Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

When Simplicity Offers a Benefit, Not a Cost: Closed-Form Estimation of the GARCH(1,1) Model that Enhances the Efficiency of Quasi-Maximum Likelihood

Todd Prono

2019-030

Please cite this paper as:
Prono, Todd (2019). "When Simplicity Offers a Benefit, Not a Cost: Closed-Form Estimation of the GARCH(1,1) Model that Enhances the Efficiency of Quasi-Maximum Likelihood," Finance and Economics Discussion Series 2019-030. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2019.030.

When Simplicity Offers a Benefit, Not a Cost: Closed-Form Estimation of the GARCH(1,1) Model that Enhances the Efficiency of Quasi-Maximum Likelihood[1]

Todd Prono[2]

This Version: February 2019

Abstract

Simple, multi-step estimators are developed for the popular GARCH$(1,1)$ model, where these estimators are either available entirely in closed form or dependent upon a preliminary estimate from, for example, quasi-maximum likelihood. Identification sources to asymmetry in the model's innovations, casting skewness as an instrument in a linear, two-stage least squares estimator. Properties of regular variation coupled with point process theory establish the distributional limits of these estimators as stable, though highly non-Gaussian, with slow convergence rates relative to the $\sqrt{n}$-case. Moment existence criteria necessary for these results are consistent with the heavy-tailed features of many financial returns. In light-tailed cases that support asymptotic normality for these simple estimators, conditions are discovered where the simple estimators can enhance the asymptotic efficiency of quasi-maximum likelihood estimation. In small samples, extensive Monte Carlo experiments reveal these efficiency enhancements to be available for (very) heavy tailed cases. Consequently, the proposed simple estimators are members of the class of multi-step estimators aimed at improving the efficiency of the quasi-maximum likelihood estimator.

Keywords: GARCH models, closed form estimation, heavy tails, instrumental variables, regular variation. JEL codes: C13, C22, C58.

[1] The views expressed in this paper are those of the author and do not necessarily reflect those of the Federal Reserve Board.

[2] Federal Reserve Board. (202) 973-6955, todd.a.prono@frb.gov.

1.1. Introduction

The linear GARCH$(1, 1)$ model of Bollerslev (1986) is a workhorse of conditional volatility forecasting in financial economics, its applications spanning portfolio formation, derivative pricing, and risk management. Despite its parsimony, this model is shown to outperform (in terms of out-of-sample forecasting) more complicated alternative specifications (see; e.g., Hansen and Lunde, 2005). The most common estimator for this model is quasi-maximum likelihood (QML), which is based on a Gaussian likelihood function. Pioneering works by Lee and Hansen (1994), and Lumsdaine (1996) establish the QMLE as consistent and asymptotically normal under a variety of (unknown) densities for the model's innovations. Berkes, Horváth, and Kokoszka (2003) and Francq and Zakoïan (2004) extend this result to the GARCH(p, q) model under milder conditions, including a well-defined fourth moment of the model's innovations. Hall and Yao (2003) establish the distributional limit of the QMLE in cases when the fourth moment of these innovations is ill-defined.

The first aim of this paper is to temporarily part ways with QMLE to propose simple, moment-based alternatives for GARCH$(1, 1)$ model estimation. The definition of a simple estimator heralds from Lewbel (2004) and is subsequently applied in Dong and Lewbel (2015).

DEFINITION. *A simple estimator closely resembles (or consists of steps that each resemble) estimators that are already in common use and involves few or no numerical searches or numerical optimizations.*

Consistent with this definition, the estimators developed herein are available in closed form and, therefore, comparable to those proposed by Kristensen and Linton (2006), although under milder conditions. Collectively, these simple estimators are instrumental variables (IV) estimators that apply separately to the model's ARCH and GARCH parameters and are implemented via applications of linear, two-stage least squares. These simple estimators are shown to be strongly consistent and to weakly converge to stable, though highly non-Gaussian, limits in empirically-relevant cases. Specifically, these results require (slightly stronger than) third moment existence for the raw return sequence being modeled and a well-defined ith moment of the GARCH innovations, where $i \in (3,\ 6)$. Convergence rates for these simple estimators tend to be (much) slower than $\sqrt{n}$ and depend on the tail-thickness of the raw returns being modeled.

Simple estimators tend to be associated with inefficient estimators. Indeed, relative to the case of maximum likelihood estimation (MLE), which relies upon knowledge of the true innovation density, this association is (many times) justified. QMLE, on the other hand, while still consistent, can also be considerably inefficient in cases where the true (and unknown) innovation density deviates from normality. While one possible fix for this inefficiency loss is to specify a heavier-tailed density for the model's innovations, like the student-t (see; e.g., Baillie and Bollerslev, 1989), consistency is lost if the true innovation density happens to reside outside of the student-t family. Consequently, a literature on GARCH estimation has emerged aimed at defining multi-step estimators that improve upon the QMLE, as implemented in a preliminary first step, but also maintain robustness in terms of consistency (see; e.g., Drost and Klaassen, 1997, Francq and Zakoïan, 2011, Fan, Qi and Xiu, 2014, and Preminger and Storti, 2017).[3] Collectively, by better targeting the scale of the true (and unknown) innovation density, these estimators offer efficiency enhancements over QMLE.

Identification of the simple estimators proposed herein relies on non-zero skewness in the raw returns being modeled. Essentially, skewness is the instrument upon which these estimators are based. In a linear GARCH context, skewness in the raw returns necessarily sources to skewness in the true (and unknown) innovation density. If that skewness represents a prominent feature of the innovation density, explicitly targeting it may very-well provide efficiency gains, just as (better) targeting scale does. Additionally, the simple estimators proposed herein are also multi-step estimators reliant upon preliminary estimates from a first step. The second aim of this paper, then, is to investigate the advantages of sourcing the requisite preliminary estimates for the proposed simple estimators to QMLE. From that investigation, it is found that for raw return processes characterized by no more than a well-defined third moment, the proposed simple estimators are asymptotically more efficient with QMLE-based preliminary estimates than closed-form, moments-based alternatives with slower convergence rates. In thin-tailed cases that support asymptotic normality for the simple estimators, conditions are found under which the simple estimators are actually asymptotically more efficient than QMLE.[4] Lastly (and, perhaps, most surprisingly) it is also found that in small samples, the simple estimators can enhance the efficiency of QMLE even when the model's innovations are (very) heavy tailed. Explaining this enhancement in heavy-tailed cases is the same factor at work under asymptotic normality; namely, skewness in the model's innovations. Consequently, the simple estimators proposed herein are, in fact, comparable to the aforementioned class of multi-step estimators aimed at enhancing the efficiency of QMLE; with the added benefit of not requiring any numerical optimization in their final step.

[3] Specifically, Drost and Klaassen (1997) investgate the possibility for adaptive GARCH estimation; that is, a semiparametric estimator matching the efficiency of MLE. Recognizing adaptive estimation to be, generally, infeasible, the more recent cited works look to improve upon the efficiency of QMLE, thereby narrowing (but not eliminating) the efficiency loss relative to MLE.

[4] Monte Carlo studies find that these conditions can be satisfied for (at least) certain regions of the parameter space for the GARCH$(1, 1)$ model.

1.2 Background and Motivation

For the linear GARCH$(1,1)$ model of

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\epsilon_t \sim i.i.d.\ D\left(0,\ 1\right)$ and D is unknown, it is well known that

$$Y_t^2 = \omega + \phi Y_{t-1}^2 - \beta W_{t-1} + W_t, \qquad \phi = \alpha + \beta, \qquad W_t = \sigma_t^2 \left(\epsilon_t^2 - 1\right), \tag{1}$$

where $\{W_t\}$ is a Martingale difference sequence (MDS); that is, the GARCH$(1,1)$ model implies an ARMA$(1,1)$ model for the second-order sequence $\left\{Y_t^2\right\}$. When thinking about simple estimators for this second-order ARMA$(1,1)$ model, two sets of possible instruments spring to mind:

$$\mathbf{Z}_{t-1}^{(i)} = \left(\ Y_{t-1}^i, \quad \ldots, \quad Y_{t-h}^i \ \right), \qquad i = 1, 2.$$

The case where $i = 2$ covers the estimators proposed by Kristensen and Linton (2006) and Giraitis and Robinson (2000). This paper investigates the (up until this point) overlooked case of $i = 1$. In order for $\mathbf{Z}_{t-1}^{(1)}$ to serve as a valid instrument for Y_{t-1}^2 requires both that $E\left(Y_t^3\right) < \infty$ and that $E\left(Y_t^3\right) \neq 0$.

STYLIZED FACT. *Many financial returns seem to be characterized by heavy-tailed processes for which the fourth moment is not well-defined...(see Figure 1 and; e.g., Hill and Renault, 2012).*

Plotted in Panel A of Figure 1 are Hill (1975) tail index estimates together with 95% confidence bands (the latter coming from Hill, 2010, Theorem 4) for daily S&P 500 Index log returns. Recalling that a tail index $\kappa > 0$ for a regularly varying random variable is a moment supremum (i.e., if Y_t is regularly varying, then $E\left|Y_t\right|^p < \infty$ if and only if $p < \kappa$), empirical evidence does not (strongly) support well-defined fourth moments for these returns. In fact, in many instances, even the upper bounds of the 95% Confidence intervals fall below 4. This lack of support for $E\left(Y_t^4\right) < \infty$ is problematic if $\mathbf{Z}_{t-1}^{(2)}$ is to serve as instruments for (1) since identification and, hence, consistency hinges upon this criterion.

STYLIZED FACT. *"There is now good evidence that on short time scales, and using long time series, the tail index for stocks is around 3 on several markets (U.S., Japan, Germany)"...Bouchard and Potters (2003)*

Panel A of Figure 1 is much more supportive of the claim that $E\left(Y_t^3\right) < \infty$. In addition, the skewness statistic for the returns is -0.26, which is highly significant against a null of normality, given the sample size.[5] Table 1 illustrates additional instances where (very high frequency) financial returns evidence very significant skewness statistics that are also quite large in absolute terms. Collectively then, empirical evidence seems to support $\mathbf{Z}_{t-1}^{(1)}$ as a viable set of instruments for (1).

The empirical evidence from the previous paragraph also illustrates the impracticality of simple estimators for (1) based even on $\mathbf{Z}_{t-1}^{(1)}$ being asymptotically normal: the well-defined, higher moments necessary for such a result simply aren't supported empirically. Fortunately, these simple estimators can be shown to weakly converge in distribution to a heavy-tailed mixture of stable random variables using results from Davis and Hsing (1995) that are also applied in, for example, Davis and Mikosch (1998) and Mikosch and Stărică (2000).[6] Applicability of these results depends on $\{Y_t\}$ being regularly varying in the case where ϵ_t is drawn for a skewed distribution. In addition, as mentioned in the introduction, a second requirement is for $E\left|\epsilon_t\right|^i < \infty$, where $i \in (3,\ 6)$. From Panel B of Figure 1 (which depicts tail index estimates for the innovations to a GARCH$(1,1)$ model applied to daily S&P 500 Index log returns), this second requirement also enjoys empirical support.[7]

2.1. Simple Estimation of the GARCH(1,1) Model

Consider the model

$$Y_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim i.i.d.\ D\left(0,\ 1\right), \tag{2}$$

where

$$\begin{aligned} \sigma_t^2 &= \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \\ &= \omega_0 + \sigma_{t-1}^2 \left(\alpha_0 \epsilon_{t-1}^2 + \beta_0\right) \\ &= \omega_0 + \sigma_{t-1}^2 A_t \end{aligned} \tag{3}$$

Here, ω_0 denotes the true value, ω any one of a set of possible values, $\widehat{\omega}$ an estimate, and parallel definitions hold for all other parameter values. The model of (2) and (3) describes a strong GARCH process (see Drost and Nijman, 1993). Consistency of the simple estimators studied in this paper holds if $\{\epsilon_t\}$ is, instead,

[5] In fact, the skewness statistic for daily S&P 500 Index log returns can be as high as -1.02, depending on the length of the data sample used.

[6] This mixture of stable random variables has an ill-defined variance.

[7] Specifically, the linear GARCH$(1,1)$ model applied also includes an AR(1) component in the conditional mean.

a weakly dependent, MDS (see; e.g., Prono, 2014). The distributional limits for these simple estimators, however, require $\{\epsilon_t\}$ to be $i.i.d.$ (Mikosch and Straumann, 2002, 2006, and Vaynman and Beare, 2014, impose this same requirement).

Mikosch and Stărică (2000) establish (3) as a stochastic recurrence equation (SRE). Most linear GARCH processes are afforded this characterization (see; e.g., Basrak, Davis and Mikosch, 2002), which is important for establishing them as regularly varying. For instance, conditional on (3) being a SRE, both $\{\sigma_t^2\}$ and $\{Y_t\}$ are regularly varying sequences (see Lemmas 3 and 5, respectively, in the Supplemental Appendix). Specifically, for $0 \leq h < \infty$, consider $\mathbf{Y}_t = \begin{pmatrix} Y_t, & \ldots, & Y_{t+h} \end{pmatrix}$, or $\mathbf{Y} = \mathbf{Y}_0 = \begin{pmatrix} Y_0, & \ldots, & Y_h \end{pmatrix}$ for short. $\mathbf{Y}$ is regularly varying in $\mathbb{R}^{h+1}$ with tail index κ_0, meaning there exists a sequence of constants $\{a_n\}$ such that

$$nP\left(|\mathbf{Y}| > a_n\right) \longrightarrow 1, \qquad n \to \infty,$$

where $|\mathbf{Y}| = \max_{m=0,\ldots,h} |Y_m|$,

$$a_n = n^{1/\kappa_0} L\left(n\right),$$

and $L\left(\cdot\right)$ is slowly-varying at ∞. Mikosch and Stărică (2000, Theorem 2.3) demonstrate $\mathbf{Y}$ to be regularly varying in the case where D is symmetric. Lemma 5 in the Supplemental Appendix is a more general result that establishes $\mathbf{Y}$ as regularly varying regardless of whether D is symmetric or skewed by combining certain elements from the proofs of Mikosch and Stărică (2000, Theorem 2.3) and Basrak et al. (2002, Corollary 3.5(B)), respectively (see Remark 6 in the Supplemental Appendix). This generalization is important because a necessary condition for identifying the simple estimators in this paper is $E\left(\epsilon_t^3\right) \neq 0$, and given (2) and (3), this condition implies $E\left(Y_t^3\right) \neq 0$.

ASSUMPTION A1: *The distribution D has an unbounded support. In addition, for some $\delta > 0$, $E\left|\epsilon_t\right|^{i+\delta} < \infty$, where $3 \leq i < \overline{h} < \infty$, while $E\left|\epsilon_t\right|^{\overline{h}} = \infty$, and for $j \leq i$, $E\left|\epsilon_t\right|^{j} = c_j$.*

Under A1, $\{\epsilon_t\}$ is lighter-tailed than $\{\sigma_t\}$. This distinction is important because it limits the heavy-tailed features of $\{Y_t\}$ to stem from $\{\sigma_t\}$, which, in turn, enables $\{Y_t\}$ to be established as regularly varying.

ASSUMPTION A2: *The parameter space is given by*

$$\Theta = \left\{\boldsymbol{\theta} = \begin{pmatrix} \omega, & \alpha, & \beta \end{pmatrix} \in \mathbb{R}^3 \mid \omega \geq \underline{\omega},\ \alpha > 0,\ \beta \geq 0\right\},$$

for some $\underline{\omega} > 0$.

The strictly positive lower bound on ω heralds from Kristensen and Rahbek (2005). Notice as well that Θ is non-compact.

ASSUMPTION A3: $E\left(\epsilon_t^3\right) = c_3^* \neq 0$.

A3 passes skewness onto the unconditional distribution of Y_t. The direction of skewness is unconstrained.[8] Skewness in (high frequency) returns is considered a stylized fact. This fact is exogenous to the model under consideration, yet (as will be shown) can be harnessed to identify the model. Examples where an asymmetric D is used to account for skewness in returns include Hansen (1994), Harvey and Siddique (1999), and Jondeau and Rockinger (2003).

ASSUMPTION A4: $E\left(A^{3/2}\right) < 1$.

A4 is sufficient for $\{Y_t\}$ to have a strictly stationary solution (see; e.g., Mikosch, 1999, Corollary 1.4.38 and Remark 1.4.39). Throughout this and the remaining sections, assume that this strictly stationary solution is the one being observed. From (2) and (3) follows that

$$Y_t^2 = \sigma_t^2 + W_t, \tag{4}$$

where $\{W_t\}$ is an MDS. Let $X_t \equiv Y_t^2 - \gamma_0$, where

$$\gamma_0 \equiv E\left(Y_t^2\right) = \frac{\omega_0}{1-\phi_0}, \qquad \phi_0 = \alpha_0 + \beta_0,$$

and $\phi_0 < 1$, given A4. Then from (4) follows that

$$\begin{aligned} X_t &= \phi_0 X_{t-1} - \beta_0 W_{t-1} + W_t \\ &= \phi_0 X_{t-1} + V_t \end{aligned} \tag{5}$$

which relates the GARCH$(1,1)$ model to an ARMA$(1,1)$ model of the (centered) second-order sequence $\left\{Y_t^2\right\}$. Also given A4, $E\left(Y_t^3\right) = E\left(\sigma_t^3\right) \times c_3^*$ is well defined (see Prono, 2018, Lemma 1). Consequently, given the law of iterated expectations, multiplying both sides of (5) by Y_{t-m} for a $m \geq 1$ and taking expectations produces

$$E\left(X_t Y_{t-m}\right) = \alpha_0 \phi_0^{m-1} E\left(Y_t^3\right),$$

[8] For equity returns, as an example, skewness can be of either sign for single names and tends to be negative for portfolios.

in which case,

$$E\left(X_tY_{t-1}\right) = \alpha_0 E\left(Y_t^3\right), \tag{6}$$

and

$$E\left(X_tY_{t-m}\right) = \phi_0 E\left(X_tY_{t-m+1}\right), \qquad m \geq 2. \tag{7}$$

From (6), an exactly identified estimator for α_0 in (5) is

$$\widehat{\alpha}_{IV} = \widehat{F}\left(n^{-1}\sum_t \widehat{X}_tY_{t-1}\right), \qquad \widehat{F} = \left(n^{-1}\sum_t \widehat{X}_{t-1}Y_{t-1}\right)^{-1}, \tag{8}$$

where

$$\widehat{X}_t = Y_t^2 - \widehat{\gamma}, \qquad \widehat{\gamma} = n^{-1}\sum_t Y_t^2. \tag{9}$$

Notice that (8) is a linear TSLS estimator applied to the feasible version of (5) (see (25) in the Appendix) using Y_{t-1} as an instrument for $\widehat{X}_{t-1}$. Notice as well that Y_{t-1} is not a proper instrument, since

$$E\left(W_{t-1}Y_{t-1}\right) = E\left(Y_{t-1}^3\right) \neq 0.$$

Nonetheless, Y_{t-1} is sufficient for identifying α_0 from (5) as the following Theorem demonstrates.

Theorem 1 *Consider the estimator in (8). Let $F_0 = E\left(X_{t-1}Y_{t-1}\right)^{-1}$, and let Assumptions A1–A4 hold. Then*

$$\widehat{\alpha}_{IV} \xrightarrow{a.s.} \alpha_0,$$

and

$$na_n^{-3}\left(\widehat{\alpha}_{IV} - \alpha_0\right) \xrightarrow{d} \alpha_0^{-1}F_0\left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right), \tag{10}$$

where $\kappa_0 \in (3,\ 6)$, "$\xrightarrow{d}$" is weak, and the limiting random variables $\left(V_{i,\mathbf{y}}\right)_{i=1,2}$ defined in Lemma 11 are jointly $(\kappa_0/3)$ −stable. If $\kappa_0 \in (6,\ \infty)$, in which case, $E\left(Y_t^6\right) < \infty$, then

$$\sqrt{n}\left(\widehat{\alpha}_{IV} - \alpha_0\right) \xrightarrow{d} N\left(0,\ E\left(Y_t^3\right)^{-2}\Sigma_{VY_{-1}}\right), \tag{11}$$

where

$$\Sigma_{VY_{-1}} = E\left(\left(V_tY_{t-1}\right)^2\right) + 2\sum_{s=1}^{\infty} E\left(V_tY_{t-1}V_{t-s}Y_{t-1-s}\right).$$

Proof. See the Appendix for proofs of all theorems and corollaries stated here in the main text. See the

Supplemental Appendix for statements and proofs of all supporting lemmas as well as additional theorems and corollaries. ■

Remark 2 *Asymptotically, $\widehat{\gamma}$ does not affect the limiting distribution of $\widehat{\alpha}_{IV}$. Also, consistency of $\widehat{\alpha}_{IV}$ does not require consistency of $\widehat{\gamma}$ (see (26) in the Appendix). In thin-tailed cases where $E\left(Y_t^6\right) < \infty$ (which is equivalent to $E\left(A^3\right) < 1$), there is an inverse relationship between the required asymmetry in the distribution of Y_t and the asymptotic variance of $\widehat{\alpha}_{IV}$. Specifically, as $\left|E\left(Y_t^3\right)\right| \to 0$, the asymptotic variance of $\widehat{\alpha}_{IV}$ increases without bound. The limiting case where $E\left(Y_t^3\right) = 0$ corresponds to the case where this asymptotic variance is ill-defined, rendering $\widehat{\alpha}_{IV}$ unidentified. Analogously, in heavy-tailed cases where $\kappa_0 \in (3,\ 6)$ and, consequently, $E\left(Y_t^6\right) = \infty$, the stable limit is ill-defined when $E\left(Y_t^3\right) = 0$. In addition, away from symmetric innovations, the rate of convergence in (10) is $n^{\frac{\kappa_0-3}{\kappa_0}}$, which is quite a bit slower than $\sqrt{n}$ in (11), especially for empirically-relevant values of κ_0.*

Remark 3 *The distributional limit in (10) depends on $i \in (3,\ 6)$ in A1. This requirement is both consistent with existing limit theory for alternative GARCH estimators like the QMLE as well as the empirical features of many GARCH processes (see; e.g., Figure 1 as well as Hill and Renault, 2012). In contrast, a requirement more analogous to one employed in related works (see; e.g., Mikosch and Stărică, 2000, and Kristensen and Linton, 2006) would be for $i \geq 6$, which is both much stronger and not as well supported by empirical evidence.*[9]

Remark 4 *In the special case where $\beta_0 = 0$ (i.e., the ARCH(1) case), the distributional limits in (10) and (11) reduce to those in Prono (2018a, Theorem 1) with a single, lagged instrument.*

Owing to its dependence on κ_0, the exact convergence rate in (10) is unknown. This feature complicates bootstrapping a confidence interval for $\widehat{\alpha}_{IV}$.[10] Consider, then, the estimator $\widehat{\tau}_n^2 = n^{-1}\sum_t Y_t^6$. Given this estimator,

$$na_n^{-6}\widehat{\tau}_n^2 = a_n^{-6}\sum_t Y_t^6 \xrightarrow{d} V_{0,\mathbf{y}}^{(2)},$$

where $V_{0,\mathbf{y}}^{(2)}$ is $(\kappa_0/6)$ $-$stable following the method of proof given for Davis and Hsing (1995, Theorem 3.1(i)). Since $\left(\begin{array}{cc} V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}, & V_{0,\mathbf{y}}^{(2)} \end{array}\right)$ is jointly-stable following the arguments given for Vaynman and

[9]In each of these two referenced cases, second-order autocovariances are considered; i.e., $E\left(X_t X_{t-m}\right)$ for $m \geq 1$, in which case, the analogous condition is $i = 8$.

[10]The distributional limit in (10) has an awkward characteristic function that does not readily admit the construction of confidence intervals.

Beare (2014, Theorem 4),

$$\sqrt{n}\left(\frac{\widehat{\alpha}_{IV}-\alpha_0}{\widehat{\tau}_n}\right)\xrightarrow{d}\alpha_0^{-1}F_0\left(\frac{V_{2,\mathbf{y}}-\beta_0V_{1,\mathbf{y}}}{V_{0,\mathbf{y}}^{(2)}}\right),$$

by the continuous mapping theorem. Advantages of this normalization are twofold. First, the bootstrap method of Hall and Yao (2003) can now be applied.[11] Second, a bridge is provided for understanding the transition from (10) to (standard) asymptotic normality in (11). Specifically, if $\kappa_0\in(6,\ \infty)$, the limit of $\widehat{\tau}_n$ is degenerate, and the linear combination of $V_{1,\mathbf{y}}$ and $V_{2,\mathbf{y}}$ is Gaussian.

Next, let

$$R_t=X_t-\phi_0X_{t-1}$$

so that, given (5),

$$R_t=-\beta_0W_{t-1}+W_t, \tag{12}$$

making R_t a MA(1) process. Recursive substitution into (12) produces

$$R_t=-\beta_0R_{t-1}+U_t,\qquad U_t=W_t-\beta_0^2W_{t-2}. \tag{13}$$

From (13), an exactly identified IV estimator for β_0 is

$$\widehat{\beta}_{IV}\left(\widehat{\phi}\right)=-\widehat{G}\left(n^{-1}\sum_t\widehat{R}_tY_{t-1}\right),\qquad \widehat{G}=\left(n^{-1}\sum_t\widehat{R}_{t-1}Y_{t-1}\right)^{-1}, \tag{14}$$

where

$$\widehat{R}_t=\widehat{X}_t-\widehat{\phi}\widehat{X}_{t-1}.$$

The estimator in (14) is highly analogous to the one in (8), since

$$E\left(R_{t-1}Y_{t-1}\right)=E\left(X_{t-1}Y_{t-1}\right)=E\left(Y_{t-1}^3\right),$$

thereby linking instrument strength directly to the level of skewness in D. Differentiating (14) from (8) is

1. Y_{t-1} being a proper instrument for R_{t-1} (it is an improper instrument for X_{t-1} as previously discussed)

[11]This type of normalization applies to all of the simple estimators discussed in this paper and enables the calculation of confidence intervals for the, respective, parameter estimates, even in heavy-tailed cases.

2. the estimator depending on $\widehat{\phi}$

This second differentiating feature likens (14) to the closed-form estimator proposed by Kristensen and Linton (2006) and necessitates, in turn, a consistent estimator for ϕ_0. In order to satisfy this condition and preserve (14) being entirely closed form, consider

$$\widehat{\phi}_{IV} = \widehat{\mathbf{F}} \left(n^{-1} \sum_t \widehat{X}_t \mathbf{Z}_{t-2} \right), \qquad \widehat{\mathbf{F}} = \frac{\left(n^{-1} \sum_t \widehat{X}_{t-1} \mathbf{Z}_{t-2} \right)' \widehat{\mathbf{\Lambda}}}{\left(n^{-1} \sum_t \widehat{X}_{t-1} \mathbf{Z}_{t-2} \right)' \widehat{\mathbf{\Lambda}} \left(n^{-1} \sum_t \widehat{X}_{t-1} \mathbf{Z}_{t-2} \right)}, \tag{15}$$

with the instrument vector

$$\mathbf{Z}_{t-2} = \begin{pmatrix} Y_{t-2}, & \ldots, & Y_{t-h} \end{pmatrix}'.$$

Given A3, (7), and

$$E \left(W_{t-i} \mathbf{Z}_{t-2} \right) = 0, \qquad i = 0, 1,$$

which follows by the law of iterated expectations, $\mathbf{Z}_{t-2}$ is a proper set of instruments for X_{t-1} in (5).

ASSUMPTION A5: $\widehat{\mathbf{\Lambda}} \xrightarrow{a.s.} \mathbf{\Lambda}_0$, *a positive definite matrix*.

If $\widehat{\mathbf{\Lambda}} = \left(n^{-1} \sum_t \mathbf{Z}_{t-2} \mathbf{Z}'_{t-2} \right)^{-1}$, then $\widehat{\phi}^{IV}$ is a TSLS estimator. The advantage of this choice of a weighting matrix is that

$$\widehat{\mathbf{\Lambda}} = \left(n^{-1} \sum_t \mathbf{Z}_{t-2} \mathbf{Z}'_{t-2} \right)^{-1} \xrightarrow{a.s.} \gamma_0^{-1} \mathbf{I}_h,$$

where $\mathbf{I}_h$ is a $h \times h$ identity matrix, given Assumptions A1–A4. Alternatively, if

$$\widehat{\mathbf{\Lambda}} = \left(n^{-1} \sum_t \left(X_t - \widetilde{\phi} X_{t-1} \right)^2 \mathbf{Z}_{t-2} \mathbf{Z}'_{t-2} \right)^{-1},$$

where $\widetilde{\phi}$ is a preliminary estimator, then $\widehat{\phi}_{IV}$ is a two-step GMM estimator. While this second choice of a weighting matrix is certainly preferable on efficiency grounds, $\widehat{\mathbf{\Lambda}} \xrightarrow{a.s.} \mathbf{\Lambda}_0$ now requires $E \left(A^3 \right) < 1,$ which is a rather tall order given the empirical features of many financial return series (see; e.g., Figure 1). Consequently, this paper focuses on the TSLS interpretation of (15). Theorem 13 in the Supplemental Appendix shows that $\widehat{\phi}_{IV} \xrightarrow{a.s.} \phi_0$ and that $\widehat{\phi}_{IV}$ converges (weakly) in distribution to a limit that is qualitatively similar to (10) when $\kappa_0 \in (3,\ 6)$, with the same rate of convergence. The following theorem depends on these results.

Theorem 5 *Consider the estimator in (14) with $\widehat{\phi} = \widehat{\phi}_{IV}$ as defined in (15). Let*

$$\mathbf{A}_0 = \mathbf{\Lambda}_0 E\left(X_{t-1}\mathbf{Z}_{t-2}\right), \qquad B_0 = E\left(X_{t-1}\mathbf{Z}_{t-2}\right)' \mathbf{A}_0, \qquad G_0 = E\left(R_{t-1}Y_{t-1}\right)^{-1},$$

and let Assumptions A1–A5 hold. Then

$$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right) \xrightarrow{a.s.} \beta_0,$$

and

$$na_n^{-3}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right) - \beta_0\right) \xrightarrow{d} \alpha_0^{-1}\left(\mathbf{F}_0\mathbf{S} + G_0\left(\alpha_0\phi_0 V_{0,\mathbf{y}} - \left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right)\right)\right), \tag{16}$$

where $\kappa_0 \in (3,\ 6)$, "$\xrightarrow{d}$" is weak, $V_{0,\mathbf{y}}$ is defined in Lemma 12, $\left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right)$ heralds from Theorem 1, and $F_0 S$ is defined in Theorem 13 of the Supplemental Appendix. This (weak) distributional limit is also $(\kappa_0/3)$ −stable. If $\kappa_0 \in (6,\ \infty)$, in which case, $E\left(Y_t^6\right) < \infty$, then

$$\sqrt{n}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right) - \beta_0\right) \xrightarrow{d} N\left(0,\ \Sigma_\beta\right),$$

where

$$\Sigma_\beta = E\left(Y_t^3\right)^{-2}\left[B_0^2 \Sigma_{UY_{-1}} + E\left(Y_t^3\right)^2\left(\mathbf{A}_0' \mathbf{\Sigma}_{V\mathbf{Z}_{-2}} \mathbf{A}_0 - 2B_0 \mathbf{\Sigma}_{UVY_{-1}\mathbf{Z}_{-2}}' \mathbf{A}_0\right)\right],$$

A_0, B_0, and $\Sigma_{V\mathbf{Z}_{-2}}$ are defined in Theorem 13,

$$\Sigma_{UY_{-1}} = E\left(\left(U_t Y_{t-1}\right)^2\right) + 2E\left(U_t U_{t-1} Y_{t-1} Y_{t-2}\right),$$

$$\mathbf{\Sigma}_{V\mathbf{Z}_{-2}} = E\left(V_t^2 \mathbf{Z}_{t-2}\mathbf{Z}_{t-2}'\right) + 2E\left(V_t V_{t-1}\mathbf{Z}_{t-2}\mathbf{Z}_{t-3}'\right),$$

and

$$\mathbf{\Sigma}_{UVY_{-1}\mathbf{Z}_{-2}} = E\left(U_t V_t Y_{t-1}\mathbf{Z}_{t-2}\right) + 2E\left(U_t V_{t-1} Y_{t-1}\mathbf{Z}_{t-3}\right).$$

Remark 6 *As is true in Theorem 1, (i) $\widehat{\gamma}$ impacts neither consistency nor the asymptotic variance of $\widehat{\beta}_{IV}$, (ii) a necessary condition for the limiting result in (16) is that $i \in (3,\ 6)$ in A1, and (iii) in order for both the heavy-tailed and thin-tailed distributional limits to be stable, $E\left(Y_t^3\right) \neq 0$; otherwise, $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ is not identified.*

Remark 7 *The limiting result in (16) is a linear combination of the results in (10) and (28) in the Supplemental Appendix; in which case, the limiting distribution of $\widehat{\phi}$ impacts the asymptotic distribution of $\widehat{\beta}_{IV}$ as* *it does in the Kristensen and Linton (2006) estimator.*

Corollary 8 *Consider the estimator in (14). Let Assumptions A1–A4 hold. In addition, assume there exists a $\widehat{\phi}$ such that $\widehat{\phi} \xrightarrow{a.s.} \phi_0$ with a rate of convergence of to a stable limiting distribution of n^l, where $l > \frac{\kappa_0 - 3}{\kappa_0}$. Then,*

$$\widehat{\beta}_{IV}\left(\widehat{\phi}\right) \xrightarrow{a.s.} \beta_0$$

and

$$na_n^{-3}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right) - \beta_0\right) \xrightarrow{d} G_0\left(\phi_0 V_{0,Y} - \alpha_0^{-1}\left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right)\right). \tag{17}$$

Remark 9 *Relative to (16), (17) has a (substantial) source of variation in the asymptotic limit removed. The implications of this result are threefold. First, for any consistent $\widehat{\phi}$ that converges faster than $\widehat{\phi}_{IV}$ (see Theorem 13 in the Supplemental Appendix), asymptotically there is no difference between using this estimate and the true value ϕ_0. Second, in this case, $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ is asymptotically more efficient than $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$. Third, it is natural to consider $\widehat{\phi} = \widehat{\phi}_{QMLE}$, since $\widehat{\phi}_{QMLE}$ is both consistent and n^l asymptotically normal under conditions supported by the Corollary (see; e.g., Berkes, Horváth, and Kokoszka, 2003, Hall and Yao, 2003, and Francq and Zakoïan, 2004).*

Given $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ in Corollary 8, consider the alternative estimator for α_0

$$\widehat{\alpha}_{IV}\left(\widehat{\phi}\right) = \widehat{\phi} - \widehat{\beta}_{IV}\left(\widehat{\phi}\right). \tag{18}$$

Corollary 10 *Consider the estimator in (18). Let Assumptions A1–A5 hold. Then for both the $\widehat{\phi}$ defined in Corollary 8 and $\widehat{\phi} = \widehat{\phi}_{IV}$,*

$$\widehat{\alpha}_{IV}\left(\widehat{\phi}\right) \xrightarrow{a.s.} \alpha_0,$$

and

$$na_n^{-3}\left(\widehat{\alpha}_{IV}\left(\widehat{\phi}\right) - \alpha_0\right) \xrightarrow{d} -G_0\left(\phi_0 V_{0,Y} - \alpha_0^{-1}\left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right)\right). \tag{19}$$

From Theorem 1 and Corollary 10, if

$$na_n^{-3}\left(\widehat{\alpha}_{IV} - \alpha_0\right) \xrightarrow{d} X,$$

then

$$na_n^{-3}\left(\widehat{\alpha}_{IV}\left(\widehat{\phi}\right) - \alpha_0\right) \xrightarrow{d} X - Y$$

Drawing upon what is (generally) known for two-step estimators, necessary for $\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)$ to display asymptotic efficiency gains over $\widehat{\alpha}_{IV}$ is a (strong) positive association between X and Y (see; e.g., Newey and McFadden, 1994). Specifically, given (10) and (19), there needs to be a (strong) and positive association between $V_{0,Y}$ and $\left(V_{2,\mathbf{y}} - \beta_0 V_{1,\mathbf{y}}\right)$. From Lemma 12 in the Supplemental Appendix, $V_{0,Y}$ is the limit to $n^{-1}\sum_t Y_t^3$, while $V_{m,Y}$ is the limit to $n^{-1}\sum_t Y_t Y_{t+m}^2$ for $m = 1, 2$ (see Lemma 11). Given (6) and (7), supporting such a positive association in large samples and thin-tailed cases where $\kappa_0 \in (6,\ \infty)$ is then

$$
\begin{aligned}
Cov\left(n^{-1}\sum_t Y_t^3,\ n^{-1}\sum_t Y_t Y_{t+2}^2 - \beta_0 n^{-1}\sum_t Y_t Y_{t+1}^2\right) &= Cov\left(n^{-1}\sum_t Y_t^3,\ n^{-1}\sum_t Y_t Y_{t+2}^2\right) \\
&\quad -\beta_0 Cov\left(n^{-1}\sum_t Y_t^3,\ n^{-1}\sum_t Y_t Y_{t+1}^2\right) \\
&\approx Cov\left(n^{-1}\sum_t Y_t^3,\ \alpha_0\phi_0 n^{-1}\sum_t Y_t^3\right) \\
&\quad -\beta_0 Cov\left(n^{-1}\sum_t Y_t^3,\ \alpha_0 n^{-1}\sum_t Y_t^3\right) \\
&= \alpha_0^2 Var\left(n^{-1}\sum_t Y_t^3\right) \\
&> 0
\end{aligned}
$$

It can be anticipated, however, that any efficiency gains in $\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)$ over $\widehat{\alpha}_{IV}$ will be muted relative to the gains in $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ over $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$.

2.2. Potential for Efficiency Gains

Consider the estimator $\widehat{\theta}$.

ASSUMPTION A6: $\widehat{\theta} \xrightarrow{a.s.} \theta_0$ *and* $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0,\ \Sigma_\theta\right)$*, where*

$$
\Sigma_\theta = \begin{pmatrix} \Sigma_\omega & \Sigma_{\omega,\ \alpha} & \Sigma_{\omega,\ \beta} \\ \Sigma_{\omega,\ \alpha} & \Sigma_\alpha & \Sigma_{\alpha,\ \beta} \\ \Sigma_{\omega,\ \beta} & \Sigma_{\alpha,\ \beta} & \Sigma_\beta \end{pmatrix}.
$$

In addition,

$$
\Sigma_\alpha + 2\Sigma_{\alpha,\ \beta} < 0. \tag{20}
$$

Since $\begin{pmatrix} 0, & 1, & 1 \end{pmatrix} \left(\widehat{\theta} - \theta_0\right) = \left(\widehat{\phi} - \phi_0\right)$, from A6 follows that

$$\sqrt{n}\left(\widehat{\phi} - \phi_0\right) \xrightarrow{d} N\left(0, \Sigma_\beta + \Sigma_\alpha + 2\Sigma_{\alpha,\ \beta}\right).$$

Consequently, given (20), $\Sigma_\phi < \Sigma_\beta$. Moreover, since $\Sigma_\alpha > 0$, necessary for (20) is $\Sigma_{\alpha,\ \beta} < 0$. Given the discussion that follows Corollary 8, a natural candidate for $\widehat{\theta}$ is $\widehat{\theta}_{QMLE}$. The next section considers a wide range of Monte Carlo simulation designs for the model in (2) and (3). Under all of these designs (without exception), including ones conducted using (very) large sample sizes, (20) holds when $\widehat{\theta} = \widehat{\theta}_{QMLE}$. Consequently, because the Monte Carlo designs conform both with the magnitudes of ARCH and GARCH parameters and the stylized facts of GARCH innovations encountered empirically, A6 appears to (at least) enjoy broad empirical support when applied to the QMLE.

Consider next the estimator in (8). Given Theorem 1,

$$\sqrt{n}\left(\widehat{\alpha}_{IV} - \alpha_0\right) \xrightarrow{d} N\left(0,\ \Sigma_\alpha^*\right)$$

in thin-tailed cases where $\kappa_0 \in (6,\ \infty)$. Further consider

$$\Sigma_{\alpha,\ \beta}^* = nE\left(\left(\widehat{\alpha}_{IV} - \alpha_0\right)\left(\widehat{\beta} - \beta_0\right)\right), \qquad \Omega_\alpha^* = nE\left(\left(\widehat{\alpha}_{IV} - \alpha_0\right)\left(\widehat{\alpha} - \alpha_0\right)\right).$$

ASSUMPTION A7:

$$\Sigma_\alpha^* - 2\left(\Omega_\alpha^* + \Sigma_{\alpha,\ \beta}^*\right) < 0 \tag{21}$$

In all the Monte Carlo experiments considered in the next section, $\Omega_\alpha^* + \Sigma_{\alpha,\ \beta}^* < 0$ when $\widehat{\theta} = \widehat{\theta}_{QMLE}$. Consequently, the relative size of Σ_α^* is important for determining whether (21) holds. Recall from (11) that as $\left|E\left(Y_t^3\right)\right|$ increases, Σ_α^* decreases. As a result, the likelihood of (21) holding increases along with the magnitude of skewness in Y_t. Empirical evidence supports elevated skewness levels in (very) high frequency returns (see Table 1). These elevated levels, however, correspond with heavy-tailed return processes (see Figure 1). An interesting question, then, is whether there exist thin-tailed GARCH processes in the context of Theorem 1 with sufficient skewness to satisfy (21). Evidencing why this question is interesting is the following theorem.

Theorem 11 *Consider the estimator in (14) under thin-tailed cases where* $\kappa_0 \in (6, \infty)$. *Let Assumptions A1–A4 and A6–A7 hold. Then,* $\widehat{\beta}_{IV}\left(\widehat{\phi}\right) \xrightarrow{a.s.} \beta_0$,

$$\sqrt{n}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right) - \beta_0\right) \xrightarrow{d} N\left(0, \Sigma^*_\beta\right),$$

and

$$\Sigma^*_\beta < \Sigma_\beta.$$

Remark 12 *From Theorem 11, there exist conditions under which* $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ *has a smaller asymptotic variance than* $\widehat{\beta}$. *Under these conditions,* $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ *can be interpreted as enhancing the efficiency of* $\widehat{\beta}$, *generally, and* $\widehat{\beta}_{QMLE}$, *specifically, in an analogous way that the estimators of Francq et al. (2011) and Fan et al. (2014) enhance the efficiency of* $\widehat{\beta}_{QMLE}$. *These latter estimators achieve efficiency gains (i.e., smaller asymptotic variances) by targeting the scale of the unknown innovation density* D. *Rather than targeting scale,* $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ *targets the skewness of* D. *To the extent that this skewness is pronounced (i.e., a prevalent feature of* D*), efficiency gains should result, especially if the initial estimator* $\widehat{\theta}$ *ignores this skewness, as is the case, generally, for many GARCH estimators, and, certainly,* $\widehat{\theta}_{QMLE}$, *specifically.*

Corollary 13 *Consider the estimator in (18) under thin-tailed cases where* $\kappa_0 \in (6, \infty)$. *Let Assumptions A1–A4 hold. Then*

$$\sqrt{n}\left(\widehat{\alpha}_{IV}\left(\widehat{\phi}\right) - \alpha_0\right) \xrightarrow{d} N\left(0, \Sigma^*_\alpha\right),$$

the same distributional limit as $\widehat{\alpha}_{IV}$ *in (11).*

Remark 14 *In thin tailed cases where* $\kappa_0 \in (6, \infty)$, $\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)$ *and* $\widehat{\alpha}_{IV}$ *share the same distributional limit. Consequently, if* $\Sigma^*_\alpha < \Sigma_\alpha$, *which likely sources to heavy skewness in* D, $\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)$ *offers no improvement over* $\widehat{\alpha}_{IV}$, *a result which, owing to simulation evidence, stands in contrast to the improvement* $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$ *affords over* $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$.

3.1. Monte Carlo Design

This section considers the GARCH$(1, 1)$ model from Section 2.1, where $\{\epsilon_t\}$ is drawn from the skewed student's-t density of Hansen (1994). This density has two parameters, λ and η, where the former governs skewness, the latter governs the tails, and up to the ηth moment of the distribution is well defined. Values for these parameters considered in the Monte Carlo experiments are

$$\lambda = -0.20, \ -0.40, \ -0.80, \ -0.90, \ -0.99, \qquad \eta = 64.5, \ 8.1, \ 4.5, \ 3.5. \tag{22}$$

As λ increases, skewness increases, while as η decreases, tail thickness increases. The Monte Carlo experiments summarized in this section involve the GARCH specification of

$$\omega_0 = 0.005, \qquad \alpha_0 = 0.10, \qquad \beta_0 = 0.80. \tag{23}$$

For robustness, other specifications are also considered, the results for which are summarized in the Supplemental Appendix. The estimators under study are the simple IV estimators from Section 2.1, with both the closed-form Kristensen and Linton (2006) estimator (KL) and the QMLE serving as benchmarks. Recalling that m denotes the number of lags used as instruments, for $\widehat{\alpha}_{IV}$, $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, $m = 1$. For $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$, $m = 20, \ 10, \ 5$, so as to investigate the effects of the number of instruments on the performance of the estimator. Table 2 summarizes the skewness statistics and tail index estimates for $\{Y_t\}_{t=1}^{T}$, given the GARCH specification in (23) and the different values for λ and η in (22). Notice that skewness levels in the simulations are consistent with skewness levels encountered empirically (see Table 1). When $\eta = 64.5$, $\widehat{\alpha}_{IV}$, $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, $\widehat{\alpha}_{QMLE}$, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, and $\widehat{\beta}_{QMLE}$ are all asymptotically normal, while $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ (likely) are not.[12] When $\eta = 8.1$ and $\eta = 4.5$, only $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{QMLE}$ are asymptotically normal. When $\eta = 3.5$, none of the estimators are asymptotically normal.[13].When $\eta = 4.5$ and $\eta = 3.5$, $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ are not consistent; in which case, they are not considered in the experiments.[14]

Samples sizes for the simulations are $100,000$ and 500 observations, respectively, with the former investigating the large-sample properties of the simple IV estimators (given their slow convergence rates), and the latter investigating their small-sample properties. All simulations are conducted over $10,000$ trials, with the first 200 observations dropped to avoid initialization effects. Summary statistics for the simulations are the root mean squared error, mean absolute error, and median absolute error (each measured relative to the true parameter value) divided by the corresponding efficiency measure for the QMLE. These ratios are termed "efficiency ratios," and benchmark the performance of the simple IV estimators against the QMLE. Additional details on the simulations are contained in the notes to the relevant Tables.

[12]Necessary for $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ to be asymptotically normal is $E\left(Y_t^8\right) < \infty$ (see Kristensen and Linton, 2006), which does not appear to be true, given the results in Table 2.

[13]A necessary condition for $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{QMLE}$ to be asymptotically normal is $E\left(\epsilon_t^4\right) < \infty$ (see; e.g., Hall and Yao, 2003).

[14]Necessary for consistency of $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ is $E\left(Y_t^4\right) < \infty$, which (very likely) does not hold, given the results in Table 2.

3.2. Results

Tables 3–4 report large-sample results for the simple estimators proposed herein at varying levels of tail-thickness for the GARCH$(1,1)$ model's innovation density. Beginning with Table 3, in the thin-tailed case of $\eta = 64.5$, the relative performance of all simple estimators improves as λ increases (in absolute terms). The opposite is true for the KL estimator, which sees its performance meaningfully degrade with increasing levels of skewness At low skewness levels, the KL estimator is more efficient than the simple estimators. At high skewness levels, this tendency is reversed, with the KL estimator notably lagging the simple estimators. Amongst the simple estimators, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ performs the best. However, all simple estimators, including the KL estimator, notably lag the QMLE.

Moving to the heavier-tailed case of $\eta = 8.1$ (still Table 3), the same trends mentioned above continue to hold. There are, however, some notable points of departure from these trends. Specifically, the relative performance of $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ materially degrades in this heavier-tailed case. Also, its performance now appears to worsen as λ increases (in absolute terms). In contrast, the relative performance of $\widehat{\alpha}_{IV}$, $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ remains much more stable across the two cases.

In the heavy-tailed case of $\eta = 4.5$ (now Table 4), a substantial relative performance drop is, again, evidenced for $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$. Consequently, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ now notably lags in performance relative to the other simple estimators at all skewness levels considered. Overall, there is a general tendency for the relative performance of the other simple estimators to also decline between the cases of $\eta = 8.1$ and $\eta = 4.5$; however, this decline is decidedly more modest. In the heaviest-tailed case of $\eta = 3.5$, the relative performance of all simple estimators notably improves, likely owing to the fact that while under the cases of $\eta = 8.1$ and $\eta = 4.5$, QMLE is asymptotically normal, under the case of $\eta = 3.5$, the distributional limit of QMLE is qualitatively (much) more similar to that of the simple estimators.

Table 5 summarizes results of an investigation into whether Assumptions A6 and A7 of Theorem 11 can hold in thin-tailed cases that support asymptotic normality for both $\widehat{\beta}_{QMLE}$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$. The design of this investigation is as follows. For different ARCH and GARCH parameter values (listed in the Table) and the highest possible λ values (in absolute terms), determine the lowest possible η value that still supports asymptotic normality for $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$. In each of these cases, measure the efficiency of $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ relative to $\widehat{\beta}_{QMLE}$. Across all of the cases summarized in Table 5 (and, in fact, all large and small sample cases considered in this section), A6 holds.[15] Consequently, as hypothesized in Section 2.2, the prediction of Theorem 11 critically hinges upon the validity of A7. Table 5 demonstrates that A7 can, in fact, hold, in

[15]This finding, though not explicitly evident in the Table, can be seen through performance comparisons of $\widehat{\phi}_{QMLE}$ and $\widehat{\beta}_{QMLE}$. These comparisons are available upon request.

which case, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ is more efficient than $\widehat{\beta}_{QMLE}$.

It is tempting to conclude from the results in Table 5 that efficiency gains of $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ over $\widehat{\beta}_{QMLE}$ are limited to GARCH processes with low persistence. Such a conclusion under-weights the fact that cases where $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ is more efficient than $\widehat{\beta}_{QMLE}$ correspond with the highest skewness levels in $\{Y_t\}_{t=1}^{T}$. Evidently, lower GARCH persistent levels are associated with higher skewness levels in the Table. However, to the extent that is possible to generate skewness levels $\geq$ (in absolute terms) the highest levels observed in Table 5 for more persistent GARCH processes, then it seems likely that $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ will exceed the performance of $\widehat{\beta}_{QMLE}$ in these cases as well.

Tables 3–5 evidence that the simple estimators record their best performance relative to QMLE in cases where skewness is high. In Table 5, the level of skewness achievable is limited by the constraint to only consider thin-tailed densities. Relaxing this constraint allows for materially higher skewness levels (see Table 2), the effects of which are demonstrated in Tables 3–4 for (very) large samples. From Section 2.1, the convergence rate of the simple estimators is slow, implying that their distributional limits as proxied for in the large sample results might offer poor predictions for how the simple estimators behave in small samples under the same heavy-tailed and skewed data generating processes. It is also generally known that asymptotic normality offers a poor proxy for QMLE in small samples of non-normally distributed data. Consequently, Table 6 summarizes small sample results from the cases of $\eta = 4.5$ and $\eta = 3.5$. The results are striking. In particular, $\widehat{\alpha}_{IV}$, $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ are now shown to outperform their QMLE counterparts when skewness is high. Moreover, this outperformance can be (very) substantial.

3.3. Interpretation

$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ are analogous to the multi-step estimators of Preminger and Storti (2017) and Fan, Qi and Xiu (2014) (see also Francq, Lepage and Zakoïan, 2011), the latter of which are intended to improve upon the efficiency of $\widehat{\boldsymbol{\theta}}_{QMLE}$ by using $\widehat{\boldsymbol{\theta}}_{QMLE}$ as first-step inputs. The relative performance of these simple estimators against the QMLE strengthens this analogy. Simulation evidence in Preminger and Storti (2017) for their least squares estimator (LSE) benchmarked against the Fan et al. (2014) non-gaussian quasi-maximum likelihood estimator (NGQMLE) for the same sample size and $(\alpha_0,\ \beta_0)$ values considered here shows $\widehat{\alpha}_{NGQMLE}$ to best $\widehat{\alpha}_{QMLE}$ (in terms of root mean squared error) by less than $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ bests $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{LSE}$ to best $\widehat{\beta}_{QMLE}$ to a comparable degree as $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ bests $\widehat{\beta}_{QMLE}$.

To understand why $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ can improve upon their QMLE inputs, begin by noting that both LSE and NGQMLE attempt such improvements by better targeting the scale of the unknown GARCH$(1,1)$ innovation density. The simple estimators $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ also target a particular feature of this innovation density; namely, its skewness. To the extent that skewness figures prominently in this density (as it does in the high λ cases), it is certainly possible that estimators focused on this feature can outperform alternatives that ignore it (like QMLE), even if those alternatives possess better large sample properties, so long as those large sample properties have yet to apply. Case-in-point, the small sample distributions of both $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{QMLE}$ evidenced in Table 6 are characterized by elevated levels of both skewness and excess kurtosis, making normality a (very) poor approximation for these distributions and, consequently, rendering the large-sample properties of $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{QMLE}$ uninformative of their, respective, small-sample behavior. Noting further that the target of both LSE and NGQMLE is also insensitive to skewness, it is, perhaps, less surprising that $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ would (at least) perform comparably to these alternatives, in cases where skewness is a feature worth targeting.

4. Empirical Application

Estimators for the GARCH$(1,1)$ model from Section 3 are applied to intra-day Japanese Yen log returns measured against the USD at 15-, 10-, 5-, and 1-minute sampling frequencies over the time period January 1, 2015 through July 1, 2015.[16] Using the approach in Hecq, Laurent, and Palm (2012, Eq. 4.1), all log returns are pre-filtered for the U-shaped intra-day periodicity noted by Anderson and Bollerslev (1997). The QMLE serves to benchmark the simple IV estimators.

Tables 7–8 summarize the estimation results. With the exception of the 1-minute frequency, $\widehat{\alpha}_{IV}$ and $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ exist inside the 95% confidence intervals for $\widehat{\alpha}_{QMLE}$.[17] In contrast, and also with the exception of the 1-minute frequency, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ exists well outside the 95% confidence intervals for $\widehat{\beta}_{QMLE}$, while $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ exists inside them. These estimation results conform with the Monte Carlo experiments in that $\widehat{\alpha}_{IV}$ tends to perform better than $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ at all sample sizes and across all degrees of tail thickness. Specifically, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ tends to be (severely) biased, where the source of the bias is $\widehat{\phi}_{IV}$. The estimation results confirm this finding, since $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ tends to be (well) inside the 95% confidence interval for $\widehat{\beta}_{QMLE}$. Notice, however, that at the 1-minute frequency (which, corresponds with the largest

[16] Japanese Yen log returns are selected because of the forecast comparisons conducted by Hansen and Lunde (2005), which show that the GARCH$(1,1)$ model provides the best volatility forecast for these returns over more complicated GARCH specifications.

[17] These confidence intervals are based on asymptotic normality. Given the, respective, tail index estimates, asymptotic normality for the QMLE is suspect (see; e.g., Hall and Yao, 2003). Consequently, the true confidence intervals are likely to be wider.

data sample), $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ is also well inside the 95% confidence interval for $\widehat{\beta}_{QMLE}$. Consequently, in this case, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ enjoys a sizable advantage over $\widehat{\beta}_{QMLE}$ in terms of computation time (the former is faster to compute by orders-of-magnitude over the latter), with no seeming cost in terms of sacrificed precision.

5. Conclusion

This paper proposes simple estimators for the popular GARCH$(1,1)$ model and studies their properties. Simple, in this context, means available in closed form. Consequently, all such estimators are linear TSLS estimators. In some cases, the instruments upon which these estimators depend, in turn, depend on preliminary parameter estimates from the GARCH$(1,1)$ model that may, or may not, be available in closed form. An example of the latter case is preliminary QML estimates; in which case, the linear TSLS estimators based on these estimates are shown to improve upon the efficiency (either asymptotically, in thin-tailed cases, or when applied to small samples, in heavy-tailed cases) of QMLE. As a result, these simple estimators are members of the (growing) class of multi-step estimators aimed at improving the performance of QMLE by better aligning the parameter estimates with certain aspects of the unknown GARCH$(1,1)$ innovation density.

Established in this paper are the desirable properties of the simple estimators over the QMLE alternative. These desirable properties relate to in-sample fit. It would be interesting to investigate the degree to which these desirable properties translate into improved out-of-sample volatility forecasts. That is, to what extent (and under what conditions) can the simple estimators beat the QMLE in out-of-sample forecasting? This investigation is left for future research.

References

[1] Amado, C. & T. Teräsvirta (2013) Modelling volatility by variance decomposition. *Journal of Econometrics* 175, 142-153.

[2] Amado, C. & T. Teräsvirta (2014) Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance* 25, 15-35.

[3] Andersen, T.G. & T. Bollerslev (1997) Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 4, 115-158.

[4] Baillie, R.T., & T. Bollerslev (1989) The message in daily exchange rates: a conditional-variance tale. *Journal of Business and Economic Statistics* 7, 297-305.

[5] Berkes, I., L. Horváth & P. Kokoszka (2003) GARCH processes: structure and estimation. *Bernoulli* 9, 201-227.

[6] Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

[7] Bouchaud, J.P. & M. Potters (2003) *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Second Edition). Cambridge University Press.

[8] Davis, R.A. & T. Hsing (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability* 23, 879-917.

[9] Dong, Y. & A. Lewbel (2015) A simple estimator for binary choice models with endogenous regressors. *Econometric Reviews* 34, 82-105.

[10] Drost, F.C & C.A.J. Klaassen (1997) Efficient estimation in semiparametric garch models. *Journal of Econometrics* 81, 193-221.

[11] Drost, F.C. & T.E. Nijman (1993) Temporal aggregation of garch processes. *Econometrica* 61, 909-927.

[12] Engle, R.F. & J.G. Ranel (2008) The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies* 21, 1187-1222.

[13] Fan, J., L. Qi and D. Xiu (2014) Quasi-maximum likelihood estimation of garch models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics* 32, 178-191.

[14] Francq, C., & J.-M. Zakoïan (2004) Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli* 10, 605-637.

[15] Francq, C., G. Lepage & J.M. Zakoïan (2011), Two-stage non gaussian qml estimation of garch models and testing the efficiency of the gaussian qmle. *Journal of Econometrics* 165, 246-257.

[16] Giraitis, L. & P.M. Robinson (2001) Whittle estimation of arch models. *Econometric Theory* 17, 608-631.

[17] Glosten, L.R., R. Jagannathan & D.E. Runkle (1993) On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779-1801.

[18] Hall, P. & Q. Yao (2003) Inference in arch and garch models with heavy-tailed errors. *Econometrica* 71, 285-317.

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

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

[21] Harvey, C.R. & A. Siddique (1999) Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34, 465-487.

[22] Hecq, A., S. Laurent & F.C. Palm (2012) Common intraday periodicity. *Journal of Financial Econometrics* 10, 325-353.

[23] Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 5, 1163-1174.

[24] Hill, J.B. (2010) On tail index estimation for dependent, heterogeneous data." *Econometric Theory* 26(5): 1398-1436.

[25] Hill, J.B. & E. Renault (2012) Variance targeting for heavy tailed time series. Unpublished manuscript.

[26] Jondeau, E. & M. Rockinger (2003) Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27, 1699-1737.

[27] Kristensen, D. & O. Linton (2006) A closed form estimator for the garch(1,1)-model. *Econometric Theory* 22, 323-327.

[28] Kristensen, D. & A. Rahbek (2005) Asymptotics of the qmle for a class of arch(q) models. *Econometric Theory* 21, 946-961.

[29] Lee, S.W. & B.E. Hansen (1994) Asymptotic theory for the garch(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29-52.

[30] Lewbel, A. (2004) Simple estimators for hard problems: endogeneity in discrete choice related models. unpublished manuscript.

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

[32] Mikosch, T. (1999) Regular variation, subexponentiality and their applications in probability theory. Lecture notes for the workshop "Heavy Tails and Queques," EURANDOM, Eindhoven, Netherlands.

[33] Mikosch, T. & C. Stărică (2000) Limit theory for the sample autocorrelations and extremes of a garch(1,1) process. *The Annals of Statistics* 28, 1427-1451.

[34] Mikosch, T. & D. Straumann (2002) Whittle estimation in a heavy-tailed garch(1,1) model. *Stochastic Processes and their Applications* 100, 187-222.

[35] Mikosch, T. & D. Straumann (2006) Quasi-maximum-likelihood estimation in conditionally heteroskedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34, 2449-2495.

[36] Newey, W.K. & 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.

[37] Preminger, A. & G. Storti (2017) Least-squares estimation of garch(1,1) models with heavy-tailed errors. *Econometrics Journal* 20, 221-258.

[38] Prono, T. (2014) Simple estimators for the GARCH(1,1) model. Available at SSRN: http://ssrn.com/abstract=1511720.

[39] Prono, T. (2018a) Closed-form estimators for finite-order arch models as simple and competitive alternatives to qmle. Forthcoming in *Studies in Nonlinear Dynamics and Econometrics*.

[40] Prono, T. (2018b) Closed-form estimators for finite-order arch models as simple and competitive alternatives to qmle: supplemental appendix. Forthcoming in *Studies in Nonlinear Dynamics and Econometrics*.

[41] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*. New York: Springer-Verlag.

[42] Samorodnitsky, G. & M.S. Taqqu (1994) *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Stochastic Modeling*. New York: Chapmand and Hall.

[43] Shao, Y. & M. Zhou (2010) A characterization of multivariate normality through univariate projections. *Journal of Multivariate Analysis* 101, 2637-2640.

[44] Vaynman, I. & B.K. Beare (2014) Stable limit theory for the variance targeting estimator, in Y. Chang, T.B. Fomby & J.Y. Park (eds), *Essays in Honor of Peter C.B. Phillips*, vol. 33 of *Advances in Econometrics:* Emerald Group Publishing Limited, chapter 24, 639-672.

Appendix

All Lemmas, upon which the proofs of the Theorems rely, are both stated and proved in the Supplemental Appendix.

Proof of Theorem 1: To begin, note that

$$\widehat{X}_t = X_t - (\widehat{\gamma} - \gamma_0), \tag{24}$$

so that given (5),

$$\widehat{X}_t = c_0 + \phi_0 \widehat{X}_{t-1} - \beta_0 W_{t-1} + W_t, \qquad c_0 = (\widehat{\gamma} - \gamma_0) \times (\phi_0 - 1). \tag{25}$$

Since by Carrasco and Chen (2002, Corollary 6), $\{Y_t\}$ is strongly mixing,

$$\begin{aligned} n^{-1}\sum_t \widehat{X}_{t-1} Y_{t-1} &= n^{-1}\sum_t X_{t-1} Y_{t-1} - (\widehat{\gamma} - \gamma_0)\, n^{-1}\sum_t Y_{t-1} \\ &\xrightarrow{a.s.} E\left(X_{t-1} Y_{t-1}\right) \end{aligned} \tag{26}$$

by the Ergodic Theorem, and, given (25) and (26),

$$\begin{aligned} n^{-1}\sum_t \widehat{X}_t Y_{t-1} &= c_0 n^{-1}\sum_t Y_{t-1} + \phi_0 n^{-1}\sum_t \widehat{X}_{t-1} Y_{t-1} - \beta_0 n^{-1}\sum_t W_{t-1} Y_{t-1} + n^{-1}\sum_t W_t Y_{t-1} \\ &\xrightarrow{a.s.} \phi_0 E\left(X_{t-1} Y_{t-1}\right) - \beta_0 E\left(W_{t-1} Y_{t-1}\right) \\ &= \alpha_0 E\left(X_{t-1} Y_{t-1}\right), \end{aligned}$$

where the final equality follows since

$$E\left(X_{t-1}Y_{t-1}\right) = E\left(\left(\sigma_{t-1}^{2}-W_{t-1}\right)Y_{t-1}\right) = E\left(W_{t-1}Y_{t-1}\right). \tag{27}$$

Next,

$$\begin{aligned} na_{n}^{-3}\left(\widehat{\alpha}_{IV}-\alpha_{0}\right) &= F_{0}\left(a_{n}^{-3}\sum_{t}X_{t}Y_{t-1}-E\left(X_{t}Y_{t-1}\right)\right)+o_{p}\left(1\right) \\ &= F_{0}\left(a_{n}^{-3}\sum_{t}Y_{t}^{2}Y_{t-1}-E\left(Y_{t}^{2}Y_{t-1}\right)-\gamma_{0}a_{n}^{-3}\sum_{t}Y_{t-1}\right)+o_{p}\left(1\right) \\ &= F_{0}\left(a_{n}^{-3}\sum_{t}Y_{t}^{2}Y_{t-1}-E\left(Y_{t}^{2}Y_{t-1}\right)\right)+o_{p}\left(1\right) \\ &\xrightarrow{d} \alpha_{0}^{-1}F_{0}\left(V_{2,\mathbf{y}}-\beta_{0}V_{1,\mathbf{y}}\right), \end{aligned} \tag{28}$$

where "$\xrightarrow{d}$" follows from Lemma 11, which itself depends on Lemmas 9–10, the CLT in Lemma 7, and Lemma 5, and the final equality follows since

$$a_{n}^{-1}\sum_{t}Y_{t-1}\xrightarrow{d}V_{0}, \tag{29}$$

with V_0 being κ_0-stable (see the proof of Lemma 10). Finally, if $\kappa_0 \in (6,\ \infty)$ so that $E\left(Y_t^6\right) < \infty$, then given (5), (9), and (24),

$$\begin{aligned} \sqrt{n}\left(\widehat{\alpha}_{IV}-\alpha_{0}\right) &= \sqrt{n}\left(\frac{n^{-1}\sum_{t}X_{t}Y_{t-1}}{n^{-1}\sum_{t}\widehat{X}_{t-1}Y_{t-1}}-\alpha_{0}+o_{p}\left(1\right)\right) \\ &= \sqrt{n}\left(\beta_{0}+\frac{n^{-1}\sum_{t}V_{t}Y_{t-1}}{E\left(X_{t-1}Y_{t-1}\right)}+o_{p}\left(1\right)\right) \\ &= E\left(Y_{t}^{3}\right)^{-1}\sqrt{n}\left(n^{-1}\sum_{t}V_{t}Y_{t-1}-E\left(V_{t}Y_{t-1}\right)+o_{p}\left(1\right)\right) \\ &\xrightarrow{d} N\left(0,\ E\left(Y_{t}^{3}\right)^{-2}\Sigma_{VY_{-1}}\right) \end{aligned}$$

where the limiting result follows from Ibragimov and Linnik (1971, Theorem 18.5.3).■

Proof of Theorem 5: For notational ease, let $\widehat{\beta}_{IV} = \widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$. Given (24),

$$\widehat{R}_t = R_t - \left(\widehat{\phi} - \phi_0\right) X_{t-1} - (\widehat{\gamma} - \gamma_0)\left(1 - \widehat{\phi}\right). \tag{30}$$

If $\widehat{\phi} \xrightarrow{a.s.} \phi_0$ (as is the case when $\widehat{\phi} = \widehat{\phi}_{IV}$, see Theorem 13 in the Supplemental Appendix), then, given that $\{Y_t\}$ is strongly mixing,

$$\begin{aligned} n^{-1}\sum_t \widehat{R}_{t-1} Y_{t-1} &= n^{-1}\sum_t R_{t-1}Y_{t-1} - \left(\widehat{\phi} - \phi_0\right) n^{-1}\sum_t X_{t-2}Y_{t-1} - (\widehat{\gamma} - \gamma_0)\left(1 - \widehat{\phi}\right) n^{-1}\sum_t Y_{t-1} \\ &\xrightarrow{a.s.} E\left(W_{t-1}Y_{t-1}\right) \end{aligned}$$

and

$$\begin{aligned} n^{-1}\sum_t \widehat{R}_t Y_{t-1} &= n^{-1}\sum_t R_t Y_{t-1} - \left(\widehat{\phi} - \phi_0\right) n^{-1}\sum_t X_{t-1}Y_{t-1} - (\widehat{\gamma} - \gamma_0)\left(1 - \widehat{\phi}\right) n^{-1}\sum_t Y_{t-1} \\ &\xrightarrow{a.s.} -\beta_0 E\left(W_{t-1}Y_{t-1}\right) \end{aligned}$$

by the Ergodic Theorem, since also given (27),

$$\begin{aligned} E\left(R_{t-1}Y_{t-1}\right) &= E\left(X_{t-1}Y_{t-1}\right) - \phi_0 E\left(X_{t-2}Y_{t-1}\right) \\ &= E\left(X_{t-1}Y_{t-1}\right) \end{aligned}$$

and

$$\begin{aligned} E\left(R_t Y_{t-1}\right) &= E\left(X_t Y_{t-1}\right) - \phi_0 E\left(X_{t-1}Y_{t-1}\right) \\ &= E\left(Y_t^2 Y_{t-1}\right) - \phi_0 E\left(Y_{t-1}^2 Y_{t-1}\right) \\ &= E\left(\sigma_t^2 Y_{t-1}\right) - \phi_0 E\left(W_{t-1}Y_{t-1}\right) \\ &= \alpha_0 E\left(Y_{t-1}^2 Y_{t-1}\right) - \phi_0 E\left(W_{t-1}Y_{t-1}\right). \end{aligned}$$

Next, since

$$\widehat{\beta}_{IV} = -\widehat{G}\left(n^{-1}\sum_t R_t Y_{t-1} - \left(\widehat{\phi} - \phi_0\right) n^{-1}\sum_t X_{t-1}Y_{t-1} + o_p(1)\right),$$

then

$$\begin{aligned}
\widehat{\beta}_{IV}-\beta_0 &= -\widehat{G}\left(n^{-1}\sum_t R_tY_{t-1}-E\left(R_tY_{t-1}\right)\right)+\left(\widehat{\phi}-\phi_0\right)\widehat{G}\left(n^{-1}\sum_t X_{t-1}Y_{t-1}\right)\\
&\quad -\left(\widehat{G}-G_0\right)E\left(R_tY_{t-1}\right)\\
&= \left(\widehat{\phi}-\phi_0\right)-\widehat{G}\left(n^{-1}\sum_t X_tY_{t-1}-E\left(X_tY_{t-1}\right)-\phi_0 n^{-1}\sum_t X_{t-1}Y_{t-1}-E\left(X_{t-1}Y_{t-1}\right)\right)+o_p\left(1\right)
\end{aligned}$$

such that

$$\begin{aligned}
na_n^{-3}\left(\widehat{\beta}_{IV}-\beta_0\right) &= na_n^{-3}\left(\widehat{\phi}-\phi_0\right) \qquad (31)\\
&\quad -\widehat{G}\left(a_n^{-3}\sum_t Y_t^2Y_{t-1}-E\left(Y_t^2Y_{t-1}\right)-\phi_0 a_n^{-3}\sum_t Y_{t-1}^3-E\left(Y_{t-1}^3\right)+o_p\left(1\right)\right)\\
&\xrightarrow{d} \alpha_0^{-1}\left(\mathbf{F}_0\mathbf{S}+G_0\left(\alpha_0\phi_0V_{0,Y}-\left(V_{2,\mathbf{y}}-\beta_0V_{1,\mathbf{y}}\right)\right)\right),
\end{aligned}$$

where this last equality follows given (29), and "$\xrightarrow{d}$" follows given Lemmas 11 and 12. That this limit is jointly $(\kappa_0/3)$ $-$stable follows from Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Finally,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_{IV}-\beta_0\right) &= \sqrt{n}\left(\frac{-n^{-1}\sum_t U_tY_{t-1}}{E\left(R_tY_{t-1}\right)}+\left(\widehat{\phi}-\phi_0\right)\frac{E\left(X_{t-1}Y_{t-1}\right)}{E\left(R_tY_{t-1}\right)}+o_p\left(1\right)\right)\\
&= \sqrt{n}\left(-E\left(Y_t^3\right)^{-1}\left(-n^{-1}\sum_t U_tY_{t-1}\right)+\left(\widehat{\phi}-\phi_0\right)+o_p\left(1\right)\right)\\
&\xrightarrow{d} N\left(0,\ \Sigma_\beta\right),
\end{aligned}$$

where "$\xrightarrow{d}$" follows using the CLT from the proof of Theorem 1 as well as Shao and Zhou (2010, Theorem 1).■

Proof of Corollary 8 Almost sure convergence is established in the proof of Theorem 5. From (31),

$$na_n^{-3}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right)-\beta_0\right)=na_n^{-3}\left(\widehat{\phi}-\phi_0\right)+O_p\left(na_n^{-3}\right),$$

where the second term on the right-hand-side of the equality follows from Lemmas 11 and 12. For the first term,

$$na_n^{-3}\left(\widehat{\phi}-\phi_0\right)=\left(n^{\frac{\kappa_0-3}{\kappa_0}-l}\right)n^l\left(\widehat{\phi}-\phi_0\right)=o_p\left(1\right). \qquad (32)$$

Lemmas 11 and 12, again, then establish (17).■

Proof of Corollary 10 Almost sure convergence follows from the proof of Theorem 5 here and the proof of Theorem 13 in the Supplemental Appendix. Given (18),

$$na_n^{-3}\left(\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)-\alpha_0\right)=na_n^{-3}\left(\widehat{\phi}-\phi_0\right)-na_n^{-3}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right)-\beta_0\right). \tag{33}$$

If $\widehat{\phi}=\widehat{\phi}_{IV}$, then (19) is established by Theorem 2 here and Theorem 13 in the Supplemental Appendix, noting that the shared (asymptotic) dependence of $\widehat{\phi}_{IV}$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$ on $\mathbf{S}$ cancels out. Lastly, from (33) and given (32), $\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)$ shares the same distributional limit (excluding a sign change) with $\widehat{\beta}_{IV}\left(\widehat{\phi}\right)$.■

Proof of Theorem 11 Almost sure convergence follows from the proof of Theorem 5. Given (30) and (13),

$$\begin{aligned}\widehat{\beta}_{IV}\left(\widehat{\phi}\right) &= -\widehat{G}\left(n^{-1}\sum_t \widehat{R}_t Y_{t-1}\right) \\ &= \beta_0-\widehat{G}\left(n^{-1}\sum_t U_t Y_{t-1}\right)+\left(\widehat{\phi}-\phi_0\right)\widehat{G}\left(n^{-1}\sum_t X_{t-1}Y_{t-1}\right)+o_p(1)\end{aligned} \tag{34}$$

Recalling also that $R_t=X_t-\phi_0X_{t-1}$, from the proof of Theorem 5,

$$\begin{aligned}n^{-1}\sum_t \widehat{R}_{t-1}Y_{t-1} &= n^{-1}\sum_t X_{t-1}Y_{t-1}-\widehat{\phi}n^{-1}\sum_t X_{t-2}Y_{t-1}-\left(\widehat{\gamma}-\gamma_0\right)\left(1-\widehat{\phi}\right)n^{-1}\sum_t Y_{t-1} \\ &= n^{-1}\sum_t X_{t-1}Y_{t-1}+o_p(1),\end{aligned}$$

in which case,

$$\widehat{G}\left(n^{-1}\sum_t X_{t-1}Y_{t-1}\right)=1, \qquad n\to\infty.$$

Next, given the definitions of V_t and U_t in (5) and (13), respectively,

$$U_t=-\beta_0^2W_{t-2}+\beta_0W_{t-1}+V_t.$$

Consequently,

$$
\begin{aligned}
\widehat{G}\left(n^{-1}\sum_t U_t Y_{t-1}\right) &= \frac{n^{-1}\sum_t V_t Y_{t-1} + \beta_0 n^{-1}\sum_t W_{t-1}Y_{t-1} + o_p(1)}{n^{-1}\sum_t X_{t-1}Y_{t-1} + o_p(1)} \\
&= \frac{n^{-1}\sum_t V_t Y_{t-1}}{n^{-1}\sum_t X_{t-1}Y_{t-1}} + \beta_0 + o_p(1) \\
&= \left(n^{-1}\sum_t X_{t-1}Y_{t-1}\right)^{-1} \times \{n^{-1}\sum_t V_t Y_{t-1} - E\left(V_t Y_{t-1}\right) \\
&\quad +\beta_0\left(n^{-1}\sum_t X_{t-1}Y_{t-1} - E\left(W_{t-1}Y_{t-1}\right)\right)\} \\
&= E\left(Y_t^3\right)^{-1}\left(n^{-1}\sum_t V_t Y_{t-1} - E\left(V_t Y_{t-1}\right) + o_p(1)\right) \\
&= \left(\widehat{\alpha}_{IV} - \alpha_0\right),
\end{aligned}
$$

where given (5), the third equality relies on $E\left(V_t Y_{t-1}\right) = -\beta_0 E\left(W_{t-1}Y_{t-1}\right)$ and $E\left(X_{t-1}Y_{t-1}\right) = E\left(W_{t-1}Y_{t-1}\right)$, and the fourth equality follows from the proof of Theorem 1. Then, given (34),

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right) - \beta_0\right) &= \sqrt{n}\left(\widehat{\phi} - \phi_0\right) - \sqrt{n}\left(\widehat{\alpha}_{IV} - \alpha_0\right) + o_p(1) \\
&= \sqrt{n}\left(\widehat{\beta} - \beta_0\right) + \sqrt{n}\left(\widehat{\alpha} - \alpha_0\right) - \sqrt{n}\left(\widehat{\alpha}_{IV} - \alpha_0\right) + o_p(1)
\end{aligned} \tag{35}
$$

Since $\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0,\ \Sigma_\theta\right)$ by assumption, and $\sqrt{n}\left(\widehat{\alpha}_{IV} - \alpha_0\right) \xrightarrow{d} N\left(0,\ \Sigma_\alpha^*\right)$ by Theorem 1,

$$
\sqrt{n}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right) - \beta_0\right) \xrightarrow{d} N\left(0,\ \Sigma_\beta^*\right),
$$

where

$$
\begin{aligned}
\Sigma_\beta^* &= \Sigma_\beta + \Sigma_\alpha + 2\Sigma_{\alpha,\ \beta} + \Sigma_\alpha^* - 2\left(\Omega_\alpha^* + \Sigma_{\alpha,\ \beta}^*\right) \\
&< \Sigma_\beta + \Sigma_\alpha^* - 2\left(\Omega_\alpha^* + \Sigma_{\alpha,\ \beta}^*\right) \\
&< \Sigma_\beta.
\end{aligned}
$$

■

Proof of Corollary 13 From (18),

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)-\alpha_0\right) &= \sqrt{n}\left(\widehat{\phi}-\widehat{\beta}_{IV}\left(\widehat{\phi}\right)-\alpha_0\right) \\
&= \sqrt{n}\left(\widehat{\phi}-\widehat{\beta}_{IV}\left(\widehat{\phi}\right)-\alpha_0-\beta_0+\beta_0\right) \\
&= \sqrt{n}\left(\widehat{\phi}-\phi_0\right)-\sqrt{n}\left(\widehat{\beta}_{IV}\left(\widehat{\phi}\right)-\beta_0\right) \\
&= \sqrt{n}\left(\widehat{\alpha}_{IV}-\alpha_0\right)+o_p\left(1\right),
\end{aligned}
$$

where the final equality follows from (35). Then from Theorem 1,

$$
\sqrt{n}\left(\widehat{\alpha}_{IV}\left(\widehat{\phi}\right)-\alpha_0\right)\xrightarrow{d} N\left(0,\ \Sigma_{\alpha}^{*}\right),
$$

which completes the proof.■

TABLE 1: Skewness Estimates

	JPY Returns		SPX Returns		DJIA Returns	
freq.	obs.	skew.	obs.	skew.	obs.	skew.
1-min	174,997	-2.68	46,551	-1.75	46,557	-1.25
		(0.01)		(0.01)		(0.01)
5-min	35,028	-1.94	9,312	-3.17	9,315	-2.68
		(0.01)		(0.03)		(0.03)
10-min	17,523	-1.51				
		(0.02)				
15-min	11,685	-3.10				
		(0.02)				
20-min	8,766	-2.10				
		(0.03)				

Notes to Tables 1. The data source is Bloomberg Finace LP. JPY is the Yen/USD exchange rate. SPX and DJIA is the S&P 500 and Dow Jones Industrial Average, respectively. The date range for all return series is 7/19/2015–12/31/2015. Skew is the standard estimate of the (unconditionally) standardized third moment. Standard errors for the skewness, measured against the null of normality, are in parentheses .

TABLE 2: Simulation Designs

	$\eta = 64.5$		$\eta = 8.1$		$\eta = 4.5$		$\eta = 3.5$	
λ	skew.	κ	skew.	κ	skew.	κ	skew.	κ
-0.20	-0.34	7.25	-0.56	5.15	-1.10	3.76	-1.92	3.14
-0.40	-0.65	6.60	-1.06	4.73	-2.05	3.53	-3.50	2.98
-0.80	-1.03	6.08	-1.67	4.37	-3.18	3.29	-5.26	2.81

Notes to Tables 2. Reported are the skewness statistics and tail index values for the Monte Carlo simulation designs that study the linear GARCH$(1,1)$ model when $\omega = 0.005$, $\alpha = 0.10$, and $\beta = 0.80$. The rescaled errors from this model follow the skewed student's t density of Hansen (1994) normalized so that $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$. This density has two parameters, λ and η, with the former governing skewness, the latter governing the tails, and up to the ηth moment being well defined. Both the skewness statistics and tail index values κ for the simulated raw returns are themselves determined through simulation as the mean estimate from $\{Y_t\}_{t=1}^{10{,}000}$ across $10,000$ trials for the given design. The estimator for κ is Hill (1975) with a constant threshold of 0.5%.

TABLE 3: Large Sample Results I

est.	m	$\lambda=-0.20$ efficiency ratio rmse	mae	mdae	$\lambda=-0.40$ efficiency ratio rmse	mae	mdae	$\lambda=-0.80$ efficiency ratio rmse	mae	mdae
I. $\eta=64.5$										
$\widehat{\alpha}_{IV}$	1	7.35	7.27	7.22	4.23	4.17	4.09	3.15	3.05	2.99
$\widehat{\alpha}_{KL}$	4	3.00	2.85	2.72	3.46	3.19	2.98	4.23	3.76	3.41
$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	8.16	8.12	8.03	4.27	4.24	4.22	2.71	2.68	2.65
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$	20	4.96	4.87	4.78	4.04	3.84	3.66	4.66	4.13	3.71
	10	4.80	4.72	4.64	3.83	3.59	3.41	4.36	3.73	3.33
	5	5.11	5.05	4.98	4.19	4.03	3.85	4.67	4.21	3.88
$\widehat{\beta}_{KL}$	4	4.49	4.39	4.25	5.10	4.92	4.71	6.08	5.77	5.40
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	3.91	3.90	3.88	2.18	2.17	2.16	1.54	1.53	1.50
II. $\eta=8.1$										
$\widehat{\alpha}_{IV}$	1	7.14	6.82	6.61	4.74	4.31	4.02	4.24	3.53	3.25
$\widehat{\alpha}_{KL}$	4	5.65	4.96	4.33	6.21	5.43	4.68	6.59	5.67	5.06
$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	7.89	7.52	7.25	4.63	4.25	3.95	3.46	3.03	2.74
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$	20	8.94	7.40	6.10	10.69	8.30	6.17	13.42	10.94	8.32
	10	8.21	6.76	5.64	9.64	7.20	5.06	11.96	9.16	6.23
	5	8.49	7.18	6.28	9.32	7.27	5.61	11.13	8.55	6.16
$\widehat{\beta}_{KL}$	4	7.86	7.40	6.78	8.50	8.00	7.33	9.06	8.50	8.02
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	4.07	3.85	3.65	2.49	2.29	2.12	2.02	1.79	1.64

TABLE 4: Large Sample Results II

est.	m	$\lambda=-0.20$ efficiency ratio rmse	mae	mdae	$\lambda=-0.40$ efficiency ratio rmse	mae	mdae	$\lambda=-0.80$ efficiency ratio rmse	mae	mdae
III. $\eta=4.5$										
$\widehat{\alpha}_{IV}$	1	7.83	6.92	6.34	5.18	4.65	4.36	3.86	3.84	3.76
$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	8.86	7.83	7.13	5.20	4.71	4.25	3.32	3.36	3.09
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$	20	17.60	16.65	13.91	17.41	17.34	15.08	16.97	17.89	16.99
	10	15.98	14.36	10.88	15.60	14.64	11.80	15.11	15.04	13.29
	5	14.62	12.79	9.56	14.00	12.56	9.21	13.29	12.46	9.68
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	5.25	4.61	4.03	3.27	2.83	2.42	2.34	2.16	1.91
IV. $\eta=3.5$										
$\widehat{\alpha}_{IV}$	1	3.87	5.30	5.21	2.57	3.51	3.87	1.83	2.82	3.45
$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	4.29	5.94	5.71	2.63	3.45	3.58	1.52	2.33	2.59
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$	20	13.97	16.72	18.18	12.73	15.80	17.85	11.65	14.82	17.50
	10	12.63	14.38	14.58	11.45	13.49	14.40	10.49	12.69	14.21
	5	11.23	12.15	10.93	10.11	11.20	10.44	9.12	10.33	10.20
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	4.16	4.07	3.75	2.43	2.45	2.28	1.69	1.80	1.76

Notes to Tables 3–4. Simulations are conducted on sample sizes of $T = 100,000$ across $10,000$ trials, where, within each trial, the first 200 observations are dropped to avoid initialization effects. The linear GARCH$(1,1)$ model under study is parameterized as

$$\omega_0 = 0.005, \qquad \alpha_0 = 0.10, \qquad \beta_0 = 0.80.$$

The simple IV estimators are considered, along with the Kristensen and Linton (2006) estimator (KL) and quasi-maximum likelihood estimator (QMLE), both of which serve as benchmarks. For the simple IV estimators, m is the number of lagged instruments used. The innovations from the GARCH$(1,1)$ model follow Hansen's (1994) skewed student's-t density, where

$$\lambda = -0.20, -0.40, -0.80, \qquad \eta = 64.5, 8.1, 4.5, 3.5.$$

Higher values of λ correspond with more skewness (in absolute terms) in $\{Y_t\}_{t=1}^{T}$, while higher values of η correspond with heavier tails in the simulated return sample. In the thin-tailed case of $\eta = 64.5$, $\widehat{\alpha}_{IV}$, $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, $\widehat{\alpha}_{QMLE}$, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$, $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, and $\widehat{\beta}_{QMLE}$ are all asymptotically normal, while $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ (likely) are not.[18] In the heavy-tailed cases of $\eta = 8.1, 4.5$, only $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{QMLE}$ are asymptotically normal. In the (very) heavy-tailed case of $\eta = 3.5$, none of the estimators are asymptotically normal.[19]. In the heavy-tailed cases of $\eta = 4.5, 3.5$, $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ are not consistent ; consequently, they are not considered in those cases.[20] Summary statistics for the simulations are the root mean squared error, mean absolute error, and median absolute error (each measured relative to the true parameter value) divided by the corresponding efficiency measure for the QMLE. These ratios are termed "efficiency ratios."

[18]Necessary for $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ to be asymptotically normal is $E\left(Y_t^8\right) < \infty$ (see Kristensen and Linton, 2006), which does not appear to be true, given the results in Table 2.

[19]A necessary condition for $\widehat{\alpha}_{QMLE}$ and $\widehat{\beta}_{QMLE}$ to be asymptotically normal is $E\left(\epsilon_t^4\right) < \infty$ (see; e.g., Hall and Yao, 2003).

[20]Necessary for consistency of $\widehat{\alpha}_{KL}$ and $\widehat{\beta}_{KL}$ is $E\left(Y_t^4\right) < \infty$, which (very likely) does not hold, given the results in Table 2.

TABLE 5: Thin-Tailed Efficiency Comparisons (Large Sample)

λ	η	skew.	κ	α	β	ϕ	efficiency ratio rmse	mae	mdae
-0.90	16.5	-1.21	6.03	0.05	0.50	0.55	1.00	0.99	0.99
-0.99	16.5	-1.24	5.99	0.05	0.50	0.55	1.00	0.99	0.99
-0.90	20.5	-1.17	6.04	0.10	0.50	0.60	1.01	1.00	1.00
-0.99	20.5	-1.21	6.00	0.10	0.50	0.60	1.01	1.00	1.00
-0.90	16.5	-1.21	6.01	0.05	0.60	0.65	0.99	0.99	0.98
-0.99	16.5	-1.25	5.97	0.05	0.60	0.65	0.99	0.99	0.98
-0.90	21.5	-1.17	6.01	0.10	0.60	0.70	1.03	1.01	1.00
-0.99	21.5	-1.20	5.97	0.10	0.60	0.70	1.03	1.01	1.00
-0.90	18.5	-1.19	6.07	0.05	0.80	0.85	1.07	1.01	1.01
-0.99	18.5	-1.22	6.03	0.05	0.80	0.85	1.07	1.01	1.01
-0.90	64.5	-1.08	6.01	0.10	0.80	0.90	1.50	1.49	1.49
-0.99	64.5	-1.11	5.96	0.10	0.80	0.90	1.49	1.48	1.47
-0.90	24.5	-1.15	6.04	0.05	0.90	0.95	1.68	1.68	1.71
-0.99	24.5	-1.19	6.00	0.05	0.90	0.95	1.65	1.66	1.67

Table 6: Small Sample Results

		$\lambda = -0.20$ efficiency ratio			$\lambda = -0.40$ efficiency ratio			$\lambda = -0.80$ efficiency ratio		
est.	m	rmse	mae	mdae	rmse	mae	mdae	rmse	mae	mdae
III. $\eta = 4.5$										
$\widehat{\alpha}_{IV}$	1	2.41	2.49	2.05	1.17	1.31	1.48	0.68	0.85	1.15
$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	2.34	2.52	2.18	1.18	1.34	1.50	0.61	0.77	1.03
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$	20	2.69	3.95	7.06	2.77	4.14	7.35	2.66	3.95	6.95
	10	2.50	3.60	6.38	2.54	3.73	6.62	2.44	3.55	6.28
	5	2.22	3.07	5.30	2.25	3.15	5.56	2.15	3.00	5.33
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	1.29	1.47	1.74	1.06	1.15	1.37	0.97	0.99	1.09
IV. $\eta = 3.5$										
$\widehat{\alpha}_{IV}$	1	1.63	1.73	1.62	0.86	0.98	1.20	0.54	0.68	1.01
$\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	1.60	1.79	1.76	0.82	0.97	1.19	0.45	0.58	0.85
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$	20	2.51	3.61	6.61	2.55	3.74	6.70	2.43	3.54	6.29
	10	2.35	3.33	6.07	2.37	3.42	6.16	2.25	3.24	5.80
	5	2.11	2.87	5.17	2.11	2.91	5.28	2.01	2.78	5.04
$\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$	1	1.20	1.33	1.55	1.02	1.08	1.24	0.95	0.95	1.00

Notes to Table 5. Simulations are conducted on sample sizes of $T = 100,000$ across $10,000$ trials, where, within each trial, the first 200 observations are dropped to avoid initialization effects. The linear GARCH$(1,1)$ model is studied under different parameter values for α_0 and β_0 and different specifications of Hansen's (1994) skewed student's-t density for the model's innovations (see the Table).[21] The estimators considered are $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{QMLE}$. Different specifications of Hansen's (1994) skewed student's-t density are selected to maximize the amount (in absolute terms) of skewness in $\{Y_t\}_{t=1}^T$, while maintaining asymptotic normality for $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$. "skew." and κ are the skewness and tail index value of $\{Y_t\}_{t=1}^T$, respectively, under the given simulation design, while $\phi = \alpha + \beta$. Summary statistics for the simulations are the root mean squared error, mean absolute error, and median absolute error (each measured relative to the true parameter value) for $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ divided by the corresponding efficiency measure for $\widehat{\beta}_{QMLE}$, with these ratios being termed "efficiency ratios," as in Tables 3 and 4.

Notes to Table 6. Simulations are conducted on sample sizes of $T = 500$ across $10,000$ trials, where, within each trial, the first 200 observations are dropped to avoid initialization effects. The linear GARCH$(1,1)$ model under study is parameterized as

$$\omega_0 = 0.005, \qquad \alpha_0 = 0.10, \qquad \beta_0 = 0.80.$$

The simple IV estimators are considered (where m is the number of lagged instruments used to construct the estimator), along with the quasi-maximum likelihood estimator (QMLE), which serves as a benchmark. The innovations from the GARCH$(1,1)$ model follow Hansen's (1994) skewed student's-t density, where

$$\lambda = -0.20, -0.40, -0.80, \qquad \eta = 4.5, 3.5.$$

Higher values of λ correspond with more skewness (in absolute terms) in $\{Y_t\}_{t=1}^T$, while higher values of η correspond with heavier tails in the simulated return sample. In these small-sample experiments, only (very) heavy-tailed innovation densities are considered. Summary statistics for the simulations are the root mean squared error, mean absolute error, and median absolute error (each measured relative to the true parameter value) divided by the corresponding efficiency measure for the QMLE. These ratios are termed "efficiency ratios."

[21] In all cases, $\omega_0 = 0.005$.

TABLE 7: GARCH Model Estimates I

			tail		estimator			
					TSLS			QMLE
freq.	obs.	skew.	index	para.	$m = 5$	$m = 10$	$\widehat{\phi}_{QMLE}$	
15-min	12,680	-1.27	2.93	α	0.15	0.15	0.19	0.20
		(0.02)						(0.06, 0.34)
				ϕ	0.32	0.25	0.90	0.94
				β	0.17	0.10	0.71	0.74
								(0.58, 0.89)
10-min	19,021	-1.33	2.83	α	0.19	0.19	0.17	0.12
		(0.02)						(0.05, 0.20)
				ϕ	0.53	0.53	0.94	0.95
				β	0.35	0.35	0.77	0.83
								(0.73, 0.93)

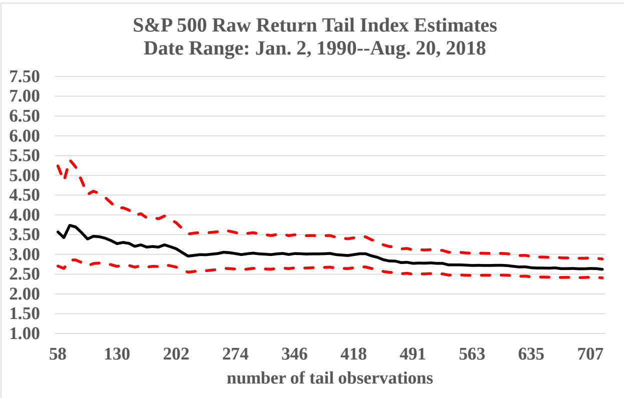
TABLE 8: GARCH Model Estimates II

			tail		estimator			
					TSLS			QMLE
freq.	obs.	skew.	index	para.	$m = 5$	$m = 10$	$\widehat{\phi}_{QMLE}$	
5-min	38,035	-1.40	2.83	α	0.09	0.09	0.15	0.12
		(0.01)						(0.08, 0.17)
				ϕ	0.77	0.78	0.96	0.99
				β	0.64	0.65	0.82	0.87
								(0.82, 0.91)
1-min	190,058	-1.81	3.05	α	0.03	0.03	0.03	0.13
		(0.01)						(0.08, 0.18)
				ϕ	0.84	0.91	1.04	1.00
				β	0.86	0.94	1.01	0.87
								(0.82, 0.91)

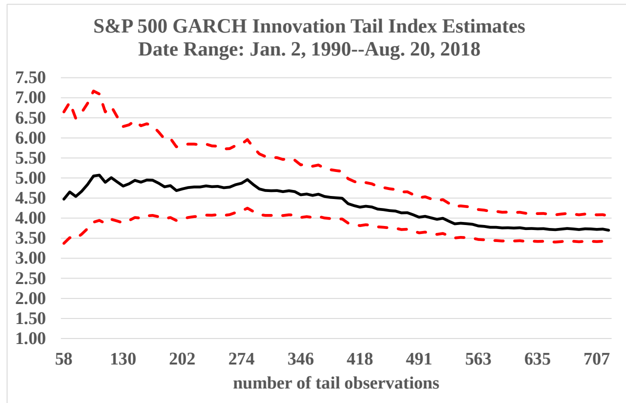
Notes to Tables 7–8. All data ranges from January 1, 2015 through July 1, 2015 and sources to Bloomberg Finance LP. Log return data is intra-daily at the stated frequency measured from traded Japanese Yen exchange rates relative to the USD. Using the approach in Hecq, Laurent, and Palm (2012, Eq. 4.1), all log returns are pre-filtered for the U-shaped intra-day periodicity noted by Anderson and Bollerslev (1997). Skew is the unconditional skewness of the log returns. Moving from left to right in the columns under the TSLS heading, the first two columns show $\widehat{\alpha}_{IV}$, $\widehat{\phi}_{IV}$, and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$, respectively. For $\widehat{\alpha}_{IV}$, it is always the case that $m = 1$ (where m denotes the number of lags used as instruments). For $\widehat{\phi}_{IV}$, and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{IV}\right)$, it is either the case that $m = 5$ or $m = 10$. The third column shows $\widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right)$ and $\widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$, where, it is also always that case that $m = 1$, and, additionally, $\phi = \widehat{\alpha}_{IV}\left(\widehat{\phi}_{QMLE}\right) + \widehat{\beta}_{IV}\left(\widehat{\phi}_{QMLE}\right)$. The final column of the Table shows estimates from the QMLE, together with the lower- and upper-bounds of their associated 95% confidence interval.

FIGURE 1

PANEL A



PANEL B



Notes to Figure1: Hill (1975) tail index estimates at varying thresholds are depicted for S&P 500 Index log returns and the innovations from a GARCH(1,1) model applied to these returns. Th thresholds are determined as proportions of the ranked data ranging from 1% to 10%. The underlying price data is daily and sources to Bloomberg Finance LP.