Simple GMM Estimation of the Semi-Strong GARCH(1,1) Model

Todd, Prono

Commodity Futures Trading Commission

11 November 2009

Simple GMM Estimation of the Semi-Strong GARCH(1,1) Model[1]

Todd Prono[2]
Commodity Futures Trading Commission

First Version: November 2009
This Version: January 2011

Abstract

IV estimators for the semi-strong ARCH(1) model that rely on past squared residuals alone as instruments do not extend to the GARCH case. Efficient IV estimators of the semi-strong GARCH(1,1) model require the derivative of the conditional variance as well as both the third and fourth conditional moments to be included within the instrument vector. This paper proposes IV estimators for the semi-strong GARCH(1,1) model that only rely on past residuals and past squared residuals as instruments. These estimators are based on the autocovariances of squared residuals, as in the ARCH(1) case described above, as well as on the covariances between past residuals and current squared residuals. These latter covariances are nonzero if the residuals are skewed. Jackknife GMM estimators and jackknife continuous updating estimators (CUE) eliminate the bias caused by many (weak) instruments. The jackknife CUE is new and applies to cases where the optimal weighting matrix is unavailable out of a concern over the existence of higher moments. In these cases, a robust analog to the variance-covariance matrix determines the weighting matrix. A Monte Carlo study shows that a CUE based on the optimal weighting matrix as well as the jackknife CUE outperforms QMLE in finite samples. An empirical application involving Australian Dollar and Japanese Yen spot returns is also included.

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

[1] 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, 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. 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 are what 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.

These simple GMM estimators 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. The estimators are shown to be CAN under less restrictive moment existence criteria than in Weiss (1986) and Rich, Raymond, and Butler (1991). 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), which deletes contemporaneous observations from the double sum that forms the GMM objective function. 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 though, in this case, from the CUE objective function. 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.[3] 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.

The remainder of this paper is organized as follows. Section 2 outlines the model's assumptions and states two lemmas that define a set of moment conditions for identifying the GARCH(1,1) model. From these moment conditions, Section 3 establishes simple GMM estimators, develops their properties, and proposes a data dependent weighting matrix for the moment conditions that does not require higher moment existence criteria for consistency. Section 4 discusses bias-free estimation given many (potentially weak) instruments and gives a consistency result for the JGMM and JCUE. Section 5 discusses Monte Carlo results for the proposed estimators. Section 6 details an empirical application involving Australian Dollar and Japanese Yen spot returns, and Section 7 concludes.

2. The Model and Implications

For the sequence $\{Y_t\}_{t\in\mathbb{Z}}$, let $\mathcal{F}_t$ be the associated σ-algebra where $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \cdots \subseteq \mathcal{F}$.

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

The first two conditional moments of Y_t are parameterized as

$$E\left[Y_t \mid F_{t-1}\right] = 0, \qquad E\left[Y_t^2 \mid F_{t-1}\right] = h_t, \tag{1}$$

where

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

In what follows, ω_0 denotes the true value, ω any one of a set of possible values, and $\widehat{\omega}$ an estimate. Parallel definitions hold for all other parameter values. The model of (1) 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 $\frac{Y_t}{h_t^{1/2}}$ is iid and drawn from a known distribution nests as a special case. Consider the following additional assumptions for the model of (1) and (2).

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

The restrictions on θ ensure that h_t is everywhere strictly positive. From Lumsdaine (1996), α is strictly positive because if $\alpha = 0$, then h_t is completely deterministic, in which case ω_0 and β_0 are not separately identified. Since $\beta \geq 0$, A1 nests the ARCH(1) model. Implicit in A1 is the condition that $\alpha_0 + \beta_0 < 1$, in which case Y_t is covariance stationary with $E\left[Y_t^2\right] = \sigma_0^2$ following from Theorem 1 of Bollerslev (1986).[4]

The mean-adjusted form of (2) is

$$\widetilde{h}_t = \alpha_0 \widetilde{Y}_{t-1}^2 + \beta_0 \widetilde{h}_{t-1}, \tag{3}$$

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

$$\widetilde{Y}_t^2 = \widetilde{h}_t + W_t, \tag{4}$$

[4]Covariance stationarity implies additional restrictions on θ, namely that $\{(\alpha,\ \beta):\ \alpha + \beta < 1\}$.

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

$$\widetilde{h}_t = \sum_{i=0}^{t-1} \alpha_0 \beta_0^i \widetilde{Y}_{t-1-i}^2 + \beta_0^t \widetilde{h}_0, \tag{5}$$

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

$$\widetilde{Y}_t^2 = W_t + \alpha_0 \sum_{i=1}^{t-1} \left(\alpha_0 + \beta_0\right)^{i-1} W_{t-i} + \beta_0 \left(\alpha_0 + \beta_0\right)^{t-1} \widetilde{h}_0, \tag{6}$$

which shows that the GARCH(1,1) model relates $\widetilde{Y}_t^2$ to a weighted sum of current and past innovations. A similar recursion is found for the ARCH(p) model in Guo and Phillips (2001). Moment properties for W_t are central to defining simple GMM estimators for (3) and are the subject of the following two assumptions.

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

Since Y_t is a MDS, (4) and an application of the law of iterated expectations assuming stationarity grants that

$$\begin{aligned} E\left[Y_t^3\right] &= E\left[\widetilde{Y}_t^2 Y_t\right] \\ &= E\left[\left(\widetilde{h}_t + W_t\right) Y_t\right] \\ &= E\left[W_t Y_t\right]. \end{aligned} \tag{7}$$

Given A2(i), Y_t is asymmetric 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 of the random variable 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 $W_tY_t - \gamma_0$ 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).

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

Suppose

$$Y_t = h_t^{1/2}\epsilon_t, \tag{8}$$

where ϵ_t is iid with a mean of zero and a unit variance. Then A3(i) is equivalent to assuming that

$$(\kappa+1)\,\alpha_0^2 + 2\alpha_0\beta_0 + \beta_0^2 < 1, \quad \kappa = E\left[\epsilon_t^4\right] - 1, \tag{9}$$

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] A3(i), of course, implies covariance stationarity for Y_t. Moreover, it imposes additional restrictions on the parameter set $(\alpha,\ \beta)$, comparable to $\left\{(\alpha,\ \beta):\ (\kappa+1)\,\alpha^2 + 2\alpha\beta + \beta^2 < 1\right\}$ but of an unknown form, owing to potential dependence in ϵ_t^4. 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 A2ii) also governs the fourth. This assumption is supported empirically by the results of Hansen (1994).

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

$$E\left[\widetilde{Y}_t^2 Y_{t-1}\right] = \alpha_0 E\left[W_tY_t\right], \tag{10}$$

[5] If ϵ_t is normally distributed, then this inequality follows from Theorem 2 of Bollerslev (1986) with $\kappa = 2$.

and

$$E\left[\widetilde{Y}_t^2 Y_{t-(k+1)}\right] = (\alpha_0 + \beta_0)\, E\left[\widetilde{Y}_t^2 Y_{t-k}\right]. \tag{11}$$

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.[6] 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 (10) that is only a function of the data and of α_0. Separation of α_0 from β_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.

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,*

$$E\left[\widetilde{Y}_t^2 \widetilde{Y}_{t-(k+1)}^2\right] = (\alpha_0 + \beta_0) E\left[\widetilde{Y}_t^2 \widetilde{Y}_{t-(k)}^2\right]. \tag{12}$$

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. Such an approach requires these squared residuals to be covariance stationary. Lemma 2 provides necessary and sufficient conditions for this result and is closely related to Theorem 3 of Hafner (2003).

(12), like (11), provide moment conditions in terms of the parameters α_0 and β_0. Lemma 2 shows that while sufficient for identifying the ARCH(1) model (and, in general, the ARCH(p) model), the autocovariances of squared residuals alone are not sufficient for identifying the

[6] See (24) in the Appendix.

GARCH(1,1) model, since these moment conditions only involve the parameters α_0 and β_0 jointly, not separately. Neither are these autocovariances necessary for identification in the case of either the ARCH(1) or GARCH(1,1) model, given skewed residuals and Lemma 1. (12) does, however, provide an expanded set of moment conditions for a GMM estimator that should improve efficiency in cases where the fourth moment is stationary.

3. Estimation

3.1. Notation

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

$$g_{1,t}\left(Y;\ \lambda, \sigma^2\right) = \left(Y_t^2 - \sigma^2\right) Y_{t-1} - \alpha Y_t^3, \tag{13}$$

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

$$g_{3,t}\left(Y;\ \lambda, \sigma^2\right) = \left(Y_t^2 - \sigma^2\right)\left(Z_{t-2} - (\alpha + \beta) Z_{t-1}\right),$$

and the following definitions

$$
\begin{aligned}
g_{i,t}\left(Y;\ \lambda,\sigma^2\right) &= g_{i,t}\left(\lambda,\sigma^2\right),\ i=1,2,3,\\
g_t\left(\lambda,\sigma^2\right) &= \left[g_{i,t}\left(\lambda,\sigma^2\right)\right],\ i=1,\ldots,\max(i),\ 2\leq\max(i)\leq 3,\\
g_{m,t}\left(\lambda,\sigma^2\right) &= m\text{th element of } g_t\left(\lambda,\sigma^2\right),\\
\widehat{g}\left(\lambda,\sigma^2\right) &= T(k)^{-1}\sum_{t=k+1}^{T} g_t\left(\lambda,\sigma^2\right), \qquad \overline{g}\left(\lambda,\sigma^2\right)=E\left[g_t\left(\lambda,\sigma^2\right)\right],\\
m_t\left(\sigma^2\right) &= Y_t^2-\sigma^2, \qquad \widehat{m}\left(\sigma^2\right)=T^{-1}\sum_{t=1}^{T}Y_t^2-\sigma^2,\\
\widetilde{g}_t\left(\lambda,\sigma^2\right) &= g_t\left(\lambda,\sigma^2\right)+S_{\sigma^2}\left(\lambda,\sigma^2\right)m_t\left(\sigma^2\right),\\
\widehat{S}_\lambda\left(\lambda,\sigma^2\right) &= \frac{\partial\widehat{g}\left(\lambda,\sigma^2\right)}{\partial\lambda}, \qquad S_\lambda\left(\lambda,\sigma^2\right)=E\left[\frac{\partial g_t\left(\lambda,\sigma^2\right)}{\partial\lambda}\right],\\
\widehat{S}_{\sigma^2}\left(\lambda,\sigma^2\right) &= \frac{\partial\widehat{g}\left(\lambda,\sigma^2\right)}{\partial\sigma^2}, \qquad S_{\sigma^2}\left(\lambda,\sigma^2\right)=E\left[\frac{\partial g_t\left(\lambda,\sigma^2\right)}{\partial\sigma^2}\right],\\
\Omega\left(\lambda,\sigma^2\right) &= \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{t-s}\left(\lambda,\sigma^2\right)g_t\left(\lambda,\sigma^2\right)'\right],\ L\geq 1,\\
\widehat{\Omega}\left(\lambda,\sigma^2\right) &= \sum_{s=-(L-1)}^{s=(L-1)} T(k)^{-1}\sum_{t=k+s+1}^{T} g_{t-s}\left(\lambda,\sigma^2\right)g_t\left(\lambda,\sigma^2\right)',\\
R\left[g_{m,t}\left(\lambda,\sigma^2\right)\right] &= \text{rank of } g_{m,t}\left(\lambda,\sigma^2\right)\ \text{ in } g_{m,k+1}\left(\lambda,\sigma^2\right),\ldots,g_{m,T}\left(\lambda,\sigma^2\right),\\
\widehat{\rho}_{t,s}^{(m,n)}\left(\lambda,\sigma^2\right) &= 1-\frac{6}{T(k,s)\left(T(k,s)^2-1\right)}\sum_{t=k+s+1}^{T}\left(R\left[g_{m,t}\left(\lambda,\sigma^2\right)\right]-R\left[g_{n,t-s}\left(\lambda,\sigma^2\right)\right]\right)^2,\\
\widehat{\Sigma}\left(\lambda,\sigma^2\right) &= \sum_{s=-(L-1)}^{s=(L-1)}\left[\widehat{\rho}_{t,s}^{(m,n)}\left(\lambda,\sigma^2\right)\right],
\end{aligned}
$$

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

$$\widehat{\lambda}=\arg\min_{\lambda\in\Lambda}\ \widehat{g}\left(\lambda,\widehat{\sigma}^2\right)'M_T\widehat{g}\left(\lambda,\widehat{\sigma}^2\right), \tag{14}$$

for some sequence of positive semi-definite M_T, which is the familiar GMM estimator of Hansen (1982) with $\widehat{\sigma}^2$ 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). If $M_T=M_T\left(\widetilde{\lambda},\ \widehat{\sigma}^2\right)$, where $\widetilde{\lambda}$ is a preliminary (and consistent) estimator of λ_0, then (14) is a two-step GMM estimator. If $M_T = M_T\left(\lambda,\ \widehat{\sigma}^2\right)$, 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

$$\widetilde{Y}_t^2 = X_{-1}'\lambda_0 + W_t, \tag{15}$$

where $X_{-1} = \begin{pmatrix} \widetilde{Y}_{t-1}^2, & \widetilde{h}_{t-1} \end{pmatrix}'$. Next, let $Z_{-1} \in \mathcal{F}_{t-1}$. Since W_t is a MDS,

$$E\left[Z_{-1}\left(\widetilde{Y}_t^2 - X_{-1}'\lambda_0\right)\right] = 0, \tag{16}$$

which are the population moment conditions for an infeasible IV estimator of $\widetilde{h}_t$; where, in this case, and throughout the ensuing discussions of potential IV estimators, infeasible references the fact that $\widetilde{h}_{t-1}$ is not observed at time t.

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

$$E\left[Z_{-1}\left(\widetilde{Y}_t^2 - X_{-1}'\lambda_0\right)\right] = \overline{g}\left(\lambda_0, \sigma_0^2\right).$$

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

$$Cov\left[Y_t^2;\ Y_{t-k}\right] = Cov\left[h_t;\ Y_{t-k}\right], \quad Cov\left[Y_t^2;\ Y_{t-k}^2\right] = Cov\left[h_t;\ Y_{t-k}^2\right],$$

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 λ_0 because (16) is not linear in λ_0, owing to the dependence of h_{t-1} on λ_0.

The Proposition uncovers an instrument vector that permits feasible estimation of (16).

Notice that this instrument vector omits $\widetilde{Y}_{t-1}^2$. If $\widetilde{Y}_{t-1}^2$ is included as an instrument, then feasible estimation of (16) is no longer possible. To see this, append $\widetilde{Y}_{t-1}^2$ to the end of Z_{-1} as $\dot{Z}_{-1} = \begin{pmatrix} Z_{-1} \\ \widetilde{Y}_{t-1}^2 \end{pmatrix}$, and then substitute $\dot{Z}_{-1}$ for Z_{-1} in (16). The final row of $E\left[\dot{Z}_{-1}X'_{-1}\lambda_0\right]$ is

$$\alpha_0 E\left[\widetilde{Y}_{t-1}^4\right] + \beta_0 E\left[\widetilde{h}_{t-1}\widetilde{Y}_{t-1}^2\right]. \tag{17}$$

Expanding the left term in (17) using (4) produces

$$\begin{aligned} E\left[\widetilde{Y}_{t-1}^4\right] &= E\left[\left(\widetilde{h}_{t-1} + W_{t-1}\right)\widetilde{Y}_{t-1}^2\right] \\ &= E\left[\widetilde{h}_{t-1}\widetilde{Y}_{t-1}^2\right] + E\left[W_{t-1}\widetilde{Y}_{t-1}^2\right] \\ &\neq E\left[\widetilde{h}_{t-1}\widetilde{Y}_{t-1}^2\right], \end{aligned}$$

in general, since $E\left[W_{t-1}\widetilde{Y}_{t-1}^2\right] \neq 0$. As a consequence, (17) can only be simplified to

$$(\alpha_0 + \beta_0) E\left[\widetilde{Y}_t^4\right] - \beta_0 E\left[W_t\widetilde{Y}_t^2\right],$$

given stationarity, which preserves the explicit dependence of (16) on the conditional variance through the contemporaneous covariance between W_t and $\widetilde{Y}_t^2$.

The move from Z_{-1} to $\dot{Z}_{-1}$ 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

$$\widehat{\vartheta} = \underset{\vartheta\in\Theta}{\arg\min}\ \widehat{f}(\vartheta)'\Lambda_T\widehat{f}(\vartheta),$$

where $\vartheta = (\omega,\ \alpha,\ \beta)'$,

$$\begin{aligned}
\widehat{f}(\vartheta) &= T^{-1}\sum_{t=1}^{T} f_t(\vartheta) = T^{-1}\sum_{t=1}^{T}\left[f_{i,t}(\vartheta)\right] \text{ for } i = 1,2,3, \\
f_{i,t}(\vartheta) &= \frac{1}{\Delta_t}\left(\frac{\partial h_t}{\partial \vartheta_i}\right) h_t^{1/2}\left[\left(\frac{Y_t}{h_t^{1/2}}\right) E\left[Y_t^3 \mid \digamma_{t-1}\right] - h_t^{3/2}\left(\left(\frac{Y_t^2}{h_t}\right) - 1\right)\right], \\
\Delta_t &= h_t^3\left(\frac{E\left[Y_t^4 \mid \digamma_{t-1}\right]}{h_t^2} - 1\right) - E\left[Y_t^3 \mid \digamma_{t-1}\right]^2, \\
\Lambda_T &= \left(T^{-1}\sum_t f_t(\vartheta)\, f_t(\vartheta)'\right)^{-1}.
\end{aligned}$$

The estimator $\widehat{\vartheta}$ 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, $\widehat{\lambda}$, 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, $\widehat{\lambda}$ depends on the third and fourth moments of Y_t only unconditionally, meaning that $\widehat{\lambda}$ 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 $\widehat{\lambda}$ a simple estimator for the GARCH(1,1) model within the class of IV estimators discussed above.

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

Given A1, 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 stationarity around a nonzero mean is necessary for this result. When $\max(i) = 3$, fourth moment stationarity also becomes necessary, owing to the consideration of autocovariances between squared residuals. Since estimators for the ARCH(1) model in Theorem 4.4 of Weiss (1986), in Rich et al. (1991), as well as in Theorems 2.1 and 3.1 of Guo and Phillips (2001) also involve the autocovariances of squared residuals, fourth moment stationarity is so, too, required. Through skewness, therefore, Theorem 1 shows that it is possible to (i) extend feasible IV estimation from the ARCH(1) to the GARCH(1,1) case and (ii) do so using a milder set of moment existence criteria than is required for the ARCH(1) case given a symmetric distribution.

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

$$\begin{aligned}
\widehat{\alpha} &= \left\{ \left(\sum_t \widehat{U}_t \right)' M_T \left(\sum_t \widehat{U}_t \right) \right\}^{-1} \left(\sum_t \widehat{U}_t \right)' M_T \left(\sum_t \widehat{V}_t \right), \\
\widehat{U}_t &= \begin{pmatrix} Y_t^3 \\ \left(Y_t^2 - \widehat{\sigma}^2\right) X_{t-1} \\ \left(Y_t^2 - \widehat{\sigma}^2\right) \widehat{Z}_{t-1} \end{pmatrix}, \qquad \widehat{V}_t = \begin{pmatrix} \left(Y_t^2 - \widehat{\sigma}^2\right) Y_{t-1} \\ \left(Y_t^2 - \widehat{\sigma}^2\right) X_{t-2} \\ \left(Y_t^2 - \widehat{\sigma}^2\right) \widehat{Z}_{t-2} \end{pmatrix},
\end{aligned} \tag{18}$$

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

$$\ddot{\alpha} = \left\{ \left(\sum_t \widehat{Z}_{-1} \left(Y_{t-1}^2 - \widehat{\sigma}^2\right) \right)' N_T \left(\sum_t \widehat{Z}_{-1} \left(Y_{t-1}^2 - \widehat{\sigma}^2\right) \right) \right\}^{-1} \left(\sum_t \widehat{Z}_{-1} \left(Y_{t-1}^2 - \widehat{\sigma}^2\right) \right)' N_T \left(\sum_t \widehat{Z}_{-1} \left(Y_t^2 - \widehat{\sigma}^2\right) \right)$$

if $N_T \xrightarrow{p} M_0$, where $\ddot{\alpha}$ is a generalized IV estimator based on the population moment conditions $E\left[Z_{-1}\left(\widetilde{Y}_t^2 - \alpha_0 \widetilde{Y}_{t-1}^2\right)\right] = 0$. In the special case of an ARCH(1) process, $\dot{Z}_{-1}$ 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

$$g_{4,t}\left(\alpha, \widehat{\sigma}^2\right) = \left(Y_t^2 - \widehat{\sigma}^2\right)\left(\left(Y_{t-1}^2 - \widehat{\sigma}^2\right) - \alpha\left(Y_t^2 - \widehat{\sigma}^2\right)\right) \tag{19}$$

to $g_t\left(\alpha, \widehat{\sigma}^2\right)$.

The principal advantage to (18) is its computational simplicity. Simulation results in Section 5 suggest that (14) with $M_T = M_T\left(\lambda,\ \widehat{\sigma}^2\right)$ is the better estimator in terms of bias and efficiency even in the special case of an ARCH(1) process. Moreover, evaluation of (14) as a CUE when $\beta_0 = 0$ is relative straightforward, since the parameter vector is one dimensional. In this case, (14) can be evaluated using a grid search, which requires neither a consistent starting value for α nor the computation of numerical derivatives of the CUE objective function.

THEOREM 2 (Asymptotic Normality). *Consider the estimator in (14) for the model of (1) and (2), letting* $\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} Y_t^2$. *Assume (i)* $M_T \xrightarrow{p} M_0$, *a positive semi-definite matrix and that* $M_0\overline{g}\left(\lambda, \sigma_0^2\right) = 0$ *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)* $S_\lambda\left(\lambda_0, \sigma_0^2\right)' M_0 \times S_\lambda\left(\lambda_0, \sigma_0^2\right)$ *is nonsingular; (iv)* $\sqrt{T(k)}\widehat{g}\left(\lambda_0, \sigma_0^2\right) \xrightarrow{d} N\left(\ 0,\ \ \Omega\left(\lambda_0, \sigma_0^2\right)\ \right)$. *Then*

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

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

As a VTE, (14) is a two-step estimator, since the objective function is minimized conditional on a preliminary, or first-step, estimator $\widehat{\sigma}^2$. 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

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

which is the variance-covariance matrix of

$$\sqrt{T(k)}\widehat{g}\left(\lambda_0, \widehat{\sigma}^2\right) = \sqrt{T(k)}\left\{\widehat{g}\left(\lambda_0, \sigma_0^2\right) + S_{\sigma^2}\left(\lambda_0, \sigma_0^2\right) \widehat{m}\left(\sigma_0^2\right)\right\}, \tag{20}$$

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 $\widehat{\sigma}^2$ on the asymptotic variance of $\widehat{\lambda}$. Given Lemma 4 stated in the Appendix, however, $S_{\sigma^2}(\lambda_0, \sigma_0^2) = 0$, which means that $\widehat{g}(\lambda_0, \widehat{\sigma}^2) = \widehat{g}(\lambda_0, \sigma_0^2)$, $\widetilde{\Omega}(\lambda_0, \sigma_0^2) = \Omega(\lambda_0, \sigma_0^2)$, and, as a consequence, nothing is lost (asymptotically) by plugging $\widehat{\sigma}^2$ into (14) as opposed to σ_0^2. 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. This result does, however, stand in contrast to the VTE studied by Francq, Horath, and Zakoian (2009), where the variance of $\widehat{\sigma}^2$ does, in fact, impact the variance of $\widehat{\lambda}$ asymptotically.

Condition (iv) under Theorem 2 is, of course, a high level assumption. If $g_t(\lambda_0, \sigma_0^2)$ were a MDS (the assumption made in Sections 5 and 6), then (iv) would follow if $E\left[\|g_t(\lambda_0, \sigma_0^2)\|^2\right] < \infty$. Other CLTs for dependent data may also prove applicable, depending on the process for $g_t(\lambda_0, \sigma_0^2)$. This process (be it an MDS or other) depends, in turn, on the processes governing W_tY_t and W_t^2. The fact that these processes are specified only generally motivates condition (iv).

Theorem 4.4 of Weiss (1986) demonstrates the CAN property of an autocovariance-based estimator for the ARCH model under the condition of a finite eighth moment for the residuals. Theorem 2 requires this same condition if $\max(i) = 3$ (i.e., if the autocovariances of squared residuals are considered). If, on the other hand, $\max(i) = 2$, this condition is replaced by the relatively milder requirement of a finite sixth moment. When skewness is present, therefore, feasible IV estimators for the semi-strong ARCH(1) and GARCH(1,1) models are CAN under relatively milder moment existence criteria than in the case of comparable IV estimators for the ARCH(1) model given a symmetric distribution.

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). Additionally, the proof to Theorem 2 is based on the two-step GMM estimator. For the CUE, although the first order condition analogous to (30) 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).

Use of the optimal weighting matrix under Theorem 2 requires at least sixth moment stationarity. Such an assumption may prove overly restrictive, especially for certain financial data. A key question, therefore, is what weighting matrix to choose in the context of Theorem 1, so that $\widehat{\lambda}$ is consistent under, at most, fourth moment stationarity. One option, of course, is to use a non data dependent weighting matrix like the identity matrix. Skoglund (2001), however, reports that the identity 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 here) in Monte Carlo studies of (14). Alternatively, one can consider using a robust analog to $\widehat{\Omega}\left(\widehat{\theta}\right)$ when constructing the weighting matrix. One such alternative is $\widehat{\Sigma}\left(\widehat{\theta}\right)$. The matrix $\left[\widehat{\rho}_{t,s}^{(m,n)}\left(\widehat{\theta}\right)\right]$ is Spearman's (1904) correlation matrix for the vector valued functions $g_t\left(\widehat{\theta}\right)$ and $g_{t-s}\left(\widehat{\theta}\right)$. The matrix $\widehat{\Sigma}\left(\widehat{\theta}\right)$, therefore, reflects rank dependent measures of contemporaneous and lagged association between the sequences of vector valued functions that comprise the moment conditions. The following lemma is useful for establishing consistency of $\widehat{\Sigma}\left(\widehat{\theta}\right)$.

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

Consistency of $\widehat{\rho}_{t,s}^{(m,n)}\left(\widehat{\theta}\right)$ $\forall\, m,n$ follows from Lemma 5 and selected results in Schmid and Schmidt (2007).[7] 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 $\widehat{\Sigma}\left(\widehat{\theta}\right)$ can be thought of as robust.

For simple GMM estimators based only on Theorem 1, standard errors can be computed via the parametric bootstrap. Suppose that the data generating process for Y_t is characterized by (1), (2), and (8), where $E\left[\epsilon_t \mid \mathcal{F}_{t-1}\right] = 0$, $E\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right] = 1$, and the higher moments of ϵ_t follow L^{th} order Markov processes with a finite $L << T$. Use (14) to obtain $\widehat{h}_t$. Let

[7] These results are Theorem 5 and the fact that $\lim_{n\to\infty}\sqrt{n}\left\{\widehat{\rho}_{1,n} - \widehat{\rho}_{S,n}\right\} = 0$, where $\widehat{\rho}_{S,n}$ relates to $\widehat{\rho}_{t,s}^{(m,n)}(\theta_0)$.

$\widehat{\epsilon}_t = Y_t / \sqrt{\widehat{h}_t}$, and apply the nonoverlapping block bootstrap method of Carlstein (1986) to these standardized residuals to obtain the bootstrap sample $\widehat{\epsilon}_t^*$. Use these bootstrap residuals to construct the series $\widehat{Y}_t^* = \sqrt{\widehat{h}_t^*} \widehat{\epsilon}_t^*$, where $\widehat{h}_t^*$ depends on the parameter estimates from the original data sample. Estimate the model of (1) and (2) on $\widehat{Y}_t^*$, 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 $\widehat{\theta}$ that are robust to higher moment dynamics in ϵ_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.

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) 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 $E\left[g_t(\theta) \mid \mathcal{F}_{t-L}\right]$ is constant.[8] Let $s^* = \{S: \ s \geq t + L \text{ or } s \leq t - L; \ s = 1, \ldots, T\}$. Then, the expectation of the GMM objec-

[8] $g_t(\theta)$ can be thought of as a vector of residuals. The requirement is satisified if these residuals follow an MA process of order $L - 1$.

tive function $\widehat{g}(\theta)' M_T \widehat{g}(\theta)$ for a nonrandom weighting matrix M_T is

$$
\begin{aligned}
E\left[\widehat{g}(\theta)' M_T \widehat{g}(\theta)\right] &= T(k)^{-2} E\left[\sum_{t\neq s} g_t(\theta)' M_T g_s(\theta) + \sum_t g_t(\theta)' M_T g_t(\theta)\right] \\
&= T(k)^{-2} E\left[\sum_{t\in s^*} g_t(\theta)' M_T g_{s^*}(\theta) + \sum_{s=-(L-1)}^{s=(L-1)} \sum_t g_t(\theta)' M_T g_{t-s}(\theta)\right] \\
&= \left(1-\frac{L}{T(k)}\right)\overline{g}(\theta)' M_T \overline{g}(\theta) + T(k)^{-1} \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_t(\theta)' M_T g_{t-s}(\theta)\right] \\
&= \left(1-\frac{L}{T(k)}\right)\overline{g}(\theta)' M_T \overline{g}(\theta) + T(k)^{-1} tr\left(M_T \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{t-s}(\theta) g_t(\theta)'\right]\right),
\end{aligned}
$$

which is an adaptation of (2) in Newey and Windmeijer (2009) to dependent time series data.[9]

In the language of Newey and Windmeijer (2009), $\left(1-\frac{L}{T(k)}\right)\overline{g}(\theta)' M_T \overline{g}(\theta)$ is a "signal" term minimized at θ_0. The second term is a "noise" term that is, generally, not minimized at θ_0 if $\frac{\partial g_t(\theta)}{\partial \theta}$ is correlated with $g_t(\theta)$ and is increasing in k.[10] If k is increasing with T, this bias term need not even vanish asymptotically (see Han and Phillips 2006).[11]

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

$$
T(k)^{-1} tr\left(M_T \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{t-s}(\theta) g_t(\theta)'\right]\right) = \frac{m(k)}{T(k)}, \quad m(k) = 2k-1,
$$

[9]This expansion is also valid under a random M_T because estimation of M_T does not effect the limiting distribution.

[10]This "noise" or bias term is analogous to the higher order bias term B_G in Newey and Smith (2004).

[11]Under Theorem 1, however, k is treated as fixed so that (14) is consistent.

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

$$\begin{aligned}
\widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right)' M_T \widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right) &= T\left(k\right)^{-2}\{\sum_{t\neq s} g_t\left(\lambda,\widehat{\sigma}^{2}\right) M_T g_s\left(\lambda,\widehat{\sigma}^{2}\right) + \sum_{t} g_t\left(\lambda,\widehat{\sigma}^{2}\right) M_T g_t\left(\lambda,\widehat{\sigma}^{2}\right)\} \\
&= T\left(k\right)^{-2}\sum_{t\in s^*} g_t\left(\lambda,\widehat{\sigma}^{2}\right)' M_T g_{s^*}\left(\lambda,\widehat{\sigma}^{2}\right) \\
&\quad + T\left(k\right)^{-2}\sum_{s=-(L-1)}^{s=(L-1)}\sum_{t} g_t\left(\lambda,\widehat{\sigma}^{2}\right) M_T g_{t-s}\left(\lambda,\widehat{\sigma}^{2}\right) \\
&= T\left(k\right)^{-2}\sum_{t\in s^*} g_t\left(\lambda,\widehat{\sigma}^{2}\right)' M_T g_{s^*}\left(\lambda,\widehat{\sigma}^{2}\right) \\
&\quad + T\left(k\right)^{-1} tr\left(M_T\left\{\sum_{s=-(L-1)}^{s=(L-1)} T\left(k\right)^{-1}\sum_{t} g_{t-s}\left(\lambda,\widehat{\sigma}^{2}\right) g_t\left(\lambda,\widehat{\sigma}^{2}\right)'\right\}\right)
\end{aligned}$$

If $M_T = \widehat{\Omega}\left(\lambda,\widehat{\sigma}^{2}\right)^{-1}$, the feasible version of $\Omega\left(\lambda,\widehat{\sigma}^{2}\right)^{-1}$, then

$$\widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right)' M_T \widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right) = T\left(k\right)^{-2}\sum_{t\in s^*} g_t\left(\lambda,\widehat{\sigma}^{2}\right)' M_T g_{s^*}\left(\lambda,\widehat{\sigma}^{2}\right) + \frac{m\left(k\right)}{T\left(k\right)},$$

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) $M_T = \widehat{\Sigma}\left(\lambda,\widehat{\sigma}^{2}\right)^{-1}$, in which case $\widehat{\lambda}$ is a robust CUE, (ii) $M_T = \widehat{\Omega}\left(\widetilde{\lambda},\widehat{\sigma}^{2}\right)^{-1}$, in which case $\widehat{\lambda}$ is the optimal two-step GMM estimator, or (iii) $M_T = \widehat{\Sigma}\left(\widetilde{\lambda},\widehat{\sigma}^{2}\right)^{-1}$, in which case $\widehat{\lambda}$ is a robust two-step GMM estimator, (14) will be biased. The expansion of $\widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right)' M_T \widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right)$ offers a way to correct for this bias. Namely, consider the alternative estimator

$$\breve{\lambda} = \arg\min_{\lambda\in\Lambda}\ \breve{Q}\left(\lambda,\widehat{\sigma}^{2}\right), \tag{21}$$

where

$$\begin{aligned}
\breve{Q}\left(\lambda,\widehat{\sigma}^{2}\right) &= T\left(k\right)^{-2}\sum_{t\in s^*} g_t\left(\lambda,\widehat{\sigma}^{2}\right)' M_T g_{s^*}\left(\lambda,\widehat{\sigma}^{2}\right) \qquad (22) \\
&= \widehat{Q}\left(\lambda,\widehat{\sigma}^{2}\right) - T\left(k\right)^{-1} tr\left(M_T\left\{\sum_{s=-(L-1)}^{s=(L-1)} T\left(k\right)^{-1}\sum_{t} g_{t-s}\left(\lambda,\widehat{\sigma}^{2}\right) g_t\left(\lambda,\widehat{\sigma}^{2}\right)'\right\}\right),
\end{aligned}$$

and $\widehat{Q}\left(\lambda,\widehat{\sigma}^{2}\right) = \widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right)' M_T \widehat{g}\left(\lambda,\widehat{\sigma}^{2}\right)$. 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. $\breve{\lambda}$ is consistent given the following corollary.

COROLLARY (Consistency). *Consider the estimator in (21) for the model of (1) and (2). Let* $\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T} Y_t^2$, *and assume that (i)* $M_T \xrightarrow{p} M_0$, *a positive semi-definite matrix, (ii)* $M_0 \overline{g}(\lambda, \sigma_0^2) = 0$ *only if* $\lambda = \lambda_0$, *(iii)* $L = 1$. *If* $\max(i) = 2$, *then* $\breve{\lambda} \xrightarrow{p} \lambda_0$ *given Assumptions A1–A2. If* $\max(i) = 3$, *then* $\breve{\lambda} \xrightarrow{p} \lambda_0$ *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.[12] As a result, if either $M_T = \widehat{\Sigma}\left(\lambda, \widehat{\sigma}^2\right)^{-1}$ or $M_T = \widehat{\Sigma}\left(\widetilde{\lambda}, \widehat{\sigma}^2\right)^{-1}$, $\breve{\lambda}$ is robust in the dual sense that it (i) requires the same moment existence criteria as Theorem 1, and (ii) is free of many (weak) moments bias. Following from Newey and Windmeijer (2009) p. 702, $\breve{\lambda}$ is asymptotically normal if $L = 1$.

If $\beta_0 = 0$ and either M_T is nonrandom or $M_T = M_T\left(\widetilde{\alpha}, \widehat{\sigma}^2\right)$, then the solution to (21) is

$$\breve{\alpha} = \left\{\sum_{t \in s^*} \widehat{U}_t' M_T \widehat{U}_{s^*}\right\}^{-1} \sum_{t \in s^*} \widehat{U}_t' M_T \widehat{V}_{s^*},$$

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

5. Monte Carlo

Consider the data generating process in (1), (2), and (8), where ϵ_t is the negative of a standardized Gamma(2,1) random variable. The skewness and kurtosis of ϵ_t is $-2/\sqrt{2}$ and

[12] This result assumes, of course, that M_T is not constructed from $\Omega\left(\lambda, \widehat{\sigma}^2\right)$.

6, respectively. Values for θ_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 ϵ_t support a finite fourth moment for Y_t according to (9). 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 λ 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 $\widehat{\sigma}^2 = T^{-1} \sum_t Y_t^2$, (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 $L = 1$.[13]

For estimating α_0 and β_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 $\widehat{\alpha}$ and $\widehat{\beta}$. The same can be said for JGMM2 relative to GMM2 for $\widehat{\beta}$, with mixed results (in terms of bias reduction) evidenced for $\widehat{\alpha}$. JCUE3 records less bias than CUE3 for both $\widehat{\alpha}$ and $\widehat{\beta}$. JCUE2 records less bias than CUE2 for $\widehat{\beta}$ but mixed results (in terms of bias reduction) for $\widehat{\alpha}$. 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 $\widehat{\alpha}$ under the jackknifed CUE. Though not reported here, the bias of non jackknifed estimators for $\widehat{\beta}$ tends to increase with k. The level of this bias is most noticeable for high values of β_0.

[13] In some of the simulations, an alternative rank dependent correlation matrix based on Kendall's (1938) 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.

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 $\widehat{\beta}$. JGMM2 is more efficient than GMM2 for $\widehat{\alpha}$, with mixed results appearing for $\widehat{\beta}$. 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 α_0 but a less efficient estimator of β_0 when compared to QMLE. Figure 1 compares the efficiency of JCUE3 relative to QMLE (for both $\widehat{\alpha}$ and $\widehat{\beta}$) for various lag lengths out to $k = 40$. As is evidenced, $\widehat{\alpha}$ remains more efficient under JCUE3 as opposed to QMLE for all lag lengths considered. Moreover, the efficiency of $\widehat{\beta}$ 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).

Of the parameter values considered, $\theta_0 = (1.0,\ 0.05,\ 0.94)'$ is the most likely to support a finite eighth moment.[14] Figure 2, therefore, compares the efficiency of JCUE3, OCUE3, and QMLE for lags lengths out to $k = 40$. Similar to Figure 1, $\widehat{\alpha}$ remains more efficiently estimated under JCUE3 than under QMLE for all lag lengths considered. Interestingly, at low levels of k, $\widehat{\alpha}$ is less efficiently estimated under OCUE3 than under either JCUE3 or QMLE. As k increases, however, the performance of $\widehat{\alpha}$ under OCUE3 converges to that of JCUE3, therefore passing that of QMLE. In terms of $\widehat{\beta}$, 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 $\widehat{\beta}$ under JCUE3 approaches that under QMLE, while the performance of $\widehat{\beta}$ under OCUE3 betters that of QMLE. Therefore, both

[14] 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 α_0 is a principal constraint on the existence of higher moments.

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, benchmarked against the QMLE) in the case where ϵ_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 α_0 and closest to QMLE, in terms of both bias and efficiency, in estimating β_0. CUE3 no longer dominates JGMM3 in terms of dispersion as it does in Table 1. To the contrary, $\widehat{\alpha}$ and $\widehat{\beta}$ tend to be less dispersed under JGMM3 (very noticeably so for $\widehat{\beta}$ when $\beta_0 = 0.85$ and $\beta_0 = 0.94$). JGMM3, however, displays significantly higher bias in $\widehat{\alpha}$ under both $\alpha_0 = 0.15$ and $\alpha_0 = 0.10$ when ϵ_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 $\widehat{\alpha}$, and most consistently for JGMM3. Specifically for JGMM3, the Ratio statistic for both the Decile Range and SD of $\widehat{\alpha}$ is less than one for all the cases considered. This result, perhaps, is not so surprising given that skewness is what identifies α_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 $Y_{i,t} = \log\left(S_{i,t}/S_{i,t-1}\right)$. This section fits the GARCH(1,1) model of (1) and (2) to $\left\{Y_{i,t}\right\}_{t=1}^{T}$.[15] 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 $\max(i) = 3$ 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 $\widehat{\alpha}$. The $\widehat{\beta}$ standard errors are more comparable. The higher standard errors under OCUE3 may relate to the fact that $\widehat{\alpha} + \widehat{\beta}$ 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, $\left\|\widehat{\lambda}_j - \widehat{\lambda}_{QMLE}\right\|$, where $j =$ JCUE3 or

[15] 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.

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, $\left\|\widehat{\lambda}_{JCUE3}-\widehat{\lambda}_{QMLE}\right\|$ tends to vary less and be of a smaller magnitude than $\left\|\widehat{\lambda}_{OCUE3}-\widehat{\lambda}_{QMLE}\right\|$ with lag length, especially at low levels of k. The same observation seems generally true in Figure 4, with three notable exceptions for $\left\|\widehat{\lambda}_{JCUE3}-\widehat{\lambda}_{QMLE}\right\|$ occurring at $k=25,26,34$. Apparent from both figures, $\widehat{\lambda}_{JCUE3}\rightarrow\widehat{\lambda}_{QMLE}$ and $\widehat{\lambda}_{OCUE3}\rightarrow\widehat{\lambda}_{QMLE}$ as k increases. However, in all cases considered, $\min_{k\in K}\left\|\widehat{\lambda}_{j}-\widehat{\lambda}_{QMLE}\right\|$ 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 observed at time $t-1$. 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.

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: Recall that both Y_t and W_t are MDS. Then, applications of the law of iterated expectations, the result from (7), and A2(i) grant that

$$\begin{aligned} E\left[\widetilde{Y}_t^2 Y_{t-1}\right] &= E\left[\left(\widetilde{h}_t + W_t\right) Y_{t-1}\right] \\ &= E\left[\left(\alpha_0 \widetilde{Y}_{t-1}^2 + \beta_0 \widetilde{h}_{t-1}\right) Y_{t-1}\right] \\ &= \alpha_0 E\left[W_t Y_t\right] \end{aligned} \tag{23}$$

and that

$$\begin{aligned} E\left[\widetilde{Y}_t^2 Y_{t-2}\right] &= E\left[\widetilde{h}_t Y_{t-2}\right] \\ &= \alpha_0 E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right] + \beta_0 E\left[\widetilde{h}_{t-1} Y_{t-2}\right] \\ &= \left(\alpha_0 + \beta_0\right) E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right]. \end{aligned}$$

Since application of the same expansion in (23) to $E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right]$ reveals that

$$E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right] = \alpha_0 E\left[W_t Y_t\right],$$

it follows that

$$E\left[\widetilde{Y}_t^2 Y_{t-2}\right] = \alpha_0 \left(\alpha_0 + \beta_0\right) E\left[W_t Y_t\right].$$

Repeated applications of recursive substitution into $E\left[\widetilde{Y}_t^2 Y_{t-k}\right]$ demonstrates, in general, that

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

Solving (24) for $k = k + 1$ and comparing the result to $E\left[\widetilde{Y}_t^2 Y_{t-k}\right]$ produces (11).■

PROOF OF LEMMA 2: From (4) follows that

$$E\left[\widetilde{Y}_t^4\right] = E\left[\left(\widetilde{h}_t + W_t\right)^2\right] = E\left[\widetilde{h}_t^2\right] + E\left[W_t^2\right].$$

Given (3),

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

Recursive substitution into (25) produces

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

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, $E\left[\widetilde{h}_t^2\right] \to \left(\frac{\alpha_0^2}{1-(\alpha_0+\beta_0)^2}\right)\lambda_0$ as $\tau \to \infty$ if and only if A1 holds. Let $E\left[\widetilde{h}_t^2\right] = \eta_0$. For $k=1$,

$$\begin{aligned} E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-1}^2\right] &= E\left[E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-1}^2 \mid \mathcal{F}_{t-1}\right]\right] \\ &= E\left[\left(\alpha_0\widetilde{Y}_{t-1}^2 + \beta_0\widetilde{h}_{t-1}\right)\widetilde{Y}_{t-1}^2\right] \\ &= \alpha_0\lambda_0 + (\alpha_0+\beta_0)\eta_0 \end{aligned}$$

For $k \geq 2$,

$$\begin{aligned} E\left[\widetilde{h}_t \mid \mathcal{F}_{t-k}\right] &= \alpha_0 E\left[\widetilde{Y}_{t-1}^2 \mid \mathcal{F}_{t-k}\right] + \beta_0 E\left[\widetilde{h}_{t-1} \mid \mathcal{F}_{t-k}\right] \\ &= (\alpha_0+\beta_0)E\left[\widetilde{h}_{t-1} \mid \mathcal{F}_{t-k}\right] \\ &= (\alpha_0+\beta_0)^2 E\left[\widetilde{h}_{t-2} \mid \mathcal{F}_{t-k}\right] \\ &\vdots \\ &= (\alpha_0+\beta_0)^{\tau-1} E\left[h_{t-(k-1)} \mid \mathcal{F}_{t-k}\right] \\ &= (\alpha_0+\beta_0)^{\tau-1}\left[\alpha_0 Y_{t-k}^2 + \beta_0 h_{t-k}\right] \end{aligned}$$

and, therefore,

$$\begin{aligned} E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right] &= E\left[E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2 \mid \mathcal{F}_{t-k}\right]\right] \\ &= E\left[E\left[\widetilde{h}_t \mid \mathcal{F}_{t-k}\right]\widetilde{Y}_{t-k}^2\right] \\ &= (\alpha_0+\beta_0)^{k-1}\left[\alpha_0\lambda_0 + (\alpha_0+\beta_0)\eta_0\right]. \end{aligned} \tag{26}$$

Given (26), $E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right] \to 0$ as $k \to \infty$. Solving (26) for $k = k+1$ and comparing the result to $E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right]$ grants (12).■

PROOF OF THE PROPOSITION: From (16),

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

and

$$E\left[Z_{-1}X_{-1}^{'}\lambda_0\right] = \begin{bmatrix} \alpha_0 E\left[\widetilde{Y}_{t-1}^2 Y_{t-1}\right] + \beta_0 E\left[\widetilde{h}_{t-1} Y_{t-1}\right] \\ \alpha_0 E\left[\widetilde{Y}_{t-1}^2 X_{t-2}\right] + \beta_0 E\left[\widetilde{h}_{t-1} X_{t-2}\right] \\ \alpha_0 E\left[\widetilde{Y}_{t-1}^2 \widetilde{Z}_{t-2}\right] + \beta_0 E\left[\widetilde{h}_{t-1} \widetilde{Z}_{t-2}\right] \end{bmatrix}.$$

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

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

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

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

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

$$E\left[Z_{-1}X_{-1}^{'}\lambda_0\right] = \begin{bmatrix} \alpha_0 E\left[Y_t^3\right] \\ \left(\alpha_0 + \beta_0\right) E\left[\widetilde{Y}_t^2 X_{t-1}\right] \\ \left(\alpha_0 + \beta_0\right) E\left[\widetilde{Y}_t^2 \widetilde{Z}_{t-1}\right] \end{bmatrix},$$

and $E\left[Z_{-1}\left(\widetilde{Y}_t^2 - X_{-1}^{'}\lambda_0\right)\right] = \overline{g}\left(\lambda_0, \sigma_0^2\right)$.■

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: $T^{-1}\sum_{t=1}^{T} W_t \xrightarrow{p} 0$

CONDITION C4: $T^{-1}\sum_{t=1}^{T} W_t Y_t \xrightarrow{p} \gamma_0$

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

CONDITION C6: $(T-k)^{-1} \sum_{t=k+1}^{T} W_t W_{t-k} \xrightarrow{p} 0 \; \forall \; k \geq 1$

CONDITION C7: $T^{-1}\sum_{t=1}^{T} W_t^2 \xrightarrow{p} \lambda_0$

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

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 $E\left[\widetilde{Y}_t^2 \widetilde{Y}_{t-k}^2\right] \to 0$ as $k \to \infty$ (see 26). C2 then also follows from the LLN, as does C3, given $E\left[W_t \mid \mathcal{F}_{t-1}\right] = 0$, $E\left[W_t W_{t-k}\right] = 0$, 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

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

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

$$|X_t| = \left|C^t\right| |Z_t| < |Z_t|,$$

and

$$|X_t| \times I\left(|X_t| \geq c\right) \leq |Z_t| \times I\left(|Z_t| \geq c\right).$$

As a consequence,

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

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

PROOF OF THEOREM 1: By C1 and C2,

$$\text{p}\lim\left(T\left(k\right)^{-1}\sum_t g_{1,t}\left(\lambda,\widehat{\sigma}^2\right)\right) = \text{p}\lim\left(T\left(k\right)^{-1}\sum_t Y_t^2 Y_{t-1}\right) - \alpha \text{p}\lim\left(T\left(k\right)^{-1}\sum_t Y_t^3\right).$$

Given (6),

$$\begin{aligned} T\left(k\right)^{-1}\sum_t Y_t^2 Y_{t-1} &= T\left(k\right)^{-1}\sum_t\left(W_t + \alpha_0\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i} + \beta_0\left(\alpha_0+\beta_0\right)^{t-1}\widetilde{h}_0 + \sigma_0^2\right)Y_{t-1} \\ &= \alpha_0 T\left(k\right)^{-1}\sum_t\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}Y_{t-1} + \left(3 \text{ additional terms}\right), \end{aligned}$$

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

$$\begin{aligned} T\left(k\right)^{-1}\sum_t\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}Y_{t-1} &= T\left(k\right)^{-1}\sum_t W_{t-1}Y_{t-1} + \left(\alpha_0+\beta_0\right)T\left(k\right)^{-1}\sum_t W_{t-2}Y_{t-1} \\ &\quad + \left(\alpha_0+\beta_0\right)^2 T\left(k\right)^{-1}\sum_t W_{t-3}Y_{t-1} + \cdots + o_p\left(1\right), \end{aligned}$$

for which $\text{p}\lim\left(T\left(k\right)^{-1}\sum_t\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}Y_{t-1}\right) = \gamma_0$ by C4 and C5,

$$\text{p}\lim\left(T\left(k\right)^{-1}\sum_t Y_t^2 Y_{t-1}\right) = \alpha_0\gamma_0.$$

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

$$\text{p}\lim\left(T\left(k\right)^{-1}\sum_t Y_t^3\right) = \text{p}\lim\left(T\left(k\right)^{-1}\sum_t W_t Y_t\right) = \gamma_0$$

by C4, with the end result being that

$$\begin{aligned}
\text{p}\lim\left(T(k)^{-1}\sum_t g_{1,t}\left(\lambda,\widehat{\sigma}^2\right)\right) &= \left(\alpha_0-\alpha\right)\gamma_0 \\
&= E\left[g_{1,t}\left(\lambda,\sigma_0^2\right)\right].
\end{aligned} \tag{27}$$

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

$$g_{2,t}^{(l)}\left(\lambda,\widehat{\sigma}^2\right)=\left(Y_t^2-\widehat{\sigma}^2\right)\left(Y_{t-(l+1)}-(\alpha+\beta)\,Y_{t-l}\right).$$

$$\text{p}\lim\left(T(k)^{-1}\sum_t g_{2,t}^{(l)}\left(\lambda,\widehat{\sigma}^2\right)\right)=\text{p}\lim\left(T(k)^{-1}\sum_t Y_t^2Y_{t-(l+1)}\right)-(\alpha+\beta)\,\text{p}\lim\left(T(k)^{-1}\sum_t Y_t^2Y_{t-l}\right)$$

by C1 and C2. Given (6),

$$\begin{aligned}
T(k)^{-1}\sum_t Y_t^2Y_{t-(l+1)} &= \alpha_0T(k)^{-1}\sum_t\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}Y_{t-(l+1)}+(3\text{ additional terms}) \\
&= \alpha_0\left(\alpha_0+\beta_0\right)^l T(k)^{-1}\sum_t W_{t-(l+1)}Y_{t-(l+1)} \\
&\quad +\alpha_0T(k)^{-1}\sum_t\sum_{i\neq l+1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}Y_{t-(l+1)}+(3\text{ additional terms}).
\end{aligned}$$

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

$$\begin{aligned}
\text{p}\lim\left(T(k)^{-1}\sum_t g_{2,t}^{(l)}\left(\lambda,\widehat{\sigma}^2\right)\right) &= \alpha_0\left[\left(\alpha_0+\beta_0\right)-(\alpha+\beta)\right]\left(\alpha_0+\beta_0\right)^{l-1}\gamma_0 \\
&= E\left[g_{2,t}^{(l)}\left(\lambda,\sigma_0^2\right)\right].
\end{aligned} \tag{28}$$

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

$$g_{3,t}^{(l)}\left(\lambda,\widehat{\sigma}^2\right)=\left(Y_t^2-\widehat{\sigma}^2\right)\left(Y_{t-(l+1)}-\widehat{\sigma}^2\right)-(\alpha+\beta)\left(Y_t^2-\widehat{\sigma}^2\right)\left(Y_{t-l}-\widehat{\sigma}^2\right)$$

and considering the $\operatorname{p\,lim}\left(T(k)^{-1}\sum_t g_{3,t}^{(l)}\left(\lambda,\widehat{\sigma}^2\right)\right)$, given (6),

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

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

$$\begin{aligned}
\operatorname{p\,lim}\left(T(k)^{-1}\sum_t\sum_{i\neq l}(\alpha_0+\beta_0)^{i-1}W_{t-i}W_{t-l}\right) &= 0 \\
\operatorname{p\,lim}\left(T(k)^{-1}\sum_t\sum_{i\neq j}(\alpha_0+\beta_0)^{(i+j)-2}W_{t-i}W_{t-l-j}\right) &= 0.
\end{aligned}$$

The term

$$\begin{aligned}
& T(k)^{-1}\sum_t\sum_{j=l}^{t-1}(\alpha_0+\beta_0)^{2j-l}W_{t-j-1}^2 = \\
& T(k)^{-1}\sum_t\left((\alpha_0+\beta_0)^l W_{t-l-1}^2 + (\alpha_0+\beta_0)^{l+2}W_{t-l-2}^2+\cdots+(\alpha_0+\beta_0)^{2t-(l+2))}W_1^2\right) \\
= \;& (\alpha_0+\beta_0)^l T(k)^{-1}\sum_t W_{t-l-1}^2 + (\alpha_0+\beta_0)^{l+2}T(k)^{-1}\sum_t W_{t-l-2}^2+\cdots+o_p(1).
\end{aligned}$$

By C7,

$$\begin{aligned}
& \operatorname{p\,lim}\left(T\left(k\right)^{-1}\sum_t\sum_{j=l}^{t-1}\left(\alpha_0+\beta_0\right)^{2j-l}W_{t-j-1}^2\right) = \\
& \left(\alpha_0+\beta_0\right)^l\lambda_0\left(1+\left(\alpha_0+\beta_0\right)^2+\left(\alpha_0+\beta_0\right)^4+\cdots\right) \\
= \; & \left(\alpha_0+\beta_0\right)^l\frac{\lambda_0}{1-\left(\alpha_0+\beta_0\right)^2},
\end{aligned}$$

and

$$\operatorname{p\,lim}\left(T\left(k\right)^{-1}\sum_t Y_t^2Y_{t-l}^2\right) = \left(\sigma_0^2\right)^2+\left(\alpha_0+\beta_0\right)^{l-1}\left(\alpha_0\lambda_0+\left(\alpha_0+\beta_0\right)\eta_0\right),$$

where $\eta_0 = E\left[\widetilde{h}_t^2\right]$ from Lemma 2. Therefore,

$$\begin{aligned}
\operatorname{p\,lim}\left(T\left(k\right)^{-1}\sum_t g_{3,t}\left(\lambda,\widehat{\sigma}^2\right)\right) &= \left(\left(\alpha_0+\beta_0\right)-\left(\alpha+\beta\right)\right)\times \\
& \quad \left(\alpha_0+\beta_0\right)^{l-1}\left(\alpha_0\lambda_0+\left(\alpha_0+\beta_0\right)\eta_0\right) \\
&= E\left[g_{3,t}\left(\lambda,\sigma_0^2\right)\right].
\end{aligned} \tag{29}$$

For $\max(i) = 2$, (27) and (28) establish $\widehat{g}\left(\lambda,\widehat{\sigma}^2\right) \xrightarrow{p} \overline{g}\left(\lambda,\sigma_0^2\right)$. For $\max(i) = 3$, (27)–(29) establish the same result. Under either specification, let $\overline{Q}\left(\lambda,\sigma_0^2\right) = \overline{g}\left(\lambda,\sigma_0^2\right)' M_0\overline{g}\left(\lambda,\sigma_0^2\right)$, and $\widehat{Q}\left(\lambda,\widehat{\sigma}^2\right) = \widehat{g}\left(\lambda,\widehat{\sigma}^2\right)' M_T\widehat{g}\left(\lambda,\widehat{\sigma}^2\right)$. Then $\widehat{Q}\left(\lambda,\widehat{\sigma}^2\right) \xrightarrow{p} \overline{Q}\left(\lambda,\sigma_0^2\right)$ by continuity of multiplication. For $\max(i) = 2$, (27) and (28) establish that the only $\lambda \in \Lambda$ satisfying $\overline{g}\left(\lambda,\sigma_0^2\right) = 0$ is $\lambda = \lambda_0$, since $\gamma_0 \neq 0$ and $\alpha_0+\beta_0$ is strictly positive. As a consequence, $Q\left(\lambda,\sigma_0^2\right)$ is uniquely minimized at $\lambda = \lambda_0$. A parallel result holds for $\max(i) = 3$, given the aforementioned conditions plus (29) and the fact that $\alpha_0\lambda_0+\left(\alpha_0+\beta_0\right)\eta_0$ is strictly positive.■

LEMMA 4: *$\widehat{S}_\lambda\left(\widehat{\lambda},\widehat{\sigma}^2\right) \xrightarrow{p} S_\lambda\left(\lambda_0,\sigma_0^2\right)$, and $\widehat{S}_{\sigma^2}\left(\lambda_0,\widehat{\sigma}^2\right) \xrightarrow{p} S_{\sigma^2}\left(\lambda_0,\sigma_0^2\right) = 0$ given (i) Assumptions A1 and A2, if $\max(i) = 2$ or (ii) Assumptions A1–A3, if $\max(i) = 3$.*

PROOF OF LEMMA 4: Define $\widehat{s}_{\lambda,ij}\left(\widehat{\lambda},\widehat{\sigma}^2\right)$ as the element in the ith row and jth column of $\widehat{S}_\lambda\left(\widehat{\lambda},\widehat{\sigma}^2\right)$. Let $\ddot{Z}_{t-2}=\left[Y_{t-2}^2\cdots Y_{t-k}^2\right]'$ for $k\geq 2$, and ι be a $(k-1)$-vector of ones. For $\max(i)=3$,

$$\widehat{S}_\lambda\left(\widehat{\lambda},\widehat{\sigma}^2\right)=-T(k)^{-1}\begin{bmatrix}\sum_t Y_t^3 & 0\\ \sum_t\left(Y_t^2-\widehat{\sigma}^2\right)X_{t-1} & \sum_t\left(Y_t^2-\widehat{\sigma}^2\right)X_{t-1}\\ \sum_t\left(Y_t^2-\widehat{\sigma}^2\right)\ddot{Z}_{t-1} & \sum_t\left(Y_t^2-\widehat{\sigma}^2\right)\ddot{Z}_{t-1}\end{bmatrix},$$

and

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

The following results follow from the proof to Theorem 1.

RESULT R1:

$$\begin{aligned}p\lim\left(\widehat{s}_{\lambda,11}\left(\widehat{\lambda},\widehat{\sigma}^2\right)\right)&=-p\lim\left(T(k)^{-1}\sum_t Y_t^2Y_t\right)\\&=-p\lim\left(T(k)^{-1}\sum_t W_tY_t\right)\\&=-\gamma_0\end{aligned}$$

RESULT R2:

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

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

RESULT R3:

$$
\begin{aligned}
p\lim\left(\widehat{s}_{\lambda,31}^{(l)}\left(\widehat{\lambda},\widehat{\sigma}^2\right)\right) &= -p\lim\left(T(k)^{-1}\sum_t Y_t^2 Y_{t-l}\right)+\left(\sigma_0^2\right)^2 \\
&= \left(\alpha_0+\beta_0\right)^{l-1}\left(\alpha_0\lambda_0+\left(\alpha_0+\beta_0\right)\eta_0\right),
\end{aligned}
$$

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

Given R1–R3, $\widehat{s}_{\lambda,ij}\left(\widehat{\lambda},\widehat{\sigma}^2\right) \xrightarrow{p} s_{\lambda,ij}\left(\lambda_0,\sigma_0^2\right)\ \forall\ i,j$. Next, $p\lim\left(\widehat{s}_{\sigma^2,11}\left(\lambda_0,\widehat{\sigma}^2\right)\right)=0$, and $p\lim\left(\widehat{s}_{\sigma^2,21}\left(\lambda_0,\widehat{\sigma}^2\right)\right)=0$ both by C1. Finally, $p\lim\left(\widehat{s}_{\sigma^2,31}\left(\lambda_0,\widehat{\sigma}^2\right)\right)=0$ by C2.■

PROOF OF THEOREM 2: Let $M_T=M_T\left(\widetilde{\lambda},\widehat{\sigma}^2\right)$. Then the first order condition from (14) is

$$
\widehat{S}_\lambda\left(\widehat{\lambda},\widehat{\sigma}^2\right)' M_T \widehat{g}\left(\widehat{\lambda},\widehat{\sigma}^2\right)=0. \tag{30}
$$

Let $H\left(\widehat{\lambda},\overline{\lambda},\sigma_0^2\right)=\widehat{S}_\lambda\left(\widehat{\lambda},\widehat{\sigma}^2\right)' M_T\ \widehat{S}_\lambda\left(\overline{\lambda},\widehat{\sigma}^2\right)$, where $\overline{\lambda}$ is between $\widehat{\lambda}$ and λ_0. Expanding $\widehat{g}\left(\widehat{\lambda},\widehat{\sigma}^2\right)$ first around λ_0, then around σ_0^2, and then solving for $\left(\widehat{\lambda}-\lambda_0\right)$ produces

$$
\begin{aligned}
\sqrt{T(k)}\left(\widehat{\lambda}-\lambda_0\right) &= -H\left(\widehat{\lambda},\overline{\lambda},\sigma_0^2\right)^{-1}\ \widehat{S}_\lambda\left(\widehat{\lambda},\widehat{\sigma}^2\right)' M_T\sqrt{T(k)}\left(\widehat{g}\left(\lambda_0,\sigma_0^2\right)+\widehat{S}_{\sigma^2}\left(\lambda_0,\overline{\sigma}^2\right)\widehat{m}\left(\sigma_0^2\right)\right) \\
&= -\ H\left(\lambda_0,\sigma_0^2\right)^{-1} S_\lambda\left(\lambda_0,\sigma_0^2\right)' M_0\sqrt{T(k)}\widehat{g}\left(\lambda_0,\sigma_0^2\right),
\end{aligned}
$$

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

PROOF OF LEMMA 5: From the definition of $\widehat{\rho}_{t,s}^{(m,n)}(\theta)$,

$$
\widehat{\rho}_{t,s}^{(m,n)}\left(\widehat{\theta}\right)-\widehat{\rho}_{t,s}^{(m,n)}\left(\theta_0\right)=\frac{-6}{T(k,s)^2-1}\left\{T(k,s)^{-1}\sum_t a_{t,s}\left(\widehat{\theta}\right)-a_{t,s}\left(\theta_0\right)\right\}.
$$

By the consistency of $\widehat{\theta}$ established under Theorem 1, $\exists$ a $\delta_t\to 0$ such that $\left\|\widehat{\theta}-\theta_0\right\|\leq \delta_t$. By the triangle inequality,

$$
\left\|T(k,s)^{-1}\sum_t a_{t,s}\left(\widehat{\theta}\right)-a_{t,s}\left(\theta_0\right)\right\|\leq T(k,s)^{-1}\sum_t\left\|a_{t,s}\left(\widehat{\theta}\right)-a_{t,s}\left(\theta_0\right)\right\|\leq T(k,s)^{-1}\sum_t\Delta_{t,s}(\theta).
$$

Finally, by a WLLN, $T(k,s)^{-1}\sum_{t}\Delta_{t,s}(\theta)\xrightarrow{p}E\left[\Delta_{t,s}(\theta)\right]$, which establishes the result.■

PROOF OF THE COROLLARY:

$$\begin{aligned}
\breve{Q}\left(\lambda,\widehat{\sigma}^2\right) &= T(k)^{-2}\sum_{s=1}^{T}\sum_{t\neq s}^{T}g_t\left(\lambda,\widehat{\sigma}^2\right)'M_T g_s\left(\lambda,\widehat{\sigma}^2\right)\\
&= T(k)^{-1}\sum_{s=1}^{T}T(k)^{-1}\sum_{t\neq s}^{T}g_t\left(\lambda,\widehat{\sigma}^2\right)'M_T g_s\left(\lambda,\widehat{\sigma}^2\right)\\
&= T(k)^{-1}\sum_{s=1}^{T}A_s\left(\lambda,\widehat{\sigma}^2\right)g_s\left(\lambda,\widehat{\sigma}^2\right),
\end{aligned}$$

where

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

From the proof to Theorem 1, $\widehat{g}\left(\lambda,\widehat{\sigma}^2\right)\xrightarrow{p}\overline{g}\left(\lambda,\sigma_0^2\right)$ if $\max(i)=2$ or 3, which means that each $A_s\left(\lambda,\widehat{\sigma}^2\right)$ has the same probability limit. As a consequence, $\breve{Q}\left(\lambda,\widehat{\sigma}^2\right)\xrightarrow{p}\overline{Q}\left(\lambda,\sigma_0^2\right)$, which has a unique minimum at $\lambda=\lambda_0$ given Theorem 1.■

References

[1] Andrews, D.W.K., 1988, Laws of large numbers for dependent non-identically distributed random variables, Econometric Theory, 4, 458-467.

[2] Angrist, J., G. Imbens and A. Kreuger 1999, Jackknife instrumental variables estimation, Journal of Applied Econometrics, 14, 57-67.

[3] Bodurtha, J.N. and N.C. Mark, 1991, Testing the CAPM with time-varying risks and returns, Journal of Finance, 46, 1485-1505.

[4] Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307–327.

[5] Brown, B.W. and W.K. Newey, 2002, Generalized method of moments, efficient bootstrapping, and improved inference, Journal of Business and Economic Statistics, 20, 507-571.

[6] Carlstein, E., 1986, The use of subseries methods for estimating the variance of a general statistic from a stationary time series, Annals of Statistics, 14, 1171-1179.

[7] Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics, 18, 5-46.

[8] Cragg, J.G., 1983, More efficient estimation in the presence of heteroskedasticity of unknown form, Econometrica, 51, 751-764.

[9] Donald, S.G., G. Imbens and W.K Newey, 2008, Choosing the number of moments in conditional moment restriction models, MIT working paper.

[10] Donald, S.G., and W.K. Newey, 2000, A jackknife interpretation of the continuous updating estimator, Economic Letters, 67, 239 - 243.

[11] Drost, F.C. and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, Econometrica, 61, 909-927.

[12] Engle, R.F., and J. Mezrich, 1996, GARCH for groups, Risk, 9, 36-40.

[13] Escanciano, J.C., 2009, Quasi-maximum likelihood estimation of semi-strong GARCH models, Econometric Theory, 25, 561-570.

[14] Francq, C., L. Horath and J.M. Zakoian, 2009, Merits and drawbacks of variance targeting in GARCH models, NBER-NSF Time Series Conference proceedings.

[15] Guo, B. and P.C.B Phillips, 2001, Efficient estimation of second moment parameters in ARCH models, unpublished manuscript.

[16] Hafner, C.M., 2003, Fourth moment structure of multivariate GARCH models, Journal of Financial Econometrics, 1, 26-54.

[17] Hall, P. and J.L. Horowitz, 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators, Econometrica, 64, 891-916.

[18] Han, C. and P.C.B. Phillips, 2006, GMM with many moment conditions, Econometrica, 74, 147-192.

[19] Hansen, B., 1994, Autoregressive conditional density estimation, International Economic Review, 35, 705-730.

[20] Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, Econometrica, 50, 1029-1054.

[21] Hansen, L.P., J. Heaton and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business and Economic Statistics, 14, 262-280.

[22] Hansen, P.R. and A. Lunde, 2005, A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, Journal of Applied Econometrics, 20, 873-889.

[23] Harvey, C. and A. Siddique, 1999, Autoregressive conditional skewness, Journal of Financial and Quantitative Analysis, 34, 465-487.

[24] Kendall, M., 1938, A new measure of rank correlation, Biometrica, 30, 81-89.

[25] Lee, S.W, B.E. Hansen, 1994, Asymptotic theory for the GARCH(1,1) qausi-maximum likelihood estimator, Econometric Theory, 10, 29-52.

[26] 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.

[27] Mark, N.C, 1988, Time-varying betas and risk premia in the pricing of forward foreign exchange contracts, Journal of Financial Economics, 22, 335-354.

[28] Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in R.F. Engle and D. McFadded, eds, Handbook of Econometrics, Vol. 4, Amsterdam North Holland, chapter 36, 2111-2245.

[29] Newey, W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica, 72, 219-255.

[30] Newey, W.K. and D.G. Steigerwald, 1997, Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, Econometrica, 65, 587-599.

[31] Newey, W.K and F. Windmeijer, 2009, Generalized method of moments with many weak moment conditions, Econometrica, 77, 687-719.

[32] Pakes, A. and D. Pollard, 1989, Simulation and the asymptotics of optimization estimators, Econometrica, 57, 1027-1057.

[33] Rich, R.W., J. Raymond and J.S. Butler, 1991, Generalized instrumental variables estimation of autoregressive conditional heteroskedastic models, Economics Letters, 35, 179-185.

[34] Schmid, F. and R. Schmidt, 2007, Multivariate extensions of Spearman's rho and related statistics, Statistics and Probability Letters, 77, 407-416.

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

[36] Spearman, C., 1904, The proof and measurement of association between two things, American Journal of Psychology, 15, 72-101.

[37] Stock, J. and J. Wright, 2000, GMM with weak identification, Econometrica, 68, 1055-1096.

[38] Weiss, A.A., 1986, Asymptotic theory for ARCH models: estimation and testing, Econometric Theory, 2, 107-131.

[39] White, H., 1982, Instrumental variables regression with independent observations, Econometrica, 50, 483-499.

[40] Zadrozny, P.A., 2005, Necessary and sufficient restrictions for existence of a unique fourth moment of a univariate GARCH(p,q) process, CESIFO Working Paper No. 1505.

TABLE 1

		True Theta											
		(1.0, 0.15, 0.75)				(1.0, 0.10, 0.85)				(1.0, 0.05, 0.94)			
		Med	Dec			Med	Dec			Med	Dec		
Para.	Est.	Bias	Rge	SD	MDAE	Bias	Rge	SD	MDAE	Bias	Rge	SD	MDAE
Var	QMLE	-0.005	0.242	0.094	0.063	-0.008	0.283	0.111	0.074	-0.022	0.581	0.309	0.156
	MM	-0.018	0.235	0.100	0.060	-0.022	0.289	0.129	0.076	-0.066	0.501	0.272	0.148
Alpha	QMLE	-0.001	0.054	0.021	0.013	0.000	0.039	0.015	0.010	0.000	0.022	0.008	0.005
	JCUE2	-0.016	0.091	0.042	0.028	-0.009	0.067	0.031	0.020	0.000	0.048	0.022	0.011
	JCUE3	-0.001	0.029	0.027	0.006	0.000	0.014	0.011	0.002	0.000	0.004	0.005	0.001
	JGMM2	-0.017	0.109	0.046	0.032	-0.011	0.082	0.035	0.025	0.001	0.067	0.029	0.016
	JGMM3	-0.015	0.090	0.043	0.027	-0.006	0.070	0.034	0.016	-0.001	0.039	0.019	0.005
	CUE2	-0.011	0.109	0.050	0.027	-0.005	0.084	0.043	0.019	-0.004	0.081	0.033	0.018
	CUE3	-0.006	0.040	0.036	0.009	-0.002	0.024	0.026	0.003	-0.001	0.005	0.007	0.001
	GMM2	-0.013	0.112	0.051	0.031	-0.009	0.094	0.041	0.025	-0.007	0.083	0.032	0.021
	GMM3	-0.016	0.113	0.053	0.031	-0.012	0.093	0.042	0.026	-0.010	0.071	0.027	0.019
Beta	QMLE	0.000	0.081	0.033	0.020	0.000	0.056	0.022	0.013	-0.001	0.023	0.009	0.006
	JCUE2	0.010	0.173	0.076	0.043	0.009	0.137	0.061	0.036	-0.008	0.144	0.154	0.031
	JCUE3	0.000	0.104	0.058	0.022	0.000	0.063	0.036	0.015	0.000	0.035	0.022	0.009
	JGMM2	0.011	0.198	0.093	0.053	0.010	0.167	0.077	0.047	-0.030	0.386	0.235	0.043
	JGMM3	0.011	0.158	0.077	0.040	0.006	0.114	0.059	0.029	0.002	0.068	0.035	0.015
	CUE2	-0.040	0.227	0.110	0.051	-0.051	0.211	0.147	0.053	-0.115	0.833	0.325	0.115
	CUE3	-0.024	0.152	0.095	0.031	-0.020	0.130	0.090	0.022	-0.014	0.054	0.086	0.015
	GMM2	-0.053	0.242	0.106	0.061	-0.075	0.272	0.120	0.075	-0.214	0.618	0.247	0.214
	GMM3	-0.031	0.217	0.099	0.044	-0.026	0.144	0.081	0.035	-0.025	0.108	0.059	0.031

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector theta = (Var, Alpha, Beta), where 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.

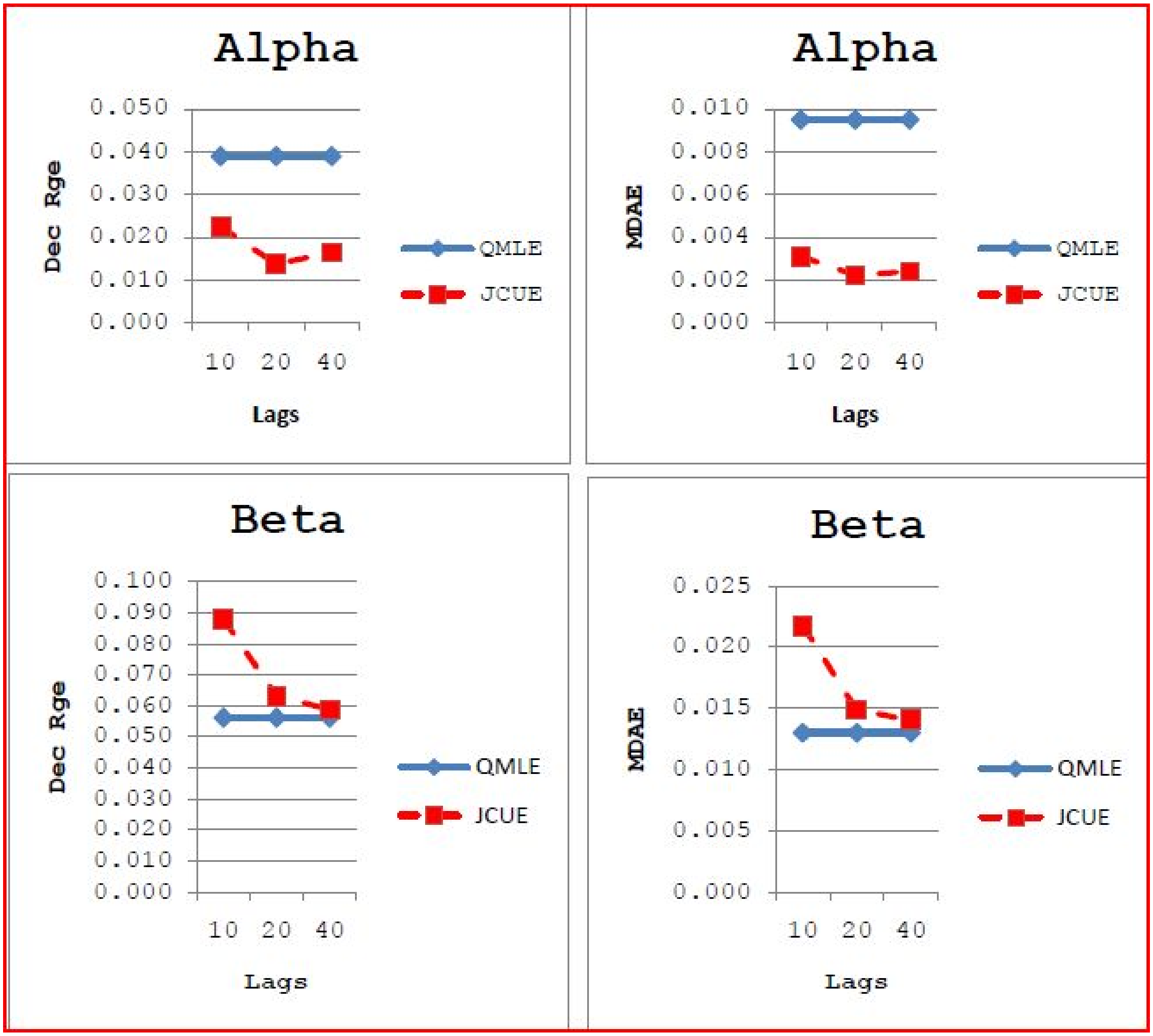

FIGURE 1

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.

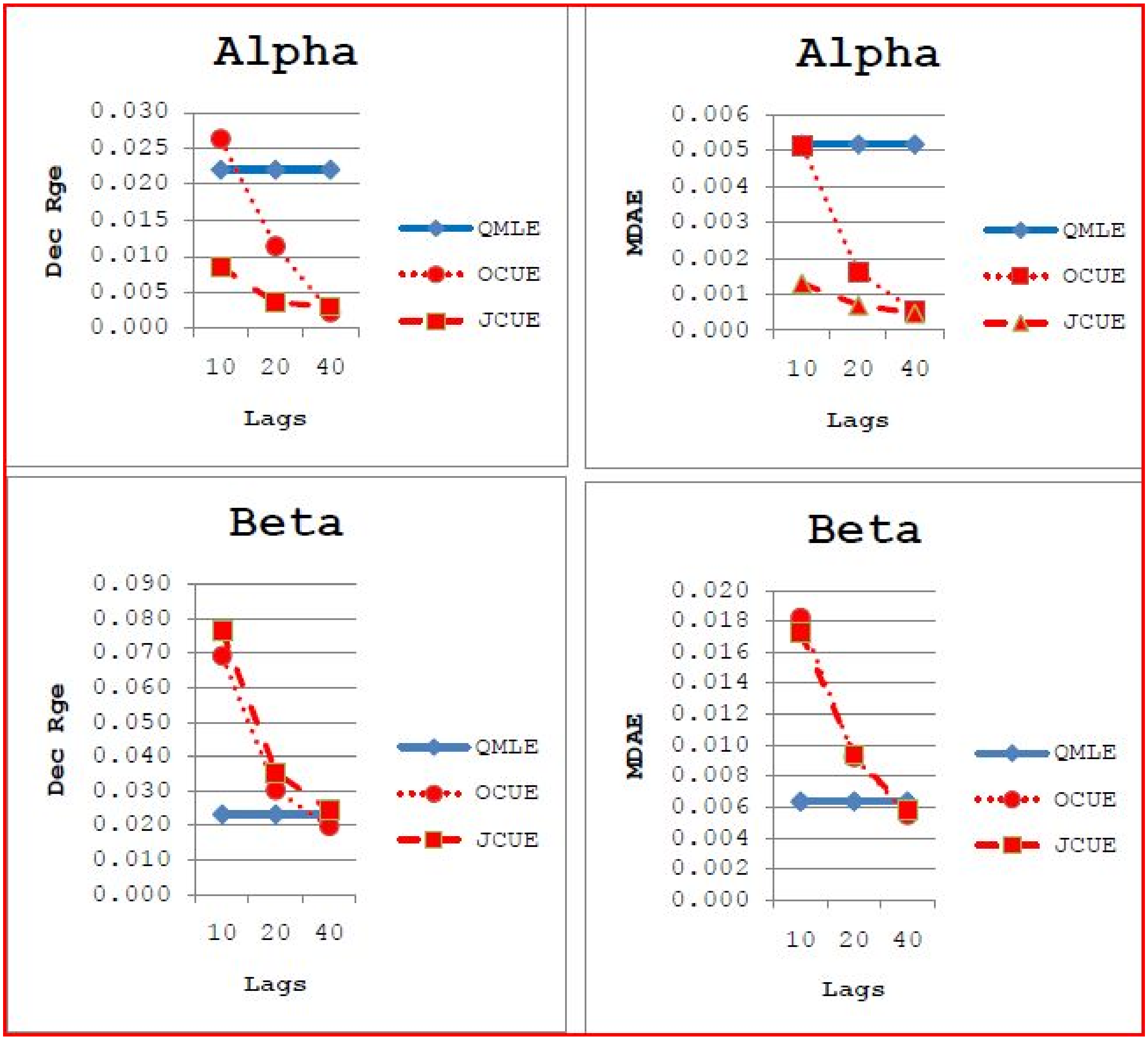

FIGURE 2

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 = 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

		True Theta											
		(1.0, 0.15, 0.75)				(1.0, 0.10, 0.85)				(1.0, 0.05, 0.94)			
		Med	Dec			Med	Dec			Med	Dec		
Para.	Est.	Bias	Rge	SD	MDAE	Bias	Rge	SD	MDAE	Bias	Rge	SD	MDAE
Var	QMLE	-0.006	0.326	0.130	0.088	-0.005	0.388	0.158	0.103	-0.036	0.844	1.155	0.208
	-Ratio		1.338	1.375	1.348		1.333	1.396	1.284		1.404	3.278	1.281
	MM	-0.032	0.314	0.132	0.084	-0.043	0.358	0.150	0.093	-0.090	0.619	0.332	0.174
	-Ratio		1.338	1.329	1.413		1.238	1.163	1.229		1.236	1.223	1.178
Alpha	QMLE	-0.002	0.066	0.026	0.017	-0.001	0.047	0.019	0.012	0.000	0.024	0.009	0.006
	-Ratio		1.225	1.242	1.316		1.218	1.200	1.275		1.109	1.153	1.191
	JCUE3	-0.003	0.040	0.022	0.007	0.000	0.019	0.012	0.003	0.000	0.006	0.009	0.001
	-Ratio		1.344	0.836	1.228		1.393	1.135	1.213		1.670	1.781	1.170
	JGMM3	-0.026	0.089	0.041	0.033	-0.017	0.068	0.033	0.022	-0.004	0.038	0.016	0.009
	-Ratio		0.987	0.950	1.220		0.975	0.953	1.391		0.979	0.821	1.770
	CUE3	-0.011	0.052	0.030	0.013	-0.005	0.042	0.041	0.008	-0.002	0.028	0.033	0.003
	-Ratio		1.288	0.847	1.560		1.795	1.545	2.278		5.172	4.828	2.689
Beta	QMLE	-0.001	0.096	0.039	0.023	0.000	0.064	0.025	0.014	-0.002	0.027	0.011	0.007
	-Ratio		1.183	1.182	1.134		1.138	1.168	1.102		1.158	1.129	1.056
	JCUE3	0.001	0.121	0.061	0.025	0.000	0.074	0.056	0.016	0.000	0.046	0.074	0.010
	-Ratio		1.164	1.049	1.164		1.172	1.556	1.097		1.291	3.312	1.097
	JGMM3	0.018	0.195	0.089	0.047	0.012	0.123	0.074	0.031	0.003	0.080	0.042	0.018
	-Ratio		1.231	1.161	1.181		1.077	1.248	1.089		1.181	1.221	1.263
	CUE3	-0.037	0.187	0.104	0.041	-0.043	0.220	0.120	0.043	-0.030	0.320	0.147	0.031
	-Ratio		1.231	1.101	1.325		1.688	1.325	1.996		5.956	1.721	2.035

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector theta = (Var, Alpha, Beta), where 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 max(i) = 3. JGMM3 is the jackknife two-step generalized method of moments estimator, also with 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

Currency	Para.	JCUE3	OCUE3	QMLE
	k	40	40	
	Var	0.5579	0.5579	0.4957
	Alpha	0.050	0.0890	0.0532
AUD			(0.0648)	(0.0088)
	Beta	0.922	0.9081	0.9382
			(0.0211)	(0.0101)
	Sum	0.9726	0.9971	0.9914
	k	40	40	
	Var	0.4963	0.4963	0.5057
	Alpha	0.049	0.0901	0.0486
JPY			(0.0448)	(0.0095)
	Beta	0.916	0.8864	0.9361
			(0.0147)	(0.0123)
	Sum	0.9650	0.9764	0.9848

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. JCUE3 and OCUE3 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.

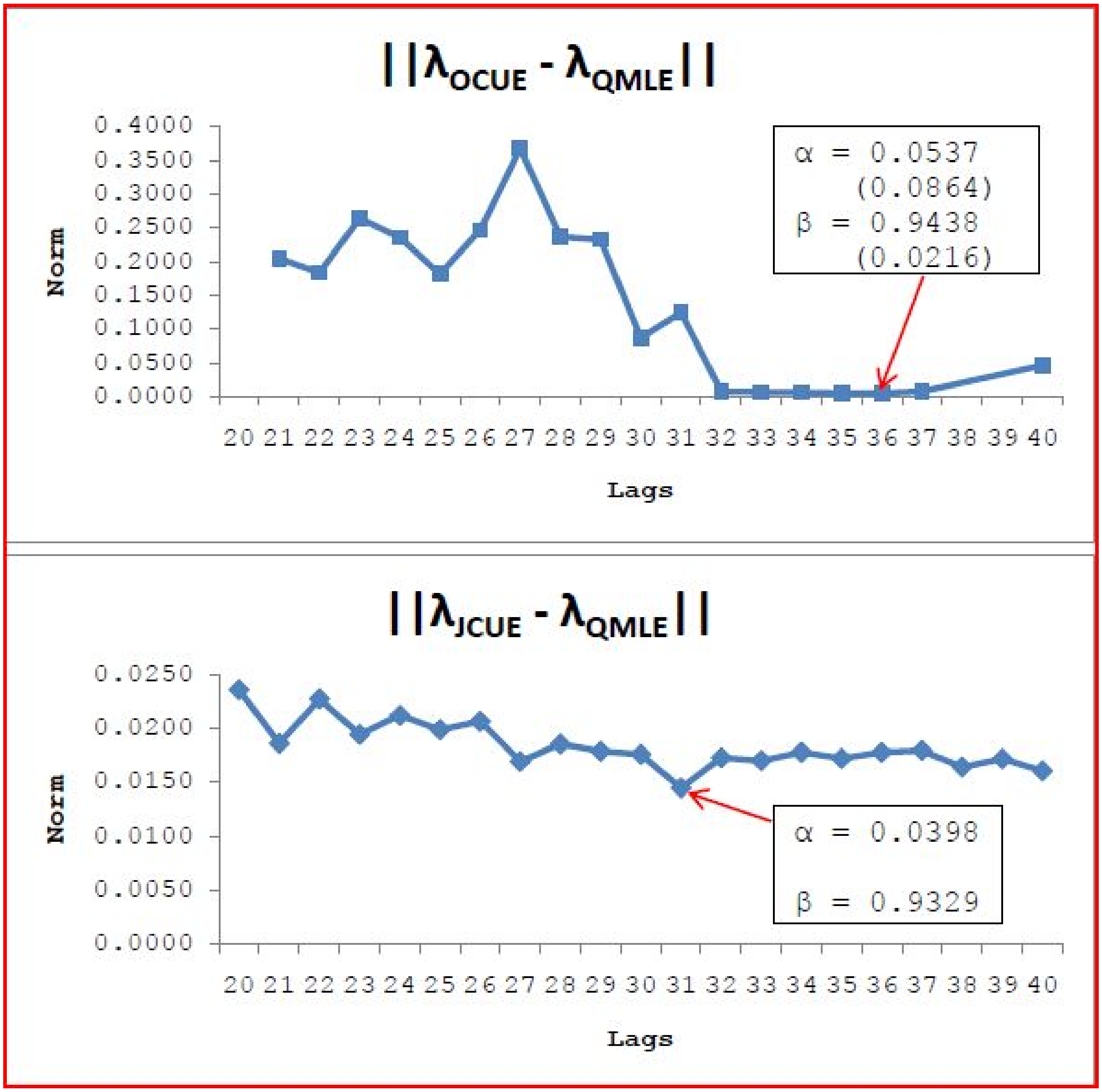

FIGURE 3

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, . . ., 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.

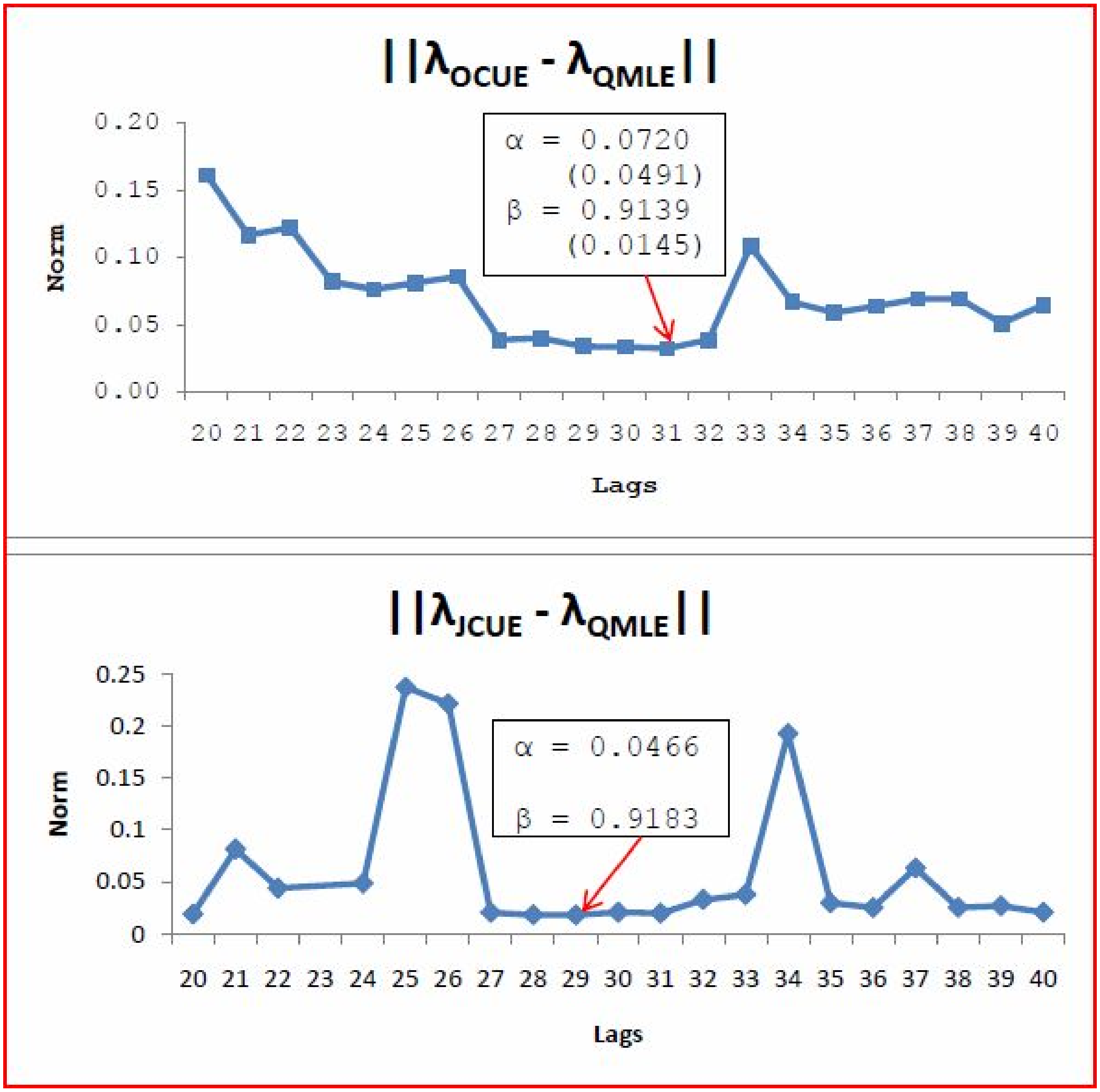

FIGURE 4

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, . . ., 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.