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

Todd, Prono

Commodity Futures Trading Commission

11 November 2009

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

Todd Prono[2]

Commodity Futures Trading Commission

First Version: November 2009
This Version: July 2011

Abstract

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

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

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. The most common estimator for this model is the QMLE. Properties of this estimator are well-studied. For example, 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) case. In this paper, I also consider estimation of the semi-strong GARCH(1,1) model, but I do so through the lens of GMM estimators. In particular, I propose simple GMM estimators constructed from (i) the covariances between past residuals and current squared residuals, and possibly (ii) the autocovariances between squared residuals. These estimators are IV-like, where the instrument vector is comprised of past residuals and past squared residuals.

Weiss (1986) and Guo and Phillips (2001) discuss IV estimators for the ARCH model based on the autocovariances between squared residuals. These estimators do not extend to the GARCH(1,1) case, however, because autocovariances of squared residuals alone are insufficient for identifying the model. I show that the covariances between past residuals and current squared residuals are sufficient for identifying the GARCH(1,1) model if the residuals are skewed, which differentiates my results from Baillie and Chung (2001) and Kristensen and Linton (2006), who both show that autocorrelations of squared residuals can be used to identify the GARCH(1,1) model. Like Kristensen and Linton (2006), the simple GMM estimators I propose also have closed-form expressions that when combined with an iterative GLS estimator have the same asymptotic variance as the QMLE. By the nature of their reliance on third moment properties, however, these simple estimators are CAN under less restrictive moment existence criteria than Kristensen and Linton (2006) and Baillie and Chung (2001) in the GARCH(1,1) case, and Weiss (1986) and Guo and Phillips (2001) in the ARCH(1) case. Additionally, there are cases where the asymptotic variance of these estimators decreases as the absolute value of residual skewness increases (i.e., as the

distribution of residuals moves farther away from normality, these estimators become more efficient).

Meddahi and Renault (1998) recognize that the covariance between the mean and the variance, or skewness, is important for efficiency reasons when considering estimators of ARCH-type processes. This work builds on their results by linking skewness to identification. 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. In the semi-strong GARCH(1,1) case, however, his estimator necessitates the conditional variance function, its first derivative, as well as the third and fourth conditional moments to be included within the moment conditions. The GMM estimators I propose, in contrast, require none of these features. Specifically, neither does the conditional variance function enter the moment conditions nor do the dynamics of the third and fourth moments need to be estimated. These omissions render my estimators simple. Such simplicity, of course, comes at the cost of diminished efficiency. However, even these simple estimators are shown to be serious competitors to the QMLE.

The proposed estimators are overidentified. As a consequence, the choice of a weighting matrix is a material concern. Following Hansen (1982), the optimal weighting matrix involves the variance-covariance matrix of the functions comprising the moment conditions. Since the estimators I propose define moment conditions in terms of the third and possibly the fourth moments, however, use of the variance-covariance matrix 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. This robust analog requires no more than fourth moment existence for consistency, and provides superior finite sample

performance over simple GMM estimators that utilize a non data dependent weighting matrix like the identity matrix.

Finally, the proposed estimators (potentially) involve many moment conditions. From Newey and Windmeijer (2009), the CUE of Hansen, Heaton, and Yaron (1996) with the optimal weighting matrix is robust to the biases caused by many (potentially weak) instruments. The finite sample properties of this estimator 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. The JCUE removes the term responsible for many (weak) moments bias from the CUE objective function. Consistency of the JCUE 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 JCUE a robust alternative. Monte Carlo studies uncover cases where both the OCUE and the JCUE are more efficient than the QMLE. These efficiency gains relate to the number of instruments used in constructing the respective estimators.

2. The Model and Implications

For $\{Y_t\}_{t\in\mathbb{Z}}$, let $\mathcal{F}_t$ be the associated σ-algebra where $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \cdots \subseteq \mathcal{F}$. The first two conditional moments are

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

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 $Y_t/h_t^{1/2}$ is iid and

4

drawn from a known distribution nests as a special case. Consider the following additional assumptions.

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

Given A1, h_t is everywhere strictly positive. Lumsdaine (1996) supplies the individual bounds on ω, α, and β. Since $\beta \geq 0$, A1 nests the ARCH(1) model.

Given $\alpha + \beta < 1$, Y_t is covariance stationary with $E[Y_t^2] = \sigma_0^2$ (see Theorem 1 of Bollerslev 1986). Therefore, the mean-adjusted form of (2) is

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

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

$$\widetilde{X}_t = \widetilde{h}_t + W_t, \tag{4}$$

where W_t is a martingale difference sequence (MDS), with $E[W_t \mid \mathcal{F}_{t-1}] = 0$ and $E[W_t W_{t-k}] = 0 \, \forall \, k \geq 1$.

ASSUMPTION A2: *(i) $E[Y_t^3] = \gamma_0 \neq 0$. (ii) $E|W_t Y_t| < \infty$. (iii) $\left\{\widetilde{U}_{t,k}\right\}$ is uniformly integrable, where $\widetilde{U}_{t,k} \equiv \widetilde{X}_t Y_{t-k} - E\left[\widetilde{X}_t Y_{t-k}\right]$ for $k = 1, \ldots, K$.*

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

$$E\left[\widetilde{X}_t Y_{t-1}\right] = \alpha_0 E[Y_t^3], \tag{5}$$

and

$$E\left[\widetilde{X}_t Y_{t-(k+1)}\right] = \phi_0 E\left[\widetilde{X}_t Y_{t-k}\right], \tag{6}$$

where $\phi_0 = \alpha_0 + \beta_0$.

Proof. All proofs are stated in the Appendix. ∎

Lemma 1 relates the covariance between $\widetilde{X}_t$ and Y_{t-k} to the third moment of Y_t (see (22) in the Appendix). Lemma 1 of Guo and Phillips (2001) establishes an analogous result for the ARCH(p) model. In contrast to Guo and Phillips, the Lemma presented here is central to identification because it provides the moment condition in (5) 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 GMM estimators 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 GMM estimators that are not identified without such skewness.

ASSUMPTION A3: *(i)* $E\left[W_t^2\right] = \lambda_0$. *(ii)* $\left\{\widetilde{V}_{t,k}\right\}$ *is uniformly integrable, where* $\widetilde{V}_{t,k} \equiv \widetilde{X}_t \widetilde{X}_{t-k} - E\left[\widetilde{X}_t \widetilde{X}_{t-k}\right]$ *for* $k = 2, \dots, K$.

Suppose

$$Y_t = h_t^{1/2}\epsilon_t, \quad \epsilon_t \sim iid\left(0,\ 1\right). \tag{7}$$

Then A3(i) is equivalent to assuming that $E\left[\left(\beta + \alpha\epsilon_t^2\right)^2\right] < 1$, which grants Y_t to have a finite fourth moment (see Carrasco and Chen 2002, Corollary 6) and so strengthens A1.[3] Finally, A3(i) is sufficient for both A2(ii) and A2(iii). These latter two assumptions are only necessary when A3 does not hold.

It is straight-forward to express (4) as

$$\widetilde{X}_t = \phi_0 \widetilde{X}_{t-1} + W_t - \beta_0 W_{t-1}. \tag{8}$$

---

[3]Of course, in the semi-strong GARCH case, A3(i) also strengthens A1, but in an unknown way owing to possible dependence in the fourth moment of ϵ_t.

Multiplying both sides of (8) by $\widetilde{X}_{t-1}$ and taking expectations produces

$$
\begin{aligned}
E\left[\widetilde{X}_t\widetilde{X}_{t-1}\right] &= \phi_0 E\left[\widetilde{h}_{t-1}^2\right] + \alpha_0 E\left[W_{t-1}^2\right] & (9) \\
&= \left[\frac{(1-\phi_0\beta_0)(\phi_0-\beta_0)}{1-\phi_0^2}\right]\lambda_0,
\end{aligned}
$$

where the second equality follows from Lemma 2 (see the Appendix). Multiplying both sides of (8) by $\widetilde{X}_{t-k}$ for $k \geq 2$ and taking expectations then produces

$$
E\left[\widetilde{X}_t\widetilde{X}_{t-k}\right] = \phi_0 E\left[\widetilde{X}_t\widetilde{X}_{t-k+1}\right]. \tag{10}
$$

Even given (10), (9) does not identify β_0 owing to the presence of λ_0. Autocovariances of $\widetilde{X}_t$ alone, therefore, are insufficient for identifying the GARCH(1,1) model.

Let $\rho(k) = \frac{E[\widetilde{X}_t\widetilde{X}_{t-k}]}{E[\widetilde{X}_t^2]}$ for $k \geq 1$. Then

$$
\rho(1) = \frac{(1-\phi_0\beta_0)(\phi_0-\beta_0)}{1+\beta_0^2-2\phi_0\beta_0}, \tag{11}
$$

and $\rho(k) = \phi_0\rho(k-1)$ for $k \geq 2$.[4] Kristensen and Linton (2006) show that (11) can be expressed as a quadratic equation in β_0 with a unique solution based on ϕ_0 and $\rho(1)$ if and only if $\beta_0 > 0$. Autocorrelations of $\widetilde{X}_t$ do, therefore, identify the GARCH(1,1) model.

Lemma 1 identifies the GARCH(1,1) model in an analogous fashion to (11) and $\phi_0 = \rho(2)/\rho(1)$. Advantages of basing identification on Lemma 1 include allowing β_0 to be zero and not requiring the fourth moment of Y_t to be finite.

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 $Z_{1,t-2} = \left[Y_{t-2}, \cdots, Y_{t-k}\right]'$ and $Z_{2,t-2} =$

---

[4]These equations are derived in Bollerslev (1988) and He and Teräsvirta (1999).

$\left[Y_{t-2}^2 - \sigma^2, \cdots, Y_{t-k}^2 - \sigma^2 \right]'$ for $2 \leq k \leq K$. Consider the following vector valued functions

$$g_{1,t}\left(Y;\ \lambda, \sigma^2\right) = \left(Y_t^2 - \sigma^2\right) Y_{t-1} - \alpha Y_t^3, \tag{12}$$

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

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

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\left(i\right),\ 2 \leq \max\left(i\right) \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^{-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], \\
\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^{-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\left(T^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 $m, n = 1, \ldots, 2k - 1$.

3.2. CAN and Robust Estimators

Consider

$$\widehat{\lambda} = \arg\min_{\lambda \in \Lambda}\ \widehat{g}\left(\lambda, \widehat{\sigma}^2\right)' M_T \widehat{g}\left(\lambda, \widehat{\sigma}^2\right), \tag{13}$$

where M_T is positive semi-definite. (13) is the familiar GMM estimator of Hansen (1982) with $\widehat{\sigma}^2$ plugged-in. Given this plug-in feature, (13) is also a VTE similar to that studied by Engle and Mezrich (1996) as well as by Francq, Horath, and Zakoian (2009). The moment conditions in (13) are the finite sample analogs to (5), (6), and (10). Depending on the choice for M_T, (13) supports either the traditional two-step GMM estimator or the CUE, the latter of which is shown to be a member of the class of Generalized Empirical Likelihood (GEL) estimators by Newey and Smith (2004). Newey and Windmeijer (2009) show GEL estimators to be more efficient than (jackknife) GMM estimators under many (potentially weak) moments. Given the reliance of $\widehat{g}\left(\lambda, \widehat{\sigma}^2\right)$ on k, the association of (13) to the CUE is important both asymptotically as well as for finite sample performance.

If $\beta = 0$, then (13) has a closed-form solution. Moreover, even if $\beta > 0$, (13) retains a closed-form solution; namely

$$
\widehat{\sigma}^2 = T^{-1} \sum_t Y_t^2, \qquad \widehat{\alpha} = \frac{\sum_t \widehat{\widetilde{X}}_t Y_{t-1}}{\sum_t Y_t^3}, \tag{14}
$$

$$
\widehat{\beta} = \left(\left[\sum_t \widehat{\widetilde{X}}_t \widehat{Z}_{t-1}\right]' M_T \left[\sum_t \widehat{\widetilde{X}}_t \widehat{Z}_{t-1}\right]\right)^{-1} \left[\sum_t \widehat{\widetilde{X}}_t \widehat{Z}_{t-1}\right]' M_T \left[\sum_t \widehat{\widetilde{X}}_t \left(\widehat{Z}_{t-2} - \widehat{\alpha} \widehat{Z}_{t-1}\right)\right],
$$

where $\widehat{Z}_{t-2} = \begin{pmatrix} Z_{1,t-2} \\ \widehat{Z}_{2,t-2} \end{pmatrix}$ and M_T is $2(k-1) \times 2(k-1)$, making it comparable to the GARCH(1,1) estimator in Kristensen and Linton (2006).

ASSUMPTION A4: *(i) $\exists$ a neighborhood N of θ_0 such that $E\left[\sup_{\theta \in N} \left\|g_t\left(\lambda_0, \sigma_0^2\right) g_t\left(\lambda_0, \sigma_0^2\right)'\right\|\right] < \infty$; or (ii) given (7), $E\left[\left(\beta + \alpha\epsilon_t^2\right)^s\right] < 1$ for $s \geq 3$.*

ASSUMPTION A5: *$S_\lambda\left(\lambda_0, \sigma_0^2\right)' M_0 S_\lambda\left(\lambda_0, \sigma_0^2\right)$ is nonsingular.*

ASSUMPTION A6: *The conditions relating to an L^2 mixingale in Assumption 1 of De Jong (1997) hold.*

THEOREM. *Consider the estimator in (13) for the model of (1) and (2). Let $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2$, and assume that $M_T \overset{p}{\to} M_0$, a positive definite matrix. If $\max(i) = 2$,*

then $\widehat{\lambda} \xrightarrow{p} \lambda_0$ given Assumptions A1 and A2. If $\max(i) = 3$, then $\widehat{\lambda} \xrightarrow{p} \lambda_0$ given Assumptions A1–A3. If, in addition, Assumptions A4(i), A5, and A6 hold, then

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

(15)

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

The first part of the Theorem establishes weak consistency of (13) through the properties of L^1 mixingales (see Andrews 1988). When $\max(i) = 2$, third moment existence is necessary for this result. When $\max(i) = 3$, fourth moment existence becomes necessary, owing to the consideration of autocovariances between squared residuals.[5] Theorem 4.4 of Weiss (1986), the estimator in Rich et al. (1991), as well as Theorems 2.2 and 4.1 of Guo and Phillips (2001) all require fourth moment existence for the consistency of their, respective, ARCH model estimators. Baillie and Chung (2001) and Kristensen and Linton (2006) require the same condition for autocorrelation-based estimators of the GARCH(1,1) model. The Theorem replaces necessary with sufficient for the condition of a finite fourth moment by nature of the fact that identification links to properties of the third moment.

Given (4), it is straight-forward to show that

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

(16)

where $X_{-1} = \left[\begin{array}{cc} \widetilde{X}_{t-1}, & \widetilde{h}_{t-1} \end{array}\right]'$ and $Z_{-1} = \left[\begin{array}{cc} Y_{t-1}, & \widetilde{Z}_{t-2}' \end{array}\right]'$, thus linking (13) to IV estimation. The sample moment conditions associated with the left-hand-side of (16), however, are infeasible, since they involve elements not included in the time-t information set. The sample moment conditions associated with the right-hand-side of (16), on the other hand, are feasible, since they are only a function of $\{Y_t\}_{t=1}^T$. As a consequence, (13) can be regarded as a feasible IV-like estimator for the GARCH(1,1) model constructed using an "instrument vector" of past residual and squared residual values.

The second part of the Theorem establishes the traditional asymptotic result for GMM

---

[5]Such consideration is made for efficiency reasons, since the introduction of autocovariances of squared residuals provides additional moment conditions without adding parameters.

estimators using the CLT for L^2 mixingales developed by De Jong (1997). This result, of course, is also efficient if $M_0 = \Omega\left(\lambda_0, \sigma_0^2\right)^{-1}$.[6] In the efficient case, $\widehat{\Omega}\left(\widehat{\lambda}, \widehat{\sigma}^2\right) \xrightarrow{p} \Omega\left(\lambda_0, \sigma_0^2\right)$ given $\left\{g_{t-s}\left(\lambda_0, \sigma_0^2\right) g_t\left(\lambda_0, \sigma_0^2\right)'\right\}_{t=1}^{T}$ satisfying the UWLLN and Lemma 4.3 of Newey and McFadden (1994) applied to $a\left(z, \theta\right) = g_{t-s}\left(\lambda, \sigma^2\right) g_t\left(\lambda, \sigma^2\right)'$.[7] Also worthy of note is that the asymptotic variance of $\widehat{\sigma}^2$ does not impact the asymptotic variance of $\widehat{\lambda}$, meaning that nothing is lost (asymptotically) by plugging $\widehat{\sigma}^2$ into (13) as opposed to σ_0^2. This result stands in contrast to the VTE studied by Francq, Horath, and Zakoian (2009).

COROLLARY 1. *For the estimator in (13), let* $\widehat{\sigma}^2 = T^{-1}\sum\limits_{t=1}^{T}Y_t^2$, *and* $M_T \xrightarrow{p} M_0$, *a positive definite matrix. If* $\max\left(i\right) = 2$, *then (15) holds given A4(ii) with* $s = 3$ *and A5. If* $\max\left(i\right) = 3$, *then (15) holds given A4(ii) with* $s = 4$ *and A5.*

Corollary 1 facilitates comparison of the asymptotic properties of (13) to those of the estimator in Kristensen and Linton (2006). Establishing $\sqrt{T}$-asymptotic normality for the latter case requires existence of the eighth moment, or, specifically, A4(ii) to hold with $s = 4$. $\sqrt{T}$-asymptotic normality of (13) can result, on the other hand, given existence of only the sixth moment, since the estimator relies on third moment properties for identification.[8]

Rather than relying on asymptotic approximations (and the higher moment existence criteria those approximations entail), standard errors for (13) can, alternatively, be computed via the parametric bootstrap. Suppose that the data generating process for Y_t is characterized by (1), (2), and (7), where $E\left[\epsilon_t \mid F_{t-1}\right] = 0$, $E\left[\epsilon_t^2 \mid F_{t-1}\right] = 1$, and the higher moments of ϵ_t follow L^{th} order Markov processes with a finite $L << T$. Use (13) to obtain $\widehat{h}_t$. Let $\widehat{\epsilon}_t = Y_t/\sqrt{\widehat{h}_t}$, 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

---

[6] The proof of this result is based on the two-step GMM estimator. For the CUE, although the first order condition contains an additional term, this term does not distort the limiting distribution in (15). Pakes and Pollard (1989) discuss this result in detail as do Donald and Newey (2000).

[7] The UWLLN replaces Khintchine's law of large numbers in the proof of Lemma 4.3.

[8] Since (13) nests the ARCH(1) model, this same condition (also shared by the Theorem) relaxes the moment existence criteria necessary for asymptotic normality under Theorem 4.4 of Weis (1986).

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.

3.3. Efficiency Issues

From (15), let $V_{GMM} = H(\theta_0)^{-1}$ when $M_0 = \Omega(\theta_0)^{-1}$. If $\max(i) = 2$, then

$$V_{GMM} = \frac{1}{\gamma_0^2} \left(\Phi_0' M_0 \Phi_0\right)^{-1},$$

where the individual entrees of Φ_0 are functions of α_0, β_0, and k. This expression illustrates the underidentification of λ_0 when $\{Y_t\}$ is symmetrically distributed.

ASSUMPTION A7: *For an $r > 0$, $|\gamma_0| < \frac{2}{r}$ excluding an open set around zero. For any $x \neq 0$, (i) $x' \frac{\partial \Omega(\theta_0)}{\gamma} x \leq r x' \Omega(\theta_0) x$ if $\gamma_0 > 0$, while (ii) $x' \frac{\partial \Omega(\theta_0)}{\gamma} x \geq -r x' \Omega(\theta_0) x$ if $\gamma_0 < 0$.*

PROPOSITION. *Let Assumption A7 hold. Then V_{GMM} decreases as $|\gamma_0|$ increases.*

As skewness increases in absolute value, (13) becomes more efficient. When $\gamma_0 > 0$, $\frac{\partial \Omega(\theta_0)}{\gamma}$ can be expected to be positive definite, since a positive change in γ_0 can be expected to increase the variance of the moment conditions through an increase in the higher moments of $\{Y_t\}$.[9] Conversely, when $\gamma_0 < 0$, $\frac{\partial \Omega(\theta_0)}{\gamma}$ can be expected to be negative definite, since positive changes in γ_0 can be expected to decrease the variance of the moment conditions by decreasing the higher moments of $\{Y_t\}$. The substantive assumption of the Proposition, therefore, is that the size of $\frac{\partial \Omega(\theta_0)}{\gamma}$ is bounded by the size of $\Omega(\theta_0)$.

Populate the parameter vector $\widehat{\vartheta} = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})'$ using (14) and $\widehat{\omega} = \widehat{\sigma}^2 \left(1 - \widehat{\phi}\right)$. Define the iterative GLS iterative estimator,

$$\widehat{\vartheta}_{l+1}^{GLS} = \left(\sum_t \widehat{h}_{l,t}^{-2} X_{l,t-1} X_{l,t-1}'\right)^{-1} \left(\sum_t \widehat{h}_{l,t}^{-2} X_{l,t-1} Y_t^2\right), \quad l \geq 1, \tag{17}$$

---

[9] Given (7), examples where this statement is true include ϵ_t being distributed as as a standardized $\Gamma(\kappa, \vartheta)$ or $\chi(\kappa)$ distribution for strictly positive and decreasing values of κ.

where $\widehat{h}_{l,t} = \widehat{\omega}_l^{GLS} + \widehat{\alpha}_l^{GLS} Y_{t-1}^2 + \widehat{\beta}_l^{GLS} \widehat{h}_{l,t-1}$, $X_{l,t-1} = \left[1, \; Y_{t-1}^2, \; \widehat{h}_{l,t-1}\right]'$, and $\widehat{\vartheta}_1^{GLS} = \widehat{\vartheta}$. From Kristensen and Linton (2006, Theorem 3),

$$\sqrt{T}\left(\widehat{\vartheta}_{l+1}^{GLS} - \vartheta_0\right) \xrightarrow{d} N\left(\; 0, \;\; H^{-1}\Sigma H^{-1}\;\right),$$

where H is the Hessian of the QMLE for the semi-strong GARCH(1,1) model, and Σ is the variance-covariance matrix of the score. Given (14) and (17), it is possible to define a semi-strong GARCH(1,1) estimator that does not require any numerical optimization and has the same asymptotic variance as the QMLE (see, e.g., Bollerslev and Wooldridge 1992 and Lee and Hansen 1994).

3.4. The Weighting Matrix

The estimator in (13) requires specification of a weighting matrix. Use of the optimal weighting matrix requires existence of, at least, the sixth moment and as high as the eighth if autocovariances are also considered. Such an assumption may prove overly restrictive, especially for certain financial data. A key question, therefore, is what potential weighting matrices exist that economize on the number of higher moment existence criteria needed for consistency. One option, of course, is to use a non data dependent weighting matrix like the identity matrix. Skoglund (2001), however, reports that the identity matrix used in an Efficient IV estimator for the strong GARCH(1,1) model results in quite poor finite sample performance. This result is also found (though not reported) in Monte Carlo studies of (13). 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 3. *Let* $a_{t,s}\left(\theta\right) = \left\{R\left[g_{m,t}\left(\theta\right)\right] - R\left[g_{n,t-s}\left(\theta\right)\right]\right\}^2$. *For a* $\delta_t \to 0$, *define* $\Delta_{t,s}\left(\theta\right) = \sup_{\|\theta-\theta_0\|\leq\delta_t} \left\|a_{t,s}\left(\theta\right) - a_{t,s}\left(\theta_0\right)\right\|$. *Assume that* $\left\{\Delta_{t,s}\left(\theta\right)\right\}$ *satisfies the UWLLN. Then for*

$$\widehat{\theta} \xrightarrow{p} \theta_0, \ \widehat{\rho}_{t,s}^{(m,n)}\left(\widehat{\theta}\right) - \widehat{\rho}_{t,s}^{(m,n)}\left(\theta_0\right) \xrightarrow{p} 0.$$

Consistency of $\widehat{\rho}_{t,s}^{(m,n)}\left(\widehat{\theta}\right)$ follows from Lemma 5 and selected results in Schmid and Schmidt (2007).[10] Conditions for consistency involve the copula for $g_{m,t}\left(\theta_0\right)$ and $g_{n,t-s}\left(\theta_0\right)$ (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.

3.5. Many (Weak) Moments Bias Correction

For the estimator in (13), k (the number of lags, which corresponds to the number of instruments) needs to be specified. Standard GMM asymptotics point to efficiency gains from increasing k. Work by Stock and Wright (2000), Newey and Smith (2004), Han and Phillips (2006), and Newey and Windmeijer (2009), however, discuss the biases of GMM estimators when the instrument vector is large, (possibly) inclusive of (many) weak instruments, and allowed to grow with the sample size. To see how these biases relate to k, suppose that there exists a finite L such that $E\left[g_t\left(\theta\right) \mid F_{t-L}\right]$ is constant.[11] Let $s^* = \{S: \ s \geq t + L \text{ or } s \leq t - L; \ s = 1, \ldots, T\}$. Then, the expectation of the GMM objective function $\widehat{g}\left(\theta\right)' M_T \widehat{g}\left(\theta\right)$ for a nonrandom weighting matrix M_T is

$$
\begin{aligned}
E\left[\widehat{g}\left(\theta\right)' M_T \widehat{g}\left(\theta\right)\right] &= T^{-2} E\left[\sum_{t \in s^*} g_t\left(\theta\right)' M_T g_{s^*}\left(\theta\right) + \sum_{s=-(L-1)}^{s=(L-1)} \sum_t g_t\left(\theta\right)' M_T g_{t-s}\left(\theta\right)\right] \quad (18) \\
&= \left(1 - \frac{L}{T}\right) \overline{g}\left(\theta\right)' M_T \overline{g}\left(\theta\right) + T^{-1} tr\left(M_T \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{t-s}\left(\theta\right) g_t\left(\theta\right)'\right]\right),
\end{aligned}
$$

which is an adaptation of (2) in Newey and Windmeijer (2009) to dependent time series data.[12]

---

[10] 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 $\widehat{\rho}_{t,s}^{(m,n)}\left(\theta_0\right)$.

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

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

In the language of Newey and Windmeijer (2009), $\left(1 - \frac{L}{T}\right) \bar{g}\left(\theta\right)' M_T \bar{g}\left(\theta\right)$ 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 $g_t\left(\theta\right)$, as is the case, generally, in the IV setting, and is increasing in k.[13] From (18), if $M_T = \Omega\left(\theta\right)^{-1}$, then the "noise" term is no longer a function of θ, and the GMM objective function is minimized at the truth. This result shows that (13) specified as the optimal CUE (OCUE) is robust to many (potentially weak) instruments.

If $M_T \neq \widehat{\Omega}\left(\lambda, \widehat{\sigma}^2\right)^{-1}$ (e.g., $M_T = \widehat{\Omega}\left(\widetilde{\lambda}, \widehat{\sigma}^2\right)^{-1}$ for some preliminary consistent estimator $\widetilde{\lambda}$ or $M_T = \widehat{\Sigma}\left(\lambda, \widehat{\sigma}^2\right)^{-1}$), then (13) will be biased and increasingly so at large values of k. To correct for this problem, consider the estimator

$$\breve{\lambda} = \arg\min_{\lambda \in \Lambda} \breve{Q}\left(\lambda, \widehat{\sigma}^2\right), \tag{19}$$

where

$$
\begin{aligned}
\breve{Q}\left(\lambda, \widehat{\sigma}^2\right) &= T^{-2} \sum_{t \in s^*} g_t\left(\lambda, \widehat{\sigma}^2\right)' M_T g_{s*}\left(\lambda, \widehat{\sigma}^2\right) \\
&= \widehat{Q}\left(\lambda, \widehat{\sigma}^2\right) - T^{-1} tr\left(M_T \left\{ \sum_{s=-(L-1)}^{s=(L-1)} T^{-1} \sum_t g_{t-s}\left(\lambda, \widehat{\sigma}^2\right) g_t\left(\lambda, \widehat{\sigma}^2\right)' \right\}\right),
\end{aligned}
\tag{20}
$$

and $\widehat{Q}\left(\lambda, \widehat{\sigma}^2\right) = \widehat{g}\left(\lambda, \widehat{\sigma}^2\right)' M_T \widehat{g}\left(\lambda, \widehat{\sigma}^2\right)$. (19) removes the "noise" term from the GMM objective function. It will be referred to as the jackknife CUE (JCUE) when $M_T = \widehat{\Sigma}\left(\lambda, \widehat{\sigma}^2\right)^{-1}$ because, as seen through (20), it leaves out contemporaneous and certain lagged observations from the CUE objective function.

COROLLARY 2. *Consider the estimator in (13) for the model of (1) and (2). Let $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2$, and assume that $M_T \overset{p}{\to} M_0$, a positive definite matrix. In addition, assume that $L = 1$. If $\max\left(i\right) = 2$, then $\breve{\lambda} \overset{p}{\to} \lambda_0$ given Assumptions A1–A2. If $\max\left(i\right) = 3$, then $\breve{\lambda} \overset{p}{\to} \lambda_0$ given Assumptions A1–A3.*

When $L = 1$, a straightforward way of demonstrating consistency of (19) is by examining

---

[13]This "noise" or bias term is analogous to the higher order bias term B_G in Newey and Smith (2004). If k is increasing with T, this term need not even vanish asymptotically (see Han and Phillips 2006), although this case is beyond the scope of the paper.

the second equality in (20), in which case, the conditions under the Theorem (including A4–A6) 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 (19). Corollary 2, therefore, bases consistency on the first equality in (20) and shows that A1–A3 are sufficient.

Following from Newey and Windmeijer (2009, p. 702), the two-step version of $\breve{\lambda}$ is asymptotically normal (provided that the requisite moment existence criteria hold) if $L = 1$. If $\beta_0 = 0$, $L = 1$, and $\breve{\lambda}$ is the two-step GMM estimator, then the solution to (19) is JIVE2 from Angrist, Imbens, and Krueger (1999).

From (14), the closed-form estimator is susceptible to many moments bias through $\widehat{\beta}$. Following the discussion above, one solution to this problem is to estimate $\widehat{\beta}$ using JIVE2. Alternatively, one can estimate $\widehat{\beta}$ using either the OCUE or the JCUE. In these cases, a closed-form solution for $\widehat{\beta}$ is no longer available; however, minimization of the relevant objective function via a grid search is feasible, thus bypassing the need for numerical optimization techniques. Since JIVE2 is a special case of JGMM, and Newey and Windmeijer (2009) show the CUE to be more efficient JGMM under many moments, it is likely that the alternative involving CUE for $\widehat{\beta}$ will be preferable.

4. Monte Carlo

Consider the data generating process in (1), (2), and (7) for different values of θ_0, where ϵ_t is the negative of a standardized $\Gamma(\kappa, 1)$ random variable, with values of κ ranging from 2 to 5. Simulations consider the OCUE and JCUE benchmarked against the QMLE. Two-step GMM and JGMM estimators are not considered because of the results from Newey and Windmeijer (2009). All simulations are conducted with 5,000 observations across 500 trials. When generating those observations, the first 200 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 efficiency, 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. Finally, MM denotes the method of moments plug-in estimator $\widehat{\sigma}^2$.

Table 1A summarizes the results for the OCUE and JCUE (13 and 19, respectively) at various lag lengths k, when $\max(i) = 2$ and $\kappa = 2$. For this specification of ϵ_t and the three values of θ_0 considered, Y_t has at least a finite fourth moment. MM estimates $\widehat{\sigma}^2$ with more bias than does QMLE, but also with less dispersion. With near uniformity, the dispersion of $\widehat{\alpha}(k)$ and $\widehat{\beta}(k)$ for each estimator is decreasing in k. JCUE tends to be less biased than OCUE for $\widehat{\beta}(k)$, although the magnitudes of the bias for OCUE tend to be small. JCUE is significantly more dispersed than OCUE. The dispersion of $\widehat{\alpha}(k)$ for OCUE is less than that of $\widehat{\alpha}$ for $k = 20, 40$. The dispersion of $\widehat{\beta}(k)$ for OCUE approaches that of $\widehat{\beta}$ as k increases, exceeding it for $\theta_0^{(1)}$ and $\theta_0^{(2)}$. In these latter two cases, however, the bias of $\widehat{\beta}(k)$ is higher than that of $\widehat{\beta}$. In summary, when $\max(i) = 2$, OCUE becomes comparable to QMLE as k increases. JCUE does not.

Table 1B summarizes the results for the OCUE and JCUE under the same conditions as Table 1A except that $\max(i) = 3$. In this case, for all values of θ_0 considered, $\widehat{\alpha}(k)$ from the OCUE is more efficient than $\widehat{\alpha}$ for all k considered. For $\theta_0^{(3)}$, $\widehat{\beta}(k)$ from the OCUE is more efficient than $\widehat{\beta}$ for $k = 40$. For $\theta_0^{(1)}$ and $\theta_0^{(2)}$, $\widehat{\beta}(k)$ from the OCUE is less dispersed than $\widehat{\beta}$, but with higher biases. For the JCUE, $\widehat{\alpha}(k)$ is more efficient than $\widehat{\alpha}$ for all k considered, and $\widehat{\beta}(k)$ is seen to approach the efficiency of $\widehat{\beta}$ as k increases. In summary, when $\max(i) = 3$, OCUE and JCUE are now seen to both be serious competitors to the QMLE, with the OCUE able to deliver more efficient individual point estimates than its QMLE counterpart.

Table 2 summarizes the results for the OCUE at various values of both κ and k, when $\max(i) = 2$, for values of θ_0 considered in the previous tables. Note that higher values of κ correspond to lower levels of skewness. For the QMLE, as skewness increases, so, too, does the dispersion of the parameter estimates. For the OCUE, in contrast, as skewness increases, the dispersion of the parameter estimates decreases, confirming the result of the Proposition in section 3.3. When making comparisons at a given level of κ, $\widehat{\alpha}(40)$ is less dispersed than $\widehat{\alpha}$ in all cases considered, while $\widehat{\alpha}(20)$ is less dispersed than $\widehat{\alpha}$ in nearly all

cases. The dispersion of $\widehat{\beta}(40)$, on the other hand, is only less dispersed than $\widehat{\beta}$ when $\kappa = 2$. For levels of κ higher than 2, $\widehat{\beta}(k)$ is generally more dispersed than $\widehat{\beta}$.

Table 3A summarizes additional simulation results for the OCUE and JCUE when $\max(i) = 2$, and $\kappa = 2$. In this case, values for θ_0 are selected that support a finite variance of Y_t but not a finite fourth moment. The degree to which θ_0 violates $E\left[(\beta + \alpha \epsilon_t^2)^2\right] < 1$ increases from $\theta_0^{(4)}$ to $\theta_0^{(6)}$. In these simulations, neither the QMLE nor the MM estimator is particularly apt at estimating $\widehat{\sigma}^2$. As before, the QMLE displays relatively less bias but is significantly more dispersed. As expected, JCUE is unbiased for $\widehat{\beta}(k)$ across the different specifications. Unexpected for the JCUE, however, is the finding that $\widehat{\alpha}(k)$ evidences non-neglible biases (much larger than those of the OCUE) for specifications $\theta_0^{(5)}$ and $\theta_0^{(6)}$. In addition, JCUE is far less efficient than the QMLE. Also unexpected is the finding that OCUE appears to be a serious competitor to the QMLE in the case of $\theta_0^{(4)}$. Equally a surprise is the finding that the OCUE maintains its previous tendency of providing relatively more efficient estimates than $\widehat{\alpha}$.

Finally, Table 3B replicates the conditions from Table 3A but for $\max(i) = 3$. In this case, a surprising result given the non-existence of the fourth moment is the finding that both $\widehat{\alpha}(20)$ and $\widehat{\alpha}(40)$ for the OCUE and JCUE are more efficient than $\widehat{\alpha}$ for $\theta_0^{(4)}$ and $\theta_0^{(5)}$. Equally surprising for $\theta_0^{(4)}$ is that the OCUE and JCUE remain serious competitors to the QMLE generally, since, in each case, $\widehat{\beta}(20)$ and $\widehat{\beta}(40)$ are quite comparable to $\widehat{\beta}$. In general, for the OCUE and JCUE, $\widehat{\beta}(20)$ and $\widehat{\beta}(40)$ tend to be less dispersed than $\widehat{\beta}$ across the three specifications. The biases of $\widehat{\beta}(20)$ and $\widehat{\beta}(40)$, however, increase significantly for $\theta_0^{(5)}$ and $\theta_0^{(6)}$. While contrary to what theory predicts, the results from Tables 3A and 3B are supported by simulation results in Kristensen and Linton (2006), where for data lacking a finite fourth moment, their autocorrelation-based estimator continued to display descent finite sample performance.

5. 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 model of (1) and (2) to $\{Y_{i,t}\}_{t=1}^{T}$.[14] 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 4 reports the estimation results for the JCUE, OCUE, and QMLE. For both the JCUE and OCUE, $L = 1$. For the JCUE, only the specification with $\max(i) = 3$ is considered. For the OCUE, both $\max(i) = 2$ and $\max(i) = 3$ are considered. Also for the OCUE, when $\max(i) = 2$, k is twice as large as when $\max(i) = 3$ so that the total number of moment conditions being used in each case is the same. Starting values for the JCUE and OCUE are the QMLE estimates.

The JCUE estimates are closer to the QMLE estimates than are the OCUE estimates. For the AUD series, the OCUE with $\max(i) = 3$ implies appreciably higher ARCH and appreciably lower GARCH effects than does the QMLE. For the JPY series, the OCUE generally implies much larger ARCH and much smaller GARCH estimates than the QMLE. Across both exchange rate series, however, differences in point estimates are accompanied by significantly higher standard errors than in the QMLE case. These higher standard errors are likely related to the near proximity of $\widehat{\alpha}+\widehat{\beta}$ to one.

6. 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. The moment conditions from these estimators are stated entirely in terms of covariates observable at time t, and while

---

[14]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.

they rely on skewness for identification, these estimators do not require treatment of the third and fourth conditional moments. Standard $\sqrt{T}$-asymptotics apply to these estimators given moment existence criteria no stronger than those required for comparable moment estimators discussed in the literature. These criteria can even be relaxed somewhat by nature of the fact that identification links to properties of the third as opposed to the fourth moment. These simple estimators (can) involve many (potentially weak) moments, the bias from which can be eliminated by using either a CUE with the optimal weighting matrix or what this paper terms the JCUE. Both the OCUE and JCUE can outperform QMLE in finite samples.

The identification result in this paper can be extended to a GARCH(1,1) model with a leverage effect. Suppose that $h_t = \omega_0 + \left(\alpha_0 + \alpha_0^- \times 1\left(Y_{t-1} < 0\right)\right) Y_{t-1}^2 + \beta_0 h_{t-1}$. Then (5) can be divided into the set of moment conditions $E\left[\widetilde{Y}_t^2 Y_{t-1}\right] = \left(\alpha_0 + \alpha_0^- \times P\left(Y_t < 0\right)\right) E\left[W_t Y_t\right]$, and $E\left[\widetilde{Y}_t^2 Y_{t-1} \times \left(1 - 1\left(Y_{t-1} < 0\right)\right)\right] = \alpha_0 \left(1 - P\left(Y_t < 0\right)\right) E\left[W_t Y_t\right]$, which can be used to identify a semi-parametric IV estimator of the semi-strong GARCH(1,1) model with a leverage effect. Such an estimator would be applicable to stock returns given the results of Hansen and Lunde (2005) and would expand the set of empirical applications to which simple IV estimators of the GARCH(1,1) model can apply.

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

Finally, since the estimators proposed in this paper are IV estimators with (potentially) many instruments, methods for selecting the number of instruments like those proposed by Donald, Imbens, and Newey (2008) are, therefore, of interest. Future research may look to relax the symmetry assumption in Donald, Imbens, and Newey (2008) and define criteria that are not (necessarily) dependent upon the variance-covariance matrix of the moment conditions.

Appendix

PROOF OF LEMMA 1: From (1) , (2), $E\left[W_t \mid F_{t-1}\right] = 0$, and the law of iterated expectations,

$$
\begin{aligned}
E\left[\widetilde{X}_t Y_{t-1}\right] &= E\left[\left(\widetilde{h}_t + W_t\right) Y_{t-1}\right] \\
&= E\left[\left(\alpha_0 \widetilde{X}_{t-1} + \beta_0 \widetilde{h}_{t-1}\right) Y_{t-1}\right] \\
&= \alpha_0 E\left[Y_{t-1}^3\right],
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
E\left[\widetilde{X}_t Y_{t-2}\right] &= E\left[\widetilde{h}_t Y_{t-2}\right] \\
&= \phi_0 E\left[\widetilde{X}_{t-1} Y_{t-2}\right] \\
&= \alpha_0 \phi_0 E\left[Y_{t-2}^3\right],
\end{aligned}
$$

and

$$
\begin{aligned}
E\left[\widetilde{X}_t Y_{t-3}\right] &= \phi_0 E\left[\widetilde{X}_{t-1} Y_{t-3}\right] \\
&= \phi_0^2 E\left[\widetilde{X}_{t-2} Y_{t-3}\right] \\
&= \alpha_0 \phi_0^2 E\left[Y_{t-3}^3\right].
\end{aligned}
$$

Given A2(i), these results imply that

$$
E\left[\widetilde{X}_t Y_{t-k}\right] = \alpha_0 \phi_0^{k-1} E\left[Y_t^3\right].
\tag{22}
$$

Solving (22) for $k = k + 1$ and comparing the result to $E\left[\widetilde{X}_t Y_{t-k}\right]$ produces (6).∎

LEMMA 2. *Given the model of (1) and (2), let Assumptions A1 and A3(i) hold. Then*

$$
E\left[\widetilde{h}_t^2\right] = \left(\frac{\alpha_0^2}{1 - \phi_0^2}\right) \lambda_0.
\tag{23}
$$

PROOF OF LEMMA 2: Given (4), $E\left[\widetilde{X}_t^2\right] = E\left[\widetilde{h}_t^2\right] + E\left[W_t^2\right]$. Given (3),

$$E\left[\widetilde{h}_t^2\right] = \phi_0^2 E\left[\widetilde{h}_{t-1}^2\right] + \alpha_0^2 \lambda_0. \tag{24}$$

Recursive substitution into (24) using (3) produces

$$E\left[\widetilde{h}_t^2\right] = \left(1 + \phi_0^2 + \cdots + \phi_0^{2(\tau-1)}\right)\alpha_0^2 \lambda_0 + \phi_0^{2\tau} E\left[\widetilde{h}_{t-\tau}^2\right]$$

for $\tau \geq 1$. It is well known that $\phi_0^{2\tau} \to 0$ as $\tau \to \infty$ if and only if $\phi_0 < 1$, which establishes (23). ∎

PROOF OF THE THEOREM: Given Lemma 2, Y_t^2 is covariance stationary. As a consequence, $\widehat{\sigma}^2 \xrightarrow{p} \sigma_0^2$ by a law of large numbers. Recursive substitution into (8) produces

$$\widetilde{X}_t = \sum_{i=0}^{\infty} \psi_i W_{t-i}, \tag{25}$$

where $\psi_0 = 1$ and $\psi_i = \alpha_0 \phi_0^{i-1}$ for $i = 1, 2, \ldots$. Given (25) and A3(i), $\widetilde{V}_{t,k}$ is an L^1 mixingale (see Andrews 1988 for a definition and Hamilton 1994 p. 192-193 for a proof). Given A3(ii), $T^{-1}\sum_t \widetilde{V}_{t,k} \xrightarrow{p} 0$ (see Theorem 1 of Andrews 1988). Similarly, $\widetilde{U}_{t,k}$ is an L^1 mixingale given (25) and either A2(ii) or A3(i) for which $T^{-1}\sum_t \widetilde{V}_{t,k} \xrightarrow{p} 0$ given either A2(iii) or A3(i). It then follows that (a) $\widehat{g}_{1,t}\left(\lambda, \widehat{\sigma}^2\right) \xrightarrow{p} (\alpha_0 - \alpha)\gamma_0$, (b) $\widehat{g}_{2,t}^{(k)}\left(\lambda, \widehat{\sigma}^2\right) \xrightarrow{p} \alpha_0(\phi_0 - \phi)\phi_0^{k-1}\gamma_0$, and (c) $\widehat{g}_{3,t}^{(k)}\left(\lambda, \widehat{\sigma}^2\right) \xrightarrow{p} (\phi_0 - \phi)\phi_0^{k-1}(\alpha_0\lambda_0 + \phi_0\eta_0)$, where $g_{2,t}^{(k)}\left(\lambda, \widehat{\sigma}^2\right)$ and $g_{3,t}^{(k)}\left(\lambda, \widehat{\sigma}^2\right)$ are the kth elements of $g_{2,t}\left(\lambda, \widehat{\sigma}^2\right)$ and $g_{3,t}\left(\lambda, \widehat{\sigma}^2\right)$, respectively, for $k = 2, \ldots, K$ and $\eta_0 = E\left[\widetilde{h}_t^2\right]$. 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 $\widehat{Q}\left(\lambda, \widehat{\sigma}^2\right) = \widehat{g}\left(\lambda, \widehat{\sigma}^2\right)' M_T \widehat{g}\left(\lambda, \widehat{\sigma}^2\right)$. Given (a)–(c) and continuity of multiplication, $\widehat{Q}\left(\lambda, \widehat{\sigma}^2\right) \xrightarrow{p} \overline{Q}\left(\lambda, \sigma_0^2\right)$. For $\max(i) = 2$, (a) and (b) 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 ϕ_0 is strictly positive. For $\max(i) = 3$, (a)–(c) establish the same result with parallel reasoning given that $\alpha_0\lambda_0 + \phi_0\eta_0$ is also strictly positive. $Q\left(\lambda, \sigma_0^2\right)$ is then uniquely minimized at $\lambda = \lambda_0$. Next, let $M_T = M_T\left(\widetilde{\lambda}, \widehat{\sigma}^2\right)$. Then the first order condition from (13) is $\widehat{S}_\lambda\left(\widehat{\lambda}, \widehat{\sigma}^2\right)' M_T \widehat{g}\left(\widehat{\lambda}, \widehat{\sigma}^2\right) = 0$.

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. Given A5, 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

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

where the second equality follows from $\widehat{S}_\lambda\left(\widehat{\lambda}, \widehat{\sigma}^2\right) \xrightarrow{p} S_\lambda\left(\lambda_0, \sigma_0^2\right)$ given that either $\widetilde{U}_{t,k}$ is an L^1 mixingale that is uniformly integrable if $\max(i) = 2$ or $\widetilde{V}_{t,k}$ is an L^1 mixingale that is uniformly integrable if $\max(i) = 3$ and Theorem 1 of Andrews (1988), and $\widehat{S}_{\sigma^2}\left(\lambda_0, \widehat{\sigma}^2\right) \xrightarrow{p} 0$ given that Y_t^2 is covariance stationary. From A4(i) and (25), $g_t\left(\lambda_0, \sigma_0^2\right)$ is an L^2 mixingale.[15] Given A6, $\sqrt{T}\widehat{g}\left(\lambda_0, \sigma_0^2\right) \xrightarrow{d} N\left(0, \ \Omega\left(\lambda_0, \sigma_0^2\right)\right)$ by Theorem 1 of De Jong (1997). The conclusion then follows from the Slutzky Theorem. ■

PROOF OF COROLLARY 1: A4(ii) grants h_t to be β-mixing with decreasing mixing coefficients (see Corollary 6 of Carrasco and Chen 2002). Theorem 17.0.1 of Meyn and Tweedie (1993) then establishes $\sqrt{T}\widehat{g}\left(\lambda_0, \sigma_0^2\right) \xrightarrow{d} N\left(0, \ \Omega\left(\lambda_0, \sigma_0^2\right)\right)$. The rest follows from the proof of the Theorem. ■

PROOF OF THE PROPOSITION: Given the results for derivatives of inverse matrices,

$$
\frac{\partial V_{GMM}}{\partial \gamma} = \frac{1}{\gamma_0^2}\left\{-\frac{2}{\gamma_0}\left(\Phi_0' M_0 \Phi_0\right)^{-1} + \left(\Phi_0' M_0 \Phi_0\right)^{-1} \Phi_0' M_0 \frac{\partial \Omega\left(\theta_0\right)}{\partial \gamma} M_0 \Phi_0 \left(\Phi_0' M_0 \Phi_0\right)^{-1}\right\}.
$$

---

[15] The proof of this result follows closely with those of $\widetilde{U}_{t,k}$ and $\widetilde{V}_{t,k}$ being L^1 mixingales and is available upon request.

Consider first the case where $\gamma_0 > 0$, and let $x = M_0\Phi_0 (\Phi_0'M_0\Phi_0)^{-1}$. Then

$$-\frac{2}{\gamma_0}(\Phi_0'M_0\Phi_0)^{-1} + (\Phi_0'M_0\Phi_0)^{-1}\Phi_0'M_0\frac{\partial\Omega(\theta_0)}{\gamma}M_0\Phi_0(\Phi_0'M_0\Phi_0)^{-1} \leq$$

$$-\frac{2}{\gamma_0}(\Phi_0'M_0\Phi_0)^{-1} + r(\Phi_0'M_0\Phi_0)^{-1}\Phi_0'M_0\Omega(\theta_0)M_0\Phi_0(\Phi_0'M_0\Phi_0)^{-1} =$$

$$\left(r - \frac{2}{\gamma_0}\right)(\Phi_0'M_0\Phi_0)^{-1} < 0.$$

Next, consider the case where $\gamma_0 < 0$. Then

$$-\frac{2}{\gamma_0}(\Phi_0'M_0\Phi_0)^{-1} + (\Phi_0'M_0\Phi_0)^{-1}\Phi_0'M_0\frac{\partial\Omega(\theta_0)}{\gamma}M_0\Phi_0(\Phi_0'M_0\Phi_0)^{-1} \geq$$

$$-\frac{2}{\gamma_0}(\Phi_0'M_0\Phi_0)^{-1} - r(\Phi_0'M_0\Phi_0)^{-1}\Phi_0'M_0\Omega(\theta_0)M_0\Phi_0(\Phi_0'M_0\Phi_0)^{-1} =$$

$$-\left(r + \frac{2}{\gamma_0}\right)(\Phi_0'M_0\Phi_0)^{-1} > 0.$$

■

PROOF OF LEMMA 3: From the definition of $\widehat{\rho}_{t,s}^{(m,n)}(\theta)$,

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

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,

$$\left\|T^{-1}\sum_t a_{t,s}\left(\widehat{\theta}\right) - a_{t,s}(\theta_0)\right\| \leq T^{-1}\sum_t \left\|a_{t,s}\left(\widehat{\theta}\right) - a_{t,s}(\theta_0)\right\| \leq T^{-1}\sum_t \Delta_{t,s}(\theta) \xrightarrow{p} E\left[\Delta_{t,s}(\theta)\right]$$

establishing the result.■

PROOF OF COROLLARY 2:

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

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

$$=\; T^{-1}\sum_{s=1}^{T} A_s\left(\lambda,\widehat{\sigma}^2\right) g_s\left(\lambda,\widehat{\sigma}^2\right),$$

where

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

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

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] Baillie, R.T., and H. Chung, 2001, Estimation of GARCH models from the autocorrelations of the squares of a process, Journal of Time Series Analysis, 22, 631-650.

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

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

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

[7] 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.

[8] Carrasco, M. and X. Chen, 2002, Mixing and moment properties of various GARCH and stochastic volatility models, Econometric Theory, 18, 17-39.

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

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

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

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

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

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

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

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

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

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

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

[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] He, C. and T. Teräsvirta, 1999, Properties of moments of a family of GARCH processes, Journal of Econometrics, 92, 173-192.

[24] Kristensen, D. and O. Linton, 2006, A closed-form estimator for the GARCH(1,1)-model, Econometric Theory, 22, 323-327.

[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] Meddahi, N. and E. Renault, 1998, Quadratic M-estimators for ARCH-type processes, working paper CIRANO 98s-29.

[29] 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.

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

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

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

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

[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.

Para.	Est.	$\theta_0^{(1)}= (1.0,\ 0.15,\ 0.75)$ Med Bias	Dec Rge	SD	MDAE	$\theta_0^{(2)}= (1.0,\ 0.10,\ 0.85)$ Med Bias	Dec Rge	SD	MDAE	$\theta_0^{(3)}= (1.0,\ 0.05,\ 0.94)$ Med Bias	Dec Rge	SD	MDAE
σ^2	QMLE	-0.003	0.244	0.095	0.065	-0.004	0.291	0.113	0.080	-0.028	0.602	0.352	0.163
	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
α	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
$\alpha(10)$	OCUE	-0.009	0.053	0.028	0.014	-0.006	0.040	0.018	0.010	-0.004	0.034	0.018	0.008
	JCUE	-0.017	0.101	0.045	0.030	-0.008	0.077	0.034	0.020	0.002	0.064	0.028	0.014
$\alpha(20)$	OCUE	-0.006	0.040	0.021	0.010	-0.003	0.022	0.014	0.005	-0.002	0.019	0.011	0.004
	JCUE	-0.016	0.091	0.042	0.028	-0.009	0.067	0.031	0.020	0.000	0.048	0.022	0.011
$\alpha(40)$	OCUE	-0.003	0.035	0.019	0.007	-0.001	0.017	0.014	0.003	0.000	0.007	0.008	0.001
	JCUE					-0.003	0.071	0.036	0.016				
β	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
$\beta(10)$	OCUE	-0.013	0.109	0.056	0.029	-0.012	0.107	0.050	0.029	-0.021	0.144	0.079	0.034
	JCUE	0.009	0.229	0.096	0.058	0.005	0.203	0.099	0.049	-0.038	0.316	0.187	0.050
$\beta(20)$	OCUE	-0.015	0.091	0.042	0.024	-0.009	0.067	0.031	0.018	-0.010	0.077	0.045	0.019
	JCUE	0.010	0.173	0.076	0.043	0.009	0.137	0.061	0.036	-0.008	0.144	0.154	0.031
$\beta(40)$	OCUE	-0.015	0.079	0.039	0.022	-0.010	0.055	0.028	0.014	-0.004	0.038	0.020	0.010
	JCUE					0.005	0.126	0.065	0.030				

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector θ_0 = $(\sigma_0^2,\ \alpha_0,\ \beta_0)'$, and $\kappa = 2$. $\widehat{\alpha}(k)$ and $\widehat{\beta}(k)$ are the α and β estimates, respectively, based on k lags. QMLE is the quasi-maximum likelihood estimator. MM is the method of moments estimator. OCUE and JCUE are the optimal and jackknife continuous updating estimator, respectively, with $\max(i) = 2$, $k = 10, 20, 40$, and $L = 1$. 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.

TABLE 1B

Para.	Est.	$\theta_0^{(1)}= (1.0,\ 0.15,\ 0.75)$				$\theta_0^{(2)}= (1.0,\ 0.10,\ 0.85)$				$\theta_0^{(3)}= (1.0,\ 0.05,\ 0.94)$			
		Med Bias	Dec Rge	SD	MDAE	Med Bias	Dec Rge	SD	MDAE	Med Bias	Dec Rge	SD	MDAE
α	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
$\alpha\,(10)$	OCUE	-0.009	0.041	0.021	0.011	-0.006	0.033	0.022	0.009	-0.003	0.026	0.013	0.005
	JCUE	-0.003	0.044	0.023	0.007	-0.001	0.023	0.014	0.003	0.000	0.009	0.006	0.001
$\alpha\,(20)$	OCUE	-0.006	0.032	0.017	0.009	-0.003	0.015	0.009	0.004	-0.001	0.011	0.009	0.002
	JCUE	-0.001	0.029	0.027	0.006	0.000	0.014	0.011	0.002	0.000	0.004	0.005	0.001
$\alpha\,(40)$	OCUE	-0.002	0.037	0.024	0.007	-0.001	0.017	0.013	0.003	0.000	0.002	0.004	0.001
	JCUE	-0.001	0.029	0.014	0.005	0.000	0.013	0.011	0.002	0.000	0.003	0.005	0.000
β	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
$\beta\,(10)$	OCUE	-0.018	0.091	0.042	0.027	-0.011	0.077	0.035	0.021	-0.006	0.069	0.027	0.018
	JCUE	0.005	0.108	0.053	0.025	0.004	0.088	0.041	0.022	0.000	0.077	0.038	0.017
$\beta\,(20)$	OCUE	-0.026	0.078	0.036	0.028	-0.014	0.049	0.022	0.016	-0.006	0.030	0.013	0.009
	JCUE	0.000	0.104	0.058	0.022	0.000	0.063	0.036	0.015	0.000	0.035	0.022	0.009
$\beta\,(40)$	OCUE	-0.028	0.079	0.038	0.029	-0.018	0.052	0.023	0.018	-0.004	0.020	0.008	0.006
	JCUE	0.000	0.089	0.040	0.020	0.000	0.059	0.032	0.014	-0.001	0.025	0.029	0.006

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector θ_0 = $(\sigma_0^2,\ \alpha_0,\ \beta_0)'$, and $\kappa = 2$. $\widehat{\alpha}\,(k)$ and $\widehat{\beta}\,(k)$ are the α and β estimates, respectively, based on k lags. QMLE is the quasi-maximum likelihood estimator. OCUE and JCUE are the optimal and jackknife continuous updating estimator, respectively, with $\max(i) = 3$, $k = 10, 20, 40$, and $L = 1$. 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.

TABLE 2

Para.	Est.	$\theta_0^{(1)}= (1.0,\ 0.15,\ 0.75)$ Med Bias	Dec Rge	SD	MDAE	$\theta_0^{(2)}= (1.0,\ 0.10,\ 0.85)$ Med Bias	Dec Rge	SD	MDAE	$\theta_0^{(3)}= (1.0,\ 0.05,\ 0.94)$ Med Bias	Dec Rge	SD	MDAE
α	QMLE (5)	-0.001	0.044	0.017	0.011	0.000	0.033	0.013	0.008	0.000	0.018	0.007	0.005
	QMLE (4)	-0.001	0.046	0.018	0.011	0.000	0.034	0.013	0.008	0.000	0.019	0.007	0.005
	QMLE (3)	-0.001	0.051	0.019	0.012	0.000	0.036	0.014	0.009	0.000	0.020	0.008	0.005
	QMLE (2)	-0.001	0.054	0.021	0.013	0.000	0.039	0.015	0.010	0.000	0.022	0.008	0.005
$\alpha\,(20)$	OCUE (5)	-0.003	0.039	0.025	0.009	-0.001	0.024	0.018	0.005	0.000	0.019	0.013	0.004
	OCUE (4)	-0.004	0.038	0.025	0.009	-0.002	0.023	0.017	0.005	0.000	0.018	0.014	0.004
	OCUE (3)	-0.005	0.035	0.023	0.009	-0.002	0.023	0.017	0.005	-0.001	0.020	0.012	0.004
	OCUE (2)	-0.006	0.040	0.021	0.010	-0.003	0.022	0.014	0.005	-0.002	0.019	0.011	0.004
$\alpha\,(40)$	OCUE (5)	0.001	0.032	0.024	0.007	0.001	0.014	0.018	0.003	0.000	0.008	0.008	0.001
	OCUE (4)	0.000	0.031	0.020	0.007	0.000	0.014	0.014	0.003	0.000	0.007	0.009	0.001
	OCUE (3)	-0.001	0.031	0.018	0.007	0.000	0.013	0.014	0.003	0.000	0.007	0.008	0.001
	OCUE (2)	-0.003	0.035	0.019	0.007	-0.001	0.017	0.014	0.003	0.000	0.007	0.008	0.001
β	QMLE (5)	0.000	0.071	0.028	0.017	0.000	0.048	0.019	0.011	-0.002	0.021	0.008	0.006
	QMLE (4)	0.000	0.073	0.029	0.018	0.000	0.049	0.019	0.012	-0.001	0.022	0.009	0.006
	QMLE (3)	0.000	0.077	0.030	0.020	0.000	0.051	0.020	0.012	-0.001	0.022	0.009	0.006
	QMLE (2)	0.000	0.081	0.033	0.020	0.000	0.056	0.022	0.013	-0.001	0.023	0.009	0.006
$\beta\,(20)$	OCUE (5)	-0.010	0.107	0.049	0.025	-0.006	0.077	0.047	0.020	-0.007	0.097	0.081	0.019
	OCUE (4)	-0.010	0.098	0.051	0.025	-0.007	0.078	0.035	0.018	-0.007	0.067	0.067	0.019
	OCUE (3)	-0.012	0.089	0.046	0.023	-0.007	0.070	0.037	0.018	-0.008	0.085	0.052	0.020
	OCUE (2)	-0.015	0.091	0.042	0.024	-0.009	0.067	0.031	0.018	-0.010	0.077	0.045	0.019
$\beta\,(40)$	OCUE (5)	-0.009	0.098	0.050	0.021	-0.006	0.064	0.036	0.014	-0.002	0.048	0.042	0.010
	OCUE (4)	-0.010	0.085	0.045	0.020	-0.006	0.058	0.030	0.013	-0.002	0.043	0.041	0.010
	OCUE (3)	-0.012	0.084	0.039	0.021	-0.008	0.056	0.030	0.013	-0.003	0.041	0.022	0.010
	OCUE (2)	-0.015	0.079	0.039	0.022	-0.010	0.055	0.028	0.014	-0.004	0.038	0.020	0.010

Notes: See Table 1A. QMLE (κ) and OCUE (κ) refer to the QMLE and CUE estimator, respectively, applied to data where $\kappa = 2, \ldots, 5$.

Para.	Est.	$\theta_0^{(4)} = (1.0,\ 0.10,\ 0.88)$				$\theta_0^{(5)} = (1.0,\ 0.20,\ 0.78)$				$\theta_0^{(6)} = (1.0,\ 0.30,\ 0.68)$			
		Med Bias	Dec Rge	SD	MDAE	Med Bias	Dec Rge	SD	MDAE	Med Bias	Dec Rge	SD	MDAE
σ^2	QMLE	-0.017	0.718	0.352	0.171	-0.082	1.991	6.941	0.338	-0.223	3.729	4.517	0.464
	MM	-0.090	0.579	0.333	0.157	-0.293	0.795	1.218	0.337	-0.463	0.768	2.104	0.485
α	QMLE	0.000	0.035	0.014	0.009	0.000	0.056	0.022	0.014	-0.003	0.076	0.029	0.019
$\alpha\,(20)$	OCUE	-0.003	0.021	0.013	0.005	-0.009	0.028	0.022	0.011	-0.020	0.036	0.036	0.021
	JCUE	-0.009	0.071	0.034	0.019	-0.034	0.133	0.066	0.046	-0.058	0.188	0.099	0.073
$\alpha\,(40)$	OCUE	-0.001	0.009	0.018	0.002	-0.006	0.014	0.017	0.007	-0.018	0.024	0.021	0.018
	JCUE	-0.004	0.064	0.031	0.015								
β	QMLE	-0.002	0.037	0.015	0.010	0.000	0.052	0.020	0.012	0.001	0.063	0.025	0.015
$\beta\,(20)$	OCUE	-0.014	0.065	0.030	0.019	-0.026	0.070	0.050	0.029	-0.040	0.069	0.049	0.042
	JCUE	0.006	0.147	0.083	0.038	0.007	0.210	0.108	0.052	0.004	0.286	0.143	0.067
$\beta\,(40)$	OCUE	-0.009	0.038	0.025	0.011	-0.023	0.047	0.026	0.023	-0.039	0.053	0.032	0.039
	JCUE	0.001	0.119	0.055	0.027								

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector $\theta_0 = (\sigma_0^2,\ \alpha_0,\ \beta_0)'$, and $\kappa = 2$. $\widehat{\alpha}(k)$ and $\widehat{\beta}(k)$ are the α and β estimates, respectively, based on k lags. QMLE is the quasi-maximum likelihood estimator. MM is the method of moments estimator. OCUE and JCUE are the optimal and jackknife continuous updating estimator, respectively, with $\max(i) = 2$, $k = 20, 40$, and $L = 1$. 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.

TABLE 3B

Para.	Est.	$\theta_0^{(4)}=(1.0,\ 0.10,\ 0.88)$				$\theta_0^{(5)}=(1.0,\ 0.20,\ 0.78)$				$\theta_0^{(6)}=(1.0,\ 0.30,\ 0.68)$			
		Med Bias	Dec Rge	SD	MDAE	Med Bias	Dec Rge	SD	MDAE	Med Bias	Dec Rge	SD	MDAE
α	QMLE	0.000	0.035	0.014	0.009	0.000	0.056	0.022	0.014	-0.003	0.076	0.029	0.019
$\alpha\,(20)$	OCUE	-0.002	0.008	0.008	0.003	-0.008	0.014	0.017	0.008	-0.020	0.024	0.016	0.021
	JCUE	-0.001	0.007	0.008	0.001	-0.006	0.022	0.017	0.006	-0.018	0.049	0.030	0.018
$\alpha\,(40)$	OCUE	-0.001	0.004	0.005	0.001	-0.007	0.011	0.016	0.007	-0.020	0.022	0.012	0.020
	JCUE												
β	QMLE	-0.002	0.037	0.015	0.010	0.000	0.052	0.020	0.012	0.001	0.063	0.025	0.015
$\beta\,(20)$	OCUE	-0.011	0.036	0.016	0.012	-0.026	0.042	0.022	0.026	-0.042	0.051	0.025	0.042
	JCUE	-0.004	0.039	0.020	0.009	-0.018	0.048	0.043	0.019	-0.031	0.085	0.057	0.032
$\beta\,(40)$	OCUE	-0.011	0.027	0.012	0.012	-0.023	0.047	0.026	0.023	-0.045	0.053	0.023	0.045
	JCUE												

Notes: Simulations are conducted using 5,000 observations across 500 trials. The true parameter vector θ_0 = $(\sigma_0^2,\ \alpha_0,\ \beta_0)'$, and $\kappa = 2$. $\widehat{\alpha}\,(k)$ and $\widehat{\beta}\,(k)$ are the α and β estimates, respectively, based on k lags. QMLE is the quasi-maximum likelihood estimator. OCUE and JCUE are the optimal and jackknife continuous updating estimator, respectively, with $\max(i) = 3$, $k = 20, 40$, and $L = 1$. 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.

TABLE 4

Currency	Para.	JCUE	OCUE		QMLE
	$\max(i)$	3	2	3	
	k	40	80	40	
	$\widehat{\sigma}^2$	0.5579	0.5579	0.5579	0.4957
	$\widehat{\alpha}$	0.0511	0.0706	0.1192	0.0532
AUD			(0.0979)	(0.0608)	(0.0088)
	$\widehat{\beta}$	0.9215	0.9180	0.8772	0.9382
			(0.0276)	(0.0231)	(0.0101)
	$\widehat{\alpha}+\widehat{\beta}$	0.9726	0.9887	0.9964	0.9914
	$\max(i)$	3	2	3	
	k	40	80	40	
	$\widehat{\sigma}^2$	0.4963	0.4963	0.4963	0.5057
	$\widehat{\alpha}$	0.0539	0.1984	0.1192	0.0486
JPY			(0.0517)	(0.0608)	(0.0095)
	$\widehat{\beta}$	0.9099	0.7523	0.8772	0.9361
			(0.0288)	(0.0231)	(0.0123)
	$\widehat{\alpha}+\widehat{\beta}$	0.9638	0.9507	0.9964	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. JCUE and OCUE are the jackknife and optimal continuous updating estimator, respectively, where $L = 1$. k is the number of lags used in the given estimator (if applicable). $\max(i)$ specifies whether the given estimator is based on properties of the third moment only ($\max(i) = 2$) or also on properties of the fourth ($\max(i) = 3$). $\widehat{\sigma}^2$ is the unconditional variance estimate for the given spot return. $\widehat{\alpha}$ is the ARCH estimate, and $\widehat{\beta}$ is the GARCH estimate.