SUPPORTING INFORMATION: QUANTILE-BASED SMOOTH TRANSITION VALUE-AT-RISK ESTIMATION

STEFAN HUBNER AND PAVEL ČÍŽEK

A1. PROOFS OF RESULTS

Within the following derivations, let $\mathbb{P}_n$ be the empirical distribution that puts mass $d\mathbb{P}_n = n^{-1}$ to each observation $u_1, \ldots, u_n$, such that $\mathbb{P}_n f(u_t) = \int f(u_t) d\mathbb{P}_n = n^{-1} \sum_{t=1}^{n} f(u_t)$ for any measurable function f. Also, let the vector $z_t^m = (|u_{t-1}|, \ldots, |u_{t-m}|)^T$, and $z_t^m(\boldsymbol{\zeta}) = \left[G_t(\boldsymbol{\zeta})\, z_t^m, (1 - G_t(\boldsymbol{\zeta}))z_t^m\right]^T$, where the latter is defined as a function of $\boldsymbol{\zeta}$ to highlight the dependence on the transition function. Furthermore, let the data considered in the second stage be $z_t(\boldsymbol{\alpha}) = [1, |u_{t-1}|, \ldots, |u_{t-p}|, \sigma_{t-1}(\boldsymbol{\alpha}), \ldots, \sigma_{t-q}(\boldsymbol{\alpha})]^T$, which is a function of $\boldsymbol{\alpha} := [\boldsymbol{\alpha}^{\mathrm{I,T}}, \boldsymbol{\alpha}^{\mathrm{II,T}}, \boldsymbol{\zeta}^T]^T$, where $\boldsymbol{\zeta}$ refers to the location and scale parameters entering the first stage. Similarly, we define $z_t(\boldsymbol{\alpha}, \boldsymbol{\zeta}) = [G_t(\boldsymbol{\zeta})\, z_t(\boldsymbol{\alpha}), (1 - G_t(\boldsymbol{\zeta}))z_t(\boldsymbol{\alpha})]^T$ to be the vector that stacks both regimes' weighted data. Also recall that $\boldsymbol{a} = [\boldsymbol{\alpha}^{\mathrm{I,T}}, \boldsymbol{\alpha}^{\mathrm{II,T}}, \boldsymbol{q}^T, \boldsymbol{\zeta}^T]^T$.

Now the right-side derivative of the check-function $\rho_{\tau_k}(u) = u(\tau_k - \mathbb{1}\{u \le 0\})$ is given by $\psi_{\tau_k,t}(u) := (\tau_k - \mathbb{1}\{u \le 0\})$ so that the directional derivative of the objective function defined in equation (3.1) is given as

$$g_n(\boldsymbol{a}) = g_n(\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}, \boldsymbol{q}, \boldsymbol{\zeta}) := \mathbb{P}_n \sum_{k=1}^{K} \underbrace{\begin{bmatrix} z_t^m(\boldsymbol{\zeta}) q_k \\ \mathbb{1}\{k=1\}\, [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T z_t^m(\boldsymbol{\zeta}) \\ \vdots \\ \mathbb{1}\{k=K\}\, [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T z_t^m(\boldsymbol{\zeta}) \\ q_k(\boldsymbol{\alpha}^{\mathrm{I}} - \boldsymbol{\alpha}^{\mathrm{II}})^T z_t^m(\boldsymbol{\zeta}) \dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \end{bmatrix}}_{\boldsymbol{x}_{t,k}(\boldsymbol{a})} (\tau_k - \mathbb{1}\left\{u_t \le q_k [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T z_t^m(\boldsymbol{\zeta})\right\}).$$

Similarly, without making any statements about convergence yet, the corresponding population equivalent can be written as

$$g(\boldsymbol{a}) = g(\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}, \boldsymbol{q}, \boldsymbol{\zeta}) := \mathbb{E} \sum_{k=1}^{K} \begin{bmatrix} z_t^m(\boldsymbol{\zeta}) q_k \\ \mathbb{1}\{k=1\}\, [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T z_t^m(\boldsymbol{\zeta}) \\ \vdots \\ \mathbb{1}\{k=K\}\, [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T z_t^m(\boldsymbol{\zeta}) \\ q_k(\boldsymbol{\alpha}^{\mathrm{I}} - \boldsymbol{\alpha}^{\mathrm{II}})^T z_t^m(\boldsymbol{\zeta}) \frac{\partial G_t(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \end{bmatrix} (\tau_k - F_{u_t|\mathcal{F}_{t-1}}(q_k [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T z_t^m(\boldsymbol{\zeta})))$$

by applying the law of iterated expectations. In addition, we will frequently make use of the identity $\partial F_{u_t|\mathcal{F}_{t-1}}(x)/\partial x = \sigma_t{}^{-1} \partial F_\varepsilon(x)/\partial x = \sigma_t^{-1} f_\varepsilon(x)$: because equation (2.4) establishes $F^{-1}_{u_t|\mathcal{F}_{t-1}}(x) = \sigma_t F_\varepsilon^{-1}(x)$ and because both $F_{u_t|\mathcal{F}_{t-1}}$ and F_ε are monotone and differentiable by Assumptions 4.3 and 4.4, the expression follows by applying the inverse function theorem.

Proof of Theorem 4.1: The proof is split into three parts: we first discuss the identification of the first-stage parameters, then their consistency, and finally, the convergence rate is derived. First, we show that for $\boldsymbol{\alpha}(\tau) = [\boldsymbol{\alpha}^{\mathrm{I,T}}(\tau), \boldsymbol{\alpha}^{\mathrm{II,T}}(\tau)]$,[15] the vector $\left[\boldsymbol{\alpha}_0^T(\tau), \boldsymbol{\zeta}_0^T\right]^T$ is in fact the minimum of the objective function $\mathbb{E}\rho_\tau(u)$.

[15] In the first stage, the parameters of the transition function are estimated globally by using the objective function at quantiles $\tau_1, \ldots, \tau_K$ and are therefore not included in the locally estimated $\boldsymbol{\alpha}(\tau)$.

© 2019 Royal Economic Society.

We show this for an arbitrary quantile $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$, from which the composite quantile result follows as $K > 1$ by Assumption 4.7.

To see this, let the objective function $\mathfrak{m}(\boldsymbol{\alpha}(\tau), \boldsymbol{\zeta}) = \mathbb{E}\rho_\tau(u_t - \boldsymbol{\alpha}^T(\tau) z_t^m(\boldsymbol{\zeta}))$. Then, we have global identification if and only if $\mathfrak{m}(\boldsymbol{\alpha}(\tau), \boldsymbol{\zeta}) - \mathfrak{m}(\boldsymbol{\alpha}_0(\tau), \boldsymbol{\zeta}_0) > 0$ for any $\left[\boldsymbol{\alpha}^T(\tau), \boldsymbol{\zeta}^T\right] \neq \left[\boldsymbol{\alpha}_0^T(\tau), \boldsymbol{\zeta}_0^T\right]$.

Let $\boldsymbol{\alpha}(\tau)^\Delta = \boldsymbol{\alpha}^{\mathrm{I}}(\tau) - \boldsymbol{\alpha}^{\mathrm{II}}(\tau)$, and also let v_t be the probability measure of u_t conditional upon the filtration $\mathcal{F}_{t-1}$. Then, we have to prove

$$\inf_{\substack{\boldsymbol{\alpha}(\tau)\,:\,\|\boldsymbol{\alpha}(\tau) - \boldsymbol{\alpha}_0(\tau)\| > \delta \\ \boldsymbol{\zeta}\,:\,\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_0\| > \delta}} \mathbb{E}\left[\int \rho_\tau(u_t - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t - \int \rho_\tau(u_t - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t\right] > \varepsilon_\delta, \tag{A.1}$$

where we can discuss the inside part of the expectation into two cases. First, consider $\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) > \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})$:

$$\int \rho_\tau(u_t - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t - \int \rho_\tau(u_t - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t =$$

$$(\tau - 1)\int_{-\infty}^{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}(u_t - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t + \tau \int_{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)} (u_t - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t + \tau \int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}^{+\infty} (u_t - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t$$

$$- (\tau - 1)\int_{-\infty}^{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}(u_t - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t - \tau \int_{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)} (u_t - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t - \tau \int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}^{+\infty} (u_t - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t$$

$$= (1 - \tau)\int_{-\infty}^{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}(\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t + \Omega_1 + \tau \int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}^{+\infty} (\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t$$

$$\geq (1 - \tau)\int_{-\infty}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}(\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t + \tau \int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}^{+\infty} (\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t$$

$$= \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) \left[(1 - \tau) \int_{-\infty}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)} dv_t + \tau \int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}^{+\infty} dv_t\right] \tag{A.2}$$

$$= \left(\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})\right)\left[\tau - v_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)]\right], \tag{A.3}$$

where we use the fact that $v_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)] + v_t(F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau), +\infty) = 1$ (v_t is a probability measure) and where the inequality in equation (A.2) follows from the fact that

$$\Omega_1 := \int_{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)} (u_t - \tau\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - (1 - \tau)\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t \geq (1 - \tau)\int_{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)} (\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t.$$

Similarly, for $\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) < \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})$, it holds

$$= (1 - \tau)\int_{-\infty}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0)}(\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0))dv_t - \Omega_2 + \tau \int_{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}^{+\infty} (\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta}_0) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t$$

© 2019 Royal Economic Society.

$$\geq (1-\tau)\int_{-\infty}^{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0})}(z_t^m(\boldsymbol{\zeta})\boldsymbol{\alpha}(\tau)^T - z_t^m(\boldsymbol{\zeta_0})\boldsymbol{\alpha}_0(\tau)^T)dv_t + \tau\int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0})}^{+\infty}(\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0}) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t$$

$$\left(\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0})\right)\left[\tau - v_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)]\right], \tag{A.4}$$

where we use

$$\Omega_2 := \int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0})}^{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}(u_t - \tau\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - (1-\tau)\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0}))dv_t \leq \tau\int_{\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0})}^{\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})}(\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0}) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}))dv_t.$$

Then, by the definition of the τ^{th} quantile, $v_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)] \leq \tau$, and the final expressions in both equation (A.3) and equation (A.4) are thus non-negative. Thus, the expectation in (A.1) with respect to the measure of z_t is zero for parameters other than the true parameter if and only if it holds for all z_t that $\left(\boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0}) - \boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta})\right) = 0$. The parameters are thus identified if the following identification statement holds:

$$\mathbb{E}\left[\boldsymbol{\alpha}(\tau)^T z_t^m(\boldsymbol{\zeta}) - \boldsymbol{\alpha}_0(\tau)^T z_t^m(\boldsymbol{\zeta_0})\right] = \mathbb{E}\left[(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)^T F_\varepsilon^{-1}(\tau) z_t^m(\boldsymbol{\zeta}) + \boldsymbol{\alpha}_0^\Delta F_\varepsilon^{-1}(\tau) z_t^m(G_t(\boldsymbol{\zeta}) - G_t(\boldsymbol{\zeta}_0))\right]$$
$$= F_\varepsilon^{-1}(\tau)\mathbb{E}\left[\begin{pmatrix}\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\\ \boldsymbol{\alpha}^\Delta\end{pmatrix}^T\begin{pmatrix} z_t^m(\boldsymbol{\zeta})\\ z_t^m(G_t(\boldsymbol{\zeta}) - G_t(\boldsymbol{\zeta}_0))\end{pmatrix}\right] = 0$$

if and only if $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ and $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0$. This, however, follows from the global identification Assumption 4.6, which states that $\mathbb{E}[G_t(\boldsymbol{\zeta})z_t^m, (1-G_t(\boldsymbol{\zeta}))z_t^m, z_t^m(G_t(\boldsymbol{\zeta}) - G_t(\boldsymbol{\zeta}_0))]^T[G_t(\boldsymbol{\zeta})z_t^m, (1-G_t(\boldsymbol{\zeta}))z_t^m, z_t^m(G_t(\boldsymbol{\zeta}) - G_t(\boldsymbol{\zeta}_0))]$ has full rank for any $\boldsymbol{\zeta} \neq \boldsymbol{\zeta_0}$. Thus, $\boldsymbol{\alpha}(\tau) = \boldsymbol{\alpha} F_\varepsilon^{-1}(\tau)$ is identified for an arbitrary $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$. Note that the result does not hold for $\tau = \frac{1}{2}$, for which $F_\varepsilon^{-1}(\tau)$ is zero.

Given this identification result, the consistency for a fixed dimension m would follow once the sample objective function is shown to converge uniformly in probability to its (continuous) population counterpart, which we just analysed above (Theorem 2.1, p. 2121 in Newey and McFadden, 1994); note that for $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$, the parameter space $\boldsymbol{\Theta}_1^\tau$ is a compact subset of $\mathbb{R}^{2(m+1)+2}$. As $m \to \infty$, the consistency of the proposed sieve estimator is obtained by Theorem 1 in Chen and Shen (1998, p. 297), for which we have to check conditions A.1–A.4 therein. Let us first denote $\mathfrak{m} = \rho_\tau \circ \omega_t$, where $\omega_t : (\boldsymbol{\alpha}^{\text{I}}(\tau), \boldsymbol{\alpha}^{\text{II}}(\tau), \boldsymbol{\zeta}) \mapsto z_t^{m,T}\boldsymbol{\alpha}^{\text{II}}(\tau) + (\boldsymbol{\alpha}^{\text{I}}(\tau) - \boldsymbol{\alpha}^{\text{II}}(\tau))^T z_t^m G(\xi_t, \zeta, \eta)$, $\boldsymbol{\alpha}(\tau) = [\boldsymbol{\alpha}^{\text{I}}(\tau, T), \boldsymbol{\alpha}^{\text{II,T}}(\tau)]^T$, and $\boldsymbol{\zeta} = [\zeta, \eta]^T$. For condition A.4, we have to show that the function $m = \rho_\tau{\circ}\omega_t$ is Lipschitz. By Assumption 2.1, the transition function $G_t(\boldsymbol{\zeta})$ is Lipschitz in $\boldsymbol{\zeta}$, and by construction, ω_t is Lipschitz in $\boldsymbol{\alpha}$ and $G()$. Thus, ω_t is Lipschitz in $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ because the property is preserved under function composition.[16] The piecewise linear function ρ_τ is Lipschitz as well, and thus so is $m = \rho_\tau{\circ}\omega_t$. Furthermore, note that we actually have a special case of s = 1 in the Hölder condition A.4, which by Chen and Shen (1998, Remark 1(c)) implies their condition A.2. In addition, condition A.1 holds by Assumption 4.5, which states that u_t is β-mixing with a decay rate satisfying $\beta_s \leq \beta_0 s^{-(2+\delta)}$ for some $\delta > 0$. Finally, we note that, by Chen (2008, p. 5595), it holds for Lipschitz functions that the bracketing numbers $\log N_{[]}(\epsilon^s, \mathcal{F}_n, \|\bullet\|_2) \leq \log N(\epsilon, \Theta_1^\tau, \|\bullet\|) \leq Cm\log(\frac{1}{\epsilon})$ for some constant $C > 0$, where m is the dimension of the sieve parameter space, and that this implies their condition A.3. Therefore, we can apply Theorem 1 in Chen and Shen (1998) and conclude that the first-stage sieve estimator is consistent.

Next, to derive the convergence rate of the first-stage estimator, we have to show that the directional derivative of the nondifferentiable g_n around the true value $\boldsymbol{a}_0$ is positive in every direction with probability

[16] If both f and g are Lipschitz, so is $f{\circ}g$ because $(f{\circ}g)(x) - (f{\circ}g)(x_0) = f(g(x)) - f(g(x_0)) \leq C_f(g(x) - g(x_0)) \leq C_f C_g(x - x_0)$ for finite constants C_f and C_g.

© 2019 Royal Economic Society.

tending to one. Let $\varphi_t(\boldsymbol{a}) = \sum_{k=1}^{K} \boldsymbol{x}_t(\boldsymbol{a})(\tau_k - \mathbb{1}\left\{u_t \leq \boldsymbol{a}^T \boldsymbol{x}_t(\boldsymbol{a})\right\})$. Then, we have to prove that

$$\forall \epsilon > 0 : \exists B < \infty : \lim_{n\to\infty} \mathbf{P}\left[\inf_{\boldsymbol{\lambda}\in\mathbb{R}^{2(m+1)+K+2}:\|\boldsymbol{\lambda}\|=1} \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t\left(\boldsymbol{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) > 0 \right] > 1 - \epsilon. \tag{A.5}$$

Adding and subtracting $\mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) \middle| \mathcal{F}_{t-1}\right] - \mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1}\right]$ as well as $\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t\left(\boldsymbol{a}_0\right)$, the random quantity in (A.5) can be rewritten as:

$$\mathbb{P}_n \boldsymbol{\lambda}^T \ \varphi_t\left(\boldsymbol{a}_0 + B\left(\tfrac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) \tag{A.6}$$

$$= \left\{\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t\left(\boldsymbol{a}_0\right)\right\} \tag{A.7}$$

$$+ \left\{\left(\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t\left(\boldsymbol{a}_0 + B\left(\tfrac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) - \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t\left(\boldsymbol{a}_0\right)\right)\right. \tag{A.8}$$

$$\left. - \left(\mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0 + B\left(\tfrac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) \middle| \mathcal{F}_{t-1}\right] - \mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1}\right]\right)\right\}$$
$$+ \left\{\mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0 + B\left(\tfrac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) \middle| \mathcal{F}_{t-1}\right] - \mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1}\right]\right\}. \tag{A.9}$$

This expansion can be now analysed term by term. Term (A.8) will turn out to be stochastically negligible, whereas terms (A.7) and (A.9) can be made explicit. Let us start by rewriting (A.9) in the following way, using the definition of φ_t:

$$\begin{aligned}
\mathbb{P}_n \boldsymbol{\lambda}^T \ & \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right) \middle| \mathcal{F}_{t-1}\right] - \mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\varphi_t\left(\boldsymbol{a}_0\right) \middle| \mathcal{F}_{t-1}\right] \\
&= \mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\sum_{k=1}^{K} \left(\tau_k - \mathbb{1}\left\{u_t \leq \left(\boldsymbol{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right)^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)\right\}\right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \middle| \mathcal{F}_{t-1}\right] \\
&- \mathbb{P}_n \boldsymbol{\lambda}^T \mathbb{E}\left[\sum_{k=1}^{K} \left(\tau_k - \mathbb{1}\left\{u_t \leq \boldsymbol{a}_0^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)\right\}\right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \middle| \mathcal{F}_{t-1}\right] \\
&= -\mathbb{P}_n \boldsymbol{\lambda}^T \sum_{k=1}^{K} \left(F_{u_t|\mathcal{F}_{t-1}}\left(\left(\boldsymbol{a}_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}\right)^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)\right) - F_{u_t|\mathcal{F}_{t-1}}\left(\boldsymbol{a}_0^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)\right)\right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0).
\end{aligned}$$

Applying the Taylor expansion around $\boldsymbol{a}_0$ to the first term for each $t \in \mathcal{I}_{m,n}$ yields:

$$\begin{aligned}
-\mathbb{P}_n \boldsymbol{\lambda}^T & \sum_{k=1}^{K} f_{u_t|\mathcal{F}_{t-1}}\left(\boldsymbol{a}_0^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)\right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)^T B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} + \mathcal{O}_p\left(m^1 n^{-1}\right) \\
&= -B\left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}^T \mathbb{P}_n \left[\sum_{k=1}^{K} \frac{1}{\sigma_t} f_\varepsilon\left(F_\varepsilon^{-1}(\tau_k)\right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)^T\right] \boldsymbol{\lambda} + \mathcal{O}_p\left(m^1 n^{-1}\right) \\
&= \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}^T \boldsymbol{D}_{1,n,m} \boldsymbol{\lambda} + \mathcal{O}_p\left(m^1 n^{-1}\right),
\end{aligned}$$

© 2019 Royal Economic Society.

using equations (2.6) and (2.4) for the last step and defining $\boldsymbol{D}_{1,n,m} =$

$$-\mathbb{E}\mathbb{P}_n \frac{1}{\sigma_t} \left[\begin{array}{c|c|c} \boldsymbol{\iota}_K^T h(\boldsymbol{q}) \nu(\boldsymbol{\zeta}) \nu(\boldsymbol{\zeta})^T \otimes z_t z_t^T & z_t^T \bar{\alpha}(\boldsymbol{\zeta}) (\nu(\boldsymbol{\zeta}) \otimes z_t) \otimes h(\boldsymbol{\iota}_K)^T & \boldsymbol{\iota}_K^T h(\boldsymbol{q}) \boldsymbol{\alpha}^{\Delta} z_t (\nu(\boldsymbol{\zeta}) \otimes z_t) \frac{\partial G}{\partial \boldsymbol{\zeta}^T} \\ \hdashline & \mathrm{diag}(\boldsymbol{s}) (\bar{\boldsymbol{\alpha}}^T z_t)^2 & \bar{\boldsymbol{\alpha}}^T z_t(\boldsymbol{\zeta}) \boldsymbol{\alpha}^{\Delta} z_t h(\boldsymbol{\iota}_K) \frac{\partial G}{\partial \boldsymbol{\zeta}^T} \\ \hdashline & & \boldsymbol{\iota}_K^T h(\boldsymbol{q}) (\boldsymbol{\alpha}^{\Delta} z_t)^2 \frac{\partial G}{\partial \boldsymbol{\zeta}} \frac{\partial G}{\partial \boldsymbol{\zeta}^T} \end{array} \right],$$

where $\bar{\boldsymbol{\alpha}} = G_t(\boldsymbol{\zeta}) \boldsymbol{\alpha}^{\mathrm{I}} + (1 - G_t(\boldsymbol{\zeta})) \boldsymbol{\alpha}^{\mathrm{II}}$, $\nu(\boldsymbol{\zeta}) = [G_t(\boldsymbol{\zeta}), 1 - G_t(\boldsymbol{\zeta})]^T$, $\boldsymbol{s} = (s_1, \ldots, s_K)^T$, $\boldsymbol{q} = (q_1, \ldots, q_K)^T$, $h(\boldsymbol{\chi}) = \boldsymbol{s} \odot \boldsymbol{q} \odot \boldsymbol{\chi}$, and $\odot$ is the Hadamard product.

To analyse the remaining terms, let $\eta_t(\nu) = \varphi_t(\boldsymbol{a}_0 + \boldsymbol{\nu}) - \varphi_t(\boldsymbol{a}_0)$. Then, term (A.8) is negligible in probability with rate $\left(\frac{m}{n}\right)^{-\frac{1}{2}}$ if

$$\sup_{\|\nu\| \le B\left(\frac{m}{n}\right)^{\frac{1}{2}}} \left| \mathbb{P}_n \boldsymbol{\lambda}^T \left(\eta_t(\nu) - \mathbb{E}\left[\eta_t(\nu) \middle| \mathcal{F}_{t-1} \right] \right) \right| = o_p\left(\frac{1}{\sqrt{mn}} \right).$$

The considered process is a martingale difference sequence. The next step is to divide the ball defined as $\left\{ \boldsymbol{\nu} \in \mathbb{R}^{2(m+1)+K+2} : \|\boldsymbol{\nu}\| \le B\left(\frac{m}{n}\right)^{\frac{1}{2}} \right\}$ in equation (A.5) into cubes $\mathcal{C}_j \subset \mathbb{R}^{2(m+1)+K+2}$ centred at ν_j and with side-length $m^{\frac{1}{2}} n^{-\frac{5}{2}}$. The resulting cardinality for $2(m+1)+K+2$ dimensions is then $N(n) := \left\| \left\{ \mathcal{C}_j \right\} \right\| = (2n)^{2(m+1)+K+2}$. Now for each $k \in \mathcal{I}_{1,K}$, the term $\eta_t(\nu)$ can be bounded by

$$\eta_t(\nu) \le \eta_t(\nu_j) + b_{k,t}(\nu_j) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0), \tag{A.10}$$

and similarly,

$$\eta_t(\nu) \ge \eta_t(\nu_j) + \left(b_{k,t}(\nu_j) - d_{k,t}(\nu_j) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0), \tag{A.11}$$

with $b_{k,t}(\nu_j) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)$ and $d_{k,t}(\nu_j) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)$ being the process η_t evaluated at the maximum possible distance on each axis from the centre to the boundary of the cube and the maximum possible distance on each axis between the boundaries of the cube $\mathcal{C}_j$, respectively:

$$\begin{aligned} b_{k,t}(\nu_j) = \mathbb{1} & \left\{ u_t < \left(\boldsymbol{a}_0 + \nu_j \right)^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\} \\ & - \mathbb{1} \left\{ u_t < \left(\boldsymbol{a}_0 + \nu_j \right)^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) + B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}} \right) \left\| \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\| \right\}, \end{aligned}$$

$$\begin{aligned} d_{\tau,t}(\nu_j) = \mathbb{1} & \left\{ u_t < \left(\boldsymbol{a}_0 + \nu_j \right)^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) + B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}} \right) \left\| \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\| \right\} \\ & - \mathbb{1} \left\{ u_t < \left(\boldsymbol{a}_0 + \nu_j \right)^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) - B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}} \right) \left\| \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\| \right\}. \end{aligned}$$

Taking expectations of (A.11) and subtracting it from (A.10) implies that, for all $\nu \in \mathcal{C}_j$, for all t, and for all k, it holds that

$$\begin{aligned} \left(\eta_t(\nu) - \mathbb{E}\left[\eta_t(\nu) \middle| \mathcal{F}_{t-1} \right] \right) \le & \left(\eta(\nu_j) - \mathbb{E}\left[\eta_t(\nu_j) \middle| \mathcal{F}_{t-1} \right] \right) \\ & + \left(b_{k,t}(\nu_j) - \mathbb{E}\left[b_{k,t}(\nu_j) \middle| \mathcal{F}_{t-1} \right] \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) + \mathbb{E}\left[d_{k,t}(\nu_t) \middle| \mathcal{F}_{t-1} \right] \boldsymbol{x}_{t,k}(\boldsymbol{a}_0), \end{aligned}$$

which implies that

$$\sup_{\|\nu\| \le B\left(\frac{m}{n}\right)^{\frac{1}{2}}} \left| \mathbb{P}_n \boldsymbol{\lambda}^T \left(\eta_t(\nu) - \mathbb{E}\left[\eta_t(\nu) \middle| \mathcal{F}_{t-1} \right] \right) \right|$$

$$\le \max_{j \in \mathcal{I}_{1,N(n)}} \left| \mathbb{P}_n \left\| \boldsymbol{\lambda}^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\| \left(b_{k,t}(\nu_j) - \mathbb{E}\left[b_{k,t}(\nu_j) \middle| \mathcal{F}_{t-1} \right] \right) \right| \tag{A.12}$$

© 2019 Royal Economic Society.

$$+ \max_{j \in \mathcal{I}_{1,N(n)}} \left| \mathbb{P}_n \left\| \boldsymbol{\lambda}^T \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\| \mathbb{E}\left[d_{k,t}(\nu_t) \middle| \mathcal{F}_{t-1} \right] \right| \tag{A.13}$$

$$+ \max_{j \in \mathcal{I}_{1,N(n)}} \left| \mathbb{P}_n \boldsymbol{\lambda}^T \left(\eta(\nu_j) - \mathbb{E}\left[\eta_t(\nu_j) \middle| \mathcal{F}_{t-1} \right] \right) \right|. \tag{A.14}$$

Expressions (A.12), (A.13), and (A.14) are equivalent to expressions (A.5) through (A.7) in Xiao and Koenker (2009), who show that these terms are asymptotically negligible of order $\sqrt{m/n}$ in probability. As their proof is general and relies on the existence of the exponential bound on innovations imposed in Assumption 4.8, the results also apply to the present analysis because the current problem is piecewise linear with a bounded transition function, with finite second moments of the conditional volatility process, and satisfying Assumption 4.8.

Hence, with term (A.8) being negligible, equation (A.6) can be written as

$$\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) = \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t \left(\boldsymbol{a}_0 \right) + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda}^T \boldsymbol{D}_{1,n,m} \boldsymbol{\lambda} + o_p \left(\left(\frac{m}{n} \right)^{\frac{1}{2}} \right). \tag{A.15}$$

Whenever the right-hand side of this equation exceeds zero, it is implied that the left-hand side does, as well. The left-hand side is, however, positive, with probability tending to one as $B \to \infty$ and $n \to \infty$ because the following equation holds by Assumption 4.6:

$$\inf_{\boldsymbol{\lambda} \in \mathbb{R}^{2(m+1)+K+2} : \|\boldsymbol{\lambda}\| = 1} \left(\frac{m}{n} \right)^{-\frac{1}{2}} \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t \left(\boldsymbol{a}_0 \right) > -\frac{B}{2} \lambda_{n,\min} - o_p(1) < 0,$$

noting that $\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t(\boldsymbol{a}_0) = \mathcal{O}_p(\sqrt{m/n})$ and $\lambda_{n,\min} > 0$ as $n \to \infty$. Statement (A.5) and the convergence rate following from it are thus verified. □

Proof of Theorem 4.2: Let $\widehat{\nu} = \widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0$, where $\widehat{\boldsymbol{a}}_n$ solves the objective function defined in equation (3.16). By Theorem 4.1, we can write $\widehat{\nu}$ as $B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda}$, B in a compact set, with a probability arbitrarily close to 1. This substitution in equation (A.15) leads to

$$\mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t \left(\widehat{\boldsymbol{a}}_n \right) = \mathbb{P}_n \boldsymbol{\lambda}^T \varphi_t \left(\boldsymbol{a}_0 \right) + \boldsymbol{\lambda}^T \boldsymbol{D}_{1,n,m} (\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0) + o_p \left(\left(\frac{m}{n} \right)^{\frac{1}{2}} \right).$$

By construction, the moment function on the left-hand side is zero at the estimate $\widehat{\boldsymbol{a}}_n$. Thus, the right-hand side satisfies for all $\lambda \in \mathbb{R}^m$, $\|\lambda\| = 1$,

$$\boldsymbol{\lambda}^T \left[\mathbb{P}_n \varphi_t \left(\boldsymbol{a}_0 \right) + \boldsymbol{D}_{1,n,m} (\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0) + o_p \left(\frac{1}{\sqrt{n}} \right) \right] = 0,$$

and hence, the expression inside the bracket must be zero. After premultiplying it by $\sqrt{n}$ and $\boldsymbol{D}_{1,n,m}$, the Bahadur representation for $\sqrt{n}(\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0)$ follows as well as the one for $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)$ by considering only the first $2(m+1)$ and the last 2 elements. The submatrix consisting of the corner blocks (upper right, upper left, bottom right, and bottom left corners) of the matrix $\boldsymbol{D}_{1,n,m}$ is denoted $\boldsymbol{D}_m$ here.

Additionally, applying the central limit theorem (Theorem 18.5.3 in Ibragimov and Linnik, 1971) to $\frac{1}{\sqrt{n}} \sum_{t=1}^{N} \varphi_t(\boldsymbol{a}_0)$, it follows for any linear combination of the components of the $(2(m+1)+2)$-dimensional

© 2019 Royal Economic Society.

Bahadur representation that

$$\sqrt{n}\boldsymbol{\mu}_n^T(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)\mathcal{N}\left(\mathbf{0}\,,\ \frac{\sum_{k=1}^{K}\sum_{k'=1}^{K} q_k q_{k'}(\tau_k \wedge \tau_{k'})(1-\tau_k \vee \tau_{k'})}{\left(\sum_{k=1}^{K} q_k^2 f_\varepsilon\left(F_\varepsilon^{-1}(\tau_k)\right)\right)^2}\lim_{m\to\infty}\boldsymbol{\mu}_m^T \boldsymbol{D}_m^{-1}\boldsymbol{\mu}_m\right) \tag{A.16}$$

where $\boldsymbol{\mu}_m \in \mathbb{R}^{2(m+1)+2}$ is such that the limit $\lim_{m\to\infty}\boldsymbol{\mu}_m^T\boldsymbol{D}_m^{-1}\boldsymbol{\mu}_m$ exists. The assumptions of the central limit theorem are satisfied because of the moment and mixing conditions stated in Assumption 4.5, which ensure that the $(2+\delta)$ moments of the data exist and that the mixing coefficients satisfy: $\beta_s \to 0$ and $\sum_{s=1}^{\infty}\beta_s^{\delta/(2+\delta)} < +\infty$. □

For the second stage, the directional derivative of the objective function as in equation (3.18) is re-defined as

$$g_n(\boldsymbol{\theta},\boldsymbol{\alpha}) = g_n\left((\boldsymbol{\theta}^{\mathrm{I}},\boldsymbol{\theta}^{\mathrm{II}},\boldsymbol{\zeta}),(\boldsymbol{\alpha}^{\mathrm{I}},\boldsymbol{\alpha}^{\mathrm{II}},\boldsymbol{\zeta})\right) := \mathbb{P}_n\begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\boldsymbol{\zeta}) \\ z_t(\boldsymbol{\alpha})(1-G_t(\boldsymbol{\zeta})) \\ \boldsymbol{\theta}^{\Delta T} z_t(\boldsymbol{\alpha})\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}}\end{bmatrix}\left(\tau - \mathbb{1}\left\{u_t \le [\boldsymbol{\theta}^{\mathrm{I}},\boldsymbol{\theta}^{\mathrm{II}}]^T z_t(\boldsymbol{\alpha},\boldsymbol{\zeta})\right\}\right)$$

and the one for the population, using the law of iterative expectations, as

$$g(\boldsymbol{\theta},\boldsymbol{\alpha}) = g\left((\boldsymbol{\theta}^{\mathrm{I}},\boldsymbol{\theta}^{\mathrm{II}},\boldsymbol{\zeta}),(\boldsymbol{\alpha}^{\mathrm{I}},\boldsymbol{\alpha}^{\mathrm{II}},\boldsymbol{\zeta})\right) := \mathbb{E}\left[\begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\boldsymbol{\zeta}) \\ z_t(\boldsymbol{\alpha})(1-G_t(\boldsymbol{\zeta})) \\ \boldsymbol{\theta}^{\Delta T} z_t(\boldsymbol{\alpha})\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}}\end{bmatrix}\left(\tau - F_{u|\mathcal{F}_{t-1}}\left([\boldsymbol{\theta}^{\mathrm{I}},\boldsymbol{\theta}^{\mathrm{II}}]^T z_t(\boldsymbol{\alpha},\boldsymbol{\zeta})\right)\right)\right].$$

Note that we re-estimate $\boldsymbol{\zeta}$, and therefore, we consider only the first-order conditions for the parameter $\boldsymbol{\zeta}$ that is a part of $\boldsymbol{\theta}$; parameters within $\boldsymbol{\alpha}$ are fixed and will be substituted for by the first-stage estimates. Because of Assumption 4.3, the population derivative $g(\boldsymbol{\theta},\boldsymbol{\alpha})$ is differentiable, and the partial derivatives with respect to the two parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ evaluated at the true parameters are given by $\Gamma_{\theta,0} = \Gamma_\theta(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0;\mathbf{P})$ and $\Gamma_{\alpha,m,0} := \Gamma_{\alpha,m}(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0;\mathbf{P})$, with

$$\Gamma_\theta(\boldsymbol{\theta},\boldsymbol{\alpha};\mu) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\times$$

$$\int\left[\begin{array}{c:c}\begin{bmatrix} G_t(\boldsymbol{\zeta})^2 & G_t(\boldsymbol{\zeta})(1-G_t(\boldsymbol{\zeta})) \\ G_t(\boldsymbol{\zeta})(1-G_t(\boldsymbol{\zeta})) & (1-G_t(\boldsymbol{\zeta}))^2\end{bmatrix}\otimes z_t(\boldsymbol{\alpha})z_t(\boldsymbol{\alpha})^T & \boldsymbol{\theta}^{\Delta T} z_t(\boldsymbol{\alpha})\begin{bmatrix} G_t(\boldsymbol{\zeta}) \\ 1-G_t(\boldsymbol{\zeta})\end{bmatrix}\otimes z_t(\boldsymbol{\alpha})\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^T} \\ \hdashline & \left(\boldsymbol{\theta}^{\Delta T} z_t(\boldsymbol{\alpha})\right)^2\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}}\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^T}\end{array}\right]d\mu$$

and

$$\Gamma_{\alpha,m}(\boldsymbol{\theta},\boldsymbol{\alpha};\mu) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\int\begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\boldsymbol{\zeta}) \\ z_t(\boldsymbol{\alpha})(1-G_t(\boldsymbol{\zeta})) \\ \boldsymbol{\theta}^{\Delta T} z_t(\boldsymbol{\alpha})\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}}\end{bmatrix}\boldsymbol{\theta}^T\begin{bmatrix}\begin{bmatrix} G_t(\boldsymbol{\zeta}) \\ 1-G_t(\boldsymbol{\zeta})\end{bmatrix}\otimes\begin{bmatrix}\mathbf{0}_{p+1,m} \\ [L^1,\ldots,L^q]^T\otimes z_t^{\mathrm{m}}\end{bmatrix} \\ \boldsymbol{\iota}_q^T[L^1,\ldots,L^q]^T\otimes z_t^{\mathrm{m}T}\boldsymbol{\alpha}^\Delta\dfrac{\partial G}{\partial\zeta}\end{bmatrix}d\mu,$$

respectively, where L is the lag operator. These expectations are well defined and exist by Assumptions 2.1–4.5. In addition, $\Gamma_{\theta,0}$ is positive definite and has full rank because of the invertibility of the conditional scale process in Assumptions 2.2 and 4.9. Finally, let $\Gamma_{\theta,n}(\boldsymbol{\theta},\boldsymbol{\alpha}) = \Gamma_\theta(\boldsymbol{\theta},\boldsymbol{\alpha};\mathbb{P}_n)$ and $\Gamma_{\alpha,n}(\boldsymbol{\theta},\boldsymbol{\alpha}) := \Gamma_{\alpha,m}(\boldsymbol{\theta},\boldsymbol{\alpha};\mathbb{P}_n)$ be the corresponding sample analogues.

© 2019 Royal Economic Society.

Proof of Theorem 4.3: It needs to be shown that $\left\|\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. For this, note that $g(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is differentiable for any $\boldsymbol{\theta}$. Thus, the first-order Taylor expansion around $\boldsymbol{\theta}_0(\tau)$ can be applied, and because of continuity of Γ_θ in $\boldsymbol{\theta}$, it follows that

$$g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - g\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right) = \left(\Gamma_{\theta,0} + o_p(1)\right)\left(\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau)\right).$$

Taking norms, a bound for the right-hand side is obtained with a probability arbitrarily close to 1 for $n \to \infty$:

$$\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - g\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\right\| \geq \frac{1}{2}\lambda_{\min}\left(\Gamma_{\theta,0}\right)\left\|\left(\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau)\right)\right\|, \tag{A.17}$$

with $\lambda_{\min}(\Gamma_{\theta,0})$ being the the smallest eigenvalue of $\Gamma_{\theta,0}$, which is strictly positive as argued above. Because $g(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0) = 0$, it is sufficient to show that $\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ to prove the theorem. Using the triangle inequality, it follows that

$$\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0)\right)\right\| \leq \left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\| + \left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\|$$

$$\leq \left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\| \tag{A.18}$$

$$+ \left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right) - g\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right) - g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right) + g_n\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\right\| \tag{A.19}$$

$$+ \left\|g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\| \tag{A.20}$$

$$+ \left\|g_n\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\right\|, \tag{A.21}$$

where $g(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0) = 0$ was subtracted within the second norm (A.19). By the central limit theorem (Theorem 18.5.3 in Ibragimov and Linnik, 1971), the existence of the $(2+\delta)$ moments of z_t^m, $z_t^m G_t\left(\boldsymbol{\zeta}_0\right)$, and $z_t^m \partial G_t(\boldsymbol{\zeta})/\partial\boldsymbol{\zeta}$, respectively, and the boundedness of both $G_t\left(\boldsymbol{\zeta}_0\right)$ and $\partial G_t\left(\boldsymbol{\zeta}_0\right)/\partial\boldsymbol{\zeta}$ implies that the expression (A.21) is tight, and it holds that $\left\|g_n\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. The remaining terms (A.18), (A.19), and (A.20) can again be analysed separately. Starting with the first term, again using the triangle inequality, and changing the signs within the norm, term (A.18) can be bounded by

$$\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\| \leq \left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right) - g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - \Gamma_{\alpha,n}\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\| \tag{A.22}$$

$$+ \left\|\Gamma_{\alpha,n}\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right) - \Gamma_{\alpha,n}\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\| \tag{A.23}$$

$$+ \left\|\Gamma_{\alpha,n}\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\|.$$

Applying the Taylor series expansion of $g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)$ around $\boldsymbol{\alpha}_0$ in (A.22), and using the fact that $\Gamma_{\alpha,n}$ is Lipschitz in $\boldsymbol{\alpha}$ (because $\partial G_t\left(\boldsymbol{\zeta}_0\right)/\partial\boldsymbol{\zeta}$ is Lipschitz and σ_t is bounded as it is invertible to an ARCH model by Assumption 2.2), term (A.22) is negligible in probability with respect to $\|\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\|^2 = \mathcal{O}_p(m/n) = o_p\left(n^{-1/2}\right)$ by Theorem 4.1 and Assumption 4.7. Similarly, for equation (A.23), we use that $\Gamma_{\alpha,n}\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right)$ is Lipschitz in $\boldsymbol{\theta}_0(\tau)$, which has bounded parameter space. Thus, (A.22) reduces to

$$\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \boldsymbol{\alpha}_0\right) - g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\| \tag{A.24}$$

© 2019 Royal Economic Society.

$$\leq \mathcal{O}_p\left(\|\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\|^2\right) + \mathcal{O}_p\left(\|\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\| \left\|\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau)\right\|\right) + \left\|\Gamma_{\alpha,m,0}\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\|$$

$$= \left\|\Gamma_{\alpha,m,0}\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\|(1 + o_p(1)) = \mathcal{O}_p\left(n^{-1/2}\right),$$

where the the last term follows from (1) the fact that the elementwise $\boldsymbol{\Gamma}_{\alpha,n} \to \boldsymbol{\Gamma}_{\alpha,0}$ in probability by law of large numbers (the respective moments exist by Assumption 4.5) and Slutsky's lemma and (2) equation (A.16), which applies because of Assumption 9.

In a next step, we analyse the remaining terms (A.19) and (A.20), for which we have to check the conditions of Lemma 4.2 in Chen (2008). For this, let

$$\mathfrak{m}_\tau(z_t, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \begin{bmatrix} z_t(\boldsymbol{\alpha})G_t(\boldsymbol{\zeta}) \\ z_t(\boldsymbol{\alpha})(1 - G_t(\boldsymbol{\zeta})) \\ \boldsymbol{\theta}^{\Delta T} z_t(\boldsymbol{\alpha})\dfrac{\partial G_t(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \end{bmatrix}\left(\tau - \mathbb{1}\left\{u_t \leq [\boldsymbol{\theta}^{\mathrm{I}}, \boldsymbol{\theta}^{\mathrm{II}}]^T z_t(\boldsymbol{\alpha}, \boldsymbol{\zeta})\right\}\right) \tag{A.25}$$

so that $g_n(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \mathbb{P}_n \mathfrak{m}_\tau(z_t(\boldsymbol{\alpha}), \boldsymbol{\theta}, \boldsymbol{\alpha})$ and $g(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \mathbb{E}\mathfrak{m}_\tau(z_t, \boldsymbol{\theta}, \boldsymbol{\alpha})$. Then, if z_t is stationary, which is true by Assumption 2.2 and 4.3, and has β-mixing decay rate as in Assumption 4.5 (see, e.g., Carrasco and Chen (2002), Meitz and Saikkonen (2008)), and $\boldsymbol{\Theta}_2^\tau$ is a compact subset of $\mathbb{R}^{2(p+q+1)+2}$ and $\boldsymbol{\Theta}_1$ one of $\mathbb{R}^{2(m+1)} \times \mathbb{R} \times \mathbb{R}_+$, we have to verify for each j^{th} component $\mathfrak{m}_{\tau,j}$, $j \in \mathcal{I}_{2(p+q+1)+2}$, of $\mathfrak{m}_\tau$ that

$$\left(\mathbb{E}\left[\sup_{(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \in \mathfrak{U}_\delta((\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0)} \left|\mathfrak{m}_{\tau,j}(z_t, \boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - \mathfrak{m}_{\tau,j}(z_t, \boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0)\right|^r\right]\right)^{\frac{1}{r}} \leq K_j \delta^{s_j},$$

where $\boldsymbol{\alpha}' = [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^T$ and $\boldsymbol{\theta}'' = [\boldsymbol{\theta}^{\mathrm{I}}, \boldsymbol{\theta}^{\mathrm{II}}]^T$, for some s_j that is bounded by the degree of smoothness of $G_t(\boldsymbol{\zeta})$, for some constant $K_j > 0$, and for $r = 2 + \delta$ satisfying the restriction in Assumption 4.5 to claim that:

$$\sup_{(\boldsymbol{\theta}, \boldsymbol{\alpha}) \in \mathfrak{U}_\delta(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0)} \|g(\boldsymbol{\theta}, \boldsymbol{\alpha}) - g(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0) - g_n(\boldsymbol{\theta}, \boldsymbol{\alpha}) + g_n(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0)\| = o_p\left(n^{-\frac{1}{2}}\right)$$

with $\mathfrak{U}_\delta := \left\{(\boldsymbol{\theta}, \boldsymbol{\alpha}) \in \boldsymbol{\Theta}_2^\tau \times \boldsymbol{\Theta}_1 : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0(\tau)\| < \delta, \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| < \delta\right\}$ by Lemma 4.2 in Chen (2008). As discussed in Chen (2008), $\mathfrak{m}_{\tau,j}$ needs to be a member of a function class with covering numbers satisfying condition $\int_0^\infty \sqrt{\log N(\epsilon^{1/s_j}, \mathcal{H}, \|\bullet\|_{\mathcal{H}})}d\epsilon < \infty$, where the degree of smoothness satisfies $d = 1 \geq 2/(2s_j)$, with $s_j = 1$ in our case. Alternatively, we can make use of the class of monotone functions, which is sufficient for the former condition; for details, see Chen (2008). Consequently, we either need continuity of $\frac{\partial G_t(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}$ (Theorem 2.7.1 in van der Vaart and Wellner, 1996) or monotonicity (Theorem 2.7.5 in van der Vaart and Wellner, 1996) of $\frac{\partial G_t(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}$ with respect to ξ_t, and Assumption 2.1 ensures that the transition function belongs to one of these classes.

The uniform boundedness relative to the L^r-norm of the distance between values of (A.25) evaluated at any two parameter values within a neighbourhood of the true parameters can be shown as follows.

By definition,

$$z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') = \begin{cases} G(\boldsymbol{\zeta}'')\left[z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}')\right]_j & \text{if} \quad 0 < j \leq p + q + 1, \\ 1 - G(\boldsymbol{\zeta}''))\left[z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}')\right]_{j-(p+q+1)} & \text{if} \quad p + q + 1 < j \leq 2(p + q + 1), \\ \boldsymbol{\theta}''^{\Delta T} z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}')\dfrac{\partial G(\zeta'', \eta'')}{\partial \zeta} & \text{if} \quad j = 2(p + q + 1) + 1, \\ \boldsymbol{\theta}''^{\Delta T} z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}')\dfrac{\partial G(\zeta'', \eta'')}{\partial \eta} & \text{if} \quad j = 2(p + q + 1) + 2. \end{cases}.$$

In addition, it holds that

$$\left|\mathfrak{m}_{\tau,j}(z_t, \boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - \mathfrak{m}_{\tau,j}(z_t, \boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0)\right|^r$$

© 2019 Royal Economic Society.

$$\leq \tau \left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \right|^r \tag{A.26}$$

$$+ \left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}''^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'') \right\} - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0) \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0) \right\} \right|^r. \tag{A.27}$$

We start by expanding equation (A.26) and bounding each term individually:

$$\tau \mathbb{E} \left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0) \right|^r \leq \tau \mathbb{E} \left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \right|^r \tag{A.28}$$

$$+\tau \mathbb{E} \left| z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}_0', \boldsymbol{\zeta}') \right|^r \tag{A.29}$$

$$+\tau \mathbb{E} \left| z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}_0', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}') \right|^r \tag{A.30}$$

$$+\tau \mathbb{E} \left| z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0) \right|^r. \tag{A.31}$$

Given that $\left\| \boldsymbol{\theta}'' - \boldsymbol{\theta}_0 \right\| < \delta$, a bound for the first term (A.28) can be obtained by noting that we have finite $(2 + \delta)$ moments of u_t (for all t) and finite derivatives of the transition function G by Assumptions 2.1 and 4.6, such that $\tau^r \mathbb{E} \left| (\boldsymbol{\theta}''^{\Delta} - \boldsymbol{\theta}_0^{\Delta})^T z_{t,j}(\boldsymbol{\alpha}', \boldsymbol{\zeta}') \frac{\partial G(\zeta'', \eta'')}{\partial \bullet} \right|^r \leq K_{1,1,j} \delta^r$. Because $z_{t,j}$ is linear in u_t, the same bound also applies to the expectation of the supremum of the absolute value. For the second term (A.29), if $\|\boldsymbol{\alpha}' - \boldsymbol{\alpha}_0\| < \delta$, a bound denoted by $K_{1,2,j}\delta^r$ follows immediately from the linearity of $z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}')$ with respect to $\boldsymbol{\alpha}$, finite second moments of u_t, and Assumption 2.1. For the remaining terms (A.30) and (A.31), with the transition parameters $\left\| \boldsymbol{\zeta}'' - \boldsymbol{\zeta}_0 \right\| < \delta$ and $\left\| \boldsymbol{\zeta}' - \boldsymbol{\zeta}_0 \right\| < \delta$, the differentiability of G and the Lipschitz continuity and boundedness of $\frac{\partial G}{\partial \zeta}$ (both stated in Assumption 2.1), along with previous arguments, imply that their respective suprema and the expectations thereof can also be bounded by $K_{1,3,j}\delta^r$ and $K_{1,4,j}\delta^r$, respectively.

Putting all these terms together, we get the bound with constant $K_{1,j} = 4 \sup_l K_{1,l,j}$

$$\tau^r \mathbb{E} \sup_{(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \in \mathfrak{U}_\delta((\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0)} \left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'' \boldsymbol{\alpha}', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0) \right|^r \leq K_{1,j}\delta^r. \tag{A.32}$$

Returning to the original inequality in (A.26) and (A.27), for the second term (A.27), we note that

$$\left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}''^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'') \right\} - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0) \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0) \right\} \right|^r$$

$$\leq \left| z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') \left(\mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}''^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'') \right\} - \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0) \right\} \right) \right|^r \tag{A.33}$$

$$+ \left| \left(z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}') - z_{t,j}(\boldsymbol{\theta}_0'', \boldsymbol{\zeta}_0, \boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0) \right) \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0) \right\} \right|^r. \tag{A.34}$$

Although the bound of the expectation of term (A.34) follows from (A.32), we need to take care of equation (A.33). Let $z_{t,j}' = z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}')$. Taking expectations in neighbourhoods of the true parameter, it follows that

$$\mathbb{E} \left| z_{t,j}' \left(\mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}''^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'') \right\} - \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}_0', \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0) \right\} \right) \right|^r$$

$$\leq \left\{ \mathbb{E} \left| z_{t,j}' \right|^r \left(\mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}''^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'') \right\} - \mathbb{1}\left\{ u_t \leq \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'') \right\} \right) \right.$$

© 2019 Royal Economic Society.

$$+\mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'')\right\} - \mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}', \boldsymbol{\zeta}'')\right\}\right)$$

$$+2\mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}', \boldsymbol{\zeta}'')\right\} - \mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}', \boldsymbol{\zeta}_0)\right\}\right)$$

$$+2\mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}', \boldsymbol{\zeta}_0)\right\} - \mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0)\right\}\right)\Bigg\}$$

$$\le \mathbb{E}\Bigg\{\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\boldsymbol{\phi}}_1\right) z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'')^T\left(\boldsymbol{\theta}'' - \boldsymbol{\theta}''_0\right) \tag{A.35}$$

$$+\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\boldsymbol{\phi}}_2\right)\left(G(\boldsymbol{\zeta}'')\sum_{j=1}^{q}\theta^{\mathrm{I}}_{p+j+1}L^j z_t^m(\boldsymbol{\zeta}') + (1 - G(\boldsymbol{\zeta}''))\sum_{j=1}^{q}\theta^{\mathrm{II}}_{p+j+1}L^j z_t^m(\boldsymbol{\zeta}')\right)^T\left(\boldsymbol{\alpha}' - \boldsymbol{\alpha}'_0\right) \tag{A.36}$$

$$+2\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\boldsymbol{\phi}}_3\right)\left(z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}', \boldsymbol{\zeta}'')^T\boldsymbol{\theta}_0''^{\Delta}\frac{\partial G(\boldsymbol{\zeta}_0)}{\partial\boldsymbol{\zeta}}\right)^T\left(\boldsymbol{\zeta}'' - \boldsymbol{\zeta}_0\right) \tag{A.37}$$

$$+2\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\boldsymbol{\phi}}_4\right)\Bigg(G(\boldsymbol{\zeta}'')\sum_{j=1}^{q}\theta^{\mathrm{I}}_{p+j+1}L^j z_t^m(\boldsymbol{\zeta}')^T\boldsymbol{\alpha}^{\Delta}\tfrac{\partial L^j G_t(\zeta_0)}{\partial\zeta}$$

$$+(1 - G(\boldsymbol{\zeta}''))\sum_{j=1}^{q}\theta^{\mathrm{II}}_{p+j+1}L^j z_t^m(\boldsymbol{\zeta}')^T\boldsymbol{\alpha}^{\Delta}\tfrac{\partial L^j G_t(\zeta_0)}{\partial\zeta}\Bigg)^T\left(\boldsymbol{\zeta}' - \boldsymbol{\zeta}_0\right)\Bigg\}, \tag{A.38}$$

where the first inequality follows from the triangle inequality and the fact that $\omega \mapsto \mathbb{1}\{u_t \le \omega\}$ is monotone in ω, which is in turn linear in $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. In addition, $\mathbb{1} \circ G$ is monotone in the first parameter (location) of $\boldsymbol{\zeta}$, namely ζ, and piecewise monotone (increasing over half of the domain and decreasing over the other half) in its second parameter η (scale). For the second inequality, we apply the law of iterated expectations and the mean value theorem, for which we require the density $f_{u_t|\mathcal{F}_{t-1}}$ to exist and to be bounded (Assumption 4.4). The variables $\tilde{\boldsymbol{\phi}}_j$ for $j \in \mathcal{I}_{1,4}$ refer to the elements of small neighbourhoods of the respective parameters at which we applied the mean value theorem. Although the existence and boundedness of the density and the finiteness of $(2+\delta)$ moments of $z_{t,j}$ (Assumption 4.5) are sufficient for the terms (A.35) and (A.36) not to diverge, the final two terms (A.37) and (A.38) additionally require the bound on $\frac{\partial G_t(\zeta)}{\partial\zeta}$ (Assumption 2.1). Then, for any $(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta})$ in a neighbourhood of their true counterparts $\mathfrak{U}_\delta = \mathfrak{U}_\delta(\boldsymbol{\theta}''_0, \boldsymbol{\zeta}_0, \boldsymbol{\alpha}'_0, \boldsymbol{\zeta}_0)$ (i.e., where $\left\|\boldsymbol{\theta}'' - \boldsymbol{\theta}_0\right\| < \delta$, $\left\|\boldsymbol{\zeta}'' - \boldsymbol{\zeta}_0\right\| < \delta$, $\|\boldsymbol{\alpha}' - \boldsymbol{\alpha}_0\| < \delta$, and $\left\|\boldsymbol{\zeta}' - \boldsymbol{\zeta}_0\right\| < \delta$), there exists a $K_{2,j} > 0$, such that the second term in equation (A.27) can be bounded by

$$\mathbb{E}\sup_{(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}')\in\mathfrak{U}_\delta}\left|z_{t,j}(\boldsymbol{\theta}'', \boldsymbol{\zeta}'', \boldsymbol{\alpha}', \boldsymbol{\zeta}')\mathbb{1}\left\{u_t \le \boldsymbol{\theta}''^T z_t(\boldsymbol{\alpha}', \boldsymbol{\zeta}', \boldsymbol{\zeta}'')\right\} - z_{t,j}(\boldsymbol{\theta}''_0, \boldsymbol{\zeta}_0, \boldsymbol{\alpha}'_0, \boldsymbol{\zeta}_0)\mathbb{1}\left\{u_t \le \boldsymbol{\theta}_0^T z_t(\boldsymbol{\alpha}'_0, \boldsymbol{\zeta}_0, \boldsymbol{\zeta}_0)\right\}\right|^r,$$

which can then be bounded by $K_{2,j}\delta^r$. Subsequently, by Lemma 4.2 in Chen (2008), the following expression holds:

$$\sup_{(\boldsymbol{\theta}, \boldsymbol{\alpha})\in\mathfrak{U}_\delta(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0)}\left\|g\left(\boldsymbol{\theta}, \boldsymbol{\alpha}\right) - g\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right) - g_n\left(\boldsymbol{\theta}, \boldsymbol{\alpha}\right) + g_n\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\right\| = o_p\left(n^{-1/2}\right).$$

Thus, (A.19) reduces to $\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right) - g\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right) + g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right) - g_n\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right)\right\| = o_p\left(n^{-\frac{1}{2}}\right)$, whereas for (A.20) we have by definition $\left\|g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau), \widehat{\boldsymbol{\alpha}}_n\right)\right\| = o_p\left(n^{-\frac{1}{2}}\right)$, and we immediately get from

© 2019 Royal Economic Society.

equation (A.17):

$$\lambda_{\min}\left(\Gamma_{\theta,0}\right)\left\|\left(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau)\right)\right\| \leq \left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right),$$

which completes the proof. □

COROLLARY 1.1. *Under Assumptions 2.1–4.5, the following linearisation holds as $n \to \infty$:*

$$g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g_n\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-\Gamma_{\theta,0}(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau))-\Gamma_{\boldsymbol{\alpha},0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)=o_p\left(n^{-\frac{1}{2}}\right). \tag{A.39}$$

Proof. By adding and subtracting, equation (A.39) can be rewritten as:

$$\begin{aligned}
&g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g_n\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-\Gamma_{\theta,0}(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau))-\Gamma_{\boldsymbol{\alpha},0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\\
&=g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g_n\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-\Gamma_{\theta,0}(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau))-\Gamma_{\boldsymbol{\alpha},0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\\
&+g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)+g\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-g\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)\\
&+g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)-g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\\
&+\Gamma_{\boldsymbol{\alpha},n}\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)-\Gamma_{\boldsymbol{\alpha},n}\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)
\end{aligned}$$

Again by taking norms, rearranging the terms on the right-hand side, and using the triangle inequality, the following bound is obtained as $n \to \infty$:

$$\begin{aligned}
&\left\|g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g_n\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-\Gamma_{\theta,0}(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau))-\Gamma_{\boldsymbol{\alpha},0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\|\\
&\leq\left\|g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g_n\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-\left(g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)\right)\right\|\\
&+\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)-g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\right)-\Gamma_{\boldsymbol{\alpha},n}\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\|\\
&+\left\|g\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)-g\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)-\Gamma_{\theta,0}\left(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau)\right)\right\|\\
&+\left\|\Gamma_{\boldsymbol{\alpha},n}\left(\widehat{\boldsymbol{\theta}}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)-\Gamma_{\boldsymbol{\alpha},0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right\|=o_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where we use stochastic equicontinuity verified in the previous lemma for the first term, and we reason along the lines of (A.22), (A.23), and (A.24), using Lipschitz continuity of $\Gamma_{\alpha,0}$ and $\Gamma_{\theta,0}$, as well as the law of large numbers for $\Gamma_{\alpha,n}$ and $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\theta}}_n(\tau)$ for the remaining terms. □

Proof of Theorem 4.4: The first-order condition $g_n\left(\boldsymbol{\theta}(\tau),\widehat{\boldsymbol{\alpha}}_n\right)=0$ is solved by $\widehat{\boldsymbol{\theta}}_n(\tau)$ so that

$$0=g_n\left(\widehat{\boldsymbol{\theta}}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)=g_n\left(\boldsymbol{\theta}_0,\boldsymbol{\alpha}_0\right)+\Gamma_{\theta,0}(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau))+\Gamma_{\alpha,0}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)+o_p\left(n^{-\frac{1}{2}}\right),$$

using the linearisation from Corollary 8.1. Because $\Gamma_{\theta,0}$ has full rank by Assumption 4.9, by premultiplying $\sqrt{n}$, an asymptotic representation of the second-stage estimator is obtained for $n \to \infty$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n(\tau)-\boldsymbol{\theta}_0(\tau)\right)=-\boldsymbol{\Gamma}_{\theta,0}^{-1}\left[\sqrt{n}g_n\left(\boldsymbol{\theta}_0(\tau),\boldsymbol{\alpha}_0\right)\right)+\boldsymbol{\Gamma}_{\alpha,0}\sqrt{n}\left(\widehat{\boldsymbol{\alpha}}_n-\boldsymbol{\alpha}_0\right)\right]+o_p(1).$$

© 2019 Royal Economic Society.

Finally, we apply the α-mixing central limit theorem (Theorem 18.5.3 in Ibragimov and Linnik, 1971), for which the $(2+\delta)$ moments of the data have to exist and mixing coefficients have to satisfy $\beta_s \to 0$ and $\sum_{s=1}^{\infty} \beta_s^{\delta/(2+\delta)} < +\infty$. These conditions are guaranteed by Assumption 4.5. After stacking the two summands in the last equation, we can therefore write as $n \to \infty$

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n(\tau) - \boldsymbol{\theta}_0(\tau)\right) = \Gamma_{\theta,0}^{-1}\left[\boldsymbol{I}_{2(p+q+1)}, \Gamma_{\alpha,m,0}\right]\sqrt{n}\begin{bmatrix} g_n\left(\boldsymbol{\theta}_0(\tau), \boldsymbol{\alpha}_0\right) \\ \widehat{\boldsymbol{\alpha}}_n^{\mathrm{I}} - \boldsymbol{\alpha}_0^{\mathrm{I}} \\ \widehat{\boldsymbol{\alpha}}_n^{\mathrm{II}} - \boldsymbol{\alpha}_0^{\mathrm{II}} \\ \widehat{\boldsymbol{\zeta}}_n - \boldsymbol{\zeta}_0 \end{bmatrix} + o_P(1)$$

$$\approx \frac{1}{\sqrt{n}}\Gamma_{\theta,0}^{-1}\left[\boldsymbol{I}_{2(p+q+1)}, \frac{\Gamma_{\alpha,m,0}\boldsymbol{D}_n^{-1}}{\sum_{k=1}^{K} s_k q_k^2}\right]\sum_{t=m+1}^{T}\begin{bmatrix} \begin{bmatrix} z_t(\boldsymbol{\alpha}_0)G_t\left(\boldsymbol{\zeta}_0\right) \\ z_t(\boldsymbol{\alpha}_0)(1-G_t\left(\boldsymbol{\zeta}_0\right)) \\ \boldsymbol{\theta}_0^{\Delta T} z_t(\boldsymbol{\alpha}_0)\dfrac{\partial G_t\left(\boldsymbol{\zeta}_0\right)}{\partial \boldsymbol{\zeta}} \end{bmatrix}\left(\mathbb{1}\left\{u_t \le F_{u_t|\mathcal{F}_{t-1}}^{-1}(\tau)\right\} - \tau\right) \\ \begin{bmatrix} G_t\left(\boldsymbol{\zeta}_0\right) z_t^m \\ 1-G_t\left(\boldsymbol{\zeta}_0\right)) z_t^m \\ \boldsymbol{\alpha}_0^{\Delta} z_t^m \dfrac{\partial G_t\left(\boldsymbol{\zeta}_0\right)}{\partial \boldsymbol{\zeta}} \end{bmatrix}\sum_{k=1}^{K} q_k\left(\mathbb{1}\left\{u_t \le F_{u_t|\mathcal{F}_{t-1}}^{-1}(\tau_k)\right\} - \tau_k\right) \end{bmatrix}$$

$$\mathcal{N}\left(0, \lim_{n\to\infty} \boldsymbol{\Gamma}_{\theta,0}^{-1}\mathbb{E}\left[\boldsymbol{M}_t \boldsymbol{\Xi}^{\tau} \boldsymbol{M}_t^T\right]\boldsymbol{\Gamma}_{\theta,0}^{-1}\right),$$

where we use independence of the innovations ε_t and z_t^m and $z_t(\boldsymbol{\alpha}_0)$, respectively. The matrices $\boldsymbol{\Xi}^{\tau}$, $\boldsymbol{M}_t$, and $\boldsymbol{\Gamma}_{\theta,0}$ are defined in Assumption 4.9, which also postulates the existence of the asymptotic variance matrix. □

Proof of Lemma 5.1: We show this by induction and omit the parameters of sigma for readability. By definition, we have

$$\sigma_t = \bar{\mathcal{P}}(L^0)\sigma_{t-1} + \bar{\mathcal{P}}(L^0)|u_{t-1}|.$$

Lagging this equation and plugging in for σ_{t-1}, we get by construction of $\bar{\mathcal{P}}(L^1)$:

$$\sigma_t = \bar{\mathcal{P}}(L^0)\left[L\bar{\mathcal{P}}(L^0)\sigma_{t-2} + L\bar{\mathcal{P}}(L^0)|u_{t-2}|\right] + \bar{\mathcal{P}}(L^0)|u_{t-1}|$$

$$= \bar{\mathcal{P}}(L^1)\sigma_{t-2} + \bar{\mathcal{P}}(L^1)|u_{t-2}| + \bar{\mathcal{P}}(L^0)|u_{t-1}|.$$

For $m \mapsto m+1$, we note that $\sigma_{t-j} = L^j\bar{\mathcal{P}}(L^0)\sigma_{t-j-1} + L^j\bar{\mathcal{P}}(L^0)|u_{t-j-1}|$. Thus, substituting for σ_{t-m-1} in equation (5.35), we obtain

$$\sigma_t = \bar{\mathcal{P}}(L^m)\left[L^{m+1}\bar{\mathcal{P}}(L^0)\sigma_{t-m} + L^{m+1}\bar{\mathcal{P}}(L^0)|u_{t-m}|\right] + \sum_{k=0}^{m}\bar{\mathcal{P}}(L^k)|u_{t-k-1}|,$$

and after using $L^{m+1}\bar{\mathcal{P}}(L^0) = L\bar{\mathcal{P}}(L^m)$ we get

$$\sigma_t = \bar{\mathcal{P}}(L^{m+1})\sigma_{t-(m+1)-1} + \bar{\mathcal{P}}(L^{m+1})|u_{t-(m+1)-1}| + \sum_{k=0}^{m}\bar{\mathcal{P}}(L^k)|u_{t-k-1}|$$

$$= \bar{\mathcal{P}}(L^{m+1})\sigma_{t-(m+1)-1} + \sum_{k=0}^{m+1}\bar{\mathcal{P}}(L^k)|u_{t-k-1}|,$$

where the first term is vanishing because of the parameter restrictions in Assumption 2.2, and $G_t(\boldsymbol{\zeta}) \in [0, 1]$ by Assumption 2.1, which completes the proof. □

© 2019 Royal Economic Society.

A2. ALGORITHM FOR ESTIMATION

Algorithm 1 Two-stage estimation procedure

$(l^1_{\min}, \widehat{\alpha}^{\mathrm{I}}_n, \widehat{\alpha}^{\mathrm{II}}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (\infty, 0, 0, 0, 0)$
for all $\eta \in \{\eta_1, \ldots, \eta_{k_\eta}\}$ **do**
  Define $G(\xi_t, \zeta, \eta)$, using $\xi_t \leftarrow \xi(\mathbf{z}_t)$ for given scale η as a function of ζ
  Estimate $\widehat{\alpha}^{\mathrm{I}}_{n,k_\eta}, \widehat{\alpha}^{\mathrm{II}}_{n,k_\eta}, \widehat{q}_{k_\eta}, \widehat{\zeta}_{k_\eta}$ and obtain loss $l^1_{k_\eta}$ according to (3.9)
  ∼∼∼by (smoothed) composite quantile regression
  if $l^1_{k_\eta} \le l^1_{\min}$ **then**
    $(l^1_{\min}, \widehat{\alpha}^{\mathrm{I}}_n, \widehat{\alpha}^{\mathrm{I}}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (l^1_{k_\eta}, \widehat{\alpha}^{\mathrm{I}}_{n,k_\eta}, \widehat{\alpha}^{\mathrm{II}}_{n,k_\eta}, \widehat{\zeta}_{k_\eta}, \eta)$
  end if
end for
Calculate $\sigma_t(\widehat{\alpha}_n)$ according to equation (3.10) using $\widehat{\alpha}_n = (\widehat{\alpha}^{\mathrm{I}T}_n, \widehat{\alpha}^{\mathrm{II}T}_n, \widehat{\zeta}_n, \widehat{\eta}_n)^T$
Construct $\mathbf{z}_t(\widehat{\alpha}_n) = (\sigma_{t-1}(\widehat{\alpha}_n), \ldots, \sigma_{t-p}(\widehat{\alpha}_n), |u_{t-1}|, \ldots, |u_{t-q}|)^T$
$(l^2_{\min}, \widehat{\theta}^{\mathrm{I}}_n, \widehat{\theta}^{\mathrm{II}}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (\infty, 0, 0, 0, 0)$
for all $(\zeta, \eta) \in \{\zeta_1, \ldots \zeta_{k_\zeta}\} \times \{\eta_1, \ldots, \eta_{k_\eta}\} \cap \mathfrak{Z}$ **do**
  Calculate $G(\xi_t, \zeta, \eta)$ using $\xi_t \leftarrow \xi(\mathbf{z}_t)$ for given location ζ and scale η
  Estimate $\widehat{\theta}^{\mathrm{I}}_{n,k_{\zeta,\eta}}, \widehat{\theta}^{\mathrm{II}}_{n,k_{\zeta,\eta}}$ and obtain loss $l^2_{k_{\zeta,\eta}}$ according to (3.11)
  ∼∼∼by linear (inequality constrained) quantile regression
  if $l^2_{k_{\zeta,\eta}} \le l^2_{\min}$ **then**
    $(l^2_{\min}, \widehat{\theta}^{\mathrm{I}}_n, \widehat{\theta}^{\mathrm{II}}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (l^2_{k_{\zeta,\eta}}, \widehat{\theta}^{\mathrm{I}}_{n,k_{\zeta,\eta}}, \widehat{\theta}^{\mathrm{II}}_{n,k_{\zeta,\eta}}, \zeta, \eta)$
  end if
end for
Set the final estimate to $\widehat{\theta}_n = (\widehat{\theta}^{\mathrm{I}T}_n, \widehat{\theta}^{\mathrm{II}T}_n, \widehat{\zeta}_n, \widehat{\eta}_n)^T$.

A3. SUFFICIENT CONDITIONS FOR ABSOLUTE REGULARITY

Let us recall existing results with regard to the (nonlinear) GARCH processes. Meitz and Saikkonen (2008) derive Theorems 1 and 2, stating sufficient conditions for the geometric ergodicity of Markov processes following various nonlinear GARCH models and implying their stationarity and absolute regularity (see Meitz and Saikkonen, 2008, Section 2.2 and Theorem 3). We limit ourselves for simplicity to the models of order 1 here, with the transition variable being the lagged dependent variable.

First, various regularity assumptions (Meitz and Saikkonen, 2008, Assumption 2) have to hold so that the volatility process defined by (2.3) or a GARCH model has some basic properties, such as irreducibility and aperiodicity. These regularity assumptions are, however, satisfied in our model and GARCH models because of the imposed assumptions (i.e., Assumption 4.3) and the model definitions, implying the boundedness of the volatility function on compact subsets of the support, its positivity on the compact subsets of the support, and its monotonicity and differentiability in the transition variable (Assumption 2.1). More importantly, Theorems 1–3 of Meitz and Saikkonen (2008) require that the volatility is bounded by a linear function of the past volatility and its slope has a finite expectation smaller than 1. Denoting the volatility in (2.1) by σ_t for the sake of conciseness, and recalling that all parameter values are assumed to be non-negative, the standard GARCH(1,1) volatility process in model (2.1) is, for example, bounded by

$$\sigma_t^2 = \beta_0 + \beta_1 \sigma_{t-1}^2 + \gamma_1 u_{t-1}^2 \le \beta_0 + [\beta_1 + \gamma_1 \varepsilon_{t-1}^2]\sigma_{t-1}^2.$$

Assuming that $E(\beta_1 + \gamma_1 \varepsilon_{t-1}^2) < 1$ holds (i.e., $\beta_1 + \gamma_1 < 1$ if the variance of ε_{t-1} is normalised to 1), Meitz and Saikkonen (2008, Theorems 1–3) with $V(\sigma) = 1 + \sigma$ imply that the GARCH(1,1) volatility process is V-geometrically ergodic and the GARCH(1,1) process itself is geometrically ergodic and β-mixing.

© 2019 Royal Economic Society.

Under the abovementioned regularity assumptions, Theorems 1–3 of Meitz and Saikkonen (2008) allow us to directly obtain the same results also for the single-regime model in (2.1) and (2.2), where for the first-order model,

$$\sigma_t = \beta_0 + \beta_1\sigma_{t-1} + \gamma_1|u_{t-1}| \leq \beta_0 + [\beta_1 + \gamma_1|\varepsilon_{t-1}|]\sigma_{t-1}$$

and it has to hold that $E(\beta_1 + \gamma_1|\varepsilon_{t-1}|) < 1$ and $\beta_1 + \gamma_1 < 1$ if $E|\varepsilon_{t-1}|$ is normalised to 1. Finally, a similar condition can be found for the general model (2.8) (again of the first order), where

$$\begin{aligned}\sigma_t &= G_t(\zeta)[\beta_0^I + \beta_1^I\sigma_{t-1} + \gamma_1^I|u_{t-1}|] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II}\sigma_{t-1} + \gamma_1^{II}|u_{t-1}|]\\ &\leq \max\{\beta_0^I, \beta_0^{II}\} + [\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\}|\varepsilon_{t-1}|]\sigma_{t-1}\end{aligned}$$

because $G_t(\zeta) \in [0, 1]$. To apply Meitz and Saikkonen (2008, Theorems 1–3), it has to hold that $E([\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\}|\varepsilon_{t-1}|]) < 1$ (i.e., $\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\} < 1$ if $E|\varepsilon_{t-1}| = 1$).

The theorems above can be also extended to model (2.3), but we have to formulate the volatility as a multivariate process as the volatility σ_t depends on volatilities σ_t^I and σ_t^{II} of each regime:

$$\begin{aligned}\sigma_t^I &= \beta_0^I + \beta_1^I\sigma_{t-1}^I + \gamma_1^I|u_{t-1}|\\ \sigma_t^{II} &= \beta_0^{II} + \beta_1^{II}\sigma_{t-1}^{II} + \gamma_1^{II}|u_{t-1}|\\ \sigma_t &= G_t(\zeta)[\beta_0^I + \beta_1^I\sigma_{t-1}^I + \gamma_1^I|u_{t-1}|] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II}\sigma_{t-1}^{II} + \gamma_1^{II}|u_{t-1}|].\end{aligned}$$

Defining $V(\sigma^I, \sigma^{II}, \sigma) = 1 + \sigma$, the proof of Meitz and Saikkonen (2008, Theorem 1) applies to $(\sigma_t^I, \sigma_t^{II}, \sigma_t)$ using $V(\sigma^I, \sigma^{II}, \sigma)$ because by definition (2.3),

$$\begin{aligned}\sigma_t &= G_t(\zeta)[\beta_0^I + \beta_1^I\sigma_{t-1}^I + \gamma_1^I|u_{t-1}|] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II}\sigma_{t-1}^{II} + \gamma_1^{II}|u_{t-1}|]\\ &= G_t(\zeta)[\beta_0^I + \beta_1^I\sigma_{t-1}^I + \gamma_1^I|\varepsilon_{t-1}|\sigma_{t-1}] + [1 - G_t(\zeta)][\beta_0^{II} + \beta_1^{II}\sigma_{t-1}^{II} + \gamma_1^{II}|\varepsilon_{t-1}|\sigma_{t-1}]\\ &\leq \max\{\beta_0^I, \beta_0^{II}\} + [\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\}|\varepsilon_{t-1}|]\sigma_{t-1}.\end{aligned}$$

Hence, if $E[\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\}|\varepsilon_{t-1}|] < 1$ (i.e., $\max\{\beta_1^I, \beta_1^{II}\} + \max\{\gamma_1^I, \gamma_2^{II}\} < 1$ after normalisation of $E|\varepsilon_{t-1}|$ to 1), $(\sigma_t^I, \sigma_t^{II}, \sigma_t)$ is V-geometrically ergodic, and following Meitz and Saikkonen (2008, Theorems 2–3), we can obtain the geometric ergodicity and β-mixing properties for the ANST-GARCH process.

A4. OUT-OF-SAMPLE TESTS

The Kupiec (1995) test assumes that the Bernoulli process I_t is an i.i.d. sequence, and as such, the probability of having exactly x hits is given by the binomial probability mass function

$$f_{\text{bin}}(x, n, p) = \binom{n}{x} p^x(1-p)^{n-x}.$$

Using $x = Nh_{\text{oos}}$, the likelihood ratio is then given by

$$\mathcal{L}_N = -2\Big[\log f_{\text{bin}}(Nh_{\text{oos}}, N, \tau) - \log f_{\text{bin}}(Nh_{\text{oos}}, N, h_{\text{oos}})\Big]$$

$$= -2N\left[(1 - h_{\text{oos}})\log\left(\frac{1-\tau}{1-h_{\text{oos}}}\right) + h_{\text{oos}}\log\left(\frac{\tau}{h_{\text{oos}}}\right)\right],$$

which is asymptotically chi-square distributed with one degree of freedom, $\mathcal{L}_n \sim \chi_1^2$. The corresponding 10%, 5%, and 1% critical values are thus given by 2.7055, 3.8415, and 6.6349, respectively.

© 2019 Royal Economic Society.

For the second test, we use the observation that $I_{t+1} - \tau$ is a martingale difference sequence with zero mean and variance $N\tau(1-\tau)$. Hence, its cumulative sum converges to a normal distribution (Campbell, 2007; Xiao and Koenker, 2009). Formally, the test statistic can be defined as

$$Z_N := (N\tau(1-\tau))^{-\frac{1}{2}} \sum_{i=n-N+1}^{n} (I_{t+1} - \tau),$$

which asymptotically follows the standard normal distribution, and thus the 10%, 5%, and 1% critical values for the absolute value of Z_N are given by 1.6449, 1.9600, and 2.5758, respectively.

A5. SIMULATION RESULTS

In this section we summarise the results of a comprehensive simulation study. The study is divided into two main parts. First, the proposed asymmetric nonlinear smooth transition generalised autoregressive conditional quantile (shortly referred to here and in tables as GACQ instead of ANST-GACQ) procedure will be analysed with respect to different choices of the sample sizes and auxiliary parameters. Later, results are compared with the regime-switching GARCH model of Anderson et al. (1999) (abbreviated as GARCH) for various error distributions, including distributions contaminated by outliers.[17]

By default, the estimation is performed for time series of length $n = 1{,}000$, the number of simulations per experiment is $s = 100$, the composite quantile regression employs by default $k = 9$ quantiles for $\tau \in [0.05, 0.25] \cup [0.75, 0.95]$, the truncation parameter for the ARCH approximation is set to $m = \lceil \frac{5}{2} n^{\frac{1}{4}} \rceil$, and the grid size is $(k_\zeta, k_\eta) = (30, 30)$. The true global parameter vector for both processes is chosen to be $\boldsymbol{\theta}_0 = (\beta_0^{\mathrm{I}}, \beta_1^{\mathrm{I}}, \gamma_1^{\mathrm{I}}, \beta_0^{\mathrm{II}}, \beta_1^{\mathrm{II}}, \gamma_1^{\mathrm{II}})_0^T = (0.50, 0.15, 0.60, 0.25, 0.30, 0.15)^T$, and the location-scale parameter pair equals $\boldsymbol{\zeta}_0 = (\zeta, \eta)_0^T = (0.00, 0.2)^T$. Although β_0^{I} and β_0^{II} are determining only the unconditional variances of the respective regimes, we chose γ_1^{I} and γ_1^{II} in a way that is consistent with findings in the two-regime conditional heteroscedasticity literature (Gonzales-Rivera, 1998; Lubrano, 2001; Wago, 2004; Khemiri, 2011). Unfortunately, the findings on regime-specific parameter values for β_1^{I} and β_1^{II} are rather limited, and there is also no clear link to their single-regime counterparts. Thus, coming up with a sensible prior is somewhat *ad hoc*. We approached this by choosing their values in a way that generates both a higher- and a lower-persistence regime. Unreported simulations show that different DGPs work similarly well, although, perhaps unsurprisingly, numerical stability deteriorates as one of the regimes' processes becomes close to being integrated.

If not stated otherwise, we will assume the innovations to be standard normally distributed: $\varepsilon_t \sim N(0, 1)$. When running simulations using different innovation distributions, to ensure comparability, their variances will always be normalised to one. This implies that there is one high- and one low-variance regime with unconditional variances, defined by $\beta_0^{\mathrm{r}}/(1 - \beta_1^{\mathrm{r}} - \gamma_1^{\mathrm{r}})$ for r $\in \{\mathrm{I}, \mathrm{II}\}$, of 2 and 0.45, respectively. All of the presented results use a specification with the logistic function G_{logistic}. However, unreported simulations confirmed that the GACQ estimation is insensitive to the misspecification of the transition function (e.g., if the logistic transition function is used although the true underlying model follows the linear or threshold function). Finally, note that we have to restrict the grid for both location and scale. We introduce the data-driven criterion ensuring that location satisfies $\zeta \in [\underline{\zeta}, \bar{\zeta}]$ with unconditional sample quantiles $\underline{\zeta} = \widehat{F}_{u_t}^{-1}(0.1)$ and $\bar{\zeta} = \widehat{F}_{u_t}^{-1}(0.9)$. Similarly, the scale is restricted to $\eta \in [\underline{\eta}, \bar{\eta}(\underline{\zeta}, \bar{\zeta})]$ with fixed $\underline{\eta} = 0.1$ and $\bar{\eta}(\underline{\zeta}, \bar{\zeta}) = \left[\log(0.1^{-1} - 1)(0.5\bar{\zeta} - 0.5\underline{\zeta})\right]^{-1}$. The latter bound represents the inverse of the logistic function with respect to the scale evaluated at 0.1 and the location at the centre of the considered location grid.

To evaluate the procedures, we report the biases and root mean squared errors (RMSEs) of all estimates. As the focus of the quantile regression modelling is on the estimation of quantiles such as value at risk (VaR) rather than parameters, the performance is measured by the mean (absolute) prediction error averaged

[17] All experiments are conducted in Ox (Doornik, 2009), with extensions written in C for the computationally expensive parts.

© 2019 Royal Economic Society.

Table A.1. The bias and RMSE of GACQ for different sample sizes n.

	$n = 1{,}000$		$n = 2{,}000$		$n = 4{,}000$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1797	0.4120	0.1223	0.2489	0.0749	0.2123
β_1^{I}	− 0.0129	0.3559	0.0065	0.2779	− 0.0068	0.1179
γ_1^{I}	− 0.1196	0.3270	− 0.0653	0.2276	− 0.0228	0.1404
β_0^{II}	0.0319	0.1650	0.0337	0.1231	0.0236	0.1006
β_1^{II}	0.0263	0.2274	0.0272	0.1701	0.0187	0.1021
γ_1^{II}	0.0050	0.1310	0.0011	0.0984	0.0150	0.0891
ζ	0.3798	0.4923	0.3252	0.4703	0.4595	0.5632
$\zeta(\tau)$	0.1063	0.3730	0.1149	0.3138	0.0541	0.2547
η	0.0870	0.1341	0.0627	0.1170	0.0743	0.1229
$\eta(\tau)$	− 0.0622	0.0867	− 0.0327	0.0807	− 0.0353	0.0826
MPE	0.0082	0.0482	0.0027	0.0447	0.0032	0.0203
MAPE	0.1259	0.1310	0.0974	0.1004	0.0682	0.0700
MAFE	0.1129	0.1420	0.0848	0.1182	0.0530	0.0718
Coverage	0.0008	0.0013	0.0004	0.0007	0.0001	0.0003

over the sample, denoted as M(A)PE, absolute one-period-ahead out-of-sample forecast errors (MAFE), and by the coverage ratio, each of them referring to the estimated 5% VaR. Note that the coverage (ratio) is defined as the proportion of observations falling below the estimated VaR and should thus be close to $\tau = 0.05$ for the 5% VaR. It should be mentioned that although coverage, MPE, MAPE, and MAFE are reported in the Bias column for the purpose of a tidy exposition, their values represent the mean deviations from the value 0.0, which corresponds to the perfect fit of the model; for example, coverage value 0 represents the exactly correct coverage level 0.05, and MAPE value 0 would represent the exact fit. The RMSEs of these quantities additionally depict their corresponding Monte Carlo standard deviations. We will use these metrics to compare different estimators with each other as well as the impact of different features of the data-generating process on prediction and forecasting.

Our first simulation experiment considers the rate of convergence of the proposed estimator by studying its performance for different sample sizes $n = 1{,}000$, $n = 2{,}000$, and $n = 4{,}000$; Table A.1 summarises the results. It is comforting to report that the RMSEs of the parameter estimates decrease, at a rate that is consistent with our theoretical conclusions, as the sample size increases. With regard to the second-stage transition parameters, although they are estimated more precisely as the sample size increases as well, their RMSEs seem to go down slower than expected. This issue, which will be even more pronounced in the case of the standard GARCH model later, can be caused by the nonlinearity of the model with respect to the transition parameters, which makes them difficult to estimate from a numerical point of view. This is most pronounced in the first stage, where we use a smooth approximation of the quantile loss function, which is often very flat around the true parameters. Similarly, mean absolute prediction errors (MAPE) and mean absolute forecast errors (MAFE) are decreasing, and unsurprisingly, coverage ratios are accurate by construction of the quantile regression estimator.[18,19]

Moving on to study the influence of auxiliary parameters, to begin with we look at the number of lags in the ARCH(m) approximation by considering different multiples c of $n^{\frac{1}{4}}$, all satisfying the required order of

[18] The reason for reporting the coverage ratio is to allow for a direct comparison to the GARCH models in the second part of the simulation study, for which this property does not necessarily hold.

[19] To get an intuition for the magnitude of MAPE and MAFE, which are of order 10^{-1}, note that the 5th unconditional quantile of u_t is given by -2.4 for a typical series. The reported statistics refer to integrated absolute deviations of the predictions our model makes for the conditional quantile process, which is centred at this value.

© 2019 Royal Economic Society.

Table A.2. The bias and RMSE of GACQ as a function of the order of the first-stage ARCH(m) approximation with $m = \lceil cn^{1/4} \rceil$.

	$c = 2.0$		$c = 2.5$		$c = 3.0$		$c = 3.5$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.2070	0.4318	0.1727	0.3901	0.2041	0.4427	0.1837	0.4149
β_1^{I}	− 0.0031	0.3767	0.0173	0.3590	0.0037	0.3643	− 0.0286	0.3606
γ_1^{I}	− 0.1409	0.3316	− 0.1264	0.3310	− 0.1441	0.3568	− 0.1154	0.3110
β_0^{II}	0.0377	0.1666	0.0295	0.1714	0.0317	0.1714	0.0205	0.1777
β_1^{II}	0.0091	0.2168	0.0444	0.3932	0.0075	0.2022	0.0181	0.1984
γ_1^{II}	0.0047	0.1319	0.0085	0.1354	0.0070	0.1265	0.0123	0.1359
ζ	0.3779	0.4898	0.3939	0.4912	0.4618	0.5484	0.4818	0.5685
$\zeta(\tau)$	0.1380	0.3901	0.0955	0.3617	0.1198	0.3849	0.0855	0.3644
η	0.0573	0.1175	0.0412	0.0970	0.0740	0.1319	0.0635	0.1197
$\eta(\tau)$	− 0.0531	0.0864	− 0.0608	0.0882	− 0.0503	0.0860	− 0.0605	0.0888
MPE	0.0104	0.0504	0.0117	0.0507	0.0128	0.0516	0.0104	0.0516
MAPE	0.1273	0.1321	0.1248	0.1297	0.1270	0.1324	0.1270	0.1316
MAFE	0.1009	0.1307	0.1074	0.1409	0.1070	0.1502	0.1091	0.1473
Coverage	0.0008	0.0013	0.0004	0.0013	0.0008	0.0014	0.0007	0.0014

the ARCH(m) approximation rate. The results for $c = 2.0$, 2.5, 3.0, and 3.5 multiples of $n^{\frac{1}{4}}$, which translate to $m = 12$, 15, 17, and 20 for $n = 1{,}000$, are reported in Table A.2. We conclude that they are fairly constant with respect to c, and thus m, although there is a slight U-shaped pattern with the optimum in terms of MAPE around $c = 2.5$, which we will use for the remainder of the experiments and the following empirical application.

Furthermore, as discussed in Section 3, the model parameters are not identified at the median, and we thus suggested to estimate the first-stage composite quantile regression without using quantiles $\tau \in (0.5 - \delta/2, 0.5 + \delta/2)$ for some $\delta > 0$, as they could introduced extra noise into estimation. In Table A.3, results for different values of δ are collected, indicating that the precision of the estimates seems rather insensitive to a particular choice of δ. For the remainder of the simulations and the empirical application, we use $\delta = 0.25$, which corresponds to considering the first and fourth quartiles of the data to approximate conditional volatility.

In the second part of the simulation study, we compare smooth transition estimates of conditional quantiles (GACQ) with traditional smooth transition GARCH estimates. In particular, we consider the maximum likelihood estimators of the latter on the basis of both Normal (GARCH-N) and Student t_4 distribution (GARCH-t).

Naturally, the correctly specified GARCH maximum likelihood estimator yields the best parameter estimates for the case in which the data-generating process exhibits standard normally distributed innovations, $\varepsilon_t \sim N(0, 1)$; see Table A.4. Neglecting the parameters of the transition function, the GARCH-t model also performs relatively well in terms of RMSEs of the parameter estimates. However, the wrong assumption about the innovation distribution has serious negative consequences on the calculation of conditional quantiles and thus its prediction and forecast errors, as can be seen by looking at the GARCH-t estimates in Table A.4. The proposed GACQ model comes with the price of an efficiency loss in the parameter estimates, but the model outperforms both GARCH-N and GARCH-t in terms of predictions errors. Although in-sample prediction errors of our model are only slightly smaller than those of GARCH-N and GARCH-t, with respect to out-of-sample forecasting we see a substantial improvement with GACQ over the other two models. This could be partially explained by more precise estimates of the transition function, and in the case of forecasting errors, by directly modelling and fitting the quantiles of the innovation distribution.

© 2019 Royal Economic Society.

Table A.3. The bias and RMSE of GACQ if quantiles $\tau \in (0.5 - \delta/2, 0.5 + \delta/2)$ are not used in estimation.

	$\delta = 0.15$		$\delta = 0.20$		$\delta = 0.30$		$\delta = 0.50$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1748	0.4086	0.1897	0.4217	0.1816	0.4322	0.2021	0.4270
β_1^{I}	0.0106	0.3652	0.0133	0.3825	– 0.0035	0.3550	– 0.0078	0.3897
γ_1^{I}	– 0.1275	0.3335	– 0.1374	0.3503	– 0.1258	0.3465	– 0.1385	0.3509
β_0^{II}	0.0281	0.1610	0.0415	0.1655	0.0260	0.1611	0.0391	0.1589
β_1^{II}	0.0281	0.2159	0.0283	0.2271	0.0474	0.2331	0.0191	0.2095
γ_1^{II}	0.0040	0.1213	– 0.0024	0.1125	– 0.0030	0.1222	0.0004	0.1206
ζ	0.3641	0.4925	0.3341	0.4811	0.3311	0.4334	0.3259	0.4628
$\zeta(\tau)$	0.1073	0.3647	0.1211	0.3811	0.1017	0.3608	0.1348	0.3727
η	0.0639	0.1226	0.0717	0.1210	0.0748	0.1223	0.0606	0.1148
$\eta(\tau)$	– 0.0532	0.0863	– 0.0581	0.0849	– 0.0571	0.0853	– 0.0536	0.0868
MPE	0.0116	0.0496	0.0105	0.0518	0.0092	0.0484	0.0111	0.0498
MAPE	0.1228	0.1275	0.1262	0.1310	0.1260	0.1308	0.1263	0.1318
MAFE	0.1042	0.1392	0.1123	0.1408	0.1134	0.1426	0.0955	0.1208
Coverage	0.0008	0.0014	0.0008	0.0014	0.0009	0.0014	0.0006	0.0012

Table A.4. The bias and RMSE of the GACQ and GARCH estimators in the case of normally distributed errors.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1797	0.4120	0.0343	0.0986	– 0.1489	0.2789
β_1^{I}	– 0.0129	0.3559	– 0.1351	0.1410	– 0.1020	0.1405
γ_1^{I}	– 0.1196	0.3270	– 0.1507	0.1752	– 0.0145	0.2500
β_0^{II}	0.0319	0.1650	– 0.0820	0.1007	0.0180	0.2037
β_1^{II}	0.0263	0.2274	– 0.0190	0.0991	– 0.0938	0.1928
γ_1^{II}	0.0050	0.1310	– 0.0571	0.1094	0.1146	0.2908
ζ	0.3798	0.4923	0.1248	0.3144	0.4136	0.6883
$\zeta(\tau)$	0.1063	0.3730	0.1248	0.3144	0.4136	0.6883
η	0.0870	0.1342	– 0.1478	0.1593	73247	248720
$\eta(\tau)$	– 0.0622	0.0867	– 0.1478	0.1593	73247	248720
MPE	0.0082	0.0482	0.0771	0.0861	– 0.2142	0.3402
MAPE	0.1259	0.1310	0.1461	0.1568	0.3320	0.3877
MAFE	0.1129	0.1420	0.2939	0.4278	0.4325	0.5591
Coverage	0.0008	0.0013	0.0071	0.0094	– 0.0150	0.0219

The picture is similar for the data-generating process with Student errors, $\varepsilon_t \sim t_4/\sqrt{2}$. Again, the correctly specified model, in this case GARCH-t, provides the most precise coefficient estimates. Interestingly, the GARCH-N estimator performs better in terms of prediction errors (MAPE) than GARCH-t despite the Student errors. Although the GACQ parameter estimates are less precise than the GARCH ones (with the exception of η), GACQ has similar MAPE as GARCH-N but outperforms GARCH-N (and thus GARCH-t) in terms of out-of-sample forecasting.

© 2019 Royal Economic Society.

Table A.5. The bias and RMSE of the GACQ and GARCH estimators in the case of Student t_4 distributed errors.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.2086	0.5754	0.0209	0.1717	0.0298	0.1314
β_1^{I}	0.0547	0.6786	– 0.0723	0.1494	– 0.1037	0.1293
γ_1^{I}	– 0.1722	0.4184	– 0.2043	0.2504	– 0.1763	0.2143
β_0^{II}	0.0455	0.2513	– 0.0824	0.1638	– 0.1066	0.1271
β_1^{II}	0.0429	0.3827	– 0.0279	0.1677	– 0.0270	0.1181
γ_1^{II}	0.0024	0.1858	– 0.0749	0.1085	– 0.0753	0.0939
ζ	0.3625	0.4689	0.1240	0.3008	0.1386	0.2587
$\zeta(\tau)$	0.1940	0.4330	0.1240	0.3008	0.1386	0.2587
η	0.0304	0.0781	– 0.1453	0.1657	– 0.1267	0.1560
$\eta(\tau)$	– 0.0845	0.0927	– 0.1453	0.1657	– 0.1267	0.1560
MPE	0.0046	0.0596	– 0.0325	0.0676	– 0.3239	0.3368
MAPE	0.1536	0.1585	0.1409	0.1517	0.3271	0.3397
MAFE	0.1286	0.1798	0.3164	0.5886	0.4561	0.6462
Coverage	0.0008	0.0013	– 0.0044	0.0081	– 0.0254	0.0259

Table A.6. The bias and RMSE of the GACQ and GARCH estimators in the case of errors following the re-centered Type 1 Gumbel distribution.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.1368	0.2631	– 0.0183	0.0286	– 0.0113	0.0694
β_1^{I}	– 0.0295	0.2081	– 0.0857	0.1405	– 0.1261	0.2264
γ_1^{I}	– 0.0838	0.1921	– 0.0076	0.0573	0.0045	0.0550
β_0^{II}	0.0615	0.1230	0.1304	0.1903	0.1162	0.1783
β_1^{II}	– 0.0365	0.1255	– 0.1858	0.2224	– 0.1296	0.1966
γ_1^{II}	– 0.0025	0.1153	0.0949	0.1826	0.0571	0.1329
ζ	0.2723	0.4382	– 0.1837	0.2432	0.4859	2.4011
$\zeta(\tau)$	0.1346	0.2908	– 0.1837	0.2432	0.4859	2.4011
η	0.0575	0.0967	– 0.0476	0.1115	– 0.0615	0.0898
$\eta(\tau)$	– 0.0418	0.0751	– 0.0476	0.1115	– 0.0615	0.0898
MPE	0.0108	0.0279	0.0351	0.0414	– 0.2860	0.2961
MAPE	0.0727	0.0750	0.1045	0.1161	0.2905	0.2999
MAFE	0.0611	0.0836	0.1261	0.1815	0.5775	0.6107
Coverage	0.0006	0.0014	0.0017	0.0064	– 0.0263	0.0269

We finalise the innovation-distribution–related group of experiments by studying a member of the class of asymmetric distributions. Table A.6 shows results for the case where innovations follow a Gumbel distribution, which is parameterised by location parameter $\mu_G = 0$ and scale parameter $\beta_G = \sqrt{6}/\pi$. We re-centred the innovations by subtracting $\beta_G e^1$ from each realisation so that ε has mean zero. Because it is distribution agnostic, it should come as no surprise that the performance of the proposed GACQ model is similar to the previous experiments with symmetric errors. Being misspecified, the GARCH-N and GARCH-

© 2019 Royal Economic Society.

Table A.7. The bias and RMSE of the GACQ and GARCH estimators in the case of normally and Student distributed errors with 2.5% outliers.

	GACQ: $\varepsilon \sim \mathcal{N}$		GARCH-N: $\varepsilon \sim \mathcal{N}$		GACQ: $\varepsilon \sim t(4)$		GARCH-t: $\varepsilon \sim t(4)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^{I}	0.2656	0.6630	– 0.3962	0.4538	0.2758	0.7464	– 0.0975	0.2817
β_1^{I}	0.0358	0.4642	0.0647	0.3166	0.1187	0.6819	– 0.0614	0.1429
γ_1^{I}	– 0.0749	0.2867	0.0804	0.3435	– 0.0615	0.3810	– 0.0911	0.2344
β_0^{II}	0.0668	0.2337	0.4326	0.4719	0.1801	0.5841	0.1139	0.2651
β_1^{II}	0.0315	0.2780	– 0.0313	0.2028	0.0659	0.5414	– 0.0089	0.1635
γ_1^{II}	– 0.0010	0.1379	– 0.1096	0.1308	– 0.0290	0.1701	– 0.0689	0.0972
ζ	0.1545	0.5263	1.0059	1.0553	0.1686	0.5461	0.3365	0.5907
$\zeta(\tau)$	0.1832	0.5171	1.0059	1.0553	0.1599	0.5316	0.3365	0.5907
η	0.0678	0.1588	– 0.1480	0.3122	0.0262	0.1183	– 0.0419	0.5081
$\eta(\tau)$	– 0.0445	0.0929	– 0.1480	0.3122	– 0.0805	0.0939	– 0.0419	0.5081
MPE	– 0.1023	0.1291	–175.09	1008.6	– 0.1528	0.1940	– 0.9166	0.9684
MAPE	0.2180	0.2311	175.22	1008.6	0.2749	0.2933	0.9182	0.9696
MAFE	0.2175	0.4203	0.5545	0.6154	0.2791	0.5217	0.9664	1.1980
Coverage	0.0007	0.0013	– 0.0144	0.0164	0.0009	0.0014	– 0.0269	0.0273

t models now provide less precise estimates in regime II, and both their MAPEs and MAFEs are larger than those of GACQ.

Finally, we look at the case in which normally distributed innovations are contaminated by outliers. We define them as follows. Let $\varepsilon_t \sim N(0, 1)$ or t_4 and $r_t \sim \mathcal{U}[0, 1]$ are independent and uniformly distributed. Then, for each $u_t(\boldsymbol{\theta}_0) = \sigma_t(z_t, \boldsymbol{\theta}_0)\varepsilon_t$, the contaminated series $\{u'\}_{t=1}^n$ is defined as

$$u_t' := u_t + \mathbb{1}_{\{r_t \leq 0.025\}} \mathrm{sgn}(\varepsilon_t) 3\sigma_\varepsilon$$

with $\sigma_\varepsilon = 1$. Note that this might be considered a very small contamination, but we report estimates for the 5% VaR, and thus these contaminated values form a large proportion of the data used for estimation. Predictably, the RMSEs of all the parameters' estimates increase in the presence of contaminations irrespective of the considered model. Considering MAPE, the increase in the prediction errors is relatively limited in the case of the GACQ method. On the other hand, the MAPE of GARCH-N increases substantially, and the model exhibits large prediction biases. The situation is similar for GARCH-t, which, however, partially compensates for fatter tails by underestimating its degrees-of-freedom parameter. Most importantly, the coverage ratio in the conditional variance GARCH models are on average off by 1.4 and 2.7 percentage points, respectively, which is a rather significant deviation given that we consider the 5% VaR, whereas the coverage ratio of GACQ is unaffected by the contamination.

© 2019 Royal Economic Society.