Supplemental Appendix

S.A.1 Detailed proofs

In order to prove Proposition 1, we use the following five lemmas. First, we recall the definition of stochastic equicontinuity.

Definition 2 *(Andrews (1994)) The empirical process* $\{\mathbf{h}_T(\cdot) : T \geq 1\}$ *is stochastically equicontinuous if* $\forall\ \varepsilon > 0$ *and* $\eta > 0, \exists\ \delta > 0$ *such that*

$$\limsup_{T\to\infty} P\left[\sup_{\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|<\delta} \|\mathbf{h}_T(\boldsymbol{\theta}_1) - \mathbf{h}_T(\boldsymbol{\theta}_2)\| > \eta\right] < \varepsilon \tag{2}$$

Lemma 1 *Under Assumptions 1 and 2,*

(i) $\frac{1}{T}\sum_{t=1}^{T} \hat{F}_i(\hat{\eta}_{it}) \hat{F}_j(\hat{\eta}_{jt}) \xrightarrow{p} \int\int uv dC_{\eta_i,\eta_j}(u, v; \boldsymbol{\theta}_0)$ *as* $T \to \infty$

(ii) $\frac{1}{T}\sum_{t=1}^{T} 1\left\{\hat{F}_i(\hat{\eta}_{it}) \leq q, \hat{F}_j(\hat{\eta}_{jt}) \leq q\right\} \xrightarrow{p} C_{\eta_i,\eta_j}(q, q; \boldsymbol{\theta}_0)$ *as* $T \to \infty$

(iii) $\frac{1}{S}\sum_{s=1}^{S} \hat{G}_i(x_{is}(\boldsymbol{\theta})) \hat{G}_j(x_{js}(\boldsymbol{\theta})) \xrightarrow{p} \int\int uv dC_{\eta_i,\eta_j}(u, v; \boldsymbol{\theta})$ *for* $\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$ *as* $S \to \infty$

(iv) $\frac{1}{S}\sum_{s=1}^{S} 1\left\{\hat{G}_i(x_{is}(\boldsymbol{\theta})) \leq q, \hat{G}_j(x_{js}(\boldsymbol{\theta})) \leq q\right\} \xrightarrow{p} C_{\eta_i,\eta_j}(q, q; \boldsymbol{\theta})$ *for* $\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$ *as* $S \to \infty$

(v) $\frac{1}{S}\sum_{s=1}^{S} G_i(x_{is}(\boldsymbol{\theta})) G_j(x_{js}(\boldsymbol{\theta})) \xrightarrow{p} \int\int uv dC_{\eta_i,\eta_j}(u, v; \boldsymbol{\theta})$ *for* $\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$ *as* $S \to \infty$

(vi) $\frac{1}{S}\sum_{s=1}^{S} 1\{G_i(x_{is}(\boldsymbol{\theta})) \leq q, G_j(x_{js}(\boldsymbol{\theta})) \leq q\} \xrightarrow{p} C_{\eta_i,\eta_j}(q, q; \boldsymbol{\theta})$ *for* $\forall\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$ *as* $S \to \infty$

Proof of Lemma 1. Under Assumption 1, parts (iii) and (iv) of Lemma 1 can be proven by Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004). Under Assumption 2, Corollary 1 of Rémillard (2010) proves that the empirical copula process constructed by the standardized residuals $\hat{\boldsymbol{\eta}}_t$ weakly converges to the limit of that constructed by the innovations $\boldsymbol{\eta}_t$, which combined with Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004) yields parts (i) and (ii) above. In the case where it is possible to simulate directly from the copula rather than the joint distribution, e.g. Clayton/Gaussian copula in Section 3 or where we only can simulate from the joint distribution but know the marginal distribution G_i in closed form, it is not necessary to estimate marginal distribution G_i. In this case, instead of (iii) and (iv), (v) and (vi) are used for the later proofs. (v) and (vi) are proven by the standard law of large numbers. ■

Lemma 2 *(Lemma 2.8 of Newey and McFadden (1994)) Suppose* $\mathbf{\Theta}$ *is compact and* $\mathbf{g}_0(\boldsymbol{\theta})$ *is continuous. Then* $\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\| \xrightarrow{p} 0$ *as* $T, S \to \infty$ *if and only if* $\mathbf{g}_{T,S}(\boldsymbol{\theta}) \xrightarrow{p} \mathbf{g}_0(\boldsymbol{\theta})$ *for any* $\boldsymbol{\theta} \in \mathbf{\Theta}$ *as* $T, S \to \infty$ *and* $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ *is stochastically equicontinuous.*

Lemma 2 states that sufficient and necessary conditions for uniform convergence are pointwise convergence and stochastic equicontinuity. The following lemma shows that uniform convergence of the moment functions $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ implies uniform convergence of the objective function $Q_{T,S}(\boldsymbol{\theta})$.

Lemma 3 *If* $\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\| \xrightarrow{p} 0$ *as* $T, S \to \infty$*, then* $\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |Q_{T,S}(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \xrightarrow{p} 0$ *as* $T, S \to \infty$.

Proof of Lemma 3. By the triangle inequality and Cauchy-Schwarz inequality

$$
\begin{aligned}
|Q_{T,S}(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \leq\; & \left|[\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})]' \hat{\mathbf{W}}_T [\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})]\right| \\
& + \left|\mathbf{g}_0(\boldsymbol{\theta})' \left(\hat{\mathbf{W}}_T + \hat{\mathbf{W}}_T'\right) [\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})]\right| + \left|\mathbf{g}_0(\boldsymbol{\theta})' \left(\hat{\mathbf{W}}_T - \mathbf{W}_0\right) \mathbf{g}_0(\boldsymbol{\theta})\right| \\
\leq\; & \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\|^2 \left\|\hat{\mathbf{W}}_T\right\| + 2\|\mathbf{g}_0(\boldsymbol{\theta})\| \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\| \left\|\hat{\mathbf{W}}_T\right\| \\
& + \|\mathbf{g}_0(\boldsymbol{\theta})\|^2 \left\|\hat{\mathbf{W}}_T - \mathbf{W}_0\right\|
\end{aligned}
\tag{3}
$$

Then note that $\mathbf{g}_0(\boldsymbol{\theta})$ is bounded, $\hat{\mathbf{W}}_T$ is $O_p(1)$ and converges to $\mathbf{W}_0$ by Assumption 3(iv), and $\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\| = o_p(1)$ is given. So

$$
\begin{aligned}
\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |Q_{T,S}(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \leq\; & \left(\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\|\right)^2 O_p(1) \\
& + 2O(1) \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\| O_p(1) + o_p(1) = o_p(1)
\end{aligned}
\tag{4}
$$

■

Lemma 4 *Under Assumption 1, Assumption 2, and Assumption 3(iii),*

(i) $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ *is stochastic Lipschitz continuous, i.e.*

$\exists B_{T,S} = O_p(1)$ *such that for all* $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, $\|\mathbf{g}_{T,S}(\boldsymbol{\theta}_1) - \mathbf{g}_{T,S}(\boldsymbol{\theta}_2)\| \leq B_{T,S} \cdot \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$

(ii) There exists $\delta > 0$ *such that*

$$\lim \sup_{T,S\to\infty} E\left(B_{T,S}^{2+\delta}\right) < \infty \text{ for some } \delta > 0$$

Proof of Lemma 4. Without loss of generality, assume that $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ is scalar. By Lemma 1, we know that

$$\tilde{\mathbf{m}}_S(\boldsymbol{\theta}) = \mathbf{m}_0(\boldsymbol{\theta}) + o_p(1) \tag{5}$$

Also, by Assumption 3(iii) and the fact that $\mathbf{m}(\boldsymbol{\theta})$ consists of a function of Lipschitz continuous $C_{ij}(\boldsymbol{\theta})$, $\mathbf{m}_0(\boldsymbol{\theta})$ is Lipschitz continuous, i.e. $\exists K$ such that

$$|\mathbf{m}_0(\boldsymbol{\theta}_1) - \mathbf{m}_0(\boldsymbol{\theta}_2)| \leq K\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \tag{6}$$

Then,

$$\begin{aligned} |\mathbf{g}_{T,S}(\boldsymbol{\theta}_1) - \mathbf{g}_{T,S}(\boldsymbol{\theta}_2)| &= |\tilde{\mathbf{m}}_S(\boldsymbol{\theta}_1) - \tilde{\mathbf{m}}_S(\boldsymbol{\theta}_2)| = |\mathbf{m}_0(\boldsymbol{\theta}_1) - \mathbf{m}_0(\boldsymbol{\theta}_2) + o_p(1)| \qquad (7) \\ &\leq |\mathbf{m}_0(\boldsymbol{\theta}_1) - \mathbf{m}_0(\boldsymbol{\theta}_2)| + |o_p(1)| \\ &\leq K\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| + |o_p(1)| \\ &= \underbrace{\left(K + \frac{|o_p(1)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|}\right)}_{=O_p(1)}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \end{aligned}$$

and let $B_{T,S} = K + \frac{|o_p(1)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|}$. Then for some $\delta > 0$

$$\lim \sup_{T,S\to\infty} E\left(B_{T,S}^{2+\delta}\right) = \lim \sup_{T,S\to\infty} E\left(K + \frac{|o_p(1)|}{\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|}\right)^{2+\delta} < \infty \tag{8}$$

■

Lemma 5 *(Theorem 2.1 of Newey and McFadden (1994)) Suppose that (i)* $Q_0(\boldsymbol{\theta})$ *is uniquely minimized at* $\boldsymbol{\theta}_0$; *(ii)* $\boldsymbol{\Theta}$ *is compact; (iii)* $Q_0(\boldsymbol{\theta})$ *is continuous (iv)* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\hat{Q}_T(\boldsymbol{\theta})-Q_0(\boldsymbol{\theta})\right| \xrightarrow{p} 0$. *Then* $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$

Proof of Proposition 1. We prove this proposition by checking the conditions of Lemma 5.

(i) $Q_0(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$ by Assumption 3(i) and Assumption 3(iv).

(ii) $\boldsymbol{\Theta}$ is compact by Assumption 3(ii).

(iii) $Q_0(\boldsymbol{\theta})$ consists of linear combinations of rank correlations and quantile dependence measures that are functions of pair-wise copula functions. Therefore, $Q_0(\boldsymbol{\theta})$ is continuous by Assumption 3(iii).

(iv) The pointwise convergence of $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ to $\mathbf{g}_0(\boldsymbol{\theta})$ and the stochastic Lipschitz continuity of $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ are shown by Lemma 1 and by Lemma 4(i), respectively. By Lemma 2.9 of Newey and McFadden (1994), the stochastic Lipschitz continuity of $\mathbf{g}_{T,S}(\boldsymbol{\theta})$ ensures the stochastic equicontinuity of $\mathbf{g}_{T,S}(\boldsymbol{\theta})$, and under Assumption 3, $\boldsymbol{\Theta}$ is compact and $\mathbf{g}_0(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$. Therefore, $\mathbf{g}_{T,S}$ uniformly converges in probability to $\mathbf{g}_0$ by Lemma 2. This implies that $Q_{T,S}$ uniformly converges in probability to Q_0 by Lemma 3. ■

The proof of Proposition 2 uses the following three lemmas.

Lemma 6 *Let the dependence measures of interest include rank correlation and quantile dependence measures, and possibly linear combinations thereof. Then under Assumptions 1 and 2,*

$$\sqrt{T}\left(\hat{\mathbf{m}}_T-\mathbf{m}_0(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N(0,\boldsymbol{\Sigma}_0) \ \ \textit{as } T \ \rightarrow \ \infty \tag{9}$$

$$\sqrt{S}\left(\tilde{\mathbf{m}}_S(\boldsymbol{\theta}_0)-\mathbf{m}_0(\boldsymbol{\theta}_0)\right) \xrightarrow{d} N(0,\boldsymbol{\Sigma}_0) \ \ \textit{as } S \ \rightarrow \ \infty \tag{10}$$

Proof of Lemma 6. Follows from Theorem 3 and Theorem 6 of Fermanian, Radulović and Wegkamp (2004) and Corollary 1, Proposition 2 and Proposition 4 of Rémillard (2010). ■

We use Theorem 7.2 of Newey & McFadden (1994) to establish the asymptotic normality of our estimator, and this relies on showing the stochastic equicontinuity of $\mathbf{v}_{T,S}(\boldsymbol{\theta})$ defined below.

Lemma 7 *Suppose that Assumptions 1, 2, and 3(iii) hold. Then when* $S/T \to \infty$ *or* $S/T \to k \in (0,\infty)$, $\mathbf{v}_{T,S}(\boldsymbol{\theta}) = \sqrt{T}\left[\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\right]$ *is stochastically equicontinuous and when* $S/T \to 0$, $\mathbf{v}_{T,S}(\boldsymbol{\theta}) = \sqrt{S}\left[\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_0(\boldsymbol{\theta})\right]$ *is stochastically equicontinuous.*

Proof of Lemma 7. By Lemma 4(i), $\{\mathbf{g}_{\cdot,\cdot}(\boldsymbol{\theta}) : \theta \in \Theta\}$ is a type II class of functions in Andrews (1994). By Theorem 2 of Andrews (1994), $\{\mathbf{g}_{\cdot,\cdot}(\boldsymbol{\theta}) : \theta \in \Theta\}$ satisfies Pollard's entropy condition with envelope $1 \vee \sup_{\theta} \|\mathbf{g}_{\cdot,\cdot}(\boldsymbol{\theta})\| \vee B_{\cdot,\cdot}$, so Assumption A of Andrews (1994) is satisfied. Since $\mathbf{g}_{\cdot,\cdot}(\boldsymbol{\theta})$ is bounded and by the condition of $\limsup_{T,S\to\infty} E\left(B_{T,S}^{2+\delta}\right) < \infty$ for some $\delta > 0$ by Lemma 4(ii), the Assumption B of Andrews (1994) is also satisfied. Therefore, $\mathbf{v}_{T,S}(\boldsymbol{\theta})$ is stochastically equicontinuous by Theorem 1 of Andrews (1994). ■

Lemma 8 *(Theorem 7.2 of Newey & McFadden (1994)) Suppose that* $\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}\right)' \hat{\mathbf{W}}_T \mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}\right) \leq \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbf{g}_{T,S}(\boldsymbol{\theta})' \hat{\mathbf{W}}_T \mathbf{g}_{T,S}(\boldsymbol{\theta}) + o_p\left(T^{-1}\right), \hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ *and* $\hat{\mathbf{W}}_T \xrightarrow{p} \mathbf{W}_0, \mathbf{W}_0$ *is positive semi-definite, where there is* $\mathbf{g}_0(\boldsymbol{\theta})$ *such that (i)* $\mathbf{g}_0(\boldsymbol{\theta}_0) = 0$*,(ii)* $\mathbf{g}_0(\boldsymbol{\theta})$ *is differentiable at* $\boldsymbol{\theta}_0$ *with derivative* $\mathbf{G}_0$ *such that* $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ *is nonsingular,(iii)* $\boldsymbol{\theta}_0$ *is an interior point of* $\boldsymbol{\Theta}$*,(iv)* $\sqrt{T}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_0)$*,(v)* $\exists\delta$ *such that* $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta} \sqrt{T}\left\|\mathbf{g}_{T,S}(\boldsymbol{\theta}) - \mathbf{g}_{T,S}(\boldsymbol{\theta}_0) - \mathbf{g}_0(\boldsymbol{\theta})\right\| / \left[1 + \sqrt{T}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\right] \xrightarrow{p} 0$*. Then* $\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(0, \left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Sigma}_0\mathbf{W}_0\mathbf{G}_0\left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\right)$.

Proof of Proposition 2. We prove this proposition by checking conditions of Lemma 8.

(i) $\mathbf{g}_0(\boldsymbol{\theta}_0) = 0$ by construction of $\mathbf{g}_0(\boldsymbol{\theta}) = \mathbf{m}_0(\boldsymbol{\theta}_0) - \mathbf{m}_0(\boldsymbol{\theta})$

(ii) $\mathbf{g}_0(\boldsymbol{\theta})$ is differentiable at $\boldsymbol{\theta}_0$ with derivative $\mathbf{G}_0$ such that $\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0$ is nonsingular by Assumption 4(ii).

(iii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}$ by Assumption 4(i).

(iv) If $S/T \to \infty$ as $T, S \to \infty$,

$$\begin{aligned} \sqrt{T}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) &= \sqrt{T}\left(\hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S(\boldsymbol{\theta}_0)\right) \\ &= \sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\boldsymbol{\theta}_0)\right) - \sqrt{T}\left(\tilde{\mathbf{m}}_S(\boldsymbol{\theta}_0) - \mathbf{m}_0(\boldsymbol{\theta}_0)\right) \\ &= \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\boldsymbol{\theta}_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} - \underbrace{\frac{\sqrt{T}}{\sqrt{S}}}_{=o(1)} \times \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\boldsymbol{\theta}_0) - \mathbf{m}_0(\boldsymbol{\theta}_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} \end{aligned} \tag{11}$$

Therefore,

$$\sqrt{T}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma_0) \text{ as } T, S \to \infty.$$

If $S/T \to k \in (0, \infty)$ as $T, S \to \infty$,

$$\sqrt{T}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) = \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\boldsymbol{\theta}_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} - \underbrace{\frac{\sqrt{T}}{\sqrt{S}}}_{\to 1/\sqrt{k}} \times \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\boldsymbol{\theta}_0) - \mathbf{m}_0(\boldsymbol{\theta}_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}}$$

Therefore,

$$\sqrt{T}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{k}\right)\Sigma_0\right) \text{ as } T, S \to \infty.$$

If $S/T \to 0$ as $T, S \to \infty$,

$$\sqrt{S}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) = \underbrace{\frac{\sqrt{S}}{\sqrt{T}}}_{=o(1)} \times \underbrace{\sqrt{T}\left(\hat{\mathbf{m}}_T - \mathbf{m}_0(\boldsymbol{\theta}_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}} - \underbrace{\sqrt{S}\left(\tilde{\mathbf{m}}_S(\boldsymbol{\theta}_0) - \mathbf{m}_0(\boldsymbol{\theta}_0)\right)}_{\xrightarrow{d} N(0,\Sigma_0) \text{ by Lemma 6}}$$

Therefore,

$$\sqrt{S}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma_0) \text{ as } T, S \to \infty$$

Consolidating these results across all three combinations of divergence rates for S and T we obtain:

$$\frac{1}{\sqrt{1/S + 1/T}}\mathbf{g}_{T,S}(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma_0) \text{ as } T, S \to \infty.$$

(v) We established the stochastic equicontinuity of $\mathbf{v}_{T,S}(\boldsymbol{\theta}) = \sqrt{T}\left[\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_0(\boldsymbol{\theta})\right]$ when $S/T \to \infty$ or $S/T \to k$ by Lemma 7, i.e. for $\forall\varepsilon > 0$, $\eta > 0, \exists\delta$ such that

$$\begin{aligned}
&\limsup_{T\to\infty} P\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta} \|\mathbf{v}_{T,S}(\boldsymbol{\theta}) - \mathbf{v}_{T,S}(\boldsymbol{\theta}_0)\| > \eta\right] \\
= &\limsup_{T\to\infty} P\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta} \sqrt{T}\left\|\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_{T,S}(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})\right\| > \eta\right] < \varepsilon
\end{aligned} \tag{12}$$

and from the following inequality

$$\sqrt{T}\left\|\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_{T,S}(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})\right\| / \left[1 + \sqrt{T}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\right] \le \sqrt{T}\left\|\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_{T,S}(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})\right\| \tag{13}$$

we know that

$$\begin{aligned}
&\limsup_{T\to\infty} P\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta} \sqrt{T}\left\|\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_{T,S}(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})\right\| / \left[1 + \sqrt{T}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\right] > \eta\right] \\
\le &\limsup_{T\to\infty} P\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta} \sqrt{T}\left\|\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_{T,S}(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})\right\| > \eta\right] < \varepsilon
\end{aligned} \tag{14}$$

Similarly, it can be shown that when $S/T \to 0$,

$$\limsup_{S\to\infty} P\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta} \sqrt{S}\left\|\boldsymbol{g}_{T,S}(\boldsymbol{\theta}) - \boldsymbol{g}_{T,S}(\boldsymbol{\theta}_0) - \boldsymbol{g}_0(\boldsymbol{\theta})\right\| / \left[1 + \sqrt{S}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\right] > \eta\right] < \varepsilon. \tag{15}$$

■

Proof of Proposition 3. First, we prove the consistency of the numerical derivatives $\hat{\mathbf{G}}_{T,S}$. This part of the proof is similar to that of Theorem 7.4 in Newey and McFadden (1994). We will consider one-sided derivatives first, with the same arguments applying to two-sided derivatives. First we consider the case where $S/T \to \infty$ or $S/T \to k > 0$ as $T, S \to \infty$. We know that $\left\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right\| = O_p\left(T^{-1/2}\right)$ by the conclusion of Proposition 2. Also, by assumption we have $\varepsilon_{T,S} \to 0$ and $\varepsilon_{T,S}\sqrt{T} \to \infty$, so

$$\left\|\hat{\boldsymbol{\theta}}_{T,S} + \mathbf{e}_k\varepsilon_{T,S} - \boldsymbol{\theta}_0\right\| \le \left\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right\| + \|\mathbf{e}_k\varepsilon_{T,S}\| = O_p\left(T^{-1/2}\right) + O(\varepsilon_{T,S}) = O_p(\varepsilon_{T,S})$$

(Recall that $\mathbf{e}_k$ is the k^{th} unit vector.) In the proof of Proposition 2, it is shown that $\exists\delta$ such that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta}\sqrt{T}\left\|\mathbf{g}_{T,S}\left(\boldsymbol{\theta}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)-\mathbf{g}_0\left(\boldsymbol{\theta}\right)\right\|/\left[1+\sqrt{T}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right\|\right]=o_p\left(1\right)$$

Substituting $\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}$ for $\boldsymbol{\theta}$, then for T,S large, it follows that

$$\sqrt{T}\left\|\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)-\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)\right\|/\left[1+\sqrt{T}\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|\right]\leq o_p\left(1\right)$$

$$\begin{aligned}
&\text{so}\quad\left\|\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)-\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)\right\|\\
\leq&\quad\left[1+\sqrt{T}\underbrace{\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|}_{=O_p\left(\varepsilon_{T,S}\right)}\right]o_p\left(\frac{1}{\sqrt{T}}\right)\\
=&\quad\sqrt{T}O_p\left(\varepsilon_{T,S}\right)o_p\left(\frac{1}{\sqrt{T}}\right)=O_p\left(\varepsilon_{T,S}\right)o_p\left(1\right)\\
=&\quad o_p\left(\varepsilon_{T,S}\right)
\end{aligned}\tag{16}$$

On the other hand, since $\mathbf{g}_0\left(\boldsymbol{\theta}\right)$ is differentiable at $\boldsymbol{\theta}_0$ with derivative $\mathbf{G}_0$ by Assumption 4(ii), a Taylor expansion of $\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)$ around $\boldsymbol{\theta}_0$ is

$$\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)=\mathbf{g}_0\left(\boldsymbol{\theta}_0\right)+\mathbf{G}_0\cdot\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right)+o\left(\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|\right)$$

with $\mathbf{g}_0\left(\boldsymbol{\theta}_0\right)=0$. Then divide by $\varepsilon_{T,S}$,

$$\begin{aligned}
\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}&=\mathbf{G}_0\cdot\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right)/\varepsilon_{T,S}+o\left(\varepsilon_{T,S}^{-1}\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|\right)\\
\text{so}\quad\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k&=\mathbf{G}_0\cdot\left(\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right)/\varepsilon_{T,S}+o\left(\varepsilon_{T,S}^{-1}\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|\right)
\end{aligned}$$

The triangle inequality implies that

$$\begin{aligned}
\left\|\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k\right\|&\leq\left\|\mathbf{G}_0\cdot\left(\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right)/\varepsilon_{T,S}\right\|+o\left(\varepsilon_{T,S}^{-1}\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|\right)\\
&=\frac{1}{\sqrt{T}\varepsilon_{T,S}}\left\|\mathbf{G}_0\cdot\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right)\right\|\\
&\quad+\varepsilon_{T,S}^{-1}\left\|\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}-\boldsymbol{\theta}_0\right\|o\left(1\right)\\
&=o\left(1\right)O_p\left(1\right)+\varepsilon_{T,S}^{-1}O_p\left(\varepsilon_{T,S}\right)o\left(1\right)=o_p\left(1\right)
\end{aligned}\tag{17}$$

Combining the inequalities in equations (16) and (17) gives

$$\begin{aligned}
\left(\frac{\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)}{\varepsilon_{T,S}}-\mathbf{G}_0\mathbf{e}_k\right) &= \left(\frac{\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)-\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)}{\varepsilon_{T,S}}\right) \\
&\quad +\left(\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k\right) \\
\left\|\frac{\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)}{\varepsilon_{T,S}}-\mathbf{G}_0\mathbf{e}_k\right\| &\leq \left\|\frac{\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)-\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)}{\varepsilon_{T,S}}\right\| \\
&\quad +\left\|\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)/\varepsilon_{T,S}-\mathbf{G}_0\mathbf{e}_k\right\| \\
&\leq o_p\left(1\right)
\end{aligned}$$

Then,

$$\frac{\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)}{\varepsilon_{T,S}}\xrightarrow{p}\mathbf{G}_0\mathbf{e}_k$$

and the same arguments can be applied to the two-sided derivative:

$$\frac{\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}+\mathbf{e}_k\varepsilon_{T,S}\right)-\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}-\mathbf{e}_k\varepsilon_{T,S}\right)}{2\varepsilon_{T,S}}\xrightarrow{p}\mathbf{G}_0\mathbf{e}_k$$

This holds for each column $k=1,2...,p$. Thus $\hat{\mathbf{G}}_{T,S}\xrightarrow{p}\mathbf{G}_0$.

In the case where $S/T\rightarrow 0$ as $T,S\rightarrow\infty$, the proof for the consistency of $\hat{\mathbf{G}}_{T,S}$ is done in the similar way using the following facts:

$$\left\|\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right\|=O_p\left(S^{-1/2}\right) \tag{18}$$

and $\exists\delta$

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\delta}\sqrt{S}\left\|\mathbf{g}_{T,S}\left(\boldsymbol{\theta}\right)-\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right)-\mathbf{g}_0\left(\boldsymbol{\theta}\right)\right\|/\left[1+\sqrt{S}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\right\|\right]=o_p\left(1\right) \tag{19}$$

Next, we show the consistency of $\hat{\boldsymbol{\Sigma}}_{T,B}$. If $\boldsymbol{\mu}_t$ and $\boldsymbol{\sigma}_t$ are known constant, or if $\boldsymbol{\phi}_0$ is known, then the result follows from Theorems 5 and 6 of Fermanian, Radulović and Wegkamp (2004). When $\boldsymbol{\phi}_0$ is estimated, the result is obtained by combining the results in Fermanian, Radulović and Wekkamp with those of Rémillard (2010), see the Proof of Proposition 3 in the paper for details. ■

Proof of Proposition 4. First consider $S/T \to \infty$ or $S/T \to k > 0$. A Taylor expansion of $\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right)$ around $\boldsymbol{\theta}_0$ yields

$$\sqrt{T}\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right) = \sqrt{T}\mathbf{g}_0\left(\boldsymbol{\theta}_0\right) + \mathbf{G}_0 \cdot \sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right) + o\left(\sqrt{T}\left\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right\|\right) \tag{20}$$

and since $\mathbf{g}_0\left(\boldsymbol{\theta}_0\right) = 0$ and $\sqrt{T}\left\|\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right\| = O_p\left(1\right)$

$$\sqrt{T}\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right) = \mathbf{G}_0 \cdot \sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right) + o_p\left(1\right) \tag{21}$$

Then consider the following expansion of $\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right)$ around $\boldsymbol{\theta}_0$

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) = \sqrt{T}\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right) + \hat{\mathbf{G}}_{T,S} \cdot \sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right) + \mathbf{R}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) \tag{22}$$

where the remaining term is captured by $\mathbf{R}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right)$. Combining equations (21) and (22) we obtain

$$\sqrt{T}\left[\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) - \mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right) - \mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right)\right] = \left(\hat{\mathbf{G}}_{T,S} - \mathbf{G}_0\right) \cdot \sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right) + \mathbf{R}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) + o_p\left(1\right)$$

Lemma 7 shows the stochastic equicontinuity of $\mathbf{v}_{T,S}\left(\boldsymbol{\theta}\right)$, which implies (see proof of Proposition 2) that

$$\sqrt{T}\left[\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) - \mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right) - \mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right)\right] = o_p\left(1\right)$$

By Proposition 3, $\hat{\mathbf{G}}_{T,S} - \mathbf{G}_0 = o_p\left(1\right)$, which implies $\mathbf{R}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) = o_p\left(1\right)$. Thus, we obtain the expansion of $\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right)$ around $\boldsymbol{\theta}_0$:

$$\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) = \sqrt{T}\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right) + \hat{\mathbf{G}}_{T,S} \cdot \sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S} - \boldsymbol{\theta}_0\right) + o_p\left(1\right) \tag{23}$$

The remainder of the proof is the same as in standard GMM applications: From the proof of Proposition 2, we have $\sqrt{T}\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(0, \boldsymbol{\Sigma}_0\right)$ and rewrite this as $-\boldsymbol{\Sigma}_0^{-1/2}\sqrt{T}\mathbf{g}_{T,S}\left(\boldsymbol{\theta}_0\right) \equiv$

$\mathbf{u}_{T,S}\xrightarrow{d}\mathbf{u}\sim N\left(0,\mathbf{I}\right)$, and from Proposition 2, we have $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right)=\left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Sigma}_0^{1/2}\mathbf{u}_{T,S}+o_p\left(1\right)$. By these two equations and Proposition 3, equation (23) becomes

$$\begin{aligned}\sqrt{T}\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) &= -\hat{\boldsymbol{\Sigma}}_{T,B}^{1/2}\mathbf{u}_{T,S}+\hat{\mathbf{G}}_{T,S}\left(\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,S}\right)^{-1}\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\boldsymbol{\Sigma}}_{T,B}^{1/2}\mathbf{u}_{T,S}+o_p\left(1\right) \\ &= -\hat{\boldsymbol{\Sigma}}_{T,B}^{1/2}\hat{\mathbf{R}}\mathbf{u}_{T,S}+o_p\left(1\right)\end{aligned} \tag{24}$$

where $\hat{\mathbf{R}}\equiv\left(\mathbf{I}-\hat{\boldsymbol{\Sigma}}_{T,B}^{-1/2}\hat{\mathbf{G}}_{T,S}\left(\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\mathbf{G}}_{T,S}\right)^{-1}\hat{\mathbf{G}}_{T,S}'\hat{\mathbf{W}}_T\hat{\boldsymbol{\Sigma}}_{T,B}^{1/2}\right)$. The test statistic is

$$\begin{aligned}T\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right)'\hat{\mathbf{W}}_T\mathbf{g}_{T,S}\left(\hat{\boldsymbol{\theta}}_{T,S}\right) &= \mathbf{u}_{T,S}'\hat{\mathbf{R}}'\hat{\boldsymbol{\Sigma}}_{T,B}^{1/2\prime}\hat{\mathbf{W}}_T\hat{\boldsymbol{\Sigma}}_{T,B}^{1/2}\hat{\mathbf{R}}\mathbf{u}_{T,S}+o_p\left(1\right) \\ &= \mathbf{u}'\mathbf{R}_0'\boldsymbol{\Sigma}_0^{1/2\prime}\mathbf{W}_0\boldsymbol{\Sigma}_0^{1/2}\mathbf{R}_0\mathbf{u}+o_p\left(1\right)\end{aligned} \tag{25}$$

where $\mathbf{R}_0\equiv\left(\mathbf{I}-\boldsymbol{\Sigma}_0^{-1/2}\mathbf{G}_0\left(\mathbf{G}_0'\mathbf{W}_0\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{W}_0\boldsymbol{\Sigma}_0^{1/2}\right)$. When $\hat{\mathbf{W}}_T=\hat{\boldsymbol{\Sigma}}_{T,B}^{-1}$, $\hat{\mathbf{R}}$ is symmetric and idempotent with $rank\left(\hat{\mathbf{R}}\right)=tr\left(\hat{\mathbf{R}}\right)=m-p$, and the test statistic converges to a χ^2_{m-p} random variable, as usual. In general, the asymptotic distribution is a sample-dependent combination of m independent standard Normal variables, namely that of $\mathbf{u}'\mathbf{R}_0'\boldsymbol{\Sigma}_0^{1/2\prime}\mathbf{W}_0\boldsymbol{\Sigma}_0^{1/2}\mathbf{R}_0\mathbf{u}$ where $\mathbf{u}\sim N\left(0,\mathbf{I}\right)$.

When $S/T\rightarrow 0$, a similar proof can be given using Taylor expansion of $\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right)$

$$\sqrt{S}\mathbf{g}_0\left(\hat{\boldsymbol{\theta}}_{T,S}\right)=\sqrt{S}\mathbf{g}_0\left(\boldsymbol{\theta}_0\right)+\mathbf{G}_0\cdot\sqrt{S}\left(\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right)+o\left(\sqrt{S}\left\|\hat{\boldsymbol{\theta}}_{T,S}-\boldsymbol{\theta}_0\right\|\right) \tag{26}$$

■

S.A.2 Implementation of the SMM estimator

This section provides further details on the constricution of the SMM objective function and the estimation of the parameter.

Our estimator is based on matching sample dependence measures (rank correlation, quantile dependence, etc) to measures of dependence computed on simulated data from the model evaluated at a given parameter $\boldsymbol{\theta}$. The sample dependence measures are stacked into a vector $\hat{\mathbf{m}}_T$, and the corresponding measures on the simulated data are stacked into a vector $\tilde{\mathbf{m}}_S(\boldsymbol{\theta})$. Re-stating equation (9) from the paper, our estimator is:

$$\hat{\boldsymbol{\theta}}_{T,S} \equiv \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\ \mathbf{g}'_{T,S}(\boldsymbol{\theta})\,\hat{\mathbf{W}}_T\mathbf{g}_{T,S}(\boldsymbol{\theta}) \tag{27}$$

where $\mathbf{g}_{T,S}(\boldsymbol{\theta}) \equiv \hat{\mathbf{m}}_T - \tilde{\mathbf{m}}_S(\boldsymbol{\theta})$.

We now describe the construction of the SMM objective function. All dependence measures used in this paper are based on the estimated standardized residuals, which are constructed as:

$$\hat{\boldsymbol{\eta}}_t \equiv \boldsymbol{\sigma}_t^{-1}(\hat{\boldsymbol{\phi}})[\mathbf{Y}_t - \boldsymbol{\mu}_t(\hat{\boldsymbol{\phi}})] \tag{28}$$

We then compute pair-wise dependence measures such as those in equations (4) and (5) of the paper, e.g., $\hat{\rho}^{ij}$ and $\hat{\lambda}_q^{ij}$. For quantile dependence we set $q \in \{0.05, 0.10, 0.90, 0.95\}$.

The copula models we consider all satisfy an "exchangeability" property, and we use that when constructing the moments to use in the estimator. Specifically, we calculate moments $\hat{\mathbf{m}}_T$ as:

$$\hat{\mathbf{m}}_T = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left[\begin{array}{ccccc}\hat{\rho}^{ij} & \hat{\lambda}_{0.05}^{ij} & \hat{\lambda}_{0.10}^{ij} & \hat{\lambda}_{0.90}^{ij} & \hat{\lambda}_{0.95}^{ij}\end{array}\right]' \tag{29}$$

Next we simulate data $\{\mathbf{X}_s(\boldsymbol{\theta})\}_{s=1}^{S}$ from distribution $\mathbf{F}_x(\boldsymbol{\theta})$, and compute the vector of dependence measures $\tilde{\mathbf{m}}_S(\boldsymbol{\theta})$. It is critically important in this step to keep the random number generator seed *fixed* across simulations, see Gouriéroux and Monfort (1996, *Simulation-Based Econometric Methods*, Oxford University Press). Failing to do so makes the simulated data "jittery" across function evaluations, and the numerical optimization algorithm will fail to converge.

Finally, we specify the weight matrix. In this paper we choose either $\hat{\mathbf{W}}_T = \mathbf{I}$ or $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}_{T,B}^{-1}$. Note that for our estimation problem the estimated efficient weight matrix, $\hat{\mathbf{\Sigma}}_{T,B}^{-1}$, depends on the covariance matrix of the vector of sample dependence measures, and not on the parameters of the model. Thus unlike some GMM or SMM estimation problems, this estimator does not require an initial estimate of the unknown parameter.

We use numerical optimization procedure to find $\hat{\boldsymbol{\theta}}_{T,S}$. As our objective function is not differentiable we cannot use procedures that rely on analytical or numerical derivatives (such as familiar Newton or "quasi-Newton" algorithms). We use "fminsearch" in MATLAB, which is a simplex search algorithm that does not require derivatives. As with all numerical optimization procedures, some care is required to ensure that a global optimum has been found. In each estimation, we consider many different starting values for the algorithm, and choose the resulting parameter estimate that leads to the smallest value of the objective function. The models considered here are relatively small, with up to three unknown parameters, but when the number of unknown parameters is large more care is required to ensure that a global optimum has been found, see Judd (1998, *Numerical Methods in Economics*, MIT Press) for more discussion.

S.A.3 Implementation of MLE for factor copulas

Consider a simple factor model:

$$\begin{aligned} X_i &= Z + \varepsilon_i, \quad i = 1, 2, ..., N \\ Z &\sim F_Z, \ \varepsilon_i \sim iid \ F_\varepsilon, \ \ \varepsilon_i \perp\!\!\!\perp Z \ \forall \ i \\ [X_1, ..., X_N]' &\equiv \mathbf{X} \sim \mathbf{F}_x = \mathbf{C}\left(G, ..., G\right) \end{aligned}$$

To obtain the copula density $\mathbf{c}$ we must first obtain the joint density, $\mathbf{f}_x$, and the marginal density, g. These can be obtained using numerical integration to "integrate out" the latent common factor, Z. First, note that

$$\begin{aligned} f_{x_i|z}\left(x_i|z\right) &= f_\varepsilon\left(x_i - z\right) \\ F_{x_i|z}\left(x_i|z\right) &= F_\varepsilon\left(x_i - z\right) \\ \text{and } \mathbf{f}_{x|z}\left(x_1,\ldots,x_N|z\right) &= \prod_{i=1}^{N} f_\varepsilon\left(x_i - z\right) \end{aligned}$$

Then the marginal density and marginal distribution of X_i are:

$$\begin{aligned} g\left(x\right) &= \int_{-\infty}^{\infty} f_{x|z}\left(x|z\right) f_z\left(z\right) dz = \int_{-\infty}^{\infty} f_\varepsilon\left(x - z\right) f_z\left(z\right) dz \\ G\left(x\right) &= \int_{-\infty}^{\infty} \Pr\left[X \leq x | Z = z\right] f_z\left(z\right) dz = \int_{-\infty}^{\infty} F_\varepsilon\left(x - z\right) f_z\left(z\right) dz \end{aligned}$$

The joint density is similarly obtained:

$$\mathbf{f}_x\left(x_1,\ldots,x_N\right) = \int_{-\infty}^{\infty} \mathbf{f}_{x|z}\left(x_1,\ldots,x_N|z\right) f_z\left(z\right) dz = \int_{-\infty}^{\infty} \prod_{i=1}^{N} f_\varepsilon\left(x_i - z\right) f_z\left(z\right) dz$$

From these, we obtain the copula density:

$$\mathbf{c}\left(u_1,\ldots,u_N\right) = \frac{\mathbf{f}_x\left(G^{-1}\left(u_1\right),\ldots,G^{-1}\left(u_N\right)\right)}{\prod_{i=1}^{N} g\left(G^{-1}\left(u_i\right)\right)}$$

We approximate the above integrals using Gauss-Legendre quadrature, see Judd (1998) for details and discussion. We use the probability integral transformation of Z to convert the above unbounded integals to integrals on $[0,1]$, for example:

$$g\left(x\right) = \int_{-\infty}^{\infty} f_\varepsilon\left(x - z\right) f_z\left(z\right) dz = \int_0^1 f_\varepsilon\left(x - F_z^{-1}\left(u\right)\right) du$$

A key choice in quadrature methods is the number of "nodes" to use in approximating the integral. We ran simulations using 50, 150, and 250 nodes, and found that the accuracy of the resulting MLE was slightly better for 150 than 50 nodes, and not different for 250 compared with 150 nodes. Thus in the paper we report results for MLE based on quadrature using 150 nodes.

S.A.4 Additional tables

Table S1: Simulation results for iid data with optimal weight matrix

	Clayton			*Normal*		*Factor copula*		
	GMM	SMM	SMM*	GMM	SMM	SMM		
	κ	κ	κ	ρ	ρ	σ^2	ν^{-1}	λ
True	1.00	1.00	1.00	0.5	0.5	1.00	0.25	-0.50
				$N = 2$				
Bias	-0.018	-0.020	-0.018	-0.001	0.000	0.016	-0.026	-0.094
St dev	0.085	0.092	0.091	0.025	0.026	0.144	0.119	0.189
Median	0.984	0.977	0.981	0.497	0.500	0.999	0.200	-0.557
90-10%	0.224	0.247	0.233	0.070	0.069	0.374	0.332	0.447
Time	0.07	515	51	0.41	0.67		112	
				$N = 3$				
Bias	0.008	0.010	0.006	-0.003	-0.003	0.022	-0.009	-0.057
St dev	0.063	0.073	0.068	0.021	0.022	0.110	0.103	0.146
Median	0.996	1.008	1.002	0.495	0.498	1.006	0.238	-0.540
90-10%	0.160	0.172	0.165	0.054	0.061	0.294	0.261	0.366
Time	0.12	1398	59	0.29	1.60		138	
				$N = 10$				
Bias	-0.003	-0.004	-0.005	-0.004	-0.004	0.019	-0.010	-0.023
St dev	0.047	0.049	0.050	0.014	0.015	0.097	0.078	0.085
Median	0.993	0.997	0.997	0.497	0.495	1.006	0.251	-0.514
90-10%	0.121	0.126	0.127	0.036	0.037	0.248	0.189	0.165
Time	1	22521	170	0.34	3		358	

Notes: The simulation design is the same as that of Table 1 in the paper except that we use the efficient weight matrix, $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}_{T,B}^{-1}$.

Table S2: Simulation results for AR-GARCH data with optimal weight matrix

	Clayton			*Normal*		*Factor copula*		
	GMM	SMM	SMM*	GMM	SMM	SMM		
	κ	κ	κ	ρ	ρ	σ^2	ν^{-1}	λ
True	1.00	1.00	1.00	0.5	0.5	1.00	0.25	-0.50
				$N=2$				
Bias	-0.021	-0.017	-0.014	-0.002	-0.001	0.018	-0.022	-0.083
St dev	0.087	0.097	0.097	0.026	0.026	0.154	0.121	0.188
Median	0.980	0.989	0.987	0.498	0.498	0.997	0.209	-0.553
90-10%	0.225	0.247	0.258	0.070	0.069	0.399	0.346	0.485
Time	0.06	531	60	0.39	0.69		119	
				$N=3$				
Bias	0.002	-0.004	-0.001	-0.003	-0.003	0.021	-0.009	-0.061
St dev	0.063	0.066	0.068	0.021	0.023	0.114	0.106	0.151
Median	0.995	0.990	0.991	0.495	0.497	1.018	0.243	-0.548
90-10%	0.153	0.166	0.164	0.052	0.058	0.299	0.278	0.336
Time	0.12	1613	76	0.33	1.50		135	
				$N=10$				
Bias	-0.006	-0.005	-0.007	-0.005	-0.005	0.014	-0.013	-0.027
St dev	0.047	0.051	0.050	0.014	0.015	0.093	0.078	0.097
Median	0.991	0.997	0.993	0.496	0.494	1.000	0.250	-0.513
90-10%	0.120	0.136	0.134	0.037	0.040	0.229	0.193	0.187
Time	2	25492	175	0.41	4		361	

Notes: The simulation design is the same as that of Table 2 in the paper except that we use the efficient weight matrix, $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}_{T,B}^{-1}$.

Table S3: Simulation results on coverage rates with optimal weight matrix

	Clayton		Normal		Factor copula			
	κ	J	ρ	J	σ^2	ν^{-1}	λ	J
				$N=2$				
$\varepsilon_{T,S}$								
0.1	89	95	93	99	97	99	96	96
0.01	56		93		95	99	97	
0.001	9		80		77	79	80	
0.0001	1		16		40	54	56	
				$N=3$				
$\varepsilon_{T,S}$								
0.1	91	98	88	95	98	99	97	99
0.01	70		88		98	99	96	
0.001	10		82		88	86	86	
0.0001	0		41		51	59	48	
				$N=10$				
$\varepsilon_{T,S}$								
0.1	93	100	87	97	95	96	94	100
0.01	79		87		94	94	93	
0.001	20		87		89	84	92	
0.0001	5		64		70	70	73	

Notes: The simulation design is the same as that of Table 3 in the paper except that we use the efficient weight matrix, $\hat{\mathbf{W}}_T = \hat{\mathbf{\Sigma}}_{T,B}^{-1}$. The numbers in column J present the percentage of simulations for which the test statistic of over-identifying restrictions test described in Section 2 was smaller than its critical value from chi square distribution under 95% confidence level (this test does not require a choice of step size for the numerical derivative, $\varepsilon_{T,S}$, and so we have only one value per model).

Table S4: Summary statistics on the daily stock returns

	Bank of America	Bank of N.Y.	Citi Group	Goldman Sachs	JP Morgan	Morgan Stanley	Wells Fargo
Mean	0.038	0.015	-0.020	0.052	0.041	0.032	0.047
Std dev	3.461	2.797	3.817	2.638	2.966	3.814	2.965
Skewness	1.048	0.592	1.595	0.984	0.922	4.982	2.012
Kurtosis	28.190	18.721	43.478	18.152	16.006	119.757	30.984

Notes: This table presents some summary statistics of the seven daily equity returns data used in the empirical analysis.

Table S5: Parameter estimates for the conditional mean and variance models

	Bank of America	Bank of N.Y.	Citi Group	Goldman Sachs	JP Morgan	Morgan Stanley	Wells Fargo
Constant (ϕ_0)	0.038	0.017	-0.019	0.058	0.043	0.031	0.051
$r_{i,t-1}$	0.020	-0.151	0.053	-0.156	-0.035	0.004	-0.078
$r_{m,t-1}$	-0.053	-0.011	0.029	0.282	-0.141	0.063	-0.099
Constant (ω)	0.009	0.069	0.019	0.034	0.014	0.036	0.008
$\sigma^2_{i,t-1}$	0.931	0.895	0.901	0.953	0.926	0.922	0.926
$\varepsilon^2_{i,t-1}$	0.031	0.017	0.036	0.000	0.025	0.002	0.021
$\varepsilon^2_{i,t-1} \cdot \mathbf{1}_{\{\varepsilon_{i,t-1} \leq 0\}}$	0.048	0.079	0.123	0.077	0.082	0.135	0.108
$\varepsilon^2_{m,t-1}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\varepsilon^2_{m,t-1} \cdot \mathbf{1}_{\{\varepsilon_{m,t-1} \leq 0\}}$	0.068	0.266	0.046	0.012	0.064	0.077	0.013

Notes: This table presents the estimated models for the conditional mean (top panel) and conditional variance (lower panel).