Econometrica Supplementary Material

SUPPLEMENT TO "TIME-VARYING RISK PREMIUM IN LARGE CROSS-SECTIONAL EQUITY DATA SETS"

(*Econometrica*, Vol. 84, No. 3, May 2016, 985–1046)

BY PATRICK GAGLIARDINI, ELISA OSSOLA, AND OLIVIER SCAILLET

THESE SUPPLEMENTARY MATERIALS PROVIDE the derivation of Equations (9)–(12) (Appendix C), the proofs of technical lemmas used in the paper (Appendix D), the link of our no-arbitrage pricing restrictions with Chamberlain and Rothschild (1983) results (Appendix E), the check that the high-level assumptions in the paper hold under block-dependence (Appendix F), and the results of Monte Carlo experiments that investigate the finite sample properties of the estimators and test statistics (Appendix G). Finally, we investigate the effects of model misspecification on risk premia estimation and give estimates of the pseudo-true values (Appendix H).

APPENDIX C: DERIVATION OF EQUATIONS (9)–(12)

C.1. *Derivation of Equations (9) and (10)*

From Equation (8) and by using $\text{vec}[ABC] = [C' \otimes A]\text{vec}[B]$ (MN, Theorem 2, p. 35), we get $Z'_{t-1}B'_i f_t = \text{vec}[Z'_{t-1}B'_i f_t] = [f'_t \otimes Z'_{t-1}]\text{vec}[B'_i]$, and $Z'_{i,t-1}C'_i f_t = [f'_t \otimes Z'_{i,t-1}]\text{vec}[C'_i]$, which gives $Z'_{t-1}B'_i f_t + Z'_{i,t-1}C'_i f_t = x'_{2,i,t}\beta_{2,i}$.

Let us now consider the first two terms in the RHS of Equation (8).

(a) By definition of matrix X_t in Section 3.1, we have

$$\begin{aligned} Z'_{t-1}B'_i(\Lambda - F)Z_{t-1} &= \frac{1}{2}Z'_{t-1}\big[B'_i(\Lambda - F) + (\Lambda - F)'B_i\big]Z_{t-1} \\ &= \frac{1}{2}\text{vech}[X_t]'\,\text{vech}\big[B'_i(\Lambda - F) + (\Lambda - F)'B_i\big]. \end{aligned}$$

By using the Moore–Penrose inverse of the duplication matrix D_p, we get

$$\begin{aligned} &\text{vech}\big[B'_i(\Lambda - F) + (\Lambda - F)'B_i\big] \\ &\quad = D_p^+\big[\text{vec}\big[B'_i(\Lambda - F)\big] + \text{vec}\big[(\Lambda - F)'B_i\big]\big]. \end{aligned}$$

Finally, by the properties of the vec operator and the commutation matrix W_p, and the definition of matrix N_p, we obtain

$$\begin{aligned} &\frac{1}{2}D_p^+\big[\text{vec}\big[B'_i(\Lambda - F)\big] + \text{vec}\big[(\Lambda - F)'B_i\big]\big] \\ &\quad = \frac{1}{2}D_p^+(I_{p^2} + W_p)\,\text{vec}\big[B'_i(\Lambda - F)\big] \\ &\quad = N_p\big[(\Lambda - F)' \otimes I_p\big]\text{vec}\big[B'_i\big]. \end{aligned}$$

 DOI: 10.3982/ECTA11069

(b) By the properties of the tr and vec operators, we have

$$\begin{aligned}Z'_{i,t-1}C'_i(\Lambda-F)Z_{t-1} &= \mathrm{tr}\big[Z_{t-1}Z'_{i,t-1}C'_i(\Lambda-F)\big]\\ &= \mathrm{vec}\big[Z_{i,t-1}Z'_{t-1}\big]'\,\mathrm{vec}\big[C'_i(\Lambda-F)\big]\\ &= (Z_{t-1}\otimes Z_{i,t-1})'\big[(\Lambda-F)'\otimes I_q\big]\mathrm{vec}\big[C'_i\big].\end{aligned}$$

By combining (a) and (b), we get $Z'_{t-1}B'_i(\Lambda-F)Z_{t-1}+Z'_{i,t-1}C'_i(\Lambda-F)Z_{t-1}=x'_{1,i,t}\beta_{1,i}$ and $\beta_{1,i}=((N_p[(\Lambda-F)'\otimes I_p]\mathrm{vec}[B'_i])',([(\Lambda-F)'\otimes I_q]\mathrm{vec}[C'_i])')'$.

C.2. *Derivation of Equation (11)*

We use $\beta_{1,i}=((\frac{1}{2}D_p^+[\mathrm{vec}[B'_i(\Lambda-F)]+\mathrm{vec}[(\Lambda-F)'B_i]])',(\mathrm{vec}[C'_i(\Lambda-F)])')'$ from Section C.1. (a) From the properties of the vec operator and the commutation matrix W_p, we get

$$\begin{aligned}&\mathrm{vec}\big[B'_i(\Lambda-F)\big]+\mathrm{vec}\big[(\Lambda-F)'B_i\big]\\ &\quad=(W_p+I_{p^2})\,\mathrm{vec}\big[(\Lambda-F)'B_i\big]\\ &\quad=(W_p+I_{p^2})\big(B'_i\otimes I_p\big)\mathrm{vec}\big[\Lambda'-F'\big].\end{aligned}$$

From $\nu=\mathrm{vec}[\Lambda'-F']$ we obtain

$$\begin{aligned}&\frac{1}{2}D_p^+\big[\mathrm{vec}\big[B'_i(\Lambda-F)\big]+\mathrm{vec}\big[(\Lambda-F)'B_i\big]\big]\\ &\quad=\frac{1}{2}D_p^+(I_{p^2}+W_p)\big(B'_i\otimes I_p\big)\nu=N_p\big(B'_i\otimes I_p\big)\nu.\end{aligned}$$

(b) From the properties of the vec operator and the commutation matrix $W_{p,q}$, we get

$$\mathrm{vec}\big[C'_i(\Lambda-F)\big]=W_{p,q}\,\mathrm{vec}\big[(\Lambda-F)'C_i\big]=W_{p,q}\big(C'_i\otimes I_p\big)\nu.$$

C.3. *Derivation of Equation (12)*

We use $\mathrm{vec}[\beta'_{3,i}]=(\mathrm{vec}[\{N_p(B'_i\otimes I_p)\}']',\mathrm{vec}[\{W_{p,q}(C'_i\otimes I_p)\}']')'$ from Equation (11).

(a) By MN, Theorem 2, p. 35 and Exercise 1, p. 56, and by writing $I_{pK}=I_K\otimes I_p$, we obtain

$$\begin{aligned}&\mathrm{vec}\big[N_p\big(B'_i\otimes I_p\big)\big]\\ &\quad=(I_{pK}\otimes N_p)\,\mathrm{vec}\big[B'_i\otimes I_p\big]\\ &\quad=(I_{pK}\otimes N_p)\big\{I_K\otimes\big[(W_p\otimes I_p)\big(I_p\otimes\mathrm{vec}[I_p]\big)\big]\big\}\,\mathrm{vec}\big[B'_i\big]\\ &\quad=\big\{I_K\otimes\big[(I_p\otimes N_p)(W_p\otimes I_p)\big(I_p\otimes\mathrm{vec}[I_p]\big)\big]\big\}\,\mathrm{vec}\big[B'_i\big].\end{aligned}$$

Moreover, $\mathrm{vec}[\{N_p(B_i' \otimes I_p)\}'] = W_{p(p+1)/2,\,pK}\,\mathrm{vec}[N_p(B_i' \otimes I_p)]$.

(b) Similarly, $\mathrm{vec}[W_{p,q}(C_i' \otimes I_p)] = \{I_K \otimes [(I_p \otimes W_{p,q})(W_{p,q} \otimes I_p) \times (I_q \otimes \mathrm{vec}[I_p])]\}\,\mathrm{vec}[C_i']$ and $\mathrm{vec}[\{W_{p,q}(C_i' \otimes I_p)\}'] = W_{pq,\,pK}\,\mathrm{vec}[W_{p,q}(C_i' \otimes I_p)]$.

By combining (a) and (b), the conclusion follows.

APPENDIX D: PROOFS OF STATEMENTS AND TECHNICAL LEMMAS

D.1. *Proof of Lemma 2*

Let vector $(z_1, \dots, z_n)$ be such that $\sum_i z_i^2 = 1$. From Equation (25), we have

$$\begin{aligned} (35)\qquad &\sum_i \sum_j z_i [\Sigma_{\tilde{\varepsilon},1,n}]_{i,j} z_j \\ &= \sum_k \sum_l \sum_i \sum_j z_{k,i}^* z_{l,j}^* \,\mathrm{Cov}\big(\varepsilon\big[G_k^{-1}(\gamma_i)\big], \varepsilon\big[G_l^{-1}(\gamma_j)\big] | \mathcal{F}_0\big), \end{aligned}$$

where $z_{k,i}^* = w_k[G_k^{-1}(\gamma_i)]z_i$. Now, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned} &\sum_i \sum_j z_{k,i}^* z_{l,j}^* \,\mathrm{Cov}\big(\varepsilon\big[G_k^{-1}(\gamma_i)\big], \varepsilon\big[G_l^{-1}(\gamma_j)\big] | \mathcal{F}_0\big) \\ &= \mathrm{Cov}\bigg(\sum_i z_{k,i}^* \varepsilon\big[G_k^{-1}(\gamma_i)\big], \sum_j z_{l,j}^* \varepsilon\big[G_l^{-1}(\gamma_j)\big] \bigg| \mathcal{F}_0\bigg) \\ &\le V\bigg(\sum_i z_{k,i}^* \varepsilon\big[G_k^{-1}(\gamma_i)\big] \bigg| \mathcal{F}_0\bigg)^{1/2} V\bigg(\sum_j z_{l,j}^* \varepsilon\big[G_l^{-1}(\gamma_j)\big] \bigg| \mathcal{F}_0\bigg)^{1/2} \\ &= \bigg(\sum_i \sum_j z_{k,i}^* z_{k,j}^* \,\mathrm{Cov}\big(\varepsilon\big[G_k^{-1}(\gamma_i)\big], \varepsilon\big[G_k^{-1}(\gamma_j)\big] | \mathcal{F}_0\big)\bigg)^{1/2} \\ &\quad \times \bigg(\sum_i \sum_j z_{l,i}^* z_{l,j}^* \,\mathrm{Cov}\big(\varepsilon\big[G_l^{-1}(\gamma_i)\big], \varepsilon\big[G_l^{-1}(\gamma_j)\big] | \mathcal{F}_0\big)\bigg)^{1/2}. \end{aligned}$$

Moreover,

$$\begin{aligned} &\sum_i \sum_j z_{k,i}^* z_{k,j}^* \,\mathrm{Cov}\big(\varepsilon\big[G_k^{-1}(\gamma_i)\big], \varepsilon\big[G_k^{-1}(\gamma_j)\big] | \mathcal{F}_0\big) \\ &\le \bigg(\sum_i \big(z_{k,i}^*\big)^2\bigg) \mathrm{eig}_{\max}\big(\Sigma_{\varepsilon,1,n}(G_k)\big) \\ &\le \bar{w}_k^2 \,\mathrm{eig}_{\max}\big(\Sigma_{\varepsilon,1,n}(G_k)\big). \end{aligned}$$

Thus, for any vector $(z_1, \dots, z_n)$ such that $\sum_i z_i^2 = 1$, we have

$$\sum_i \sum_j z_i [\Sigma_{\tilde{\varepsilon},1,n}]_{i,j} z_j$$
$$\leq \sum_k \sum_l \bar{w}_k \bar{w}_l \operatorname{eig}_{\max}\big(\Sigma_{\varepsilon,1,n}(G_k)\big)^{1/2} \operatorname{eig}_{\max}\big(\Sigma_{\varepsilon,1,n}(G_l)\big)^{1/2}.$$

Since the largest eigenvalue of a symmetric matrix is equal to the sup of the associated quadratic form w.r.t. vectors with unit length, the conclusion follows.

D.2. *Proof of Lemma 3(iii)*

We have $\hat{w}_i - w_i = \mathbf{1}_i^\chi((\operatorname{diag}[\hat{v}_i])^{-1} - (\operatorname{diag}[v_i])^{-1}) + (\mathbf{1}_i^\chi - 1)(\operatorname{diag}[v_i]^{-1})$ and $(\operatorname{diag}[\hat{v}_i])^{-1} - (\operatorname{diag}[v_i])^{-1} = -(\operatorname{diag}[\hat{v}_i])^{-1} \operatorname{diag}[\hat{v}_i - v_i](\operatorname{diag}[v_i])^{-1}$. Since $\|(\operatorname{diag}[v_i])^{-1}\|$ is uniformly lower bounded from part (ii), we have $\frac{1}{n}\sum_i \|\hat{w}_i - w_i\| \leq C\frac{1}{n}\sum_i \mathbf{1}_i^\chi \frac{\|\hat{v}_i - v_i\|}{C - \|\hat{v}_i - v_i\|} + C\frac{1}{n}\sum_i (1 - \mathbf{1}_i^\chi)$. The second term in the RHS is $o_p(1)$ from Lemma 7. To prove that the first term is $o_p(1)$, it is sufficient to show

$$\sup_i \mathbf{1}_i^\chi \|\hat{v}_i - v_i\| = o_p(1). \tag{36}$$

We use Equation (30). Since $\hat{\nu}_1 - \nu = O_p(T^{-c})$, for some $c > 0$ (by repeating the proof of Proposition 3 with known weights equal to 1), $\mathbf{1}_i^\chi \|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$, $\mathbf{1}_i^\chi \tau_{i,T} \leq \chi_{2,T}$, $\|S_{ii}\| \leq M$, and by using Assumption B.5, the uniform bound in (36) follows if we prove

$$\sup_i \mathbf{1}_i^\chi \|\hat{S}_{ii} - S_{ii}\| = O_p\big(T^{-c}\big), \tag{37}$$

$$\sup_i \mathbf{1}_i^\chi \big\|\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}\big\| = O_p\big(T^{-c}\big), \tag{38}$$

$$\sup_i \mathbf{1}_i^\chi |\tau_{i,T} - \tau_i| = O_p\big(T^{-c}\big), \tag{39}$$

for some $c > 0$. To prove the uniform bound (37), we use Equation (32). As in the proof of Lemma 3(i), we have $\sup_i T^{-1/2}\|Y_{i,T}\| = O_{p,\log}(T^{-\eta/2})$ from Assumption B.1(c), and similarly $\sup_i T^{-1/2}\|W_{1,i,T} + W_{2,i,T}\| = O_{p,\log}(T^{-\eta/2})$ and $\sup_i T^{-1/2}\|W_{3,i,T}\| = O_p(T^{-\eta/2})$, from Assumptions B.1(e) and (f), respectively. Moreover, $\|\hat{Q}_{x,i}^{(4)}\| \leq M$, $\mathbf{1}_i^\chi \|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$ and $\mathbf{1}_i^\chi \tau_{i,T} \leq \chi_{2,T}$. Thus, from Assumption B.5, bound (37) follows. To prove (38), we use Equation (33) where $W_{i,T}$ is such that $\sup_i \|W_{i,T}\| = O_{p,\log}(T^{-\eta/2})$ from Assumption B.1(b). Finally, (39) follows from $|\tau_{i,T} - \tau_i| \leq \tau_{i,T}\tau_i|\frac{1}{T}\sum_t (I_{i,t} - E[I_{i,t}|\gamma_i])|$, $\mathbf{1}_i^\chi \tau_{i,T} \leq \chi_{2,T}$,

$\tau_i \leq M$, and by using $\sup_i |\frac{1}{T}\sum_t (I_{i,t} - E[I_{i,t}|\gamma_i])| = O_{p,\log}(T^{-\eta/2})$ from Assumption B.1(d).

D.3. *Proof of Lemma 4*

By applying MN, Theorem 2, p. 35, Theorem 10, p. 55, and using $W_{n,1} = I_n$, we have

$$\begin{aligned}
Ab &= \text{vec}[Ab] \\
&= (b' \otimes A)\,\text{vec}[I_n] \\
&= \text{vec}\big[(b' \otimes A)\,\text{vec}[I_n]\big] \\
&= \big(\text{vec}[I_n]' \otimes I_m\big)\,\text{vec}\big[b' \otimes A\big] \\
&= \big(\text{vec}[I_n]' \otimes I_m\big)(I_{n^2} \otimes I_m)\,\text{vec}\big[\text{vec}[A]b'\big] \\
&= \big(\text{vec}[I_n]' \otimes I_m\big)\,\text{vec}\big[\text{vec}[A]b'\big].
\end{aligned}$$

D.4. *Proof of Lemma 6*

D.4.1. *Part (i)*

Let us write I_{131} as $I_{131} = (I_{d_1} \otimes E_2')\tilde{I}_{131}$ and

$$\begin{aligned}
\tilde{I}_{131} &= \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2\big(\hat{w}_i \otimes \big[\hat{Q}_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)\hat{Q}_{x,i}^{-1}\big]\big) \\
&= \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2\big(\hat{w}_i \otimes \big[Q_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)Q_{x,i}^{-1}\big]\big) \\
&\quad + \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2\big(\hat{w}_i \otimes \big[\big(\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}\big)\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)Q_{x,i}^{-1}\big]\big) \\
&\quad + \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2\big(\hat{w}_i \otimes \big[Q_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)\big(\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}\big)\big]\big) \\
&\quad + \frac{1}{\sqrt{n}}\sum_i \tau_{i,T}^2\big(\hat{w}_i \otimes \big[\big(\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}\big)\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big) \\
&\quad \times \big(\hat{Q}_{x,i}^{-1} - Q_{x,i}^{-1}\big)\big]\big) \\
&=: I_{1311} + I_{1312} + I_{1312}' + I_{1313}.
\end{aligned}$$

We control the terms separately.

Proof that $I_{1311} = \frac{1}{\sqrt{n}}\sum_i \tau_i^2(w_i \otimes [Q_{x,i}^{-1}(Y_{i,T}Y_{i,T}' - S_{ii,T})Q_{x,i}^{-1}]) + O_{p,\log}(\sqrt{n}/T) = O_p(1) + O_{p,\log}(\sqrt{n}/T)$. We use a decomposition similar to term I_{111} in the proof of Lemma 5:

$$
\begin{aligned}
I_{1311} = & \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \big(w_i \otimes \big[Q_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)Q_{x,i}^{-1}\big]\big) \\
& + \frac{1}{\sqrt{n}} \sum_i \tau_i^2 \big(\mathbf{1}_i^{\chi} - 1\big)\big(w_i \otimes \big[Q_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)Q_{x,i}^{-1}\big]\big) \\
& + \frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^{\chi}\big(\tau_{i,T}^2 - \tau_i^2\big)\big(w_i \otimes \big[Q_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)Q_{x,i}^{-1}\big]\big) \\
& + \frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^{\chi}\tau_{i,T}^2\big(\big(\mathrm{diag}[\hat{v}_i]^{-1} - \mathrm{diag}[v_i]^{-1}\big) \\
& \otimes \big[Q_{x,i}^{-1}\big(Y_{i,T}Y_{i,T}' - S_{ii,T}\big)Q_{x,i}^{-1}\big]\big) \\
=: & \, I_{13111} + I_{13112} + I_{13113} + I_{13114}.
\end{aligned}
$$

To simplify the notation, let us treat $x_{i,t}$ as a scalar. We first prove $I_{13111} = O_p(1)$. We have

$$
\begin{aligned}
& E\big[I_{13111}^2 | \mathcal{F}_{\underline{T}}, \big\{I_{\underline{T}}(\gamma_i), \gamma_i\big\}\big] \\
& = \frac{1}{n} \sum_{i,j} w_i w_j \tau_i^2 \tau_j^2 Q_{x,i}^{-2} Q_{x,j}^{-2} \operatorname{cov}\big(Y_{i,T}^2, Y_{j,T}^2 | \mathcal{F}_{\underline{T}}, I_{\underline{T}}(\gamma_i), I_{\underline{T}}(\gamma_j), \gamma_i, \gamma_j\big) \\
& = \frac{1}{nT^2} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} w_i w_j \tau_i^2 \tau_j^2 Q_{x,i}^{-2} Q_{x,j}^{-2} \\
& \quad \times \operatorname{cov}(\varepsilon_{i,t_1}\varepsilon_{i,t_2}, \varepsilon_{j,t_3}\varepsilon_{j,t_4} | \mathcal{F}_{\underline{T}}, \gamma_i, \gamma_j) I_{i,t_1} I_{i,t_2} I_{j,t_3} I_{j,t_4} x_{i,t_1} x_{i,t_2} x_{j,t_3} x_{j,t_4}.
\end{aligned}
$$

From Assumptions B.3(b) and B.4, it follows that $E[I_{13111}^2] = O(1)$. Hence, $I_{13111} = O_p(1)$. We can prove that $I_{13112} = o_p(1)$ and $I_{13113} = o_p(1)$ by using arguments similar to terms I_{1112} and I_{1113} in the proof of Lemma 5. Finally, let us prove that $I_{13114} = O_{p,\log}(\sqrt{n}/T)$. Similarly to I_{1114} in the proof of Lemma 5, we use

$$
\hat{v}_i^{-1} - v_i^{-1} = -v_i^{-2}(\hat{v}_i - v_i) + \hat{v}_i^{-1} v_i^{-2}(\hat{v}_i - v_i)^2, \tag{40}
$$

and Equation (30). We focus on the term

$$
I_{131141} = -\frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^{\chi} v_i^{-2} \tau_{i,T}^3 C_{\hat{\nu}_1}' \hat{Q}_{x,i}^{-1}(\hat{S}_{ii} - S_{ii})\hat{Q}_{x,i}^{-1} C_{\hat{\nu}_1} Q_{x,i}^{-2}\big(Y_{i,T}^2 - S_{ii,T}\big);
$$

the other contributions to I_{13114} can be controlled similarly. Now, we use Equation (32). We have

$$
\begin{aligned}
I_{131141} &= -\frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\tau_{i,T}^4 C_{\hat{\nu}_1}'\hat{Q}_{x,i}^{-1}W_{1,i,T}\hat{Q}_{x,i}^{-1}C_{\hat{\nu}_1}Q_{x,i}^{-2}\big(Y_{i,T}^2 - S_{ii,T}\big)\\
&\quad - \frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\tau_{i,T}^4 C_{\hat{\nu}_1}'\hat{Q}_{x,i}^{-1}W_{2,i,T}\hat{Q}_{x,i}^{-1}C_{\hat{\nu}_1}Q_{x,i}^{-2}\big(Y_{i,T}^2 - S_{ii,T}\big)\\
&\quad + 2\frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\tau_{i,T}^5 C_{\hat{\nu}_1}'\hat{Q}_{x,i}^{-1}W_{3,i,T}\hat{Q}_{x,i}^{-1}Y_{i,T}\\
&\quad \times \hat{Q}_{x,i}^{-1}C_{\hat{\nu}_1}Q_{x,i}^{-2}\big(Y_{i,T}^2 - S_{ii,T}\big)\\
&\quad - \frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\tau_{i,T}^6 C_{\hat{\nu}_1}'\hat{Q}_{x,i}^{-1}\hat{Q}_{x,i}^{(4)}\hat{Q}_{x,i}^{-1}Y_{i,T}^2\\
&\quad \times \hat{Q}_{x,i}^{-2}C_{\hat{\nu}_1}Q_{x,i}^{-2}\big(Y_{i,T}^2 - S_{ii,T}\big)\\
&=: -C_{\hat{\nu}_1}'(I_{1311411} + I_{1311412} + I_{13211413} + I_{1311414})C_{\hat{\nu}_1}.
\end{aligned}
$$

Let us focus on term $I_{1311411}$ and prove that it is $O_{p,\log}(\sqrt{n}/T)$. We have

$$
\begin{aligned}
I_{1311411} &= \frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\tau_{i,T}^4 \hat{Q}_{x,i}^{-2}Q_{x,i}^{-2}W_{1,i,T}Y_{i,T}^2\\
&\quad - \frac{1}{\sqrt{nT}}\sum_i \mathbf{1}_i^{\chi} v_i^{-2}\tau_{i,T}^4 \hat{Q}_{x,i}^{-2}Q_{x,i}^{-2}W_{1,i,T}S_{ii,T}\\
&=: I_{13114111} + I_{13114112}.
\end{aligned}
$$

Term $I_{13114111}$ is such that

$$
\begin{aligned}
&\big|E\big[I_{13114111}|\mathcal{F}_{\underline{T}}, \big\{I_{\underline{T}}(\gamma_i), \gamma_i\big\}\big]\big|\\
&\quad \le \frac{C\chi_{1,T}^4\chi_{2,T}^4}{\sqrt{n}T^2}\sum_i\sum_{t_1,t_2,t_3}\big|E[\eta_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{i,t_3}|\mathcal{F}_{\underline{T}}, \gamma_i]\big|,
\end{aligned}
$$

and

$$
\begin{aligned}
&V\big[I_{13114111}|\mathcal{F}_{\underline{T}}, \big\{I_{\underline{T}}(\gamma_i), \gamma_i\big\}\big]\\
&\quad \le \frac{C\chi_{1,T}^8\chi_{2,T}^8}{nT^4}\sum_{i,j}\sum_{t_1,\ldots,t_6}\big|\mathrm{cov}(\eta_{i,t_1}\varepsilon_{i,t_2}\varepsilon_{i,t_3}, \eta_{j,t_4}\varepsilon_{j,t_5}\varepsilon_{j,t_6}|\mathcal{F}_{\underline{T}}, \gamma_i, \gamma_j)\big|.
\end{aligned}
$$

From Assumptions B.2, B.3(f), and B.5, we get $E[I_{13114111}] = O_{\log}(\sqrt{n}/T)$ and $V[I_{13114111}] = o(1)$, which implies $I_{13114111} = O_{p,\log}(\sqrt{n}/T)$. The other terms making I_{13114} can be controlled similarly, and we get $I_{13114} = O_{p,\log}(\sqrt{n}/T)$.

Proof that $I_{1312} = o_p(1)$. We have

$$\begin{aligned} I_{1312} &= \frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^{\chi} \tau_{i,T}^2 \big(\text{diag}[v_i]^{-1} \\ &\quad \otimes \big[\big(\hat{Q}_{x,i}^{-1} - \hat{Q}_{x,i}^{-1} \big) \big(Y_{i,T} Y_{i,T}' - S_{ii,T} \big) Q_{x,i}^{-1} \big] \big) \\ &\quad + \frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^{\chi} \tau_{i,T}^2 \big(\big(\text{diag}[\hat{v}_i]^{-1} - \text{diag}[v_i]^{-1} \big) \\ &\quad \otimes \big[\big(\hat{Q}_{x,i}^{-1} - \hat{Q}_{x,i}^{-1} \big) \big(Y_{i,T} Y_{i,T}' - S_{ii,T} \big) Q_{x,i}^{-1} \big] \big) \\ &=: I_{13121} + I_{13122}. \end{aligned}$$

We focus on term I_{13121}, use Equation (33), and treat $x_{i,t}$ as a scalar to ease notation. We have $I_{13121} = -\frac{1}{\sqrt{n}} \sum_i \mathbf{1}_i^{\chi} v_i^{-1} \tau_{i,T}^3 \hat{Q}_{x,i}^{-1} W_{i,T} Q_{x,i}^{-2} (Y_{i,T}^2 - S_{ii,T})$. Then,

$$\begin{aligned} &E\big[\|I_{13121}\|^2 | \mathcal{F}_{\underline{T}}, \{ I_{\underline{T}}(\gamma_i), \gamma_i \} \big] \\ &\quad \le \frac{C \chi_{1,T}^4 \chi_{2,T}^6}{nT^2} \sum_{i,j} \sum_{t_1,\dots,t_4} \|W_{i,T}\| \|W_{j,T}\| \\ &\qquad \times \big| \text{cov}(\varepsilon_{i,t_1} \varepsilon_{i,t_2}, \varepsilon_{j,t_3} \varepsilon_{j,t_4} | \mathcal{F}_{\underline{T}}, \gamma_i, \gamma_j) \big|. \end{aligned}$$

By the Cauchy–Schwarz inequality, we get

$$\begin{aligned} &E\big[\|I_{13121}\|^2 | \{\gamma_i\} \big] \\ &\quad \le C \chi_{1,T}^4 \chi_{2,T}^6 \sup_i E\big[\|W_{i,T}\|^4 | \gamma_i \big]^{1/2} \\ &\qquad \times \frac{1}{nT^2} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} E\big[\big| \text{cov}(\varepsilon_{i,t_1} \varepsilon_{i,t_2}, \varepsilon_{j,t_3} \varepsilon_{j,t_4} | \mathcal{F}_{\underline{T}}, \gamma_i, \gamma_j) \big|^2 | \gamma_i, \gamma_j \big]^{1/2}. \end{aligned}$$

From Assumptions B.1(b), B.3(b), B.4(a), and B.5, we deduce $E[\|I_{13121}\|^2] = o(1)$, which implies $I_{13121} = o_p(1)$. Similar arguments can be used to prove that the other terms making I_{1312} are $o_p(1)$.

Proof that $I_{1313} = o_p(1)$. This step uses arguments similar to those for I_{1312}.

D.4.2. *Part (ii)*

Let us treat $x_{i,t}$ as a scalar to ease notation. We have $I_{132} = (I_{d_1} \otimes E_2') \tilde{I}_{132}$, where $\tilde{I}_{132} = \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_{i,T}^2 \hat{Q}_{x,i}^{-1} W_{1,i,T} \hat{Q}_{x,i}^{-1}$, and $W_{1,i,T}$ is as in Equation (32).

Write

$$\tilde{I}_{132} = \frac{1}{\sqrt{nT}} \sum_i \mathbf{1}_i^\chi v_i^{-1} \tau_{i,T}^2 \hat{Q}_{x,i}^{-1} W_{1,i,T} \hat{Q}_{x,i}^{-1} + \frac{1}{\sqrt{nT}} \sum_i \mathbf{1}_i^\chi \big(\hat{v}_i^{-1} - v_i^{-1}\big) \tau_{i,T}^2 \hat{Q}_{x,i}^{-1} W_{1,i,T} \hat{Q}_{x,i}^{-1} =: I_{1321} + I_{1322}.$$

Let us first consider I_{1321}. We have

$$E\big[\|I_{1321}\|^2 | \mathcal{F}_{\underline{T}}, \{I_{\underline{T}}(\gamma_i), \gamma_i\}\big] \leq C\chi_{1,T}^8 \chi_{2,T}^4 \frac{1}{nT^2} \sum_{i,j} \sum_{t_1,t_2} \big|\mathrm{cov}(\eta_{i,t_1}, \eta_{j,t_2} | \mathcal{F}_{\underline{T}}, \gamma_i, \gamma_j)\big|.$$

From Assumptions B.3(a) and B.5, it follows that $E[\|I_{1321}\|^2] = O_{\log}(1/T)$, and thus $I_{1321} = O_{p,\log}(1/\sqrt{T})$.

Let us now consider term I_{1322}. We use Equation (40), and plug in the decompositions (30) and (32). We focus on term $C_{\hat{\nu}_1}^2 I_{13221}$ of the resulting expansion, where $I_{13221} = -\frac{1}{\sqrt{nT}} \sum_i \mathbf{1}_i^\chi v_i^{-2} \tau_{i,T}^4 \hat{Q}_{x,i}^{-4} W_{1,i,T}^2$. The other terms can be treated similarly. We have

$$E\big[I_{13221} | \mathcal{F}_{\underline{T}}, \{I_{\underline{T}}(\gamma_i), \gamma_i\}\big] \leq C\chi_{1,T}^8 \chi_{2,T}^4 \frac{1}{\sqrt{n}T^2} \sum_i \sum_{t_1,t_2} \big|\mathrm{cov}\big(\varepsilon_{i,t_1}^2, \varepsilon_{i,t_2}^2 | \mathcal{F}_{\underline{T}}, \gamma_i\big)\big|,$$

and

$$V\big[I_{13221} | \mathcal{F}_{\underline{T}}, \{I_{\underline{T}}(\gamma_i), \gamma_i\}\big] \leq C\chi_{1,T}^{16} \chi_{2,T}^8 \frac{1}{nT^4} \sum_{i,j} \sum_{t_1,t_2,t_3,t_4} \big|\mathrm{cov}(\eta_{i,t_1}\eta_{i,t_2}, \eta_{j,t_3}\eta_{j,t_4} | \mathcal{F}_{\underline{T}}, \gamma_i, \gamma_j)\big|.$$

From Assumptions B.3(a) and B.5, it follows that $E[I_{13221}] = O_{\log}(\sqrt{n}/T)$. By Assumptions B.3(d) and B.5, we can prove that $V[I_{13221}] = o(1)$, and it follows that $I_{13221} = O_p(\sqrt{n}/T)$.

D.4.3. *Part (iii)*

We have $I_{133} = (I_{d_1} \otimes E_2')\tilde{I}_{133}$, where $I_{133} = -\frac{2}{\sqrt{nT}} \sum_i \hat{w}_i \tau_{i,T}^3 \hat{Q}_{x,i}^{-3} W_{3,i,T} Y_{i,T} + \frac{1}{\sqrt{nT}} \sum_i \hat{w}_i \tau_{i,T}^4 \hat{Q}_{x,i}^{-4} \hat{Q}_{x,i}^{(4)} Y_{i,T}^2$ and $W_{3,i,T}$ and $\hat{Q}_{x,i}^{(4)}$ are as in Equation (32) and we treat $x_{i,t}$ as a scalar to ease notation. By similar arguments as in part (ii), we can prove that $I_{133} = O_{p,\log}(\sqrt{n}/T)$.

D.4.4. *Part (iv)*

The statement follows from Lemma 3(ii)–(iii), $\mathbf{1}_i^\chi \tau_{i,T} \leq \chi_{2,T}$, $\mathbf{1}_i^\chi \|\hat{Q}_{x,i}^{-1}\| \leq C\chi_{1,T}^2$, bound (37), $\|S_{ii}\| \leq M$, and Assumption B.5.

D.4.5. *Part (v)*

The statement follows from Equation (28), Lemma 3(iv), $I_{11} = O_p(1)$, and $\frac{1}{n}\sum_i \hat{w}_i \tau_{i,T}^2 \hat{Q}_{x,i}^{-1} Y_{i,T} Y_{i,T}' \hat{Q}_{x,i}^{-1} = O_{p,\log}(1)$.

D.5. *Proof of Lemma 7*

We have $P[\mathbf{1}_i^\chi = 0] \leq P[\tau_{i,T} \geq \chi_{2,T}] + P[CN(\hat{Q}_{x,i}) \geq \chi_{1,T}] =: P_{1,nT} + P_{2,nT}$. Let us first control $P_{1,nT}$. We have $P_{1,nT} \leq P[\frac{1}{T}\sum_t I_{i,t} \leq \chi_{2,T}^{-1}] \leq P[\frac{1}{T}\sum_t (I_{i,t} - \tau_i^{-1}) \leq \chi_{2,T}^{-1} - M^{-1}]$, where we use $\tau_i \leq M$ for all i (Assumption B.4(c)). Then, for $0 < \delta < M^{-1}/2$ and T large such that $M^{-1} - \chi_{2,T}^{-1} > \delta$, we get the upper bound $P_{1,nT} \leq P[|\frac{1}{T}\sum_t (I_{i,t} - \tau_i^{-1})| \geq \delta]$. By using that $\tau_i^{-1} = E[I_{i,t}|\gamma_i]$, and $P[|\frac{1}{T}\sum_t (I_{i,t} - \tau_i^{-1})| \geq \delta] = E[P[|\frac{1}{T}\sum_t (I_{i,t} - E[I_{i,t}|\gamma_i])| \geq \delta|\gamma_i]] \leq \sup_{\gamma \in [0,1]} P[|\frac{1}{T}\sum_t (I_t(\gamma) - E[I_t(\gamma)])| \geq \delta]$, from Assumption B.1(d), it follows that $P_{1,nT} = O(T^{-\bar{b}})$, $\forall \bar{b} > 0$.

Let us now consider $P_{2,nT}$. By using $\|\hat{Q}_{x,i}\| \leq M$ (Assumption B.4(a)), we get $\mathrm{eig}_{\max}(\hat{Q}_{x,i}) \leq M$, and thus $CN(\hat{Q}_{x,i}) \leq M^{1/2}[\mathrm{eig}_{\min}(\hat{Q}_{x,i})]^{-1/2}$. Hence $P_{2,nT} \leq P[\mathrm{eig}_{\min}(\hat{Q}_{x,i}) \leq M/\chi_{1,T}^2]$. By using that $\mathrm{eig}_{\min}(\hat{Q}_{x,i}) \geq \mathrm{eig}_{\min}(Q_{x,i}) - \|\hat{Q}_{x,i} - Q_{x,i}\|$, we get $P_{2,nT} \leq P[\|\hat{Q}_{x,i} - Q_{x,i}\| \geq \mathrm{eig}_{\min}(Q_{x,i}) - M/\chi_{1,T}^2]$. Now, let $\delta > 0$ be such that $\mathrm{eig}_{\min}(Q_{x,i}) - M/\chi_{1,T}^2 > \delta$ uniformly in i for large T (see Assumption B.4(d)). Then, by using $P[\|\hat{Q}_{x,i} - Q_{x,i}\| \geq \delta] \leq P[|\frac{1}{T}\sum_t I_{i,t}(x_{i,t}x_{i,t} - Q_{x,i})| \geq \sqrt{\delta}] + P[\tau_{i,T} \geq \sqrt{\delta}]$, we get $P_{2,nT} \leq P[|\frac{1}{T}\sum_t I_{i,t}(x_{i,t}x_{i,t} - Q_{x,i})| \geq \sqrt{\delta}] + O(T^{-\bar{b}})$. The first term in the RHS is $O(T^{-\bar{b}})$ by using $P[|\frac{1}{T}\sum_t I_{i,t}(x_{i,t}x_{i,t} - Q_{x,i})| \geq \sqrt{\delta}] \leq \sup_{\gamma \in [0,1]} P[|\frac{1}{T}\sum_t I_t(\gamma)(x_t(\gamma)x_t(\gamma)' - E[x_t(\gamma)x_t(\gamma)'])| \geq \sqrt{\delta}]$ and Assumption B.1(b). Then, $P_{2,nT} = O(T^{-\bar{b}})$, for any $\bar{b} > 0$.

D.6. *Proof of Lemma 8*

Let $W_T(\gamma) := \frac{1}{T}\sum_t (I_t(\gamma) - E[I_t(\gamma)])$ and $r_T := T^{-a}$ for $0 < a < \eta/2$. Since $|W_T(\gamma)| \leq 1$ for all $\gamma \in [0,1]$, and from Assumption B.1(d), we have

$$\sup_{\gamma \in [0,1]} E\big[\big|W_T(\gamma)\big|^4\big]$$

$$\leq \sup_{\gamma \in [0,1]} E\big[\big|W_T(\gamma)\big|\big] = \sup_{\gamma \in [0,1]} \int_0^1 P\big[\big|W_T(\gamma)\big| \geq \delta\big]\, d\delta$$

$$\leq r_T + \sup_{\gamma\in[0,1]} \int_{r_T}^{1} P\big[\big|W_T(\gamma)\big| \geq \delta\big]\, d\delta$$
$$\leq r_T + C_1 T \int_{r_T}^{1} \exp\big\{-C_2\delta^2 T^{\eta}\big\}\, d\delta + C_3 \exp\big\{-C_4 T^{\bar\eta}\big\} \int_{r_T}^{1} \frac{1}{\delta}\, d\delta$$
$$\leq r_T + C_1 T \exp\big\{-C_2 r_T^2 T^{\eta}\big\} + C_3 \exp\big\{-C_4 T^{\bar\eta}\big\} \log(1/r_T) = o(1).$$

D.7. *Proof of Lemma 9*

By definition of $\tilde S_{ij}$, we have

$$\frac{1}{n}\sum_{i,j} \|\tilde S_{ij} - S_{ij}\|$$
$$= \frac{1}{n}\sum_{i,j} \|\hat S_{ij}\mathbf{1}_{\{\|\hat S_{ij}\|\geq\kappa\}} - S_{ij}\|$$
$$\leq \frac{1}{n}\sum_{i,j} \|S_{ij}\mathbf{1}_{\{\|S_{ij}\|\geq\kappa\}} - S_{ij}\| + \frac{1}{n}\sum_{i,j} \|\hat S_{ij}\mathbf{1}_{\{\|\hat S_{ij}\|\geq\kappa\}} - S_{ij}\mathbf{1}_{\{\|S_{ij}\|\geq\kappa\}}\|$$
$$=: I_{31} + I_{32}.$$

By Assumption A.4,

$$I_{31} = \frac{1}{n}\sum_{i,j} \|S_{ij}\|\mathbf{1}_{\{\|S_{ij}\|<\kappa\}} \leq \max_i \sum_j \|S_{ij}\|^{\bar q}\kappa^{1-\bar q} \tag{41}$$
$$\leq \kappa^{1-\bar q} c_0(n) = O_p\big(\kappa^{1-\bar q} n^{\bar{\bar\delta}}\big),$$

where $c_0(n) := \max_i \sum_j \|S_{ij}\|^{\bar q} = O_p(n^{\bar{\bar\delta}})$.

Let us now consider I_{32}:

$$I_{32} = \frac{1}{n}\sum_{i,j} \|\hat S_{ij}\|\mathbf{1}_{\{\|\hat S_{ij}\|\geq\kappa, \|S_{ij}\|<\kappa\}} + \frac{1}{n}\sum_{i,j} \|S_{ij}\|\mathbf{1}_{\{\|\hat S_{ij}\|<\kappa, \|S_{ij}\|\geq\kappa\}}$$
$$+ \frac{1}{n}\sum_{i,j} \|\hat S_{ij} - S_{ij}\|\mathbf{1}_{\{\|\hat S_{ij}\|\geq\kappa, \|S_{ij}\|\geq\kappa\}}$$
$$\leq \max_i \sum_j \|\hat S_{ij}\|\mathbf{1}_{\{\|\hat S_{ij}\|\geq\kappa, \|S_{ij}\|<\kappa\}} + \max_i \sum_j \|S_{ij}\|\mathbf{1}_{\{\|\hat S_{ij}\|<\kappa, \|S_{ij}\|\geq\kappa\}}$$
$$+ \max_i \sum_j \|\hat S_{ij} - S_{ij}\|\mathbf{1}_{\{\|\hat S_{ij}\|\geq\kappa, \|S_{ij}\|\geq\kappa\}}$$
$$=: I_{33} + I_{34} + I_{35}.$$

From Assumption A.4, we have

$$I_{35} \leq \max_{i,j} \|\hat{S}_{ij} - S_{ij}\| \max_i \sum_j \|S_{ij}\|^{\bar{q}} \kappa^{-\bar{q}} = O_p\big(\psi_{nT} c_0(n) \kappa^{-\bar{q}}\big). \tag{42}$$

Let us study I_{33}:

$$\begin{aligned} I_{33} &\leq \max_i \sum_j \|\hat{S}_{ij} - S_{ij}\| \mathbf{1}_{\{\|\hat{S}_{ij}\| \geq \kappa, \|S_{ij}\| < \kappa\}} + \max_i \sum_j \|S_{ij}\| \mathbf{1}_{\{\|S_{ij}\| < \kappa\}} \\ &=: I_{36} + I_{37}. \end{aligned}$$

By Assumption A.4,

$$I_{37} \leq \kappa^{1-\bar{q}} c_0(n). \tag{43}$$

Now take $v \in (0, 1)$. Let $N_i(\epsilon) := \sum_j \mathbf{1}_{\{\|\hat{S}_{ij} - S_{ij}\| > \epsilon\}}$, for $\epsilon > 0$; then

$$\begin{aligned} I_{36} &= \max_i \sum_j \|\hat{S}_{ij} - S_{ij}\| \mathbf{1}_{\{\|\hat{S}_{ij}\| \geq \kappa, \|S_{ij}\| \leq v\kappa\}} \\ &\quad + \max_i \sum_j \|\hat{S}_{ij} - S_{ij}\| \mathbf{1}_{\{\|\hat{S}_{ij}\| \geq \kappa, v\kappa < \|S_{ij}\| < \kappa\}} \\ &\leq \max_{i,j} \|\hat{S}_{ij} - S_{ij}\| \max_i N_i\big((1-v)\kappa\big) + \max_{i,j} \|\hat{S}_{ij} - S_{ij}\| c_0(n) (v\kappa)^{-\bar{q}}. \end{aligned}$$

Moreover, by the Chebyschev inequality, for any positive sequence R_{nT}, we have

$$\begin{aligned} P\Big[\max_i N_i(\epsilon) \geq R_{nT}\Big] &\leq nP\big[N_i(\epsilon) \geq R_{nT}\big] \leq \frac{n}{R_{nT}} E\big[N_i(\epsilon)\big] \\ &\leq \frac{n^2}{R_{nT}} \max_{i,j} P\big[\|\hat{S}_{ij} - S_{ij}\| \geq \epsilon\big], \end{aligned}$$

which implies $\max_i N_i(\epsilon) = O_p(n^2 \max_{i,j} P[\|\hat{S}_{ij} - S_{ij}\| \geq \epsilon])$. Thus,

$$I_{36} = O_p\big(\psi_{nT} n^2 \Psi_{nT}\big((1-v)\kappa\big) + \psi_{nT} c_0(n) (v\kappa)^{-\bar{q}}\big). \tag{44}$$

Finally, we consider I_{34}. We have

$$\begin{aligned} I_{34} &\leq \max_i \sum_j \big(\|\hat{S}_{ij} - S_{ij}\| + \|\hat{S}_{ij}\|\big) \mathbf{1}_{\{\|\hat{S}_{ij}\| < \kappa, \|S_{ij}\| \geq \kappa\}} \\ &\leq \max_{i,j} \|\hat{S}_{ij} - S_{ij}\| \max_i \sum_j \mathbf{1}_{\{\|S_{ij}\| \geq \kappa\}} + \kappa \max_i \sum_j \mathbf{1}_{\{\|S_{ij}\| \geq \kappa\}} \\ &= O_p\big(\psi_{nT} c_0(n) \kappa^{-\bar{q}} + c_0(n) \kappa^{1-\bar{q}}\big). \end{aligned} \tag{45}$$

Combining (41)–(45), the result follows.

D.8. *Proof of Lemma 10*

By using $\hat{\varepsilon}_{i,t} = \varepsilon_{i,t} - x'_{i,t}(\hat{\beta}_i - \beta_i)$ and $\hat{S}^0_{ij} = \frac{1}{T_{ij}} \sum_t I_{ij,t}\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}$, we have

$$\begin{aligned}\hat{S}_{ij} &= \hat{S}^0_{ij} - \frac{1}{T_{ij}}\sum_t I_{ij,t}\varepsilon_{i,t}x'_{j,t}(\hat{\beta}_j - \beta_j)x_{i,t}x'_{j,t} \\ &\quad - \frac{1}{T_{ij}}\sum_t I_{ij,t}\varepsilon_{j,t}x'_{i,t}(\hat{\beta}_i - \beta_i)x_{i,t}x'_{j,t} \\ &\quad + \frac{1}{T_{ij}}\sum_t I_{ij,t}(\hat{\beta}_i - \beta_i)'x_{i,t}x'_{j,t}(\hat{\beta}_j - \beta_j)x_{i,t}x'_{j,t} \\ &=: \hat{S}^0_{ij} - A_{ij} - B_{ij} + C_{ij},\end{aligned}$$

where $A_{ij} = B_{ji}$. Then, for any i, j, we have $\|\hat{S}_{ij} - S_{ij}\| \le \|\hat{S}^0_{ij} - S_{ij}\| + \|A_{ij}\| + \|B_{ij}\| + \|C_{ij}\|$. We get, for any $\xi \ge 0$,

$$\begin{aligned}\Psi_{nT}(\xi) &\le \max_{i,j} P\left[\|\hat{S}^0_{ij} - S_{ij}\| \ge \frac{\xi}{4}\right] + \max_{i,j} P\left[\|A_{ij}\| \ge \frac{\xi}{4}\right] \\ &\quad + \max_{i,j} P\left[\|B_{ij}\| \ge \frac{\xi}{4}\right] + \max_{i,j} P\left[\|C_{ij}\| \ge \frac{\xi}{4}\right] \\ &= \Psi^0_{nT}(\xi/4) + 2P_{1,nT}(\xi/4) + P_{2,nT}(\xi/4),\end{aligned} \tag{46}$$

where $\Psi^0_{nT}(\xi/4) := \max_{i,j} P[\|\hat{S}^0_{ij} - S_{ij}\| \ge \frac{\xi}{4}]$, $P_{1,nT}(\xi/4) := \max_{i,j} P[\|A_{ij}\| \ge \frac{\xi}{4}]$, and $P_{2,nT}(\xi/4) := \max_{i,j} P[\|C_{ij}\| \ge \frac{\xi}{4}]$. Let us bound the three terms in the RHS of inequality (46).

(a) *Bound of* $\Psi^0_{nT}(\xi/4)$. We use that

$$\begin{aligned}\hat{S}^0_{ij} - S_{ij} &= \frac{1}{T_{ij}}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t} - S_{ij}\big) \\ &= \tau_{ij,T}\frac{1}{T}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t} - E[\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}|\gamma_i, \gamma_j]\big)\end{aligned}$$

and $\tau_{ij} \le M$. Then:

$$\begin{aligned}&\|\hat{S}^0_{ij} - S_{ij}\| \\ &\quad \le M\left\|\frac{1}{T}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t} - E[\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}|\gamma_i, \gamma_j]\big)\right\| \\ &\qquad + |\tau_{ij,T} - \tau_{ij}|\left\|\frac{1}{T}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t} - E[\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}|\gamma_i, \gamma_j]\big)\right\|.\end{aligned}$$

We deduce:

$$
\begin{aligned}
&\Psi^0_{nT}(\xi/4)\\
&\quad\le \max_{i,j} P\Big[\Big\| \frac{1}{T}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}\\
&\qquad - E\big[\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}|\gamma_i,\gamma_j\big]\big)\Big\| \ge \frac{\xi}{8M}\Big]\\
&\qquad + \max_{i,j} P\Big[|\tau_{ij,T}-\tau_{ij}| \ge \sqrt{\frac{\xi}{8}}\Big]\\
&\qquad + \max_{i,j} P\Big[\Big\| \frac{1}{T}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}\\
&\qquad - E\big[\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}|\gamma_i,\gamma_j\big]\big)\Big\| \ge \sqrt{\frac{\xi}{8}}\Big]\\
&\quad\le 2\max_{i,j} P\Big[\Big\| \frac{1}{T}\sum_t I_{ij,t}\big(\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}\\
&\qquad - E\big[\varepsilon_{i,t}\varepsilon_{j,t}x_{i,t}x'_{j,t}|\gamma_i,\gamma_j\big]\big)\Big\| \ge \frac{\xi}{8M}\Big]\\
&\qquad + \max_{i,j} P\Big[|\tau_{ij,T}-\tau_{ij}| \ge \sqrt{\frac{\xi}{8}}\Big]\\
&\quad =: 2P_{3,nT} + P_{4,nT},
\end{aligned}
$$

for small ξ. We use

$$
\begin{aligned}
P_{3,nT} \le \sup_{\gamma,\tilde{\gamma}\in[0,1]} P\Big[\Big\| \frac{1}{T}\sum_t I_t(\gamma)I_t(\tilde{\gamma})\big(\varepsilon_t(\gamma)\varepsilon_t(\tilde{\gamma})x_t(\gamma)x_t(\tilde{\gamma})'\\
- E\big[\varepsilon_t(\gamma)\varepsilon_t(\tilde{\gamma})x_t(\gamma)x_t(\tilde{\gamma})'\big]\big)\Big\| \ge \frac{\xi}{8M}\Big]
\end{aligned}
$$

and Assumption B.1(e) to get $P_{3,nT} \le C_1 T \exp\{-C_2^*\xi^2 T^{\eta}\} + C_3^*\xi^{-1}\exp\{-C_4 T^{\bar{\eta}}\}$, for some constants $C_1, C_2^*, C_3^*, C_4 > 0$. To bound $P_{4,nT}$, we use $\tau_{ij} \le M$ and $|\tau_{ij,T}-\tau_{ij}| \le \tau_{ij}\tau_{ij,T}|\tau_{ij,T}^{-1}-\tau_{ij}^{-1}| \le \tau_{ij}\frac{|\tau_{ij,T}^{-1}-\tau_{ij}^{-1}|}{\tau_{ij}^{-1}-|\tau_{ij,T}^{-1}-\tau_{ij}^{-1}|} \le 2M^2|\tau_{ij,T}^{-1}-\tau_{ij}^{-1}|$, if $|\tau_{ij,T}^{-1}-\tau_{ij}^{-1}| \le M^{-1}/2$. Thus, we have $P_{4,nT} \le 2\max_{i,j} P[|\tau_{ij,T}^{-1}-\tau_{ij}^{-1}| \ge \frac{1}{2M^2}\sqrt{\frac{\xi}{8}}]$, for small ξ. By using $\tau_{ij,T}^{-1} = \frac{1}{T}\sum_t I_{ij,t}$ and $\tau_{ij}^{-1} = E[I_{ij,t}|\gamma_i,\gamma_j]$, from Assumption B.1(d) we get

$$\max_{i,j} P\left[\left|\tau_{ij,T}^{-1} - \tau_{ij}^{-1}\right| \geq \frac{1}{2M^2}\sqrt{\frac{\xi}{8}}\right]$$
$$\leq \sup_{\gamma,\tilde{\gamma}\in[0,1]} P\left[\left|\frac{1}{T}\sum_t \big(I_t(\gamma)I_t(\tilde{\gamma}) - E\big[I_t(\gamma)I_t(\tilde{\gamma})\big]\big)\right| \geq \frac{1}{2M^2}\sqrt{\frac{\xi}{8}}\right]$$
$$\leq C_1 T \exp\{-C_2^* \xi T^{\eta}\} + C_3^* \xi^{-1/2} \exp\{-C_4 T^{\bar{\eta}}\}.$$

We deduce

$$\Psi_{nT}^0(\xi/4) \leq C_1^* T \exp\{-C_2^* \xi^2 T^{\eta}\} + C_3^* \xi^{-1} \exp\{-C_4 T^{\bar{\eta}}\}. \tag{47}$$

(b) *Bound of* $P_{1,nT}(\xi/4)$. For some constant C, we have

$$\|A_{ij}\| \leq C\tau_{ij,T} \max_{k,l,m}\left|\frac{1}{T}\sum_t I_{ij,t}\varepsilon_{i,t}x_{i,t,k}x_{i,t,l}x_{j,t,m}\right| \|\hat{\beta}_j - \beta_j\|.$$

Let $\chi_{3,T} = (\log T)^a$, for $a > 0$. From a similar argument as in the proof of Lemma 7, and Assumption B.1(d), we have $\max_{i,j} P[\tau_{ij,T} \geq \chi_{3,T}] = O(T^{-\bar{b}})$, for any $\bar{b} > 0$. Thus,

$$P_{1,nT}(\xi/4) \tag{48}$$
$$\leq \max_{i,j} P\left[\tau_{ij,T} \max_{k,l,m}\left|\frac{1}{T}\sum_t I_{ij,t}\varepsilon_{i,t}x_{i,t,k}x_{i,t,l}x_{j,t,m}\right| \|\hat{\beta}_j - \beta_j\| \geq \frac{\xi}{4C}\right]$$
$$\leq \max_{i,j} P[\tau_{ij,T} \geq \chi_{3,T}]$$
$$+ \max_{i,j} P\left[\max_{k,l,m}\left|\frac{1}{T}\sum_t I_{ij,t}\varepsilon_{i,t}x_{i,t,k}x_{i,t,l}x_{j,t,m}\right| \geq \sqrt{\frac{\xi}{4\chi_{3,T}C}} \text{ and } \tau_{ij,T} \leq \chi_{3,T}\right]$$
$$+ \max_{i,j} P\left[\|\hat{\beta}_j - \beta_j\| \geq \sqrt{\frac{\xi}{4\chi_{3,T}C}} \text{ and } \tau_{ij,T} \leq \chi_{3,T}\right]$$
$$\leq d^3 \max_{i,j}\max_{k,l,m} P\left[\left|\frac{1}{T}\sum_t I_{ij,t}\varepsilon_{i,t}x_{i,t,k}x_{i,t,l}x_{j,t,m}\right| \geq \sqrt{\frac{\xi}{4\chi_{3,T}C}}\right]$$
$$+ P\left[\|\hat{\beta}_j - \beta_j\| \geq \sqrt{\frac{\xi}{4\chi_{3,T}C}} \text{ and } \tau_{j,T} \leq \chi_{3,T}\right] + O\big(T^{-\bar{b}}\big).$$

By Assumption B.1(f),

$$\max_{i,j} \max_{k,l,m} P\left[\left| \frac{1}{T} \sum_t I_{ij,t} \varepsilon_{i,t} x_{i,t,k} x_{i,t,l} x_{j,t,m} \right| \geq \sqrt{\frac{\xi}{4\chi_{3,T} C}} \right] \tag{49}$$
$$\leq C_1 T \exp\left\{ -\frac{C_2^* \xi}{\chi_{3,T}} T^{\eta} \right\} + C_3^* \sqrt{\frac{\chi_{3,T}}{\xi}} \exp\{-C_4 T^{\bar{\eta}}\}.$$

Let us now focus on $P[\|\hat{\beta}_j - \beta_j\| \geq \sqrt{\frac{\xi}{4\chi_{3,T}C}}$ and $\tau_{j,T} \leq \chi_{3,T}]$. By using

$$\|\hat{\beta}_j - \beta_j\| \leq \chi_{3,T} \left\| Q_{x,j}^{-1} \right\| \left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\|$$
$$+ \chi_{3,T} \left\| \hat{Q}_{x,j}^{-1} - Q_{x,j}^{-1} \right\| \left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\|$$

when $\tau_{j,T} \leq \chi_{3,T}$, we get

$$P\left[\|\hat{\beta}_j - \beta_j\| \geq \sqrt{\frac{\xi}{4\chi_{3,T} C}} \text{ and } \tau_{j,T} \leq \chi_{3,T} \right] \tag{50}$$
$$\leq P\left[\left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\| \geq \frac{1}{2} \sqrt{\frac{\xi}{4\chi_{3,T} C}} \chi_{3,T}^{-1} \left\| Q_{x,j}^{-1} \right\|^{-1} \right]$$
$$+ P\left[\left\| \hat{Q}_{x,j}^{-1} - Q_{x,j}^{-1} \right\| \left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\| \geq \frac{1}{2} \sqrt{\frac{\xi}{4\chi_{3,T} C}} \chi_{3,T}^{-1} \right]$$
$$\leq P\left[\left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\| \geq \sqrt{\frac{\xi}{16\chi_{3,T}^3 C}} \left\| Q_{x,j}^{-1} \right\|^{-1} \right]$$
$$+ P\left[\left\| \hat{Q}_{x,j}^{-1} - Q_{x,j}^{-1} \right\| \geq \left(\frac{\xi}{16\chi_{3,T}^3 C} \right)^{1/4} \right]$$
$$+ P\left[\left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\| \geq \left(\frac{\xi}{16\chi_{3,T}^3 C} \right)^{1/4} \right]$$
$$\leq 2P\left[\left\| \frac{1}{T} \sum_t I_{j,t} x_{j,t} \varepsilon_{j,t} \right\| \geq \sqrt{\frac{\xi}{16\chi_{3,T}^3 C}} \left\| Q_{x,j}^{-1} \right\|^{-1} \right]$$
$$+ P\left[\left\| \hat{Q}_{x,j}^{-1} - Q_{x,j}^{-1} \right\| \geq \left(\frac{\xi}{16\chi_{3,T}^3 C} \right)^{1/4} \right],$$

for small ξ. From Assumption B.4(d), $\|Q_{x,j}^{-1}\|$ is bounded uniformly in j. Then, from Assumption B.1(c), the first probability in the RHS of inequality (51) is such that

$$P\left[\left\|\frac{1}{T}\sum_t I_{j,t}x_{,jt}\varepsilon_{j,t}\right\| \geq \sqrt{\frac{\xi}{16\chi_{3,T}^3 C}}\left\|Q_{x,j}^{-1}\right\|^{-1}\right] \leq C_1 T\exp\left\{-\frac{C_2^*\xi}{\chi_{3,T}^3}T^{\eta}\right\} + C_3^*\sqrt{\frac{\chi_{3,T}^3}{\xi}}\exp\left\{-C_4 T^{\bar{\eta}}\right\}. \tag{51}$$

To bound the second probability in the RHS of inequality (51), we use the next lemma.

LEMMA 13: *For any two nonsingular matrices A and B such that $\|A-B\| < \frac{1}{2}\|A^{-1}\|^{-1}$, we have*

$$\left\|B^{-1}-A^{-1}\right\| \leq 2\left\|A^{-1}\right\|^2\|A-B\|.$$

From Lemma 13, we get

$$\begin{aligned}
&P\left[\left\|\hat{Q}_{x,j}^{-1}-Q_{x,j}^{-1}\right\| \geq \left(\frac{\xi}{16\chi_{3,T}^3 C}\right)^{1/4}\right] \\
&\leq P\left[\|\hat{Q}_{x,j}-Q_{x,j}\| \geq \frac{1}{2}\left(\frac{\xi}{16\chi_{3,T}^3 C}\right)^{1/4}\left\|Q_{x,j}^{-1}\right\|^{-2}\right] \\
&\quad + P\left[\|\hat{Q}_{x,j}-Q_{x,j}\| \geq \frac{1}{2}\left\|Q_{x,j}^{-1}\right\|^{-1}\right] \\
&\leq 2P\left[\|\hat{Q}_{x,j}-Q_{x,j}\| \geq \frac{1}{2}\left(\frac{\xi}{16\chi_{3,T}^3 C}\right)^{1/4}\left\|Q_{x,j}^{-1}\right\|^{-2}\right],
\end{aligned}$$

for small $\xi > 0$. From Assumptions B.1(b) and B.1(c),

$$P\left[\|\hat{Q}_{x,j}-Q_{x,j}\| \geq \frac{1}{2}\left(\frac{\xi}{16\chi_{3,T}^3 C}\right)^{1/4}\left\|Q_{x,j}^{-1}\right\|^{-2}\right] \leq C_1 T\exp\left\{-C_2^*\sqrt{\frac{\xi}{\chi_{3,T}^3}}T^{\eta}\right\} + 2C_3^*\left(\frac{\chi_{3,T}^3}{\xi}\right)^{1/4}\exp\left\{-C_4 T^{\bar{\eta}}\right\}. \tag{52}$$

Then, from (48)–(52), we get

$$P_{1,nT}(\xi/4) \leq C_1^* T\exp\left\{-C_2^*\xi T^{\eta}/\chi_{3,T}^3\right\} + \frac{C_3^*\chi_{3,T}^{3/2}}{\sqrt{\xi}}\exp\left\{-C_4 T^{\bar{\eta}}\right\} + O\left(T^{-\bar{b}}\right), \tag{53}$$

for small $\xi > 0$ and some constants $C_1^*, C_2^*, C_3^*, C_4 > 0$.

(c) *Bound of* $P_{2,nT}(\xi/4)$. We have, from Assumption B.4,

$$\|C_{ij}\| \leq \|\hat{\beta}_i - \beta_i\| \|\hat{\beta}_j - \beta_j\| \sup_{k,l,m,p} \left| \frac{1}{T_{ij}} \sum_t I_{ij,t} x_{i,t,k} x_{j,t,l} x_{i,t,m} x_{j,t,p} \right|$$
$$\leq C \|\hat{\beta}_i - \beta_i\| \|\hat{\beta}_j - \beta_j\|.$$

Thus, we have

$$P_{2,nT}(\xi/4) \leq \max_{i,j} P\left[C \|\hat{\beta}_i - \beta_i\| \|\hat{\beta}_j - \beta_j\| \geq \frac{\xi}{4} \right]$$
$$\leq 2P\left[\|\hat{\beta}_i - \beta_i\| \geq \left(\frac{\xi}{4C} \right)^{1/2} \right].$$

By the same arguments as above, we get

$$P_{2,nT}(\xi/4) \leq C_1^* T \exp\{-C_2^* \xi T^\eta / \chi_{3,T}^3\} + \frac{C_3^* \chi_{3,T}^{3/2}}{\sqrt{\xi}} \exp\{-C_4 T^{\bar{\eta}}\}, \tag{54}$$

for small $\xi > 0$ and some constants $C_1^*, C_2^*, C_3^*, C_4 > 0$.

(d) *Conclusion.* From inequalities (46), (47), (53), and (54), we deduce

$$\Psi_{nT}(\xi) \leq C_1^* T \exp\{-C_2^* \xi_T^2 T^\eta\} + \frac{C_3^*}{\xi_T} \exp\{-C_4 T^{\bar{\eta}}\} + O(T^{-\bar{b}}),$$

where $\xi_T := \min\{\xi, \sqrt{\xi/\chi_{3,T}^3}\}$, for small $\xi > 0$, and constants C_1^*, C_2^*, C_3^*, $C_4 > 0$. For $\xi = (1-v)\kappa$ and $\kappa = M\sqrt{\frac{\log n}{T^\eta}}$, we get $\xi_T = (1-\nu)\kappa$ for large T and

$$n^2 \Psi_{nT}((1-v)\kappa) \leq C_1^* n^2 T \exp\{-C_2^* M^2 (1-v)^2 \log n\}$$
$$+ \frac{n^2 C_3^*}{(1-v)M} \sqrt{\frac{T^\eta}{\log n}} \exp\{-C_4^* T^{\bar{\eta}}\}$$
$$+ O(n^2 T^{-\bar{b}})$$
$$= O(1),$$

for $\bar{b}$ and M sufficiently large, when $n, T \to \infty$ such that $n = O(T^{\bar{\gamma}})$ for $\bar{\gamma} > 0$.

Finally, let us prove that $\psi_{nT} = O_p(\sqrt{\frac{\log n}{T^\eta}})$. Let $\epsilon > 0$. Then,

$$P\left[\psi_{nT} \geq \sqrt{\frac{\log n}{T^\eta}}\epsilon\right] \leq n^2 \max_{i,j} P\left[\|\hat{S}_{ij} - S_{ij}\| \geq \sqrt{\frac{\log n}{T^\eta}}\epsilon\right]$$

$$= n^2 \Psi_{nT}\left(\sqrt{\frac{\log n}{T^\eta}}\epsilon\right) \leq n^2 \Psi_{nT}\big((1-\upsilon)\kappa\big) = O(1),$$

for large ϵ. The conclusion follows.

D.9. *Proof of Lemma 11*

Under the null hypothesis $\mathcal{H}_0$, and by definition of the fitted residual $\hat{e}_i$, we have

$$\begin{aligned} \hat{e}_i &= \beta_{1,i} - \beta_{3,i}\hat{\nu} + C'_{\hat{\nu}}(\hat{\beta}_i - \beta_i) \\ &= \beta_{1,i} - \beta_{3,i}\nu + C'_{\hat{\nu}}(\hat{\beta}_i - \beta_i) - \beta_{3,i}(\hat{\nu} - \nu) \\ &= C'_{\hat{\nu}}(\hat{\beta}_i - \beta_i) - \beta_{3,i}(\hat{\nu} - \nu). \end{aligned} \tag{55}$$

By definition of $\hat{Q}_e$, it follows that

$$\begin{aligned} \hat{Q}_e &= \frac{1}{n}\sum_i (\hat{\beta}_i - \beta_i)' C_{\hat{\nu}} \hat{w}_i C'_{\hat{\nu}} (\hat{\beta}_i - \beta_i) \\ &\quad - 2(\hat{\nu} - \nu)' \frac{1}{n}\sum_i \beta'_{3,i} \hat{w}_i C'_{\hat{\nu}} (\hat{\beta}_i - \beta_i) \\ &\quad + (\hat{\nu} - \nu)' \frac{1}{n}\sum_i \beta'_{3,i} \hat{w}_i \beta_{3,i} (\hat{\nu} - \nu) \\ &=: \frac{1}{n}\sum_i (\hat{\beta}_i - \beta_i)' C_{\hat{\nu}} \hat{w}_i C'_{\hat{\nu}} (\hat{\beta}_i - \beta_i) - 2I_{71} + I_{72}. \end{aligned}$$

Let us study the second term in the RHS:

$$I_{71} = \frac{1}{\sqrt{nT}}(\hat{\nu} - \nu)' \frac{1}{\sqrt{n}} \sum_i \tau_{i,T} \beta'_{3,i} \hat{w}_i C'_{\hat{\nu}} \hat{Q}_{x,i}^{-1} Y_{i,T} =: \frac{1}{\sqrt{nT}}(\hat{\nu} - \nu)' I_{711},$$

where $I_{711} = O_p(1)$ by the same arguments used to control term I_{11} in the proof of Proposition 4. We have $\hat{\nu} - \nu = O_{p,\log}(\frac{1}{\sqrt{nT}} + \frac{1}{T})$ and $C_{\hat{\nu}} = O_p(1)$ by Lemma 6(v). Thus, $I_{71} = O_{p,\log}(\frac{1}{nT} + \frac{1}{T\sqrt{nT}})$.

Let us now consider I_{72}. From Lemma 3(ii)–(iii) and Lemma 6(v), we have $I_{72} = O_{p,\log}(\frac{1}{nT} + \frac{1}{T^2})$. The conclusion follows.

D.10. *Proof of Lemma 12*

Under $\mathcal{H}_1$, and using Equation (55), we have $\hat{e}_i = e_i + C'_{\hat{\nu}}(\hat{\beta}_i - \beta_i) - \beta_{3,i}(\hat{\nu} - \nu_\infty)$. By definition of $\hat{Q}_e$, it follows that

$$\begin{aligned}\hat{Q}_e = {} & \frac{1}{n}\sum_i e_i'\hat{w}_i e_i + 2\frac{1}{n}\sum_i(\hat{\beta}_i - \beta_i)'C_{\hat{\nu}}\hat{w}_i e_i \\ & - 2(\hat{\nu} - \nu_\infty)'\frac{1}{n}\sum_i \beta_{3,i}'\hat{w}_i e_i \\ & + \frac{1}{n}\sum_i(\hat{\beta}_i - \beta_i)'C_{\hat{\nu}}\hat{w}_i C'_{\hat{\nu}}(\hat{\beta}_i - \beta_i) \\ & - 2(\hat{\nu} - \nu_\infty)'\frac{1}{n}\sum_i \beta_{3,i}'\hat{w}_i C'_{\hat{\nu}}(\hat{\beta}_i - \beta_i) \\ & + (\hat{\nu} - \nu_\infty)'\frac{1}{n}\sum_i \beta_{3,i}'\hat{w}_i \beta_{3,i}(\hat{\nu} - \nu_\infty) \\ =: {} & I_{81} + I_{82} + I_{83} + I_{84} + I_{85} + I_{86}.\end{aligned} \tag{56}$$

From Equations (30) and (32) and similar arguments as in Section B.4(c), we have $I_{81} = \frac{1}{n}\sum_i w_i e_i^2 + O_{p,\log}(\frac{1}{\sqrt{T}})$. By similar arguments as for term I_{11} in the proof of Proposition 4, we have $I_{82} = \frac{2}{\sqrt{nT}}(\frac{1}{\sqrt{n}}\sum_i \tau_{i,T} Y_{i,T}' \hat{Q}_{x,i}^{-1} C_{\hat{\nu}}\hat{w}_i e_i) = O_p(\frac{1}{\sqrt{nT}})$. By using $\frac{1}{n}\sum_i \beta_{3,i}'\hat{w}_i e_i = \frac{1}{n}\sum_i \beta_{3,i}' w_i e_i + O_{p,\log}(\frac{1}{\sqrt{T}}) = O_p(\frac{1}{\sqrt{n}}) + O_{p,\log}(\frac{1}{\sqrt{T}})$ and $\hat{\nu} - \nu_\infty = O_{p,\log}(\frac{1}{\sqrt{n}} + \frac{1}{T})$, we get $I_{83} = O_{p,\log}(\frac{1}{n} + \frac{1}{\sqrt{nT}} + \frac{1}{\sqrt{T^3}})$. Similarly as for I_{82}, we have $I_{85} = O_{p,\log}(\frac{1}{n\sqrt{T}} + \frac{1}{\sqrt{nT^3}})$. From $\hat{\nu} - \nu_\infty = O_{p,\log}(\frac{1}{\sqrt{n}} + \frac{1}{T})$, we have $I_{86} = O_{p,\log}(\frac{1}{n} + \frac{1}{T^2})$. The conclusion follows.

D.11. *Proof of Lemma 13*

Write $B^{-1} - A^{-1} = [A(I - A^{-1}(A-B))]^{-1} - A^{-1} = \{[I - A^{-1}(A-B)]^{-1} - I\}A^{-1}$, and use that, for a square matrix C such that $\|C\| < 1$, we have $(I-C)^{-1} = I + C + C^2 + C^3 + \cdots$ and $\|(I-C)^{-1} - I\| \leq \|C\| + \|C\|^2 + \cdots \leq \frac{\|C\|}{1-\|C\|}$. Thus, we get

$$\begin{aligned}\|B^{-1} - A^{-1}\| &\leq \frac{\|A^{-1}(A-B)\|}{1 - \|A^{-1}(A-B)\|}\|A^{-1}\| \leq \frac{\|A^{-1}\|^2\|A-B\|}{1 - \|A^{-1}\|\|A-B\|} \\ &\leq 2\|A^{-1}\|^2\|A-B\|,\end{aligned}$$

if $\|A-B\| < \frac{1}{2}\|A^{-1}\|^{-1}$.

APPENDIX E: LINK TO CHAMBERLAIN AND ROTHSCHILD (1983)

In this appendix, we establish the link between the no-arbitrage conditions and asset pricing restrictions in CR on the one hand, and the asset pricing restriction (3) on the other hand. As in Appendix B.1, for any sequence (γ_i) in Γ let $\mathcal{P}_n$ be the set of portfolios investing in the n assets $\gamma_1, \gamma_2, \dots, \gamma_n$ with $\mathcal{F}_0$-measurable shares. By assuming that the shares are finite P-a.s., we have $E[p_n^2|\mathcal{F}_0] < \infty$, P-a.s., and we can build on the framework of Hansen and Richard (1987) with conditionally square integrable payoffs. Moreover, we denote by $\mathcal{P} = \bigcup_{n=1}^{\infty} \mathcal{P}_n$ the set of finite portfolios with conditionally square integrable payoff.

Let $\mathcal{J}^* \subset \Gamma$ be the set of countable collections of assets (γ_i) such that Conditions (i) and (ii) hold for any portfolio sequence $(p_n) \in \mathcal{P}$, where Conditions (i) and (ii) are: (i) If $V[p_n|\mathcal{F}_0] \overset{\text{a.s.}}{\to} 0$ and $C(p_n) \overset{\text{a.s.}}{\to} 0$, then $E[p_n|\mathcal{F}_0] \overset{\text{a.s.}}{\to} 0$. (ii) If $V[p_n|\mathcal{F}_0] \overset{\text{a.s.}}{\to} 0$, $C(p_n) \geq 0$, P-a.s., $\limsup_{n\to\infty} |C(p_n)| \geq \epsilon$ on a set of nonzero measure, for a constant $\epsilon > 0$, and $E[p_n|\mathcal{F}_0] \overset{\text{a.s.}}{\to} \bar{\delta}$, for a constant $\bar{\delta}$, then $\bar{\delta} > 0$. Condition (i) means that, if the conditional variability and cost vanish, so does the conditional expected return. Condition (ii) means that, if the conditional variability vanishes and the cost is positive, the conditional expected return is positive. They correspond to Conditions A.1(i) and (ii) in CR written conditionally on $\mathcal{F}_0$ and for a given countable collection of assets (γ_i). Hence, the set $\mathcal{J}^*$ is the set permitting no asymptotic arbitrage opportunities in the sense of CR in a conditional setting (see also Chamberlain (1983)). We use the convergence of conditional expectations as in Hansen and Richard (1987), and focus on a.s. convergence as opposed to convergence in probability (see Hansen and Richard (1987, footnote 5 on p. 594)) since this helps when defining the extension of the cost function $C(\cdot)$ to the completion of set $\mathcal{P}$. Let $\mathcal{J}^{**} \subset \Gamma$ be the set of sequences (γ_i) such that $\inf_{\nu\in\mathbb{R}^K} \sum_{i=1}^{\infty} [a(\gamma_i) - b(\gamma_i)'\nu]^2 < \infty$, P-a.s. These sequences met the summability condition of CR in a conditional setting. In the proof of the following proposition, we assume that β is bounded on $[0, 1] \times \Omega$ and $E[f_1|\mathcal{F}_0]$ is bounded on Ω.

PROPOSITION APR: *Under Assumptions* APR.1–APR.3, *and*

(i) $\inf_{n\geq 1} \text{eig}_{\min}(\Sigma_{\varepsilon,t,n}) > 0$, *P-a.s., for a.e.* (γ_i) *in* Γ,

(ii) $\text{eig}_{\min}(V[f_t|\mathcal{F}_{t-1}]) > 0$, *P-a.s.*,

we have: either $\bar{\mu}_\Gamma(\mathcal{J}^*) = \bar{\mu}_\Gamma(\mathcal{J}^{**}) = 1$, *or* $\bar{\mu}_\Gamma(\mathcal{J}^*) = \bar{\mu}_\Gamma(\mathcal{J}^{**}) = 0$. *The former case occurs if, and only if, the asset pricing restriction* (3) *holds.*

When we condition on $\mathcal{F}_0$, the fact that the set of sequences such that $\inf_{\nu\in\mathbb{R}^K} \sum_{i=1}^{\infty} [a(\gamma_i) - b(\gamma_i)'\nu]^2 < \infty$ has μ_Γ-measure equal to either 1, or 0, is a consequence of the Kolmogorov zero–one law (e.g., Billingsley (1995)). Indeed, $\inf_{\nu\in\mathbb{R}^K} \sum_{i=1}^{\infty} [a(\gamma_i) - b(\gamma_i)'\nu]^2 < \infty$ if, and only if, $\inf_{\nu\in\mathbb{R}^K} \sum_{i=n}^{\infty} [a(\gamma_i) - b(\gamma_i)'\nu]^2 < \infty$, for any $n \in \mathbb{N}$. Thus, the zero–one law applies since the event $\inf_{\nu\in\mathbb{R}^K}\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2<\infty$ belongs to the tail sigma-field $\mathcal{T}=\bigcap_{n=1}^{\infty}\sigma(\gamma_i, i=n,n+1,\ldots)$, and the variables γ_i are i.i.d. under measure μ_Γ. Proposition APR shows that this zero–one measure property applies also for the set $\mathcal{J}^{**}$. Proposition APR shows that the asset pricing (3) characterizes the functions $\beta=(a,b')'$ defined on $[0,1]\times\Omega$ that are compatible with absence of asymptotic arbitrage opportunities in the continuum economy under the definitions of arbitrage used in CR and in Hansen and Richard (1987). Moreover, Proposition APR also provides a reverse implication compared to Proposition 1: when the asset pricing restriction (3) does not hold, asymptotic arbitrage in the sense of Assumption APR.4, or of Assumptions A.1(i) and (ii) of CR, exists for $\bar{\mu}_\Gamma$-almost any countable collection of assets.

PROOF OF PROPOSITION APR: The proof involves four steps.

Step 1: If the asset pricing restriction (3) holds, then $\bar{\mu}_\Gamma(\mathcal{J}^{**})=1$. Indeed, if the asset pricing restriction (3) holds for some $\mathcal{F}_0$-measurable function ν, we have for a.e. $\omega\in\Omega$: $a(\gamma,\omega)-b(\gamma,\omega)'\nu(\omega)=0$ for a.e. $\gamma\in[0,1]$. Since functions a and b are jointly measurable on $[0,1]\times\Omega$, this implies that for a.e. $\gamma\in[0,1]$: $a(\gamma,\omega)-b(\gamma,\omega)'\nu(\omega)=0$ for a.e. $\omega\in\Omega$. Then, the set $\{(\gamma_i)\in\Gamma:\sum_{i=1}^{\infty}[a(\gamma_i)-b(\gamma_i)'\nu]^2=0, P\text{-a.s.}\}=\bigcap_{i=1}^{\infty}\{(\gamma_i)\in\Gamma:a(\gamma_i,\omega)-b(\gamma_i,\omega)'\nu(\omega)=0, \text{for a.e.}\,\omega\in\Omega\}$ has μ_Γ-measure 1. Since this set is a subset of $\mathcal{J}^{**}$, it follows that $\bar{\mu}_\Gamma(\mathcal{J}^{**})=1$.

Step 2: If the asset pricing restriction (3) does not hold, then $\bar{\mu}_\Gamma(\mathcal{J}^{**})=0$. If the asset pricing restriction (3) does not hold, the quantity $\delta=\inf_{\nu\in\mathbb{R}^K}\int[a(\gamma)-b(\gamma)'\nu]^2\,d\gamma$ is such that $\delta(\omega)\geq\underline{\delta}$ for all $\omega\in A$, for a set $A\in\mathcal{F}_0$ with $P(A)>0$ and a scalar $\underline{\delta}>0$. To prove $\bar{\mu}_\Gamma(\mathcal{J}^{**})=0$, we show $\mathcal{J}_1\cap\mathcal{J}^{**}=\emptyset$, where $\mathcal{J}_1$ is the set with μ_Γ-measure 1 defined in Lemma 1. Indeed, $\mathcal{J}_1\cap\mathcal{J}^{**}=\emptyset$ implies that $\mathcal{J}^{**}\subset\mathcal{J}_1^c$ is a negligible set under measure μ_Γ, and thus has $\bar{\mu}_\Gamma$ measure 0. The proof of $\mathcal{J}_1\cap\mathcal{J}^{**}=\emptyset$ is by contradiction. Let us assume that sequence (γ_i) is in $\mathcal{J}_1\cap\mathcal{J}^{**}$, and let $\xi_n:=\inf_{\nu\in\mathbb{R}^K}\frac{1}{n}\sum_{i=1}^{n}[a(\gamma_i)-b(\gamma_i)'\nu]^2$. Since $(\gamma_i)\in\mathcal{J}_1$, from inequality (19), we have $\xi_n 1_{A\cap S_n^*}\geq 2^{-1}\underline{\delta}1_{A\cap S_n^*}$, where the set S_n^* defined in the proof of Proposition 1 is such that $P(S_n^*)\to1$ as $n\to\infty$. This implies that $E[\xi_n^2]\geq E[\xi_n^2|1_{A\cap S_n^*}=1]P(A\cap S_n^*)\geq(\underline{\delta}^2/4)P(A\cap S_n^*)\to(\underline{\delta}^2/4)P(A)$, and thus

$$\liminf_{n\to\infty}E\big[\xi_n^2\big]>0. \tag{57}$$

Since $(\gamma_i)\in\mathcal{J}^{**}$, we have $\xi_n\to0$, P-a.s. Moreover, since function β is bounded, we have $|\xi_n|\leq C$, P-a.s., for some constant C. Then, by the Lebesgue dominated convergence theorem, it follows that $E[\xi_n^2]\to0$. This is impossible, if (57) holds.

Step 3: If the asset pricing restriction (3) holds, then $\bar{\mu}_\Gamma(\mathcal{J}^{*})=1$. If (3) holds, it follows that $\mu_n=B_n\lambda$, P-a.s., for all n, for μ_Γ-almost all sequences (γ_i), where $\lambda=\nu+E[f_1|\mathcal{F}_0]$. Then, for any portfolio sequence (p_n), we get $E[p_n|\mathcal{F}_0] = R_0 C(p_n) + \alpha_n' B_n \lambda$. From Assumption APR.2(iv) and boundedness of $E[f_1|\mathcal{F}_0]$, it follows that λ is bounded on Ω. Moreover, we have $V[p_n|\mathcal{F}_0] = (B_n'\alpha_n)' V[f_1|\mathcal{F}_0](B_n'\alpha_n) + \alpha_n' \Sigma_{\varepsilon,1,n}\alpha_n \geq \text{eig}_{\min}(V[f_1|\mathcal{F}_0])\|B_n'\alpha_n\|^2$, where $\text{eig}_{\min}(V[f_1|\mathcal{F}_0]) > 0$, P-a.s. Then, Conditions (i) and (ii) in the definition of set $\mathcal{J}^*$ follow, for μ_Γ-almost any sequence (γ_i), that is, $\mu_\Gamma(\mathcal{J}^*) = \bar{\mu}_\Gamma(\mathcal{J}^*) = 1$.

Step 4: If the asset pricing restriction (3) does not hold, then $\bar{\mu}_\Gamma(\mathcal{J}^*) = 0$. To prove that $\bar{\mu}_\Gamma(\mathcal{J}^*) = 0$, we show that $\mathcal{J}^* \cap \mathcal{J} \cap \mathcal{J}_1 = \emptyset$, where $\mathcal{J}$ and $\mathcal{J}_1$ are the sets with μ_Γ-measure 1 defined in Assumption APR.3 and in Lemma 1, respectively. The proof is by contradiction. Let us assume that sequence (γ_i) is in set $\mathcal{J}^* \cap \mathcal{J} \cap \mathcal{J}_1$. By following the same arguments as in CR on pp. 1292 and 1295, we have

$$\mu_n' \Sigma_n^{-1} \mu_n = \sup_{p_n \in \mathcal{P}_n : C(p_n) = 0} E[p_n|\mathcal{F}_0]^2 / V[p_n|\mathcal{F}_0], \tag{58}$$

$$\Sigma_n^{-1} \geq \text{eig}_{\max}(\Sigma_{\varepsilon,1,n})^{-1}\left[I_n - B_n\left(B_n'B_n\right)^{-1}B_n'\right], \tag{59}$$

P-a.s. Let us prove that the RHS of (58) is upper bounded uniformly in n. We use Hilbert space methods as in Hansen and Richard (1987) applied to the conditional economy generated by the countable collection of assets (γ_i). Let $\langle p, q\rangle_{\mathcal{F}_0} = E[pq|\mathcal{F}_0]$ and $\|p\|_{\mathcal{F}_0} = \langle p, p\rangle_{\mathcal{F}_0}^{1/2}$ be the conditional scalar product and norm in the linear space of $\mathcal{F}_1$-measurable random variables, which are square integrable conditionally to $\mathcal{F}_0$. Conditional convergence of (p_n) to p is defined as $\|p_n - p\|_{\mathcal{F}_0} \overset{\text{a.s.}}{\to} 0$ for $n \to \infty$. Conditional Cauchy sequences are defined similarly. Since $(\gamma_i) \in \mathcal{J}^*$, Condition (ii) is satisfied for any portfolio sequence in $\mathcal{P}$. This implies that Condition (iii): If $E[p_n^2|\mathcal{F}_0] \overset{\text{a.s.}}{\to} 0$, then $C(p_n) \overset{\text{a.s.}}{\to} 0$, holds for any portfolio sequence (p_n) in $\mathcal{P}$. Indeed, suppose that (p_n) is such that $E[p_n^2|\mathcal{F}_0] \overset{\text{a.s.}}{\to} 0$ but $C(p_n)$ does not converge to 0 a.s. Define the new portfolio sequence (p_n'), such that $p_n' = p_n$ if $C(p_n) \geq 0$, and $p_n' = -p_n$ otherwise. Then, portfolio sequence (p_n') violates Condition (ii), which is impossible. Condition (iii) implies conditional continuity of function $C(\cdot)$ at the zero payoff in $\mathcal{P}$, and corresponds to Assumption 2.3 in Hansen and Richard (1987). Now, by using Condition (iii), we can extend the cost function $C(\cdot)$ to the linear space $\bar{\mathcal{P}}$, that is, the conditional completion of $\mathcal{P}$ w.r.t. the limits of conditional Cauchy sequences. Indeed, let $p \in \bar{\mathcal{P}}$, and let (p_n) be a conditional Cauchy sequence in $\mathcal{P}$ converging conditionally to p. Then, $C(p_n)$ is a Cauchy sequence in $\mathbb{R}$, P-a.s. By the completeness property of $\mathbb{R}$, this Cauchy sequence converges to a unique value, P-a.s., which we define as $C(p)$. For any $p \in \bar{\mathcal{P}}$, random variable $C(p)$ is $\mathcal{F}_0$-measurable by Theorem 20.A in Halmos (1950). This extension of the function $C(\cdot)$ on $\bar{\mathcal{P}}$ is conditionally linear and conditionally continuous at the zero payoff. By Theorem 2.1 in Hansen and Richard (1987), there exists a $\mathcal{F}_1$-measurable random variable c such that $E[c^2|\mathcal{F}_0] < \infty$ and $C(p) = E[cp|\mathcal{F}_0]$, P-a.s., for any portfolio $p \in \bar{\mathcal{P}}$. This property is the conditional analogue of the Riesz Representation Theorem. Any portfolio $p \in \bar{\mathcal{P}}$ can be written as $p = \pi_0 + \pi_1 c + \tilde{p}$, where π_0 and π_1 are $\mathcal{F}_0$-measurable, and $\tilde{p}$ is conditionally orthogonal to 1 and c, namely, $E[\tilde{p}|\mathcal{F}_0] = E[c\tilde{p}|\mathcal{F}_0] = 0$. If the portfolio p has zero cost, that is, $C(p) = 0$, then $p = \pi_0(1 - E[c|\mathcal{F}_0]E[c^2|\mathcal{F}_0]^{-1}c) + \tilde{p} =: \pi_0 p^* + \tilde{p}$. The payoff p^* is the residual of the conditional projection of the constant payoff 1 on the payoff c. Since the component $\tilde{p}$ contributes to the conditional variance of portfolio p but not to its conditional mean, we deduce that, for any portfolio $p \in \bar{\mathcal{P}}$ such that $C(p) = 0$, we get

$$E[p|\mathcal{F}_0]^2/V[p|\mathcal{F}_0] \leq E\big[p^*|\mathcal{F}_0\big]^2/V\big[p^*|\mathcal{F}_0\big] =: \rho^2 < \infty, \tag{60}$$

P-a.s. (see CR, Corollary 1, for a similar result in their unconditional framework). From (58), (59), and (60), we get $\rho^2 \operatorname{eig}_{\max}(\Sigma_{\varepsilon,1,n}) \geq \mu_n'(I_n - B_n(B_n'B_n)^{-1}B_n')\mu_n = \min_{\lambda\in\mathbb{R}^K} \|\mu_n - B_n\lambda\|^2 = \min_{\nu\in\mathbb{R}^K} \|A_n - B_n\nu\|^2 = \min_{\nu\in\mathbb{R}^K} \sum_{i=1}^n [a(\gamma_i) - b(\gamma_i)'\nu]^2$, for any $n \in \mathbb{N}$, P-a.s. Hence, we deduce that $\xi_n = \min_{\nu\in\mathbb{R}^K} \frac{1}{n}\sum_{i=1}^n [a(\gamma_i) - b(\gamma_i)'\nu]^2$ is such that $\xi_n \leq \rho^2 \frac{1}{n} \operatorname{eig}_{\max}(\Sigma_{\varepsilon,1,n})$, for any n, P-a.s. Since $(\gamma_i) \in \mathcal{J}$, from Assumption APR.3, the RHS converges in L^2 to 0. Then, we get $E[\xi_n^2] \to 0$ as $n \to \infty$. However, since the asset pricing restriction (3) does not hold and $(\gamma_i) \in \mathcal{J}_1$, we know from inequality (57) that $E[\xi_n^2]$ is bounded away from 0, and we get a contradiction. *Q.E.D.*

APPENDIX F: CHECK OF ASSUMPTIONS UNDER BLOCK-DEPENDENCE

In this appendix, we verify that the eigenvalue condition in Assumption APR.3, and the cross-sectional/time series dependence and CLT conditions in Assumptions A.1–A.5, are satisfied under a block-dependence structure in a time-invariant and serially i.i.d. framework. We start by providing the main result (Section F.1), we prove it (Section F.2), and then prove two auxiliary lemmas (Sections F.3 and F.4).

F.1. *Main Result*

Let us assume that:

BD.1. The errors $\varepsilon_t(\gamma)$ are i.i.d. over time with $E[\varepsilon_t(\gamma)] = 0$ and $E[\varepsilon_t(\gamma)^3] = 0$, for all $\gamma \in [0,1]$. For any n, there exists a partition of the interval $[0,1]$ into $J_n \leq n$ subintervals $I_1, \ldots, I_{J_n}$, such that $\varepsilon_t(\gamma)$ and $\varepsilon_t(\gamma')$ are independent if γ and γ' belong to different subintervals, and $J_n \to \infty$ as $n \to \infty$.

BD.2. The blocks are such that $n\sum_{m=1}^{J_n} B_m^2 = O(1)$, $n^{3/2}\sum_{m=1}^{J_n} B_m^3 = o(1)$, where $B_m = \int_{I_m} dG(\gamma)$.

BD.3. The factors (f_t) and the indicators $(I_t(\gamma))$, $\gamma \in [0,1]$, are i.i.d. over time, mutually independent, and independent of the errors $(\varepsilon_t(\gamma))$, $\gamma \in [0,1]$.

BD.4. There exists a constant M such that $\|f_t\| \leq M$, P-a.s. Moreover, $\sup_{\gamma\in[0,1]} E[|\varepsilon_t(\gamma)|^6] < \infty$, $\sup_{\gamma\in[0,1]} \|\beta(\gamma)\| < \infty$ and $\inf_{\gamma\in[0,1]} E[I_t(\gamma)] > 0$.

The block-dependence structure as in Assumption BD.1 is satisfied, for instance, when there are unobserved industry-specific factors independent among industries and over time, as in Ang, Liu, and Schwarz (2008). In empirical applications, blocks can match industrial sectors. Then, the number J_n of blocks amounts to a couple of dozens, and the number of assets n amounts to a couple of thousands. There are approximately nB_m assets in block m, when n is large. In the asymptotic analysis, Assumption BD.2 on block sizes and block number requires that the largest block size shrinks with n and that there are not too many large blocks, that is, the partition in independent blocks is sufficiently fine grained asymptotically. Within blocks, covariances do not need to vanish asymptotically.

LEMMA 14: *Let Assumptions* BD.1–BD.4 *on block-dependence and Assumptions* SC.1–SC.2 *on random sampling hold*. *Then*, *Assumptions* APR.3, A.1, A.2, A.3, A.4 (*with any* $\bar{q} \in (0, 1)$ *and* $\bar{\delta} \in (1/2, 1)$), *and* A.5 *are satisfied*.

The proof of Lemma 14 uses a result on almost sure convergence in Stout (1974), a large deviation theorem based on the Hoeffding inequality in Bosq (1998), and CLTs for martingale difference arrays in Davidson (1994) and White (2001).

Instead of a block structure, we can also assume that the covariance matrix is full, but with off-diagonal elements vanishing asymptotically. We could also accommodate weak serial dependence and conditioning information. In those settings, we can carry out similar checks, although at the cost of increased notational complexity.

F.2. *Proof of Lemma 14*

F.2.1. *Assumption APR.3*

We use that $\mathrm{eig}_{\max}(A) \leq \max_{i=1,\ldots,n} \sum_{j=1}^n |a_{i,j}|$ for any matrix $A = [a_{ij}]_{i,j=1,\ldots,n}$. Then, for any sequence (γ_i) in $[0, 1]$, we have

$$\begin{aligned}(61)\qquad \mathrm{eig}_{\max}(\Sigma_{\varepsilon,1,n}) &\leq \max_{i=1,\ldots,n} \sum_{j=1}^n \big|\mathrm{Cov}\big[\varepsilon_t(\gamma_i), \varepsilon_t(\gamma_j)\big]\big| \\ &\leq C \max_{m=1,\ldots,J_n} \sum_{j=1}^n 1\{\gamma_j \in I_m\},\end{aligned}$$

where $C := \sup_{\gamma\in[0,1]} E[\varepsilon_t(\gamma)^2]$. Define $\mathcal{J} = \{(\gamma_i) : \max_{m=1,\ldots,J_n} \frac{1}{n} \sum_{i=1}^n 1\{\gamma_i \in I_m\} = o(1)\}$. Then Assumption APR.3(ii) holds if $\mu_\Gamma(\mathcal{J}) = 1$. From Theorem 2.1.1 in Stout (1974), it is enough to show that $\sum_{n=1}^\infty \mu_\Gamma(\max_{m=1,\ldots,J_n} \frac{1}{n} \times \sum_{i=1}^n 1\{\gamma_i \in I_m\} > \varepsilon) < \infty$, for any $\varepsilon > 0$. Now, since $\max_{m=1,\ldots,J_n} B_m = o(1)$, we have $\mu_\Gamma(\max_{m=1,\ldots,J_n} \frac{1}{n}\sum_{i=1}^n 1\{\gamma_i \in I_m\} > \varepsilon) \le \mu_\Gamma(\max_{m=1,\ldots,J_n} |\frac{1}{n}\sum_{i=1}^n 1\{\gamma_i \in I_m\} - B_m| > \varepsilon/2)$, for large n. Thus, we get

$$\mu_\Gamma\left(\max_{m=1,\ldots,J_n} \frac{1}{n}\sum_{i=1}^n 1\{\gamma_i \in I_m\} > \varepsilon\right)$$
$$\le J_n \max_{m=1,\ldots,J_n} \mu_\Gamma\left(\left|\frac{1}{n}\sum_{i=1}^n 1\{\gamma_i \in I_m\} - B_m\right| > \varepsilon/2\right),$$

for large n. To bound the probability in the RHS, we use $|1\{\gamma_i \in I_m\} - B_m| \le 1$ and the Hoeffding inequality (see Bosq (1998, Theorem 1.2)) to get

$$\mu_\Gamma\left(\left|\frac{1}{n}\sum_{i=1}^n 1\{\gamma_i \in I_m\} - B_m\right| > \varepsilon/2\right) \le 2\exp\left(-n\varepsilon^2/8\right).$$

Then, since $J_n \le n$, we get

$$\sum_{n=1}^\infty \mu_\Gamma\left(\max_{m=1,\ldots,J_n} \frac{1}{n}\sum_{i=1}^n 1\{\gamma_i \in I_m\} > \varepsilon\right) \le 2\sum_{n=1}^\infty n\exp\left(-n\varepsilon^2/8\right) < \infty,$$

and the conclusion follows.

F.2.2. *Assumption A.1*

Conditions (a) and (b) are clearly satisfied under Assumptions BD.1, BD.3, and BD.4. Let us now consider condition (c). We have $\sigma_{ij,t} = E[\varepsilon_t(\gamma_i)\varepsilon_t(\gamma_j)| \gamma_i, \gamma_j] =: \sigma_{ij}$ independent of t. Thus, $E[\sigma_{ij,t}^2|\gamma_i, \gamma_j]^{1/2} = \sigma_{ij}$. By Assumptions BD.1, BD.4, and the Cauchy–Schwarz inequality, $\sigma_{ij} = \sum_{m=1}^{J_n} 1\{\gamma_i, \gamma_j \in I_m\} \times E[\varepsilon_t(\gamma_i)\varepsilon_t(\gamma_j)|\gamma_i, \gamma_j] \le C\sum_{m=1}^{J_n} 1\{\gamma_i, \gamma_j \in I_m\}$, where $C = \sup_{\gamma\in[0,1]} E[\varepsilon_t(\gamma)^2]$. Hence, we get

$$E\left[\frac{1}{n}\sum_{i,j} E\left[\sigma_{ij,t}^2|\gamma_i, \gamma_j\right]^{1/2}\right]$$
$$\le C\frac{1}{n}\sum_i \sum_{m=1}^{J_n} E\left[1\{\gamma_i \in I_m\}\right] + C\frac{1}{n}\sum_{i\neq j}\sum_{m=1}^{J_n} E\left[1\{\gamma_i, \gamma_j \in I_m\}\right]$$
$$= C\sum_{m=1}^{J_n} B_m + C(n-1)\sum_{m=1}^{J_n} B_m^2 = O\left(1 + n\sum_{m=1}^{J_n} B_m^2\right).$$

From Assumption BD.2, the RHS is $O(1)$, and condition (c) in Assumption A.1 follows.

F.2.3. *Assumption A.2*

Let us consider condition (a). In the time-invariant case under Assumptions BD.1 and BD.3, we have $S_{ij} = \sigma_{ij} Q_x$ and $v_3 = w_i b_i$, where $Q_x = E[x_t x_t']$. Then, Assumption A.2(a) is equivalent to $\frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \tau_i Y_{i,T} \otimes b_i \Rightarrow N(0, S_b)$, where $S_b := \lim_{n\to\infty} E[\frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} \sigma_{ij}(Q_x \otimes b_i b_j')]$. This limit is finite (if it exists), since from Assumption BD.4 we have $\|\frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} \sigma_{ij}(Q_x \otimes b_i b_j')\| \leq C \frac{1}{n} \sum_{i,j} |\sigma_{i,j}|$, and $E[\frac{1}{n} \sum_{i,j} |\sigma_{i,j}|] = O(1)$ from Assumption A.1. Moreover,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \tau_i Y_{i,T} \otimes b_i = \frac{1}{\sqrt{Tn}} \sum_{t=1}^T \sum_{i=1}^n w_i \tau_i I_{i,t}(x_t \otimes b_i)\varepsilon_{i,t}$$
$$= \frac{1}{\sqrt{T}} \sum_{t=1}^T \xi_{n,t},$$

where $\xi_{n,t} = \frac{1}{\sqrt{n}} \sum_{i=1}^n w_i \tau_i I_{i,t}(x_t \otimes b_i)\varepsilon_{i,t}$. The triangular array $(\xi_{n,t})$ is a martingale difference sequence w.r.t. the sigma-field $\mathcal{F}_{n,t} = \{f_{\underline{t}}, \varepsilon_{i,\underline{t}}, \gamma_i, i = 1, \dots, n\}$. From a multivariate version of Corollary 5.26 in White (2001), the CLT in condition (a) follows if we show:

(i) $\frac{1}{T} \sum_{t=1}^T E[\xi_{n,t} \xi_{n,t}'] \to S_b$,
(ii) $\frac{1}{T} \sum_{t=1}^T (\xi_{n,t} \xi_{n,t}' - E[\xi_{n,t} \xi_{n,t}']) = o_p(1)$,
(iii) $\sup_{t=1,\dots,T} E[\|\xi_{n,t}\|^{2+\delta}] = O(1)$, for some $\delta > 0$.

Moreover, we prove the alternative characterization of the asymptotic variance–covariance matrix:

(iv) $S_b = \text{a.s.-}\lim_{n\to\infty} \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} \sigma_{ij}(Q_x \otimes b_i b_j')$.

Let us check these conditions. (i) Let $\mathcal{G}_n = \{\gamma_i, i = 1, \dots, n\}$. We have

$$\frac{1}{T} \sum_t E[\xi_{n,t} \xi_{n,t}' | \mathcal{G}_n]$$
$$= \frac{1}{Tn} \sum_t \sum_{i,j} w_i w_j \tau_i \tau_j E[I_{i,t} I_{j,t}(x_t x_t' \otimes b_i b_j')\varepsilon_{i,t}\varepsilon_{j,t} | \gamma_i, \gamma_j]$$
$$= \frac{1}{Tn} \sum_t \sum_{i,j} w_i w_j \tau_i \tau_j$$
$$\times E[I_{i,t} I_{j,t} | \gamma_i, \gamma_j](E[x_t x_t'] \otimes b_i b_j') E[\varepsilon_{i,t}\varepsilon_{j,t} | \gamma_i, \gamma_j]$$
$$= \frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{i,j}} \sigma_{ij}(Q_x \otimes b_i b_j').$$

By taking expectation on both sides, condition (i) follows.

Let us now consider condition (ii). Define $\zeta_{n,T} = \frac{1}{T}\sum_t(\xi_{n,t,k}\xi_{n,t,l} - E[\xi_{n,t,k}\xi_{n,t,l}])$, where $\xi_{n,t,k}$ is the kth element of $\xi_{n,t}$. Since $E[\zeta_{n,T}] = 0$, it is enough to show $V[\zeta_{n,T}] = o(1)$, for any k, l. We show this for $k = l$; the proof for $k \neq l$ is similar. For expository purposes, we omit the index k, and we write $x_{t,k}^2 \equiv x_t^2$. We have

$$V[\zeta_{n,T}] = \frac{1}{T^2}\sum_t V\big[\xi_{n,t}^2\big] + \frac{1}{T^2}\sum_{t\neq s}\operatorname{Cov}\big(\xi_{n,t}^2, \xi_{n,s}^2\big), \tag{62}$$

where $\xi_{n,t}^2 = \frac{1}{n}\sum_{i,j} w_i w_j \tau_i\tau_j I_{i,t} I_{j,t} x_t^2 b_i b_j \varepsilon_{i,t}\varepsilon_{j,t}$.

• Consider first the terms $\operatorname{Cov}(\xi_{n,t}^2, \xi_{n,s}^2)$ for $t \neq s$. By the variance decomposition formula,

$$\operatorname{Cov}\big(\xi_{n,t}^2, \xi_{n,s}^2\big) = E\big[\operatorname{Cov}\big(\xi_{n,t}^2, \xi_{n,s}^2|\mathcal{G}_n\big)\big] + \operatorname{Cov}\big[E\big(\xi_{n,t}^2|\mathcal{G}_n\big), E\big(\xi_{n,s}^2|\mathcal{G}_n\big)\big].$$

We have $\operatorname{Cov}(\xi_{n,t}^2, \xi_{n,s}^2|\mathcal{G}_n) = 0$ from the i.i.d. assumption over time. Moreover,

$$E\big[\xi_{n,t}^2|\mathcal{G}_n\big] = \frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i\tau_j}{\tau_{ij}} Q_x \sigma_{ij} b_i b_j = \frac{1}{n}\sum_{m=1}^{J_n}\sum_{i,j}\alpha_{ij}\sigma_{ij}1\{\gamma_i, \gamma_j \in I_m\},$$

where $\alpha_{ij} = w_i w_j \frac{\tau_i\tau_j}{\tau_{ij}} b_i b_j E[x_t^2]$. Thus,

$$\begin{aligned}
&\operatorname{Cov}\big[E\big(\xi_{n,t}^2\big|\mathcal{G}_n\big), E\big(\xi_{n,s}^2\big|\mathcal{G}_n\big)\big] \\
&\quad = \frac{1}{n^2}\sum_{m,p=1}^{J_n}\sum_{i,j,k,l}\operatorname{Cov}\big(\alpha_{ij}\sigma_{ij}1\{\gamma_i, \gamma_j \in I_m\}, \alpha_{kl}\sigma_{kl}1\{\gamma_k, \gamma_l \in I_p\}\big).
\end{aligned}$$

In the above sum, the terms such that sets $\{i, j\}$ and $\{k, l\}$ do not have a common element, vanish. Consider now the sum of the terms such that $i = k$ (terms such that $i = l$, or $j = k$, or $j = l$ are symmetric). Therefore, let us focus on the sum

$$\begin{aligned}
S_n &:= \frac{1}{n^2}\sum_{m,p=1}^{J_n}\sum_{i,j,l}\operatorname{Cov}\big(\alpha_{ij}\sigma_{ij}1\{\gamma_i, \gamma_j \in I_m\}, \alpha_{il}\sigma_{il}1\{\gamma_i, \gamma_l \in I_p\}\big) \\
&= \frac{1}{n^2}\sum_{m=1}^{J_n}\sum_{i,j,l}\operatorname{Cov}\big(\alpha_{ij}\sigma_{ij}1\{\gamma_i, \gamma_j \in I_m\}, \alpha_{il}\sigma_{il}1\{\gamma_i, \gamma_l \in I_m\}\big) \\
&\quad - \frac{1}{n^2}\sum_{m,p=1,m\neq p}^{J_n}\sum_{i,j,l}E\big[\alpha_{ij}\sigma_{ij}1\{\gamma_i, \gamma_j \in I_m\}\big]E\big[\alpha_{il}\sigma_{il}1\{\gamma_i, \gamma_l \in I_p\}\big].
\end{aligned}$$

From Assumption BD.4, we have $\alpha_{ij} \leq C$ and $\sigma_{ij} \leq C$. Thus, we get $S_n = O(\frac{1}{n^2}\sum_{m=1}^{J_n}\sum_{i,j,l}E[1\{\gamma_i, \gamma_j, \gamma_l \in I_m\}]) + O(\frac{1}{n^2}\sum_{m,p=1,m\neq p}^{J_n}\sum_{i,j,l}E[1\{\gamma_i, \gamma_j \in$

$I_m\}]E[1\{\gamma_i, \gamma_l \in I_p\}])$. By using that $\sum_{i,j,l} E[1\{\gamma_i, \gamma_j, \gamma_l \in I_m\}] = O(nB_m + n^2B_m^2 + n^3B_m^3)$ and $\sum_{i,j,l} E[1\{\gamma_i, \gamma_j \in I_m\}]E[1\{\gamma_i, \gamma_l \in I_p\}] = O(nB_mB_p + n^2(B_m^2B_p + B_mB_p^2) + n^3B_m^2B_p^2))$, we get $S_n = O(1/n + \sum_{m=1}^{J_n} B_m^2 + n\sum_{m=1}^{J_n} B_m^3 + n(\sum_{m=1}^{J_n} B_m^2)^2)$. The RHS is $o(1)$ from Assumption BD.2. Thus, we have shown that

$$\text{Cov}(\xi_{n,t}^2, \xi_{n,s}^2) = o(1), \tag{63}$$

uniformly in $t \neq s$.

- Consider now $V[\xi_{n,t}^2]$. By the variance decomposition formula, $V[\xi_{n,t}^2] = E[V(\xi_{n,t}^2|\mathcal{G}_n)] + V[E(\xi_{n,t}^2|\mathcal{G}_n)]$. By similar arguments as above, we have $V[E(\xi_{n,t}^2|\mathcal{G}_n)] = o(1)$ uniformly in t. Consider now term $E[V(\xi_{n,t}^2|\mathcal{G}_n)]$. We have

$$V(\xi_{n,t}^2|\mathcal{G}_n) = \frac{1}{n^2}\sum_{i,j,k,l} w_iw_jw_kw_l\tau_i\tau_j\tau_k\tau_l b_ib_jb_kb_l \times \text{Cov}(I_{i,t}I_{j,t}x_t^2\varepsilon_{i,t}\varepsilon_{j,t}, I_{k,t}I_{l,t}x_t^2\varepsilon_{k,t}\varepsilon_{l,t}|\gamma_i, \gamma_j, \gamma_k, \gamma_l).$$

Moreover,

$$\begin{aligned}
&\text{Cov}(I_{i,t}I_{j,t}x_t^2\varepsilon_{i,t}\varepsilon_{j,t}, I_{k,t}I_{l,t}x_t^2\varepsilon_{k,t}\varepsilon_{l,t}|\gamma_i, \gamma_j, \gamma_k, \gamma_l) \\
&\quad = E[I_{i,t}I_{j,t}I_{k,t}I_{l,t}|\gamma_i, \gamma_j, \gamma_k, \gamma_l]E[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{k,t}\varepsilon_{l,t}|\gamma_i, \gamma_j, \gamma_k, \gamma_l]E[x_t^4] \\
&\qquad - \sigma_{ij}\sigma_{kl}\tau_{ij}^{-1}\tau_{kl}^{-1}E[x_t^2]^2.
\end{aligned}$$

From the block-dependence structure in Assumption BD.1, the expectation $E[\varepsilon_{i,t}\varepsilon_{j,t}\varepsilon_{k,t}\varepsilon_{l,t}|\gamma_i, \gamma_j, \gamma_k, \gamma_l]$ is different from zero only if a pair of indices is in a same block I_m, and the other pair is also in a same block I_p, say, possibly with $m = p$. Similarly, $\sigma_{ij}\sigma_{kl}$ is different from zero only if γ_i and γ_j are in the same block and γ_k and γ_l are in the same block. From Assumption BD.4, we deduce that $V(\xi_{n,t}^2|\mathcal{G}_n) \leq C\frac{1}{n^2}\sum_{i,j,k,l}\sum_{m,p=1}^{J_n} 1\{\gamma_i, \gamma_j \in I_m\}1\{\gamma_k, \gamma_l \in I_p\}$, where in the double sum the elements with $m \neq p$ are not zero only if the pairs (γ_i, γ_j) and (γ_k, γ_l) have no element in common. Thus,

$$\begin{aligned}
&E[V(\xi_{n,t}^2|\mathcal{G}_n)] \\
&\quad = O\left(\frac{1}{n^2}\sum_{i,j,k,l}\sum_{m=1}^{J_n} E[1\{\gamma_i, \gamma_j, \gamma_k, \gamma_l \in I_m\}]\right) \\
&\qquad + O\left(\frac{1}{n^2}\sum_{i,j,k,l: i\neq k,l; j\neq k,l}\ \sum_{m,p=1: m\neq p}^{J_n} E[1\{\gamma_i, \gamma_j \in I_m\}]E[1\{\gamma_k, \gamma_l \in I_p\}]\right).
\end{aligned}$$

By using $\sum_{i,j,k,l}\sum_{m=1}^{J_n} E[1\{\gamma_i, \gamma_j, \gamma_k, \gamma_l \in I_m\}] = O(\sum_{m=1}^{J_n}(nB_m + n^2B_m^2 + n^3B_m^3 + n^4B_m^4))$ and $\sum_{i,j,k,l}\sum_{m,p=1}^{J_n} E[1\{\gamma_i, \gamma_j \in I_m\}]E[1\{\gamma_k, \gamma_l \in I_p\}] =$

$O(\sum_{m,p=1}^{J_n}(n^2B_mB_p + n^3B_m^2B_p + n^4B_m^2B_p^2))$, we get

$$E\big[V\big(\xi_{n,t}^2|\mathcal{G}_n\big)\big] = O\left(1 + n\sum_{m=1}^{J_n}B_m^2 + \left(n\sum_{m=1}^{J_n}B_m^2\right)^2 + n^2\sum_{m=1}^{J_n}B_m^4\right).$$

By Assumption BD.2, $n\max_{m=1,\dots,J_n}B_m^2 = O(1)$, and we get $E[V(\xi_{n,t}^2|\mathcal{G}_n)] = O(1)$.

Thus, we have shown

$$V\big(\xi_{n,t}^2\big) = O(1), \tag{64}$$

uniformly in t.

From (62), (63), and (64), we get $V[\zeta_{nT}] = o(1)$, and condition (ii) follows. From (64) and by using $E[\xi_{n,t}^2] = O(1)$, condition (iii) follows for $\delta = 2$. Finally, condition (iv) follows from $\frac{1}{n}\sum_{i,j}w_iw_j\frac{\tau_i\tau_j}{\tau_{ij}}\sigma_{ij}b_ib_j' = (1+\lambda'V[f_t]\lambda)^{-2} \times \frac{1}{n}\sum_{i,j}\frac{1}{\tau_{ij}}\frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}}b_ib_j'$ and the next Lemma 15.

LEMMA 15: *Under Assumptions* BD.1–BD.4: $\frac{1}{n}\sum_{i,j}\frac{1}{\tau_{ij}}\frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}}b_ib_j' \to L$, *P-a.s., where*

$$\begin{aligned} L &= \lim_{n\to\infty}E\left[\frac{1}{n}\sum_{i,j}\frac{1}{\tau_{ij}}\frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}}b_ib_j'\right] \\ &= \int_0^1\omega(\gamma)\,d\gamma + \lim_{n\to\infty}n\sum_{m=1}^{J_n}\int_{I_m}\int_{I_m}\omega\big(\gamma,\gamma'\big)\,d\gamma\,d\gamma', \end{aligned}$$

with $\omega(\gamma,\gamma') := E[I_t(\gamma)I_t(\gamma')]\frac{E[\varepsilon_t(\gamma)\varepsilon_t(\gamma')]}{E[\varepsilon_t(\gamma)^2]E[\varepsilon_t(\gamma')^2]}b(\gamma)b(\gamma')'$ *and* $\omega(\gamma) := \omega(\gamma,\gamma)$.

Then, we have proved part (a). Part (b) follows by a standard CLT.

F.2.4. *Assumption A.3*

Assumption A.3 is satisfied since the errors are i.i.d. and have zero third moment (Assumption BD.1).

F.2.5. *Assumption A.4*

We have to show that $\max_i\sum_j\|S_{ij}\|^{\bar{q}} = O_p(n^{\bar{\delta}})$, for any $\bar{q}\in(0,1)$ and $\bar{\delta} > 1/2$. From $S_{ij} = \sigma_{ij}Q_x$, and an argument similar to (61),

$$\begin{aligned} \max_i\sum_j\|S_{ij}\|^{\bar{q}} &\le C\max_{m=1,\dots,J_n}\sum_{j=1}^n 1\{\gamma_j\in I_m\} \\ &\le Cn\max_{m=1,\dots,J_n}B_m + C\max_{m=1,\dots,J_n}\left|\sum_{j=1}^n\big[1\{\gamma_j\in I_m\} - B_m\big]\right|, \end{aligned}$$

for any $\bar{q} > 0$. Let us derive (probability) bounds for the two terms in the RHS. From Assumption BD.2,

$$n \max_m |B_m| \leq \sqrt{n}\left(n \sum_m |B_m|^2\right)^{1/2} = O(\sqrt{n}).$$

Let $\varepsilon_n := n^{\bar{\delta}}$, with $\bar{\delta} > 1/2$. Then,

$$\begin{aligned} &P\left[\max_{m=1,\ldots,J_n}\left|\sum_{j=1}^{n}\left[1\{\gamma_j \in I_m\} - B_m\right]\right| \geq \varepsilon_n\right] \\ &\leq J_n \max_{m=1,\ldots,J_n} P\left[\left|\sum_{j=1}^{n}\left[1\{\gamma_j \in I_m\} - B_m\right]\right| \geq \varepsilon_n\right] \\ &\leq 2J_n \exp\left(-\varepsilon_n^2/(2n)\right) = o(1), \end{aligned}$$

from the Hoeffding inequality (see Bosq (1998, Theorem 1.2)), and $J_n \leq n$. Thus, we have shown that $\max_{m=1,\ldots,J_n} |\sum_{j=1}^{n}[1\{\gamma_j \in I_m\} - B_m]| = o_p(n^{\bar{\delta}})$, and the conclusion follows.

F.2.6. *Assumption A.5*

In the time-invariant i.i.d. case, we have $S_{ii,T} = \sigma_{ii}\hat{Q}_{x,i}$ and $S_{ij} = \sigma_{ij}Q_x$. Then, Assumption A.5 boils down to $Y_{nT} := \frac{1}{\sqrt{n}}\sum_i w_i\tau_i^2[Y_{i,T} \otimes Y_{i,T} - \widetilde{S}_{ii,T}] \Rightarrow N(0, \Omega)$, as $n, T \to \infty$, where $\widetilde{S}_{ii,T} = \sigma_{ii}\,\text{vec}(\hat{Q}_{x,i})$ and $\Omega = \lim_{n\to\infty} E[\frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\sigma_{ij}^2] \times [Q_x \otimes Q_x + (Q_x \otimes Q_x)W_{K+1}]$. Let us denote by $\mathcal{H} = \sigma((f_t), (I_t(\gamma)), \gamma \in [0,1], \gamma_i, i = 1, 2, \ldots)$ the information in the factor path, the indicators paths, and the individual random effects. The proof is in two steps.

Step 1: We first show that, conditional on $\mathcal{H}$, we have

$$Y_{nT} \Rightarrow N(0, \Omega), \quad n, T \to \infty, \tag{65}$$

P-a.s. For this purpose, we apply the Lyapunov CLT for heterogeneous independent arrays (see Davidson (1994, Theorem 23.11)). Write

$$Y_{nT} = \frac{1}{\sqrt{n}}\sum_i \sum_{m=1}^{J_n} 1\{\gamma_i \in I_m\}w_i\tau_i^2[Y_{i,T} \otimes Y_{i,T} - \widetilde{S}_{ii,T}] = \frac{1}{\sqrt{J_n}}\sum_{m=1}^{J_n} W_{m,nT},$$

where

$$W_{m,nT} := \sqrt{\frac{J_n}{n}}\sum_i 1\{\gamma_i \in I_m\}w_i\tau_i^2[Y_{i,T} \otimes Y_{i,T} - \widetilde{S}_{ii,T}].$$

Conditional on $\mathcal{H}$, the variables $W_{m,nT}$, for $m = 1, \ldots, J_n$, are independent, with zero mean. The conclusion follows if we prove:

(i) $\lim_{n,T\to\infty} \frac{1}{J_n}\sum_m V[W_{m,nT}|\mathcal{H}] = \Omega$, P-a.s., and

(ii) $\lim_{n,T\to\infty} \frac{1}{J_n^{3/2}}\sum_m E[\|W_{m,nT}\|^3|\mathcal{H}] = 0$, P-a.s.

To show (i), we use

$$\begin{aligned} &V[W_{m,nT}|\mathcal{H}] \\ &\quad = \frac{J_n}{n}\sum_{i,j\in I_m} w_i w_j \tau_i^2\tau_j^2 \operatorname{Cov}[Y_{i,T}\otimes Y_{i,T}, Y_{j,T}\otimes Y_{j,T}|\mathcal{H}] \\ &\quad = \frac{J_n}{n}\sum_{i,j\in I_m} w_i w_j \tau_i^2\tau_j^2 \{E[(Y_{i,T}\otimes Y_{i,T})(Y_{j,T}\otimes Y_{j,T})'|\mathcal{H}] - \widetilde{S}_{ii,T}\widetilde{S}'_{jj,T}\}, \end{aligned}$$

where $\sum_{i,j\in I_m}$ denotes double sum over all $i, j = 1, \ldots, n$ such that $\gamma_i, \gamma_j \in I_m$. Now, we have, by the independence property over time,

$$\begin{aligned} &E[(Y_{i,T}\otimes Y_{i,T})(Y_{j,T}\otimes Y_{j,T})'|\mathcal{H}] \\ &\quad = \frac{1}{T^2}\sum_t\sum_s\sum_p\sum_q E[\varepsilon_{i,t}\varepsilon_{i,p}\varepsilon_{j,s}\varepsilon_{j,q}|(f_t), \gamma_i, \gamma_j] \\ &\qquad \times I_{i,t}I_{i,p}I_{j,s}I_{j,q}(x_t x_s'\otimes x_p x_q') \\ &\quad = E[\varepsilon_{it}^2\varepsilon_{jt}^2|\gamma_i, \gamma_j]\frac{1}{T^2}\sum_t I_{i,t}I_{j,t}(x_t x_t'\otimes x_t x_t') \\ &\qquad + \sigma_{ij}^2\frac{1}{T^2}\sum_t\sum_{p\neq t} I_{ij,t}I_{ij,p}(x_t x_t'\otimes x_p x_p') \\ &\qquad + \sigma_{ii}^2\sigma_{jj}^2\frac{1}{T^2}\sum_t\sum_{s\neq t} I_{i,t}I_{j,s}(x_t x_s'\otimes x_t x_s') \\ &\qquad + \sigma_{ij}^2\frac{1}{T^2}\sum_t\sum_{s\neq t} I_{ij,t}I_{ij,s}(x_t x_s'\otimes x_s x_t') \\ &\quad =: E[\varepsilon_{it}^2\varepsilon_{jt}^2|\gamma_i, \gamma_j]A_{1,T} + \sigma_{ij}^2 A_{2,T} + \sigma_{ii}^2\sigma_{jj}^2 A_{3,T} + \sigma_{ij}^2 A_{4,T}. \end{aligned}$$

Moreover, $A_{1,T} = \frac{T_{ij}}{T^2}\sum_t \frac{I_{ij,t}}{T_{ij}}(x_t x_t'\otimes x_t x_t') = O(T_{ij}/T^2) = O(1/T)$, uniformly in $\mathcal{H}$. Let us define $\hat{Q}_{x,ij} = \frac{1}{T_{ij}}\sum_t I_{ij,t}x_t x_t'$; then

$$\begin{aligned} A_{2,T} &= \frac{1}{T^2}\sum_t\sum_p I_{ij,t}I_{ij,p}(x_t x_t'\otimes x_p x_p') - A_{1,T} \\ &= \frac{1}{\tau_{ij,T}^2}(\hat{Q}_{x,ij}\otimes\hat{Q}_{x,ij}) + O(1/T), \end{aligned}$$

$$
\begin{aligned}
A_{3,T} &= \frac{1}{T^2}\sum_t\sum_s I_{i,t}I_{j,s}\big(x_t x_s' \otimes x_t x_s'\big) - A_{1,T} \\
&= \mathrm{vec}(\hat{Q}_{x,i})\,\mathrm{vec}(\hat{Q}_{x,j})' + O(1/T),
\end{aligned}
$$

and

$$
\begin{aligned}
A_{4,T} &= \frac{1}{T^2}\sum_t\sum_s I_{ij,t}I_{ij,s}\big(x_t x_s' \otimes x_s x_t'\big) - A_{1,T} \\
&= \frac{1}{T^2}\sum_t\sum_s I_{ij,t}I_{ij,s}(x_t \otimes x_s)(x_s \otimes x_t)' - A_{1,T} \\
&= \frac{1}{T^2}\sum_t\sum_s I_{ij,t}I_{ij,s}(x_t \otimes x_s)(x_t \otimes x_s)' W_{K+1} - A_{1,T} \\
&= \frac{1}{\tau_{ij,T}^2}(\hat{Q}_{x,ij} \otimes \hat{Q}_{x,ij}) W_{K+1} + O(1/T).
\end{aligned}
$$

Then, it follows that

$$
\begin{aligned}
V[W_{m,nT}|\mathcal{H}] = \frac{J_n}{n}\Bigg[\sum_{i,j\in I_m} w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij,T}^2}\sigma_{ij}^2(\hat{Q}_{x,ij}\otimes\hat{Q}_{x,ij} + \hat{Q}_{x,ij}\otimes\hat{Q}_{x,ij}W_{K+1})\Bigg] \\
+ O\bigg(\frac{J_n}{n}\frac{1}{T}\sum_{i,j\in I_m} w_i w_j \tau_i^2\tau_j^2\bigg),
\end{aligned}
$$

where the O term is uniform w.r.t. $\mathcal{H}$. Thus, we get

$$
\begin{aligned}
&\frac{1}{J_n}\sum_m V[W_{m,nT}|\mathcal{H}] \\
&\quad = \bigg(\frac{1}{n}\sum_{i,j} w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\sigma_{ij}^2\bigg)(Q_x \otimes Q_x + Q_x \otimes Q_x W_{K+1}) \\
&\qquad + \frac{1}{n}\sum_m\sum_{i,j\in I_m} w_i w_j \tau_i^2\tau_j^2\sigma_{ij}^2\alpha_{ij} + O\bigg(\frac{1}{T}\frac{1}{n}\sum_m\sum_{i,j\in I_m} w_i w_j \tau_i^2\tau_j^2\bigg),
\end{aligned}
$$

where the $\alpha_{ij} = \frac{1}{\tau_{ij,T}^2}(\hat{Q}_{x,ij}\otimes\hat{Q}_{x,ij} + \hat{Q}_{x,ij}\otimes\hat{Q}_{x,ij}W_{K+1}) - \frac{1}{\tau_{ij}^2}(Q_x \otimes Q_x + Q_x \otimes Q_x W_{K+1})$ are $o(1)$ uniformly in i, j, and $w_i w_j \frac{\tau_i^2\tau_j^2}{\tau_{ij}^2}\sigma_{ij}^2 = (1+\lambda'\Sigma_f^{-1}\lambda)^{-2}\frac{\tau_i\tau_j}{\tau_{ij}^2}\frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}$.

Then, point (i) follows from $\frac{1}{n}\sum_{i,j}\frac{\tau_i\tau_j}{\tau_{ij}^2}\frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}} \to L$, P-a.s., where $L = \lim_{n\to\infty} E[\frac{1}{n}\sum_{i,j}\frac{\tau_i\tau_j}{\tau_{ij}^2}\frac{\sigma_{ij}^2}{\sigma_{ii}\sigma_{jj}}]$, which is proved by similar arguments as Lemma 15.

Let us now prove point (ii). We have

$$\begin{aligned}
&\frac{1}{J_n^{3/2}}\sum_m E\big[\|W_{m,nT}\|^3|\mathcal{H}\big] \\
&\quad\le \frac{1}{n^{3/2}}\sum_m\Big[\sum_{i\in I_m} w_i\tau_i^2\big(E\big[\big\|(Y_{i,T}\otimes Y_{i,T})\big\|^3|\mathcal{H}\big]^{1/3}+\|\widetilde{S}_{ii,T}\|\big)\Big]^3 \\
&\quad\le \frac{1}{n^{3/2}}\bigg(\sum_m\Big(\sum_{i\in I_m} w_i\tau_i^2\Big)^3\bigg) \\
&\qquad\times\Big(\sup_i E\big[\|Y_{i,T}\otimes Y_{i,T}\|^3|\mathcal{H}\big]^{1/3}+\sup_i\|\widetilde{S}_{ii,T}\|\Big)^3.
\end{aligned}$$

Now,

$$\begin{aligned}
&E\big[\|Y_{i,T}\otimes Y_{i,T}\|^3|\mathcal{H}\big] \\
&\quad\le E\big[\|Y_{i,T}\|^6|\mathcal{H}\big] \\
&\quad= E\big[\big(Y_{i,T}'Y_{i,T}\big)^3|\mathcal{H}\big] \\
&\quad= \frac{1}{T^3}\sum_{t_1,\dots,t_6} I_{i,t_1}\cdots I_{i,t_6}E[\varepsilon_{i,t_1}\cdots\varepsilon_{i,t_6}|\gamma_i]\big(x_{t_1}'x_{t_2}\big)\big(x_{t_3}'x_{t_4}\big)\big(x_{t_5}'x_{t_6}\big).
\end{aligned}$$

By the independence property, the nonzero terms $E[\varepsilon_{i,t_1}\cdots\varepsilon_{i,t_6}|\gamma_i]$ involve at most three different time indices, which implies, together with Assumption BD.4, that $\sup_i E[\|Y_{i,T}\otimes Y_{i,T}\|^3|\mathcal{H}] = O(1)$, P-a.s. Similarly $\sup_i\|\widetilde{S}_{ii,T}\| = O(1)$, P-a.s. Thus, we get

$$\frac{1}{J_n^{3/2}}\sum_{m=1}^{J_n} E\big[\|W_{m,nT}\|^3|\mathcal{H}\big] \le C\frac{1}{n^{3/2}}\sum_{m=1}^{J_n}\Big(\sum_i 1\{\gamma_i\in I_m\}\Big)^3.$$

Then, point (ii) follows from the next Lemma 16.

LEMMA 16: *Under Assumptions* BD.1–BD.4: $\frac{1}{n^{3/2}}\sum_{m=1}^{J_n}(\sum_i 1\{\gamma_i\in I_m\})^3 \to 0$, *$P$-a.s.*

Step 2: We show that (65) implies the asymptotic normality condition in Assumption A.4. Indeed, from (65), we have

$$\lim_{n,T\to\infty} P\big[\alpha' Y_{nT}\le z|\mathcal{H}\big] = \Phi\bigg(\frac{z}{\sqrt{\alpha'\Omega\alpha}}\bigg),$$

for any $\alpha \in \mathbb{R}^{2(K+1)}$ and for any $z \in \mathbb{R}$, and P-a.s. We now apply the Lebesgue dominated convergence theorem, by using that the sequence of random variables $P[\alpha' Y_{nT} \leq z|\mathcal{H}]$ is such that $P[\alpha' Y_{nT} \leq z|\mathcal{H}] \leq 1$, uniformly in n and T. We conclude that, for any $\alpha \in \mathbb{R}^{2(K+1)}$, $z \in \mathbb{R}$,

$$\lim_{n,T\to\infty} P\big[\alpha' Y_{nT} \leq z\big] = \lim_{n,T\to\infty} E\big(P\big[\alpha' Y_{nT} \leq z|\mathcal{H}\big]\big) = \Phi\bigg(\frac{z}{\sqrt{\alpha'\Omega\alpha}}\bigg),$$

since $\Phi(\frac{z}{\sqrt{\alpha'\Omega\alpha}})$ is independent of the information set $\mathcal{H}$. The conclusion follows.

F.3. *Proof of Lemma 15*

Let us denote $\xi_{i,j} = \frac{1}{\tau_{ij}}\frac{\sigma_{ij}}{\sigma_{ii}\sigma_{jj}} b_i b_j' = w(\gamma_i, \gamma_j)$. We have $\frac{1}{n}\sum_{i,j}\xi_{i,j} = \frac{1}{n}\sum_i \xi_{ii} + \frac{1}{n}\sum_{i\neq j}\xi_{i,j}$. By the LLN, we get $\frac{1}{n}\sum_i \xi_{ii} = \frac{1}{n}\sum_i \omega(\gamma_i) \to \int_0^1 \omega(\gamma)\,d\gamma$, P-a.s. Let us now consider the double sum $\frac{1}{n}\sum_{i\neq j}\xi_{i,j}$. The proof proceeds in three steps.

Step 1: We first prove that $\frac{1}{n}\sum_{i\neq j}\xi_{i,j} = L' + o_p(1)$, where $L' := \lim_{n\to\infty} n \times \sum_{m=1}^{J_n}\int_{I_m}\int_{I_m}\omega(\gamma,\gamma')\,d\gamma\,d\gamma'$. For this purpose, write $\frac{1}{n}\sum_{i\neq j}\xi_{i,j} = \sum_{m=1}^{J_n} X_m$, where $X_m := \frac{1}{n}\sum_{i\neq j}\omega(\gamma_i,\gamma_j)1\{\gamma_i,\gamma_j \in I_m\}$, by using block-dependence. Then, we have

$$\begin{aligned} E[X_m] &= \frac{1}{n}\sum_{i\neq j} E\big[\omega(\gamma_i,\gamma_j)1\{\gamma_i,\gamma_j\in I_m\}\big] \\ &= (n-1)\int_{I_m}\int_{I_m}\omega\big(\gamma,\gamma'\big)\,d\gamma\,d\gamma' =: (n-1)\bar{\omega}_m, \end{aligned}$$

which implies

$$E\bigg[\frac{1}{n}\sum_{i\neq j}\xi_{i,j}\bigg] = (n-1)\sum_{m=1}^{J_n}\bar{\omega}_m \to L'.$$

Moreover,

$$\begin{aligned} V[X_m] &= \frac{1}{n^2}\sum_{i\neq j}\sum_{k\neq l} E\big[\omega(\gamma_i,\gamma_j)\omega(\gamma_k,\gamma_l)1\{\gamma_i,\gamma_j,\gamma_k,\gamma_l\in I_m\}\big] \\ &\quad - E[X_m]^2 \\ &= \frac{1}{n^2}\big[n(n-1)(n-2)(n-3)\bar{\omega}_m^2 + O\big(n^3B_m^3\big) + O\big(n^2B_m^2\big)\big] \\ &\quad - (n-1)^2\bar{\omega}_m^2 \\ &= O\big(nB_m^4\big) + O\big(nB_m^3\big) + O\big(B_m^2\big), \end{aligned}$$

and

$$
\begin{aligned}
&\mathrm{Cov}(X_m, X_p) \\
&= \frac{1}{n^2} \sum_{i \neq j} \sum_{k \neq l} E\big[\omega(\gamma_i, \gamma_j)\omega(\gamma_k, \gamma_l) 1\{\gamma_i, \gamma_j \in I_m\} 1\{\gamma_k, \gamma_l \in I_p\}\big] \\
&\quad - E[X_m]E[X_p] \\
&= \frac{1}{n^2}\big[n(n-1)(n-2)(n-3)\bar{\omega}_m \bar{\omega}_p\big] - (n-1)^2 \bar{\omega}_m \bar{\omega}_p \\
&= O\big(n B_m^2 B_p^2\big),
\end{aligned}
$$

for $m \neq p$, which implies

$$
V\left[\frac{1}{n} \sum_{i \neq j} \xi_{i,j}\right] = \sum_{m=1}^{J_n} V[X_m] + \sum_{m,p=1, m \neq p}^{J_n} \mathrm{Cov}(X_m, X_p) = o(1),
$$

from Assumption BD.2. Then, Step 1 follows.

Step 2: There exists a random variable $\tilde{L}$ such that $\frac{1}{n} \sum_{i \neq j} \xi_{i,j} \to \tilde{L}$, P-a.s. To show this statement, we use that the event in which series $\frac{1}{n} \sum_{i \neq j} \xi_{i,j}$ converges is a tail event for the i.i.d. sequence (γ_i). Indeed, we have that $\frac{1}{n} \sum_{i \neq j} \xi_{i,j}$ converges if, and only if, $\frac{1}{n} \sum_{i,j \geq N, i \neq j} \xi_{i,j}$ converges, for any integer N. Then, by the Kolmogorov zero–one law, the event in which series $\frac{1}{n} \sum_{i \neq j} \xi_{i,j}$ converges has probability either 1 or 0. The latter case, however, is excluded by Step 1. Therefore, the sequence $\frac{1}{n} \sum_{i \neq j} \xi_{i,j}$ converges with probability 1, and Step 2 follows.

Step 3: We have $\tilde{L} = L'$, with probability 1. Indeed, by Steps 1 and 2, it follows that $\frac{1}{n} \sum_{i \neq j} \xi_{i,j} - L' = o_p(1)$ and $\frac{1}{n} \sum_{i \neq j} \xi_{i,j} - \tilde{L} = o_p(1)$. These equations imply that $\tilde{L} - L' = o_p(1)$, which holds if and only if $\tilde{L} = L$ with probability 1 (since $\tilde{L}$ and L' are independent of n).

F.4. *Proof of Lemma 16*

The proof is similar to the one of Lemma 15 and we give only the main steps. First, we prove that $\frac{1}{n^{3/2}} \sum_{m=1}^{J_n} (\sum_i 1\{\gamma_i \in I_m\})^3 = o_p(1)$. Indeed, we have

$$
\begin{aligned}
E\left[\frac{1}{n^{3/2}} \sum_{m=1}^{J_n} \left(\sum_i 1\{\gamma_i \in I_m\}\right)^3\right] &= \frac{1}{n^{3/2}} \sum_{m=1}^{J_n} \sum_{i,j,k} E[1\gamma_i, \gamma_j, \gamma_k \in I_m] \\
&= O\left(n^{3/2} \sum_{m=1}^{J_n} B_m^3\right) = o(1),
\end{aligned}
$$

from Assumption BD.2, and we can show $V[\frac{1}{n^{3/2}}\sum_{m=1}^{J_n}(\sum_i 1\{\gamma_i \in I_m\})^3] = o(1)$. Second, by using the monotone convergence theorem and the Kolmogorov zero–one law, we can show that sequence $\frac{1}{n^{3/2}}\sum_{m=1}^{J_n}(\sum_i 1\{\gamma_i \in I_m\})^3$ converges with probability 1. Third, we conclude that the limit is 0 with probability 1.

APPENDIX G: MONTE CARLO EXPERIMENTS

In this appendix, we report the results of Monte Carlo experiments to investigate the finite sample behavior of our estimators and test statistics (Section G.1) and the accuracy of the CLT asymptotic approximations underlying Assumption A.2(a) (Section G.2).

G.1. *Finite Sample Behavior of Estimators and Test Statistics*

In this section, we perform simulation exercises on balanced and unbalanced panels in order to study the properties of our estimation and testing approaches. We pay particular attention to the interaction between panel dimensions n and T in finite samples since we face conditions like $n = o(T^3)$ for inference with $\hat{\nu}$, and $n = o(T^2)$ for inference with $\hat{Q}_e$ and $\hat{Q}_a$, in the theoretical results. The simulation design mimics the empirical features of our data. The balanced case serves as a benchmark to understand when T is not sufficiently large w.r.t. n to apply the theory. The unbalanced case shows that we can exploit the guidelines found for the balanced case when we substitute the average of the sample sizes of the individual assets, that is, a kind of operative sample size, for T. To summarize our Monte Carlo findings, we do not face any finite sample distortions for the inference with $\hat{\nu}$ when $n = 1000$ and $T = 150$, and with $\hat{Q}_e$ and $\hat{Q}_a$ when $n = 1000$ and $T = 350$. In light of these results, we do not expect to face significant inference bias in our empirical application.

G.1.1. *Balanced Panel*

We simulate S data sets of excess returns from a time-invariant one-factor model (CAPM), we estimate the parameter ν, and compute the test statistics. A simulated data set includes: a vector of intercepts $a^s \in \mathbb{R}^n$, a vector of factor loadings $b^s \in \mathbb{R}^n$, and a variance–covariance matrix $\Omega^s \in \mathbb{R}^{n\times n}$. At each simulation $s = 1, \dots, S$, we randomly draw $n \leq 9904$ assets from the sample of our empirical analysis that comprises 9904 individual stocks. The assets are listed by industrial sectors. We use the classification proposed by Ferson and Harvey (1999). The vector b^s is composed by the estimated factor loadings for the n randomly chosen assets. At each simulation, we build a block-diagonal matrix Ω^s with blocks matching industrial sectors. The n elements of the main diagonal of Ω^s correspond to the variances of the estimated residuals of the individual assets. The off-diagonal elements of Ω^s are covariances computed by fixing correlations within a block equal to the average correlation of the industrial sector computed from the 9904×9904 thresholded variance–covariance matrix of estimated residuals. Hence, we get a setting in line with the block-dependence case developed in Appendix F.

In order to study the size and power properties of our procedure, we set the values of the intercepts a_i^s according to four data generating processes:

DGP1: The true parameter is $\nu_0 = 0.00\%$ and the a_i^s are generated under the null hypothesis $\mathcal{H}_0 : a_i^s = 0$.

DGP2: The true parameter is the empirical estimate of ν, $\nu_0 = 2.57\%$, and the a_i^s are generated under the null hypothesis $\mathcal{H}_0 : a_i^s = b_i^s \nu_0$.

DGP3: The a_i^s are generated under the alternative hypothesis $\mathcal{H}_a : a_i^s = (0.5 b_i^s + 0.5)\nu_0$, where $\nu_0 = 2.57\%$.

DGP4: The a_i^s are generated under the three-factor alternative hypothesis: $\mathcal{H}_a : a_i^s = b_{i,(3)}^{s\prime} \nu_{0,(3)}$ where $b_{i,(3)}^s \in \mathbb{R}^3$ and $\nu_{0,(3)} = [2.92\%, -0.63\%, -9.96\%]'$ are estimates for the Fama–French model on the CRSP data set.

DGP1 and DGP2 match two different null hypotheses. The null hypothesis for DGP1 assumes that the factor comes from a tradable asset, and for DGP2 that it does not. DGP3 and DGP4 match two different alternative hypotheses as suggested by MacKinlay (1995). DGP3 is a "nonrisk-based alternative." It represents a deviation from CAPM, which is unrelated to risk: we take the one-factor model calibrated on the data with intercepts deviating from the no-arbitrage restriction. DGP4 is a "risk-based alternative." It represents a deviation from CAPM, which comes from missing risk factors: we take intercepts from a three-factor model calibrated on the data, and then we estimate a one-factor model.

Let us define the simulated excess returns $R_{i,t}^s$ of asset i at time t as follows:

$$R_{i,t}^s = a_i^s + b_i^s f_t + \varepsilon_{i,t}^s, \quad \text{for} \quad i = 1, \dots, n, \text{ and } t = 1, \dots, T, \tag{66}$$

where f_t is the market excess return and $\varepsilon_{i,t}^s$ is the error term. The $n \times 1$ error vectors $\boldsymbol{\varepsilon}_t^s$ are independent across time and Gaussian with mean zero and variance–covariance matrix Ω^s. We apply our estimation approach on every simulated data set of excess returns. We estimate the parameter ν and we compute the statistics described in Section 3.5 of the paper. Since the panel is balanced, we do not need to fix $\chi_{2,T}$. We only use $\chi_{1,T} = 15$. However, this trimming level does not affect the number of assets n in the simulations. In order to compute the thresholded estimator of the variance–covariance matrix of $\hat{\nu}$, namely $\tilde{\Sigma}_\nu$ (see Proposition 5 in the paper), and the thresholded variance estimator $\tilde{\Sigma}_\xi$ (see Proposition 6) for the test statistics, we fix the parameter M equal to 0.0780, that is used in the empirical application. We define the parameter M using a cross-validation method as proposed in Bickel and Levina (2008). We build random subsamples from the CRSP sample. For each subsample, we minimize a risk function that exploits the difference between a thresholded variance–covariance matrix and a target variance–covariance matrix (see Bickel and Levina (2008) for details).

TABLE IV

ESTIMATION OF ν, BALANCED CASE

	DGP 1				DGP 2			
	n				n			
	1000	3000	6000	9000	1000	3000	6000	9000
				$T = 150$				
Bias($\hat{\nu}$)	−0.0742	−0.0567	−0.0585	−0.0586	−0.1630	−0.1472	−0.1484	−0.1493
Bias($\hat{\nu}_B$)	−0.0244	−0.0063	−0.0082	−0.0083	−0.0319	−0.0156	−0.0169	−0.0178
Var($\hat{\nu}_B$)	0.1167	0.0394	0.0179	0.0121	0.1140	0.0401	0.0189	0.0121
RMSE($\hat{\nu}_B$)	0.3423	0.1985	0.1340	0.1102	0.3390	0.2007	0.1383	0.1114
Coverage	0.9320	0.9290	0.9350	0.9370	0.9370	0.9290	0.9320	0.9360
				$T = 500$				
Bias($\hat{\nu}$)	−0.0587	−0.0640	−0.0687	−0.0654	−0.0847	−0.0926	−0.0972	−0.0937
Bias($\hat{\nu}_B$)	−0.0002	−0.0063	−0.0110	−0.0077	−0.0025	−0.0074	−0.0120	−0.0085
Var($\hat{\nu}_B$)	0.0343	0.0113	0.0060	0.0040	0.0341	0.0114	0.0061	0.0041
RMSE($\hat{\nu}_B$)	0.1851	0.1066	0.0781	0.0634	0.1846	0.1068	0.0788	0.0642
Coverage	0.9370	0.9340	0.9370	0.9390	0.9430	0.9370	0.9360	0.9320

In order to understand how our estimation approach works for different finite samples, we perform exercises combining different values of the cross-sectional dimension n and the time dimension T. Table IV reports estimation results for estimator $\hat{\nu}$, and for the bias-adjusted estimator $\hat{\nu}_B$, under DGP 1 and 2. The results include the bias of both estimators, the variance and the root mean square error (RMSE) of estimator $\hat{\nu}_B$, and the coverage of the 95% confidence interval for parameter ν based on Proposition 5. The bias of estimator $\hat{\nu}$ is decreasing in absolute value with time series size T and is rather stable w.r.t. cross-sectional size n. The analytical bias correction is rather effective, and the bias of estimator $\hat{\nu}_B$ is small. For instance, for sample sizes $T = 150$ and $n = 1000$, under DGP 2 the bias of estimator $\hat{\nu}_B$ is equal to −0.03, which in absolute value is about 1% of the true value of the parameter $\nu = 2.57$. The variance of estimator $\hat{\nu}_B$ is decreasing w.r.t. both time series and cross-sectional sample sizes T and n. These features reflect the large sample distribution of the estimators derived in Proposition 4. The coverage of the confidence intervals is close to the nominal level 95% across the considered designs and sample sizes.

In Table V, we display the rejection rates for the test of the null hypothesis $\nu = 0$ (tradable factor). This null hypothesis is satisfied in DGP 1, and the rejection rates are rather close to the nominal size 5% of the test, with a slight over-rejection. In DGP 2, parameter ν is different from zero, and the test features a power equal to 100%.

Tables VI and VII report the results for the tests of the null hypotheses $\mathcal{H}_0 : a(\gamma) = 0$ and $\mathcal{H}_0 : a(\gamma) = b(\gamma)'\nu$, respectively. The test statistics are based on $\hat{Q}_a$ and $\hat{Q}_e$ as defined in Proposition 6. DGP 1 satisfies the null hypothe-

TABLE V

TEST OF $\nu = 0$, BALANCED CASE

	DGP 1				DGP 2			
	n				n			
	1000	3000	6000	9000	1000	3000	6000	9000
				$T = 150$				
Rejection rate	0.0680	0.0710	0.0650	0.0630	1.0000	1.0000	1.0000	1.0000
				$T = 500$				
Rejection rate	0.0630	0.0660	0.0630	0.0610	1.0000	1.0000	1.0000	1.0000

sis for both tests. For $T = 150$, we observe an oversize, that is increasing w.r.t. cross-sectional size n. The time series dimension $T = 150$ is likely too small compared to cross-sectional size $n = 1000$ and this combination does not reflect the condition $n = o(T^2)$ for the validity of the asymptotic Gaussian approximation of the statistics. For $T = 500$ instead, the rejection rates of the tests are quite close to the nominal size. DGP 2 satisfies the null hypothesis of the test based on $\hat{Q}_e$, but corresponds to an alternative hypothesis for the test based on $\hat{Q}_a$. The former statistic features a similar behavior as under DGP 1, while the power of the latter statistic is increasing w.r.t. n. Finally, the power of both statistics under the "nonrisk-based" and "risk-based" alternatives in DGP 3 and DGP 4 is very large, with rejection rates close to 100% for all considered combinations of sample sizes n and T.

G.1.2. *Unbalanced Panel*

Let us repeat similar exercises as in the previous section, but with unbalanced characteristics for the simulated data sets. We introduce these characteristics through a matrix of observability indicators $I^s \in \mathbb{R}^{n\times T}$. The matrix gathers the indicator vectors for the n randomly chosen assets. We fix the maximal sample size $T = 546$ as in the empirical application. In the unbalanced setting, the excess returns $R^s_{i,t}$ of asset i at time t are

$$R^s_{i,t} = a^s_i + b^s_i f_t + \varepsilon^s_{i,t}, \quad \text{if} \quad I^s_{i,t} = 1, \text{ for } i = 1, \dots, n, \text{ and } t = 1, \dots, T, \tag{67}$$

where $I^s_{i,t}$ is the observability indicator of asset i at time t in simulation s.

In Tables VIII and IX, we provide the operative cross-sectional and time series sample sizes in the Monte Carlo repetitions for trimming $\chi_{1,T} = 15$ and four different levels of trimming $\chi_{2,T}$. More precisely, in Table VIII we report the average number $\bar{n}^\chi$ of retained assets across simulations, as well as the minimum $\min(n^\chi)$ and the maximum $\max(n^\chi)$ across simulations (rounded). For the lowest level of trimming $\chi_{2,T} = T/12$, all assets are kept in all simulations, while for the level of trimming $\chi_{2,T} = T/60$ on average we keep about

TABLE VI

TEST OF THE NULL HYPOTHESIS $\mathcal{H}_0 : a(\gamma) = 0$, BALANCED CASE

	DGP 1			DGP 2			DGP 3			DGP 4		
	n			n			n			n		
	500	1000	1500	500	1000	1500	500	1000	1500	500	1000	1500
						$T = 150$						
Size/power	0.1180	0.1400	0.1500	0.3850	0.5720	0.7170	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
						$T = 500$						
Size/power	0.0730	0.0610	0.0740	0.9240	0.9920	0.9970	0.9990	1.0000	1.0000	0.9990	1.0000	1.0000

TABLE VII

TEST OF THE NULL HYPOTHESIS $\mathcal{H}_0 : a(\gamma) = b(\gamma)\nu$, BALANCED CASE

	DGP 1			DGP 2			DGP 3			DGP 4		
	n			n			n			n		
	500	1000	1500	500	1000	1500	500	1000	1500	500	1000	1500
					$T = 150$							
Size/power	0.1110	0.1340	0.1460	0.1070	0.1360	0.1420	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000
					$T = 500$							
Size/power	0.0710	0.0570	0.0730	0.0730	0.0690	0.0750	0.9990	1.0000	1.0000	0.9990	1.0000	1.0000

TABLE VIII

OPERATIVE CROSS-SECTIONAL SAMPLE SIZE

	Trimming Level							
	$\chi_{2,T}=T/12$				$\chi_{2,T}=T/60$			
n	1000	3000	6000	9000	1000	3000	6000	9000
$\bar{n^\chi}$	1000	3000	6000	9000	660	2000	4000	6000
$\min(n^\chi)$	1000	3000	6000	9000	600	1900	3900	5900
$\max(n^\chi)$	1000	3000	6000	9000	700	2100	4100	6100
	Trimming Level							
	$\chi_{2,T}=T/120$				$\chi_{2,T}=T/240$			
n	1000	3000	6000	9000	1000	3000	6000	9000
$\bar{n^\chi}$	400	1250	2400	3600	140	430	850	1250
$\min(n^\chi)$	350	1100	2300	3500	100	370	800	1200
$\max(n^\chi)$	440	1300	2500	3650	170	470	900	1300

two thirds of the assets. In Table IX, we report the average across assets of the $\bar{T}_i$, that are the average time series size T_i for asset i across simulations, as well as the min and the max of the $\bar{T}_i$. Since the distribution of T_i for an asset i is right-skewed, we also report the average across assets of the median T_i. For trimming level $\chi_{2,T}=T/60$, the average mean time series size is about 180 months, while the average median time series size is 140 months.

In Table X, we display the results for estimators $\hat{\nu}$ and $\hat{\nu}_B$. The bias adjustment reduces substantially the bias for estimation of parameter ν. For trimming level $\chi_{2,T}=T/60$, the coverage of the confidence interval is close to the nominal size 95% for all considered cross-sectional sizes, while for $\chi_{2,T}=T/12$ the coverage deteriorates with increasing cross-sectional size. In comparison with Table IV, the bias and variance of estimator $\hat{\nu}_B$ are larger than the ones obtained in the balanced case with time series size $T=500$. However, for trimming level $\chi_{2,T}=T/60$, the results are similar to the ones with $T=150$ in Table IV. In fact, this time series size of the balanced panel reflects the operative

TABLE IX

OPERATIVE TIME SERIES SAMPLE SIZE

	Trimming Level			
	$\chi_{2,T}=T/12$	$\chi_{2,T}=T/60$	$\chi_{2,T}=T/120$	$\chi_{2,T}=T/240$
$\text{mean}(\bar{T}_i)$	130	180	240	360
$\min(\bar{T}_i)$	110	160	210	350
$\max(\bar{T}_i)$	140	190	260	380
$\text{mean}(\text{median}(T_i))$	90	140	197	330

TABLE X

ESTIMATION OF ν, UNBALANCED CASE

	DGP 1				DGP 2			
	n				n			
	1000	3000	6000	9000	1000	3000	6000	9000
	Trimming level: $\chi_{2,T} = T/12$							
Bias($\hat{\nu}$)	−0.3059	−0.3119	−0.3047	−0.3021	−0.4211	−0.4324	−0.4202	−0.4201
Bias($\hat{\nu}_B$)	−0.0893	−0.0954	−0.0880	−0.0854	−0.1127	−0.1233	−0.1113	−0.1113
Var($\hat{\nu}_B$)	0.1207	0.0409	0.0214	0.0124	0.1222	0.0405	0.0218	0.0124
RMSE($\hat{\nu}_B$)	0.3586	0.2235	0.1706	0.1402	0.3671	0.2360	0.1848	0.1574
Coverage	0.9230	0.9010	0.8740	0.8750	0.9180	0.8880	0.8410	0.8320
	Trimming level: $\chi_{2,T} = T/60$							
Bias($\hat{\nu}$)	−0.1703	−0.1738	−0.1675	−0.1596	−0.2454	−0.2478	−0.0411	−0.2329
Bias($\hat{\nu}_B$)	−0.0349	−0.0381	−0.0318	−0.0238	−0.0453	−0.0474	−0.0411	−0.0325
Var($\hat{\nu}_B$)	0.1294	0.0436	0.0231	0.0141	0.1281	0.0438	0.0232	0.0144
RMSE($\hat{\nu}_B$)	0.3613	0.2122	0.1551	0.1212	0.3606	0.2145	0.1578	0.1241
Coverage	0.9360	0.9310	0.9240	0.9350	0.9430	0.9310	0.9200	0.9300

sample sizes for that trimming level observed in Table IX. Similar comments apply for Table XI, where we report the results for the test of the hypothesis $\nu = 0$. For trimming level $\chi_{2,T} = T/60$, the size of the test is close to the nominal level 5% under DGP 1, and the power is 100% under DGP 2.

In Tables XII and XIII, we display the results for the tests based on $\hat{Q}_a$ and $\hat{Q}_e$, respectively. For trimming level $\chi_{2,T} = T/120$, we observe an oversize, that increases with the cross-sectional dimension. We get a similar behavior with more severe oversize with lower trimming levels (not reported). We expect these findings from the results in the previous section. Indeed, for trimming level $\chi_{2,T} = T/120$, the operative time series sample size in Table IX is around 200 months, and in Tables VI and VII, for a balanced panel with

TABLE XI

TEST OF $\nu = 0$, UNBALANCED CASE

	DGP 1				DGP 2			
	n				n			
	1000	3000	6000	9000	1000	3000	6000	9000
	Trimming level: $\chi_{2,T} = T/12$							
Rejection rate	0.0770	0.0990	0.1260	0.1250	1.0000	1.0000	1.0000	1.0000
	Trimming level: $\chi_{2,T} = T/60$							
Rejection rate	0.0640	0.0690	0.0760	0.0650	1.0000	1.0000	1.0000	1.0000

TABLE XII

TEST OF THE NULL HYPOTHESIS $\mathcal{H}_0 : a(\gamma) = 0$, UNBALANCED CASE

	DGP 1				DGP 2				DGP 3				DGP 4			
	n				n				n				n			
	1000	3000	6000	9000	1000	3000	6000	9000	1000	3000	6000	9000	1000	3000	6000	9000
Trimming level: $\chi_{2,T} = T/120$																
Size/power	0.1180	0.1710	0.2420	0.3030	0.6010	0.9410	0.9980	1.000	1.0000	1.0000	1.0000	1.0000	0.9990	1.0000	1.0000	1.0000
Trimming level: $\chi_{2,T} = T/240$																
Size/power	0.0880	0.0860	0.1020	0.1310	0.5320	0.8730	0.9920	1.0000	1.0000	1.0000	1.0000	1.0000	0.9740	1.0000	1.0000	1.0000

TABLE XIII

TEST OF THE NULL HYPOTHESIS $\mathcal{H}_0 : a(\gamma) = b(\gamma)\nu$, UNBALANCED CASE

	DGP 1				DGP 2				DGP 3				DGP 4			
	n				n				n				n			
	1000	3000	6000	9000	1000	3000	6000	9000	1000	3000	6000	9000	1000	3000	6000	9000
Trimming level: $\chi_{2,T} = T/120$																
Size/power	0.1130	0.1670	0.2370	0.3010	0.0940	0.2190	0.2590	0.3740	1.0000	1.0000	1.0000	1.0000	0.9990	1.0000	1.0000	1.0000
Trimming level: $\chi_{2,T} = T/240$																
Size/power	0.0800	0.0790	0.1000	0.1290	0.0790	0.0870	0.1080	0.1440	0.9990	1.0000	1.0000	1.0000	0.9690	1.0000	1.0000	1.0000

$T = 150$, the statistics are oversized. For trimming level $\chi_{2,T} = T/240$ with operative size of about 350 months, the oversize of the statistics is moderate. Finally, the power of the statistics is very large also in the unbalanced case, and close to 100%.

G.2. *The CLT in Assumption A.2(a)*

In this section, we provide simulation exercises to assess the empirical validity of the CLT in Assumption A.2(a). We simulate S data sets of error terms $\varepsilon_{i,t}$ from a time-invariant one-factor model (CAPM). At each simulation $s = 1, \dots, S$, we randomly draw $n \leq 9904$ assets from the sample of our empirical analysis, and we build a block-diagonal matrix Ω^s as described in the previous section. For each s, the $n \times 1$ error vectors $\boldsymbol{\varepsilon}_t^s$ are independent across time and Gaussian with mean zero and variance–covariance matrix Ω^s. We perform the exercise for the unbalanced case. We fix the maximal sample size $T = 546$ as in the empirical application. In the time-invariant one-factor framework, the statistic in Assumption A.2(a) reduces to $\frac{1}{\sqrt{n}} \sum_i w_i \tau_i Q_{x,i}^{-1} Y_{i,T} b_i$ with asymptotic variance $S_{v_3} = \lim_{n\to\infty} E[\frac{1}{n} \sum_{i,j} w_i w_j \frac{\tau_i \tau_j}{\tau_{ij}} S_{Q,ij} b_i b_j]$. At each simulation, we compute the 2×1 vector $\Psi^s = (S_{v_3}^s)^{-1/2} \frac{1}{\sqrt{n}} \sum_i w_i^s \tau_i^s (Q_{x,i}^s)^{-1} Y_{i,T}^s b_i^s$ with $Y_{i,T}^s = \frac{1}{\sqrt{T}} \sum_t I_{i,t}^s x_t \varepsilon_{i,t}^s$ and $S_{v_3}^s = \frac{1}{n} \sum_{i,j} w_i^s w_j^s \frac{\tau_i^s \tau_j^s}{\tau_{ij}^s} S_{Q,ij}^s b_i^s b_j^s$, where scalars $w_i^s, \tau_i^s, \tau_{ij}^s, b_i^s$, matrices $Q_{x,i}^s, S_{Q,ij}^s$, and indicator processes $(I_{i,t}^s)$ for draw s are those estimated for assets i and j in the empirical analysis.

Figures 3 and 4 compare the univariate distributions of the two components of simulated vectors $\Psi^s = [\Psi_1^s, \Psi_2^s]' \in \mathbb{R}^2$, $s = 1, \dots, 1000$, with the standard normal distribution through Q–Q plots. The cross-sectional size is $n = 1000$ in Figure 3, and $n = 3000$ in Figure 4. Figures 3 and 4 show that the finite sample distributions are well approximated by the asymptotic Gaussian distributions already for $n = 1000$. This finding suggests that the possible heavy tails in the cross-sectional distribution of asset characteristics should not affect the validity of our CLT assumptions.

APPENDIX H: MISSPECIFICATION ANALYSIS

In this appendix, we first present theoretical results on the role of misspecification and aggregation (Section H.1) and we derive the pseudo-true value of the risk premia parameter when we estimate a potentially misspecified time-invariant model using either individual assets (Section H.2) or portfolios (Section H.3). Then, we estimate these pseudo-true values using our data set (Section H.4).

H.1. *The Role of Misspecification and Aggregation*

A potential explanation of the differences between the results on individual stocks and portfolios, as well as between sets of portfolios, is the uneven

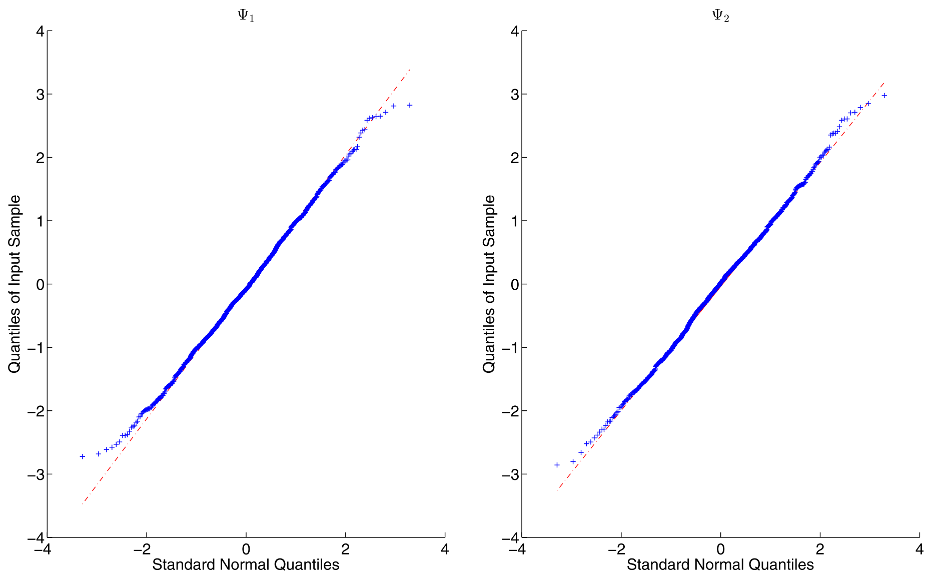

FIGURE 3.—Q–Q plots of the simulated components of Ψ for $n = 1000$. The figure compares the finite sample distributions of the two components of vector Ψ (right panel and left panel) with the standard normal distribution. We estimate the finite sample distributions with an unbalanced panel of $n = 1000$ individual stocks in the Monte Carlo exercise.

degree of misspecification of a given model across universes of assets. Using mimicking portfolio returns as observable factors and aggregating assets into portfolios may induce misspecification in the functional form of the beta dynamics. Risk premia estimated by the two-pass methodology from misspecified models converge to pseudo-true values. Estimation from individual stocks and portfolios may yield different pseudo-true values. In this section, we present theoretical and empirical arguments to support the plausibility of these claims for explaining the findings in Sections 4.2 and 4.3 of the paper.

Suppose that the data generating process (DGP) for the excess returns in the continuum economy is

$$R_t(\gamma) = c_t(\gamma) + d_t(\gamma)'h_t + \varepsilon_t(\gamma), \tag{68}$$

where h_t is an $r \times 1$ vector of "structural," or "economic," unknown factors with time-varying loadings $d_t(\gamma)$. The intercepts are $c_t(\gamma) = d_t(\gamma)'\mu_t$ for some stochastic vector μ_t because of the no-arbitrage restriction. We have $\mu_t = 0$ for tradable factors. In applying the two-pass methodology, we approximate the unobservable factors by the excess returns of some mimicking portfolios. The market, Fama–French, and momentum factors are standard examples.

Let us formalize the concept of mimicking portfolio construction. Take a weighting function $w(\gamma, \omega)$, which is $\mathcal{F}_0$-measurable w.r.t. $\omega \in \Omega$ for a.e.

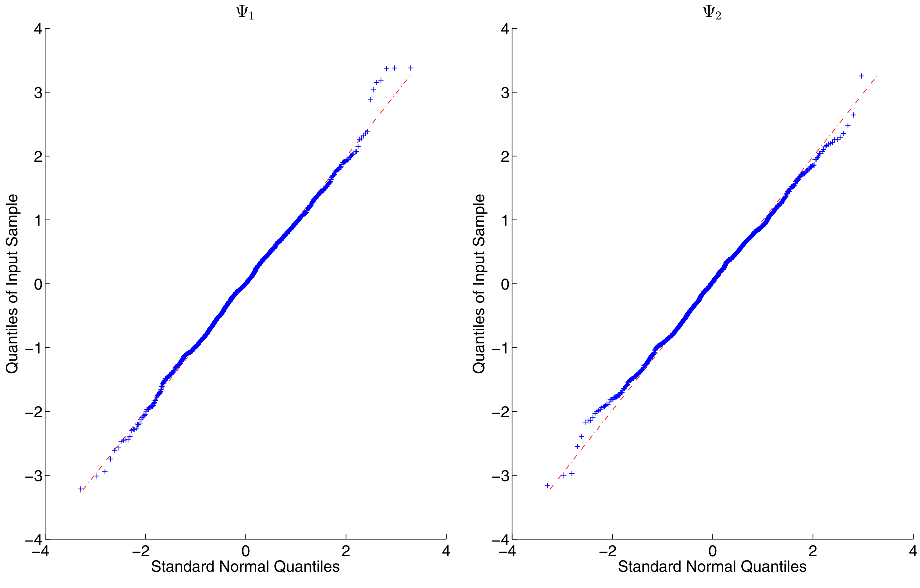

FIGURE 4.—Q–Q plots of the simulated components of Ψ for $n = 3000$. The figure compares the finite sample distributions of the two components of vector Ψ (right panel and left panel) with the standard normal distribution. We estimate the finite sample distributions with an unbalanced panel of $n = 3000$ individual stocks in the Monte Carlo exercise.

$\gamma \in [0, 1]$, and Lebesgue measurable w.r.t. $\gamma \in [0, 1]$ for a.e. $\omega \in \Omega$, such that $\int w(\gamma, \omega)\, d\gamma = 1$ for a.e. $\omega \in \Omega$. Quantities $w_t(\gamma, \omega) = w[\gamma, S^{t-1}(\omega)]$, for γ varying, yield the portfolio weights $w_t(\gamma_i)/n_t^w$ at time t, where $n_t^w = \sum_i w_t(\gamma_i)$ is the weighted number of the n sampled assets included in the portfolio w at time t. The excess return of the portfolio w is $R_t^w = \frac{1}{n_t^w} \sum_i w_t(\gamma_i) R_t(\gamma_i)$. From Equation (68), we have

$$R_t^w = \left(d_t^w\right)' (h_t + \mu_t) + \varepsilon_t^w, \tag{69}$$

with factor sensitivities $d_t^w = \frac{1}{n_t^w} \sum_i w_t(\gamma_i) d_t(\gamma_i)$ and an error term $\varepsilon_t^w = \frac{1}{n_t^w} \sum_i w_t(\gamma_i) \varepsilon_t(\gamma_i)$. We have that ε_t^w is close to zero for large n if the error terms of the individual assets feature weak cross-sectional dependence and the portfolio is sufficiently diversified. Thus, the $k \times 1$ vector f_t of excess returns from k diversified portfolios is close to $D_t(h_t + \mu_t)$, for some $k \times r$ matrix D_t which is measurable w.r.t. the information $\mathcal{F}_{t-1}$. To focus this section on specification analysis (see the online additional empirical results for discussion on missing factor impact), we assume $k = r$, namely, that the number of observable factors corresponds to the number of unknown factors, and we neglect approximation errors. Then, we have

$$h_t + \mu_t = D_t^{-1} f_t \tag{70}$$

for nonredundant observable factors. Replacing Equation (70) into model (68) shows that the asset returns satisfy model (1) with factors f_t and sensitivities $b_t(\gamma) = (D_t^{-1})' d_t(\gamma)$. By construction, we get $\nu_t = 0$ because the factors f_t are returns of tradable portfolios. Thus, model (1) is correctly specified as long as we set the correct number of factors, even if the observable factors f_t do not correspond to the unknown factors h_t. Indeed, the vector f_t dynamically spans the true factor space. However, a constrained parametric model for the economic factor sensitivities, instead of a generic unconstrained $d_t(\gamma)$, does not necessarily transmit to the observable factor sensitivities. For instance, if the economic factor sensitivities are linear functions of some instruments, the observable factor sensitivities are not necessarily linear functions of these instruments. Choosing mimicking portfolio returns as observable factors jointly with a constrained parameterization can lead to a first source of misspecification.

A second potential source of misspecification comes from the aggregation of assets into portfolios. Let w^j for $j = 1, \dots, m$ be a set of portfolios. We use the index j and the cardinality m for portfolios in order to distinguish them from the corresponding i and n for the fundamental assets. Under model (1) for the individual assets, the asset pricing restrictions yield the portfolio returns

$$R_t^j = a_t^j + \left(b_t^j\right)' f_t + \varepsilon_t^j, \tag{71}$$

with factor sensitivities

$$b_t^j = \frac{1}{n_t^j} \sum_i w_t^j(\gamma_i) b_t(\gamma_i), \tag{72}$$

intercepts $a_t^j = (b_t^j)' \nu_t$, and error terms $\varepsilon_t^j = \frac{1}{n_t^j} \sum_i w_t^j(\gamma_i) \varepsilon_t(\gamma_i)$. Model (71) is a factor model with the same factors as the original model for the individual assets, and time-varying alphas and betas. Hence, as observed in Section 2.2 for repackaging, we have robustness w.r.t. portfolio aggregation. However, if we choose a constrained parametric specification for the coefficients of a time-varying model, that parametric choice does not transmit easily under portfolio aggregation. First, the dynamics of the portfolio betas result from a combination of the dynamics of the individual stock betas and of the portfolio weights. Second, even with time-invariant portfolio weights, the aggregation of the asset-specific instruments is complex, and results in models with portfolio-specific instruments which involve unknown model parameters. For instance, let us consider the linear beta specification $b_{i,t} = B_i Z_{t-1} + C_i Z_{i,t-1}$ with a scalar stock-specific instrument estimated in our empirical analysis, and equally weighted portfolios, that is, $w_t^j = 1/|A^j|$ for $\gamma \in A^j$, and 0 otherwise, for all j and t, where $A^j \subset [0,1]$ is a measurable set with nonzero measure $|A^j|$. Then, from (72), the portfolio betas are $b_t^j = B^j Z_{t-1} + C^j Z_{t-1}^j$, where the portfolio coefficients $B^j = \frac{1}{n^j}\sum_{i:\gamma_i \in A^j} B_i$ and $C^j = \frac{1}{n^j}\sum_{i:\gamma_i \in A^j} C_i$ are averages of the individual coefficients, n_j is the number of indices i with $\gamma_i \in A^j$, and the portfolio-specific instrument $Z_{t-1}^j = \sum_{i:\gamma_i \in A^j} C_i Z_{i,t-1} / \sum_{i:\gamma_i \in A^j} C_i$ is a weighted average of the asset-specific instruments, with weights involving the unknown coefficients C_i. If we use an ad hoc aggregation scheme to define the portfolio-specific instruments, the resulting model is, in general, misspecified. If we try to replace the unknown C_i with estimates to get a proxy for the Z_{t-1}^j, we need first to estimate the model for the individual assets and face an EIV problem. For the FF portfolios, misspecification of the beta dynamics may result from the time-varying portfolio weights and the ad hoc aggregation scheme used to construct the portfolio-specific instrument, namely, the book-to-market equity of the portfolio as in Section 4.3 of the paper.

Under misspecification, the two-pass methodology may yield different pseudo-true values for the risk premia depending on the selected universe of assets. Let us assume that the DGP for the individual stock returns is given by model (1)–(3), with possibly time-varying betas and risk premia, but the researcher estimates a time-invariant model. For expository purposes, we focus on the OLS estimator in the second pass. We show in Section H.2 that the pseudo-true value of parameter ν using individual stock returns is $\nu^* = (\int b^*(\gamma) b^*(\gamma)' \, dG(\gamma))^{-1} \int b^*(\gamma) a^*(\gamma) \, dG(\gamma)$, where the pseudo-true values of sensitivities and intercepts are

$$
\begin{aligned}
b^*(\gamma) &= \big[I_K + V[f_t]^{-1} \operatorname{Cov}(f_t, \nu_t)\big] E\big[b_t(\gamma)\big] \\
&\quad + E\big[\xi_t\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big], \\
a^*(\gamma) &= E\big[\nu_t - \operatorname{Cov}(\nu_t, f_t) V[f_t]^{-1} f_t\big]' E\big[b_t(\gamma)\big] \\
&\quad - E\big[\eta_t'\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big],
\end{aligned}
$$

and the matrix and vector processes ξ_t and η_t are defined by $\xi_t = V[f_t]^{-1}(f_t - E[f_t])(\nu_t + f_t)'$ and $\eta_t = (E[f_t]' V[f_t]^{-1}(f_t - E[f_t]) - 1)(\nu_t + f_t)$. Expectations, variances, and covariances are w.r.t. the DGP. The pseudo-true value ν^* is equal to the unconditional expectation $E[\nu_t]$ if the individual betas are uncorrelated with the conditional expectations of f_t and ν_t given $\mathcal{F}_{t-1}$, and process ν_t is uncorrelated with f_t. Then the pseudo-true risk premia vector is $\lambda^* = \nu^* + E[f_t] = E[\lambda_t]$. Here, even if the model is misspecified, there is no effect on the time-averaged risk premia. However, in general, time-variation distorts risk premia estimates. Even if the factors f_t are tradable, that is, $\nu_t = 0$, we may have $\nu^* \neq 0$. The factors may appear as nontradable because of a misspecified time-invariant model as is likely in Section 4.2.

If we estimate the time-invariant model using the returns on m portfolios w^j, with $j = 1, \dots, m$, the pseudo-true value of ν becomes $\nu^{**} = (\sum_j b_j^* b_j^{*\prime})^{-1} \sum_j b_j^* a_j^*$, where (see Section H.3)

$$b_j^* = \int E\big[w_t^j(\gamma)\big] b^*(\gamma)\, dG(\gamma) + \int \operatorname{Cov}\big(\xi_t b_t(\gamma), w_t^j(\gamma)\big)\, dG(\gamma),$$

$$a_j^* = \int E\big[w_t^j(\gamma)\big] a^*(\gamma)\, dG(\gamma) - \int \operatorname{Cov}\big(\eta_t' b_t(\gamma), w_t^j(\gamma)\big)\, dG(\gamma).$$

The pseudo-true portfolio loadings b_j^* are the sum of two components. The first one is an aggregate of the pseudo-true individual loadings $b^*(\gamma)$ weighted by the time-averaged portfolio weights $E[w_t^j(\gamma)]$. The second component is induced by the time-variation of the portfolio weights and its interaction with f_t, ν_t, and factor sensitivities. A similar comment applies to the pseudo-true portfolio intercepts a_j^*. If the portfolio weights are time-invariant, building portfolios corresponds to aggregating the individual pseudo-true alphas and betas. The portfolio aggregation effect is more complex if portfolio weights are time-varying. In general, the pseudo-true value ν^{**} depends on the number m of chosen portfolios and the weights $w_t^j(\gamma)$ they are built on, and we expect the pseudo-true values ν^{**} and ν^* not to be equal, as the different estimated $\hat{\nu}$ in Table I, Panel B, may indicate. Besides, even if we observe that the portfolio betas are more stable over time, this does not imply that ν^{**} will be closer to zero than ν^*, when $\nu_t = 0$. We give a simple estimation exercise (see Section H.4) to check whether the numerical values for these pseudo-true values and their differences are compatible with the order of magnitude observed in Table I, Panel B, including values close to zero in some cases. For the value factor, time-variation in the portfolio weights can explain the large discrepancy between the pseudo-true values computed on the 25 FF portfolios and the individual stocks.

The above discussion concentrates on the impact of misspecification when the econometrician estimates a time-invariant model. Similar computations and remarks apply for estimation of misspecified time-varying models.

H.2. *Pseudo-True Value Using Individual Assets*

The pseudo-true values of the regression coefficients are $\beta^*(\gamma) = (a^*(\gamma), b(\gamma)^{*\prime})' = Q_x^{-1} E[x_t R_t(\gamma)]$, for all $\gamma \in [0, 1]$, where the expectation is w.r.t. the DGP. Let $\beta_i^* = \beta^*(\gamma_i)$. If the OLS estimator is used in the second pass, and matrix $E[b_i^* b_i^{*\prime}]$ is positive definite, the pseudo-true value of parameter ν is $\nu_1^* = E[b_i^* b_i^{*\prime}]^{-1} E[b_i^* a_i^*]$. The pseudo-true weights are $w_i^* = (v_i^*)^{-1}$ with $v_i^* = \tau_i c_{\nu_1^*}' Q_x^{-1} S_{ii}^* Q_x^{-1} c_{\nu_1^*}$, where $S_{ii}^* = E[(\varepsilon_{i,t}^*)^2 x_t x_t | \gamma_i]$ and $\varepsilon_{i,t}^* = R_{i,t} - x_t' \beta_i^*$. If the

WLS estimator is used in the second pass, and matrix $E[w_i^* b_i^* b_i^{*\prime}]$ is positive definite, the pseudo-true value of parameter ν is

$$\nu^* = E\big[w_i^* b_i^* b_i^{*\prime}\big]^{-1} E\big[w_i^* b_i^* a_i^*\big]. \tag{73}$$

Then, the pseudo-true value of the risk premia vector is $\lambda^* = \nu^* + E[f_t]$.

Let $\hat{\nu}$ be the estimator defined in Equation (14) of the paper, using the first-pass estimators $\hat{\beta}_i$ and the weights $\hat{w}_i$ for the second pass. The next lemma states that the estimators converge to the corresponding pseudo-true values and is proved at the end of this subsection.

LEMMA 17: *Suppose Assumptions* A.1(b), SC.1–SC.2, B.1, B.4, B.5 *hold. Moreover, let* $\sup_{\gamma\in[0,1]} P[\|\frac{1}{T}\sum_t I_t(\gamma) x_t \varepsilon_t^*(\gamma)\| \geq \delta]$ *satisfy the large deviation bound in Assumption* B.1, *for any* $\delta > 0$ *and* $T \in \mathbb{N}$, *where* $\varepsilon_t^*(\gamma) = R_t(\gamma) - x_t'\beta^*(\gamma)$ *is the pseudo-true error. Then, as* $n, T \to \infty$ *such that* $n = O(T^{\bar{\gamma}})$ *for* $\bar{\gamma} > 0$, *we have*: (i) $\sup_i \mathbf{1}_i^\chi \|\hat{\beta}_i - \beta_i^*\| = o_p(1)$; (ii) $\frac{1}{n}\sum_i \|\hat{w}_i - w_i^*\| = o_p(1)$; (iii) $\hat{\nu} = \nu^* + o_p(1)$.

Let us now derive more explicit expressions for the components $a^*(\gamma)$ and $b^*(\gamma)$ of the pseudo-true coefficients vector. We have

$$b^*(\gamma) = V[f_t]^{-1}\operatorname{Cov}\big(f_t, R_t(\gamma)\big), \quad a^*(\gamma) = E\big[R_t(\gamma)\big] - E[f_t]' b^*(\gamma), \tag{74}$$

for all $\gamma \in [0, 1]$. From $R_t(\gamma) = (f_t + \nu_t)' b_t(\gamma) + \varepsilon_t(\gamma)$, we have:

$$\begin{aligned} E\big[R_t(\gamma)\big] &= E\big[(f_t + \nu_t)' b_t(\gamma)\big] \\ &= E[\nu_t]' E\big[b_t(\gamma)\big] + E[f_t]' E\big[b_t(\gamma)\big] \\ &\quad + E\big[(f_t + \nu_t)'\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big], \end{aligned}$$

and

$$\begin{aligned} \operatorname{Cov}\big(f_t, R_t(\gamma)\big) &= \operatorname{Cov}\big(f_t, (f_t + \nu_t)' b_t(\gamma)\big) \\ &= \big(V[f_t] + \operatorname{Cov}(f_t, \nu_t)\big) E\big[b_t(\gamma)\big] \\ &\quad + \operatorname{Cov}\big(f_t, (f_t + \nu_t)'\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big) \\ &= \big(V[f_t] + \operatorname{Cov}(f_t, \nu_t)\big) E\big[b_t(\gamma)\big] \\ &\quad + E\big[\big(f_t - E[f_t]\big)(f_t + \nu_t)'\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big]. \end{aligned}$$

Then, by replacing into (74) and rearranging terms, we get

$$\begin{aligned} b^*(\gamma) &= \big[I_K + V[f_t]^{-1}\operatorname{Cov}(f_t, \nu_t)\big] E\big[b_t(\gamma)\big] \\ &\quad + E\big[\xi_t\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big], \end{aligned} \tag{75}$$

$$(76)\qquad a^*(\gamma) = E\big[\nu_t - \mathrm{Cov}(\nu_t, f_t)V[f_t]^{-1}f_t\big]'E\big[b_t(\gamma)\big] - E\big[\eta_t'\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\big],$$

for all $\gamma \in [0, 1]$, where $\xi_t = V[f_t]^{-1}(f_t - E[f_t])(\nu_t + f_t)'$ and $\eta_t = (E[f_t]' \times V[f_t]^{-1}(f_t - E[f_t]) - 1)(\nu_t + f_t)$.

PROOF OF LEMMA 17: We have $\hat{\beta}_i - \beta_i^* = \tau_{i,T}\hat{Q}_{x,i}^{-1}\frac{1}{T}\sum_t I_{i,t}x_t\varepsilon_{i,t}^*$. Then part (i) follows by similar arguments as in the proof of Lemma 3(i) for a well-specified time-invariant model. The proof of part (ii) is similar to the proof of Lemma 3(iii) and is omitted. Finally, using parts (i)–(ii) of this lemma, Assumption SC.2, and the LLN, we have

$$\frac{1}{n}\sum_i \hat{w}_i\hat{b}_i\hat{b}_i' = \frac{1}{n}\sum_i w_i^*b_i^*b_i^{*\prime} + o_p(1) = E\big[w_i^*b_i^*b_i^{*\prime}\big] + o_p(1),$$

and

$$\frac{1}{n}\sum_i \hat{w}_i\hat{b}_i\hat{a}_i = \frac{1}{n}\sum_i w_i^*b_i^*a_i^* + o_p(1) = E\big[w_i^*b_i^*a_i^*\big] + o_p(1).$$

Since matrix $E[w_i^*b_i^*b_i^{*\prime}]$ is invertible, part (iii) follows. *Q.E.D.*

H.3. *Pseudo-True Value Using Portfolios*

Let us now assume that we estimate the time-invariant model on a set of m portfolios w^j, with $j = 1, \ldots, m$. If the portfolios are well diversified, and the number of underlying assets n tends to infinity, the idiosyncratic error terms ε_t^j vanish in Equation (71). Then, the portfolio returns are $R_t^j = (b_t^j)'(f_t + \nu_t)$, where the portfolio sensitivities are

$$(77)\qquad b_t^j = \int w_t^j(\gamma)b_t(\gamma)\,dG(\gamma).$$

Then, the pseudo-true values of the regression coefficients are obtained along the lines of Section H.2 replacing $R_t(\gamma)$ with R_t^j, and $b_t(\gamma)$ with b_t^j. We get $\beta^{*j} = (a^{*j}, (b^{*j})')'$, where

$$(78)\qquad b^{*j} = \big[I_K + V[f_t]^{-1}\mathrm{Cov}(f_t, \nu_t)\big]E\big[b_t^j\big] + E\big[\xi_t\big(b_t^j - E\big[b_t^j\big]\big)\big],$$

$$(79)\qquad a^{*j} = E\big[\nu_t - \mathrm{Cov}(\nu_t, f_t)V[f_t]^{-1}f_t\big]'E\big[b_t^j\big] - E\big[\eta_t'\big(b_t^j - E\big[b_t^j\big]\big)\big],$$

for all $j = 1, \ldots, m$. Then, when the OLS estimator is used in the second pass, the pseudo-true value of parameter ν is $\nu_1^* = (\sum_j b^{*j}(b^{*j})')^{-1}\sum_j b^{*j}a^{*j}$. When the WLS estimator is used, the pseudo-true value of parameter ν is $\nu^* =$

$(\sum_j (v^{*j})^{-1} b^{*j} (b^{*j})')^{-1} \sum_j (v^{*j})^{-1} b^{*j} a^{*j}$, where the reciprocal of the pseudo-true weights are $v^{*j} = c'_{\nu_1^*} Q_x^{-1} S^{*j} Q_x^{-1} c_{\nu_1^*}$, with $S^{*j} = E[(\varepsilon_t^{*j})^2 x_t x_t]$ and $\varepsilon_t^{*j} = R_t^j - x_t' \beta^{*j}$.

Let us now derive the expressions of the pseudo-true regression coefficients given in Section H.1. From (77), we have

$$E\big[b_t^j\big] = \int E\big[w_t^j(\gamma)\big] E\big[b_t(\gamma)\big]\, dG(\gamma) + \int \operatorname{Cov}\big(b_t(\gamma), w_t^j(\gamma)\big)\, dG(\gamma),$$

$$\begin{aligned} b_t^j - E\big[b_t^j\big] &= \int E\big[w_t^j(\gamma)\big]\big(b_t(\gamma) - E\big[b_t(\gamma)\big]\big)\, dG(\gamma) \\ &\quad + \int \big(w_t^j(\gamma) - E\big[w_t^j(\gamma)\big]\big) b_t(\gamma)\, dG(\gamma) \\ &\quad - \int \operatorname{Cov}\big(b_t(\gamma), w_t^j(\gamma)\big)\, dG(\gamma). \end{aligned}$$

By replacing into (78), we get

$$\begin{aligned} b^{*j} = &\int E\big[w_t^j(\gamma)\big] b^*(\gamma)\, dG(\gamma) \\ &+ \big[I_K + V[f_t]^{-1} \operatorname{Cov}(f_t, \nu_t) - E[\xi_t]\big] \\ &\times \int \operatorname{Cov}\big(b_t(\gamma), w_t^j(\gamma)\big)\, dG(\gamma) \\ &+ \int \operatorname{Cov}\big(\xi_t b_t(\gamma), w_t^j(\gamma)\big)\, dG(\gamma). \end{aligned}$$

Since $E[\xi_t] = I_K + V[f_t]^{-1} \operatorname{Cov}(f_t, \nu_t)$, the second term in the RHS vanishes, and we get the expression of b^{*j} given in Section H.1. The proof of the expression of a^{*j} is similar, by using $E[\eta_t] = -E[\nu_t - \operatorname{Cov}(\nu_t, f_t) V[f_t]^{-1} f_t]$.

H.4. *Empirical Pseudo-True Values*

In Table XIV, we report the estimates of the pseudo-true values of parameter vector ν in a time-invariant four-factor model obtained with the individual stocks, the 25 FF portfolios, and the 44 Indu. portfolios. We get the estimates by replacing the expectations in Equations (73), (75)–(76), and (78)–(79) with sample averages. To assess the contributions of misspecifications along different directions, we consider several alternative assumptions on the DGP for process ν_t and factor sensitivities $b_t(\gamma)$ of the individual stocks. Specifically, we assume that the vector ν_t is either (i) time-invariant and equal to zero, or (ii) time-invariant and equal to the time-average $\bar{\nu} = [1.3772, -0.2122, -6.1630, -2.5507]'$ of the estimates $\hat{\nu}_t$ obtained with the

TABLE XIV

ESTIMATED PSEUDO-TRUE VALUES OF PARAMETER ν FOR THE FOUR-FACTOR MODEL[a]

			$n = 25$		$n = 44$	
		$n = 9936$	CW	TVW	CW	TVW
$\nu_t = 0$, $b_{i,t}$ constant	ν_{m^*}	0.0000	0.0000	−0.3427	0.0000	−0.0801
	ν_{smb^*}	0.0000	0.0000	0.6167	0.0000	0.1843
	ν_{hml^*}	0.0000	0.0000	1.1304	0.0000	−0.4866
	ν_{mom^*}	0.0000	0.0000	0.8850	0.0000	−2.3739
$\nu_t = 0$, $b_{i,t}$ time-varying	ν_{m^*}	−0.0251	1.5815	−0.0349	0.5738	−0.3040
	ν_{smb^*}	0.6486	0.7998	0.8877	0.6075	1.2729
	ν_{hml^*}	−1.1835	−4.9452	0.6012	−0.6365	−0.8209
	ν_{mom^*}	−4.5639	−1.0871	−1.4821	−3.3692	−5.5221
$\nu_t = \bar{\nu}$, $b_{i,t}$ constant	ν_{m^*}	1.3772	1.3772	0.4453	1.3772	1.0312
	ν_{smb^*}	−0.2122	−0.2122	0.4779	−0.2122	0.0657
	ν_{hml^*}	−6.1636	−6.1636	−3.0085	−6.1636	−5.8395
	ν_{mom^*}	−2.5507	−2.5507	−0.7216	−2.5507	−4.5657
$\nu_t = \bar{\nu}$, $b_{i,t}$ time-varying	ν_{m^*}	1.3406	2.6374	0.6123	1.6079	0.9199
	ν_{smb^*}	0.1490	0.1940	0.7492	0.1824	0.8432
	ν_{hml^*}	−6.5468	−9.8461	−3.4016	−6.1935	−6.4573
	ν_{mom^*}	−6.6899	−3.5831	−2.6132	−5.4675	−8.0675
$\nu_t = \hat{\nu}_t$, $b_{i,t}$ constant	ν_{m^*}	1.3788	1.3788	0.8521	1.3788	1.0816
	ν_{smb^*}	−0.2158	−0.2158	0.4970	−0.2158	0.1172
	ν_{hml^*}	−6.1291	−6.1291	−3.9565	−6.1291	−5.9395
	ν_{mom^*}	−2.4741	−2.4741	−0.9824	−2.4741	−4.2506
$\nu_t = \hat{\nu}_t$, $b_{i,t}$ time-varying	ν_{m^*}	1.0201	1.5269	−0.0080	1.4433	0.6526
	ν_{smb^*}	0.1678	0.1870	0.8511	−0.3721	0.6996
	ν_{hml^*}	−6.0848	−8.1776	−2.6871	−6.6668	−6.5043
	ν_{mom^*}	−4.8815	−3.9304	−1.6555	−6.0449	−7.4999

[a]The table contains the annualized estimates of the pseudo-true values of parameter ν for the market (ν_m^*), size (ν_{smb}^*), book-to-market (ν_{hml}^*), and momentum (ν_{mom}^*) factors. We report the estimates ν^* for individual stocks ($n = 9936$, $n^\chi = 3900$), the 25 FF, and 44 Indu. portfolios as base assets for several DGPs. For portfolios, we report both the estimates with time-varying portfolio weights (TVW) and the estimates obtained assuming time-constant weights (CW).

time-varying model applied on individual stocks in Section 4.3, or (iii) time-varying and equal to the estimates $\hat{\nu}_t$. Furthermore, we assume that the betas of the $n^\chi = 3900$ individual stocks after trimming are either (a) time-invariant and equal to the time averages of the estimates $\hat{b}_{i,t}$ obtained with the time-varying model in Section 4.3, or (b) time-varying and equal to the estimates $\hat{b}_{i,t}$. The combination of (i)–(iii) and (a)–(b) yields six alternative (empirical) DGPs. We compute the portfolio betas by aggregating the betas of the 3900 individual stocks using weights $\hat{w}_{i,t}^j$. These weights are obtained by following the methodology underlying the FF and Indu. portfolios applied to the 3900 assets of our trimmed sample. To assess the contribution of time-varying portfolio weights, we also compute the pseudo-true values using the returns of 25 and 44 portfolios with time-invariant weights equal to the time-averages of the corresponding weights $\hat{w}_{i,t}^{j}$. Thus, the pseudo-true values are computed for five different universes of assets.

For the DGPs with time constant $b_{i,t}$ and ν_t, the time-invariant model is correctly specified on individual stocks. This explains why the (pseudo-)true values of ν with individual stocks, and with time-constant portfolio weights, coincide in the first and third subpanels. Moreover, Equations (75)–(76) and (78)–(79) imply that these pseudo-true values of ν coincide also when ν_t is time-varying but the individual stock betas are constant, as observed in the fifth subpanel. Instead, the pseudo-true values with time-varying portfolio weights differ from the pseudo-true values with individual stocks for all DGPs. The largest differences across universes of assets are observed for the value and momentum factors. We get a substantial difference between $\nu_{hml}^{*} = -6.1636$ on the individual stocks and $\nu_{hml}^{**} = -3.0085$ on the 25 FF portfolios (with time-varying weights) already for the DGP with constant $\nu_t = \bar{\nu}$ and constant $b_{i,t}$. The five pseudo-true values for ν_{hml} do not change a lot when we move to DGPs with time-variation in ν_t and/or $b_{i,t}$. Moreover, the estimates of ν_{hml} on the 25 FF portfolios with time-varying weights are asymptotically larger than the estimates with constant weights. These findings suggest that, for the value factor, the difference between the results with the individual stocks and the FF portfolios is due mainly to time-variation in the portfolio weights. For the momentum factor, the largest discrepancies between individual stocks and FF portfolios are observed for the DGPs with time-varying betas and weights. The pseudo-true values for the 44 Indu. portfolios are more similar to the pseudo-true values for individual stocks.

REFERENCES

ANG, A., J. LIU, AND K. SCHWARZ (2008): "Using Individual Stocks or Portfolios in Tests of Factor Models," Working Paper. [25]

BICKEL, P. J., AND E. LEVINA (2008): "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36 (6), 2577–2604. [38]

BILLINGSLEY, P. (1995): *Probability and Measure*. New York: Wiley. [21]

BOSQ, D. (1998): *Nonparametric Statistics for Stochastic Processes*. New York: Springer-Verlag. [25,26,31]

CHAMBERLAIN, G. (1983): "Funds, Factors, and Diversification in Arbitrage Pricing Models," *Econometrica*, 51 (5), 1305–1323. [21]

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51 (5), 1281–1304. [1,21]

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Advanced Texts in Econometrics. Oxford: Oxford University Press. [25,31]

FERSON, W. E., AND C. R. HARVEY (1999): "Conditioning Variables and the Cross Section of Stock Returns," *Journal of Finance*, 54 (4), 1325–1360. [37]

HALMOS, P. (1950): *Measure Theory*. Graduate Texts in Mathematics. New York: Springer. [23]

HANSEN, L. P., AND S. F. RICHARD (1987): "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55 (3), 587–613. [21-23]

MACKINLAY, C. (1995): "Multifactor Models Do not Explain Deviations From the CAPM," *Journal of Financial Economics*, 38 (1), 3–28. [38]
STOUT, W. (1974): *Almost Sure Convergence*. New York: Academic Press. [25]
WHITE, H. (2001): *Asymptotic Theory for Econometricians*. San Diego: Academic Press. [25,27]

Faculty of Economics, Università della Svizzera Italiana (USI), Via Buffi 13, CH-6900 Lugano, Switzerland and Swiss Finance Institute; patrick.gagliardini@usi.ch,

Faculty of Economics, Università della Svizzera Italiana (USI), Via Buffi 13, CH-6900 Lugano, Switzerland; elisa.ossola@usi.ch,

and

Geneva School of Economics and Management, University of Geneva, Bd du Pont d'Arve 40, CH-1211 Geneve 4, Switzerland and Swiss Finance Institute; olivier.scaillet@unige.ch.

Co-editor Lars Hansen handled this manuscript.

Manuscript received September, 2012; final revision received October, 2015.