Simple Estimators for ARCH Models[1]

Todd Prono[2]

This Version: November 2016

Abstract

Covariances between contemporaneous squared returns and lagged returns form the basis for closed-form instrumental variables estimators of ARCH processes. These simple estimators rely on asymmetry for identification (either in the model's rescaled errors or the conditional variance function) and apply to threshold ARCH(1) and ARCH(p) with $p < \infty$ processes. Stable distributional limits for these estimators are determined in cases where the returns are regularly varying with well-defined third moments. These limits are highly non-normal in empirically relevant cases with slow rates of convergence. Surprisingly, Monte Carlo studies of heavy-tailed ARCH(1) processes show the simple IV estimator to outperform standard QMLE in both small and (relatively) moderate sample sizes. This outperformance is the most pronounced when the returns are heavily skewed.

Keywords: ARCH, closed form estimation, heavy tails, instrumental variables, regular variation, three-step estimation. JEL codes: C13, C22, C58.

[1] I owe thanks to Dennis Kristensen, Arthur Lewbel, Travis Nesmith, and Dong Hwan Oh, as well as to participants at the 2015 Annual Conference of the Royal Economic Society, the 2015 Meeting of the Midwest Econometrics Group, the Federal Reserve Board and the OCC for helpful comments and discussions. 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

This paper considers estimation of the finite-order ARCH models originated in Engle (1982). A class of simple, instrumental variables (IV) estimators for threshold ARCH(1) and ARCH(p) with $p < \infty$ models is established on the basis that the (high frequency) financial returns to which these models are often applied tend to be skewed. This well-known but (up until this point) under-used stylized fact of financial returns is shown to support simple estimators that are (almost surely) consistent with stable limiting distributions regardless of whether kurtosis in the returns is well-defined. Identification of these estimators sources to the covariances between squared returns and past level returns (hereafter referred to as cross-order covariances). For ARCH(p) models, this sourcing requires the model's rescaled errors to be skewed, while in the threshold ARCH(1) case, the requirement is for the conditional variance function to be truly asymmetric (in which case, the model's rescaled errors may or may not be skewed).

In demonstrating the large-sample properties of these simple estimators, it is useful to consider the ARCH(1) and threshold ARCH(1) models first, before moving to the general ARCH(p) case. The reason for this approach is that simple estimators of the former models have assumptions that are straightforward to verify and limiting results that only involve directly observable variables, while simple estimators in the latter case involve more complicated assumptions and limiting results, owing, in turn, to the more complex structure of the cross-order covariances when $p > 1$. However, the limiting results for simple estimators of ARCH(p) models remain (at least) qualitatively similar to the ARCH(1) case with the same rate of convergence.[3]

Before proceeding further, it is necessary to define what is meant by a simple estimator. That definition, initially given in Lewbel (2004) and later applied in Dong and Lewbel (2015), requires a simple estimator to (1) "closely resemble (or consist of steps that each resemble) estimators that are already in common use," and involve (2) "few or no numerical searches or numerical maximizations." Consistent with this definition, this paper considers ordinary least squares (OLS) and linear two-stage least squares (TSLS) estimators for ARCH processes, each of which is available in closed form. Simple estimators for the more popular class of GARCH models (introduced in Bollerslev, 1986) are not considered; however, what is learned from studying the ARCH cases both in terms of the conditions that support identification and the discovered distributional limits also apply in the GARCH context as discussed in Prono (2014) and introduced in this paper's conclusion.

[3]It is also the case that the results for ARCH(p) estimators nest the ARCH(1) case.

Almost by definition, a simple estimator is (likely) not an efficient estimator. Yet, simple estimators deserve attention because they (1) facilitate application of the bootstrap (and other resampling techniques) for determining confidence intervals, (2) avoid numerical pitfalls like flat objective functions and multiple local maxima, (3) provide consistent starting values for (more) efficient estimators (see; e.g., Francq, Lepage and Zakoïan, 2011 and Fan, Qi and Xiu, 2014, in a (G)ARCH context), (4) provide insights into the formulation of more efficient estimators (e.g., by revealing a set of over-identifiying restrictions that can be used in Hill's and Prokhorov's (2016) generalized empirical likelihood estimator), and (5) enjoy faster computation times. This first criterion is particularly relevant, since (while stable) the limiting distributions of the OLS and TSLS estimators are functionals of infinite variance random vectors, making the estimation of these distributions difficult. Given suitable normalizations, however, confidence intervals for these estimators can be obtained using the bootstrap methods of Hall and Yao (2003). The last criterion is relevant for applications requiring (many) forecasts quickly; for example, high frequency trading algorithms involving intra-day returns or (very) high dimensional conditional Value-at-Risk (VaR) estimates used to determine initial margin requirements for cleared derivatives.

1.2 Background and Motivation

For the ARCH(p) model

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2, \qquad \epsilon_t \sim i.i.d.\ D\left(0,\ 1\right), \qquad p < \infty, \tag{1}$$

where D is some zero mean, unit variance distribution, consider the simplest case where $p = 1$. In this case, it is well-known that

$$Y_t^2 = \omega + \alpha Y_{t-1}^2 + W_t, \tag{2}$$

where $\{W_t\}$ is a martingale difference sequence (MDS), which is to say that the ARCH(1) model is an AR(1) model of the second-order sequence. Based only upon information from this second-order sequence, a simple estimator for this model, then, is OLS. Weiss (1986) is among the first to consider this case. Guo and Phillips (2001) expand upon this case by considering IV estimation of (2) based on results from Kuersteiner (2002), where the instrument for Y_{t-1}^2 is an infinite, weighted sum of W_{t-1-i} for $i \geq 0$. Similarly, Giratis and Robinson (2001) and Mikosch and Straumann (2002) consider Whittle estimation for ARCH processes, which is asymptotically equivalent to constrained least squares and available in closed form when the spectral density of Y_t^2 exists.

Let $\gamma \equiv E\left(Y_t^2\right)$. From (2) follows that

$$E\left(\left(Y_t^2-\gamma\right)\left(Y_{t-m}^2-\gamma\right)\right)=\alpha^m E\left(\left(Y_t^2-\gamma\right)^2\right), \qquad m \geq 1, \tag{3}$$

provided $E\left(Y_t^4\right) < \infty$. Estimation of (2) using either Y_{t-1}^2 as the instrument (i.e., OLS) or a weighted sum of past Y_{t-1-i}^2 then selects $\widehat{\alpha}$ (the finite sample estimator of α) as the one which best accommodates the sample autocovariances of Y_t^2. Baillie and Chung (1999) argue in favor of such estimation techniques over the popular quasi-maximum likelihood estimator (QMLE), precisely because the former produces the best fit to the sample (second-order) autocovariances, while the latter is known to under-represent those same, sample autocovariances in instances where the model's rescaled errors exhibit heavy tails (see; e.g., Jacquier, Polson, and Rossi, 1994). It seems sensible, then, to craft simple ARCH estimators based on (3), where these simple estimators might perform well against the QMLE (as noted by; e.g., Bollerslev and Wooldridge, 1992, and evidenced in Baillie and Chung) in instances of excess kurtosis. There is also, though, certainly a limit to just how heavy-tailed a sequence these simple estimators can accommodate, since $E\left(Y_t^4\right) < \infty$ is required for consistency. Empirical evidence from financial returns leaves the prospect of this limit not being exceeded far from guaranteed.

Figure 1 plots the Hill (1975) tail index estimator together with 95% confidence bands from Hill (2010, Theorem 4) for three major currency returns (all measured relative the USD) sampled at 20-minute intervals. 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 $p < \kappa$ (see; e.g., Resnick, 1987, for an introduction to regular variation), empirical evidence does not (strongly) support well-defined fourth moments for these currency returns. To the contrary, for substantial sections of all three plots, even the confidence bands do not include 4. Moreover, currency returns sampled at this (very) high frequency are known to display relatively less volatility persistence (and, hence, relatively thinner tails), then returns measured at lower frequencies, like hourly or daily (see; e.g., Anderson and Bollerslev, 1997). For daily equity and FX returns, tail index estimates tend to be even less (see; e.g., Hill and Renault, 2012, Embrechts, Klüppelberg, and Mikosch, 1997, and Loretan and Phillips, 1994), sparking the conclusion that the fourth moments of these returns are not well-defined. Jondeau and Rockinger (2003) offer a (somewhat) softer view by identifying shorter intervals of time over which the fourth moments of financial returns do appear finite;

however, these same authors still discover numerous intervals over which the tails of financial time series appear heavier than can accommodate well-defined fourth moments. Overall then, it is clear that the $\sqrt{n}$ asymptotics (developed first by Weiss, 1986) for OLS applied to (2) are inconsistent with empirical findings, since those asymptotics require $E\left(Y_t^8\right) < \infty$. Moreover, it is (at least) questionable whether the OLS estimator is even consistent.

While not offering much to support well-defined fourth moments, Figure 1 does tend to support well-defined third moments. Notice that the tail index estimates for all three returns stay close to 3, and the confidence bands always include 3. Loretan and Phillips (1994) and Jondeau and Rockinger (2003) present comparable findings for daily FX and equity returns. Cont and Kan (2011, Property 3) report $\kappa \in (3,\ 6)$ for daily, credit default swap spread returns. Bouchaud and Potters (2003, p. 102) state that "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)." These same authors also report evidence supporting a comparable conclusion for (major) FX returns.

For the three currency returns in Figure 1 (JPY, EUR, and CHF), skewness is -0.32, 0.20, and 0.42, respectively, each of which is highly significant against a null of normality given the, respective, sample sizes. Table 1 illustrates additional cases where, not only is the evidenced skewness highly significant, but also quite large in absolute terms. In general, skewness in (high frequency) financial returns is prevalent enough to be considered a stylized fact along with heavy tails. This stylized fact unveils a second set of covariances based on (2) from which simple ARCH estimators can be constructed. These cross-order covariances are

$$E\left(\left(Y_t^2 - \gamma\right) Y_{t-m}\right) = \alpha^m E\left(Y_t^3\right). \tag{4}$$

Let $\mathbf{Z}_{t-1} = \left(Y_{t-1}, \ldots, Y_{t-h}\right)'$ for $h < \infty$. Given (4), $\mathbf{Z}_{t-1}$ is a valid set of instruments for Y_{t-1}^2 in (2) provided that $E\left(Y_t^3\right) \neq 0$ (see Lewbel, 1997, where skewness is also used to define valid instruments, but in a measurement error context). The appeal of a simple estimator for the ARCH(1) model that is based both on (2) and on $\mathbf{Z}_{t-1}$ is completely analogous to the appeal of the (second-order) autocovariance estimator of Baillie and Chung (which, itself, is a generalization of the OLS estimator): by being fit to a particular empirical feature of the data (in this instance, a set of cross-order covariances that map to skewness in the underlying returns), this IV estimator, too, might perform well against the QMLE in instances where this feature strays from what is predicted under normality. In support of this assertion, Monte Carlo results in Section 3 evidence the proposed

TSLS estimator to outperform the QMLE in instances where the data are (heavily) skewed. Another advantage of this TSLS estimator over OLS (and alternative estimators based on (3), generally) is that the former only requires $E\left(Y_t^3\right) < \infty$ for consistency.

The distributional limit of the proposed IV estimator is based on asymptotic theory developed for the sample cross-order covariances from (4), in the case where Y_t is regularly varying. This asymptotic theory extends results from Davis and Mikosch (1998) and Mikosch and Stărică (2000), who study the large sample properties of the sample, second-order autocovariances from (3). A unifying feature of the limit theory for both sample, cross-order covariances and sample, second-order autocovariances is reliance upon a non-standard central limit theorem (CLT) developed in Davis and Mikosch (1998) for regularly varying and dependent sequences.[4] Necessary for the applicability of this CLT in the present context is a demonstration that Y_t is regularly varying when $E\left(Y_t^3\right) \neq 0$, which is provided in this paper's Supplemental Appendix (proof of Lemma 3) and complements results in Basrak, Davis, and Mikosch (2002). For completeness, the distributional limit of the OLS estimator applied to (2) is also established, where symmetry in the rescaled ARCH errors is also not required.

In the case of the model in (2), skewness in $\{Y_t\}$ renders $\mathbf{Z}_{t-1}$ a valid set of instruments. In the case of a threshold ARCH(1) model, a slightly different story emerges. Consider

$$\sigma_t^2 = \omega + \alpha_1 Y_{t-1}^2 \times I_{\{Y_t \geq 0\}} + \alpha_2 Y_{t-1}^2 \times I_{\{Y_t < 0\}}, \tag{5}$$

which is the threshold ARCH(1) model of Glosten, Jagannathan, and Runkle (1993). For this model, a valid set of instruments turns out to be

$$\mathbf{Z}_{t-1} = \left(\left(Z_{1,t-1}, Z_{2,t-1}\right), \ldots, \left(Z_{1,t-h}, Z_{2,t-h}\right)\right)', \tag{6}$$

where

$$\begin{aligned} Z_{1,t-m} &= Y_{t-m} \times I_{\left\{Y_{t-m} \geq 0\right\}} - E\left(Y_t \times I_{\{Y_t \geq 0\}}\right), \\ Z_{2,t-m} &= Y_{t-m} \times I_{\left\{Y_{t-m} < 0\right\}} - E\left(Y_t \times I_{\{Y_t < 0\}}\right), \end{aligned}$$

for $m = 1, \ldots, h$. Validity of these instruments, which relates to a generalization of (4), chiefly

[4] Other works that rely upon this CLT for establishing the large sample properties of ARCH estimators include Mikosch and Straumann (2002), Hall and Yao (2003), and Vaynman and Beare (2014).

depends on asymmetry in the conditional variance and not skewness in the underlying returns. For instance, if $\alpha_1 \neq \alpha_2$, then identification of the IV estimator based on the instruments in (6) follows even if $E\left(Y_t^3\right) = 0$. In the special case where $\alpha_1 = \alpha_2$ (i.e., there is no conditional variance asymmetry), validity of $\mathbf{Z}_{t-1}$ as defined in (6) defaults back to requiring $E\left(Y_t^3\right) \neq 0$.

2.1. The ARCH(1) Case

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

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega_0 + \alpha_0 Y_{t-1}^2, \tag{7}$$

where 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. From (7), note that

$$\sigma_t^2 = \omega_0 + \sigma_{t-1}^2 A_t, \qquad A_t = \alpha_0 \epsilon_{t-1}^2, \tag{8}$$

which characterizes σ_t^2 as a stochastic recurrence equation (SRE). From Basrak, Davis, and Mikosch (2002), most GARCH processes can be characterized as SREs and, as such, shown to be regularly varying.

ASSUMPTION A1: *(i) The sequence $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is i.i.d. $D(0,\ 1)$ for some distribution D with unbounded support. (ii) $E\left|\epsilon_t\right|^j = c_j < \infty$ for $j > k$.*

Given A1(i), (7) reflects the strong ARCH model of Drost and Nijman (1993). Having i.i.d. rescaled errors is necessary for establishing the distributional limits and rates of convergence of the simple estimators under study. Consistency of these estimators, however, continues to follow under the semi-strong definition of ARCH, where (weak) dependence in the higher moments of the rescaled errors is allowed.

Under A1(ii), $\{\epsilon_t\}$ is relatively light-tailed, meaning that heavy-tailed features of $\{Y_t\}$ stem from $\{\sigma_t\}$. It is this distinction between the tail properties of $\{\sigma_t\}$ and $\{\epsilon_t\}$ that enables $\{Y_t\}$ to be established as regularly varying. Given A1(ii), up to the jth moment of the model's rescaled errors is well-defined. By means of comparison, $j = 4$ is assumed in Kristensen and Rahbek (2005), while Hill and Renault (2012) present empirical findings that support $j = 4$ for a (wide) range of Hill (1975) estimator threshold values. For the IV estimators developed in this paper, $k = 3$, which is consistent with the aforementioned works.

ASSUMPTION A2: *For a $d \times 1$ vector $\boldsymbol{\alpha}$ of ARCH coefficients,*

$$\Theta = \left\{ \theta = (\omega, \boldsymbol{\alpha}) \in \mathbb{R}^{d+1} \mid \omega \geq \underline{\omega},\ \alpha_i > 0 \right\}$$

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

A2 heralds from Kristensen and Rahbek (2005). For the current discussion, $d = 1$. Notice that Θ is noncompact and ω is bounded below by a nonzero value, $\underline{\omega}$.

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

Under A3, D is an asymmetric distribution, where the direction of skewness is unconstrained. Skewness in (high frequency) returns is considered a stylized fact. That fact is exogenous to the model under consideration and yet (as will be shown) can be harnessed to identify the model. Other examples where a skewed D is used to account for this stylized fact include Hansen (1994) and Harvey and Siddique (1999).

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 the (strictly) stationary solution is the one being observed.

From (7) follows that

$$Y_t^2 = \sigma_t^2 + W_t, \qquad W_t = \sigma_t^2\left(\epsilon_t^2 - 1\right), \tag{9}$$

where $\{W_t\}$ is a MDS. Let $X_t \equiv Y_t^2 - \gamma_0$, where $\gamma_0 \equiv E\left(Y_t^2\right) = \frac{\omega_0}{1-\alpha_0}$. Then

$$X_t = \alpha_0 X_{t-1} + W_t, \tag{10}$$

so the centered second-order sequence $\{X_t\}$ follows an AR(1) process. Given $E\left(Y_t^3\right) = E\left(\sigma_t^3\right)c_3^*$, A4 is also sufficient for $\left\{Y_t^3\right\}$ to have a well-defined mean (see Lemma 1 in the Appendix). As a consequence, multiplying both sides of (10) by Y_{t-m} for $m \geq 1$ and taking expectations yields

$$E\left(X_t Y_{t-m}\right) = \alpha_0^m E\left(Y_t^3\right). \tag{11}$$

Letting $\mathbf{Z}_{t-1} = \left(Y_{t-1}, \ldots, Y_{t-h}\right)'$ for $h < \infty$, then $E\left(W_t \mathbf{Z}_{t-1}\right) = 0$ by iterative expectations and, owing to (11),

$$E\left(X_{t-1}\mathbf{Z}_{t-1}\right) = E\left(Y_t^3\right) \times \left(\begin{array}{cccc} 1, & \alpha_0, & \ldots, & \alpha_0^{h-1} \end{array} \right)', \tag{12}$$

so that $\mathbf{Z}_{t-1}$ is a valid set of instruments for X_{t-1}. As a result, for the observed sequence $\{Y_t\}_{t=1}^n$, consider

$$\widehat{\alpha}^{IV} = \frac{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \widehat{X}_t \mathbf{Z}_{t-1}\right)}{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)}, \tag{13}$$

$$\widehat{\omega}^{IV} = \widehat{\gamma}\left(1 - \widehat{\alpha}^{IV}\right), \tag{14}$$

where

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

Also note that both $\widehat{\alpha}^{IV}$ and $\widehat{\omega}^{IV}$ are variance-targeted estimators (VTEs).[5]

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

Suppose $\widehat{\mathbf{\Lambda}} = \left(n^{-1}\sum_t \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)^{-1}$. Then $\widehat{\alpha}^{IV}$ is a TSLS estimator, where X_{t-1} is first regressed on $\mathbf{Z}_{t-1}$ and then X_t is regressed on the predicted value from the first stage regression. Alternatively, owing to the available overidentifying restrictions, $\widehat{\alpha}^{IV}$ is the solution to a linear, two-step GMM estimator if $\widehat{\mathbf{\Lambda}} = \left(n^{-1}\sum_t \left(X_t - \widetilde{\alpha}X_{t-1}\right)^2 \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)^{-1}$, where $\widetilde{\alpha}$ is a preliminary estimator.

ASSUMPTION A6: $E\left(A^l\right) < 1$ *for* $l \geq 2$.

A6 governs the existence of higher moments for $\{Y_t\}$. If $l = 2$, then A6 is necessary and sufficient for $E\left(Y_t^4\right) < \infty$ (see; e.g., Bollerslev, 1986, Theorem 1). Cases where $l = 3, 4$ correspond to the thin-tailed cases where $E\left(Y_t^6\right) < \infty$ and $E\left(Y_t^8\right) < \infty$, respectively.

While the two-step GMM version of (13) is certainly preferable on efficiency grounds, it requires A6 with $l = 3$ in order for A5 to hold, which is inconsistent with Figure 1. In the TSLS case, on the other hand, since $\{Y_t\}$ is strongly mixing by Carrasco and Chen (2002, Corollary 6),

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

[5] In a QMLE context, VTE for (G)ARCH models is first introduced by Engle and Mezrich (1996), while the asymptotic theory is studied by Francq, Horváth, and Zakoïan (2011) and Vaynman and Beare (2014).

by the Ergodic Theorem, where $\mathbf{I}_h$ is the $(h \times h)$ identity matrix, given only A4.

$\widehat{\alpha}^{IV}$ is related to the IV estimator proposed by Guo and Phillips (2001). There are, however, two key differences. The first difference involves instrument choice. In Guo and Phillips, the instruments are second-order lags as opposed to first-order lags, as is the case here. Second, the instruments in (13) are not efficient in the sense of Kuersteiner (2002). Making them so, however, requires A6 with $l = 3$ and, hence, is limited to the thin-tailed case.

Let $\mathbf{Y}_t = \begin{pmatrix} Y_t, & \ldots, & Y_{t+h} \end{pmatrix}$, where, for short hand, $\mathbf{Y} = \mathbf{Y}_0 = \begin{pmatrix} Y_0, & \ldots, & Y_h \end{pmatrix}$. Then, $\mathbf{Y}$ is regularly varying in $\mathbb{R}^{h+1}$ with tail index κ_0 (see Lemmas 2 and 3 in the Appendix), or, using more short hand notation, $\mathbf{Y}$ is RV(κ_0), so there exists a sequence of constants $\{a_n\}$ such that

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

where $|\mathbf{Y}| = \max_{m=0,\ldots,h} |Y_m|$; $a_n = n^{1/\kappa_0} L(n)$, and $L(\cdot)$ is slowly-varying at ∞. That $\mathbf{Y}$ is regularly varying is demonstrated in Davis and Mikosch (1998, Lemma A.1) and Mikosch and Stărică (2000, Theorem 2.3) in instances where D is symmetric (see Remark R2 in the Appendix). Regular variation of $\mathbf{Y}$ here follows minus any need for symmetry in D (see Lemma 3 in the Appendix and its proof in the Supplemental Appendix) and applies to both the ARCH(1) case in (7) as well as the threshold ARCH(1) case of (32) of the next section, making the result compatible with A3 and complementary to Basrak, Davis and Mikosch (2002, Corollary 3.5 (B)).

THEOREM 1. *Consider the estimators in (13) and (14) for the model in (7). Let* $A_0 = E\left(X_{t-1}\mathbf{Z}_{t-1}\right)' \Lambda_0$ *and* $B_0 = E\left(X_{t-1}\mathbf{Z}_{t-1}\right)' \Lambda_0 E\left(X_{t-1}\mathbf{Z}_{t-1}\right)$. *Let Assumptions A1 with* $k = 3$, *and A2–A5 hold. Then* $\widehat{\alpha}^{IV} \xrightarrow{a.s.} \alpha_0$, *and* $\widehat{\omega}^{IV} \xrightarrow{a.s.} \omega_0$. *If* $\kappa_0 \in (3,\ 6)$, *then*

$$na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) \xrightarrow{d} \mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{V}_h, \tag{15}$$

where the vector $V_h = \begin{pmatrix} V_1, & \ldots, & V_h \end{pmatrix}'$, *with components* $(V_m)_{m=1,\ldots,h}$ *defined in Lemma 5 of the Appendix, is jointly* $(\kappa_0/3)-$*stable, and*

$$na_n^{-3}\left(\widehat{\omega}^{IV} - \omega_0\right) = -\gamma_0 na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) + o_p(1). \tag{16}$$

Alternatively, if Assumption A6 with $l = 3$ holds so that $\kappa_0 \in (6, \infty)$, then

$$\sqrt{n}\left(\widehat{\alpha}^{IV} - \alpha_0\right) \xrightarrow{d} N\left(0, \Sigma_{\alpha_0}\right) \tag{17}$$

and

$$\sqrt{n}\left(\widehat{\omega}^{IV} - \omega_0\right) \xrightarrow{d} N\left(0, \Sigma_{\omega_0}\right), \tag{18}$$

where

$$\Sigma_{\alpha_0} = \mathbf{B}_0^{-2}\mathbf{A}_0 E\left(W_t^2 \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)\mathbf{A}_0', \qquad \Sigma_{\gamma_0} = E\left(X_t^2\right) + 2\sum_{s=1}^{\infty} E\left(X_t X_s\right)$$

and

$$\Sigma_{\omega_0} = \Sigma_{\gamma_0} + \gamma_0^2 \Sigma_{\alpha_0} - 2\gamma_0 \mathbf{B}_0^{-1}\mathbf{A}_0 \left(\sum_{t \le s} E\left(W_t \mathbf{Z}_{t-1} Y_s^2\right)\right). \tag{19}$$

Proof. All proofs are contained in the Appendix. ■

The IV estimator in (13) depends on the sample cross-order covariances from (11), which are all nonzero owing to A3. The (weak) distributional limits of these cross-order covariances are established using a CLT from Davis and Mikosch (1998, Theorem 2.8) for point processes of regularly varying and dependent sequences (see Vaynman and Beare, 2014, for a concise review of this CLT) and the continuous mapping theorem (see the proof of Lemma 5 in the Supplemental Appendix for the demonstration of these distributional limits). The method of proof extends results from Davis and Mikosch (1998) and Mikosch and Stărică (2000) to cross-order covariances. This extension requires dual consideration of (normalized) sums of Y_t^3, $|Y_t|^3$, and $Y_t^2 Y_{t-m}$ (as opposed to, normalized, sums of Y_t^4 and $Y_t^2 Y_{t-m}^2$, as in the case of second-order autocovariances), and also relies on a first-order Taylor Expansion of σ_t^3 around $\underline{\omega}$; in which case, the limiting results are most appropriate for a small ω_0.[6] The (weak) distributional limit in (15), then, is simply a linear combination of the distributional limits of the relevant cross-order covariances, which are jointly stable by Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Individual components of $\mathbf{V}_h$ are dependent (see Lemma 5 in the Appendix).

A sufficient condition for (15) is $j = 6$ in A1. Such a condition is a close analog to one used in both Davis and Mikosch (1998) and Mikosch and Stărică (2000). Given a result from von Bahr and Esseen (1965, Theorem 2) that is also used in Vaynman and Beare (2014), this condition is relaxed in Theorem 1 to allow, instead, that $j \in (3, 6)$ supports (15), an alternative condition

[6] Given the values of $\widehat{\omega}$ typically encountered in practice, this limitation doesn't appear to pose much of a constraint.

that is better aligned with the empirical findings for many (high frequency) financial returns. This same (relaxed) alternative condition also applies to the threshold ARCH(1) and ARCH(p) cases discussed in Sections 2.2 and 2.3, respectively.

Under (15), the rate of convergence of $\widehat{\alpha}^{IV}$ (and, by extension, $\widehat{\omega}^{IV}$) is $n^{\frac{\kappa_0 - 3}{\kappa_0}}$, which is (quite a bit) slower than the $\sqrt{n}$ case, especially for values of κ_0 near the lower-bound of its required support, which, as evidenced in Figure 1, are the most empirically relevant. Lastly, Theorem 1 omits the borderline case of $\kappa_0 = 6$ for similar reasons cited in Vaynman and Beare (2014, Section 3.2).

Consider

$$\widehat{\tau}_n^2 = n^{-1} \sum_t Y_t^6, \tag{20}$$

and continue to suppose that $\kappa_0 \in (3,\ 6)$. Then, following the same method of proof for Davis and Hsing (1995, Theorem 3.1(i)),

$$n a_n^{-6} \widehat{\tau}_n^2 \xrightarrow{d} S_0, \tag{21}$$

where S_0 is $(\kappa_0/6)$ $-$stable. Given that $\mathbf{V}_h$ and S_0 are each characterized by stable laws, $\begin{pmatrix} \mathbf{V}_h', & S_0 \end{pmatrix}$ will be multivariate stable (see; e.g., Hall and Yao, 2003, and Vaynman and Beare, 2014, Theorem 4), in which case,

$$\sqrt{n} \left(\frac{\widehat{\alpha}^{IV} - \alpha_0}{\widehat{\tau}_n} \right) \xrightarrow{d} \frac{\mathbf{B}_0^{-1} \mathbf{A}_0 \mathbf{V}_h}{S_0^{1/2}}, \tag{22}$$

by the continuous mapping theorem.

(22) enjoys the advantage relative to (15) of removing the unknown scaling factor a_n^{-3}. Given (22), confidence intervals for $\widehat{\alpha}^{IV}$ can be constructed by applying the subsampling method in Vaynman and Beare (2014, Section 4.1)to the left-hand-side of (22).[7] Moreover, confidence intervals can, alternatively, be obtained by bootstrapping this same normalized quantity as demonstrated in Hall and Yao (2003, Corollary to Theorem 3.2). These bootstrap methods display better finite sample performance relative to the subsampling method while maintaining tractability, owing (precisely) to $\widehat{\alpha}^{IV}$ being a simple estimator.

In the thin-tailed case where A6 with $l = 3$ holds, the distributional limit of $\widehat{\alpha}^{IV}$ becomes Gaussian, with the usual rate of convergence. (22) is helpful in illustrating this case; since, when

[7] This method displays (very) poor finite sample performance for $n \leq 2,500$ (see Vaynman and Beare, 2004, Section 4.2). However, given the sample sizes in Table 1 and the statement from these same authors that results for their method are improved at sample sizes of $n = 50,000$, subsampling might prove to be, generally, more feasible (empirically) for applications involving intraday returns, as envisioned here.

$E\left(Y_t^6\right) < \infty$, $\widehat{\tau}_n$ has a degenerate limit, and the variance of the joint distribution behind $\mathbf{V}_h$ is well defined. Interestingly, in this case, the asymptotic variance of $\widehat{\gamma}$ does not impact Σ_{α_0}. Moreover, owing to (12), as $c_3^* \to 0$ (i.e., as D becomes increasingly symmetric), Σ_{α_0} increases without bound. In the limit where $c_3^* = 0$, Σ_{α_0} is ill-defined, rendering $\widehat{\alpha}^{IV}$ unidentified. Finally, as is well known, $\mathbf{\Lambda}_0 = E\left(W_t^2 \mathbf{Z}_{t-1} \mathbf{Z}_{t-1}'\right)^{-1}$ produces the minimum-variance estimator. In the thin-tailed case, then, $\widehat{\alpha}^{IV}$ should be a two-step GMM estimator.

Given A6 with $l = 2$, the fourth moment analog to (11) for $m \geq 1$ is

$$E\left(X_t X_{t-m}\right) = \alpha_0^m E\left(X_t^2\right), \tag{23}$$

so that OLS estimators for α_0 and ω_0 are

$$\widehat{\alpha}^{OLS} = \frac{\sum_t \widehat{X}_t \widehat{X}_{t-1}}{\sum_t \widehat{X}_{t-1}^2}, \tag{24}$$

$$\widehat{\omega}^{OLS} = \widehat{\gamma}\left(1 - \widehat{\alpha}^{OLS}\right). \tag{25}$$

Versions of (24) were first studied by Weiss (1986) and more recently by Guo and Phillips (2001).

THEOREM 2. *Consider the estimators in (24) and (25) for the model of (7). Let Assumptions A1 with $k = 4$, A2, and A6 with $l = 2$ hold. Then $\widehat{\alpha}^{OLS} \xrightarrow{a.s.} \alpha_0$, and $\widehat{\omega}^{OLS} \xrightarrow{a.s.} \omega_0$. If $\kappa_0 \in (4,\ 8)$, then*

$$n a_n^{-4}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) \xrightarrow{d} E\left(X_{t-1}^2\right)^{-1} U_1, \tag{26}$$

where U_1 is $(\kappa_0/4)-$stable, and

$$n a_n^{-4}\left(\widehat{\omega}^{OLS} - \omega_0\right) = -\gamma_0 n a_n^{-4}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) + o_p(1). \tag{27}$$

Alternatively, if Assumption A6 with $l = 4$ holds so that $\kappa_0 \in (8,\ \infty)$, then

$$\sqrt{n}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) \xrightarrow{d} N\left(0,\ E\left(X_{t-1}^2\right)^{-2} E\left(W_t^2 X_{t-1}^2\right)\right), \tag{28}$$

and

$$\sqrt{n}\left(\widehat{\omega}^{OLS} - \omega_0\right) \xrightarrow{d} N\left(0,\ \Sigma_{\omega_0}\right), \tag{29}$$

where

$$\Sigma_{\omega_0} = \Sigma_{\gamma_0} + E\left(X_{t-1}^2\right)^{-1}\left(\gamma_0^2 E\left(X_{t-1}^2\right)^{-1} E\left(W_t^2 X_{t-1}^2\right) - \sum_{t \leq s} E\left(W_t X_{t-1} Y_s^2\right)\right). \tag{30}$$

The OLS estimator in (24) depends on the first, sample second-order autocovariance from (23). The resulting (weak) distributional limit in (26) follows immediately from Davis and Mikosch (1998) if $c_3^* = 0$, and $j = 8$ in A1. Under Theorem 1, in contrast, the asymptotic properties of $\widehat{\alpha}^{OLS}$ are unaffected by whether or not A3 holds. Moreover, given von Bahr and Esseen (1965, Theorem 2) again (see the discussion following Theorem 1), $j \in (4,\ 8)$, instead, supports (26). The distribution of U_1 is similar to that of V_1 in Theorem 1 but, nonetheless, is distinct because the former is based on fourth-order mixtures of Poisson and i.i.d. point processes (see Lemma 4 and Remark R3 in the Appendix, as well as Davis and Hsing, 1995, Theorem 3.1), while the latter depends on third-order mixtures of these same processes. The general method of proof behind Theorems 1 and 2 is analogous. Asymptotic normality under Theorem 2 mirrors Weiss (1986, Theorem 4.4). The heavy-tailed case of (26), where the rate of convergence is $n^{\frac{\kappa_0 - 4}{\kappa_0}}$, is closely related to Kristensen and Linton (2006, Theorem 2).

It is important to note that if $\kappa_0 \in (4,\ 8)$ and A3 holds, then $\widehat{\alpha}^{IV}$ converges at a faster rate than does $\widehat{\alpha}^{OLS}$. Also, if $\kappa_0 \in (4,\ 8)$, then for

$$\widehat{\tau}_n^2 = n^{-1} \sum_t Y_t^8, \qquad n a_n^{-8} \widehat{\tau}_n^2 \xrightarrow{d} \widetilde{S}_0, \tag{31}$$

where $\widetilde{S}_0$ is $(\kappa_0/8)$ $-$stable (see Davis and Mikosch, 1998, Section 4B(1), for a closely-related result). As a consequence, normalizing the left-hand-side of (26) by $\widehat{\tau}_n$ enables inference on $\widehat{\alpha}^{OLS}$ to be conducted using the subsampling and bootstrapping methods discussed above in the context of Theorem 1. Lastly, the borderline case of $\kappa_0 = 8$ is not considered for the same reason that $\kappa_0 = 6$ is excluded from consideration in Theorem 1.

In light of Figure 1, the heavy-tailed case of (26), potentially, isn't heavy-tailed enough. Herein, then, lies the principal advantage of the TSLS estimator covered by Theorem 1 over the OLS alternative. Namely, so long as A3 holds, $\widehat{\alpha}^{IV}$ requires less in the way of higher-moment existence criteria than does $\widehat{\alpha}^{OLS}$.

2.2. The Threshold ARCH(1) Case

Next, consider the model of

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega_0 + \alpha_{1,0} Y_{t-1}^2 \times I_{\left\{Y_{t-1} \geq 0\right\}} + \alpha_{2,0} Y_{t-1}^2 \times I_{\left\{Y_{t-1} < 0\right\}}, \tag{32}$$

for which the following SRE applies:

$$\sigma_t^2 = \omega_0 + \sigma_{t-1}^2 A_t, \qquad A_t = \alpha_{0,t-1} \epsilon_{t-1}^2, \qquad \alpha_{0,t-1} = \alpha_{1,0} \times I_{\left\{Y_{t-1} \geq 0\right\}} + \alpha_{2,0} \times I_{\left\{Y_{t-1} < 0\right\}}.$$

$\{Y_t\}$ then continues to have a strictly stationary solution given A4. Also, since (9) continues to hold,

$$E\left(Y_t^2\right) = \frac{\omega_0 + \alpha_{1,0} Cov\left(Y_t^2, I_{\{Y_t \geq 0\}}\right) + \alpha_{2,0} Cov\left(Y_t^2, I_{\{Y_t < 0\}}\right)}{1 - \left(\alpha_{1,0} \times P\left(Y_t \geq 0\right) + \alpha_{2,0} \times P\left(Y_t < 0\right)\right)}, \tag{33}$$

so that

$$\begin{aligned} X_t &= \alpha_{1,0} X_{1,t-1} + \alpha_{2,0} X_{1,t-1} + W_t \\ &= \mathbf{X}_{t-1}' \boldsymbol{\alpha}_0 + W_t, \end{aligned} \tag{34}$$

where

$$\begin{aligned} X_{1,t-1} &= Y_{t-1}^2 \times I_{\left\{Y_{t-1} \geq 0\right\}} - E\left(Y_t^2 \times I_{\{Y_t \geq 0\}}\right), \\ X_{2,t-1} &= Y_{t-1}^2 \times I_{\left\{Y_{t-1} < 0\right\}} - E\left(Y_t^2 \times I_{\{Y_t < 0\}}\right), \end{aligned}$$

For the second-order equation in (34), consider as instruments

$$\begin{aligned} Z_{1,t-m} &= Y_{t-m} \times I_{\left\{Y_{t-m} \geq 0\right\}} - E\left(Y_t \times I_{\{Y_t \geq 0\}}\right), \\ Z_{2,t-m} &= Y_{t-m} \times I_{\left\{Y_{t-m} < 0\right\}} - E\left(Y_t \times I_{\{Y_t < 0\}}\right), \end{aligned}$$

where $m \geq 1$. Because $E\left(Y_t^3\right)$ remains well-defined given Lemma 1, multiplying both sides of (34) by $Z_{i,t-m}$ for $i = 1, 2$ and taking expectations produces

$$E\left(X_t Z_{i,t-1}\right) = \alpha_{i,0} E\left(X_{i,t} Z_{i,t}\right), \tag{35}$$

$$E\left(X_t Z_{i,t-m}\right) = \alpha_{1,0} E\left(X_t Z_{i,t-m+1}\right) - \left(\alpha_{1,0} - \alpha_{2,0}\right) E\left(X_{2,t} Z_{i,t-m+1}\right), \qquad m \geq 2, \tag{36}$$

where

$$\begin{aligned} E\left(X_{1,t} Z_{1,t}\right) &= E\left(Y_t^3 \times I_{\{Y_t \geq 0\}}\right) - Cov\left(Y_t, I_{\{Y_t \geq 0\}}\right) E\left(Y_t^2 \times I_{\{Y_t \geq 0\}}\right), \\ E\left(X_{2,t} Z_{2,t}\right) &= E\left(Y_t^3 \times I_{\{Y_t < 0\}}\right) - Cov\left(Y_t, I_{\{Y_t < 0\}}\right) E\left(Y_t^2 \times I_{\{Y_t < 0\}}\right). \end{aligned}$$

By noting that $X_t = X_{1,t} + X_{2,t}$ and $Y_t = Z_{1,t} + Z_{2,t}$, from (35) follows that

$$\begin{aligned} \sum_{i=1}^{2} E\left(X_t Z_{i,t-1}\right) &= E\left(X_t Y_{t-1}\right) \\ &= \alpha_{1,0} E\left(X_{1,t} Z_{1,t}\right) + \alpha_{2,0} E\left(X_{2,t} Z_{2,t}\right), \end{aligned} \tag{37}$$

and from (36) follows

$$\begin{aligned} \sum_{i=1}^{2} E\left(X_t Z_{i,t-m}\right) &= E\left(X_t Y_{t-m}\right) \\ &= \alpha_{1,0} E\left(X_{1,t} Y_{t-m+1}\right) - \left(\alpha_{1,0} - \alpha_{2,0}\right)\left[E\left(X_{2,t} Z_{1,t-m+1}\right) + E\left(X_{2,t} Z_{2,t-m+1}\right)\right], \end{aligned} \tag{38}$$

in which case, if $\alpha_{1,0} = \alpha_{2,0}$, then (37) simplifies to (11) when $m = 1$, and (38) also simplifies to (11), but for $m \geq 2$.

ASSUMPTION A7: *$E\left(X_{i,t} Z_{i,t}\right) \neq 0$ for $i = 1, 2$.*

Let

$$\mathbf{Z}_{t-1} = \left(\left(Z_{1,t-1}, Z_{2,t-1}\right), \ldots, \left(Z_{1,t-h}, Z_{2,t-h}\right)\right)', \qquad h < \infty.$$

Given A7, $\mathbf{Z}_{t-1}$ is a valid set of instruments for $\mathbf{X}_{t-1}$ in (34), which is to say that the usual rank condition required for IV estimators is satisfied. Notice that so long as $\alpha_{1,0} \neq \alpha_{2,0}$ (i.e., there is a threshold effect in the conditional variance), A7 does not require A3, meaning that skewness in Y_t is not necessary for identifying a simple IV estimator for (34). Thus, it is asymmetry in the conditional variance that renders $\mathbf{Z}_{t-1}$ valid for identifying $\boldsymbol{\alpha}_0$. In the absence of this second-moment asymmetry, A7 reduces to A3; since, in this case, $E\left(X_{1,t} Z_{1,t}\right) = E\left(Y_t^3\right) \times P\left(Y \geq 0\right)$, with an analogous result holding for $E\left(X_{2,t} Z_{2,t}\right)$.

Based on the instrument vector $\mathbf{Z}_{t-1}$, the threshold ARCH(1) analog to (13) is

$$\widehat{\boldsymbol{\alpha}}^{IV} = \widehat{\mathbf{F}}\left(n^{-1} \sum_t \widehat{X}_t \widehat{\mathbf{Z}}_{t-1}\right), \tag{39}$$

where

$$\widehat{\mathbf{F}} = \left[\left(n^{-1} \sum_t \widehat{\mathbf{X}}_{t-1} \widehat{\mathbf{Z}}_{t-1}^{\prime} \right) \widehat{\boldsymbol{\Lambda}} \left(n^{-1} \sum_t \widehat{\mathbf{X}}_{t-1} \widehat{\mathbf{Z}}_{t-1}^{\prime} \right)^{\prime} \right]^{-1} \left(n^{-1} \sum_t \widehat{\mathbf{X}}_{t-1} \widehat{\mathbf{Z}}_{t-1}^{\prime} \right) \widehat{\boldsymbol{\Lambda}}, \tag{40}$$

a $2 \times 2h$ matrix. When $\widehat{\boldsymbol{\Lambda}} = \left(n^{-1} \sum_t \widehat{\mathbf{Z}}_{t-1} \widehat{\mathbf{Z}}_{t-1}^{\prime} \right)^{-1}$, (39) is a TSLS estimator for (32), with the same discussion regarding selection of $\widehat{\boldsymbol{\Lambda}}$ in Section 2 remaining applicable.

THEOREM 3. *Consider the estimator in (39) for the model in (32), and let*

$$\mathbf{F}_0 = \left[E\left(\mathbf{X}_{t-1} \mathbf{Z}_{t-1}^{\prime} \right) \boldsymbol{\Lambda}_0 E\left(\mathbf{X}_{t-1} \mathbf{Z}_{t-1}^{\prime} \right)^{\prime} \right]^{-1} E\left(\mathbf{X}_{t-1} \mathbf{Z}_{t-1}^{\prime} \right) \boldsymbol{\Lambda}_0 .$$

In addition, let Assumptions A1 with $k = 3$*, A2, A4–A5 and A7 hold. Then,* $\widehat{\boldsymbol{\alpha}}^{IV} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$*. If* $\kappa_0 \in (3,\ 6)$*, then*

$$n a_n^{-3} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) \xrightarrow{d} \mathbf{F}_0 \mathbf{W}_h^{(+,-)}, \tag{41}$$

where the vector

$$\mathbf{W}_h^{(+,-)} = \left(\begin{array}{cccccc} W_1^+, & W_1^-, & \ldots, & W_h^+, & W_h^- \end{array} \right)^{\prime},$$

with components $\left(\begin{array}{cc} W_m^+ & W_m^- \end{array} \right)_{m=1,\ldots,h}$ *defined in Lemma 6 of the Appendix, is jointly* $(\kappa_0/3)-$*stable. Alternatively, if A6 with* $l = 3$ *holds so that* $\kappa_0 \in (6,\ \infty)$*, then*

$$\sqrt{n} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) \xrightarrow{d} N\left(0,\ \mathbf{F}_0 E\left(W_t^2 \mathbf{Z}_{t-1} \mathbf{Z}_{t-1}^{\prime} \right) \mathbf{F}_0^{\prime} \right). \tag{42}$$

The main result in (41) follows from the (weak) distributional convergence of $n^{-1} \sum_t X_t \mathbf{Z}_{t-1}$ (see Lemma 6 in the Appendix), which involves cross-order sums constructed from positive and negative realizations of $\{Y_t\}$, respectively. As a consequence, the distributional limit of $\widehat{\boldsymbol{\alpha}}^{IV}$ is a linear combination of the limits to sample cross-order covariances taken from the right-hand- and left-hand-side of the distribution of Y_t. Individual components of $\mathbf{W}_h^{(+,-)}$ are dependent (see Lemma 6 in the Appendix), and W_1^+ and W_1^- jointly depend on V_1, which connects the limiting result in (41) to that in (15) of Theorem 1. Normalizing the left-hand-side of (41) by $\widehat{\tau}_n$ as it is defined in (20) enables construction of either subsample or bootstrap confidence intervals for $\widehat{\boldsymbol{\alpha}}^{IV}$ as described following Theorem 1. Lastly, in the case where A6 with $l = 3$ holds, $\boldsymbol{\Lambda}_0 = E\left(W_t^2 \mathbf{Z}_{t-1} \mathbf{Z}_{t-1}^{\prime} \right)$ produces the minimum variance estimator so that $\widehat{\boldsymbol{\alpha}}^{IV}$ should be a two-step GMM estimator.

Given A6 with $l = 2$, from (34) follows that

$$E\left(X_t X_{i,t-1}\right) = \alpha_{i,0} E\left(X_{i,t-1}^2\right), \tag{43}$$

$$E\left(X_t X_{i,t-m}\right) = \alpha_{1,0} E\left(X_t X_{i,t-m+1}\right) - \left(\alpha_{1,0} - \alpha_{2,0}\right) E\left(X_{2,t} X_{i,t-m+1}\right), \qquad m \geq 2,$$

which reduces to (23) when $\alpha_{1,0} = \alpha_{2,0}$. Based on these results,

$$\widehat{\boldsymbol{\alpha}}^{OLS} = \widehat{\mathbf{K}}\left(n^{-1}\sum_t \widehat{X}_t \widehat{\mathbf{X}}_{t-1}\right), \qquad \widehat{\mathbf{K}} = \left(n^{-1}\sum_t \widehat{\mathbf{X}}_{t-1}\widehat{\mathbf{X}}_{t-1}'\right)^{-1}, \tag{44}$$

the large sample properties of which are determined in the following Corollary.

COROLLARY. *Consider the estimator in (44) for the model in (32), and let* $K_0 = E\left(\mathbf{X}_{t-1}\mathbf{X}_{t-1}'\right)^{-1}$. *In addition, let Assumptions A1 with* $k = 4$, *A2, and A6 with* $l = 2$ *hold. Then,* $\widehat{\boldsymbol{\alpha}}^{OLS} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$. *If* $\kappa_0 \in (4,\ 8)$, *then*

$$n a_n^{-4}\left(\widehat{\boldsymbol{\alpha}}^{OLS} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} \mathbf{K}_0 \mathbf{Q}_1^{(+,-)}, \tag{45}$$

where the vector $\mathbf{Q}_1^{(+,-)} = \left(\begin{array}{cc} Q_1^+, & Q_1^- \end{array}\right)'$, *with components* Q_1^+ *and* Q_1^- *defined in Lemma 7 of the Appendix, is jointly* $(\kappa_0/4)$ −stable. *Alternatively, if A6 with* $l = 4$ *holds so that* $\kappa_0 \in (8,\ \infty)$, *then*

$$\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}^{OLS} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} N\left(0,\ \mathbf{K}_0 E\left(W_t^2 \mathbf{X}_{t-1}\mathbf{X}_{t-1}'\right)\mathbf{K}_0'\right). \tag{46}$$

The Corollary extends results from Davis and Mikosch (1998) to the threshold ARCH(1) model. Necessary for the proof of the Corollary is establishing the (weak) distributional limit of $n^{-1}\sum_t X_t \mathbf{X}_{t-1}$, (see Lemma 7 in the Appendix). Given (31), normalizing the left-hand-side of (45) by $\widehat{\tau}_n$ produces

$$\sqrt{n}\left(\frac{\widehat{\boldsymbol{\alpha}}^{OLS} - \boldsymbol{\alpha}_0}{\widehat{\tau}_n}\right) \xrightarrow{d} \frac{\mathbf{K}_0 \mathbf{Q}_1^{(+,-)}}{\widetilde{S}_0^{1/2}},$$

in which case, subsample and bootstrap confidence intervals for $\widehat{\boldsymbol{\alpha}}^{OLS}$ can also result as in the discussion that follows Theorem 2. The Corollary, like Theorem 2, also does not require D in A1 to be symmetric. As a result, the Corollary can also apply to the same processes towards which Theorem 3 is directed; provided (of course) that the requisite higher moments are well defined. However, in cases where $\kappa_0 \in (4,\ 6)$, $\widehat{\boldsymbol{\alpha}}^{IV}$ converges at a faster rate (although, to a different and stable distribution) than does $\widehat{\boldsymbol{\alpha}}^{OLS}$, and when $\kappa_0 \in [6,\ 8)$, $\widehat{\boldsymbol{\alpha}}^{IV}$ is $\sqrt{n}$ asymptotically normal. Moreover, and in contrast to the convergence rate differentials discovered between $\widehat{\alpha}^{IV}$ and $\widehat{\alpha}^{OLS}$ in the ARCH(1) case, improvements in the rate of convergence enjoyed by $\widehat{\boldsymbol{\alpha}}^{IV}$ over $\widehat{\boldsymbol{\alpha}}^{OLS}$ do not, necessarily, depend on A3.

Finally, let

$$\boldsymbol{\Gamma}_0 = \left(\ Cov\left(Y_t^2, I_{\{Y_t \geq 0\}}\right),\ \ Cov\left(Y_t^2, I_{\{Y_t < 0\}}\right)\ \right)', \qquad \mathbf{P}_0 = \left(\ P\left(Y_t \geq 0\right),\ \ P\left(Y_t < 0\right)\ \right)'.$$

Then, given (33),

$$\widehat{\omega} = \widehat{\gamma}\left(1 - \widehat{\mathbf{P}}'\widehat{\boldsymbol{\alpha}}\right) - \widehat{\boldsymbol{\Gamma}}'\widehat{\boldsymbol{\alpha}}$$

so that

$$\widehat{\omega} - \omega_0 = \left(\widehat{\gamma} - \gamma_0\right) - \left(\gamma_0 \mathbf{P}_0 + \boldsymbol{\Gamma}_0\right)'\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\right),$$

and from which comparable versions of (16) and (27) then follow. Lastly, given A6 with either $l = 3$ or $l = 4$, comparable versions of (18) and (29) follow from Theorem 3 and the Corollary, respectively.

2.3. The ARCH(p) Case

Finally, consider the model of

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \alpha_{i,0} Y_{t-i}^2, \qquad 1 \leq p < \infty. \tag{47}$$

ASSUMPTION A8: $c_3 \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_{i,0}\alpha_{j,0}^{1/2} < 1.$

A8 is the generalization of A4 to ARCH(p) processes and, as such, is sufficient for $E\left(Y_t^3\right) < \infty$ (see Lemma 8 in the Appendix).

ASSUMPTION A9: *Define* $\rho_p\left(\epsilon_t\right)$ *as the largest root of* $1 - \sum_{i=1}^{p} \lambda^i \alpha_{i,0} \epsilon_t^2$.

$$E\left(\rho_p\left(\epsilon_t\right)^{2s}\right) < 1$$

for $s = 2, 3, 4$.

Suppose $j = 2s$ in A1. Then A9 establishes $E\left(Y_t^{2s}\right) < \infty$ (see Carrasco and Chen, 2002, Proposition 13).

From Basrak, Davis, and Mikosch (2002), (47) can be recast in terms of the following SRE:

$$\widetilde{\mathbf{Y}}_t = \mathbf{A}_t \widetilde{\mathbf{Y}}_{t-1} + \mathbf{B}_t, \tag{48}$$

where

$$\widetilde{\mathbf{Y}}_t = \begin{pmatrix} \sigma_t^2, & Y_{t-1}^2, & Y_{t-2}^2, & \ldots, & Y_{t-p+1}^2 \end{pmatrix},$$

$$\mathbf{A}_t = \begin{pmatrix} \alpha_{1,0}\epsilon_{t-1}^2 & \alpha_{2,0} & \alpha_{2,0} & \cdots & \alpha_{p,0} \\ \epsilon_{t-1}^2 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}$$

$$\mathbf{B}_t = \begin{pmatrix} \omega_0, & 0, & 0, & \ldots, & 0 \end{pmatrix}'.$$

Given A8, (48), Basrak, Davis, and Mikosch (2002, Theorem 3.1(A)), and Mikosch (1999, Remark 1.4.39), $\{Y_t\}$ has a strictly stationary solution. Given Basrak, Davis, and Mikosch (2002, Theorem 3.1 (B)), $\left\{\widetilde{\mathbf{Y}}_t\right\}$ is RV($\overline{\kappa}_0$), and given Basrak, Davis, and Mikosch (2002, Corollary 3.5 (B)), $\{Y_t\}$ is RV(κ_0), where $\kappa_0 = 2\overline{\kappa}_0$.

Given the definition of X_t used in Sections 2 and 3, let

$$\mathbf{X}_{t-1} = \begin{pmatrix} X_{t-1}, & \ldots, & X_{t-p} \end{pmatrix}'. \tag{49}$$

Then the generalization of (10) is

$$X_t = \mathbf{X}_{t-1}'\boldsymbol{\alpha}_0 + W_t, \tag{50}$$

where $\boldsymbol{\alpha}_0 = \begin{pmatrix} \alpha_{1,0}, & \ldots, & \alpha_{p,0} \end{pmatrix}'$. Consider

$$\mathbf{Z}_{t-1} = \begin{pmatrix} Y_{t-1}, & \ldots, & Y_{t-h} \end{pmatrix}', \qquad p \leq h < \infty, \tag{51}$$

as a vector of instruments for $\mathbf{X}_{t-1}$. Given A3, $\mathbf{Z}_{t-1}$ identifies $\boldsymbol{\alpha}_0$ in (50) (see Lemma 9 in the

Appendix). Consider then the estimator

$$\widehat{\boldsymbol{\alpha}}^{IV} = \widehat{\mathbf{F}} \left(n^{-1} \sum_t \widehat{X}_t \mathbf{Z}_{t-1} \right), \tag{52}$$

where $\widehat{\mathbf{F}}$ is defined as in (40), but with $\mathbf{Z}_{t-1}$ everywhere replacing $\widehat{\mathbf{Z}}_{t-1}$, and $\widehat{\mathbf{X}}_{t-1}$ defined as the finite sample version of (49).

THEOREM 4. *Consider the estimator in (52) for the model in (47). Let Assumptions A1 with $k = 3$, A2–A3, A5 and A8 hold. Then, $\widehat{\boldsymbol{\alpha}}^{IV} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$. If $\kappa_0 \in (3,\ 6)$, then*

$$n a_n^{-3} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) \xrightarrow{d} \mathbf{F}_0 \mathbf{V}_{p,h}, \tag{53}$$

where the vector $\mathbf{V}_{p,h} = \left(V_{p,1}, \ \ldots, \ V_{p,h} \right)'$, with components $\left(V_{p,m}\right)_{m=1,\ldots,h}$ defined in Lemma 12 of the Appendix, is jointly $(\kappa_0/3)-$stable. Alternatively, if Assumption A9 with $s = 3$ holds so that $\kappa_0 \in (6,\ \infty)$, then (42) results with $\mathbf{F}_0$ being the population limit of $\widehat{\mathbf{F}}$ in (52) and $\mathbf{Z}_{t-1}$ being defined in (51).

Under Theorem 4, (53) reduces to (15) when $p = 1$. As a consequence, A3 is necessary for establishing the large sample properties of the IV estimator (see Lemma 9 in the Appendix). That is, in the absence of skewness, the proposed estimator neither is identified nor does it possess a stable limiting distribution. Given (20), normalization of the left-hand-side of (53) enables the application of subsampling (see Vaynman and Beare, 2014, Theorem 6) or bootstrapping (see Hall and Yao, 2003, Corollary to Theorem 3.1) techniques to $\sqrt{n}\left(\frac{\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0}{\widehat{\tau}_n}\right)$ for the purpose of determining confidence intervals for $\widehat{\boldsymbol{\alpha}}^{IV}$.

The distributional limit in (53) generally differs from the special case presented in (15) in that the former is derived, in part, from (normalized) sums of $\{\sigma_t\}$ (see Lemmas 10 and 12 in the Appendix), while the latter is derived only from (normalized) sums of $\{Y_t\}$ (see Lemma 5). The complexities that arise in the cross-order covariances generated by (47) when $p > 1$ (see; e.g., Guo and Phillips, 2001, Lemma 1) necessitate this differential approach. The limit in (53), nonetheless, reduces to the limit in (15) when $p = 1$ and establishes both a stable limit and rate of convergence for a simple IV estimator applicable to the ARCH(p) model under a method of proof that is comparable to Basrak, Davis, and Mikosch (2002, Theorem 3.6).

The differential approach in establishing (53) relative to (15) is an example of the diminished ability to easily verify the large sample properties of general ARCH(p) versus ARCH(1) processes and (by extension) the estimators that apply to each. That A4 is sufficient for establishing $\{Y_t\}$ as strictly stationary in the ARCH(1) case, while a strictly negative Lyapunov exponent for the sequence $\{\mathbf{A}_t\}$ in (48) is necessary for establishing the same result in the ARCH(p) case (see; e.g., Basrak, Davis, and Mikosch, 2002, Theorem 2.1) is another example.

If A9 with $s = 2$ holds, a simple estimator for (50) is (44) with $\mathbf{X}_{t-1}$ defined by (49). Following the same method of proof from Lemmas 9–12 in the Appendix, it can be established that

$$na_n^{-4}\left(\widehat{\boldsymbol{\alpha}}^{OLS} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} \mathbf{K}_0\mathbf{U}_{p,p},$$

where the vector $\mathbf{U}_{p,p} = \begin{pmatrix} U_{p,1}, & \ldots, & U_{p,p} \end{pmatrix}'$ is jointly $(\kappa_0/4)$ −stable, reduces to U_1 from (26) in the special case where $p = 1$, but generally is not solely an expression of the observable sequence $\{Y_t\}$. If A9 with $s = 4$ holds, then (46) is established following the same method of proof as for the Corollary and echoes the result of Weiss (1986, Theorem 4.4). Confidence intervals for $\widehat{\boldsymbol{\alpha}}^{OLS}$ can be constructed from $\sqrt{n}\left(\frac{\widehat{\boldsymbol{\alpha}}^{OLS}-\boldsymbol{\alpha}_0}{\widehat{\tau}_n}\right)$ using (31), given either the subsample or bootstrap method discussed above in the context of Theorem 4.

Lastly, since

$$\widehat{\omega} - \omega_0 = (\widehat{\gamma} - \gamma_0) - \gamma_0\boldsymbol{\iota}'(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0),$$

and given Theorem 4 and the large sample properties of $\widehat{\boldsymbol{\alpha}}^{OLS}$ discussed immediately above, the large sample properties of $\widehat{\omega}$ can be established analogously to results presented in Theorems 1 and 2, respectively.

3. Monte Carlo

This section considers the ARCH(1) model from Section 2, where $\{\epsilon_t\}$ is drawn from the skewed student's t density of Hansen (1994). 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. Table 1 summarizes the various $(\lambda,\ \eta)$ pairs considered in the simulations. Also summarized for each pair is the skewness and (tail) index of the resulting sequence $\{Y_t\}$. To provide some context for the skewness measures reported in Table 1, Table 2 summarizes skewness estimates for various intra-day Japanese Yen returns (measured relative to the USD) as well as S&P 500 Index and DJIA returns. Apparent from the Table, high frequency financial returns tend to display significant skewness that can be quite large in magnitude (see also Cont and Kan, 2011, Table 3, for comparably-sized skewness estimates for daily, 5-year credit default swap spread returns). As a consequence, even the highest level of skewness considered in the simulations has empirical support. The values of $\eta = 6.1$ and $\eta = 8.1$ are consistent with the sufficient conditions for (15) and (26), respectively, that are discussed following the statements of Theorems 1 and 2. The case of $\eta = 4.1$ seems to be the most empirically relevant (see; e.g., Hill and Renault, 2012). Lastly, for all $(\lambda,\ \eta)$ pairs, A4 is satisfied so that $E\left(Y_t^3\right) < \infty$.

Across all simulations, $\omega_0 = 0.005$ and $\alpha_0 = 0.25$. Each of these values reflects the median estimate from Euro, Swiss Franc, and Japanese Yen returns (all measured relative the USD) sampled at the daily, hourly, 5-min, and 1-min frequencies obtained using the QMLE. The estimators under study are OLS, TSLS, and the QMLE. Sample sizes for the simulations are 500, $1,000$, and $10,000$. In the Supplemental Appendix, simulation results for sample sizes of $100,000$ are also reported; which, when compared against the results presented here, provide a good indication of the reduced rate at which the TSLS and OLS estimators converge. All simulations involve $10,000$ trials. Additional details on the simulations are contained in the notes to Tables 3–5.

Tables 3 and 4 summarize simulation results for $T = 500$ and $T = 1,000$. In all cases, TSLS is less biased than OLS but more biased than the QMLE. The TSLS bias also tends to decrease as the number of instruments decrease. However, this bias difference by number of instruments tends to disappear as skewness in the time series increases. Across the different sample sizes, the bias in both TSLS and OLS decreases as T increases but slowly relative to the QMLE, evidencing the slower rates of convergence predicted by Theorems 1 and 2. For TSLS, the case where $m = 100$ tends to be more efficient than the cases where $m = 50$ and $m = 25$.[8] This finding indicates that there isn't much finite sample cost to using more instruments. At high levels of skewness, though, differences in efficiency by the number of instruments appear muted. As skewness increases, TSLS efficiency improves, which stands in contrast to both OLS and the QMLE, where efficiency degrades as skewness increases.

TSLS records its best performance against the QMLE in cases of relatively high skewness. What's more, TSLS can best the QMLE in terms of efficiency in these cases when either T is small (500) or relatively modest $(1,000)$.[9] Surprisingly, the cases where TSLS outperforms the QMLE

[8] Here efficiency is measured by either rmse, mae, or mdae (see the Notes to Tables 3–5 for the, respective, definition of each) or the, respective, efficiency ratios.

[9] An efficiency ratio < 1 indicates less dispersion around the true parameter value than in the QMLE case.

by the widest margins all occur when $T = 500$. What's more surprising is that OLS can also best the QMLE in terms of efficiency in these same small and relatively modest sample sizes even when $E\left(Y_t^4\right) = \infty$. Explaining this finding is

$$na_n^{-4} \sum_t X_t X_{t-1} \xrightarrow{d} U_1,$$

which holds when $\kappa_0 \in (0,\ 4)$ (see Davis and Mikosch, 1998, Section 4B(1)). As a consequence, OLS will continue to have a stable distributional limit in the heaviest-tailed cases considered here; although, the estimator will no longer be consistent. The simulation findings indicate that the size of the resulting bias is small relative to the reduction in dispersion achieved relative to the QMLE. As the sample size increases, however, the QMLE eclipses both TSLS and OLS in terms of efficiency and regardless of skewness level, owing to the faster rate of convergence, by (very) wide margins (see Table 6 in the Supplemental Appendix for the largest discrepancies in efficiency between TSLS and OLS relative to the QMLE). Lastly, across all cases, OLS tends to outperform TSLS at low levels of skewness. This result reverses, however, as skewness increases and even in cases where all of the Assumptions supporting Theorem 2 are satisfied.

4. Conclusion

High frequency financial returns are well-known to display both skewness and leptokurtosis. Leptokurtosis has motivated the investigation of simple estimators for ARCH processes as potential alternatives to the QMLE. No attention is paid to skewness as a motivator for simple ARCH estimators. This paper fills that void. Specifically, this paper develops closed-form IV estimators for ARCH processes that are applicable when either the raw returns being modeled are skewed, or the conditional variance function is asymmetric.

As an extension of this paper's results, consider

$$\sigma_t^2 = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2,$$

which is the popular GARCH$(1,1)$ model introduced by Bollerslev (1986). For this model, the analog to (10) is

$$X_t = \phi_0 X_{t-1} - \beta_0 W_{t-1} + W_t, \qquad \phi_0 = \alpha_0 + \beta_0.$$

Following from results in Section 2.1, $\mathbf{Z}_{t-2} = \left(Y_{t-2}, \ldots, Y_{t-h}\right)'$ is a valid set of instruments for X_{t-1} when $\{Y_t\}$ is skewed and, thus, identifies ϕ_0. From Prono (2014), skewness in $\{Y_t\}$ can be used to separately identify α_0 and β_0. An interesting investigation, therefore, is whether the simple estimators introduced in this paper can be extended to the empirically better performing GARCH(p, q) class of models. This investigation is the subject of ongoing research.

Appendix

PRELIMINARIES. Contained in this Appendix are proofs to the Theorems and the Corollary as well as statements of the supporting Lemmas. Detailed proofs of the Lemmas are contained in the Supplemental Appendix. In what follows, for a vector $\mathbf{y}$, $\delta_{\mathbf{y}}$ denotes the Dirac measure at $\mathbf{y}$.

LEMMA 1. *For ARCH processes that can be cast in terms of the SRE*

$$\sigma_t^2 = \omega_0 + \sigma_{t-1}^2 A_t, \tag{54}$$

let Assumptions A1 with $k = 3$ and A2 hold. Then Assumption A4 is sufficient for $E\left(\sigma_t^3\right) < \infty$.

LEMMA 2. *For ARCH processes consistent with (54), let Assumptions A1 with $k = 3$, A2 and A4 hold. Consider the following lagged vectors for $h \geq 0$:*

$$\mathbf{Y}_h^{(i)} = \left(\; |Y_0|^i, \quad \ldots, \quad |Y_h|^i \; \right), \; i = 1, 2,$$

$$\mathbf{E}_h^{(2)} = \left(\; \epsilon_0^2, \quad A_1\epsilon_1^2, \quad \prod_{j=1}^{2} A_j\epsilon_2^2, \quad \ldots, \quad \prod_{j=1}^{h} A_j\epsilon_h^2 \; \right).$$

If σ is $RV(\kappa_0)$, then $\mathbf{Y}_h^{(2)}$ is $RV(\kappa_0/2)$, and $\mathbf{Y}_h^{(1)}$ is $RV(\kappa_0)$.

REMARK R1: Lemma 2 summarizes a collection of results for (G)ARCH processes proved elsewhere in the literature (see; e.g., Davis and Mikosch, 1998, and Mikosch and Stărică, 2000). Note that A3 is not influential in determining $\mathbf{Y}_h^{(i)}$ to be regularly varying.

LEMMA 3. *For the threshold ARCH(1) model, let Assumptions A1 with $k = 3$, A2 and A4 hold. Consider the following lagged vectors for $h \geq 0$,*

$$\mathbf{Y}_h^i = \left(\begin{array}{ccc} Y_0^i, & \ldots, & Y_h^i \end{array} \right), \quad i = 1, 3,$$

$$\mathbf{E}_h^{(1)} = \left(\begin{array}{ccccc} \epsilon_0, & |\epsilon_0|\,\epsilon_1, & |\epsilon_0|\,|\epsilon_1|\,\epsilon_2, & \ldots, & \prod_{i=0}^{h-1} |\epsilon_i|\,\epsilon_h \end{array} \right).$$

Then for all $\mathbf{y}_h^1 \in \mathbb{R}^{h+1} \setminus \{\mathbf{0}\}$, $\mathbf{Y}_h^1$ is $RV(\kappa_0)$, and $\mathbf{Y}_h^3$ is $RV(\kappa_0/3)$.

REMARK R2: Lemma 3 also applies to the special case where $\alpha_{1,0} = \alpha_{2,0} = \alpha_0$ (i.e., the symmetric ARCH(1) model). Moreover, under Lemma 3, regular variation of $\{Y_t\}$ follows minus any need for symmetry in the distribution of rescaled errors and so is consistent with A3 and complementary to Basrak, Davis, and Mikosch (2002, Corollary 3.5(B)). If the rescaled errors are, in fact, symmetrically distributed, then regular variation of $\{Y_t\}$ can also follow from regular variation of $\{|Y_t|\}$ as given by Lemma 2 and independence of $\{|Y_t|\}$ and $\{sign\,(\epsilon_t)\}$ so that Basrak, Davis, and Mikosch (2002, Corollary A.2) applies. Both Davis and Mikosch (1998, Lemma A.1) and Mikosch and Stărică (2000, Theorem 2.3) rely on this latter argument.

LEMMA 4. *Consider the threshold ARCH(1) model under the same Assumptions as Lemma 3. For the sequence of constants $\{a_n\}$,*

$$N_n := \sum_{t=1}^{n} \delta_{a_n^{-1}\mathbf{Y}_t} \xrightarrow{d} N := \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{P_i \mathbf{Q}_{i,j}}, \tag{55}$$

where: (i) $\sum_{i=1}^{\infty} \delta_{P_i}$ is a Poisson process on $(0,\, \infty)$; (ii) For $\mathbf{Q}_{i,j} = \left(\begin{array}{ccc} Q_{ij}^{(0)}, & \ldots, & Q_{ij}^{(h)} \end{array} \right)$, $\sum_{j=1}^{\infty} \delta_{\mathbf{Q}_{i,j}}$, $i \in \mathbb{N}$, is an i.i.d. sequence of point processes on $\mathbb{R}_+^{h+1} \setminus \{\mathbf{0}\}$ with common distribution $\mathbf{Q}$; (iii) $\sum_{i=1}^{\infty} \delta_{P_i}$ and $\sum_{j=1}^{\infty} \delta_{\mathbf{Q}_{i,j}}$, $i \in \mathbb{N}$, are mutually independent.

REMARK R3: Lemma 4 is the nonstandard CLT upon which (weak) distributional convergence of the IV and OLS estimators discussed in the paper are based. A generalization of this Lemma applies to the ARCH(p) case (see Basrak, Davis, and Mikosch, 2002, Theorem 2.10). Specification of the distribution $\mathbf{Q}$ is found in Davis and Mikosch (1998, Theorem 2.8). Following from Lemma 4, for

$$\mathbf{Y}_t^{(l)} = \left(\begin{array}{ccc} Y_t^l, & \ldots, & Y_{t+h}^l \end{array} \right), \qquad l = 2, 3,$$

$$N_n := \sum_{t=1}^{n} \delta_{a_n^{-l} \mathbf{Y}_t^l} \xrightarrow{d} N := \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i^l \mathbf{Q}_{i,j}^{(l)}}, \tag{56}$$

where $\mathbf{Q}_{i,j}^{(l)} = \left(\left(Q_{ij}^{(m)} \right)^l, \ m = 0, \ldots, h \right)$ by a continuous mapping argument.

LEMMA 5. *For the ARCH*(1) *model, let Assumptions A1 with* $k = 3$*, A2 and A4 hold. For* $m = 0, \ldots, h$*, define*

$$\widehat{\gamma}_{(Y,\ Y^2)}(m) = n^{-1} \sum_{t=1}^{n-m} Y_t Y_{t+m}^2, \qquad \gamma_{(Y,\ Y^2)}(m) = E\left(Y_0 Y_m^2 \right).$$

Then for a $\kappa_0 \in (3,\ 6)$*,*

$$n a_n^{-3} \left(\widehat{\gamma}_{(Y,\ Y^2)}(m) - \gamma_{(Y,\ Y^2)}(m) \right) \xrightarrow{d} (V_m)_{m=0,\ldots,h}, \qquad h \geq 1, \tag{57}$$

where $V_0 := V_0^* - c_3^* \alpha_0^{3/2} \left(1 - c_3 \alpha_0^{3/2} \right)^{-1} V_0^{**}$*, and* $V_m := V_m^* + \alpha_0 V_{m-1}$*.*

LEMMA 6. *For the threshold ARCH*(1) *model, let Assumptions A1 with* $k = 3$*, A2 and A4 hold. For* $m = 0, \ldots, h$*, define*

$$\widehat{\gamma}_{(Y,\ Y^2)}^{+}(m) = n^{-1} \sum_{t=1}^{n-m} Y_{t+m}^2 Y_t \times I_{\{Y_t \geq 0\}}, \qquad \gamma_{(Y,\ Y^2)}^{+}(m) = E\left(Y_m^2 Y_0 \times I_{\{Y_0 \geq 0\}} \right),$$

with $\widehat{\gamma}_{(Y,\ Y^2)}^{-}(m)$ *and* $\gamma_{(Y,\ Y^2)}^{-}(m)$ *defined analogously using* $I_{\{Y_t < 0\}}$*. Then for a* $\kappa_0 \in (3,\ 6)$ *and* $h > 1$*,*

$$n a_n^{-3} \left(\widehat{\gamma}_{(Y,\ Y^2)}^{+}(m) - \gamma_{(Y,\ Y^2)}^{+}(m) \right) \xrightarrow{d} \left(W_m^{+} \right)_{m=0,\ldots,h}, \tag{58}$$

and

$$na_n^{-3}\left(\widehat{\gamma}^{-}_{(Y,\ Y^2)}(m)-\gamma^{-}_{(Y,\ Y^2)}(m)\right)\xrightarrow{d}\left(W_m^{-}\right)_{m=0,\ldots,h}, \tag{59}$$

where

$$W_m^{+}=V_m^{+}+\alpha_{1,0}W_{m-1}^{+},\qquad W_m^{-}=V_m^{-}+\alpha_{2,0}W_{m-1}^{-},$$

and both W_0^{+} and W_0^{-} jointly depend on V_0^{**} from the proof of Lemma 5.

LEMMA 7. *Let Assumptions A1 with $k=4$, A2 and A6 with $l=2$ hold. For $m=0,1$ define*

$$\widehat{\gamma}^{+}_{Y^2}(m)=n^{-1}\sum_{t=1}^{n-m}Y_{t+m}^2Y_t^2\times I_{\{Y_t\geq 0\}},\qquad \gamma^{+}_{Y^2}(m)=E\left(Y_m^2Y_0^2\times I_{\{Y_0\geq 0\}}\right),$$

with $\widehat{\gamma}^{-}_{Y^2}(m)$ and $\gamma^{-}_{Y^2}(m)$ defined analogously using $I_{\{Y_t<0\}}$. Then for a $\kappa_0\in(4,\ 8)$,

$$na_n^{-4}\left(\widehat{\gamma}^{+}_{Y^2}(m)-\gamma^{+}_{Y^2}(m)\right)\xrightarrow{d}\left(Q_m^{+}\right)_{m=0,1},$$

and

$$na_n^{-4}\left(\widehat{\gamma}^{-}_{Y^2}(m)-\gamma^{-}_{Y^2}(m)\right)\xrightarrow{d}\left(Q_m^{-}\right)_{m=0,1},$$

where

$$Q_1^{+}=U_1^{+}+\alpha_{1,0}Q_0^{+},\qquad Q_1^{-}=U_1^{-}+\alpha_{2,0}Q_0^{-},$$

jointly depend on U_1 from Theorem 2.

LEMMA 8. *For the ARCH(p) model, let Assumptions A1 with $k=3$ and A2 hold. Then Assumption A8 is sufficient for $E\left(\sigma_t^3\right)<\infty$.*

LEMMA 9. *For the ARCH(p) model let Assumptions A1 with $k=3$, A2 and A8 hold. Consider*

$$X_t=\mathbf{X}_{t-1}'\boldsymbol{\alpha}_0+W_t \tag{60}$$

as it is defined in Section 2.3 of the main text and the set of instruments

$$\mathbf{Z}_{t-1}=\left(\begin{array}{ccc} Y_{t-1}, & \ldots, & Y_{t-h}\end{array}\right)',$$

where, in this case, $h = p$. Given Assumption A3, $\mathbf{Z}_{t-1}$ *identifies* α_0.

LEMMA 10. *For the ARCH(p) model, let Assumptions A1 with $k = 3$, A2 and A8 hold. Then*

$$a_n^{-3} \sum_t \sigma_t^3 - E\left(\sigma_t^3\right) \xrightarrow{d} V_{0,\sigma}$$

when $\kappa_0 \in (3,\ 6)$, *where* $V_{0,\sigma}$ *is* $(\kappa_0/3)-$*stable.*

LEMMA 11. *For the ARCH(p) model, let Assumptions A1 with $k = 3$, A2 and A8 hold. Then*

$$a_n^{-3} \sum_t Y_t^2 Y_{t+m} \xrightarrow{d} \left(R_{p,m}\right)_{m=1,\ldots,p},$$

when $\kappa_0 \in (3,\ 6)$.

LEMMA 12. *For the ARCH(p) model, let Assumptions A1 with $k = 3$, A2 and A8 hold. Then, given the definitions of* $\widehat{\gamma}_{(Y,\ Y^2)}(m)$ *and* $\gamma_{(Y,\ Y^2)}(m)$ *in Lemma 5,*

$$na_n^{-3}\left(\widehat{\gamma}_{(Y,\ Y^2)}(m) - \gamma_{(Y,\ Y^2)}(m)\right) \xrightarrow{d} \left(V_{p,m}\right)_{m=0,\ldots,h} \tag{61}$$

for a $\kappa_0 \in (3,\ 6)$, *where* $V_{p,0} := V_{p,0}^* + c_3^* V_{0,\sigma}$, and $V_{p,m} := V_{p,m}^* - \alpha_{1,0} V_{p.m-1}$.

PROOF OF THEOREM 1. To begin, note that

$$\widehat{X}_t = X_t - (\widehat{\gamma} - \gamma_0), \tag{62}$$

and

$$\widehat{X}_t = \overline{c} + \alpha_0 \widehat{X}_{t-1} + W_t, \tag{63}$$

where $\overline{c} = (\alpha_0 - 1)(\widehat{\gamma} - \gamma_0)$. Then given (63),

$$\widehat{\alpha}^{IV} = \alpha_0 + \frac{\overline{c}\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \mathbf{Z}_{t-1}\right)}{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)} + \frac{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t W_t\mathbf{Z}_{t-1}\right)}{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)} \tag{64}$$

By Carrasco and Chen (2002, Corollary 6), $\{Y_t\}$ is strong mixing. As a consequence, given (11) and A3, $\widehat{\alpha}^{IV} \stackrel{a.s.}{\to} \alpha_0$, and $\widehat{\omega}^{IV} \stackrel{a.s.}{\to} \omega_0$ by the Ergodic Theorem. Next, given (62) and noting that the population analog to $\widehat{\alpha}^{IV}$ in (13) is α_0,

$$\begin{aligned} na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) &= \left(\frac{\mathbf{A}_0\left(a_n^{-3}\sum_t X_t\mathbf{Z}_{t-1} - E\left(X_t\mathbf{Z}_{t-1}\right)\right)}{\mathbf{B}_0}\right) + o_P(1) \\ &\xrightarrow{d} \mathbf{B}_0^{-1}\mathbf{A}_0\mathbf{V}_h, \end{aligned}$$

where $\mathbf{V}_h$ is jointly $(\kappa_0/3)$ −stable by Lemma 5 and Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)), noting that

$$\begin{aligned} a_n^{-3}\sum_t X_t\mathbf{Z}_{t-1} - E\left(X_t\mathbf{Z}_{t-1}\right) &= a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1} - E\left(Y_t^2\mathbf{Z}_{t-1}\right) \\ &\quad -\gamma_0 n^{\frac{\kappa_0-6}{2\kappa_0}}\left(n^{-1/2}\sum_t \mathbf{Z}_{t-1}\right) \\ &= a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1} - E\left(Y_t^2\mathbf{Z}_{t-1}\right) + o_P(1) \end{aligned} \tag{65}$$

by Ibragimov and Linnik (1971, Theorem 18.5.3). Next, since $\widehat{\omega}^{IV} = \widehat{\gamma}\left(1 - \widehat{\alpha}^{IV}\right)$,

$$\begin{aligned} na_n^{-3}\left(\widehat{\omega}^{IV} - \omega_0\right) &= -\gamma_0 na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) + na_n^{-3}(\widehat{\gamma} - \gamma_0) \\ &= -\gamma_0 na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) + o_P(1), \end{aligned} \tag{66}$$

where the second equality relies on

$$a_n^{-2}\sum_t Y_t^2 \xrightarrow{d} \overline{V}_0,$$

for $\kappa_0 \in (3,\ 4]$ by Davis and Mikosch (1998), where $\overline{V}_0$ is $(\kappa_0/2)$-stable, and

$$n^{-1/2} \sum_t Y_t^2 \xrightarrow{d} N\left(0,\ \Sigma_{\gamma_0}\right),$$

for $\kappa_0 \in (4,\ 6)$ by Ibragimov and Linnik, where Σ_{γ_0} is defined in Theorem 1. Finally, if $\kappa_0 \in (6,\ \infty)$, then from (64),

$$\begin{aligned}\sqrt{n}\left(\widehat{\alpha}^{IV} - \alpha_0\right) &= \mathbf{B}_0^{-1}\mathbf{A}_0\left(n^{-1/2}\sum_t W_t \mathbf{Z}_{t-1}\right) + o_P(1) \\ &\xrightarrow{d} N\left(0,\ \frac{\mathbf{A}_0 E\left(W_t^2 \mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)\mathbf{A}_0'}{\mathbf{B}_0^2}\right),\end{aligned}$$

and

$$\begin{aligned}\sqrt{n}\left(\widehat{\omega}^{IV} - \omega_0\right) &= \sqrt{n}\left(\widehat{\gamma} - \gamma_0\right) - \gamma_0\sqrt{n}\left(\widehat{\alpha}^{IV} - \alpha_0\right) \\ &\xrightarrow{d} N\left(0,\ \Sigma_{\omega_0}\right),\end{aligned}$$

with Σ_{ω_0} also defined in Theorem 1. Both of these standard convergence results rely on Ibragimov and Linnik, with the first result also depending on the Slutsky Theorem.■

PROOF OF THEOREM 2. Given (62) and (63),

$$\widehat{\alpha}^{OLS} = \alpha_0 + \left(\sum_t \widehat{X}_{t-1}^2\right)^{-1}\left(\bar{c}\sum_t \widehat{X}_{t-1} - \left(\widehat{\gamma} - \gamma_0\right)\sum_t W_t + \sum_t W_t X_{t-1}\right). \tag{67}$$

Then $\widehat{\alpha}^{OLS} \stackrel{a.s.}{\to} \alpha_0$, and $\widehat{\omega}^{OLS} \stackrel{a.s.}{\to} \omega_0$ given the same arguments that establish consistency in the proof of Theorem 1. Next, given (62),

$$\begin{aligned}na_n^{-4}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) &= E\left(X_{t-1}^2\right)^{-1}\left(a_n^{-4}\sum_t X_t X_{t-1} - E\left(X_t X_{t-1}\right)\right) + o_P(1) \\ &\xrightarrow{d} E\left(X_{t-1}^2\right)^{-1} U_1,\end{aligned} \tag{68}$$

given Lemmas 2 and 3, Davis and Mikosch (1998), and von Bahr and Essen (1965, Theorem 2), where application of the latter permits $j \in (4,\ 8)$ in A1.[10] Comparable to Theorem 1,

[10] Application of von Bahr and Esseen (1965, Theorem 2) in this instance closely mirrors that in the proof of Lemma 5 in the Supplemental Appendix.

this (weak) distributional convergence results relies on

$$a_n^{-4}\sum_t X_t X_{t-1} - E\left(X_t X_{t-1}\right) = a_n^{-4}\sum_t Y_t^2 Y_{t-1}^2 - E\left(Y_t^2 Y_{t-1}^2\right) + o_P(1)$$

since

$$a_n^{-4}\sum_t Y_t^2 - \gamma_0 = n^{\frac{\kappa_0 - 8}{2\kappa_0}}\left(n^{-1/2}\sum_t Y_t^2 - \gamma_0\right) \xrightarrow{d} 0 \tag{69}$$

by Ibragimov and Linnik (1971, Theorem 18.5.3). Also given (69),

$$n a_n^{-4}\left(\widehat{\omega}^{OLS} - \omega_0\right) = -\gamma_0 n a_n^{-4}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) + o_P(1).$$

Finally, if $\kappa_0 \in (8,\ \infty)$, then given (67),

$$\begin{aligned} \sqrt{n}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) &= E\left(X_{t-1}^2\right)^{-1}\left(n^{-1/2}\sum_t W_t X_{t-1}\right) + o_P(1) \\ &\xrightarrow{d} N\left(0,\ E\left(X_{t-1}^2\right)^{-2} E\left(W_t^2 X_{t-1}^2\right)\right) \end{aligned}$$

by Ibragimov and Linnik and the Slutsky Theorem, and

$$\begin{aligned} \sqrt{n}\left(\widehat{\omega}^{OLS} - \omega_0\right) &= \sqrt{n}\left(\widehat{\gamma} - \gamma_0\right) - \gamma_0\sqrt{n}\left(\widehat{\alpha}^{OLS} - \alpha_0\right) \\ &\xrightarrow{d} N\left(0,\ \Sigma_{\omega_0}\right) \end{aligned}$$

where Σ_{ω_0} is defined in Theorem 2.$\blacksquare$

PROOF OF THEOREM 3. Given (62), also note that

$$\widehat{\mathbf{X}}_{t-1} = \mathbf{X}_{t-1} - \left(\widehat{\mathbf{G}} - \mathbf{G}_0\right), \qquad \mathbf{G}_0 = \left(E\left(Y_t^2 \times I_{\{Y_t \geq 0\}}\right),\ E\left(Y_t^2 \times I_{\{Y_t < 0\}}\right)\right)'$$

and

$$\widehat{\mathbf{Z}}_{t-1} = \mathbf{Z}_{t-1} - \left(\widehat{\mathbf{H}} - \mathbf{H}_0\right),$$

$$\mathbf{H}_0 = \left(E\left(Y_t^2 \times I_{\{Y_t \geq 0\}}\right),\ E\left(Y_t^2 \times I_{\{Y_t < 0\}}\right),\ E\left(Y_t^2 \times I_{\{Y_t \geq 0\}}\right),\ E\left(Y_t^2 \times I_{\{Y_t < 0\}}\right),\ldots\right)'$$

so that, comparable to (63),

$$\widehat{X}_t = \overline{c} + \widehat{\mathbf{X}}_{t-1}'\boldsymbol{\alpha}_0 + W_t,$$

where $\bar{c} = \left(\widehat{\mathbf{G}} - \mathbf{G}_0\right)' \boldsymbol{\alpha}_0 - (\widehat{\gamma} - \gamma_0)$. Then

$$\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 = \widehat{\mathbf{F}} \left[\bar{c} \left(n^{-1} \sum_t \mathbf{Z}_{t-1}' \right) - \left(\widehat{\mathbf{H}} - \mathbf{H}_0 \right) \left(n^{-1} \sum_t W_t \right) \right] + \widehat{\mathbf{F}} \left(n^{-1} \sum_t W_t \mathbf{Z}_{t-1} \right), \quad (70)$$

from which $\widehat{\alpha}^{IV} \overset{a.s.}{\to} \alpha_0$, where identification follows from A9 and (almost sure) convergence in the sample moments follows from the Ergodic Theorem, since $\{Y_t\}$ is strong mixing by Carrasco and Chen (2002, Corollary 10). Next, from (39),

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 &= \widehat{\mathbf{F}} \left(n^{-1} \sum_t X_t \mathbf{Z}_{t-1} - E\left(X_t \mathbf{Z}_{t-1}\right) \right) \\ &\quad - \widehat{\mathbf{F}} \left[\left(n^{-1} \sum_t \mathbf{Z}_{t-1} \right) \left(\left(\widehat{\mathbf{H}} - \mathbf{H}_0 \right) + (\widehat{\gamma} - \gamma_0) - (\widehat{\gamma} - \gamma_0) \left(\widehat{\mathbf{H}} - \mathbf{H}_0 \right) \right) \right] \\ &\quad - \left(\widehat{\mathbf{F}} - \mathbf{F}_0 \right) E\left(X_t \mathbf{Z}_{t-1}\right) \end{aligned}$$

such that

$$n a_n^{-3} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) = \mathbf{F}_0 \left(a_n^{-3} \sum_t X_t \mathbf{Z}_{t-1} - E\left(X_t \mathbf{Z}_{t-1}\right) \right) + o_P(1).$$

Let $\mathbf{Z}_{t-1} = \mathbf{Z}_{t-1}^{(1)} - \mathbf{H}_0$. Given the arguments that support the second equalities in both (65) and (66),

$$\begin{aligned} a_n^{-3} \sum_t X_t \mathbf{Z}_{t-1} - E\left(X_t \mathbf{Z}_{t-1}\right) &= a_n^{-3} \sum_t Y_t^2 \mathbf{Z}_{t-1}^{(1)} - E\left(Y_t^2 \mathbf{Z}_{t-1}^{(1)}\right) \\ &\quad - \left(\mathbf{H}_0 a_n^{-3} \sum_t Y_t^2 - E\left(Y_t^2\right) + \gamma_0 a_n^{-3} \sum_t \mathbf{Z}_{t-1} \right) \\ &= a_n^{-3} \sum_t Y_t^2 \mathbf{Z}_{t-1}^{(1)} - E\left(Y_t^2 \mathbf{Z}_{t-1}^{(1)}\right) + o_P(1) \end{aligned}$$

such that

$$n a_n^{-3} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) \overset{d}{\longrightarrow} \mathbf{F}_0 \mathbf{W}_h^{(+,-)},$$

where $\mathbf{W}_h^{(+,-)}$ is jointly $(\kappa_0/3)-$stable by Lemma 6 and Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Finally, from (70),

$$\sqrt{n} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) \overset{d}{\longrightarrow} N\left(0,\ \mathbf{F}_0 E\left(W_t^2 \mathbf{Z}_{t-1} \mathbf{Z}_{t-1}'\right) \mathbf{F}_0'\right),$$

by Ibragimov and Linnik (1971, Theorem 18.5.3) and the Slutsky Theorem.■

PROOF OF THE COROLLARY. From (44), using the expressions for $\widehat{\mathbf{X}}_{t-1}$ and $\widehat{X}_t$ as they relate to $\mathbf{X}_{t-1}$ and W_t, respectively, in the proof to Theorem 3,

$$\widehat{\boldsymbol{\alpha}}^{OLS}-\boldsymbol{\alpha}_0=\widehat{\mathbf{K}}\left[\bar{c}\left(n^{-1}\sum_t \mathbf{X}_{t-1}\right)+\left(\widehat{\mathbf{G}}-\mathbf{G}_0\right)\left(n^{-1}\sum_t W_t-1\right)+n^{-1}\sum_t \mathbf{X}_{t-1}W_t\right]. \quad (71)$$

Then, since

$$E\left(\mathbf{X}_{t-1}\mathbf{X}_{t-1}^{\prime}\right)=\begin{pmatrix} E\left(X_{1,t-1}^2\right) & 0 \\ 0 & E\left(X_{2,t-1}^2\right)\end{pmatrix},$$

$\boldsymbol{\alpha}_0$ is identified so that $\widehat{\boldsymbol{\alpha}}^{OLS}\xrightarrow{a.s.}\boldsymbol{\alpha}_0$ follows from the same argument that establishes (almost sure) consistency in the proof of Theorem 3. Next, let $\widehat{\mathbf{X}}_{t-1}=\mathbf{Z}_{t-1}^{(2)}-\mathbf{G}_0$. In the case where $\kappa_0\in(4,\ 8)$, consider

$$\begin{aligned} na_n^{-4}\left(\widehat{\boldsymbol{\alpha}}^{OLS}-\boldsymbol{\alpha}_0\right) &= \mathbf{K}_0\left[a_n^{-4}\sum_t \mathbf{X}_{t-1}X_t-E\left(\mathbf{X}_{t-1}X_t\right)\right]+o_p(1)\\ &= \mathbf{K}_0\left[a_n^{-4}\sum_t \mathbf{Z}_{t-1}^{(2)}Y_t^2-E\left(\mathbf{Z}_{t-1}^{(2)}Y_t^2\right)\right]\\ &\quad -n^{\frac{\kappa_0-8}{2\kappa_0}}\left[\mathbf{G}_0n^{-1}\sum_t Y_t^2-E\left(Y_t^2\right)+\gamma_0n^{-1}\sum_t \mathbf{X}_{t-1}\right]+o_p(1)\\ &= \mathbf{K}_0\left[a_n^{-4}\sum_t \mathbf{Z}_{t-1}^{(2)}Y_t^2-E\left(\mathbf{Z}_{t-1}^{(2)}Y_t^2\right)\right]+o_p(1)\\ &\xrightarrow{d}\mathbf{K}_0\mathbf{Q}_1^{(+,-)}, \end{aligned}$$

where $\mathbf{Q}_1^{(+,-)}=\begin{pmatrix} Q_1^+, & Q_1^-\end{pmatrix}$; the third equality follows from the standard CLT used elsewhere in this Appendix, and (weak) convergence in distribution to a $(\kappa_0/4)-$stable limit follows from Lemma 7 and Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Finally, if $\kappa_0\in(8,\ \infty)$, then given (71), (46) follows along the same lines as given in the proof to Theorem 3.■

PROOF OF THEOREM 4. Let $\boldsymbol{\iota}$ be a $p\times 1$ vector of ones. Given (49),

$$\widehat{\mathbf{X}}_{t-1}=\mathbf{X}_{t-1}-\left(\widehat{\gamma}-\gamma_0\right)\boldsymbol{\iota}$$

Then given (62),

$$\widehat{\boldsymbol{\alpha}}^{IV}-\boldsymbol{\alpha}_0=\widehat{\mathbf{F}}\left(\bar{c}\left(n^{-1}\sum_t \mathbf{Z}_{t-1}\right)+n^{-1}\sum_t W_t\mathbf{Z}_{t-1}\right), \tag{72}$$

where $\bar{c}=\left(\boldsymbol{\iota}'\boldsymbol{\alpha}_0-1\right)\left(\widehat{\gamma}-\gamma_0\right)$. By Lemma 9, $E\left(\mathbf{X}_{t-1}\mathbf{Z}_{t-1}'\right)$ has full row rank. By Carrasco and Chen (2002, Proposition 12), $\{Y_t\}$ is strong mixing. Then by the Ergodic Theorem, $\widehat{\boldsymbol{\alpha}}^{IV}\xrightarrow{a.s.}\boldsymbol{\alpha}_0$. Next, given (62),

$$\widehat{\boldsymbol{\alpha}}^{IV}-\boldsymbol{\alpha}_0=\widehat{\mathbf{F}}\left(n^{-1}\sum_t X_t\mathbf{Z}_{t-1}-E\left(X_t\mathbf{Z}_{t-1}\right)\right)-\left(\widehat{\gamma}-\gamma_0\right)\widehat{\mathbf{F}}\left(n^{-1}\sum_t \mathbf{Z}_{t-1}\right)+\left(\widehat{\mathbf{F}}-\mathbf{F}_0\right)E\left(X_t\mathbf{Z}_{t-1}\right)$$

so that

$$na_n^{-3}\left(\widehat{\boldsymbol{\alpha}}^{IV}-\boldsymbol{\alpha}_0\right)=\mathbf{F}_0\left(a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1}-E\left(Y_t^2\mathbf{Z}_{t-1}\right)\right)+o_p(1), \tag{73}$$

since

$$a_n^{-3}\sum_t X_t\mathbf{Z}_{t-1}-E\left(X_t\mathbf{Z}_{t-1}\right)=a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1}-E\left(Y_t^2\mathbf{Z}_{t-1}\right)+o_p(1),$$

following the same argument that supports (65). Then by Lemma 12,

$$na_n^{-3}\left(\widehat{\boldsymbol{\alpha}}^{IV}-\boldsymbol{\alpha}_0\right)\xrightarrow{d}\mathbf{F}_0\mathbf{V}_{p,h},$$

where $\mathbf{V}_{p,h}$ is jointly $(\kappa_0/3)-$stable by Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Finally, from (72),

$$\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}^{IV}-\boldsymbol{\alpha}_0\right)\xrightarrow{d}N\left(0,\ \mathbf{F}_0E\left(W_t^2\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)\mathbf{F}_0'\right),$$

if $\kappa_0\in(6,\ \infty)$ by Ibragimov and Linnik (1971, Theorem 18.5.3) and the Slutsky Theorem.

References

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

[2] Baillie, R.T. & H. Chung (1999) Estimation of GARCH models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis* 22, 631-650.

[3] Basrak, B., R.A Davis & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95-115.

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

[5] Bollerslev, T. & J. M. Wooldridge (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, 143-172.

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

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

[8] Cont, R. & Y.H. Kan (2011) Statistical modeling of credit default swap portfolios. Unpublished manuscript.

[9] 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.

[10] Davis, R.A. & T. Mikosch (1998) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics* 26, 2049-2080.

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

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

[13] Embrechts, P., C. Kluppelberg & T. Mikosch (1997) *Modelling Extremal Events in Insurance and Finance*. Springer, Berlin.

[14] Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of variance of united kingdom inflation. *Econometrica* 50, 987-1007.

[15] Engle, R.F. & J. Mezrich (1996) GARCH for Groups. *Risk* 9, 36-40.

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

[17] Francq, C., L. Horváth & J.M. Zakoïan (2011) Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics* 9, 619-656.

[18] 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.

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

[20] 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.

[21] Guo, B. & P.C.B Phillips (2001) Efficient estimation of second moment parameters in ARCH models. Unpublished manuscript.

[22] Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica 71, 285-317.

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

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

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

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

[27] Hill, J.B. & A. Prokhorov (2016) GEL estimation for heavy-tailed GARCH models with robust empirical likelihood inference. *Journal of Econometrics* 190, 18-45.

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

[29] Ibragimov, I.A. and Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff: Groningen.

[30] Jacquier, E., N. Polson & P.E. Rossi (1994) Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371-417.

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

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

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

[34] Kuersteiner, G.M. (2002) Efficient instrumental variables estimation for autoregressive models with conditional heteroskedasticity. *Econometric Theory* 18, 547-583.

[35] Lewbel, A. (1997) Constructing instruments for regressors with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65, 1201-1213.

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

[37] Loretan, M. & P.C.B Phillips (1994) Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* 1, 211-248.

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

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

[40] 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.

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

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

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

[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.

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

TABLE 1

λ	η	skew.	κ
-0.10	4.1	-0.65	3.58
-0.20		-1.27	3.50
-0.40		-2.32	3.32
-0.80		-3.48	3.14
-0.10	6.1	-0.34	4.54
-0.20		-0.67	4.41
-0.40		-1.23	4.12
-0.80		-1.88	3.85
-0.10	8.1	-0.27	5.14
-0.20		-0.53	4.97
-0.40		-0.98	4.62
-0.80		-1.52	4.30

Notes to Tables 1. The Monte Carlo simulations consider $\{\epsilon_t\}$ drawn from the skewed student's t density of Hansen (1994), where λ and η are the parameters governing this density, with the former determining skewness, the latter determining the tails, and moments up to the η^{th} being well defined. Summarized for each (λ, η) pair are the skewness and (tail) index, κ, for $\{Y_t\}$, noting that $\omega_0 = 0.005$ and $\alpha_0 = 0.25$ in $\{\sigma_t^2\}$. For skewness,

$$Skew\,(Y_t) \equiv E\left(\left(\frac{Y_t}{\sigma_t}\right)^3\right) = E\left(\epsilon_t^3\right)$$

so that an analytical solution is available using results from Jondeau and Rockinger (2003). The (tail) index, κ, is obtained as the mean value across $10,000$ simulation trials of the Hill (1975) estimator applied to $10,000$ observations of $\{Y_t\}$ using a constant threshold of 0.5%.

TABLE 2

	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 2. The data source is Bloomberg. The date range for all return series is 7/19/2015–12/31/2015. Skew is an estimate of the (unconditionally) standardized third moment. While not equivalent to the skewness measure applied in Table 1, simulation evidence (using the skewed student's t density) suggests these differences to be relatively minor enough not to disrupt comparisons between the general magnitudes of skewness measures summarized here and in Table 1. Standard errors for the skewness estimates are in parentheses and are measured against the null of normality.

TABLE 3

λ	est.	m	mean bias	med. bias	sd	dec. rge.	rmse	mae	mdae	Efficiency Ratio rmse	Efficiency Ratio mae	Efficiency Ratio mdae
						$\eta = 4.1$						
-0.10	TSLS	100	-0.063	-0.082	0.122	0.313	0.138	0.116	0.110	0.95	1.05	1.19
		50	-0.047	-0.066	0.135	0.352	0.143	0.120	0.113	0.99	1.08	1.23
		25	-0.024	-0.045	0.150	0.397	0.152	0.127	0.120	1.05	1.14	1.30
	OLS		-0.077	-0.099	0.113	0.275	0.137	0.118	0.116	0.94	1.06	1.26
	QMLE		-0.005	-0.030	0.145	0.345	0.145	0.111	0.092	1.00	1.00	1.00
	TSLS	100	-0.065	-0.085	0.122	0.310	0.138	0.118	0.112	0.93	1.03	1.18
		50	-0.049	-0.070	0.133	0.342	0.142	0.119	0.112	0.95	1.04	1.17
		25	-0.035	-0.059	0.146	0.383	0.150	0.126	0.119	1.01	1.10	1.24
	OLS		-0.080	-0.102	0.113	0.276	0.138	0.119	0.119	0.93	1.05	1.24
	QMLE		-0.007	-0.035	0.149	0.352	0.149	0.114	0.096	1.00	1.00	1.00
	TSLS	100	-0.072	-0.093	0.119	0.299	0.139	0.119	0.115	0.84	0.94	1.07
		50	-0.062	-0.083	0.124	0.316	0.139	0.118	0.112	0.84	0.93	1.04
		25	-0.055	-0.078	0.130	0.331	0.141	0.119	0.113	0.86	0.94	1.05
	OLS		-0.088	-0.112	0.112	0.274	0.142	0.124	0.126	0.87	0.98	1.17
	QMLE		-0.007	-0.044	0.164	0.389	0.164	0.127	0.108	1.00	1.00	1.00
	TSLS	100	-0.087	-0.107	0.106	0.267	0.138	0.120	0.118	0.72	0.80	0.90
		50	-0.083	-0.101	0.109	0.275	0.137	0.118	0.115	0.71	0.79	0.88
		25	-0.080	-0.100	0.111	0.282	0.137	0.118	0.116	0.72	0.79	0.88
	OLS		-0.105	-0.128	0.105	0.249	0.148	0.132	0.136	0.77	0.88	1.03
	QMLE		-0.010	-0.064	0.192	0.453	0.192	0.150	0.132	1.00	1.00	1.00
						$\eta = 6.1$						
-0.10	TSLS	100	-0.046	-0.059	0.119	0.307	0.128	0.106	0.095	1.19	1.26	1.34
		50	-0.031	-0.045	0.134	0.349	0.138	0.114	0.104	1.29	1.36	1.46
		25	-0.009	-0.024	0.152	0.406	0.152	0.126	0.115	1.42	1.50	1.62
	OLS		-0.052	-0.068	0.105	0.258	0.117	0.098	0.091	1.09	1.17	1.28
	QMLE		-0.004	-0.014	0.107	0.265	0.107	0.084	0.071	1.00	1.00	1.00
	TSLS	100	-0.047	-0.061	0.120	0.308	0.129	0.107	0.098	1.17	1.24	1.35
		50	-0.032	-0.046	0.133	0.346	0.137	0.113	0.103	1.24	1.31	1.41
		25	-0.016	-0.035	0.146	0.388	0.147	0.122	0.112	1.33	1.41	1.54
	OLS		-0.055	-0.072	0.106	0.259	0.119	0.100	0.093	1.08	1.15	1.28
	QMLE		-0.006	-0.017	0.110	0.273	0.110	0.086	0.073	1.00	1.00	1.00
	TSLS	100	-0.051	-0.066	0.117	0.301	0.128	0.107	0.099	1.03	1.11	1.22
		50	-0.042	-0.056	0.124	0.318	0.131	0.108	0.099	1.05	1.12	1.22
		25	-0.035	-0.052	0.131	0.337	0.135	0.112	0.102	1.09	1.16	1.26
	OLS		-0.063	-0.081	0.107	0.265	0.124	0.105	0.100	1.00	1.09	1.23
	QMLE		-0.004	-0.021	0.124	0.300	0.124	0.096	0.081	1.00	1.00	1.00
	TSLS	100	-0.063	-0.078	0.107	0.273	0.125	0.105	0.099	0.83	0.90	1.00
		50	-0.059	-0.072	0.110	0.282	0.125	0.104	0.096	0.83	0.90	0.97
		25	-0.055	-0.070	0.113	0.290	0.126	0.105	0.097	0.83	0.90	0.98
	OLS		-0.077	-0.096	0.105	0.256	0.130	0.112	0.110	0.87	0.96	1.11
	QMLE		-0.002	-0.031	0.150	0.361	0.150	0.116	0.099	1.00	1.00	1.00
						$\eta = 8.1$						
-0.10	TSLS	100	-0.038	-0.048	0.117	0.303	0.123	0.101	0.089	1.31	1.36	1.42
		50	-0.025	-0.037	0.133	0.348	0.136	0.111	0.100	1.44	1.50	1.59
		25	-0.002	-0.015	0.152	0.405	0.152	0.125	0.113	1.62	1.69	1.80
	OLS		-0.042	-0.056	0.099	0.244	0.108	0.089	0.080	1.15	1.20	1.28
	QMLE		-0.005	-0.010	0.094	0.237	0.094	0.074	0.063	1.00	1.00	1.00
	TSLS	100	-0.040	-0.052	0.118	0.305	0.125	0.102	0.092	1.29	1.34	1.41
		50	-0.026	-0.037	0.132	0.344	0.134	0.110	0.099	1.39	1.45	1.52
		25	-0.009	-0.026	0.146	0.391	0.147	0.121	0.109	1.52	1.58	1.66
	OLS		-0.045	-0.058	0.101	0.248	0.110	0.091	0.083	1.14	1.19	1.27
	QMLE		-0.006	-0.012	0.096	0.243	0.097	0.076	0.065	1.00	1.00	1.00
	TSLS	100	-0.043	-0.055	0.116	0.298	0.124	0.102	0.093	1.16	1.21	1.30
		50	-0.034	-0.046	0.123	0.318	0.128	0.105	0.094	1.20	1.25	1.32
		25	-0.027	-0.042	0.131	0.338	0.134	0.110	0.098	1.25	1.30	1.38
	OLS		-0.052	-0.067	0.103	0.255	0.115	0.096	0.088	1.08	1.15	1.24
	QMLE		-0.005	-0.016	0.106	0.265	0.107	0.084	0.071	1.00	1.00	1.00
	TSLS	100	-0.053	-0.066	0.107	0.274	0.119	0.099	0.092	0.94	0.99	1.08
		50	-0.048	-0.060	0.110	0.282	0.120	0.099	0.090	0.94	0.99	1.05
		25	-0.045	-0.058	0.113	0.289	0.122	0.100	0.091	0.96	1.00	1.07
	OLS		-0.065	-0.082	0.103	0.254	0.122	0.103	0.099	0.96	1.03	1.16
	QMLE		-0.004	-0.021	0.127	0.315	0.127	0.100	0.085	1.00	1.00	1.00

TABLE 4

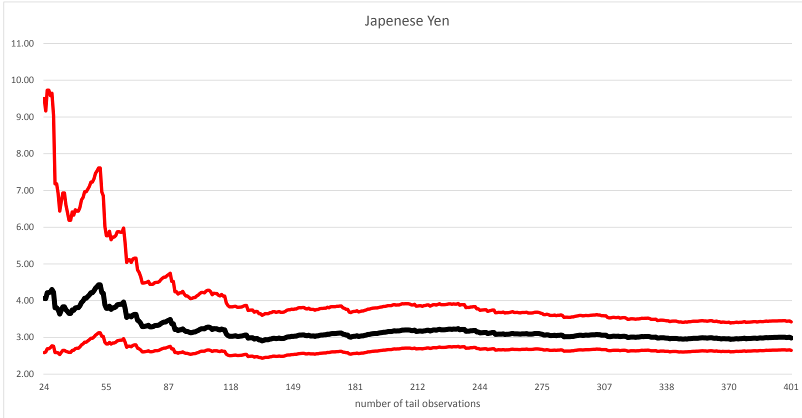
λ	est.	m	mean bias	med. bias	sd	dec. rge.	rmse	mae	mdae	Efficiency Ratio rmse	Efficiency Ratio mae	Efficiency Ratio mdae
						$\eta = 4.1$						
-0.10	TSLS	100	-0.056	-0.076	0.123	0.314	0.135	0.114	0.107	1.17	1.31	1.51
		50	-0.041	-0.059	0.134	0.349	0.140	0.117	0.108	1.22	1.35	1.53
		25	-0.022	-0.044	0.149	0.391	0.151	0.126	0.116	1.31	1.45	1.64
	OLS		-0.068	-0.088	0.106	0.257	0.126	0.108	0.104	1.10	1.24	1.46
	QMLE		-0.005	-0.022	0.115	0.269	0.115	0.087	0.071	1.00	1.00	1.00
-0.20	TSLS	100	-0.057	-0.076	0.122	0.310	0.135	0.113	0.107	1.12	1.25	1.45
		50	-0.046	-0.063	0.130	0.335	0.138	0.115	0.107	1.15	1.27	1.45
		25	-0.034	-0.054	0.140	0.363	0.144	0.119	0.110	1.20	1.32	1.49
	OLS		-0.071	-0.092	0.107	0.256	0.128	0.110	0.107	1.06	1.21	1.45
	QMLE		-0.004	-0.024	0.120	0.279	0.120	0.091	0.074	1.00	1.00	1.00
-0.40	TSLS	100	-0.064	-0.083	0.114	0.288	0.130	0.110	0.104	0.95	1.08	1.26
		50	-0.059	-0.077	0.116	0.293	0.131	0.110	0.103	0.95	1.07	1.24
		25	-0.056	-0.075	0.119	0.299	0.132	0.110	0.103	0.96	1.08	1.24
	OLS		-0.080	-0.101	0.106	0.254	0.132	0.115	0.115	0.97	1.12	1.38
	QMLE		-0.004	-0.031	0.137	0.315	0.137	0.103	0.083	1.00	1.00	1.00
-0.80	TSLS	100	-0.078	-0.095	0.100	0.246	0.127	0.109	0.106	0.78	0.89	1.04
		50	-0.076	-0.093	0.101	0.250	0.127	0.108	0.106	0.78	0.88	1.03
		25	-0.075	-0.091	0.102	0.251	0.126	0.108	0.104	0.78	0.88	1.01
	OLS		-0.096	-0.117	0.099	0.232	0.138	0.122	0.124	0.85	0.99	1.21
	QMLE		-0.005	-0.045	0.162	0.375	0.162	0.123	0.103	1.00	1.00	1.00
						$\eta = 6.1$						
-0.10	TSLS	100	-0.038	-0.050	0.119	0.305	0.124	0.102	0.090	1.58	1.66	1.77
		50	-0.024	-0.036	0.134	0.349	0.136	0.112	0.101	1.72	1.82	1.98
		25	-0.005	-0.022	0.151	0.398	0.151	0.124	0.112	1.92	2.02	2.19
	OLS		-0.042	-0.056	0.094	0.227	0.103	0.085	0.077	1.31	1.38	1.51
	QMLE		-0.003	-0.010	0.079	0.196	0.079	0.061	0.051	1.00	1.00	1.00
-0.20	TSLS	100	-0.038	-0.049	0.118	0.303	0.124	0.102	0.090	1.52	1.59	1.70
		50	-0.026	-0.038	0.129	0.336	0.132	0.108	0.097	1.61	1.69	1.82
		25	-0.014	-0.029	0.139	0.362	0.140	0.114	0.102	1.71	1.79	1.91
	OLS		-0.044	-0.060	0.096	0.230	0.106	0.087	0.079	1.29	1.37	1.50
	QMLE		-0.003	-0.010	0.082	0.201	0.082	0.064	0.053	1.00	1.00	1.00
-0.40	TSLS	100	-0.041	-0.055	0.111	0.279	0.118	0.097	0.087	1.27	1.35	1.46
		50	-0.038	-0.049	0.115	0.292	0.121	0.099	0.087	1.30	1.37	1.46
		25	-0.034	-0.046	0.118	0.297	0.123	0.100	0.089	1.32	1.39	1.49
	OLS		-0.052	-0.068	0.098	0.237	0.111	0.093	0.087	1.19	1.29	1.47
	QMLE		-0.003	-0.014	0.093	0.226	0.093	0.072	0.059	1.00	1.00	1.00
-0.80	TSLS	100	-0.051	-0.063	0.098	0.242	0.110	0.092	0.083	0.97	1.05	1.14
		50	-0.049	-0.061	0.099	0.247	0.111	0.092	0.083	0.97	1.05	1.14
		25	-0.047	-0.059	0.100	0.248	0.111	0.091	0.082	0.98	1.05	1.13
	OLS		-0.066	-0.082	0.096	0.229	0.117	0.099	0.096	1.02	1.14	1.31
	QMLE		-0.002	-0.020	0.114	0.274	0.114	0.087	0.073	1.00	1.00	1.00
						$\eta = 8.1$						
-0.10	TSLS	100	-0.030	-0.040	0.115	0.298	0.119	0.097	0.085	1.76	1.82	1.88
		50	-0.017	-0.027	0.132	0.346	0.133	0.109	0.097	1.96	2.04	2.16
		25	0.002	-0.011	0.150	0.401	0.150	0.123	0.112	2.22	2.30	2.48
	OLS		0.068	0.055	0.087	0.209	0.111	0.082	0.061	1.63	1.53	1.36
	QMLE		-0.003	-0.007	0.068	0.171	0.068	0.054	0.045	1.00	1.00	1.00
-0.20	TSLS	100	-0.030	-0.039	0.115	0.297	0.119	0.097	0.085	1.70	1.75	1.82
		50	-0.020	-0.029	0.129	0.336	0.130	0.106	0.094	1.86	1.92	2.01
		25	-0.007	-0.019	0.139	0.363	0.139	0.113	0.100	1.99	2.05	2.13
	OLS		0.066	0.053	0.089	0.215	0.111	0.082	0.060	1.58	1.47	1.28
	QMLE		-0.003	-0.007	0.070	0.176	0.070	0.055	0.047	1.00	1.00	1.00
-0.40	TSLS	100	-0.033	-0.043	0.108	0.272	0.113	0.092	0.080	1.45	1.50	1.57
		50	-0.029	-0.039	0.114	0.289	0.117	0.095	0.082	1.50	1.54	1.60
		25	-0.026	-0.036	0.117	0.295	0.120	0.097	0.084	1.53	1.57	1.64
	OLS		0.060	0.046	0.092	0.224	0.110	0.080	0.057	1.41	1.30	1.12
	QMLE		-0.003	-0.009	0.078	0.195	0.078	0.061	0.051	1.00	1.00	1.00
-0.80	TSLS	100	-0.041	-0.051	0.096	0.238	0.104	0.085	0.076	1.11	1.16	1.22
		50	-0.039	-0.048	0.098	0.242	0.105	0.085	0.075	1.12	1.16	1.21
		25	-0.037	-0.047	0.099	0.244	0.105	0.085	0.075	1.12	1.16	1.21
	OLS		-0.053	-0.068	0.093	0.222	0.107	0.090	0.084	1.14	1.22	1.36
	QMLE		-0.002	-0.013	0.094	0.234	0.094	0.074	0.062	1.00	1.00	1.00

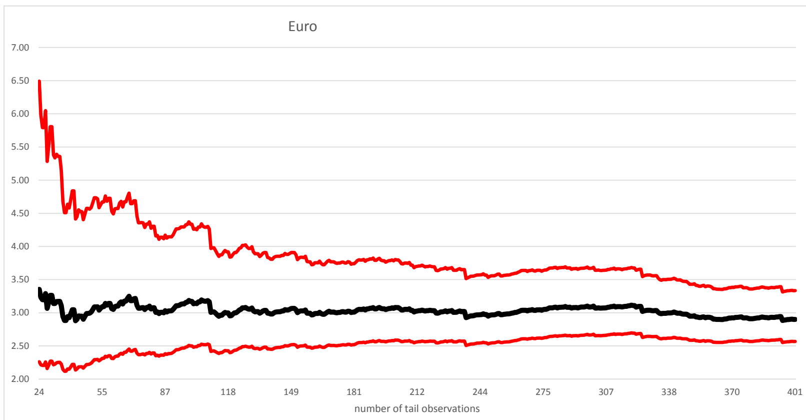
TABLE 5

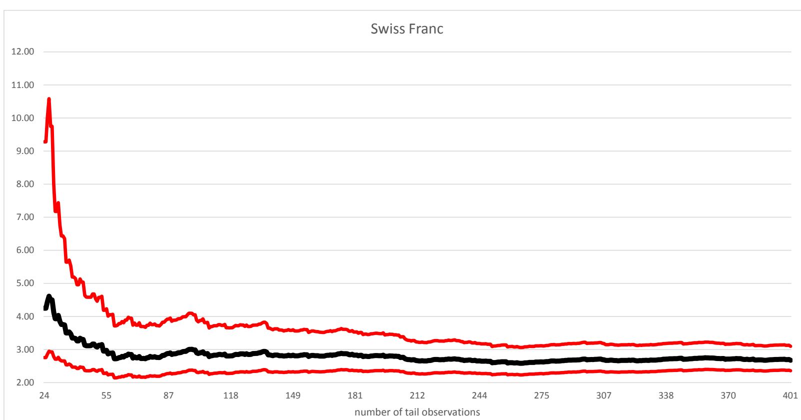
λ	est.	m	mean bias	med. bias	sd	dec. rge.	rmse	mae	mdae	Efficiency Ratio rmse	mae	mdae
						$\eta = 4.1$						
-0.10	TSLS	100	-0.034	-0.047	0.123	0.318	0.128	0.105	0.094	2.77	3.18	3.60
		50	-0.025	-0.039	0.133	0.347	0.135	0.111	0.100	2.93	3.36	3.86
		25	-0.016	-0.033	0.140	0.367	0.141	0.115	0.104	3.06	3.50	4.01
	OLS		-0.048	-0.062	0.088	0.205	0.101	0.083	0.077	2.19	2.53	2.95
	QMLE		0.000	-0.005	0.046	0.102	0.046	0.033	0.026	1.00	1.00	1.00
-0.20	TSLS	100	-0.034	-0.046	0.106	0.261	0.111	0.090	0.078	2.29	2.57	2.80
		50	-0.032	-0.045	0.110	0.271	0.115	0.092	0.081	2.36	2.64	2.91
		25	-0.030	-0.045	0.112	0.276	0.116	0.093	0.081	2.39	2.67	2.93
	OLS		-0.051	-0.066	0.089	0.210	0.103	0.086	0.079	2.11	2.46	2.86
	QMLE		0.000	-0.006	0.049	0.107	0.049	0.035	0.028	1.00	1.00	1.00
-0.40	TSLS	100	-0.038	-0.050	0.084	0.192	0.092	0.074	0.066	1.59	1.80	2.01
		50	-0.037	-0.050	0.085	0.194	0.092	0.075	0.066	1.60	1.81	2.02
		25	-0.037	-0.049	0.085	0.193	0.092	0.075	0.066	1.60	1.81	2.01
	OLS		-0.060	-0.076	0.089	0.208	0.107	0.091	0.087	1.86	2.21	2.65
	QMLE		0.000	-0.007	0.058	0.125	0.058	0.041	0.033	1.00	1.00	1.00
-0.80	TSLS	100	-0.048	-0.060	0.073	0.164	0.087	0.073	0.069	1.17	1.37	1.63
		50	-0.048	-0.059	0.073	0.165	0.087	0.073	0.069	1.18	1.38	1.63
		25	-0.047	-0.059	0.072	0.164	0.087	0.073	0.068	1.17	1.37	1.63
	OLS		-0.040	-0.054	0.092	0.224	0.101	0.083	0.076	1.36	1.57	1.81
	QMLE		0.001	-0.010	0.074	0.162	0.074	0.053	0.042	1.00	1.00	1.00
						$\eta = 6.1$						
-0.10	TSLS	100	-0.016	-0.022	0.111	0.285	0.112	0.090	0.076	4.43	4.52	4.55
		50	-0.010	-0.017	0.123	0.319	0.123	0.100	0.087	4.86	4.99	5.22
		25	-0.002	-0.011	0.131	0.341	0.131	0.106	0.092	5.17	5.29	5.46
	OLS		-0.020	-0.030	0.064	0.140	0.067	0.052	0.044	2.65	2.60	2.65
	QMLE		0.000	-0.001	0.025	0.064	0.025	0.020	0.017	1.00	1.00	1.00
-0.20	TSLS	100	-0.015	-0.022	0.090	0.221	0.091	0.071	0.059	3.44	3.42	3.34
		50	-0.014	-0.021	0.094	0.228	0.095	0.074	0.062	3.57	3.55	3.52
		25	-0.013	-0.020	0.096	0.232	0.097	0.076	0.063	3.66	3.61	3.56
	OLS		-0.022	-0.032	0.066	0.146	0.070	0.055	0.047	2.63	2.61	2.69
	QMLE		0.000	-0.001	0.027	0.066	0.027	0.021	0.018	1.00	1.00	1.00
-0.40	TSLS	100	-0.015	-0.023	0.068	0.152	0.070	0.053	0.044	2.28	2.23	2.18
		50	-0.015	-0.023	0.069	0.155	0.070	0.054	0.045	2.30	2.25	2.22
		25	-0.015	-0.023	0.069	0.156	0.071	0.054	0.044	2.31	2.25	2.19
	OLS		-0.029	-0.040	0.071	0.156	0.076	0.061	0.054	2.50	2.55	2.68
	QMLE		0.000	-0.002	0.031	0.076	0.031	0.024	0.020	1.00	1.00	1.00
-0.80	TSLS	100	-0.020	-0.029	0.059	0.132	0.063	0.049	0.042	1.63	1.64	1.68
		50	-0.020	-0.028	0.060	0.134	0.063	0.049	0.043	1.64	1.65	1.71
		25	-0.020	-0.028	0.059	0.132	0.063	0.049	0.042	1.63	1.64	1.68
	OLS		-0.041	-0.053	0.073	0.160	0.083	0.069	0.064	2.16	2.30	2.54
	QMLE		0.000	-0.002	0.039	0.095	0.039	0.030	0.025	1.00	1.00	1.00
						$\eta = 8.1$						
-0.10	TSLS	100	-0.011	-0.014	0.105	0.271	0.105	0.084	0.072	4.96	4.98	5.04
		50	-0.006	-0.011	0.117	0.307	0.118	0.095	0.082	5.53	5.60	5.76
		25	0.000	-0.007	0.126	0.327	0.126	0.101	0.087	5.93	5.99	6.07
	OLS		-0.012	-0.019	0.052	0.112	0.053	0.040	0.033	2.49	2.36	2.33
	QMLE		0.000	-0.001	0.021	0.054	0.021	0.017	0.014	1.00	1.00	1.00
-0.20	TSLS	100	-0.010	-0.014	0.083	0.200	0.083	0.065	0.053	3.77	3.68	3.53
		50	-0.009	-0.014	0.086	0.211	0.087	0.068	0.056	3.93	3.85	3.74
		25	-0.008	-0.013	0.089	0.214	0.089	0.069	0.057	4.04	3.93	3.79
	OLS		-0.013	-0.022	0.054	0.117	0.056	0.042	0.036	2.53	2.42	2.38
	QMLE		0.000	-0.001	0.022	0.056	0.022	0.018	0.015	1.00	1.00	1.00
-0.40	TSLS	100	-0.010	-0.015	0.060	0.138	0.061	0.046	0.038	2.47	2.35	2.25
		50	-0.009	-0.016	0.061	0.140	0.061	0.047	0.038	2.48	2.38	2.27
		25	-0.009	-0.015	0.061	0.141	0.062	0.047	0.038	2.49	2.38	2.26
	OLS		-0.018	-0.028	0.060	0.130	0.062	0.048	0.041	2.53	2.46	2.46
	QMLE		0.000	-0.001	0.025	0.063	0.025	0.020	0.017	1.00	1.00	1.00
-0.80	TSLS	100	-0.013	-0.020	0.053	0.118	0.054	0.041	0.035	1.79	1.73	1.71
		50	-0.012	-0.019	0.053	0.119	0.055	0.042	0.035	1.81	1.74	1.73
		25	-0.012	-0.019	0.053	0.119	0.054	0.041	0.034	1.80	1.73	1.69
	OLS		-0.028	-0.038	0.064	0.141	0.070	0.056	0.050	2.31	2.34	2.49
	QMLE		0.000	-0.001	0.030	0.077	0.030	0.024	0.020	1.00	1.00	1.00

Notes to Tables 3–5. All simulations are conducted for the ARCH(1) model with $\omega_0 = 0.005$ and $\alpha_0 = 0.25$, where each parameter is selected to match the empirical features of high frequency financial returns. Simulations are conducted on samples of 500 (Table 3), $1,000$ (Table 4) and $10,000$ (Table 5) observations across $10,000$ trials where, within each trial, the first 200 observations are dropped to avoid initialization effects. The estimators under study are TSLS, OLS, and QMLE. For TSLS, instrument vectors of 100, 50, and 25 lags are considered. Summary statistics are the mean bias and median bias, each measured relative to the true parameter value, the standard deviation, decile range (the difference between the 90th and 10th percentiles), and the root mean squared error, mean absolute error, and median absolute error, also each measured relative to the true parameter value. The Efficiency Ratio is the root mean squared error, mean absolute error, and median absolute error of the given estimator divided by the corresponding measure for the QMLE. $\{\epsilon_t\}$ is drawn from the student's t density of Hansen (1994) for the listed $(\lambda,\ \eta)$ pairs. Skewness and (tail) index estimates for $\{Y_t\}$ that correspond with each $(\lambda,\ \eta)$ pair are summarized in Table 1.

FIGURE 1
Hill Plots for Select FX (Absolute) 20-Min Log-Returns
Date Range: Jan 1, 2015--May 31, 2015







Notes to Figure 1:

This Figure depicts Hill (1975) tail index estimates for Japanese Yen, Euro, and Swiss Franc exchange rates (all measured against the US Dollar) at decreasing thresholds. The salient features of this Figure are summarized in the Introduction of the paper.