Simple Estimators for the GARCH(1,1) Model

Supplemental Appendix

Todd Prono

I. Introduction

This Supplemental Appendix contains the complete Monte Carlo results (Section II) in support of the closed-form and iterative estimators proposed in the main text. Moreover, proofs (Section III) to all Theorems and Corollaries stated in the main text are given here. Select materials from the main text that facilitate an understanding of either the Monte Carlo experiments or the proofs are restated in Section II for convenience.

I. Preliminaries

For $\{Y_t\}_{t\in\mathbb{Z}}$ with σ-algebra $\mathcal{F}_t$ such that $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \cdots \subseteq \mathcal{F}$, the model under consideration is

$$E\left[Y_t \mid \mathcal{F}_{t-1}\right] = 0, \qquad E\left[Y_t^2 \mid \mathcal{F}_{t-1}\right] = h_t, \qquad h_t = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 h_{t-1}. \tag{1}$$

The proposed closed-form estimators for this model are

$$\widehat{\sigma}^2 = T^{-1}\sum_t Y_t^2, \qquad \widehat{\alpha} = \frac{\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right) Y_{t-1}}{\sum_t Y_t^3}, \tag{2}$$

$$\widehat{\beta} = \frac{\left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{t-1}\right)' \Lambda_T \left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\left(\widehat{Z}_{t-2} - \widehat{\alpha}\widehat{Z}_{t-1}\right)\right)}{\left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{t-1}\right)' \Lambda_T \left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{t-1}\right)} \tag{3}$$

and

$$\widehat{\phi} = \frac{\left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{2,t-1}\right)' \Lambda_T \left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{2,t-2}\right)}{\left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{2,t-1}\right)' \Lambda_T \left(\sum_t \left(Y_t^2 - \widehat{\sigma}^2\right)\widehat{Z}_{2,t-1}\right)} \tag{4}$$

$$\widehat{b} = \frac{\widehat{\phi}^2 + 1 - 2\widehat{\rho}(1)\widehat{\phi}}{\widehat{\phi} - \widehat{\rho}(1)}, \quad \widehat{\beta} = \frac{\widehat{b} - \sqrt{\widehat{b}^2 - 4}}{2}, \quad \widehat{\alpha} = \widehat{\phi} - \widehat{\beta}, \tag{5}$$

where $Z_{1,t-2} = \left[Y_{t-2}, \cdots, Y_{t-(k+1)}\right]'$ and $Z_{2,t-2} = \left[Y_{t-2}^2 - \sigma^2, \cdots, Y_{t-(k+1)}^2 - \sigma^2\right]'$ both for $k = 1, \ldots, K-1$, $\theta = (\sigma^2, \alpha, \beta)'$, and $\vartheta = (\sigma^2, \phi)'$.

Closed-Form Estimator 1 (CFE1) is (2) and (3) with $\widehat{Z}_{t-1} = Z_{1,t-2}$ in (3). CFE2 is (2) and (3) with $\widehat{Z}_{t-1} = \begin{pmatrix} Z_{1,t-2} \\ Z_{2,t-2} \end{pmatrix}$ in (3). The iterative GLS estimator GLS1(2) is

$$\widehat{\lambda}_{l+1}^{GLS} = \left(\sum_t h_t^{-2}\left(\widehat{\lambda}_l^{GLS}\right) X_{l,t-1}X_{l,t-1}'\right)^{-1}\left(\sum_t h_t^{-2}\left(\widehat{\lambda}_l^{GLS}\right) X_{l,t-1}Y_t^2\right)$$

where $h_t\left(\widehat{\lambda}_l^{GLS}\right) = \widehat{\omega}_l^{GLS} + \widehat{\alpha}_l^{GLS} Y_{t-1}^2 + \widehat{\beta}_l^{GLS} h_{t-1}\left(\widehat{\lambda}_l^{GLS}\right)$, and $X_{l,t-1} = \left[1,\ Y_{t-1}^2,\ h_{t-1}\left(\widehat{\lambda}_l^{GLS}\right)\right]'$. GLS1 takes its starting values from CFE1, GLS2 from CFE2.

Given $X_t \equiv Y_t^2$ and $\widetilde{X}_t \equiv X_t - \sigma_0^2$, a useful result is

$$\widetilde{X}_t = \sum_{i=0}^{\infty} \varphi_{i,0} W_{t-i}, \tag{6}$$

where $W_t = X_t - h_t$, $\varphi_{0,0} = 1$, and $\varphi_{i,0} = \alpha_0 \phi_0^{i-1}$ for $i = 1, 2, \ldots$. In addition, note that $\{C_m : m \geq 1\}$ and $\{c_r\}_{r=1}^m$ denote sequences of constants that may take different values in different places.

II. Monte Carlo

Consider the data generating process in (1) with the additional assumptions

$$Y_t = h_t^{1/2}\epsilon_t, \qquad \epsilon_t \sim i.i.d.\ D(0,1),$$

where ϵ_t is the negative of a standardized $\Gamma(2,1)$ random variable. For this strong GARCH process, the Monte Carlo study examines different values of θ_0 that are organized into four cases and summarized in Table 1. In all cases, the value of ω_0 is chosen conditional on α_0 and β_0 so that $\sigma_0^2 = 1$. Only results for $\widehat{\alpha}$ and $\widehat{\beta}$ are reported. Under Case 1, the eighth moment of Y_t exists (see Corollary 6 of Carrasco and Chen, 2002). Under Case 2, the sixth moment of Y_t exists but not the eighth. Under Cases 3 and 4, only up to the fourth and second moments of Y_t exist, respectively. Cases 2–4 represent, to varying degrees, violations of the moment existence criteria underlying Theorem 3 and Corollary 2 as well as Theorems 4–5. These cases investigate the finite sample properties of the estimators for progressively fatter-tailed distributions of Y_t.

The estimators being considered are CFE1–3 and GLS1–2. For CFE1–2, $\Lambda_T = \widehat{\widehat{\Omega}}\left(\widehat{\sigma}^2, \widehat{\alpha}, \widetilde{\beta}\right)^{-1}$ and $L = 1$, where $\widetilde{\beta}$ is the preliminary estimator of β_0 when $\Lambda_T = I$ (see Theorem 3 for the poulation analog of $\widehat{\widehat{\Omega}}\left(\widehat{\sigma}^2, \widehat{\alpha}, \widetilde{\beta}\right)$). For CFE3, $\Lambda_T = \widehat{\Omega}\left(\widehat{\sigma}^2, \widetilde{\phi}\right)^{-1}$ and, again, $L = 1$ (see Corollary 2 for the population analog of $\widehat{\Omega}\left(\widehat{\sigma}^2, \widetilde{\phi}\right)$). For both (3) and (4), $K = 10, 20, 30, 40$. For both GLS estimators, $l = 10$. Results for the traditional QMLE are also provided as a benchmark.

All simulations are conducted with either $5,000$ or $2,000$ observations across $10,000$ trials. Results from simulations with $2,000$ observations are only reported for Case 1. Results from the remaining cases with $2,000$ observations are qualitatively similar to those being reported and are available upon request. When generating observations for the trials, the first 200 are dropped to avoid initialization effects. Estimates of α_0 and β_0 that are outside of $(0, 1]$ are discarded. Summary statistics for the simulations are the mean bias, standard deviation (SD), and root mean square error (RMSE).[1] For both CFE1 and CFE2, $\widehat{\alpha}$ does not depend on K, so only results for the smallest value ($K = 10$) are reported. For CFE3, since both $\widehat{\alpha}$ and $\widehat{\beta}$ depend on $\widehat{\phi}$, which, in turn, depends on K, results at all values of K being considered are reported. Results for all values of K being considered are also reported for the GLS1(2) estimates, since these estimates depend on CFE-generated starting values.

Results for the three CFEs are reported in Tables S2–S4. At low levels of GARCH persistence, the CFEs are close competitors to the QMLE. For instance, at the lowest persistence level of Cases 1 and 2, while $\widehat{\alpha}$ from the CFEs tends to be less efficient than the QMLE estimate, $\widehat{\beta}$ from CFE2 and CFE3 tends to be more efficient. In addition, $\widehat{\beta}$ from CFE1 is more efficient than the QMLE estimate under the lowest persistence level of Case 1. Holding α_0 fixed, as β_0 increases, the performance of the CFEs (both in terms of bias and efficiency) degrades relative to the QMLE. This statement is especially true for CFE1. The increase in bias of $\widehat{\beta}$ as β_0 increases seems to warrant either jackknife treatment of $\widehat{\beta}$ or application of the CUE to the objective function for which $\widehat{\beta}$ is the solution, at least, at high levels of K (see, for instance, CFE1 results under Cases 1 and 2). As α_0 increases, however, the level of bias in the associated $\widehat{\beta}$ decreases and, at times, markedly so (compare CFE $\widehat{\beta}$ biases moving from Case 1 to Case 4). The conclusion seems to be, therefore, that higher values of α_0 correspond with lower levels of bias in $\widehat{\beta}$ regardless of K, a result that is somewhat surprising since higher values of α_0 also correspond with fatter tails of $\{Y_t\}$ and, hence, the existence of fewer higher-order moments.

[1] In simulation evidence supporting the result from Theorem 4–5, median bias, decile range (defined as the difference between the 90th and 10th percentiles), mean absolute error (MAE), and median absolute error (MDAE) are also reported.

On the whole, CFE2 and CFE3 are broadly comparable. For each case considered, the lowest persistence level tends to favor CFE2 over CFE3 in terms of efficiency, with the opposite holding true for the highest persistence level. At the highest persistence level, however, CFE3 tends to display heightened biases relative to CFE2. With one exception, CFE1 is less efficient than CFE2 and tends to be more biased, especially at relatively high persistence levels and for high values of K. That one exception occurs at the lowest persistence level of Case 4, where CFE1 is less biased and more efficient than CFE2 for $K \geq 20$.

The asymptotic theory of Section 3 suggests that all three CFEs should perform well in Case 1. In Case 2, only CFE1 is expected to perform well, while in Cases 3 and 4, none of the estimators should fair that well. Indeed, the CFEs have their best performance against the QMLE in Case 1, and, in the other cases, both bias and efficiency suffer relative to the QMLE. When making bias and efficiency comparisons between cases, however, the performance of the CFEs do not uniformly degrade with the order of the case. To the contrary, bias and/or efficiency may actually improve (see; e.g., the performance of $\widehat{\beta}$ across the different cases). Contrary to what the asymptotic theory predicts, therefore, CFE1–3 display descent finite-sample properties even when certain requisite moments are not well defined.[2]

Results for the two GLS estimators are reported in Tables S5–S7. At low levels of GARCH persistence, the GLS estimators either (nearly) match or exceed the performance of the QMLE in terms of bias and efficiency, thereby confirming the conjecture from section 3.3. As is the case for the CFEs, as persistence increases, the performance of the GLS estimators decreases. For the highest persistence levels in Cases 2 and 3, the GLS estimators offer little if any improvement over the CFEs.[3]

Finally, Table S8 reports results for the GLS estimator applied to the ARCH(1) model at varying persistence levels benchmarked against the QMLE.[4] This set of experiments, CFE1 supplys starting values to the GLS estimator and, again, $l = 10$.[5] Apparent from these results is that the performance of the GLS and QMLE estimators is, essentially, the same. Moreover, this (near) equivalence holds across a wide range of persistence levels and, hence, progressively fatter-tailed processes. These results offer validation to Theorems 4–5.

[2] Kristensen and Linton (2006) report similar findings for a special case of CFE3.

[3] For the highest persistence level of Case 4, the number of instances where the GLS estimators failed to produce estimates for which $\widehat{\alpha} + \widehat{\beta} < 1$ was so high that results aren't reported.

[4] Specifically, $\alpha_0 = 0.10, 0.20, 0.40, 0.80$, which corresponds to Y_t having up to a finite eighth, sixth, fourth, and second moment, respectively.

[5] Because only $\widehat{\alpha}_1^{GLS}$ is required, lag length as denoted by K no longer matters.

III. Proofs

PROOF OF THEOREM 1: From (1), (6), $E\left[W_t \mid \mathcal{F}_{t-1}\right] = 0$, and the law of iterated expectations,

$$\begin{aligned} E\left[\widetilde{X}_t Y_{t-1}\right] &= E\left[\left(\phi_0 \widetilde{X}_{t-1} + W_t - \beta_0 W_{t-1}\right) Y_{t-1}\right] && (7) \\ &= \phi_0 E\left[Y_{t-1}^3\right] - \beta_0 E\left[W_{t-1} Y_{t-1}\right] && (8) \\ &= \alpha_0 E\left[Y_{t-1}^3\right], \end{aligned}$$

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

and

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

Given A2, these results imply that

$$E\left[\widetilde{X}_t Y_{t-k}\right] = \alpha_0 \phi_0^{k-1} E\left[Y_t^3\right]. \tag{9}$$

In (9), let $k = k+1$. Substitution of $E\left[\widetilde{X}_t Y_{t-k}\right]$ into the result for $E\left[\widetilde{X}_t Y_{t-(k+1)}\right]$ produces

$$E\left[\widetilde{X}_t Y_{t-(k+1)}\right] = \phi_0 E\left[\widetilde{X}_t Y_{t-k}\right], \; k \geq 1. \blacksquare$$

LEMMA 1. *Let $\widetilde{h}_t \equiv h_t - \sigma_0^2$. Given the model of (1), let Assumptions A1 and A3 hold. Then*

$$E\left[\widetilde{h}_t^2\right] = \left(\frac{\alpha_0^2}{1-\phi_0^2}\right) \kappa_0. \tag{10}$$

PROOF OF LEMMA 1: Since $\omega_0 = \sigma_0^2 (1 - \phi_0)$, $\widetilde{X}_t = \widetilde{h}_t + W_t$. Since $\{W_t\}$ is a MDS, $E\left[\widetilde{X}_t^2\right] = E\left[\widetilde{h}_t^2\right] + \kappa_0$.

$$\begin{aligned} E\left[\widetilde{h}_t^2\right] &= \phi_0^2 E\left[\widetilde{h}_{t-1}^2\right] + \alpha_0^2 \kappa_0 \\ &= \left(1 + \phi_0^2 + \cdots + \phi_0^{2(\tau-1)}\right) \alpha_0^2 \kappa_0 + \phi_0^{2\tau} E\left[\widetilde{h}_{t-\tau}^2\right] \end{aligned} \tag{11}$$

for $\tau \geq 1$. It is well known that$\phi_0^{2\tau} \to 0$ as $\tau \to \infty$ if and only if $\phi_0 < 1$.■

PROOF OF THEOREM 2: Given Lemma 1, Y_t^2 is covariance stationary so that $T^{-1} \sum_t Y_t \xrightarrow{p} 0$ and $\widehat{\sigma}^2 \xrightarrow{p} \sigma_0^2$ by the LLN. Using (6),

$$\begin{aligned} E\left[\widetilde{U}_{t,k} \mid F_{t-m}\right] &= \sum_{i=0}^{\infty} \varphi_{i,0} \left(E\left[W_{t-i} Y_{t-k} \mid F_{t-m}\right] - E\left[W_{t-i} Y_{t-k}\right]\right) \\ &= \sum_{i=m}^{\infty} \varphi_{i,0} \left(W_{t-i} Y_{t-k} - E\left[W_{t-i} Y_{t-k}\right]\right), \end{aligned}$$

where the second equality above follows if $m = k$, since $\{W_t\}$ is a MDS. Given A5(i) and the fact that $\sum_{i=0}^{\infty} \left|\varphi_{i,0}\right| < \infty$ by A1, $E\left|E\left[\widetilde{U}_{t,k} \mid F_{t-m}\right]\right| \leq C\xi_m \to 0$ as $m \to \infty$, where $\xi_m = \sum_{i=m}^{\infty} \varphi_{i,0}$. As a consequence, $\left\{\widetilde{U}_{t,k}\right\}$ is a L^1 mixingale (see Andrews, 1988 for a definition). Given A5(iii) or A3, it then follows that $T^{-1} \sum_t \widetilde{U}_{t,k} \xrightarrow{p} 0$ by Theorem 1 of Andrews. Next, $E\left[\widetilde{U}_{t,0} \mid F_{t-m}\right] = E\left[W_t Y_t \mid F_{t-m}\right] - \gamma_0$ given (6), $\{Y_t\}$ being a MDS, and the law of iterated expectations, so that $E\left|E\left[\widetilde{U}_{t,0} \mid F_{t-m}\right]\right| \leq C_m$ by A5(ii), with $T^{-1} \sum_t \widetilde{U}_{t,0} \xrightarrow{p} 0$ then following from A5(iii) or A3 and the same Theorem. Using (6) again,

$$E\left[\widetilde{V}_{t,k} \mid F_{t-m}\right] = \sum_{i=m}^{\infty} \varphi_{i,0}^2 W_{t-i} W_{t-(k+i)} + \sum_{i=m \neq j} \varphi_{i,0} \varphi_{j,0} \left(W_{t-i} W_{t-(k+j)} - E\left[W_{t-i} W_{t-(k+j)}\right]\right),$$

if $m = k$. Given A3, it then follows that $E\left|E\left[\widetilde{V}_{t,k} \mid F_{t-m}\right]\right| \leq C\overline{\xi}_m$, where $\overline{\xi}_m = \sum_{i=m}^{\infty} \varphi_{i,0}^2 + \sum_{i=m \neq j} \varphi_{i,0} \varphi_{j,0}$. Since $\lim_{m \to \infty} \overline{\xi}_m = 0$ by A1, $\left\{\widetilde{V}_{t,k}\right\}$ is also an L^1 mixingale. Given A6, $T^{-1} \sum_t \widetilde{V}_{t,k} \xrightarrow{p} 0$. Following from these weak convergence results are (i) $p\lim \widehat{\alpha} = E\left[\widetilde{X}_t Y_{t-1}\right] / E\left[Y_t^3\right] = \alpha_0$ by Theorem 1, (ii) $\widehat{g}_{2,t}^{(k)}\left(\widehat{\sigma}^2, \widehat{\alpha}, \beta\right) \xrightarrow{p} \alpha_0 \left(\phi_0 - (\alpha_0 + \beta)\right) \phi_0^{k-1} \gamma_0$, and (iii) $\widehat{g}_{3,t}^{(k)}\left(\widehat{\sigma}^2, \widehat{\alpha}, \beta\right) \xrightarrow{p} \left(\phi_0 - (\alpha_0 + \beta)\right) \phi_0^{k-1} \left(\alpha_0 \kappa_0 + \phi_0 \eta_0\right)$, where $\widehat{g}_{2,t}^{(k)}\left(\widehat{\sigma}^2, \widehat{\alpha}, \beta\right)$ and

$\widehat{g}_{3,t}^{(k)}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)$ are the kth elements of $g_{2,t}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)$ and $g_{3,t}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)$, respectively, and $\eta_0 = E\left[\widetilde{h}_t^2\right]$. Let $\overline{Q}\left(\sigma_0^2,\alpha_0,\beta\right) = \overline{g}\left(\sigma_0^2,\alpha_0,\beta\right)' \Lambda_0 \overline{g}\left(\sigma_0^2,\alpha_0,\beta\right)$, and $\widehat{Q}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right) = \widehat{g}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)' \Lambda_T \widehat{g}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)$. Given (i)–(iii) and continuity of multiplication, $\widehat{Q}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right) \xrightarrow{p} \overline{Q}\left(\sigma_0^2,\alpha_0,\beta\right)$. For $N=2$, (i) and (ii) establish that the only $\beta \in B$ satisfying $\overline{g}\left(\sigma_0^2,\alpha_0,\beta\right) = 0$ is $\beta = \beta_0$, since $\gamma_0 \neq 0$ and ϕ_0 is strictly positive. For $N=3$, (i)–(iii) establish the same result with parallel reasoning given that $\alpha_0\kappa_0 + \phi_0\eta_0$ is also strictly positive. $Q\left(\sigma_0^2,\alpha_0,\beta\right)$ is then uniquely minimized at $\beta=\beta_0$.■

PROOF OF COROLLARY 1: From the proof of Theorem 2, $\widehat{\phi} \xrightarrow{p} \phi_0$ by (iii) with ϕ substituted for $(\alpha_0+\beta)$ in the probability limit. Given (6),

$$\widetilde{V}_{t,0} = \sum_{i=0}^{\infty}\varphi_{i,0}^2\left(W_{t-i}^2-\kappa_0\right) + \sum_{i\neq j}\varphi_{i,0}\varphi_{j,0}W_{t-i}W_{t-j},$$

so that

$$E\left[\widetilde{V}_{t,0} \mid F_{t-m}\right] = \sum_{i=0}^{m-1}\varphi_{i,0}^2\left(E\left[W_{t-i}^2 \mid F_{t-m}\right]-\kappa_0\right)+\sum_{i=m}^{\infty}\varphi_{i,0}^2\left(W_{t-i}^2-\kappa_0\right)+\sum_{i=m\neq j}^{\infty}\varphi_{i,0}\varphi_{j,0}W_{t-i}W_{t-j}.$$

It then follows that

$$E\left|E\left[\widetilde{V}_{t,0} \mid F_{t-m}\right]\right| \leq \sum_{i=0}^{m-1}\varphi_{i,0}^2 C_m + 2\kappa_0\sum_{i=m}^{\infty}\varphi_{i,0}^2 + \kappa_0\sum_{i\neq j=m}^{\infty}\varphi_{i,0}\varphi_{j,0} = D_m,$$

from A3 and A7(i). Given A1 and A7(i), $\lim_{m\to\infty} D_m = 0$, thus establishing $\left\{\widetilde{V}_{t,0}\right\}$ as a L^1 mixingale. Then $T^{-1}\sum_t \widetilde{V}_{t,0} \xrightarrow{p} 0$ by A7(ii) and Theorem 1 of Andrews, which grants that $\widehat{\rho}(1) \xrightarrow{p} \rho(1)$. The remaining weak convergence results follow from the properties of probability limits.■

LEMMA 2.

$$T^{-1/2}\sum_{r=1}^{m}c_r\sum_{t=1}^{T}W_{t-r}Y_t \xrightarrow{d} N\left(\ 0,\ \ V\ \right),$$

where

$$\begin{aligned} V \;=\;& \sigma_0^2\kappa_0\sum_{r=1}^{m}c_r^2 + \sum_{r=1}^{m}c_r^2\varphi_{r,0}E\left[W_{t-r}^3\right] + \sum_{r=1}^{m}c_r^2\sum_{i=m+1}^{\infty}\varphi_{i,0}E\left[W_{t-i}W_{t-r}^2\right] \\ &+\sum_{r\neq s}^{m}c_r c_s\sum_{i=\min(r,s)}^{m}\varphi_{i,0}E\left[W_{t-i}W_{t-r}W_{t-s}\right]. \end{aligned} \qquad (12)$$

PROOF OF LEMMA 2: Let $X_t = \sum_{r=1}^{m} c_r W_{t-r} Y_t$; $V_T^2 = \sum_{t=1}^{T} E\left[X_t^2 \mid \digamma_{t-1}\right] = \sum_{t=1}^{T}\sum_{r=1}^{m}\sum_{s=1}^{m} c_r c_s W_{t-r} W_{t-s} h_t$, and $s_T^2 = E\left[V_T^2\right] = \sum_{t=1}^{T}\sum_{r=1}^{m}\sum_{s=1}^{m} c_r c_s E\left[W_{t-r} W_{t-s} Y_t^2\right]$.[6] Since given (6), $s_T^2 = V$ in (12), V is finite by A9(ii). From Theorem 2 of Brown (1971), if (i) $s_T^2 / V_T^2 \xrightarrow{p} 1$, and (ii) $s_T^{-2} \sum_{t=1}^{T} E\left[X_t^2 I\left(|X_t| \geq \epsilon s_T\right)\right] \xrightarrow{p} 0$ for any $\epsilon > 0$, then $T^{-1/2} \sum_{t=1}^{T} X_t \xrightarrow{d} N\left(\ 0,\ \ V\ \right)$. (i) follows from A9. (ii) follows from the same reasoning given in Hannan and Heyde (1972) starting with their equation (12). Specifically, that equation establishes

$$P\left(|X_t| > u\right) \leq \sum_{r=1}^{m} P\left(W_{t-r}^2 > u/mc^*\right) + mP\left(Y_t^2 > u/mc^*\right),$$

where $c^* = \max_{1 \leq r \leq m} |c_r|$. In addition, given (6),

$$\begin{aligned} P\left(Y_t^2 > u\right) &= P\left(\sum_{i=0}^{\infty} \varphi_{i,0} W_{t-i} > u^*\right), \quad u^* = u - \sigma_0^2 \\ &\leq P\left(\sum_{i=0}^{\infty} \varphi_{i,0} \left|W_{t-i}\right| > u^*\right) \\ &\leq \sum_{i=0}^{\infty} P\left(\varphi_{i,0} \left|W_{t-i}\right| > u^*\right) \\ &\leq \sum_{i=0}^{\infty} \varphi_{i,0} P\left(\left|W_{t-i}\right| > u^*\right), \end{aligned}$$

so that given A10,

$$P\left(|X_t| > u\right) \leq cm\left\{P\left(X^2 > u/mc^*\right) + dP\left(|X| > (u/mc^*)^*\right)\right\}, \tag{13}$$

where $d = \sum_{i=0}^{\infty} \varphi_{i,0}$. Finally, given (13) and using integration by parts (as noted by Hannan and Hyde),

$$\begin{aligned} E\left[X_t^2 I\left(|X_t| \geq \epsilon s_T\right)\right] &\leq \int_{\epsilon s_T}^{\infty} xP\left(|X_t| > x\right) dx \qquad (14) \\ &\leq cm\left(\int_{\epsilon s_T}^{\infty} xP\left(\left|X^2\right| > x/mc^*\right) dx + d\int_{\epsilon s_T}^{\infty} xP\left(|X| > (x/mc^*)^*\right) dx\right) \to 0 \end{aligned}$$

[6] Note that $\left\{S_T = \sum_{t=1}^{T} X_t, \digamma_T, T \geq 1\right\}$ is a martingale.

as $T \to \infty$ by Markov's Inequality.■

LEMMA 3. *Let* $E\left[\left(z_T^{(k)^*}\right)^2\right] = W = \lim_{m\to\infty} V$ *in (12), where* $z_T^{(k)^*}$ *is defined in (16). Then*

$$z_t^{(k)^*} \xrightarrow{d} N\left(\ 0,\quad W\ \right).$$

PROOF OF LEMMA 3: Let $X_{t-k} = \sum_{i=l+1}^{\infty} \varphi_{i,0} W_{t-i} Y_{t-k}$, and $R_T^{(k)^*} \equiv \left(z_T^{(k)^*} - z_{T,l}^{(k)}\right) = T^{-1/2}\left(\sum_{t=k+1}^{T} X_{t-k} - \sum_{t=1}^{T} X_t\right)$, where $z_{T,l}^{(k)}$ is defined in (17). Then $E\left[R_T^{(k)^*}\right] = 0$, and

$$\begin{aligned}
Var\left[R_T^{(k)^*}\right] &= Var\left[E\left[R_T^{(k)^*} \mid \digamma_{t-1}\right]\right] + E\left[Var\left[R_T^{(k)^*} \mid \digamma_{t-1}\right]\right] \\
&= T^{-1}\left(\sum_{t=k+1}^{T} Var\left(X_{t-k}\right) + \varphi_{k,0}^2 \sum_{t=1}^{T} E\left[X_t^2\right]\right) \\
&= \left(1+\varphi_{k,0}^2\right)\left(\sigma_0^2\kappa_0 \sum_{i=l+1}^{\infty}\varphi_{i,0}^2 + \sum_{i=l+1}^{\infty}\sum_{j=l+1}^{\infty}\sum_{n=l+1}^{\infty}\varphi_{i,0}\varphi_{j,0}\varphi_{n,0} E\left[W_{t-i}W_{t-j}W_{t-n}\right]\right) + O_p\left(T^{-1}\right) \\
&= M_l + O_p\left(T^{-1}\right).
\end{aligned}$$

Since $\lim_{l\to\infty} M_l = 0$ by A1 and A9(ii), the result then follows from Lemma 2.■

LEMMA 4.

$$T^{-1/2}\sum_{r=1}^{m} c_r \sum_{t=1}^{T} W_t Y_{t-r} \xrightarrow{d} N\left(\ 0,\quad \overline{V}\ \right),$$

where

$$\overline{V} = \sigma_0^2\kappa_0 \sum_{r=1}^{m} c_r^2 + \sum_{r=1}^{m} c_r^2 \sum_{i=0}^{\infty} \varphi_{i,0} E\left[W_{t-r-i} W_t^2\right] + \sum_{r\neq s}^{m} c_r c_s E\left[Y_{t-r}Y_{t-s}W_t^2\right].$$

PROOF OF LEMMA 4: Let $X_t = \sum_{r=1}^{m} c_r W_t Y_{t-r}$, $\overline{V}_T^2 = \sum_{t=1}^{T} E\left[X_t^2 \mid \digamma_{t-1}\right]$, and $\overline{s}_T^2 = E\left[\overline{V}_T^2\right] = \overline{V}$, where $\overline{s}_T^2 = \overline{V}$ follows from (6). As in Lemma 2, convergence to a normal law follows from Theorem 2 of Brown (1971). The first criterion for this theorem is satisfied by A9. The second criterion follows from (13) and (14).■

LEMMA 5. *Let* $c(k) = T^{-1}\sum_{t=k+1}^{T}\widetilde{X}_t Y_{t-k}$, $\gamma(k) = E\left[\widetilde{X}_t Y_{t-k}\right]$, *and* $\rho(k) = \gamma(k)/\gamma(0)$. *Consider* $\sqrt{T}\left(\frac{c(k)}{c(0)} - \rho(k)\right)$. *Then* $c(0)^{-1}\sqrt{T}\left(c(k) - \rho(k)c(0)\right)$ *converges in distribution to a normal law.*

PROOF OF LEMMA 5: From the proof of Theorem 1, $\gamma(k) = \varphi_{k,0}E[W_t Y_t]$, and $\gamma(0) = E[W_t Y_t]$ so that $\rho(k) = \varphi_{k,0}$. Given A5, $c(0) \xrightarrow{p} \gamma(0) = \gamma_0$. Let

$$
\begin{aligned}
z_T^{(k)} &= \sqrt{T}\left(c(k) - \rho(k)c(0)\right) \\
&= z_{T,k-1}^{(k)} + z_{T,k}^{(k)} + z_T^{(k)*},
\end{aligned}
$$

where

$$
z_{T,k-1}^{(k)} = T^{-1/2}\left(\sum_{i=0}^{k-1}\varphi_{i,0}\sum_{t=k+1}^{T}W_{t-i}Y_{t-k} - \sum_{i=1}^{k}\varphi_{k,0}\varphi_{i,0}\sum_{t=1}^{T}W_{t-i}Y_t\right), \tag{15}
$$

$$
\begin{aligned}
z_{T,k}^{(k)} &= T^{-1/2}\varphi_{k,0}\left(\sum_{t=k+1}^{T}W_{t-k}Y_{t-k} - \sum_{t=1}^{T}W_t Y_t\right) \\
&= O_p\left(T^{-1/2}\right),
\end{aligned}
$$

and

$$
z_T^{(k)*} = T^{-1/2}\left(\sum_{i=k+1}^{\infty}\varphi_{i,0}\sum_{t=k+1}^{T}W_{t-i}Y_{t-k} - \sum_{i=k+1}^{\infty}\varphi_{k,0}\varphi_{i,0}\sum_{t=1}^{T}W_{t-i}Y_t\right). \tag{16}
$$

Further let

$$
z_{T,l}^{(k)} = T^{-1/2}\left(\sum_{i=k+1}^{l}\varphi_{i,0}\sum_{t=k+1}^{T}W_{t-i}Y_{t-k} - \sum_{i=k+1}^{l}\varphi_{k,0}\varphi_{i,0}\sum_{t=1}^{T}W_{t-i}Y_t\right) \tag{17}
$$

for $k < l < \infty$. By adding and subtracting a finite number of terms from (17) (that number not depending on T), it is possible to show that

$$
z_{T,l}^{(k)} = T^{-1/2}\sum_{r=1}^{m}\delta_{r,l}^{(k)}\sum_{t=1}^{T}W_{t-r}Y_t + O_p\left(T^{-1/2}\right),
$$

where m is a function of l, and $\delta_{r,l}^{(k)} = \sum_{j=0}^{l} f_j\left(\varphi_{l,0}; r, k\right)$. (17) then converges in distribution to a certain normal law by Lemma 2 and $z_T^{(k)*} = \lim_{l\to\infty} z_{T,l}^{(k)}$ by Lemma 3. The second half of (15) converges in distribution to a certain normal law by Lemma 2, the first half by Lemma 4.■

LEMMA 6.

$$T^{-1/2}\sum_t g_{1,t}(\theta_0) \xrightarrow{d} N\left(0, \quad \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{1,t-s}(\theta_0)\, g_{1,t}(\theta_0)\right] \right).$$

PROOF OF LEMMA 6:

$$\begin{aligned} T^{-1/2}\sum_t g_{1,t}(\theta_0) &= T^{-1/2}\left(\sum_t \widetilde{X}_t Y_{t-1} - \alpha_0 \sum_t Y_t^3\right) \\ &= T^{-1/2}\sum_t \left(\widetilde{X}_t Y_{t-1} - \alpha_0\gamma_0\right) - \alpha_0 T^{-1/2}\sum_t \left(Y_t^3 - \gamma_0\right) \end{aligned}$$

where the first term on the right-hand-side of the second equality converges in distribution to a normal law by Lemma 5 and Theorem 1. Next,

$$\begin{aligned} T^{-1/2}\sum_t \left(Y_t^3 - \gamma_0\right) &= T^{-1/2}\sum_t \left(Y_t^2 Y_t - \gamma_0\right) \\ &= \sigma_0^2 T^{-1/2}\sum_t Y_t + T^{-1/2}\sum_t \widetilde{S}_t + T^{-1/2}\sum_{i=1}^{\infty}\varphi_{i,0}\sum_{t=1}^{T} W_{t-i}Y_t, \end{aligned}$$

where the second equality immediately above follows from (6). The third term of the second equality immediately above converges in distribution to a normal law by Lemmas 2 and 3. The first term converges in distribution to $N\left(0, \quad \sigma_0^2 \right)$ by Corollary 5.25 of White (1984). This same corollary also establishes

$$T^{-1/2}\sum_{t=1}^{T}\widetilde{S}_t \xrightarrow{d} N\left(0, \quad E\left[\widetilde{S}_t^2\right] \right),$$

where

$$E\left[\widetilde{S}_t^2\right] = \sigma_0^2\kappa_0 - \gamma_0^2 + E\left[W_t^3\right] + \sum_{i=1}^{\infty}\varphi_{i,0} E\left[W_{t-i}W_t^2\right],$$

since (i) $\left\{\widetilde{S}_t\right\}$ is a MDS; (ii) $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[\widetilde{S}_t^2 \mid F_{t-1}\right] = E\left[\widetilde{S}_t^2\right]$ by A8(ii); (iii) $T^{-1}\sum_{t=1}^{T}\widetilde{S}_t^2 \xrightarrow{p} E\left[\widetilde{S}_t^2\right]$ by A8(iii), and (iv) $E\left|\widetilde{S}_t\right|^r < \infty$ for some $r > 2$ by A9(ii).[7]■

[7] (ii)–(iv) all rely on (6).

PROOF OF THEOREM 3: From (2),

$$\begin{aligned}\sqrt{T}\left(\widehat{\alpha}-\alpha_0\right) &= \left(T^{-1}\sum_t Y_t^3\right)^{-1} T^{-1/2}\left(\sum_t \widehat{\widetilde{X}}_t Y_{t-1}-\alpha_0\sum_t Y_t^3\right) \qquad (18)\\ &= \left(T^{-1}\sum_t Y_t^3\right)^{-1} T^{-1/2}\sum_t g_{1,t}\left(\theta_0\right)+o_p\left(1\right).\end{aligned}$$

Following from Lemma 6, Theorem 2, and the Slutzky Theorem is then

$$\sqrt{T}\left(\widehat{\alpha}-\alpha_0\right)\xrightarrow{d} N\left(\ 0,\ \ \gamma_0^{-2}\sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{1,t-s}\left(\theta_0\right)g_{1,t}\left(\theta_0\right)\right]\ \right).$$

Next, let $\Lambda_T=\Lambda_T\left(\widehat{\sigma}^2,\widehat{\alpha},\widetilde{\beta}\right)$, where $\widetilde{\beta}$ is a preliminary (and consistent) estimate of β_0, and $R\left(\theta\right)=\left\{S_{\theta_1}\left(\theta\right)+S_{\theta_{21}}\left(\theta\right)\left(\widehat{\alpha}-\alpha_0\right)\right\}\left(\widehat{\sigma}^2-\sigma_0^2\right)$. The first order condition of

$$\widehat{\beta}=\arg\min_{\beta\in B}\ \widehat{g}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)'\Lambda_T\widehat{g}\left(\widehat{\sigma}^2,\widehat{\alpha},\beta\right)$$

is $\widehat{S}_{\theta_3}\left(\widehat{\theta}\right)'M_T\widehat{g}\left(\widehat{\theta}\right)=0$. Expanding $\widehat{g}\left(\widehat{\theta}\right)$ first around β_0, then around α_0, and finally around σ_0^2, noting that $\overline{\beta}$ is between β_0 and $\widehat{\beta}$ (with the same definition holding for $\overline{\alpha}$ and $\overline{\sigma}^2$, respectively), produces

$$\begin{aligned}\sqrt{T}\left(\widehat{\beta}-\beta_0\right) &= -H\left(\widehat{\sigma}^2,\widehat{\alpha},\overline{\beta}\right)^{-1}\widehat{S}_{\theta_3}\left(\widehat{\theta}\right)'\Lambda_T\sqrt{T}\left\{\widehat{g}\left(\theta_0\right)+\widehat{S}_{\theta_2}\left(\sigma_0^2,\overline{\alpha},\beta_0\right)\left(\widehat{\alpha}-\alpha_0\right)+\widehat{R}\left(\overline{\sigma}^2,\overline{\alpha},\beta_0\right)\right\}\\ &= -H\left(\theta_0\right)^{-1}S_{\theta_3}\left(\theta_0\right)'\Lambda_0\sqrt{T}\left\{\widehat{g}\left(\theta_0\right)+S_{\theta_2}\left(\theta_0\right)\left(\widehat{\alpha}-\alpha_0\right)\right\}\\ &= -H\left(\theta_0\right)^{-1}S_{\theta_3}\left(\theta_0\right)'\Lambda_0 T^{-1/2}\sum_t\widetilde{g}_t\left(\theta_0\right)+o_p\left(1\right).\end{aligned}$$

Though tedious, it is straightforward to verify that (i) $\widehat{S}_{\theta_1}\left(\widehat{\theta}\right)\xrightarrow{p} S_{\theta_1}\left(\theta_0\right)=0$, (ii) $\widehat{S}_{\theta_{21}}\left(\widehat{\theta}\right)\xrightarrow{p} S_{\theta_{21}}\left(\theta_0\right)=0$, and (iii) $\widehat{S}_{\theta_3}\left(\widehat{\theta}\right)\xrightarrow{p} S_{\theta_3}\left(\theta_0\right)$ using the results of Theorem 2. Coupled with A12, (i)–(iii) explain the second equality immediately above. The third equality immediately above follows from substituting the final result for $\sqrt{T}\left(\widehat{\alpha}-\alpha_0\right)$ in (18) into the second equality immediately above. What remains to establish is

$$T^{-1/2}\sum_t\widetilde{g}_t\left(\theta_0\right)\xrightarrow{d} N\left(\ 0,\ \ \widetilde{\Omega}\left(\theta_0\right)\ \right). \qquad (19)$$

If $N=2$, this result follows from Lemmas 5 and 6. If $N=3$, Lemmas 5 and 6 establish AN of $T^{-1/2}\sum_t \widetilde{g}_{2,t}(\theta_0)$. Theorem 2 of Hannan and Heyde (1972) establish AN of $T^{-1/2}\sum_t \widetilde{g}_{3,t}(\theta_0)$, where this result is reliant on A10 and A11. These two AN results together with Theorem 1 of Shao and Zhou (2010) produce (19). Followig from the Slutsky Theorem is then

$$\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\xrightarrow{d}N\left(\ 0,\ \ \frac{S_{\theta_3}(\theta_0)'\Lambda_0\widetilde{\Omega}(\theta_0)\Lambda_0 S_{\theta_3}(\theta_0)}{H(\theta_0)^2}\ \right).\blacksquare$$

PROOF OF COROLLARY 2: The first-order condition of

$$\widehat{\phi}=\underset{\phi\in\Phi}{\arg\min}\ \widehat{g}\left(\widehat{\sigma}^2,\phi\right)'\Lambda_T\widehat{g}\left(\widehat{\sigma}^2,\phi\right)$$

is $S_{\vartheta_2}\left(\widehat{\vartheta}\right)'\Lambda_T\widehat{g}\left(\widehat{\vartheta}\right)=0$. Expanding $\widehat{g}\left(\widehat{\vartheta}\right)$ first around ϕ_0 and then around σ_0^2, noting that $\overline{\phi}$ is between ϕ_0 and $\widehat{\phi}$ (with the same definition holding for $\overline{\sigma}^2$), produces

$$\begin{aligned}\sqrt{T}\left(\widehat{\phi}-\phi_0\right) &= \left[\widehat{S}_{\vartheta_2}\left(\widehat{\vartheta}\right)'\Lambda_T\widehat{S}_{\vartheta_2}\left(\widehat{\sigma}^2,\overline{\phi}\right)\right]^{-1}\widehat{S}_{\vartheta_2}\left(\widehat{\vartheta}\right)'\Lambda_T\sqrt{T}\left[\widehat{g}\left(\vartheta_0\right)+\widehat{S}_{\vartheta_1}\left(\overline{\sigma}^2,\phi_0\right)'\left(\widehat{\sigma}^2-\sigma_0^2\right)\right]\\ &= -H\left(\vartheta_0\right)^{-1}S_{\vartheta_2}\left(\vartheta_0\right)'\Lambda_0\sqrt{T}\widehat{g}\left(\vartheta_0\right), \qquad (20)\end{aligned}$$

where (20) follows from $\widehat{S}_{\vartheta_2}\left(\widehat{\vartheta}\right)\xrightarrow{p}S_{\vartheta_2}\left(\vartheta_0\right)$ and $\widehat{S}_{\vartheta_1}\left(\widehat{\vartheta}\right)\xrightarrow{p}S_{\vartheta_1}\left(\vartheta_0\right)=0$ (see Theorem 2), and A12. From Theorem 3, $\sqrt{T}\widehat{g}\left(\vartheta_0\right)\xrightarrow{d}N\left(\ 0,\ \ \Omega\left(\vartheta_0\right)\ \right)$, with the result

$$\sqrt{T}\left(\widehat{\phi}-\phi_0\right)\xrightarrow{d}N\left(\ 0,\ \ v_{11}\ \right)$$

then following from the Slutsky Theorem. Next,

$$\begin{aligned}\widehat{\rho}\left(1\right)-\rho\left(1\right) &= \left(\sum_t\widehat{\widetilde{X}}_t^2\right)^{-1}\left(\sum_t\widehat{\widetilde{X}}_t\widehat{\widetilde{X}}_{t-1}-\rho\left(1\right)\sum_t\widehat{\widetilde{X}}_t^2\right)\\ &= \left(T^{-1}\sum_t\widetilde{X}_t^2\right)^{-1}T^{-1}\sum_t h_t\left(\rho\left(1\right)\right)+o_p\left(1\right),\end{aligned}$$

so that

$$\sqrt{T}\left(\widehat{\rho}\left(1\right)-\rho\left(1\right)\right)=\left(E\left[\widetilde{X}_t^2\right]\right)^{-1}T^{-1}\sum_t h_t\left(\rho\left(1\right)\right)+o_p\left(1\right). \qquad (21)$$

Given (20) and (21),

$$\sqrt{T}\left(\widehat{\mu}-\mu_0\right)\overset{d}{\to}N\left(\begin{array}{cc}0, & V_\mu\end{array}\right) \tag{22}$$

by Theorem 2 of Hannan and Heyde (1972) and the continuous mapping theorem. From (22) follows

$$\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\overset{d}{\to}N\left(\begin{array}{cc}0, & \frac{\partial f(\mu_0)}{\mu}V_\mu\frac{\partial f(\mu_0)'}{\mu}\end{array}\right),\quad \sqrt{T}\left(\widehat{\alpha}-\alpha_0\right)\overset{d}{\to}N\left(\begin{array}{cc}0, & \frac{\partial g(\mu_0)}{\mu}V_\mu\frac{\partial g(\mu_0)'}{\mu}\end{array}\right),$$

given applications of the delta method. Tedious, though straightforward, differentiation and simplification reveals that

$$\begin{aligned}\frac{\partial f(\mu_0)}{\mu} &= \left(1-\frac{b_0}{\sqrt{b_0^2-4}}\right)\left[c_1(\beta_0),\ \frac{1}{2}c_2(\beta_0)\right]\\ \frac{\partial g(\mu_0)}{\mu} &= -\left(1-\frac{b_0}{\sqrt{b_0^2-4}}\right)\left[c_1(\alpha_0),\ \frac{1}{2}c_2(\beta_0)\right]\end{aligned}$$

from which

$$\sqrt{T}\left(\widehat{\beta}-\beta_0\right)\overset{d}{\to}N\left(\begin{array}{cc}0, & V_\beta\end{array}\right),\quad \sqrt{T}\left(\widehat{\alpha}-\alpha_0\right)\overset{d}{\to}N\left(\begin{array}{cc}0, & V_\alpha\end{array}\right)$$

then follows with the accompanying definitions of V_β and V_α, respectively (see the main text).■

PROOF OF THEOREM 4: $\widehat{\lambda}_1^{GLS}\overset{p}{\to}\lambda_0$ and $\widehat{\lambda}_1^{NR}\overset{p}{\to}\lambda_0$ by Theorem 2. By the triangle inequality,

$$\left\|\widehat{\lambda}_{l+1}^{GLS}-\widehat{\lambda}^{QMLE}\right\|\le\left\|\widehat{\lambda}_{l+1}^{NR}-\widehat{\lambda}_{l+1}^{GLS}\right\|+\left\|\widehat{\lambda}_{l+1}^{NR}-\widehat{\lambda}^{QMLE}\right\|. \tag{23}$$

By Robinson (1988, Theorem 2),

$$\begin{aligned}\widehat{\lambda}_{l+1}^{NR}-\widehat{\lambda}^{QMLE} &= O_p\left(\left\|\widehat{\lambda}_1^{NR}-\widehat{\lambda}^{QMLE}\right\|^{2^l}\right)\\ &= O_p\left(T^{-2^{l-1}}\right),\end{aligned} \tag{24}$$

where the second equality follows from consistency of $\widehat{\lambda}_1^{NR}$ and $\widehat{\lambda}^{QMLE}$. Under regulatory conditions,

$$\sqrt{T}\left(\widehat{\lambda}^{QMLE}-\lambda_0\right)\xrightarrow{d}N\left(\ 0,\quad H^{-1}\Sigma H^{-1}\ \right),\tag{25}$$

from Lee and Hansen (1994, Theorem 2). Given (23), combining A13, (24) and (25) produces the result.■

PROOF OF THEOREM 5: Let I_n be the $n\times n$ identity matrix.

$$\begin{aligned}\widehat{\lambda}_{l+1}^{NR} &= \widehat{H}\left(\widehat{\lambda}_l^{NR}\right)^{-1}\left[\widehat{J}\left(\widehat{\lambda}_l^{NR}\right)+\widehat{M}\left(\widehat{\lambda}_l^{NR}\right)\right]\\ &= \widehat{H}\left(\widehat{\lambda}_l^{NR}\right)^{-1}\left\{\widehat{J}\left(\widehat{\lambda}_l^{GLS}\right)+\frac{\partial\widehat{J}\left(\overline{\lambda}\right)}{\partial\lambda}\left(\widehat{\lambda}_l^{NR}-\widehat{\lambda}_l^{GLS}\right)\right\}+\widehat{H}\left(\widehat{\lambda}_l^{NR}\right)^{-1}\widehat{M}\left(\widehat{\lambda}_l^{NR}\right)\end{aligned}$$

Since $\widehat{\lambda}_{l+1}^{GLS}=\widetilde{H}\left(\widehat{\lambda}_l^{GLS}\right)^{-1}\widehat{J}\left(\widehat{\lambda}_l^{GLS}\right)$,

$$\begin{aligned}\widehat{\lambda}_{l+1}^{NR}-\widehat{\lambda}_{l+1}^{GLS} &= \widehat{H}\left(\widehat{\lambda}_l^{NR}\right)^{-1}\frac{\partial\widehat{J}\left(\overline{\lambda}\right)}{\partial\lambda}\left(\widehat{\lambda}_l^{NR}-\widehat{\lambda}_l^{GLS}\right)+\left(\widehat{H}\left(\widehat{\lambda}_l^{NR}\right)^{-1}\widetilde{H}\left(\widehat{\lambda}_l^{GLS}\right)-I_2\right)\widehat{\lambda}_{l+1}^{GLS}\\ &\quad +\widehat{H}\left(\widehat{\lambda}_l^{NR}\right)^{-1}\widehat{M}\left(\widehat{\lambda}_l^{NR}\right)\end{aligned}$$

Then given A14,

$$\begin{aligned}\widehat{\lambda}_{l+1}^{NR}-\widehat{\lambda}_{l+1}^{GLS} &= H^{-1}\dot{J}\left(\widehat{\lambda}_l^{NR}-\widehat{\lambda}_l^{GLS}\right)+O_p\left(T^{-1}\right)\\ &= O_p\left(\left\|\widehat{\lambda}_l^{NR}-\widehat{\lambda}_l^{GLS}\right\|\right)\end{aligned}$$

where the second equality holds for sufficiently large T. The result then follows from recursive substitution and consistency of both $\widehat{\lambda}_1^{GLS}$ and $\widehat{\lambda}_1^{NR}$.■

TABLE S1

Case	α_0	β_0
		0.20
1	0.10	0.40
		0.60
		0.20
2	0.15	0.40
		0.60
		0.20
3	0.20	0.40
		0.60
	0.40	0.20
4	0.35	0.40
	0.35	0.60

Notes to Table S1: The different values of α_0 and β_0 used in the Monte Carlo experiments. Under Case 1, the eighth moment of Y_t exists. Under Case 2, the sixth moment exists but not the eighth. Under Cases 3 and 4, up to only the fourth and second moment exists, respectively.

TABLE S2

$T = 5,000$			(α_0, β_0)								
			(0.10, 0.20)			(0.10, 0.40)			(0.10, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	CFE1, CFE2	10	-0.003	0.036	0.036	-0.002	0.037	0.037	-0.002	0.039	0.039
	CFE3	10	-0.007	0.033	0.034	-0.004	0.035	0.035	-0.003	0.037	0.037
		20	-0.006	0.034	0.034	-0.004	0.035	0.035	-0.002	0.036	0.037
		30	-0.006	0.034	0.034	-0.004	0.036	0.036	-0.003	0.036	0.036
		40	-0.005	0.034	0.034	-0.004	0.035	0.036	-0.002	0.036	0.036
	QMLE		-0.001	0.027	0.027	0.001	0.027	0.027	0.001	0.024	0.024
$\widehat{\beta}$	CFE1	10	0.040	0.159	0.164	-0.068	0.169	0.182	-0.107	0.161	0.193
		20	0.031	0.151	0.154	-0.087	0.161	0.184	-0.142	0.158	0.213
		30	0.016	0.138	0.139	-0.108	0.151	0.186	-0.176	0.157	0.236
		40	0.000	0.130	0.130	-0.133	0.144	0.196	-0.209	0.152	0.259
	CFE2	10	-0.004	0.128	0.129	-0.088	0.140	0.165	-0.103	0.128	0.164
		20	0.005	0.127	0.127	-0.081	0.137	0.159	-0.109	0.122	0.164
		30	0.001	0.121	0.121	-0.087	0.133	0.159	-0.119	0.122	0.170
		40	-0.005	0.115	0.115	-0.099	0.128	0.161	-0.134	0.120	0.180
	CFE3	10	-0.008	0.131	0.131	-0.096	0.140	0.170	-0.110	0.129	0.169
		20	0.010	0.135	0.136	-0.082	0.143	0.165	-0.114	0.122	0.167
		30	0.010	0.133	0.133	-0.080	0.141	0.162	-0.115	0.123	0.168
		40	0.009	0.128	0.128	-0.087	0.137	0.163	-0.124	0.122	0.174
	QMLE		0.030	0.150	0.153	-0.011	0.148	0.148	-0.012	0.101	0.102
$T = 2,000$											
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	CFE1, CFE2	10	-0.003	0.050	0.050	-0.001	0.052	0.052	-0.001	0.053	0.053
	CFE3	10	-0.012	0.043	0.045	-0.009	0.045	0.045	-0.006	0.047	0.047
		20	-0.010	0.043	0.044	-0.007	0.045	0.046	-0.005	0.048	0.048
		30	-0.009	0.044	0.045	-0.007	0.046	0.046	-0.005	0.047	0.047
		40	-0.009	0.044	0.045	-0.006	0.046	0.046	-0.005	0.047	0.047
	QMLE		0.000	0.042	0.042	0.002	0.041	0.041	0.003	0.038	0.038
$\widehat{\beta}$	CFE1	10	0.075	0.189	0.203	-0.079	0.191	0.207	-0.187	0.199	0.273
		20	0.044	0.166	0.171	-0.115	0.172	0.207	-0.239	0.183	0.301
		30	0.015	0.145	0.146	-0.149	0.155	0.215	-0.283	0.169	0.330
		40	-0.008	0.132	0.132	-0.176	0.141	0.225	-0.318	0.157	0.355
	CFE2	10	0.022	0.154	0.156	-0.111	0.162	0.197	-0.184	0.168	0.250
		20	0.012	0.140	0.140	-0.123	0.154	0.197	-0.207	0.161	0.262
		30	-0.005	0.127	0.127	-0.144	0.142	0.202	-0.232	0.156	0.280
		40	-0.017	0.117	0.118	-0.161	0.133	0.209	-0.255	0.150	0.296
	CFE3	10	0.028	0.167	0.169	-0.114	0.166	0.201	-0.189	0.169	0.254
		20	0.032	0.159	0.162	-0.109	0.165	0.197	-0.200	0.165	0.259
		30	0.019	0.147	0.148	-0.122	0.157	0.198	-0.212	0.161	0.267
		40	0.008	0.136	0.136	-0.134	0.147	0.199	-0.230	0.157	0.278
	QMLE		0.082	0.202	0.218	-0.005	0.201	0.201	-0.033	0.165	0.168

Notes to Table S2: Simulations are conducted using 5,000 and 2,000 observations across 10,000 trials. Case 1 is considered. CFE1–3 are the three closed-form estimators being considered. QMLE is the quasi-maximum likelihood estimator. For CFE1–2, K does not impact $\widehat{\alpha}$, so results for only the first value are shown. For CFE3, both $\widehat{\alpha}$ and $\widehat{\beta}$ depend on K through their shared dependence on $\widehat{\phi}$. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. RMSE is the root mean square error measured with respect to the true parameter value.

TABLE S3

CASE 2			(α_0, β_0)								
			(0.15, 0.20)			(0.15, 0.40)			(0.15, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	CFE1, CFE2	10	-0.007	0.044	0.045	-0.005	0.047	0.047	-0.007	0.052	0.053
	CFE3	10	-0.014	0.041	0.043	-0.009	0.045	0.046	-0.008	0.047	0.048
		20	-0.013	0.041	0.043	-0.010	0.045	0.046	-0.008	0.047	0.048
		30	-0.013	0.042	0.044	-0.010	0.045	0.046	-0.008	0.047	0.048
		40	-0.012	0.042	0.044	-0.010	0.045	0.046	-0.009	0.046	0.047
	QMLE		-0.001	0.032	0.032	0.000	0.031	0.031	0.001	0.027	0.027
$\widehat{\beta}$	CFE1	10	0.012	0.134	0.134	-0.056	0.144	0.154	-0.063	0.128	0.143
		20	0.015	0.131	0.132	-0.061	0.141	0.153	-0.081	0.122	0.147
		30	0.011	0.125	0.126	-0.071	0.138	0.155	-0.096	0.122	0.156
		40	0.001	0.121	0.121	-0.087	0.132	0.158	-0.117	0.121	0.168
	CFE2	10	-0.015	0.113	0.114	-0.065	0.124	0.140	-0.065	0.117	0.133
		20	0.004	0.116	0.116	-0.047	0.122	0.130	-0.064	0.110	0.127
		30	0.012	0.114	0.115	-0.040	0.120	0.127	-0.061	0.109	0.125
		40	0.013	0.111	0.112	-0.042	0.117	0.124	-0.062	0.107	0.124
	CFE3	10	-0.021	0.113	0.115	-0.076	0.123	0.144	-0.071	0.109	0.130
		20	0.002	0.119	0.119	-0.054	0.122	0.133	-0.074	0.101	0.125
		30	0.012	0.120	0.121	-0.042	0.121	0.128	-0.067	0.100	0.121
		40	0.015	0.119	0.120	-0.040	0.119	0.126	-0.065	0.099	0.118
	QMLE		0.007	0.116	0.116	-0.009	0.107	0.108	-0.006	0.068	0.069

CASE 3			(0.20, 0.20)			(0.20, 0.40)			(0.20, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	CFE1, CFE2	10	-0.012	0.052	0.053	-0.010	0.058	0.059	-0.014	0.068	0.070
	CFE3	10	-0.024	0.047	0.053	-0.019	0.053	0.056	-0.020	0.058	0.061
		20	-0.024	0.048	0.053	-0.019	0.054	0.057	-0.019	0.057	0.060
		30	-0.023	0.048	0.054	-0.020	0.054	0.057	-0.020	0.057	0.060
		40	-0.023	0.049	0.054	-0.021	0.054	0.057	-0.021	0.057	0.060
	QMLE		0.000	0.035	0.035	0.000	0.034	0.034	0.000	0.030	0.030
$\widehat{\beta}$	CFE1	10	0.001	0.121	0.121	-0.046	0.132	0.140	-0.048	0.125	0.134
		20	0.008	0.121	0.121	-0.044	0.128	0.135	-0.061	0.119	0.134
		30	0.010	0.118	0.118	-0.046	0.125	0.134	-0.069	0.118	0.137
		40	0.003	0.115	0.115	-0.055	0.122	0.134	-0.081	0.115	0.141
	CFE2	10	-0.018	0.107	0.108	-0.053	0.121	0.132	-0.049	0.125	0.134
		20	0.005	0.110	0.110	-0.033	0.119	0.123	-0.048	0.120	0.129
		30	0.018	0.110	0.112	-0.021	0.119	0.121	-0.042	0.118	0.125
		40	0.022	0.109	0.111	-0.017	0.116	0.117	-0.040	0.117	0.123
	CFE3	10	-0.022	0.105	0.107	-0.062	0.116	0.131	-0.051	0.105	0.117
		20	0.003	0.110	0.110	-0.039	0.113	0.120	-0.060	0.099	0.116
		30	0.017	0.113	0.114	-0.024	0.113	0.116	-0.052	0.099	0.112
		40	0.023	0.113	0.115	-0.018	0.111	0.113	-0.047	0.097	0.108
	QMLE		-0.001	0.096	0.096	-0.006	0.084	0.084	-0.004	0.052	0.053

Notes to Table S3: Simulations are conducted using 5,000 observations across 10,000 trials. Cases 2 and 3 are considered. CFE1–3 are the three closed-form estimators being considered. QMLE is the quasi-maximum likelihood estimator. For CFE1–2, K does not impact $\widehat{\alpha}$, so results for only the first value are shown. For CFE3, both $\widehat{\alpha}$ and $\widehat{\beta}$ depend on K through their shared dependence on $\widehat{\phi}$. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. RMSE is the root mean square error measured with respect to the true parameter value.

TABLE S4

			(α_0, β_0)								
			(0.40, 0.20)			(0.35, 0.40)			(0.35, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	CFE1, CFE2	10	-0.066	0.081	0.104	-0.051	0.094	0.107	-0.090	0.131	0.158
	CFE3	10	-0.104	0.075	0.128	-0.075	0.081	0.110	-0.103	0.086	0.134
		20	-0.102	0.078	0.129	-0.077	0.080	0.110	-0.097	0.085	0.129
		30	-0.103	0.078	0.190	-0.079	0.080	0.112	-0.095	0.085	0.127
		40	-0.103	0.078	0.130	-0.081	0.078	0.112	-0.096	0.084	0.128
	QMLE		0.000	0.046	0.046	0.000	0.041	0.041	0.000	0.036	0.036
$\widehat{\beta}$	CFE1	10	0.011	0.124	0.124	-0.022	0.144	0.146	-0.054	0.189	0.197
		20	0.020	0.124	0.126	-0.020	0.139	0.140	-0.081	0.180	0.197
		30	0.028	0.122	0.125	-0.017	0.138	0.139	-0.098	0.178	0.203
		40	0.028	0.120	0.123	-0.020	0.134	0.136	-0.108	0.176	0.206
	CFE2	10	0.009	0.123	0.123	-0.022	0.146	0.148	-0.008	0.186	0.186
		20	0.031	0.124	0.128	-0.007	0.144	0.144	-0.019	0.175	0.176
		30	0.043	0.126	0.133	0.005	0.144	0.144	-0.021	0.172	0.174
		40	0.048	0.124	0.133	0.010	0.140	0.141	-0.022	0.171	0.172
	CFE3	10	0.001	0.111	0.111	-0.030	0.125	0.128	-0.008	0.124	0.124
		20	0.028	0.114	0.117	-0.017	0.118	0.119	-0.026	0.118	0.121
		30	0.046	0.116	0.125	0.000	0.119	0.119	-0.032	0.116	0.120
		40	0.054	0.116	0.128	0.009	0.118	0.118	-0.029	0.115	0.119
	QMLE		-0.002	0.056	0.056	-0.003	0.052	0.052	-0.002	0.031	0.031

Notes to Table S4: Simulations are conducted using 5,000 observations across 10,000 trials. Case 4 is considered. CFE1–3 are the three closed-form estimators being considered. QMLE is the quasi-maximum likelihood estimator. For CFE1–2, K does not impact $\widehat{\alpha}$, so results for only the first value are shown. For CFE3, both $\widehat{\alpha}$ and $\widehat{\beta}$ depend on K through their shared dependence on $\widehat{\phi}$. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. RMSE is the root mean square error measured with respect to the true parameter value.

TABLE S5

$T = 5,000$			(α_0, β_0)								
			(0.10, 0.20)			(0.10, 0.40)			(0.10, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	GLS1	10	0.000	0.027	0.027	0.003	0.026	0.026	0.011	0.025	0.028
		20	0.001	0.027	0.027	0.007	0.026	0.026	0.011	0.025	0.028
		30	0.001	0.027	0.027	0.004	0.026	0.026	0.012	0.025	0.028
		40	0.001	0.027	0.027	0.007	0.026	0.027	0.012	0.026	0.028
	GLS2	10	-0.001	0.026	0.026	0.003	0.026	0.026	0.011	0.025	0.028
		20	0.000	0.027	0.027	0.003	0.026	0.026	0.011	0.025	0.028
		30	0.000	0.026	0.026	0.004	0.026	0.026	0.011	0.025	0.028
		40	0.000	0.027	0.027	0.003	0.026	0.026	0.011	0.025	0.028
	QMLE		-0.001	0.027	0.027	0.001	0.027	0.027	0.001	0.024	0.024
$\widehat{\beta}$	GLS1	10	0.043	0.136	0.143	-0.036	0.143	0.147	0.098	0.134	0.166
		20	0.040	0.137	0.143	0.031	0.138	0.141	0.093	0.132	0.162
		30	0.034	0.134	0.138	-0.039	0.138	0.144	0.092	0.130	0.159
		40	0.034	0.134	0.138	0.025	0.134	0.136	0.090	0.128	0.157
	GLS2	10	0.037	0.135	0.139	-0.045	0.142	0.148	0.090	0.130	0.159
		20	0.034	0.136	0.140	-0.044	0.143	0.149	0.089	0.131	0.159
		30	0.030	0.135	0.138	-0.046	0.139	0.146	0.091	0.131	0.160
		40	0.030	0.135	0.138	-0.046	0.140	0.148	0.091	0.130	0.158
	QMLE		0.030	0.150	0.153	-0.011	0.148	0.148	-0.012	0.101	0.102
$T = 2,000$											
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	GLS1	10	0.008	0.039	0.040	0.013	0.040	0.042	0.023	0.040	0.046
		20	0.008	0.040	0.041	0.014	0.040	0.043	0.024	0.040	0.047
		30	0.008	0.040	0.041	0.014	0.040	0.042	0.023	0.040	0.046
		40	0.009	0.040	0.041	0.014	0.040	0.042	0.023	0.040	0.046
	GLS2	10	0.006	0.039	0.039	0.012	0.039	0.041	0.022	0.040	0.046
		20	0.007	0.040	0.040	0.013	0.040	0.042	0.022	0.040	0.046
		30	0.007	0.040	0.040	0.013	0.040	0.042	0.022	0.040	0.046
		40	0.007	0.039	0.040	0.013	0.040	0.042	0.022	0.040	0.046
	QMLE		0.000	0.042	0.042	0.002	0.041	0.041	0.003	0.038	0.038
$\widehat{\beta}$	GLS1	10	0.070	0.168	0.182	-0.058	0.173	0.182	-0.160	0.170	0.234
		20	0.063	0.167	0.179	-0.068	0.172	0.185	-0.164	0.169	0.235
		30	0.061	0.166	0.177	-0.069	0.172	0.186	-0.168	0.169	0.238
		40	0.054	0.166	0.174	-0.071	0.172	0.186	-0.166	0.169	0.237
	GLS2	10	0.065	0.165	0.177	-0.066	0.171	0.183	-0.169	0.170	0.240
		20	0.056	0.165	0.174	-0.073	0.171	0.186	-0.168	0.170	0.239
		30	0.058	0.165	0.175	-0.074	0.171	0.187	-0.173	0.170	0.243
		40	0.052	0.165	0.173	-0.074	0.171	0.186	-0.174	0.170	0.243
	QMLE		0.082	0.202	0.218	-0.005	0.201	0.201	-0.033	0.165	0.168

Notes to Table S5: Simulations are conducted using 5,000 and 2,000 observations across 10,000 trials. Case 1 is considered. GLS1–2 use starting values from CFE1–2, respectively, and set $l = 10$. QMLE is the quasi-maximum likelihood estimator. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. RMSE is the root mean square error measured with respect to the true parameter value.

TABLE S6

CASE 2			(α_0, β_0)								
			(0.15, 0.20)			(0.15, 0.40)			(0.15, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	GLS1	10	-0.002	0.031	0.031	0.002	0.031	0.031	0.014	0.041	0.043
		20	-0.001	0.031	0.031	0.002	0.031	0.031	0.015	0.067	0.069
		30	-0.001	0.031	0.031	0.002	0.031	0.031	0.015	0.055	0.057
		40	-0.001	0.031	0.031	0.002	0.031	0.031	0.014	0.045	0.047
	GLS2	10	-0.002	0.031	0.031	0.002	0.031	0.031	0.014	0.029	0.032
		20	-0.002	0.031	0.031	0.002	0.031	0.031	0.014	0.029	0.032
		30	-0.001	0.031	0.031	0.002	0.031	0.031	0.014	0.029	0.032
		40	-0.001	0.032	0.032	0.002	0.031	0.031	0.013	0.030	0.032
	QMLE		-0.001	0.032	0.032	0.000	0.031	0.031	0.001	0.027	0.027
$\widehat{\beta}$	GLS1	10	0.020	0.112	0.114	-0.024	0.111	0.114	-0.091	0.122	0.152
		20	0.017	0.112	0.113	-0.025	0.111	0.114	-0.091	0.121	0.151
		30	0.014	0.111	0.112	-0.027	0.111	0.114	-0.095	0.118	0.151
		40	0.014	0.111	0.112	-0.027	0.109	0.113	-0.096	0.119	0.153
	GLS2	10	0.015	0.111	0.112	-0.028	0.110	0.113	-0.097	0.107	0.144
		20	0.013	0.112	0.113	-0.027	0.112	0.115	-0.095	0.108	0.144
		30	0.010	0.113	0.113	-0.027	0.112	0.115	-0.093	0.109	0.144
		40	0.010	0.112	0.113	-0.027	0.112	0.115	-0.091	0.108	0.141
	QMLE		0.007	0.116	0.116	-0.009	0.107	0.108	-0.006	0.068	0.069
CASE 3			(0.20, 0.20)			(0.20, 0.40)			(0.20, 0.60)		
Para.	Est.	K	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE	Mean Bias	SD	RMSE
$\widehat{\alpha}$	GLS1	10	-0.001	0.036	0.036	0.001	0.036	0.037	0.020	0.075	0.077
		20	-0.001	0.036	0.036	0.001	0.036	0.036	0.019	0.064	0.067
		30	-0.001	0.037	0.037	0.002	0.037	0.037	0.020	0.062	0.065
		40	-0.001	0.037	0.037	0.001	0.037	0.037	0.019	0.053	0.056
	GLS2	10	-0.001	0.036	0.036	0.001	0.036	0.036	0.021	0.069	0.072
		20	-0.001	0.036	0.036	0.001	0.036	0.036	0.019	0.056	0.059
		30	-0.001	0.037	0.037	0.001	0.038	0.038	0.019	0.071	0.073
		40	-0.001	0.037	0.037	0.001	0.039	0.039	0.019	0.067	0.070
	QMLE		0.000	0.035	0.035	0.000	0.034	0.034	0.000	0.030	0.030
$\widehat{\beta}$	GLS1	10	-0.001	0.097	0.097	-0.019	0.093	0.095	-0.091	0.109	0.142
		20	-0.001	0.099	0.099	-0.018	0.095	0.097	-0.091	0.107	0.140
		30	-0.001	0.098	0.098	-0.019	0.095	0.097	-0.095	0.106	0.142
		40	0.000	0.100	0.100	-0.019	0.096	0.097	-0.096	0.107	0.144
	GLS2	10	-0.001	0.097	0.097	-0.019	0.092	0.094	-0.096	0.107	0.144
		20	-0.001	0.099	0.099	-0.018	0.095	0.097	-0.095	0.104	0.140
		30	-0.001	0.099	0.099	-0.018	0.096	0.097	-0.094	0.107	0.143
		40	0.000	0.100	0.100	-0.017	0.096	0.098	-0.093	0.108	0.142
	QMLE		-0.001	0.096	0.096	-0.006	0.084	0.084	-0.004	0.052	0.053

Notes to Table S6: Simulations are conducted using 5,000 observations across 10,000 trials. Cases 2 and 3 are considered. GLS1–2 use starting values from CFE1–2, respectively, and set $l = 10$. QMLE is the quasi-maximum likelihood estimator. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. RMSE is the root mean square error measured with respect to the true parameter value.

TABLE S7

Para.	Est.	K	(α_0, β_0) (0.40, 0.20) Mean Bias	SD	RMSE	(0.35, 0.40) Mean Bias	SD	RMSE
$\widehat{\alpha}$	GLS1	10	-0.001	0.047	0.047	0.001	0.060	0.060
		20	-0.001	0.047	0.047	0.001	0.050	0.050
		30	-0.001	0.048	0.048	0.001	0.047	0.047
		40	-0.001	0.049	0.049	0.000	0.047	0.047
	GLS2	10	-0.001	0.047	0.047	0.001	0.048	0.048
		20	0.000	0.157	0.157	0.001	0.053	0.053
		30	-0.001	0.049	0.049	0.001	0.048	0.048
		40	-0.001	0.049	0.049	0.000	0.046	0.046
	QMLE		0.000	0.046	0.046	0.000	0.041	0.041
$\widehat{\beta}$	GLS1	10	-0.002	0.058	0.058	-0.015	0.060	0.062
		20	-0.002	0.059	0.059	-0.015	0.061	0.062
		30	-0.002	0.060	0.060	-0.015	0.061	0.063
		40	-0.002	0.061	0.061	-0.015	0.062	0.063
	GLS2	10	-0.002	0.058	0.058	-0.015	0.059	0.061
		20	-0.002	0.059	0.059	-0.015	0.061	0.063
		30	-0.002	0.060	0.060	-0.015	0.061	0.062
		40	-0.002	0.061	0.061	-0.014	0.061	0.063
	QMLE		-0.002	0.056	0.056	-0.003	0.052	0.052

Notes to Table S7: Simulations are conducted using 5,000 observations across 10,000 trials. Case 4 is considered. GLS1–2 use starting values from CFE1–2, respectively, and set $l = 10$. QMLE is the quasi-maximum likelihood estimator. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. RMSE is the root mean square error measured with respect to the true parameter value.

TABLE S8

α	Est.	Mean Bias	Med. Bias	SD	Dec. Rge	RMSE	MAE	MDAE
0.10	GLS	0.000	-0.002	0.027	0.070	0.027	0.022	0.018
	QMLE	-0.001	-0.002	0.027	0.070	0.027	0.022	0.019
0.20	GLS	0.000	-0.002	0.035	0.090	0.035	0.028	0.024
	QMLE	-0.001	-0.002	0.035	0.089	0.035	0.028	0.024
0.40	GLS	0.000	-0.002	0.045	0.115	0.045	0.036	0.030
	QMLE	-0.001	-0.002	0.045	0.114	0.045	0.036	0.030
0.80	GLS	-0.006	-0.006	0.056	0.142	0.056	0.045	0.038
	QMLE	-0.002	-0.003	0.057	0.148	0.057	0.046	0.038

Notes to Table S8: Simulations are conducted using 5,000 observations across 10,000 trials. The ARCH(1) case is considered at different levels of persistence. GLS uses starting values from CFE1 and sets $l = 10$. QMLE is the quasi-maximum likelihood estimator. Mean bias is the average difference between the parameter estimate and the true parameter value. SD is the standard deviation. Dec. Rge. is the decile range of the parameter estimates (defined as the difference between the 90th and 10th percentiles). RMSE is the root mean square error measured with respect to the true parameter value. MAE and MDAE are the mean absolute error and median absolute error, respectively, also measured with respect to the true parameter value.