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

Closed-Form Estimation of Finite-Order ARCH Models: Asymptotic Theory and Finite-Sample Performance

Todd Prono

2016-083

Please cite this paper as:
Prono, Todd (2016). "Closed-Form Estimation of Finite-Order ARCH Models: Asymptotic Theory and Finite-Sample Performance," Finance and Economics Discussion Series 2016-083. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2016.083r1.

Closed-Form Estimation of Finite-Order ARCH Models:
Asymptotic Theory and Finite-Sample Performance[1]

Todd Prono[2]

This Version: July 2017

Abstract

Strong consistency and weak distributional convergence to highly non-Gaussian limits are established for closed-form, two stage least squares (TSLS) estimators for a class of ARCH(p) models. Conditions for these results include (relatively) mild moment existence criteria that are supported empirically by many (high frequency) financial returns. These conditions are not shared by competing closed-form estimators like OLS. Identification of these TSLS estimators depends on asymmetry, either in the model's rescaled errors or in the conditional variance function. Monte Carlo studies reveal TSLS estimation to sizably outperform quasi maximum likelihood estimation in (relatively) small samples. This outperformance is most pronounced when returns are heavily skewed.

Keywords: ARCH, closed form, two stage least squares, instrumental variables, heavy tails, regular variation. JEL codes: C13, C22, C58.

[1] I owe thanks to Dennis Kristensen for detailed comments on an earlier version of this paper, to Blazej Mazur, Travis Nesmith, Dong Hwan Oh, and Olivier Scaillet, as well as to participants at the 2016 Meeting of the Midwest Econometrics Group, the 25$^{\text{th}}$ Symposium of the Society of Nonlinear Dynamics and Econometrics, and the 3$^{\text{rd}}$ International Workshop on Financial Markets and Nonlinear Dynamics 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

Since being introduced by Engle (1982), autoregressive conditional heteroskedastic (ARCH) models have become the workhorse of conditional variance modeling in financial economics. The original model has been extended and generalized in various ways (see; e.g., Bollerslev et al., 1992). The most popular estimator for these types of models is quasi-maximum likelihood (QML). The asymptotic properties of QML estimation of the linear ARCH model (Engle, 1982) are well studied (see; e.g., Weiss, 1986, and more recently, Jensen and Rahbek, 2004, and Kristensen and Rahbek, 2005). However, OLS estimation of the linear ARCH model is also possible, with the accompanying advantage over QMLE being a closed-form solution. Weiss (1986) is (among) the first to consider the asymptotic properties of OLS estimation of the linear ARCH model under very restrictive moment existence criteria, while Francq and Zakoïan (2000) provide important generalizations under comparable conditions. Since the linear ARCH model implies a set of Yule-Walker equations for the squared returns (see; e.g., Mikosch and Straumann, 2002), the Whittle estimator proposed by Giratis and Robinson (2001), the asymptotic properties for which they derive under conditions comparable to Francq and Zakoïan (2000), also fits within the paradigm of closed-form, linear ARCH estimators, because it is asymptotically equivalent to Yule-Walker estimation. More recently, Kristensen and Linton (2006) provide asymptotic theory that relaxes the restrictive conditions in Weiss (1986) and Francq and Zakoïan (2000) for establishing the distributional limit (now highly non-Gaussian) and rate of convergence of the OLS estimator for the linear ARCH model, while Mikosch and Straumann (2002) make an analogous contribution (with the same, qualitative, form for the distributional limit as in Kristensen and Linton, 2006) to the asymptotic properties of the Giratis and Robinson (2001) Whittle estimator. A necessary condition underlying even these more recent works, however, is a well-defined fourth moment for the (raw) returns being modeled. Unfortunately, and in many instances, this condition appears to be violated empirically (see; e.g., Loretan and Phillips, 1994, Embrechts, Klüppelberg, and Mikosch, 1997, and Hill and Renault, 2012).

In light of an ill-defined fourth moment for many of the financial returns to which ARCH-type models are commonly applied, this paper proposes closed-form, two stage least squares (TSLS) estimators for a class of ARCH(p) models that are comparable to Francq and Zakoïan (2000), but involve different instruments. Strong consistency and weak distributional convergence to highly non-Gaussian limits comparable (qualitatively) to those discovered in Mikosch and Straumann

(2002), Kristensen and Linton (2006), and Vaynman and Beare (2014) are established for these estimators, including under the condition where the fourth moment of the returns being modeled is ill-defined. These closed-form, TSLS estimators apply to linear ARCH models and the threshold ARCH model of Glosten, Jagannathan, and Runkle (1993). To my knowledge, no attention is paid in the literature to establishing the asymptotic properties of closed-form estimators for the threshold ARCH model.

Identification of the proposed TSLS estimators links to asymmetry, either in the distribution of rescaled errors in the linear ARCH model or in the specification of the conditional variance function in the threshold ARCH model. The large-sample properties of these estimators are derived by extending results in Davis and Mikosch (1998) and Mikosch Stărică (2000) to include this necessary asymmetry. Relative to estimators for ARCH(p) models that are asymptotically normal with a convergence rate equal to the square root of the sample size, these TSLS estimators converge (quite a bit) more slowly (especially, in empirically-relevant cases) and to a distributional limit that, while stable, lacks a well-defined variance. Not surprising, then, Monte Carlo experiments reveal QML estimation of the linear ARCH model to be (quite a bit) more efficient than TSLS estimation, in large samples. What is surprising, though, is that Monte Carlo experiments also reveal TSLS estimation of the linear ARCH model to be (quite a bit) more efficient than QML estimation, in small samples, when the return distribution is (heavily) skewed. This latter finding evidences TSLS estimators (above and beyond their relative simplicity) to possess improved finite-sample properties over the QMLE alternative.

1.2 Background and Motivation

Consider the ARCH(1) model of

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

where D is some zero-mean, unit-variance distribution. For this model, it is well known that

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

where $\{W_t\}$ is a martingale difference sequence (MDS). In other words, the ARCH(1) model implies an AR(1) model for the second-order return sequence. Given (1), it is apparent that OLS can be used to estimate the parameters of the model. Let $\gamma \equiv E\left(Y_t^2\right)$, and $X_t \equiv Y_t^2 - \gamma$. From (1), given sufficient regulatory conditions, it also follows that

$$E\left(X_t X_{t-m}\right) = \alpha^m E\left(X_t^2\right), \qquad m \geq 1, \tag{2}$$

from which it is apparent that consistency of OLS requires $E\left(Y_t^4\right) < \infty$. Based on results from Kuersteiner (2002), Guo and Phillips (2001) consider improving the efficiency of OLS by defining as an instrument for Y_{t-1}^2 an infinite, weighted sum of past W_{t-1-i} for $i \geq 0$. Given (2), either OLS applied to (1) or the instrumental variables (IV) estimator of Guo and Phillips (2001) is based upon the second-order autocovariances of returns.[3] In instances where D is heavy-tailed relative to the normal, these estimators might prove favorable to the QMLE, since the latter is known to under-represent the second-order autocovariances, in these cases (see; e.g., Jacquier, Polson, and Rossi, 1994 and Baillie and Chung, 1999). For given values of ω and α, however, there is also (certainly) a limit to how heavy-tailed D can be, while still preserving a well-defined fourth moment for Y_t. Empirical evidence suggests exceedance of this limit for many financial return series.

Figure 1 plots Hill (1975) tail index estimates 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 upper confidence band is inside of 4. Moreover, currency returns sampled at this (very) high frequency are known to display relatively less volatility persistence (and, hence, relatively thinner tails) then currency returns measured at lower frequencies (like hourly or daily) or equity returns measured at any frequency equal to or higher than daily (see; e.g., Anderson and Bollerslev, 1997). Overall then, it is clear that standard $\sqrt{n}$ asymptotics for OLS applied to (1) 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

[3] This same statement also applies to the TSLS estimator of Francq and Zakoïan (2000) and the Whittle estimator of Giratis and Robinson (2001).

to 3, and the upper confidence bands always cover (more than) 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)."

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 can be used to identify a closed-form IV estimator for the ARCH(1) model. Consider using

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

as a vector of instruments for Y_{t-1}^2 in (1). Analogous to (2), it follow that, given regulatory conditions,

$$E\left(X_t Y_{t-m}\right) = \alpha^m E\left(Y_t^3\right), \tag{3}$$

which links a set of cross-order covariances to the third moment of Y_t. If $E\left(Y_t^3\right) \neq 0$, as argued above, then $\mathbf{Z}_{t-1}$ can be shown as a valid set of instruments for Y_{t-1}^2. In this case, $\mathbf{Z}_{t-1}$ can be used in a TSLS estimator for (1), where consistency of this estimator requires $E\left(Y_t^3\right) < \infty$, a condition that is now consistent with empirical findings.

Relying on skewness to define valid instruments is not new (see; e.g., Lewbel, 1997). The benefit of doing so when estimating the ARCH(1) model is analogous to basing an estimator on (2); specifically, a TSLS estimator based on $\mathbf{Z}_{t-1}$ chooses an α that best matches (3). 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 estimator might, also, perform well against the QMLE, in instances where this feature strays from what is predicted under normality.

The (relatively) heavy-tailed asymptotics discussed in Kristensen and Linton (2006) that apply to the OLS estimator for the ARCH(1) model, rely on the large-sample properties of the sample, second-order autocovariances in (2) that are developed in Davis and Mikosch (1998). The (relatively) heavier-tailed asymptotics that apply to the proposed TSLS estimator extend these results to the sample, cross-order covariances in (3). Doing so requires the return sequence $\{Y_t\}$ to be regularly varying. While many ARCH-type processes can be shown to be regularly varying (see; Basrak, Davis, and Mikosch, 2002), an added wrinkle in the present context is the requirement that $\{Y_t\}$ be skewed. Adapting this requirement to a demonstration of regular variation for ARCH(1) and threshold ARCH(1) processes is Lemma 3 in the Supplemental Appendix.

The same logic behind the TSLS estimator described above extends to TSLS estimation of a threshold ARCH(1) model, with the interesting additional feature that $E\left(Y_t^3\right) \neq 0$ is no longer necessary for identification. Generally, threshold ARCH models posit that tomorrow's variance depends on the sign of today's return. This specification requires separate ARCH effects for positive and negative returns. Non-zero skewness in positive and negative returns occurs naturally. As a consequence, TSLS estimation of a threshold ARCH(1) model bases identification on the asymmetric specification of the conditional variance function.

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

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega_0 + \alpha_0 Y_{t-1}^2, \tag{4}$$

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 (4),

$$\sigma_t^2 = \omega_0 + \sigma_{t-1}^2 A_t, \qquad A_t = \alpha_0 \epsilon_{t-1}^2, \tag{5}$$

which characterizes σ_t^2 as a stochastic recurrence equation (SRE). Most ARCH-type processes can be characterized as SREs and, as such, shown to be regularly varying (see Basrak, Davis, and Mikosch, 2002). Specifically, for $\mathbf{Y}_t = \begin{pmatrix} Y_t, & \ldots, & Y_{t+h} \end{pmatrix}$, where, for short hand,

$$\mathbf{Y} = \mathbf{Y}_0 = \begin{pmatrix} Y_0, & \ldots, & Y_h \end{pmatrix},$$

$\mathbf{Y}$ is regularly varying in $\mathbb{R}^{h+1}$ with tail index κ_0, if 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), but only in instances where D is symmetric (see Remark R2 in the Supplemental Appendix). Regular variation of $\mathbf{Y}$ can follow minus any need for symmetry in D (see Lemma 3 in the Supplemental Appendix) and applies to both the ARCH(1) case in (4) as well as the threshold ARCH(1) case of (21), making the result compatible with Assumption A3 below and complementary to Basrak, Davis and Mikosch (2002, Corollary 3.5 (B)).

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)* $E|\epsilon_t|^j = c_j < \infty$ *for* $j > 3$.

Under A1(i), (4) is the strong ARCH(1) model of Drost and Nijman (1993). Specifying the rescaled errors as i.i.d. is necessary for establishing the distributional limits and rates of convergence of the proposed closed-form estimators. Consistency of these estimators, however, continues to follow under the semi-strong definition of ARCH (see Prono, 2014), where (weak) dependence in the higher moments of the model's 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. Kristensen and Rahbek (2005) assume $j = 4$, while Hill and Renault (2012) present empirical findings that support $j = 4$.

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

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

for some $\underline{\omega} > 0$ *and, at least, one* $\alpha_i > 0$.

A2 heralds from Kristensen and Rahbek (2005). For the ARCH(1) case, $d = 1$. Notice that Θ is noncompact and ω is bounded below by a nonzero value.

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

Under A3, D in A1(i) is an asymmetric distribution. The direction of skewness is unconstrained. Skewness in (high frequency) returns is considered a stylized fact. This fact is exogenous to the model under consideration, yet (as will be shown) can be harnessed to identify the model. Examples where an asymmetric D is used to account for skewness in returns include Hansen (1994) 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 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 (4) follows that

$$Y_t^2 = \sigma_t^2 + W_t, \qquad W_t = \sigma_t^2 \left(\epsilon_t^2 - 1\right), \tag{6}$$

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

$$X_t = \alpha_0 X_{t-1} + W_t, \tag{7}$$

in which case, the centered second-order sequence $\{X_t\}$ follows an AR(1) process. Given that

$$E\left(Y_t^3\right) = E\left(\sigma_t^3\right) c_3^*,$$

A4 is also sufficient for $\left\{Y_t^3\right\}$ to have a well-defined and stationary mean (see Lemma 1 in the Supplemental Appendix). As a consequence, multiplying both sides of (7) by Y_{t-m} for $m \geq 1$ and taking expectations produces

$$E\left(X_t Y_{t-m}\right) = \alpha_0^m E\left(Y_t^3\right). \tag{8}$$

Consider

$$\mathbf{Z}_{t-1} = \left(Y_{t-1}, \ldots, Y_{t-h}\right)' \tag{9}$$

for $h < \infty$. Then $E\left(W_t \mathbf{Z}_{t-1}\right) = 0$ by iterative expectations and, owing to (8),

$$E\left(X_{t-1}\mathbf{Z}_{t-1}\right) = E\left(Y_t^3\right) \times \begin{pmatrix} 1, & \alpha_0, & \ldots, & \alpha_0^{h-1} \end{pmatrix}', \tag{10}$$

making $\mathbf{Z}_{t-1}$ a valid set of instruments for X_{t-1}. For the observed sequence $\{Y_t\}_{t=1}^{n}$, consider then

$$\widehat{\alpha}^{IV} = \frac{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \widehat{X}_t \mathbf{Z}_{t-1}\right)}{\left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)' \widehat{\mathbf{\Lambda}} \left(\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)}, \tag{11}$$

$$\widehat{\omega}^{IV} = \widehat{\gamma}\left(1 - \widehat{\alpha}^{IV}\right), \tag{12}$$

where

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

noting that both $\widehat{\alpha}^{IV}$ and $\widehat{\omega}^{IV}$ are variance-targeted estimators (VTEs).[4]

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

Suppose $\widehat{\mathbf{\Lambda}} = \left(n^{-1}\sum_t \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}\right)^{-1}$. In this case, $\widehat{\alpha}^{IV}$ is a TSLS estimator. Alternatively, if $\widehat{\mathbf{\Lambda}} = \left(n^{-1}\sum_t \left(X_t - \widetilde{\alpha}X_{t-1}\right)^2 \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}\right)^{-1}$, $\widehat{\alpha}^{IV}$ is a two-step GMM estimator, where $\widetilde{\alpha}$ is a preliminary estimate. While the two-step GMM version of (11) is certainly preferable on efficiency grounds, it requires $E\left(A^3\right) < 1$ 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),

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

where $\mathbf{I}_h$ is the $(h \times h)$ identity matrix, by the Ergodic Theorem, given only the milder condition 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 (11) are not efficient in the sense of Kuersteiner (2002). Making them so, however, requires $E\left(A^3\right) < 1$ and, hence, is limited to the thin-tailed case.

[4] VTE for ARCH-type models is first introduced by Engle and Mezrich (1996) in a QMLE context, while the asymptotic theory for this estimator is studied by Francq, Horváth, and Zakoïan (2011) and Vaynman and Beare (2014).

THEOREM 1. *Consider the estimators in (11) and (12) for the model in (7). Let*

$$\mathbf{A}_0 = E\left(X_{t-1}\mathbf{Z}_{t-1}\right)' \mathbf{\Lambda}_0; \qquad B_0 = E\left(X_{t-1}\mathbf{Z}_{t-1}\right)' \mathbf{\Lambda}_0 E\left(X_{t-1}\mathbf{Z}_{t-1}\right).$$

Let Assumptions A1–A5 hold. Then

$$\widehat{\alpha}^{IV} \xrightarrow{a.s.} \alpha_0, \qquad \widehat{\omega}^{IV} \xrightarrow{a.s.} \omega_0.$$

In addition,

$$na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) \xrightarrow{d} B_0^{-1}\mathbf{A}_0\mathbf{V}_h \tag{13}$$

if $\kappa_0 \in (3,\ 6)$, *where the vector* $\mathbf{V}_h = \left(\begin{array}{ccc} V_1, & \ldots, & V_h \end{array}\right)'$ *is jointly* $(\kappa_0/3)$*−stable, with components* $(V_m)_{m=1,\ldots,h}$ *defined in Lemma 5 of the Supplemental Appendix, and*

$$na_n^{-3}\left(\widehat{\omega}^{IV} - \omega_0\right) = -\gamma_0 na_n^{-3}\left(\widehat{\alpha}^{IV} - \alpha_0\right) + o_p(1). \tag{14}$$

Alternatively, if $E\left(A^3\right) < 1$ *so that* $E\left(Y_t^6\right) < \infty$ *and* $\kappa_0 \in (6,\ \infty)$*, then*

$$\sqrt{n}\left(\widehat{\alpha}^{IV} - \alpha_0\right) \xrightarrow{d} N\left(0,\ \Sigma_{\alpha_0}\right) \tag{15}$$

and

$$\sqrt{n}\left(\widehat{\omega}^{IV} - \omega_0\right) \xrightarrow{d} N\left(0,\ \Sigma_{\omega_0}\right), \tag{16}$$

where

$$\Sigma_{\alpha_0} = B_0^{-2}\mathbf{A}_0 E\left(W_t^2\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)\mathbf{A}_0', \qquad \Sigma_{\gamma_0} = E\left(X_t^2\right) + 2\sum_{s=1}^{\infty} E\left(X_t X_{t-s}\right),$$

and

$$\Sigma_{\omega_0} = \Sigma_{\gamma_0} + \gamma_0^2\Sigma_{\alpha_0} - 2\gamma_0 B_0^{-1}\mathbf{A}_0\left(2\sum_{s=1}^{\infty} E\left(W_t\mathbf{Z}_{t-1}Y_{t-s}^2\right)\right). \tag{17}$$

Proof. *Proofs of all Theorems are contained in the Appendix. Statements and proofs of all Lemmas that support the Theorems are contained in the Supplemental Appendix.* ■

The IV estimator in (11) depends on the (sample) cross-order covariances from (8), 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) together with the continuous mapping theorem (see Lemma 4 and Remark R3 in the Supplemental Appendix for the CLT and Lemma 5, also in the Supplemental Appendix, for the distributional limits). The method of proof extends results from Davis and Mikosch (1998) and Mikosch and Stărică (2000) to cross-order covariances (see Lemmas 3–5 in the Supplemental Appendix) and 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.[5] The (weak) distributional limit in (13) is simply a linear combination of the distributional limits of the cross-order covariances, which are jointly stable by Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). This distributional limit consists of functionals of $\{Y_t\}$. Within this limit, the individual components of $\mathbf{V}_h$ are dependent (see Lemma 5 in the Supplemental Appendix).

A sufficient condition for (13) 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).[6] 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 $j \in (3,\ 6)$. This milder condition is better aligned with more-recent theory and empirical findings for many (high frequency) financial returns. This same milder condition also applies to the threshold ARCH(1) and ARCH(p) cases discussed in Sections 2.2 and 2.3, respectively.

In (13), the limiting distribution is not impacted by $\widehat{\gamma}$. The rate of convergence is $n^{\frac{\kappa_0 - 3}{\kappa_0}}$, which is (quite a bit) slower than the usual $\sqrt{n}$ 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. The borderline case of $\kappa_0 = 6$ is omitted for the same reasons cited in Vaynman and Beare (2014, Section 3.2).

Mentioned in the Introduction and evidenced in Theorem 1, a principal advantage of (11) over the OLS alternative is that both consistency and (weak) distributional convergence follow when $E\left(Y_t^4\right) = \infty$. This result renders (11) compatible with empirical findings for many financial return series. The cost of this result, however, is a limitation on the set of permissible distributions for the model's rescaled errors. Given this limitation, the asymptotic properties of the OLS estimator applied to (7) are derived in the Supplemental Appendix.

The distributional limit in (13) is mostly qualitative in nature, owing to a (very) awkward characteristic function that does not readily admit the construction of confidence intervals. Consider

[5] Given the values of $\widehat{\omega}$ typically encountered in practice, the described limitation does not appear to be particularly binding.

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

then

$$\widehat{\tau}_n^2 = n^{-1} \sum_t Y_t^6. \tag{18}$$

Following the same method of proof for Davis and Hsing (1995, Theorem 3.1(i)),

$$n a_n^{-6} \widehat{\tau}_n^2 \xrightarrow{d} S_0, \tag{19}$$

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,

$$\sqrt{n} \left(\frac{\widehat{\alpha}^{IV} - \alpha_0}{\widehat{\tau}_n} \right) \xrightarrow{d} \frac{\mathbf{B}_0^{-1} \mathbf{A}_0 \mathbf{V}_h}{S_0^{1/2}}, \tag{20}$$

by the continuous mapping theorem.

(20) enjoys the advantage relative to (13) of removing the unknown scaling factor a_n^{-3}. Given (20), 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 (20).[7] 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 than the subsampling method while maintaining tractability, owing to the fact that $\widehat{\alpha}^{IV}$ is closed form.

In the thin-tailed case where $E\left(A^3\right) < 1$, the distributional limit of $\widehat{\alpha}^{IV}$ becomes Gaussian, with the usual rate of convergence. (20) is helpful in illustrating this case; since, when $E\left(Y_t^6\right) < \infty$, $\widehat{\tau}_n$ 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 (10), 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.

[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.

2.2. The Threshold ARCH(1) Case

Consider next the model of

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega_0 + \alpha_{1,0} Y_{t-1}^2 \times I_{\{Y_{t-1} \geq 0\}} + \alpha_{2,0} Y_{t-1}^2 \times I_{\{Y_{t-1} < 0\}}, \tag{21}$$

which is the threshold ARCH(1) model of Glosten, Jagannathan, and Runkle (1993); henceforth, the GJR ARCH(1) model. For this model, the following SRE applies

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

As a consequence, $\{Y_t\}$ continues to have a strictly stationary solution given A4. Next, since (6) continues to hold,

$$E\left(Y_t^2\right) = \frac{\omega_0 + \alpha_{1,0} Cov\left(Y_t^2, I_{\{Y_t \geq 0\}}\right) + \alpha_{2,0} Cov\left(Y_t^2, I_{\{Y_t < 0\}}\right)}{1 - \left(\alpha_{1,0} \times P\left(Y_t \geq 0\right) + \alpha_{2,0} \times P\left(Y_t < 0\right)\right)}, \tag{22}$$

in which case,

$$\begin{aligned} X_t &= \alpha_{1,0} X_{1,t-1} + \alpha_{2,0} X_{2,t-1} + W_t \\ &= \mathbf{X}_{t-1}' \boldsymbol{\alpha}_0 + W_t, \end{aligned} \tag{23}$$

where

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

Motivated by the results in Section 2.1, consider as (potential) instruments for $\mathbf{X}_{t-1}$

$$\mathbf{Z}_{t-1} = \left(\left(Z_{1,t-1}, Z_{2,t-1}\right), \ldots, \left(Z_{1,t-h}, Z_{2,t-h}\right)\right)', \qquad h < \infty, \tag{24}$$

where

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

for $m \geq 1$.

ASSUMPTION A6: $E\left(\mathbf{Z}_{t-1}\mathbf{X}_{t-1}'\right)$ has full column rank.

A6 applies the usual rank condition for identifying IV estimators. A sufficient condition for A6 is

$$E\left(Z_{1,t-1}X_{1,t-1}\right) \times E\left(Z_{2,t-1}X_{2,t-1}\right) - E\left(Z_{1,t-1}X_{2,t-1}\right) \times E\left(Z_{2,t-1}X_{1,t-1}\right) \neq 0, \tag{25}$$

which establishes $\left(Z_{1,t-1}, Z_{2,t-1}\right)'$ as valid instruments for $\left(X_{1,t-1}, X_{2,t-1}\right)'$. Let

$$E\left(\epsilon_t^j \times I_{\{\epsilon_t \geq 0\}}\right) = c_j^+, \qquad E\left(\epsilon_t^j \times I_{\{\epsilon_t < 0\}}\right) = c_j^-, \qquad j = 1, 2, 3.$$

Given (21) then,

$$c_1^+ + c_1^- = 0, \qquad c_2^+ + c_2^- = 1, \qquad c_3^+ + c_3^- = c_3^*, \tag{26}$$

where c_3^* is defined in A3. Using (26), (25) can be restated as

$$E\left(\sigma_t^3\right) \times \left[E\left(\sigma_t^3\right) c_3^+ c_3^- - E\left(\sigma_t\right) E\left(\sigma_t^2\right) \times \left(c_1^- c_2^- c_3^+ + c_1^+ c_2^+ c_3^-\right)\right] \neq 0. \tag{27}$$

Suppose that $c_3^* = 0$, which is to say that the distribution of $\{\epsilon_t\}$ is symmetric. In this case, again using the constraints in (26), (27) is satisfied if

$$E\left(Y_t^3 \times I_{\{Y_t \geq 0\}}\right) - E\left(Y_t^2\right) \times E\left(Y_t \times I_{\{Y_t \geq 0\}}\right) \neq 0$$

and

$$E\left(Y_t^3 \times I_{\{Y_t < 0\}}\right) - E\left(Y_t^2\right) \times E\left(Y_t \times I_{\{Y_t < 0\}}\right) \neq 0,$$

depending on whether (27) is solved only in terms of c_j^+ or c_j^-, respectively. Notice that A3 is not necessary for satisfying even (25). So long as $\alpha_{1,0} \neq \alpha_{2,0}$ (i.e., there exists a threshold effect in the conditional variance), $\mathbf{Z}_{t-1}$ as defined in (24) can serve as a valid set of instruments for $\mathbf{X}_{t-1}$ regardless of whether the rescaled errors from the GJR ARCH(1) model are skewed. In this case, it is the conditional variance function itself that supplies the necessary asymmetry for identification. In the event that $\alpha_{1,0} = \alpha_{2,0}$, however, (23) reduces to (7); in which case, A3 becomes necessary for establishing validity of the instruments in (9) because, in this case, asymmetry can only come from the model's rescaled errors.[8]

[8] Notice that each instrument in (9) is a MDS. This characterization does not carry-over onto the instruments in (24). These latter instruments, while unconditionally mean-zero, are not conditionally mean-zero.

Owing to the identification condition in A6, the GJR ARCH(1) analog to (11) based upon feasible versions of $\mathbf{X}_{t-1}$ and $\mathbf{Z}_{t-1}$is

$$\widehat{\boldsymbol{\alpha}}^{IV} = \widehat{\mathbf{F}}\left(n^{-1}\sum_t \widehat{X}_t\widehat{\mathbf{Z}}_{t-1}\right), \tag{28}$$

where

$$\widehat{\mathbf{F}} = \left[\left(n^{-1}\sum_t \widehat{\mathbf{X}}_{t-1}\widehat{\mathbf{Z}}_{t-1}'\right)\widehat{\boldsymbol{\Lambda}}\left(n^{-1}\sum_t \widehat{\mathbf{X}}_{t-1}\widehat{\mathbf{Z}}_{t-1}'\right)'\right]^{-1}\left(n^{-1}\sum_t \widehat{\mathbf{X}}_{t-1}\widehat{\mathbf{Z}}_{t-1}'\right)\widehat{\boldsymbol{\Lambda}} \tag{29}$$

is a $2\times 2h$ matrix, and

$$\widehat{E}\left(Y_t^j \times I_{\{Y_t\geq 0\}}\right) = n^{-1}\sum_t Y_t^j \times I_{\{Y_t\geq 0\}}, \qquad \widehat{E}\left(Y_t^j \times I_{\{Y_t< 0\}}\right) = n^{-1}\sum_t Y_t^j \times I_{\{Y_t< 0\}}, \qquad j=1,2.$$

When $\widehat{\boldsymbol{\Lambda}} = \left(n^{-1}\sum_t \widehat{\mathbf{Z}}_{t-1}\widehat{\mathbf{Z}}_{t-1}'\right)^{-1}$, (28) is a TSLS estimator for (21), with the same discussion regarding selection of $\widehat{\boldsymbol{\Lambda}}$ in Section 2.1 remaining applicable.

THEOREM 2. *Consider the estimator in (28) for the model in (23) when $\alpha_{1,0}\neq\alpha_{2,0}$, and let*

$$\mathbf{F}_0 = \left[E\left(\mathbf{X}_{t-1}\mathbf{Z}_{t-1}'\right)\boldsymbol{\Lambda}_0 E\left(\mathbf{X}_{t-1}\mathbf{Z}_{t-1}'\right)'\right]^{-1} E\left(\mathbf{X}_{t-1}\mathbf{Z}_{t-1}'\right)\boldsymbol{\Lambda}_0.$$

In addition, let Assumptions A1–A2 and A4–A6 hold. Then,

$$\widehat{\boldsymbol{\alpha}}^{IV} \xrightarrow{a.s.} \boldsymbol{\alpha}_0.$$

In addition,

$$na_n^{-3}\left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} \mathbf{F}_0\mathbf{W}_h^{(+,-)} \tag{30}$$

if $\kappa_0 \in (3,\ 6)$, where the vector

$$\mathbf{W}_h^{(+,-)} = \begin{pmatrix} W_1^+, & W_1^-, & \ldots, & W_h^+, & W_h^- \end{pmatrix}'$$

is jointly $(\kappa_0/3)-$stable, with components $\begin{pmatrix} W_m^+ & W_m^- \end{pmatrix}_{m=1,\ldots,h}$ defined in Lemma 6 of the Supplemental Appendix. Alternatively, if $E\left(A^3\right)<1$ so that $E\left(Y_t^6\right)<\infty$ and $\kappa_0\in(6,\ \infty)$, then

$$\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} N\left(0,\ \mathbf{F}_0 E\left(W_t^2\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)\mathbf{F}_0'\right). \tag{31}$$

The main result in (30) follows from the (weak) distributional convergence of $n^{-1} \sum_t X_t \mathbf{Z}_{t-1}$ (see Lemma 6 in the Supplemental Appendix), which involves cross-order sums constructed from positive and negative realizations of $\{Y_t\}$, respectively. This result requires $\alpha_{1,0} > 0$ and $\alpha_{2,0} > 0$ (see Remark R2 in the Supplemental Appendix). 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-side and left-hand-side of the distribution of Y_t. Individual components of $\mathbf{W}_h^{(+,-)}$ are dependent (see Lemma 6 in the Supplemental Appendix). In addition, W_1^+ and W_1^- jointly depend on V_1 from Theorem 1, which connects the limiting result in (30) to that in (13). Normalizing the left-hand-side of (30) by $\widehat{\tau}_n$ as it is defined in (18) enables construction of either subsample or bootstrap confidence intervals for $\widehat{\boldsymbol{\alpha}}^{IV}$ as described following the statement of Theorem 1 in Section 2.1. In the case where $E\left(A^3\right) < 1$, $\boldsymbol{\Lambda}_0 = E\left(W_t^2 \mathbf{Z}_{t-1} \mathbf{Z}_{t-1}'\right)$ produces the minimum variance estimator so that $\widehat{\boldsymbol{\alpha}}^{IV}$ should be a two-step GMM estimator.

Note that Theorem 2 does not depend on A3. As a consequence, (28) seems to be the preferable choice for estimating (23) over OLS in the (empirically relevant) case where $E\left(Y_t^4\right) = \infty$; since, like Theorem 1, consistency and (weak) distributional convergence are supported under this case while, unlike Theorem 1, the set of permissible distributions for the model's rescaled errors includes symmetric candidates. Nevertheless, the asymptotic properties of the OLS estimator for (23) are also developed in the Supplemental Appendix.

Finally, let

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

Then, given (22),

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

so that

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

in which case, a comparable version of (14) then follows.

2.3. The ARCH(p) Case

Consider finally the model of

$$Y_t = \sigma_t \epsilon_t, \qquad \sigma_t^2 = \omega_0 + \sum_{i=1}^{p} \alpha_{i,0} Y_{t-i}^2, \qquad 1 \leq p < \infty. \tag{32}$$

ASSUMPTION A7: $c_3 \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_{i,0} \alpha_{j,0}^{1/2} < 1.$

A7 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 Supplemental Appendix).

ASSUMPTION A8: *Define* $\rho_p(\epsilon_t)$ *as the largest root of* $1 - \sum_{i=1}^{p} \lambda^i \alpha_{i,0} \epsilon_t^2$.

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

for $s = 2, 3, 4$.

Suppose $j = 2s$ in A1. Then A8 establishes $E\left(Y_t^{2s}\right) < \infty$ (see Carrasco and Chen, 2002, Proposition 13).

From Basrak, Davis, and Mikosch (2002), (32) can be recast in terms of the following SRE:

$$\widetilde{\mathbf{Y}}_t = \mathbf{A}_t \widetilde{\mathbf{Y}}_{t-1} + \mathbf{B}_t, \tag{33}$$

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 A7, (33), 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 2.1 and 2.2, let

$$\mathbf{X}_{t-1} = \begin{pmatrix} X_{t-1}, & \ldots, & X_{t-p} \end{pmatrix}'. \tag{34}$$

Then the generalization of (7) is

$$X_t = \mathbf{X}_{t-1}'\boldsymbol{\alpha}_0 + W_t, \tag{35}$$

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

$$\mathbf{Z}_{t-1} = \begin{pmatrix} Y_{t-1}, & \ldots, & Y_{t-h} \end{pmatrix}', \qquad p \leq h < \infty, \tag{36}$$

as a vector of instruments for $\mathbf{X}_{t-1}$. Given A3, $\mathbf{Z}_{t-1}$ identifies $\boldsymbol{\alpha}_0$ in (35) (see Lemma 9 in the Supplemental Appendix). Consider then the estimator

$$\widehat{\boldsymbol{\alpha}}^{IV} = \widehat{\mathbf{F}}\left(n^{-1}\sum_t \widehat{X}_t \mathbf{Z}_{t-1}\right), \tag{37}$$

where $\widehat{\mathbf{F}}$ is defined as in (29), but with $\mathbf{Z}_{t-1}$ in (36) everywhere replacing $\widehat{\mathbf{Z}}_{t-1}$, and $\widehat{\mathbf{X}}_{t-1}$ defined as the finite sample version of (34).

THEOREM 3. *Consider the estimator in (37) for the model in (35). Let Assumptions A1–A5 and A7 hold. Then,*

$$\widehat{\boldsymbol{\alpha}}^{IV} \xrightarrow{a.s.} \boldsymbol{\alpha}_0.$$

In addition,

$$na_n^{-3}\left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0\right) \xrightarrow{d} \mathbf{F}_0\mathbf{V}_{p,h} \tag{38}$$

if $\kappa_0 \in (3,\ 6)$, where the vector $\mathbf{V}_{p,h} = \begin{pmatrix} V_{p,1}, & \ldots, & V_{p,h} \end{pmatrix}'$ is jointly $(\kappa_0/3)-$stable, with components $\left(V_{p,m}\right)_{m=1,\ldots,h}$ defined in Lemma 12 of the Supplemental Appendix. Alternatively, if Assumption A8 with $s = 3$ holds so that $E\left(Y_t^6\right) < \infty$ and $\kappa_0 \in (6,\ \infty)$, then (31) results, with $\mathbf{F}_0$ being the population limit of $\widehat{\mathbf{F}}$ in (37) and $\mathbf{Z}_{t-1}$ being defined in (36).

Under Theorem 3, (38) reduces to (13) when $p = 1$. As a consequence, A3 is necessary for establishing the large sample properties of (37) (see Lemma 9 in the Supplemental Appendix). That is, in the absence of skewness, (37) neither is identified nor does it possess a stable limiting distribution. The CLT underlying (38) is Basrak, Davis, and Mikosch (2002, Theorem 2.10), which generalizes Lemma 4 in the Supplemental Appendix.[9] Application of Basrak et al. (2002, Theorem 2.10) requires $\left\{\begin{pmatrix} Y_t, & \sigma_t \end{pmatrix}\right\}$ to be regularly varying, which, in turn, is established by Basrak et. al (2002, Corollary 3.5(B)).[10] Given (18), normalization of the left-hand-side of (38) 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 $\widehat{\boldsymbol{\alpha}}^{IV}$. Lastly, A8 with $s = 3$ is the ARCH(p) analog to $E\left(A^3\right) < 1$ that is used to establish the ARCH(1) and GJR ARCH(1) estimators as asymptotically normal. A8 with $s = 2$ and $s = 4$ is used in the Supplemental Appendix to establish the large sample properties of the OLS estimator applied to (35).

The distributional limit in (38) generally differs from the special case presented in (13) in that the former is derived, in part, from (normalized) sums of $\{\sigma_t\}$ (see Lemmas 10 and 12 in the Supplemental Appendix), while the latter is derived only from (normalized) sums of $\{Y_t\}$ (see Lemma 5, also in the Supplemental Appendix). In other words, the distributional limit in (13) depends only on functionals of the observable sequence $\{Y_t\}$, while the distributional limit in (38) depends both on functionals of $\{Y_t\}$ and on functionals of the latent sequence $\{\sigma_t\}$. The complexities that arise in the cross-order covariances generated by (32) when $p > 1$ (see; e.g., Guo and Phillips, 2001, Lemma 1) necessitate this differential approach. The limit in (38), nonetheless, reduces to the limit in (13) when $p = 1$ and establishes both a stable limit and rate of convergence for (38), generally, under a method of proof that is comparable to Basrak, Davis, and Mikosch (2002, Theorem 3.6).

The differential approach in establishing (38) versus (13) 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) 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 (33) 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.

[9] Lemma 4 establishes the CLT underlying Theorems 1 and 2, respectively.

[10] In contrast, Lemma 3 in the Supplemental Appendix establishes regular variation of $\{Y_t\}$ under Theorems 1 and 2, respectively.

Lastly, since

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

and given Theorem 3, the large sample properties of $\widehat{\omega}$ can be established analogously to results presented in Theorem 1.

3. Monte Carlo

Consider the ARCH(1) model from Section 2.1, where $\{\epsilon_t\}$ is drawn from the skewed student's t density of Hansen (1994). This density has two parameters, λ and η, with the former governing skewness, the latter governing the tails, and up to the ηth moment being well defined. Table 1 summarizes the various (λ, η) 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, skewness estimates for various intra-day Japanese Yen returns (measured relative to the USD) as well as S&P 500 Index and DJIA returns are summarized in Table 2. Apparent from Table 2, 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. In light of the discussion of A1(ii) in Section 2.1, the relatively thin-tailed case of $\eta = 8.1$ is considered only to validate the large-sample properties of $\widehat{\alpha}_{IV}$ predicted by Theorem 1 and $\widehat{\alpha}_{OLS}$ predicted by Proposition 1 in the Supplemental Appendix. Given Kristensen and Rahbek (2005) and the empirical findings of Hill and Renault (2012), the case where $\eta = 4.1$ is considered more realistic. Lastly, for all (λ, η) pairs considered, A4 is satisfied so that $E\left(Y_t^3\right) < \infty$.

Across all simulations, $\omega_0 = 0.005$ and $\alpha_0 = 0.25$.[11] As noted in Table 1 by the tail indices, when $\eta = 8.1$, $E\left(Y_t^4\right) < \infty$. In these cases, the simulations study the TSLS, OLS and QML estimators of the ARCH(1) model. When $\eta = 4.1$, $E\left(Y_t^4\right) = \infty$; in which case, only the TSLS and QML estimators are studied. For the TSLS estimator, simulations consider $h = 100,\ 50,\ 25,$ where h is the longest lag included in the instrument vector. Sample sizes for the simulations are $100,000$, $1,000$, and 500, the first of which is considered to validate the large-sample properties of $\widehat{\alpha}_{IV}$ and $\widehat{\alpha}_{OLS}$, respectively. The (relatively) small sample sizes are only considered under the heavy-tailed

[11] 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.

case of $\eta = 4.1$. These cases consider the finite-sample performance of the TSLS estimator relative to the QMLE in instances (far) removed from normality that, nonetheless, remain empirically grounded. All Monte Carlo experiments are conducted across $10,000$ simulation trials. Additional details on the experiments are contained in the notes to Tables 3 and 4.

Table 3 summarizes the large sample results ($T = 100,000$). The top panel depicts the relatively thin-tailed case of $\eta = 8.1$. The bias in TSLS and OLS is small, although elevated relative to QML. In addition, OLS is more biased than TSLS, with this difference in bias widening as skewness in $\{\epsilon_t\}$ increases.[12] In a comparison of efficiency ratios (all measured against the QMLE), TSLS and OLS are both notably less efficient than QML.[13] As skewness increases, the gap in efficiency between TSLS and QML shrinks, although it remains sizable in absolute terms. The efficiency gap between OLS and QML, in contrast, widens as skewness increases. Finally, at relatively low levels of skewness, OLS appears more efficient than TSLS. At moderate to high levels of skewness, however, TSLS appears more efficient than OLS, and by fairly wide margins. Lastly, there does not appear to be much difference, either in terms of bias or in terms of dispersion, from using more lagged instruments in TSLS.

The bottom panel of Table 3 summarizes results from the heavy-tailed case where $\eta = 4.1$. In this case, OLS is not consistent, explaining its exclusion from consideration. TSLS is more biased in this case than in the case where $\eta = 8.1$.[14] Interestingly, though, the efficiency gap between TSLS and QML is smaller in this case than in the case where $\eta = 8.1$. As is true in the top panel of Table 3, this efficiency gap shrinks as skewness increases. In addition, there continues to be only modest differences in terms of bias and dispersion between TSLS with instrument vectors based on longer lag lengths.

Table 4 summarizes the small sample results ($T = 1,000$ and $T = 500$). Relative to the bottom panel of Table 3, the bias in TSLS is notably elevated, where this bias increases with the level of skewness. Interestingly, QML now also displays notable bias, where this bias, too, increases with the level of skewness. Most interestingly, the efficiency gap between TSLS and QML is now materially reduced. Moreover, in many instances, this gap is reversed, with TSLS evidencing sizable

[12] As skewness increases, the tail index decreases, thus causing the rate of convergence in $\widehat{\alpha}_{OLS}$ to also slow. Note, as well, that the convergence rate of $\widehat{\alpha}_{IV}$ should be faster than $\widehat{\alpha}_{OLS}$.

[13] This finding, perhaps, is not too surprising given the relative rates of convergence of the three estimators and the differences in distributions to which each estimator converges.

[14] This relative increase in bias is explained by the decrease in tail indices across the different levels of skewness considered (see Table 1). With each of these tail indices near 3, the rate of convergence in $\widehat{\alpha}_{IV}$ is anticipated to be rather slow overall, and slower than in the case where $\eta = 8.1$.

efficiency gains over QML. Specifically, for the sample size of $T = 1,000$, TSLS bests QML in terms of efficiency ratios at moderate and high skewness levels. For the smaller sample size of $T = 500$, TSLS bests QML in terms of efficiency ratios across all skewness levels. At the highest skewness level when $T = 500$, TSLS sizably outperforms QML. Also noteworthy, there still does not appear to be much cost in terms of sacrificed efficiency from using "many" lagged instruments.[15]

Lastly, the simulation results presented in this section immediately apply to the estimator in (28). That estimator depends on the third moment of returns conditional on those returns being either greater than or equal to or less than zero. Empirically, skewness in positive and negative equity returns is large, comparable in magnitude to the skewness levels included in the simulation designs.[16]

4. Conclusion

This paper proposes closed-form, TSLS estimators for a class of univariate ARCH(p) models. The instruments used in these estimators are not currently considered in the literature. The advantage of these instruments is that they allow the asymptotic theory for these estimators to follow under moment-existence criteria that are consistent with the empirical findings for many financial return series to which ARCH-type models are commonly applied. This characteristic renders the proposed TSLS estimators empirically feasible, a characteristic that is not shared by competing, closed-form estimators like OLS. Identification of these TSLS estimators links to asymmetry; either in the model's rescaled errors as in the ARCH(p) case, or in the specification of the conditional variance function itself as in a threshold ARCH(1) case. The asymptotic theory for these estimators extends results from Davis and Mikosch (1998) and Mikosch and Stărică (2000) to cross-order covariances (defined as covariances between contemporaneous second-order returns and lagged first-order returns), which become relevant for identification in instances of return asymmetry. These TSLS estimators are also shown to outperform QML in finite samples, confirming the conjecture of Bollerslev and Wooldridge (1992) that construction of an IV estimator for ARCH-type models more efficient than QMLE is possible.

[15] There does appear to be some increase in bias that results from using more instruments; however, this cost is counter-balanced against reductions in dispersion.

[16] The GJR ARCH model, specifically, and threshold ARCH models, generally, are applied to equity returns to account for the so called "leverage effect."

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 (7) 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 conditional on ϕ_0. An interesting investigation, therefore, is whether the closed-form TSLS 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 (Proofs of the Theorems)

PROOF OF THEOREM 1. Note that

$$\widehat{X}_t = X_t - \left(\widehat{\gamma} - \gamma_0\right), \tag{39}$$

and

$$\widehat{X}_t = \overline{c} + \alpha_0 \widehat{X}_{t-1} + W_t, \tag{40}$$

where $\overline{c} = \left(\alpha_0 - 1\right)\left(\widehat{\gamma} - \gamma_0\right)$. Then given (40),

$$\begin{aligned} \widehat{\alpha}^{IV} \quad = \quad & \alpha_0 + \left(\frac{\overline{c}\left(n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)'\widehat{\mathbf{\Lambda}}}{\left(n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)'\widehat{\mathbf{\Lambda}}\left(n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)} \right) \times \left(n^{-1}\sum_t \mathbf{Z}_{t-1}\right) \\ & + \left(\frac{\left(n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)'\widehat{\mathbf{\Lambda}}}{\left(n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)'\widehat{\mathbf{\Lambda}}\left(n^{-1}\sum_t \widehat{X}_{t-1}\mathbf{Z}_{t-1}\right)} \right) \times \left(n^{-1}\sum_t W_t\mathbf{Z}_{t-1}\right) \end{aligned} \tag{41}$$

By Carrasco and Chen (2002, Corollary 6), $\{Y_t\}$ is strong mixing. As a consequence, given (8) and A3, $\widehat{\alpha}^{IV} \overset{a.s.}{\to} \alpha_0$, and $\widehat{\omega}^{IV} \overset{a.s.}{\to} \omega_0$ by the Ergodic Theorem. Next, given (39) and noting that the population analog to $\widehat{\alpha}^{IV}$ in (11) is α_0,

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

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

$$
\begin{aligned}
a_n^{-3}\sum_t X_t\mathbf{Z}_{t-1}-E\left(X_t\mathbf{Z}_{t-1}\right) &= a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1}-E\left(Y_t^2\mathbf{Z}_{t-1}\right) \\
&\quad -\gamma_0 n^{\frac{\kappa_0-6}{2\kappa_0}}\left(n^{-1/2}\sum_t\mathbf{Z}_{t-1}\right) \\
&= a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1}-E\left(Y_t^2\mathbf{Z}_{t-1}\right)+o_P\left(1\right)
\end{aligned}
\tag{42}
$$

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

$$
\begin{aligned}
na_n^{-3}\left(\widehat{\omega}^{IV}-\omega_0\right) &= -\gamma_0 na_n^{-3}\left(\widehat{\alpha}^{IV}-\alpha_0\right)+na_n^{-3}\left(\widehat{\gamma}-\gamma_0\right) \\
&= -\gamma_0 na_n^{-3}\left(\widehat{\alpha}^{IV}-\alpha_0\right)+o_P\left(1\right),
\end{aligned}
\tag{43}
$$

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 (41),

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

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 (39), 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

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

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

so that, comparable to (40),

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

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

$$\widehat{\boldsymbol{\alpha}}^{IV}-\boldsymbol{\alpha}_0=\widehat{\mathbf{F}}\left[\overline{c}\left(n^{-1}\sum_t\mathbf{Z}_{t-1}'\right)-\left(\widehat{\mathbf{H}}-\mathbf{H}_0\right)\left(n^{-1}\sum_t W_t\right)\right]+\widehat{\mathbf{F}}\left(n^{-1}\sum_t W_t\mathbf{Z}_{t-1}\right),\quad (44)$$

from which $\widehat{\alpha}^{IV}\overset{a.s.}{\to}\alpha_0$, where identification follows from A7 and (almost sure) convergence in the sample moments follows from the Ergodic Theorem, since $\{Y_t\}$ remains strong mixing, this time by Carrasco and Chen (2002, Corollary 10). Next, from (28),

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

such that

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

Let $\mathbf{Z}_{t-1}=\mathbf{Z}_{t-1}^{(1)}-\mathbf{H}_0$. Given the arguments that support the second equalities in both (42) and (43),

$$\begin{aligned}a_n^{-3}\sum_t X_t\mathbf{Z}_{t-1}-E\left(X_t\mathbf{Z}_{t-1}\right) &= a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1}^{(1)}-E\left(Y_t^2\mathbf{Z}_{t-1}^{(1)}\right)\\ &\quad -\left(\mathbf{H}_0a_n^{-3}\sum_t Y_t^2-E\left(Y_t^2\right)+\gamma_0a_n^{-3}\sum_t\mathbf{Z}_{t-1}\right)\\ &= a_n^{-3}\sum_t Y_t^2\mathbf{Z}_{t-1}^{(1)}-E\left(Y_t^2\mathbf{Z}_{t-1}^{(1)}\right)+o_P(1)\end{aligned}$$

such that

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

where $\mathbf{W}_h^{(+,-)}$ is jointly $(\kappa_0/3)-$stable by Lemma 6 in the Supplemental Appendix and Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Finally, from (44),

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

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

PROOF OF THEOREM 3. Let $\boldsymbol{\iota}$ be a $p\times 1$ vector of ones. Given (34),

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

Then given (39),

$$\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 = \widehat{\mathbf{F}} \left(\bar{c} \left(n^{-1} \sum_t \mathbf{Z}_{t-1} \right) + n^{-1} \sum_t W_t \mathbf{Z}_{t-1} \right), \tag{45}$$

where $\bar{c} = (\boldsymbol{\iota}' \boldsymbol{\alpha}_0 - 1)(\widehat{\gamma} - \gamma_0)$. By Lemma 9 in the Supplemental Appendix, $E\left(\mathbf{Z}_{t-1}\mathbf{X}_{t-1}'\right)$ has full column rank. By Carrasco and Chen (2002, Proposition 12), $\{Y_t\}$ remains strong mixing. Then by the Ergodic Theorem, $\widehat{\boldsymbol{\alpha}}^{IV} \xrightarrow{a.s.} \boldsymbol{\alpha}_0$. Next, given (39),

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

so that

$$n a_n^{-3} \left(\widehat{\boldsymbol{\alpha}}^{IV} - \boldsymbol{\alpha}_0 \right) = \mathbf{F}_0 \left(a_n^{-3} \sum_t Y_t^2 \mathbf{Z}_{t-1} - E\left(Y_t^2 \mathbf{Z}_{t-1}\right) \right) + o_p(1), \tag{46}$$

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 (42). Then by Lemma 12 in the Supplemental Appendix,

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

where $\mathbf{V}_{p,h}$ is jointly $(\kappa_0/3)$ –stable by Samorodnitsky and Taqqu (1994, Theorem 2.1.5(c)). Finally, from (45),

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

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. R.Y. Chou & K.F. Kroner (1992) ARCH modeling in finance. *Journal of Econometrics* 52, 5–59.

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

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

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

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

[10] 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.

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

[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] Francq, C., L. Horváth & J.M. Zakoïan (2000) Estimating weak GARCH representations. *Econometric Theory* 16, 692-728.

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

[19] 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.

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

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

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

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

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

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

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

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

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

[29] Jensen, S.T. & A. Rahbek (2004) Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. *Econometrica* 72, 641-646.

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

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

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

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

[34] 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.

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

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

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

[38] 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.

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

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

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

[42] 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.

[43] von Bahr, B., & C.G. Esseen (1965) Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Annals of Mathematical Statistics* 36, 299-303.

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

TABLE 1

λ	η	skew.	κ
-0.20	4.1	-1.27	3.50
-0.40		-2.32	3.32
-0.80		-3.48	3.14
-0.20	8.1	-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 $(\lambda,\ \eta)$ 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	mae	mdae
						$\eta = 8.1$						
-0.20	TSLS	100	-0.001	-0.004	0.038	0.085	0.038	0.028	0.022	5.29	4.89	4.58
		50	-0.001	-0.004	0.039	0.085	0.039	0.028	0.022	5.33	4.93	4.68
		25	-0.001	-0.004	0.039	0.086	0.039	0.028	0.022	5.38	4.97	4.68
	OLS		-0.005	-0.010	0.034	0.064	0.034	0.023	0.018	4.76	4.12	3.89
	QMLE		0.000	0.000	0.007	0.018	0.007	0.006	0.005	1.00	1.00	1.00
-0.40	TSLS	100	-0.002	-0.005	0.028	0.059	0.028	0.020	0.016	3.54	3.14	2.95
		50	-0.002	-0.005	0.028	0.059	0.028	0.020	0.016	3.54	3.14	2.95
		25	-0.002	-0.005	0.028	0.059	0.029	0.020	0.016	3.54	3.14	2.93
	OLS		-0.008	-0.015	0.040	0.077	0.041	0.029	0.023	5.08	4.55	4.40
	QMLE		0.000	0.000	0.008	0.020	0.008	0.006	0.005	1.00	1.00	1.00
-0.80	TSLS	100	-0.003	-0.008	0.028	0.056	0.028	0.020	0.016	2.90	2.54	2.42
		50	-0.003	-0.008	0.028	0.055	0.028	0.020	0.016	2.89	2.53	2.43
		25	-0.003	-0.008	0.028	0.056	0.028	0.020	0.016	2.88	2.52	2.41
	OLS		-0.015	-0.023	0.046	0.091	0.049	0.037	0.031	5.02	4.72	4.73
	QMLE		0.000	0.000	0.010	0.025	0.010	0.008	0.007	1.00	1.00	1.00
						$\eta = 4.1$						
-0.20	TSLS	100	-0.017	-0.027	0.081	0.179	0.083	0.062	0.050	4.63	4.90	4.96
		50	-0.017	-0.027	0.082	0.181	0.083	0.063	0.049	4.66	4.92	4.87
		25	-0.017	-0.027	0.082	0.178	0.084	0.063	0.050	4.68	4.92	4.93
	QMLE		0.000	-0.002	0.018	0.039	0.018	0.013	0.010	1.00	1.00	1.00
-0.40	TSLS	100	-0.021	-0.031	0.061	0.125	0.065	0.049	0.042	2.89	3.20	3.45
		50	-0.021	-0.031	0.061	0.124	0.065	0.049	0.042	2.88	3.19	3.43
		25	-0.021	-0.031	0.061	0.124	0.065	0.049	0.041	2.87	3.18	3.42
	QMLE		0.000	-0.002	0.023	0.047	0.023	0.015	0.012	1.00	1.00	1.00
-0.80	TSLS	100	-0.030	-0.040	0.055	0.113	0.063	0.051	0.047	2.12	2.52	2.93
		50	-0.030	-0.040	0.055	0.112	0.062	0.050	0.046	2.11	2.50	2.90
		25	-0.030	-0.039	0.055	0.111	0.062	0.050	0.046	2.11	2.50	2.89
	QMLE		0.000	-0.003	0.029	0.061	0.029	0.020	0.016	1.00	1.00	1.00

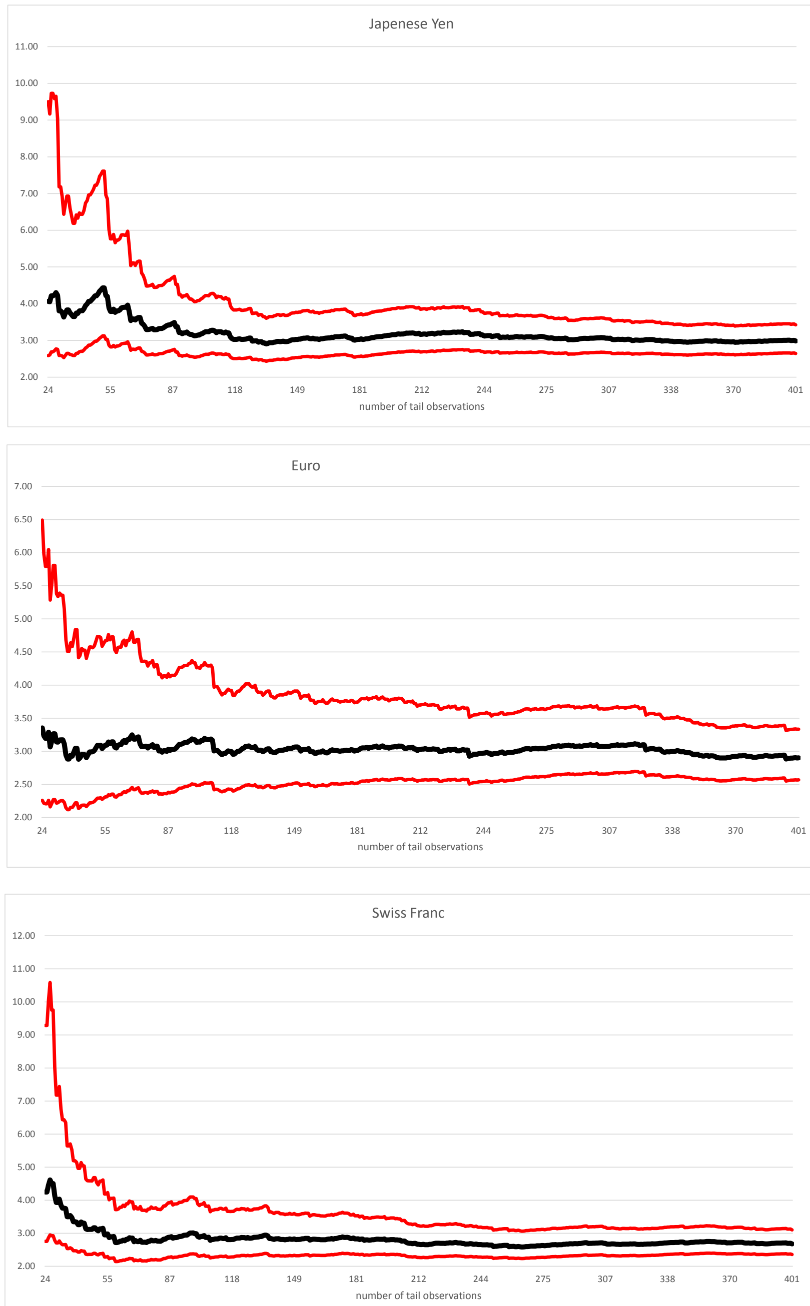
Notes to Tables 3. The ARCH(1) model is considered with $\omega_0 = 0.005$ and $\alpha_0 = 0.25$. Simulations are conducted on samples of $T = 100,000$ observations across $10,000$ trials. Within each simulation trial, the first 200 observations are dropped to avoid initialization effects. In the case where $\eta = 8.1$, the estimators under study are TSLS, OLS, and QMLE. When $\eta = 4.1$, only the TSLS and QMLE estimators are considered, owing to the insufficient existence of higher moments needed to render OLS consistent. 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 (λ, η) pairs. Skewness and (tail) index estimates for $\{Y_t\}$ that correspond with each (λ, η) pair are summarized in Table 1.

TABLE 4

λ	est.	m	mean bias	med. bias	sd	dec. rge.	rmse	mae	mdae	Efficiency Ratio rmse	mae	mdae
						$T = 1,000$						
-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
	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
	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
	QMLE		-0.005	-0.045	0.162	0.375	0.162	0.123	0.103	1.00	1.00	1.00
						$T = 500$						
-0.20	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
	QMLE		-0.007	-0.035	0.149	0.352	0.149	0.114	0.096	1.00	1.00	1.00
-0.40	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
	QMLE		-0.007	-0.044	0.164	0.389	0.164	0.127	0.108	1.00	1.00	1.00
-0.80	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
	QMLE		-0.010	-0.064	0.192	0.453	0.192	0.150	0.132	1.00	1.00	1.00

Notes to Tables 4. The ARCH(1) model is considered with $\omega_0 = 0.005$ and $\alpha_0 = 0.25$. Simulations are conducted on samples of either $T = 1,000$ or $T = 500$ observations across $10,000$ trials. Within each simulation trial, the first 200 observations are dropped to avoid initialization effects. In both panels, $\eta = 4.1$, so only the TSLS and QMLE estimators are considered, owing to the insufficient existence of higher moments needed to render OLS consistent. 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 Section 1.2 of the paper. All data sources to Bloomberg.