L-1)}^{s=(L-1)} E[ g_{t-s}(\theta)g_t(\theta)' ] \right),
\]

\(^{10}\)These results are Theorem 5 and the fact that \(\lim_{n \rightarrow \infty} \sqrt{n} \{ \hat{\rho}_{1,n} - \hat{\rho}_{S,n} \} = 0\), where \(\hat{\rho}_{S,n}\) relates to \(\hat{\rho}_{t,s}^{(m,n)}(\theta_0)\).

\(^{11}\)\(g_t(\theta)\) can be thought of as a vector of residuals. The requirement is satisfied if these residuals follow an MA process of order \(L - 1\).
which is an adaptation of (2) in Newey and Windmeijer (2009) to dependent time series data.\textsuperscript{12}

In the language of Newey and Windmeijer (2009), \( 1 - \frac{L}{T(k)} \) is a "signal" term minimized at \( \theta_0 \). The second term is a "noise" term that is, generally, not minimized at \( \theta_0 \) if \( \frac{\partial g_t(\theta)}{\partial \theta} \) is correlated with \( g_t(\theta) \) and is increasing in \( k \).\textsuperscript{13} If \( k \) is increasing with \( T \), this bias term need not even vanish asymptotically (see Han and Phillips 2006).\textsuperscript{14}

Suppose that \( M_T = \Omega(\theta)^{-1} \). In this case, the "noise" term

\[
T(k)^{-1} \text{tr} \left( M_T \sum_{s=-(L-1)}^{s=(L-1)} \mathbb{E} \left[ g_{t-s}(\theta) g_t(\theta) \right] \right) = \frac{m(k)}{T(k)}, \quad m(k) = 2k - 1,
\]

which is no longer a function of \( \theta \). For the estimator in (14),

\[
\hat{g}(\lambda, \tilde{\sigma}^2)' M_T \hat{g}(\lambda, \tilde{\sigma}^2) = T(k)^{-2} \left\{ \sum_{t \neq s} g_t(\lambda, \tilde{\sigma}^2) M_T g_s(\lambda, \tilde{\sigma}^2) + \sum_t g_t(\lambda, \tilde{\sigma}^2) M_T g_t(\lambda, \tilde{\sigma}^2) \right\}
\]

\[= T(k)^{-2} \sum_{t \in s^*} g_t(\lambda, \tilde{\sigma}^2)' M_T g_s^* (\lambda, \tilde{\sigma}^2) + T(k)^{-2} \sum_{s=-(L-1)}^{s=(L-1)} \sum_t g_t(\lambda, \tilde{\sigma}^2) M_T g_{t-s}(\lambda, \tilde{\sigma}^2)
\]

\[= T(k)^{-2} \sum_{t \in s^*} g_t(\lambda, \tilde{\sigma}^2)' M_T g_s^* (\lambda, \tilde{\sigma}^2) + T(k)^{-1} \text{tr} \left( M_T \left\{ \sum_{s=-(L-1)}^{s=(L-1)} T(k)^{-1} \sum_t g_{t-s}(\lambda, \tilde{\sigma}^2) g_t(\lambda, \tilde{\sigma}^2)' \right\} \right)
\]

If \( M_T = \hat{\Omega}(\lambda, \tilde{\sigma}^2)^{-1} \), the feasible version of \( \Omega(\lambda, \tilde{\sigma}^2)^{-1} \), then

\[
\hat{g}(\lambda, \tilde{\sigma}^2)' M_T \hat{g}(\lambda, \tilde{\sigma}^2) = T(k)^{-2} \sum_{t \in s^*} g_t(\lambda, \tilde{\sigma}^2)' M_T g_s^* (\lambda, \tilde{\sigma}^2) + \frac{m(k)}{T(k)},
\]

which shows that (14) is robust to many (potentially weak) instruments if it is specified as the OCUE. If, on the other hand, either (i) \( M_T = \hat{\Sigma}(\lambda, \tilde{\sigma}^2)^{-1} \), in which case \( \hat{\lambda} \) is a robust CUE, (ii) \( M_T = \hat{\Omega}(\hat{\lambda}, \tilde{\sigma}^2)^{-1} \), in which case \( \hat{\lambda} \) is the optimal two-step GMM estimator, or

\textsuperscript{12}This expansion is also valid under a random \( M_T \) because estimation of \( M_T \) does not effect the limiting distribution.

\textsuperscript{13}This "noise" or bias term is analogous to the higher order bias term \( B_{G} \) in Newey and Smith (2004).

\textsuperscript{14}Under both theorems, however, \( k \) is treated as fixed so that (14) is consistent.
(iii) \( M_T = \sum (\lambda, \hat{\sigma}^2)^{-1} \), in which case \( \lambda \) is a robust two-step GMM estimator, (14) will be biased. The expansion of \( \hat{g} (\lambda, \hat{\sigma}^2)' M_T \hat{g} (\lambda, \hat{\sigma}^2) \) offers a way to correct for this bias. Namely, consider the alternative estimator

\[
\lambda = \arg \min_{\lambda \in \Lambda} \tilde{Q} (\lambda, \hat{\sigma}^2),
\]

where

\[
\tilde{Q} (\lambda, \hat{\sigma}^2) = T (k)^{-2} \sum_{t \in s^*} g_t (\lambda, \hat{\sigma}^2)' M_T g_s (\lambda, \hat{\sigma}^2)
\]

\[
= \tilde{Q} (\lambda, \hat{\sigma}^2) - T (k)^{-1} \text{tr} \left\{ M_T \left\{ \sum_{s=-(L-1)}^{s=(L-1)} T (k)^{-1} \sum_{t-s} g_t (\lambda, \hat{\sigma}^2) g_t (\lambda, \hat{\sigma}^2)' \right\} \right\},
\]

and \( \tilde{Q} (\lambda, \hat{\sigma}^2) = \hat{g} (\lambda, \hat{\sigma}^2)' M_T \hat{g} (\lambda, \hat{\sigma}^2) \). Depending on the choice of \( M_T \), (21) will be referred to, generally, as either as a JGMM or a JCUE because, as seen through (22), it leaves out contemporaneous and certain lagged observations from either the GMM or CUE objective function. \( \lambda \) is consistent given the following corollary.

**COROLLARY (Consistency).** Consider the estimator in (21) for the model of (1) and (2). Let \( \hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2 \), and assume that (i) \( M_T \xrightarrow{p} M_0 \), a positive semi-definite matrix, (ii) \( M_0 \hat{g} (\lambda, \sigma_0^2) = 0 \) only if \( \lambda = \lambda_0 \), (iii) \( L = 1 \). If \( \max (i) = 2 \), then \( \lambda \xrightarrow{p} \lambda_0 \) given Assumptions A1–A2. If \( \max (i) = 3 \), then \( \lambda \xrightarrow{p} \lambda_0 \) given Assumptions A1–A3.

With \( L = 1 \), (21) is the Jackknife GMM estimator of Newey and Windmeijer (2009). A straightforward way of demonstrating consistency of this estimator is by examining the second equality in (22), in which case, conditions under Theorem 2 are sufficient. By involving the variance-covariance matrix of the moment conditions through the bias correction term, however, such a demonstration involves precisely those higher moment existence criteria that I am looking to avoid when specifying (21). The Corollary, therefore, bases consistency on the first equality in (22) and shows that the conditions under Theorem 1 are sufficient.\(^{15}\) As a result, if either \( M_T = \sum (\lambda, \hat{\sigma}^2)^{-1} \) or \( M_T = \sum (\lambda, \hat{\sigma}^2)^{-1} \), \( \lambda \) is robust in the dual sense that

\(^{15}\)This result assumes, of course, that \( M_T \) is not constructed from \( \Omega (\lambda, \hat{\sigma}^2) \).
it (i) requires the same moment existence criteria as Theorem 1 if \( M_T = I \), and (ii) is free of many (weak) moments bias. Following from Newey and Windmeijer (2009) p. 702, \( \tilde{\lambda} \) is asymptotically normal if \( L = 1 \).

If \( \beta_0 = 0 \) and either \( M_T \) is nonrandom or \( M_T = M_T (\bar{\alpha}, \delta^2) \), then the solution to (21) is

\[
\tilde{\alpha} = \left\{ \sum_{t \in s^*} \hat{U}_t' M_T \hat{U}_{s^*} \right\}^{-1} \sum_{t \in s^*} \hat{U}_t' M_T \hat{V}_{s^*},
\]

which is JIVE2 from Angrist, Imbens, and Krueger (1999) if \( L = 1 \).

5. Monte Carlo

Consider the data generating process in (1), (2), and (10), where \( \epsilon_t \) is the negative of a standardized Gamma(2,1) random variable. The skewness and kurtosis of \( \epsilon_t \) is \(-2/\sqrt{2} \) and 6, respectively. Values for \( \theta_0 \) of \((1.0, 0.15, 0.75)\)', \((1.0, 0.10, 0.85)\)', and \((1.0, 0.05, 0.94)\)' are considered. These values together with the distributional assumption for \( \epsilon_t \) support a finite fourth moment for \( Y_t \) according to (11). All simulations are conducted with 5,000 observations across 500 trials. In each simulation, the first 200 observations are dropped to avoid initialization effects. Starting values for \( \lambda \) in each simulation trial are the true parameter values. Summary statistics for the simulations include the median bias, decile range (defined as the difference between the 90th and the 10th percentiles), standard deviation, and median absolute error (measured with respect to the true parameter value) of the given parameter estimates. The median bias, decile range, and median absolute error are robust measures of central tendency, dispersion, and accuracy, respectively, reported out of a concern over the existence of higher moments. The standard deviation, while not a robust measure, provides an indication of outliers.

Table 1 summarizes the results for (14) and (21), benchmarking them against the QMLE. The forms of (14) and (21) considered: (i) utilize the method of moments plug-in estimator \( \hat{\sigma}^2 = T^{-1} \sum Y_t^2 \), (ii) rely on moments either up to the third or up to the fourth (i.e., set \( \max (i) = 2 \) or 3), (iii) use the inverse of Spearman’s correlation matrix as the data dependent weighting matrix, (iv) set \( k = 20 \) and \( L = 1 \).\(^{16}\)

\(^{16}\)In some of the simulations, an alternative rank dependent correlation matrix based on Kendall’s (1938)
For estimating $\alpha_0$ and $\beta_0$, GMM tends to be associated with the highest bias. JCUE3 has the lowest bias, most comparable to QMLE. CUE3, however, also tends to be associated with low bias. JGMM3 improves upon the bias relative to GMM3 for both $\hat{\alpha}$ and $\hat{\beta}$. The same can be said for JGMM2 relative to GMM2 for $\hat{\beta}$, with mixed results (in terms of bias reduction) evidenced for $\hat{\alpha}$. JCUE3 records less bias than CUE3 for both $\hat{\alpha}$ and $\hat{\beta}$. JCUE2 records less bias than CUE2 for $\hat{\beta}$ but mixed results (in terms of bias reduction) for $\hat{\alpha}$. In some cases, movements from $\max(i) = 2$ to $\max(i) = 3$ are associated with sizable reductions in bias. This result is particularly relevant for non-jackknifed estimators, although it also holds for $\hat{\alpha}$ under the jackknifed CUE. Though not reported here, the bias of non jackknifed estimators for $\hat{\beta}$ tends to increase with $k$. The level of this bias is most noticeable for high values of $\beta_0$.

In terms of dispersion, GMM tends to also record the highest values. However, in limited instances, the JGMM and CUE estimates can be even more dispersed (see, for instance, JGMM2 and CUE2 relative to GMM2 for the estimates of $\beta_0 = 0.94$). JCUE3 records the lowest parameter dispersion most comparable to QMLE in terms of magnitude. CUE3 also supports relatively low levels of parameter dispersion. JGMM3 is more efficient than GMM3 measured either in terms of decile range or median absolute error. The same is mostly true for both JCUE2 and JCUE3 relative to CUE2 and CUE3, with the differences being more noticeable for $\hat{\beta}$. JGMM2 is more efficient than GMM2 for $\hat{\alpha}$, with mixed results appearing for $\hat{\beta}$. In general, movements from $\max(i) = 2$ to $\max(i) = 3$ are associated with large drops in parameter dispersion (i.e., increases in efficiency).

The results from Table 1 show JCUE3 to be a more efficient estimator of $\alpha_0$ but a less efficient estimator of $\beta_0$ when compared to QMLE. Figure 1 compares the efficiency of JCUE3 relative to QMLE (for both $\hat{\alpha}$ and $\hat{\beta}$) for various lag lengths out to $k = 40$. As is evidenced, $\hat{\alpha}$ remains more efficient under JCUE3 as opposed to QMLE for all lag lengths considered. Moreover, the efficiency of $\hat{\beta}$ under JCUE3 is seen to approach that of QMLE as $k \to 40$. These results show that JCUE3 can be more efficient than QMLE given a sufficient number of instruments (still small relative to the sample size). Baillie and Chung (2001) report a similar finding for the MDS estimator they consider.

tau was also tried. The results were very similar to those based on Spearman’s measure. Since Spearman’s measure requires much less computation time, it was favored.
Of the parameter values considered, $\theta_0 = (1.0, 0.05, 0.94)'$ is the most likely to support a finite eighth moment.\footnote{If $\epsilon_t \sim N(0, 1)$, then these values would support a finite eighth moment according to Figure 2 of Bollerslev (1986). In general, for covariance stationary GARCH(1,1) processes, the magnitude of $\alpha_0$ is a principal constraint on the existence of higher moments.} Figure 2, therefore, compares the efficiency of JCUE3, OCUE3, and QMLE for lags lengths out to $k = 40$. Similar to Figure 1, $\alpha\hat{}$ remains more efficiently estimated under JCUE3 than under QMLE for all lag lengths considered. Interestingly, at low levels of $k$, $\alpha\hat{}$ is less efficiently estimated under OCUE3 than under either JCUE3 or QMLE. As $k$ increases, however, the performance of $\alpha\hat{}$ under OCUE3 converges to that of JCUE3, therefore passing that of QMLE. In terms of $\beta\hat{}$, OCUE3 is more efficient than JCUE3 for all lag lengths considered. At low levels of $k$, QMLE is more efficient than both. However, as $k \to 40$, the performance of $\beta\hat{}$ under JCUE3 approaches that under QMLE, while the performance of $\beta\hat{}$ under OCUE3 better that of QMLE. Therefore, both JCUE3 and OCUE3 can be more efficient than QMLE, again given a sufficient number of instruments. In addition, the results for OCUE3 support the claim that while strong, the moment existence criteria of Theorem 2 are not so strong as to exclude all GARCH(1,1) processes of empirical relevance.

Table 2 summarizes simulation results for the JCUE3, JGMM3, and CUE3 (again, benchmarking against the QMLE) in the case where $\epsilon_t$ is the negative of a standardized Gamma(1,1) random variable with skewness of $-2$ and kurtosis of 12. JCUE3 remains the most efficient moments estimator, more efficient than QMLE in estimating $\alpha_0$ and closest to QMLE, in terms of both bias and efficiency, in estimating $\beta_0$. CUE3 no longer dominates JGMM3 in terms of dispersion as it does in Table 1. To the contrary, $\alpha\hat{}$ and $\beta\hat{}$ tend to be less dispersed under JGMM3 (very noticeably so for $\beta\hat{}$ when $\beta_0 = 0.85$ and $\beta_0 = 0.94$). JGMM3, however, displays significantly higher bias in $\alpha\hat{}$ under both $\alpha_0 = 0.15$ and $\alpha_0 = 0.10$ when $\epsilon_t$ is the negative of a standardized Gamma(1,1) as opposed to the negative of a standardized Gamma(2,1).

The Ratio statistics in Table 2 show that dispersion tends to increase when moving to an increasingly skewed, fatter-tailed distribution for the standardized residuals. Exceptions to this tendency occur only for the moments estimators, only for $\alpha\hat{}$, and most consistently
for JGMM3. Specifically for JGMM3, the Ratio statistic for both the Decile Range and SD of $\hat{\alpha}$ is less than one for all the cases considered. This result, perhaps, is not so surprising given that skewness is what identifies $\alpha_0$.

Of all the proposed moments estimators, JCUE3 and OCUE3 have the smallest biases and are the most efficient. In general, the smallest biases are achieved using the class of estimators that are robust to many (potentially weak) instruments (i.e., JCUE, JGMM, and OCUE). The worst performing estimators both in terms of bias and in terms of efficiency are the two-step GMM estimators. Fourth moment based estimators (i.e., those with $\max(i) = 3$) tend to outperform third moment based estimators (i.e., those with $\max(i) = 2$) in terms of bias and efficiency by wide margins. For the subclass of estimators with $\max(i) = 2$, JCUE2 records the smallest bias and is the most efficient followed, for the most part, by JGMM2.

6. FX Spot Returns

Let $S_{i,t}$ be the spot rate of foreign currency $i$ measured in US Dollars, where $i = $ Australian Dollars (AUD) or Japanese Yen (JPY). Each spot series is measured daily from 1/1/90 - 12/31/09 and is obtained from Bloomberg. Consider the spot return defined as $Y_{i,t} = \log \left( \frac{S_{i,t}}{S_{i,t-1}} \right)$. This section fits the GARCH(1,1) model of (1) and (2) to $\{Y_{i,t}\}_{t=1}^T$.\(^{18}\) Engle and Gonzalez-Rivera (1999) as well as Hansen and Lunde (2005) employ similar specifications to British Pound and Deutsche Mark exchange rate series, respectively. Hansen and Lunde (2005) find no evidence that the simple GARCH(1,1) specification is outperformed by more complicated volatility models in their study of exchange rates. Their work guides the selection of financial data analyzed here.

For the AUD series, skewness is $-0.33$, and kurtosis is $15.05$. For the JPY series, skewness is $0.43$, and kurtosis is $8.34$. Both series appear decidedly non-normal with the requisite distributional asymmetry required under A2. Table 3 reports the estimation results for JCUE3, OCUE3, and QMLE. Both JCUE3 and OCUE3 utilize an, admittedly, arbitrary lag length of 40 in the specification of their instrument vector. They, additionally, set $\max(i) = 3$\(^{18}\)

\(^{18}\) Preliminary investigations fit, among other specifications, ARMA(1,1) filters to both series. For the JPY series, this filter was insignificant. For the AUD series, it proved significant; however, its removal had no meaningful impact on the GARCH estimates.
and $L = 1$. From the discussion in section 5, an application of OCUE3 is limited to high GARCH-, low ARCH-type processes. The QMLE estimates imply that such processes are appropriate characterizations of both spot return series. Starting values for JCUE3 and OCUE3 are the QMLE estimates.

From Table 3, the JCUE3 estimates are closer to the QMLE estimates than are the OCUE3 estimates. The JCUE3 estimates imply a less persistent volatility process than either the QMLE or OCUE3 estimates. The standard errors for the OCUE3 estimates are larger than their QMLE counterparts, particularly so for $\hat{\alpha}$. The $\hat{\beta}$ standard errors are more comparable. The higher standard errors under OCUE3 may relate to the fact that $\hat{\alpha} + \hat{\beta}$ is close to one.

To investigate the effects of lag length on JCUE3 and OCUE3, each were fit to the two spot return series for $k = 20, \ldots, 40$. For each $k$, $\|\hat{\lambda}_j - \hat{\lambda}_{QMLE}\|$, where $j = JCUE3$ or OCUE3, was calculated. Plots of these Euclidean norms against $k$ are shown in Figures 3 and 4, where the JCUE3 (OCUE3) estimates corresponding to the minimum value of these norms are reported. Apparent from Figure 3, $\|\hat{\lambda}_{JCUE3} - \hat{\lambda}_{QMLE}\|$ tends to vary less and be of a smaller magnitude than $\|\hat{\lambda}_{OCUE3} - \hat{\lambda}_{QMLE}\|$ with lag length, especially at low levels of $k$. The same observation seems generally true in Figure 4, with three notable exceptions for $\|\hat{\lambda}_{JCUE3} - \hat{\lambda}_{QMLE}\|$ occurring at $k = 25, 26, 34$. Apparent from both figures, $\hat{\lambda}_{JCUE3} \to \hat{\lambda}_{QMLE}$ and $\hat{\lambda}_{OCUE3} \to \hat{\lambda}_{QMLE}$ as $k$ increases. However, in all cases considered, $\min_{k \in K} \|\hat{\lambda}_j - \hat{\lambda}_{QMLE}\|$ occurs in the interior of possible lag lengths considered, suggesting that there exists an "optimal" $k$ for both JCUE3 and OCUE3.

7. Conclusion

The main contribution of this paper is to provide simple GMM estimators for the semi-strong GARCH(1,1) model with a straightforward IV interpretation. In this case, the instrument vector is populated by past residuals and past squared residuals. The resulting moment conditions are stated entirely in terms of covariates observable at time $t$. While these simple estimators rely on skewness for identification, they do not require treatment of the third and fourth conditional moments. These estimators (can) involve many (potentially weak) instruments, the bias from which can be eliminated by using either a CUE
with the optimal weighting matrix (and all the accompanying moment existence criteria it requires) or a jackknife CUE (GMM) with a robust weighting matrix based on, for example, the inverse of Spearman’s correlation matrix for the vector valued functions comprising the moment conditions of the given estimator. Versions of the optimal CUE and jackknife CUE are shown to outperform QMLE in finite samples.

The identification result in this paper can be extended to a GARCH model with a leverage effect. Suppose that 
\[ h_t = \omega_0 + (\alpha_0 + \alpha_0 \times 1 (Y_{t-1} < 0)) Y_{t-1}^2 + \beta_0 h_{t-1}. \]
Then (8) can be divided into the set of moment conditions 
\[ E \left[ \tilde{Y}_t^2 Y_{t-1} \right] = (\alpha_0 + \alpha_0 \times P (Y_t < 0)) E [W_t Y_t], \]
and 
\[ E \left[ \tilde{Y}_t^2 Y_{t-1} \times (1 - 1 (Y_{t-1} < 0)) \right] = \alpha_0 (1 - P (Y_t < 0)) E [W_t Y_t], \]
which can be used to identify a semi-parametric IV estimator of the semi-strong GARCH model with a leverage effect. Such an estimator would be applicable to stock returns given the results of Hansen and Lunde (2005) and would expand the set of empirical applications to which traditional IV estimators of the GARCH(1,1) model can apply.

Applications in empirical asset pricing involve GARCH assumptions within the GMM paradigm and are, therefore, amendable to the estimators that I propose. For instance, Mark (1988) and Bodurtha and Mark (1991) consider versions of the conditional CAPM that parameterize market betas as ARCH(1) processes. The moment conditions from the simple GMM estimators I propose can easily be appended to the moment conditions of these models to allow the market betas to display GARCH properties without the need for specifying the entire conditional distribution of asset returns.

The results of several Monte Carlo and theoretical studies are broadly consistent with those presented in this paper. Hansen, Heaton, and Yaron (1996) find, through simulation experiments, that the CUE has smaller bias than the GMM estimator. Newey and Smith (2004) show that the class of generalized empirical likelihood (GEL) estimators, of which the CUE is a member, has lower asymptotic bias than the GMM estimator when there are several instruments and zero third moments. Newey and Windmeijer (2009) show that the jackknife GMM estimator is less biased than the two-step GMM estimator but that the CUE is more efficient than the jackknife GMM estimator under many (weak) moments. For the semi-strong GARCH(1,1) model, the Monte Carlo results I present show that the CUE has smaller bias than the GMM estimator and is more efficient in the presence of a nonzero third
moment regardless of whether the weighting matrix is optimal, but for both the CUE and GMM estimators using a non-optimal weighting matrix, the associated biases grow with the size of the instrument vector. JCUE and JGMM estimators fix this problem, with JCUE proving more efficient than JGMM and both proving less efficient than the OCUE.

The estimators proposed in this paper are IV estimators with (potentially) many instruments. Methods for selecting the number of instruments for use in these estimators like those proposed by Donald, Imbens, and Newey (2008) are, therefore, of interest, especially given the results from Section 6. Future research may look to relax the symmetry assumption in Donald, Imbens, and Newey (2008) and define criteria that are not (necessarily) dependent upon the variance-covariance matrix of the moment conditions.
Appendix

**PROOF OF LEMMA 1:** From (1), (2), and \(E[W_t | F_{t-1}] = 0\),

\[
E\left[\hat{\gamma}_t^2 Y_{t-1}\right] = E\left[(\hat{h}_t + W_t)Y_{t-1}\right] \tag{23}
\]
\[
= E\left[(\alpha_0\hat{\gamma}_{t-1}^2 + \beta_0\hat{h}_{t-1})Y_{t-1}\right] \nonumber \\
= \alpha_0 E\left[Y_{t-1}^3\right] ,
\]

\[
E\left[\hat{\gamma}_t^2 Y_{t-2}\right] = E\left[h_t Y_{t-2}\right] 
= (\alpha_0 + \beta_0) E\left[\hat{\gamma}_{t-1}^2 Y_{t-2}\right] 
= \alpha_0 (\alpha_0 + \beta_0) E\left[Y_{t-2}^3\right] ,
\]

and

\[
E\left[\hat{\gamma}_t^2 Y_{t-3}\right] = (\alpha_0 + \beta_0) E\left[\hat{\gamma}_{t-1}^2 Y_{t-3}\right] 
= (\alpha_0 + \beta_0)^2 E\left[\hat{\gamma}_{t-2}^2 Y_{t-3}\right] 
= \alpha_0 (\alpha_0 + \beta_0)^2 E\left[Y_{t-3}^3\right].
\]

Given (7) and A2(i), these results imply that

\[
E\left[\hat{\gamma}_t^2 Y_{t-k}\right] = \alpha_0 (\alpha_0 + \beta_0)^{k-1} E[W_t Y_t]. \tag{24}
\]

Solving (24) for \(k = k + 1\) and comparing the result to \(E\left[\hat{\gamma}_t^2 Y_{t-k}\right]\) produces (9).

**PROOF OF LEMMA 2:** From (4) follows that

\[
E\left[\hat{\gamma}_t^4\right] = E\left[(\hat{h}_t + W_t)^2\right] = E\left[\hat{h}_t^2\right] + E\left[W_t^2\right].
\]

Given (3),

\[
E\left[\hat{h}_t^2\right] = (\alpha_0 + \beta_0)^2 E\left[\hat{h}_{t-1}^2\right] + \alpha_0^2 \lambda_0. \tag{25}
\]

28
Recursive substitution into (25) produces

\[ E \left[ \tilde{h}_t^2 \right] = (1 + (\alpha_0 + \beta_0)^2 + \cdots + (\alpha_0 + \beta_0)^{2(\tau - 1)}) \alpha_0^2 \lambda_0 + (\alpha_0 + \beta_0)^{2\tau} E \left[ \tilde{h}_{t-\tau}^2 \right] \]

for \( \tau \geq 1 \). It is well known that \( (\alpha_0 + \beta_0)^{2\tau} \to 0 \) as \( \tau \to \infty \) if and only if \( \alpha_0 + \beta_0 < 1 \). Therefore, \( E \left[ \tilde{h}_t^2 \right] \to \left( \frac{\alpha_0^2}{1 - (\alpha_0 + \beta_0)^2} \right) \lambda_0 \) as \( \tau \to \infty \) if and only if A1 holds. Let \( E \left[ \tilde{h}_t^2 \right] = \eta_0 \). For \( k = 1 \),

\[ E \left[ \tilde{Y}_t^2 \tilde{Y}_{t-1}^2 \right] = E \left[ E \left[ \tilde{Y}_t^2 \tilde{Y}_{t-1}^2 \left| F_{t-1} \right. \right] \right] = E \left[ \left( \alpha_0 \tilde{Y}_{t-1}^2 + \beta_0 \tilde{h}_{t-1} \right) \tilde{Y}_{t-1}^2 \right] = \alpha_0 \lambda_0 + (\alpha_0 + \beta_0) \eta_0 \]

For \( k \geq 2 \),

\[ E \left[ \tilde{h}_t \left| F_{t-k} \right. \right] = \alpha_0 E \left[ \tilde{Y}_{t-1}^2 \left| F_{t-k} \right. \right] + \beta_0 E \left[ \tilde{h}_{t-1} \left| F_{t-k} \right. \right] = (\alpha_0 + \beta_0) E \left[ \tilde{h}_{t-1} \left| F_{t-k} \right. \right] = (\alpha_0 + \beta_0)^2 E \left[ \tilde{h}_{t-2} \left| F_{t-k} \right. \right] \]

\[ \vdots \]

\[ = (\alpha_0 + \beta_0)^{\tau - 1} E \left[ \tilde{h}_{t-(k-1)} \left| F_{t-k} \right. \right] = (\alpha_0 + \beta_0)^{\tau - 1} \left[ \alpha_0 \tilde{Y}_{t-k}^2 + \beta_0 \tilde{h}_{t-k} \right] \]

and, therefore,

\[ E \left[ \tilde{Y}_t^2 \tilde{Y}_{t-k}^2 \right] = E \left[ E \left[ \tilde{Y}_t^2 \tilde{Y}_{t-k}^2 \left| F_{t-k} \right. \right] \right] = E \left[ E \left[ \tilde{h}_t \left| F_{t-k} \right. \right] \tilde{Y}_{t-k}^2 \right] = (\alpha_0 + \beta_0)^{k-1} \left[ \alpha_0 \lambda_0 + (\alpha_0 + \beta_0) \eta_0 \right] . \]

Given (27), \( E \left[ \tilde{Y}_t^2 \tilde{Y}_{t-k}^2 \right] \to 0 \) as \( k \to \infty \). Solving (27) for \( k = k + 1 \) and comparing the result to \( E \left[ \tilde{Y}_t^2 \tilde{Y}_{t-k}^2 \right] \) grants (12).
PROOF OF THE PROPOSITION: From (16),

\[
E \left[ \tilde{Y}_t^2 Z_{t-1} \right] = \begin{bmatrix}
E \left[ \tilde{Y}_t^2 Y_{t-1} \right] \\
E \left[ \tilde{Y}_t^2 X_{t-2} \right] \\
E \left[ \tilde{Y}_t^2 \tilde{Z}_{t-2} \right]
\end{bmatrix},
\]

and

\[
E \left[ Z_{-1} X_{-1} \lambda_0 \right] = \begin{bmatrix}
\alpha_0 E \left[ \tilde{Y}_{t-1}^2 Y_{t-1} \right] + \beta_0 E \left[ \tilde{h}_{t-1} Y_{t-1} \right] \\
\alpha_0 E \left[ \tilde{Y}_{t-1}^2 X_{t-2} \right] + \beta_0 E \left[ \tilde{h}_{t-1} X_{t-2} \right] \\
\alpha_0 E \left[ \tilde{Y}_{t-1}^2 \tilde{Z}_{t-2} \right] + \beta_0 E \left[ \tilde{h}_{t-1} \tilde{Z}_{t-2} \right]
\end{bmatrix}.
\]

\[
E \left[ \tilde{Y}_{t-1}^2 Y_{t-1} \right] = E \left[ Y_{t}^3 \right] \text{ by (7) and A2(i). Since } W_t \text{ is a MDS,}
\]

\[
E \left[ \tilde{Y}_{t-1}^2 X_{t-2} \right] = E \left[ \tilde{h}_{t-1} X_{t-2} \right] = E \left[ \tilde{Y}_t^2 X_{t-1} \right]
\]

by the law of iterated expectations and by Lemma 1. Similarly,

\[
E \left[ \tilde{Y}_{t-1}^2 \tilde{Z}_{t-2} \right] = E \left[ \tilde{h}_{t-1} \tilde{Z}_{t-2} \right] = E \left[ \tilde{Y}_t^2 \tilde{Z}_{t-1} \right]
\]

by the law of iterated expectations and by Lemma 2. Therefore,

\[
E \left[ Z_{-1} X_{-1} \lambda_0 \right] = \begin{bmatrix}
\alpha_0 E \left[ Y_{t}^3 \right] \\
(\alpha_0 + \beta_0) E \left[ \tilde{Y}_t^2 X_{t-1} \right] \\
(\alpha_0 + \beta_0) E \left[ \tilde{Y}_t^2 \tilde{Z}_{t-1} \right]
\end{bmatrix},
\]

and \( E \left[ Z_{-1} \left( \tilde{Y}_t^2 - X_{-1} \lambda_0 \right) \right] = g (\lambda_0, \sigma_0^2) \). \( \blacksquare \)

**Lemma 3.** Given Assumptions A1–A3, the following conditions hold:

**Condition C1:** \( T^{-1} \sum_{t=1}^{T} Y_t \xrightarrow{p} 0 \)

**Condition C2:** \( T^{-1} \sum_{t=1}^{T} Y_t^2 \xrightarrow{p} \sigma_0^2 \)
CONDITION C3: $T^{-1} \sum_{t=1}^{T} W_t \overset{P}{\to} 0$

CONDITION C4: $T^{-1} \sum_{t=1}^{T} W_t Y_t \overset{P}{\to} \gamma_0$

CONDITION C5: $(T - \max{(k, l)})^{-1} \sum_{t=\max{(k, l)}+1}^{T} W_{t-l} Y_{t-k} \overset{P}{\to} 0 \ \forall \ k \neq l$

CONDITION C6: $(T - k)^{-1} \sum_{t=k+1}^{T} W_t W_{t-k} \overset{P}{\to} 0 \ \forall \ k \geq 1$

CONDITION C7: $T^{-1} \sum_{t=1}^{T} W_t^2 \overset{P}{\to} \lambda_0$

CONDITION C8: For a constant $C$ where $0 < C < 1$ and a MDS $\{Z_t\}$ that is uniformly integrable, $T^{-1} \sum_{t=1}^{T} C^t Z_t \overset{P}{\to} 0$.

PROOF OF LEMMA 3: Since $Y_t$ is covariance stationary with a mean of zero, C1 follows by the LLN. Given Lemma 2, $Y_t^2$ is covariance stationary with $E\left[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2\right] \to 0$ as $k \to \infty$ (see 27). C2 then also follows from the LLN, as does C3, given $E[W_t | F_{t-1}] = 0$, $E[W_t W_{t-k}] = 0$, and A3(i). Given A2(i)-(ii), C4 follows from Theorem 1 of Andrews (1988). Since $W_{t-l} Y_{t-k}$ and $W_t W_{t-k}$ are both MDS, Theorem 1 of Andrews (1988) applies to each to establish C5 and C6, respectively, given A2(iii) and A3(ii). A3(i) and A3(iii) allow C7 to follow from Theorem 1 of Andrews (1988). Lastly, since $\{Z_t\}$ is uniformly integrable, $\exists$ a $c > 0$ for every $\epsilon > 0$ such that

$$E[|Z_t| \times I(|Z_t| \geq c)] < \epsilon,$$

where $I(|Z_t| \geq c) = 1$ if $|Z_t| \geq c$ and 0 otherwise. Let $X_t = C^t Z_t$. Then

$$|X_t| = |C^t| |Z_t| < |Z_t|,$$

and

$$|X_t| \times I(|X_t| \geq c) \leq |Z_t| \times I(|Z_t| \geq c).$$
As a consequence,

\[ E [ |X_t| \times I(|X_t| \geq c)] < \epsilon, \]

and \( \{X_t\} \) is uniformly integrable. Theorem 1 of Andrews (1988) then establishes C8.

**PROOF OF THEOREM 1:** By C1 and C2,

\[
\text{plim} \left( T(k)^{-1} \sum_t g_{1,t} (\lambda, \sigma^2) \right) = \text{plim} \left( T(k)^{-1} \sum_t Y_t^2 Y_{t-1} \right) - \text{op} \lim \left( T(k)^{-1} \sum_t Y_t^3 \right).
\]

Given (6),

\[
T(k)^{-1} \sum_t Y_t^2 Y_{t-1} = T(k)^{-1} \sum_t \left( W_t + \alpha_0 \sum_{i=1}^{t-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} + \beta_0 (\alpha_0 + \beta_0)^{t-1} \hat{h}_0 + \sigma_0^2 \right) Y_{t-1}
\]

\[
= \alpha_0 T(k)^{-1} \sum_t \sum_{i=1}^{t-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} Y_{t-1} + (3 \text{ additional terms}),
\]

where the probability limit for each of these three additional terms is zero given C1, C5, and C8, respectively. Since

\[
T(k)^{-1} \sum_t \sum_{i=1}^{t-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} Y_{t-1} = T(k)^{-1} \sum_t W_{t-1} Y_{t-1} + (\alpha_0 + \beta_0) T(k)^{-1} \sum_t W_{t-2} Y_{t-1}
\]

\[
+ (\alpha_0 + \beta_0)^2 T(k)^{-1} \sum_t W_{t-3} Y_{t-1} + \cdots + o_p(1),
\]

for which

\[
\text{plim} \left( T(k)^{-1} \sum_t \sum_{i=1}^{t-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} Y_{t-1} \right) = \gamma_0 \text{ by C4 and C5},
\]

\[
\text{plim} \left( T(k)^{-1} \sum_t Y_t^2 Y_{t-1} \right) = \alpha_0 \gamma_0.
\]

Moreover, since \( T(k)^{-1} \sum_t Y_t^3 = T(k)^{-1} \sum_t Y_t^2 Y_t \), similar expansions to those given above reveal that

\[
\text{plim} \left( T(k)^{-1} \sum_t Y_t^3 \right) = \text{plim} \left( T(k)^{-1} \sum_t W_t Y_t \right) = \gamma_0.
\]
by C4, with the end result being that

\[ \text{p lim} \left( T (k)^{-1} \sum_{t} g_{1,t} (\lambda, \hat{\sigma}^2) \right) = (\alpha_0 - \alpha) \gamma_0 \]

\[= E [g_{1,t} (\lambda, \sigma_0^2)] . \] (28)

Next, define the \( l \)th element of the vector \( g_{2,t} (\lambda, \hat{\sigma}^2) \) for \( l = 1, \ldots, K - 1 \) as

\[ g^{(l)}_{2,t} (\lambda, \hat{\sigma}^2) = (Y_t^2 - \hat{\sigma}^2) (Y_{t-(l+1)} - (\alpha + \beta) Y_{t-l}) . \]

\[ \text{p lim} \left( T (k)^{-1} \sum_{t} g^{(l)}_{2,t} (\lambda, \hat{\sigma}^2) \right) = \text{p lim} \left( T (k)^{-1} \sum_{t} Y_t^2 Y_{t-(l+1)} \right) - (\alpha + \beta) \text{p lim} \left( T (k)^{-1} \sum_{t} Y_t^2 Y_{t-l} \right) \]

by C1 and C2. Given (6),

\[ T (k)^{-1} \sum_{t} Y_t^2 Y_{t-(l+1)} = \alpha_0 T (k)^{-1} \sum_{l} \sum_{i=1}^{l-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} Y_{t-(l+1)} + (3 \text{ additional terms}) \]

\[= \alpha_0 (\alpha_0 + \beta_0)^{l-1} T (k)^{-1} \sum_{t} W_{t-(l+1)} Y_{t-(l+1)} \]

\[+ \alpha_0 T (k)^{-1} \sum_{t} \sum_{i \neq l+1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} Y_{t-(l+1)} + (3 \text{ additional terms}) . \]

The three additional terms each have probability limits equal to zero given C1, C5, and C8. Therefore, \( \text{p lim} \left( T (k)^{-1} \sum_{t} Y_t^2 Y_{t-(l+1)} \right) = \alpha_0 (\alpha_0 + \beta_0)^{l-1} \gamma_0 \), and

\[ \text{p lim} \left( T (k)^{-1} \sum_{t} g^{(l)}_{2,t} (\lambda, \hat{\sigma}^2) \right) = \alpha_0 [(\alpha_0 + \beta_0) - (\alpha + \beta)] (\alpha_0 + \beta_0)^{l-1} \gamma_0 \] (29)

\[= E [g^{(l)}_{2,t} (\lambda, \sigma_0^2)] . \]

Similarly defining the \( l \)th element of the vector \( g_{3,t} (\lambda, \hat{\sigma}^2) \) as

\[ g^{(l)}_{3,t} (\lambda, \hat{\sigma}^2) = (Y_t^2 - \hat{\sigma}^2) (Y_{t-(l+1)} - \hat{\sigma}^2) - (\alpha + \beta) (Y_t^2 - \hat{\sigma}^2) (Y_{t-l} - \hat{\sigma}^2) \]
and considering the \( \text{plim} \left( T(k)^{-1} \sum_{t} g_{3,t}^{(l)}(\lambda, \tilde{\sigma}^2) \right) \), given (6),

\[
T(k)^{-1} \sum_{t} Y_t^2 Y_{t-l} = \left( \sigma_0^2 \right)^2 + \alpha_0 T(k)^{-1} \sum_{t} \left( \sum_{i=1}^{l-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i}W_{t-l} \right)
+ \alpha_0^2 T(k)^{-1} \left( \sum_{i=1}^{l-1} (\alpha_0 + \beta_0)^{i-1} W_{t-i} \right) \left( \sum_{j=1}^{l-1} (\alpha_0 + \beta_0)^{j-1} W_{t-l-j} \right)
+ \text{6 additional terms}
= \left( \sigma_0^2 \right)^2 + \alpha_0 T(k)^{-1} \left[ (\alpha_0 + \beta_0)^{l-1} \sum_{t} W_{t-l}^2 + \sum_{t} \sum_{i \neq l} (\alpha_0 + \beta_0)^{i-1} W_{t-i}W_{t-l} \right]
+ \alpha_0^2 T(k)^{-1} \left[ \sum_{t} \sum_{i \neq j} (\alpha_0 + \beta_0)^{(i+j)-2} W_{t-i}W_{t-l-j} + \sum_{t} \sum_{j=l}^{t-1} (\alpha_0 + \beta_0)^{2j-l} W_{t-j-l}^2 \right]
+ \text{6 additional terms}.
\]

C3, C6, and C8 are used to show that the probability limits of the 6 additional terms are each zero. \( \text{plim} \left( T(k)^{-1} \sum_{t} W_{t-l}^2 \right) = \lambda_0 \), given C7. From C6, it follows that

\[
\text{plim} \left( T(k)^{-1} \sum_{t} \sum_{i \neq l} (\alpha_0 + \beta_0)^{i-1} W_{t-i}W_{t-l} \right) = 0
\]

\[
\text{plim} \left( T(k)^{-1} \sum_{t} \sum_{i \neq j} (\alpha_0 + \beta_0)^{(i+j)-2} W_{t-i}W_{t-l-j} \right) = 0.
\]

The term

\[
T(k)^{-1} \sum_{t} \sum_{j=1}^{l-1} (\alpha_0 + \beta_0)^{2j-l} W_{t-j-1}^2 =
\]

\[
T(k)^{-1} \sum_{t} \left( (\alpha_0 + \beta_0)^{l} W_{t-l-1}^2 + (\alpha_0 + \beta_0)^{l+2} W_{t-l-2}^2 + \cdots + (\alpha_0 + \beta_0)^{2t-(l+2)} W_1^2 \right)
= (\alpha_0 + \beta_0)^{l} T(k)^{-1} \sum_{t} W_{t-l-1}^2 + (\alpha_0 + \beta_0)^{l+2} T(k)^{-1} \sum_{t} W_{t-l-2}^2 + \cdots + \alpha_p (1).
\]
By C7,
\[
\begin{align*}
\lim_{t \to \infty} \left( T(k)^{-1} \sum_{i=1}^{t-1} \sum_{j=1}^{i-1} (\alpha_0 + \beta_0)^{2j-1} W_{t-j-1}^2 \right) &= (\alpha_0 + \beta_0)^{l} \lambda_0 (1 + (\alpha_0 + \beta_0)^2 + (\alpha_0 + \beta_0)^4 + \cdots) \\
&= (\alpha_0 + \beta_0)^{l} \frac{\lambda_0}{1 - (\alpha_0 + \beta_0)^2},
\end{align*}
\]
and
\[
\begin{align*}
\lim_{t \to \infty} \left( T(k)^{-1} \sum_{t} Y_t^2 Y_{t-1}^2 \right) &= (\sigma_0^2)^2 + (\alpha_0 + \beta_0)^{l-1} (\alpha_0 \lambda_0 + (\alpha_0 + \beta_0) \eta_0),
\end{align*}
\]
where \( \eta_0 = E \left[ \tilde{h}_t^2 \right] \) from Lemma 2. Therefore,
\[
\begin{align*}
\lim_{t \to \infty} \left( T(k)^{-1} \sum_{t} g_{3,t} (\lambda, \sigma_0^2) \right) &= ((\alpha_0 + \beta_0) - (\alpha + \beta)) \times (\alpha_0 + \beta_0)^{l-1} (\alpha_0 \lambda_0 + (\alpha_0 + \beta_0) \eta_0) \\
&= E \left[ g_{3,t} (\lambda, \sigma_0^2) \right].
\end{align*}
\]

For \( \max(i) = 2 \), (28) and (29) establish \( \tilde{g}(\lambda, \sigma_0^2) \xrightarrow{p} g(\lambda, \sigma_0^2) \). For \( \max(i) = 3 \), (28)–(30) establish the same result. Under either specification, let \( \tilde{Q}(\lambda, \sigma_0^2) = g(\lambda, \sigma_0^2)' M_0 g(\lambda, \sigma_0^2) \), and \( \tilde{Q}(\lambda, \sigma_0^2) = \tilde{g}(\lambda, \sigma_0^2)' M_0 \tilde{g}(\lambda, \sigma_0^2) \). Then \( \tilde{Q}(\lambda, \sigma_0^2) \xrightarrow{p} Q(\lambda, \sigma_0^2) \) by continuity of multiplication. For \( \max(i) = 2 \), (28) and (29) establish that the only \( \lambda \in \Lambda \) satisfying \( \tilde{g}(\lambda, \sigma_0^2) = 0 \) is \( \lambda = \lambda_0 \), since \( \gamma_0 \neq 0 \) and \( (\alpha_0 + \beta_0) \) is strictly positive. As a consequence, \( Q(\lambda, \sigma_0^2) \) is uniquely minimized at \( \lambda = \lambda_0 \). A parallel result holds for \( \max(i) = 3 \), given the aforementioned conditions plus (30) and the fact that \( \alpha_0 \lambda_0 + (\alpha_0 + \beta_0) \eta_0 \) is strictly positive.

**Lemma 4:** \( \tilde{S}_\lambda (\hat{\lambda}, \hat{\sigma}_0^2) \xrightarrow{p} S_\lambda (\lambda_0, \sigma_0^2) \), and \( \tilde{S}_{\sigma_2} (\lambda_0, \sigma_0^2) \xrightarrow{p} S_{\sigma_2} (\lambda_0, \sigma_0^2) = 0 \) given (i) Assumptions A1 and A2, if \( \max(i) = 2 \) or (ii) Assumptions A1–A3, if \( \max(i) = 3 \).

**Proof of Lemma 4:** Define \( \tilde{s}_{\lambda,ij} (\hat{\lambda}, \hat{\sigma}_0^2) \) as the element in the \( i \)th row and \( j \)th column.
of $\bar{S}_\lambda \left( \hat{\lambda}, \hat{\sigma}^2 \right)$. Let $\hat{Z}_{t-2} = \left[ Y_{t-2} \cdots Y_{t-k} \right]'$ for $k \geq 2$, and $\iota$ be a $(k - 1)$-vector of ones.

For $\max(i) = 3$,

$$\bar{S}_\lambda \left( \hat{\lambda}, \hat{\sigma}^2 \right) = -T(k)^{-1} \begin{pmatrix} \sum_t Y_t^2 & 0 \\ \sum_t (Y_t^2 - \hat{\sigma}^2) X_{t-1} & \sum_t (Y_t^2 - \hat{\sigma}^2) X_{t-1} \\ \sum_t (Y_t^2 - \hat{\sigma}^2) Z_{t-1} & \sum_t (Y_t^2 - \hat{\sigma}^2) Z_{t-1} \end{pmatrix},$$

and

$$\bar{S}_{\sigma^2} \left( \lambda_0, \sigma^2 \right) = -T(k)^{-1} \begin{pmatrix} \sum_t Y_{t-1} \\ \sum_t (X_{t-2} - (\alpha_0 + \beta_0) X_{t-1}) \\ \left( 2\sigma^2 T(k) - \sum_t Y_t^2 \right) \iota (1 - (\alpha_0 + \beta_0)) - \sum_t \left( \hat{Z}_{t-2} - (\alpha_0 + \beta_0) \hat{Z}_{t-1} \right) \end{pmatrix}.$$

The following results follow from the proof to Theorem 1.

**RESULT R1:**

$$p \lim \left( \bar{s}_{\lambda,11} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \right) = -p \lim \left( T(k)^{-1} \sum_t Y_t^2 Y_t \right) = -p \lim \left( T(k)^{-1} \sum_t W_t Y_t \right) = -\gamma_0$$

**RESULT R2:**

$$p \lim \left( \bar{s}_{\lambda,21}^{(l)} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \right) = -p \lim \left( T(k)^{-1} \sum_t Y_t^2 Y_{t-1} \right) = -\alpha_0 (\alpha_0 + \beta_0)^l \gamma_0,$$

where $\bar{s}_{\lambda,21}^{(l)} \left( \hat{\lambda}, \hat{\sigma}^2 \right)$ is the $l$th element of $\bar{s}_{\lambda,21} \left( \hat{\lambda}, \hat{\sigma}^2 \right)$. 

36
RESULT R3:

\[ p \lim \left( \hat{s}_{\lambda,31}^{(l)} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \right) = -p \lim \left( T \left( k \right)^{-1} \sum_{t} Y_{t}^{2} Y_{t-1} \right) + (\sigma_{0}^{2})^{2} \]

\[ = (\alpha_{0} + \beta_{0})^{l-1} (\alpha_{0} \lambda_{0} + (\alpha_{0} + \beta_{0}) \eta_{0}) , \]

where \( \hat{s}_{\lambda,31}^{(l)} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \) is the \( l \)th element of \( \hat{s}_{\lambda,31} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \).

Given R1–R3, \( \hat{s}_{\lambda,ij} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \xrightarrow{p} s_{\lambda,ij} \left( \lambda_{0}, \sigma_{0}^{2} \right) \) \( \forall \ i, j \). Next, \( p \lim \left( \hat{s}_{\sigma_{2},11} \left( \lambda_{0}, \sigma^{2} \right) \right) = 0 \), and \( p \lim \left( \hat{s}_{\sigma_{2},21} \left( \lambda_{0}, \sigma^{2} \right) \right) = 0 \) both by C1. Finally, \( p \lim \left( \hat{s}_{\sigma_{2},31} \left( \lambda_{0}, \sigma^{2} \right) \right) = 0 \) by C2.

PROOF OF THEOREM 2: Let \( M_{t} = M_{T} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \). Then the first order condition from (14) is

\[ \hat{S}_{\lambda} \left( \hat{\lambda}, \hat{\sigma}^2 \right) M_{T} \hat{g} \left( \hat{\lambda}, \hat{\sigma}^2 \right) = 0. \] (31)

Let \( H \left( \hat{\lambda}, \bar{x}, \sigma_{0}^{2} \right) = \hat{S}_{\lambda} \left( \hat{\lambda}, \hat{\sigma}^2 \right) M_{T} \hat{S}_{\lambda} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \), where \( \bar{x} \) is between \( \hat{\lambda} \) and \( \lambda_{0} \). Expanding \( \hat{g} \left( \hat{\lambda}, \hat{\sigma}^2 \right) \) first around \( \lambda_{0} \), then around \( \sigma_{0}^{2} \), and then solving for \( \hat{\lambda} - \lambda_{0} \) produces

\[ \sqrt{T \left( k \right)} \left( \hat{\lambda} - \lambda_{0} \right) = -H \left( \lambda_{0}, \sigma_{0}^{2} \right)^{-1} \hat{S}_{\lambda} \left( \hat{\lambda}, \hat{\sigma}^2 \right) M_{T} \sqrt{T \left( k \right)} \left( \hat{g} \left( \lambda_{0}, \sigma_{0}^{2} \right) + \hat{S}_{\sigma_{2}} \left( \lambda_{0}, \sigma^{2} \right) \hat{m} \left( \sigma_{0}^{2} \right) \right) \]

\[ = -H \left( \lambda_{0}, \sigma_{0}^{2} \right)^{-1} S_{\lambda} \left( \lambda_{0}, \sigma_{0}^{2} \right) M_{T} \sqrt{T \left( k \right)} \hat{g} \left( \lambda_{0}, \sigma_{0}^{2} \right) , \]

where the second equality follows from Lemma 4. The conclusion follows from the Slutsky Theorem.

PROOF OF LEMMA 5: From the definition of \( \hat{p}_{t,s}^{(m,n)} \left( \theta \right) \),

\[ \hat{p}_{t,s}^{(m,n)} \left( \theta \right) - \hat{p}_{t,s}^{(m,n)} \left( \theta_{0} \right) = \frac{-6}{T \left( k, s \right)^{2} - 1} \left\{ T \left( k, s \right)^{-1} \sum_{t} a_{t,s} \left( \hat{\theta} \right) - a_{t,s} \left( \theta_{0} \right) \right\} . \]

By the consistency of \( \hat{\theta} \) established under Theorem 1, \( \exists a \delta_{t} \rightarrow 0 \) such that \( \left\| \hat{\theta} - \theta_{0} \right\| \leq \delta_{t} \). By the triangle inequality,

\[ \left\| T \left( k, s \right)^{-1} \sum_{t} a_{t,s} \left( \hat{\theta} \right) - a_{t,s} \left( \theta_{0} \right) \right\| \leq T \left( k, s \right)^{-1} \sum_{t} \left\| a_{t,s} \left( \hat{\theta} \right) - a_{t,s} \left( \theta_{0} \right) \right\| \leq T \left( k, s \right)^{-1} \sum_{t} \Delta_{t,s} \left( \theta \right) . \]
Finally, by a WLLN, \( T(k,s)^{-1} \sum_t \Delta_{t,s}(\theta) \overset{p}{\to} E[\Delta_{t,s}(\theta)] \), which establishes the result. □

**PROOF OF THE COROLLARY:**

\[
\tilde{Q}(\lambda, \hat{\sigma}^2) = T(k)^{-2} \sum_{s=1, s \neq s}^T T(k)^{-1} \sum_{t \neq s}^T g_t(\lambda, \hat{\sigma}^2)' M_T g_s(\lambda, \hat{\sigma}^2) \\
= T(k)^{-1} \sum_{s=1}^T T(k)^{-1} \sum_{t \neq s}^T g_t(\lambda, \hat{\sigma}^2)' M_T g_s(\lambda, \hat{\sigma}^2) \\
= T(k)^{-1} \sum_{s=1}^T A_s(\lambda, \hat{\sigma}^2) g_s(\lambda, \hat{\sigma}^2),
\]

where

\[
A_s(\lambda, \hat{\sigma}^2) = \left( T(k)^{-1} \sum_{t \neq s}^T g_t(\lambda, \hat{\sigma}^2) \right)' M_T.
\]

From the proof to Theorem 1, \( g(\lambda, \hat{\sigma}^2) \overset{p}{\to} \bar{g}(\lambda, \sigma_0^2) \) if \( \max(i) = 2 \) or \( 3 \), which means that each \( A_s(\lambda, \hat{\sigma}^2) \) has the same probability limit. As a consequence, \( \tilde{Q}(\lambda, \hat{\sigma}^2) \overset{p}{\to} \tilde{Q}(\lambda, \sigma_0^2) \), which has a unique minimum at \( \lambda = \lambda_0 \) given Theorem 1. □
References


[29] Lumsdaine, R.L., 1996, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, Econometrica, 64, 575-596.


[38] Skoglund, J., 2001, A simple efficient GMM estimator of GARCH models, unpublished manuscript.


<table>
<thead>
<tr>
<th>Param. Est.</th>
<th>True Theta</th>
<th>True Theta</th>
<th>True Theta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Med Dec</td>
<td>Med Dec</td>
<td>Med Dec</td>
</tr>
<tr>
<td></td>
<td>Bias Rge SD MDAE</td>
<td>Bias Rge SD MDAE</td>
<td>Bias Rge SD MDAE</td>
</tr>
<tr>
<td>Var QMLE</td>
<td>-0.005 0.242 0.094 0.063</td>
<td>-0.008 0.283 0.111 0.074</td>
<td>-0.022 0.581 0.309 0.156</td>
</tr>
<tr>
<td>MM</td>
<td>-0.018 0.235 0.100 0.060</td>
<td>-0.022 0.289 0.129 0.076</td>
<td>-0.066 0.501 0.272 0.148</td>
</tr>
<tr>
<td>Alpha QMLE</td>
<td>-0.001 0.054 0.021 0.013</td>
<td>0.000 0.039 0.015 0.010</td>
<td>0.000 0.022 0.008 0.005</td>
</tr>
<tr>
<td>JCUE2</td>
<td>-0.016 0.091 0.042 0.028</td>
<td>-0.009 0.067 0.031 0.020</td>
<td>0.000 0.048 0.022 0.011</td>
</tr>
<tr>
<td>JCUE3</td>
<td>-0.001 0.029 0.027 0.006</td>
<td>0.000 0.014 0.011 0.002</td>
<td>0.000 0.004 0.005 0.001</td>
</tr>
<tr>
<td>JGMM2</td>
<td>-0.017 0.109 0.046 0.032</td>
<td>-0.011 0.082 0.035 0.025</td>
<td>0.001 0.067 0.029 0.016</td>
</tr>
<tr>
<td>JGMM3</td>
<td>-0.015 0.090 0.043 0.027</td>
<td>-0.006 0.070 0.034 0.016</td>
<td>-0.001 0.039 0.019 0.005</td>
</tr>
<tr>
<td>CUE2</td>
<td>-0.011 0.109 0.050 0.027</td>
<td>-0.005 0.084 0.043 0.019</td>
<td>-0.004 0.081 0.033 0.018</td>
</tr>
<tr>
<td>CUE3</td>
<td>-0.006 0.040 0.036 0.009</td>
<td>-0.002 0.024 0.026 0.003</td>
<td>-0.001 0.005 0.007 0.001</td>
</tr>
<tr>
<td>GMM2</td>
<td>-0.013 0.112 0.051 0.031</td>
<td>-0.009 0.094 0.041 0.025</td>
<td>-0.007 0.083 0.032 0.021</td>
</tr>
<tr>
<td>GMM3</td>
<td>-0.016 0.113 0.053 0.031</td>
<td>-0.012 0.093 0.042 0.026</td>
<td>-0.010 0.071 0.027 0.019</td>
</tr>
<tr>
<td>Beta QMLE</td>
<td>0.000 0.081 0.033 0.020</td>
<td>0.000 0.056 0.022 0.013</td>
<td>-0.001 0.023 0.009 0.006</td>
</tr>
<tr>
<td>JCUE2</td>
<td>0.010 0.173 0.076 0.043</td>
<td>0.009 0.137 0.061 0.036</td>
<td>-0.008 0.144 0.154 0.031</td>
</tr>
<tr>
<td>JCUE3</td>
<td>0.000 0.104 0.058 0.022</td>
<td>0.000 0.063 0.036 0.015</td>
<td>0.000 0.035 0.022 0.009</td>
</tr>
<tr>
<td>JGMM2</td>
<td>0.011 0.198 0.093 0.053</td>
<td>0.010 0.167 0.077 0.047</td>
<td>-0.030 0.386 0.235 0.043</td>
</tr>
<tr>
<td>JGMM3</td>
<td>0.011 0.158 0.077 0.040</td>
<td>0.006 0.114 0.059 0.029</td>
<td>0.002 0.068 0.035 0.015</td>
</tr>
<tr>
<td>CUE2</td>
<td>-0.040 0.227 0.110 0.051</td>
<td>-0.051 0.211 0.147 0.053</td>
<td>-0.115 0.833 0.325 0.115</td>
</tr>
<tr>
<td>CUE3</td>
<td>-0.024 0.152 0.095 0.031</td>
<td>-0.020 0.130 0.090 0.022</td>
<td>-0.014 0.054 0.086 0.015</td>
</tr>
<tr>
<td>GMM2</td>
<td>-0.053 0.242 0.106 0.061</td>
<td>-0.075 0.272 0.120 0.075</td>
<td>-0.214 0.618 0.247 0.214</td>
</tr>
<tr>
<td>GMM3</td>
<td>-0.031 0.217 0.099 0.044</td>
<td>-0.026 0.144 0.081 0.035</td>
<td>-0.025 0.108 0.059 0.031</td>
</tr>
</tbody>
</table>

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector \( \theta = (\text{Var, Alpha, Beta}) \), where \( \text{Var} \) is the unconditional variance. QMLE is the quasi-maximum likelihood estimator. MM is the method of moments estimator. (J)CUE2(3) is the (jackknife) continuous updating estimator with max(i) = 2(3). (J)GMM2(3) is the (jackknife) two-step generalized method of moments estimator with max(i) = 2(3). For all (J)CUE and (J)GMM estimators: (a) the weighting matrix is the inverse of Spearman’s correlation matrix; (b) \( k = 20 \); (c) \( L = 1 \). Med. Bias is the median bias, SD the standard deviation, and MDAE the median absolute error of the estimates. Dec Rge is the decile range of the estimates, measured as the difference between the 90th and the 10th percentiles.
FIGURE 1

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector is \((1, 0.10, 0.85)\), where \(\alpha = 0.10\) and \(\beta = 0.85\). QMLE is the quasi-maximum likelihood estimator. JCUE is the jackknife continuous updating estimator with: (a) \(\max(i) = 3\); (b) the weighting matrix as the inverse of Spearman’s correlation matrix; (c) \(k\) = the number of lags; (d) \(L = 1\). Dec Rge is the decile range of the estimates, measured as the difference between the 90th and the 10th percentiles. MDAE is the median absolute error of the estimates.
Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector is $(1, 0.05, 0.94)$, where $\alpha = 0.05$ and $\beta = 0.94$. QMLE is the quasi-maximum likelihood estimator. JCUE is the jackknife continuous updating estimator. OCUE is the optimal continuous updating estimator. For both the JCUE and OCUE: (a) $\max(i) = 3$; (b) $k$ is the number of lags; (d) $L = 1$. For the JCUE, the weighting matrix is the inverse of Spearman’s correlation matrix. For the OCUE, the weighting matrix is the inverse of the variance-covariance matrix. Dec Rge is the decile range of the estimates, measured as the difference between the 90th and the 10th percentiles. MDAE is the median absolute error of the estimates.
### TABLE 2

<table>
<thead>
<tr>
<th>Para.</th>
<th>Est.</th>
<th>True Theta</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1.0, 0.15, 0.75)</td>
<td>(1.0, 0.10, 0.85)</td>
<td>(1.0, 0.05, 0.94)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Med</td>
<td>Dec</td>
<td>Bias</td>
<td>Rge</td>
</tr>
<tr>
<td>Var</td>
<td>QMLE</td>
<td></td>
<td></td>
<td>-0.006</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.338</td>
<td>1.375</td>
</tr>
<tr>
<td>MM</td>
<td></td>
<td>-0.032</td>
<td>0.314</td>
<td>0.132</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.338</td>
<td>1.329</td>
</tr>
<tr>
<td>Alpha</td>
<td>QMLE</td>
<td></td>
<td></td>
<td>-0.002</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.225</td>
<td>1.242</td>
</tr>
<tr>
<td>JCUE3</td>
<td></td>
<td>-0.003</td>
<td>0.040</td>
<td>0.022</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.344</td>
<td>0.836</td>
</tr>
<tr>
<td>JGMM3</td>
<td></td>
<td>-0.026</td>
<td>0.089</td>
<td>0.041</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>0.987</td>
<td>0.950</td>
</tr>
<tr>
<td>CUE3</td>
<td></td>
<td>-0.011</td>
<td>0.052</td>
<td>0.030</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.288</td>
<td>0.847</td>
</tr>
<tr>
<td>Beta</td>
<td>QMLE</td>
<td></td>
<td></td>
<td>-0.001</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.183</td>
<td>1.182</td>
</tr>
<tr>
<td>JCUE3</td>
<td></td>
<td>0.001</td>
<td>0.121</td>
<td>0.061</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.164</td>
<td>1.049</td>
</tr>
<tr>
<td>JGMM3</td>
<td></td>
<td>0.018</td>
<td>0.195</td>
<td>0.089</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.231</td>
<td>1.161</td>
</tr>
<tr>
<td>CUE3</td>
<td></td>
<td>-0.037</td>
<td>0.187</td>
<td>0.104</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>-Ratio</td>
<td></td>
<td></td>
<td>1.231</td>
<td>1.101</td>
</tr>
</tbody>
</table>

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector \( \theta = (\text{Var}, \text{Alpha}, \text{Beta}) \), where \( \text{Var} \) is the unconditional variance. QMLE is the quasi-maximum likelihood estimator. MM is the method of moments estimator. (J)CUE3 is the (jackknife) continuous updating estimator with \( \text{max}(i) = 3 \). JGMM3 is the jackknife two-step generalized method of moments estimator, also with \( \text{max}(i) = 3 \). For the (J)CUE and JGMM estimators: (a) the weighting matrix is the inverse of Spearman’s correlation matrix; (b) \( k = 20 \); (c) \( L = 1 \). Ratio is the given measure of dispersion (error) for the estimator immediately above it in this table divided by the corresponding measure of dispersion (error) from Table 1. Med. Bias is the median bias, SD the standard deviation, and MDAE the median absolute error of the estimates. Dec Rge is the decile range of the estimates, measured as the difference between the 90th and the 10th percentiles.
### TABLE 3

<table>
<thead>
<tr>
<th>Currency</th>
<th>Para.</th>
<th>JCU3</th>
<th>OCU3</th>
<th>QMLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var</td>
<td>0.5579</td>
<td>0.5579</td>
<td>0.4957</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.050</td>
<td>0.0890</td>
<td>0.0532</td>
<td></td>
</tr>
<tr>
<td>AUD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta</td>
<td>0.922</td>
<td>0.9081</td>
<td>0.9382</td>
<td></td>
</tr>
<tr>
<td>(0.0211)</td>
<td></td>
<td>(0.0101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>0.9726</td>
<td>0.9971</td>
<td>0.9914</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>40</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Var</td>
<td>0.4963</td>
<td>0.4963</td>
<td>0.5057</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.049</td>
<td>0.0901</td>
<td>0.0486</td>
<td></td>
</tr>
<tr>
<td>JPY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta</td>
<td>0.916</td>
<td>0.8864</td>
<td>0.9361</td>
<td></td>
</tr>
<tr>
<td>(0.0147)</td>
<td></td>
<td>(0.0123)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>0.9650</td>
<td>0.9764</td>
<td>0.9848</td>
<td></td>
</tr>
</tbody>
</table>

Notes: GARCH(1,1) models are fit to Australian Dollar (AUD) and Japanese Yen (JPY) spot returns, where the spot rates are measured in terms of US Dollars. The time period for each series is daily from 1/1/90 - 12/31/09. JCU3 and OCU3 are the jackknife CUE and optimal CUE, where the former uses the inverse of Spearman’s correlation matrix as it’s weighting matrix, while the latter uses the inverse of the variance-covariance matrix. Both JCUE3 and OCUE3 set max(i) = 3 and L = 1. K is the number of lags used in the given estimator (if applicable). Var is the unconditional variance estimate for the given spot return. Alpha is the ARCH estimate, while Beta is the GARCH estimate. Sum is the sum of the Alpha and Beta estimates.
Notes: GARCH(1,1) models are fit to the Australian Dollar (AUD) spot return series using the jackknife CUE (JCUE) and optimal CUE (OCUE) with lag lengths from $k = 20, \ldots, 40$. The AUD spot return series is measured daily from 1/1/90 - 12/31/09. The Euclidean norm of the difference between the JCUE (OCUE) and QMLE estimates for Alpha and Beta are plotted against the lag lengths. The JCUE (OCUE) estimates closest to the QMLE estimates are shown. The weighting matrix for the JCUE is the inverse of Spearman’s correlation matrix, while the weighting matrix for OCUE is the inverse of the variance-covariance matrix. For both the JCUE and OCUE, max(i) = 3 and L = 1. For OCUE3, $k = 20$, 38, and 39 are excluded because they produce point estimates that violate covariance stationarity.
Notes: GARCH(1,1) models are fit to the Japanese Yen (JPY) spot return series using the jackknife CUE (JCUE) and optimal CUE (OCUE) with lag lengths from $k = 20, \ldots, 40$. The JPY spot return series is measured daily from 1/1/90 - 12/31/09. The Euclidean norm of the difference between the JCUE (OCUE) and QMLE estimates for Alpha and Beta are plotted against the lag lengths. The JCUE (OCUE) estimates closest to the QMLE estimates are shown. The weighting matrix for the JCUE is the inverse of Spearman’s correlation matrix, while the weighting matrix for OCUE is the inverse of the variance-covariance matrix. For both the JCUE and OCUE, $\max(i) = 3$ and $L = 1$. For JCUE3, $k = 23$ is excluded because it produces point estimates that likely violate fourth moment stationarity.