Supplement to "Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects"

(*Quantitative Economics*, Vol. 11, No. 2, May 2020, 579–608)

Jungmo Yoon
College of Economics and Finance, Hanyang University

Antonio F. Galvao
Department of Economics, University of Arizona

This Supplemental Appendix is structured as follows. Section S.1 contains Monte Carlo simulation results. Some tables appear in Section S.1.2. Section S.2 collects proofs.

S.1. Finite sample performance

We examine empirical coverage ratios of confidence intervals and sizes and powers of the Score and Wald tests. The main findings can be summarized as follows. (1) If there is no serial correlation correction, confidence intervals will suffer from severe undercoverages and the size of tests will be seriously inflated. (2) The CCM estimator delivers correct coverage ratios and sizes under widely different circumstances. (3) The bias-correction in autocovariance estimators is important to achieve good finite-sample performance.

S.1.1 *Simulation setup*

The simulation setup closely follows the real data example in Section 6. We use four models to study a wide spectrum of issues. This section discusses Models 1 and 2, defined below. Section S.3 of Yoon and Galvao (2019) examines Models 3 and 4. The first model is a location-shift model (a regression with *iid* errors)

$$\text{Model 1}: \quad y_{it} = \alpha_i + w_{it}\beta + z_{it}\gamma + e_{it},$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$ and the τth quantile of e_{it} is zero. We use Model 1 to study three issues. The first is to see the effect of bias-correction in autocovariance matrix estimation, by comparing robust confidence intervals and tests using $\widehat{J}_T$ and $\check{J}_T$. The second is to compare the Score and Wald tests. For Model 1, both are valid but the

Jungmo Yoon: jmyoon@hanyang.ac.kr
Antonio F. Galvao: agalvao@email.arizona.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE802

Score test is simpler to implement because it does not require conditional density estimation. The third objective is to compare the performance of two point estimators; the analytical bias corrected estimator $\widehat{\boldsymbol{\theta}}_{\text{BC}}$ and the jackknife estimator $\widehat{\boldsymbol{\theta}}_{\text{JK}}$ under various (N, T) and at different quantile levels.

Other specifics of the model are as follows. The true values of the slope parameters are $(\beta, \gamma) = (1.0, 0.0)$, so the effect of z_{it} on the τth conditional quantile of y_{it} is zero. To generate z_{it}, we draw one year out of T years at random and one-half out of N subjects at random. For the chosen subjects after the selected year, $z_{it} = 1$, otherwise, $z_{it} = 0$. This z_{it} can be viewed as a randomly generated fictitious law. We draw this random year and subjects for each simulation replication. Since the law is independent of the outcome of interest, any 5% significance tests should reject the null hypothesis of no effect 5 times out of 100 trials. This simulation design is taken from Bertrand, Duflo, and Mullainathan (2004). To make sure that we have enough observations before and after the selected year, we select a random year from the middle 80% of the total T years. Other variables are generated by

$$\alpha_i \overset{iid}{\sim} \text{Uniform}[-1, 1], \tag{S.1}$$

$$w_{it} = 0.3\alpha_i + \varepsilon_{it}, \qquad \varepsilon_{it} = 0.8\varepsilon_{i\,t-1} + \nu_{it}, \qquad \nu_{it} \overset{iid}{\sim} N\big(0, \big(1 - 0.8^2\big)\big), \tag{S.2}$$

$$e_{it} = \epsilon_{it} - Q_\epsilon(\tau), \qquad \epsilon_{it} = \rho\epsilon_{i\,t-1} + v_{it}, \qquad v_{it} \overset{iid}{\sim} N\big(0, 1 - \rho^2\big). \tag{S.3}$$

In (S.1), FE are generated from a uniform distribution. In (S.2), the individual effects and the included regressor are correlated by design. The control variable w_{it} is strongly positively correlated, reflecting the real data example. In (S.3), the error term is generated under varying degrees of serial correlation where ρ takes value in $\{0.0, 0.4, 0.9\}$. When $\rho = 0.0$, there is no temporal dependence and both cluster robust and nonrobust tests should have the correct size. When $\rho = 0.9$, there is a strongly positive serial correlation. In (S.3), $Q_\epsilon(\tau)$ denote the τth quintile of ϵ_{it}. The adjustment by $Q_\epsilon(\tau)$ makes the τth quantile of the error term equal to zero.

We mainly consider four combinations of N and T. The main specification is $(N, T) = (50, 50)$, the length of time-series is comparable to the number of subjects. This is common in many state-level panel data. To see any effects from a short panel, we consider $(N, T) = (50, 20)$. This represent a case where one has a relatively shorter time periods compared to the number of cross-sectional units. The third case is an example of a long panel, $(N, T) = (20, 100)$. By comparing the second and third cases, one can examine the effect of the bias-correction in short versus long panel. To see any effects of a large number of subjects, we consider $(N, T) = (100, 20)$. This is the most challenging case because it is a short panel with many nuisance parameters to estimate. The latter can make the incidental parameter problem more serious. Yoon and Galvao (2019) also considered $(N, T) = (400, 30)$ and examined how the CCM estimator performed in such case.

The second model is a location-scale model with the multiplicative errors:

$$\text{Model 2}: \quad y_{it} = \alpha_i + w_{it}\beta + z_{it}\gamma + (1 + 0.75w_{it})e_{it}.$$

The only difference between Models 1 and 2 is that the latter has heteroskedasticity. Other variables are generated in the same way. The autocorrelation structure remains the same over $1 \le i \le N$. We use Model 2 to study three issues. First, we compare the general CCM estimator $\check{J}_T$ to the simplified estimator $\widehat{J}_T^s$. In this multiplicative heteroskedastic model, both are valid, but the latter is simpler to implement. Second, the Score test is invalid and it may affect its performance. Third, we continue to examine the performance of two regression coefficient estimators.

S.1.2 *Results*

This section concentrates on results using $\widehat{\boldsymbol{\theta}}_{\mathrm{BC}}$. The result using $\widehat{\boldsymbol{\theta}}_{\mathrm{JK}}$ can be found in Section S.3 in Yoon and Galvao (2019). First, let us examine empirical coverage ratios of the 95% confidence intervals. Table S.1 shows the results for Model 1. It reports frac-

TABLE S.1. Coverage ratio, Model 1.

	$\tau=0.5$				$\tau=0.75$			
Cases	CI_0	CI_1	CI_2	CI_2^s	CI_0	CI_1	CI_2	CI_2^s
$(N,T)=(50,50)$								
$\rho=0.0$	0.952	0.938	0.953	0.952	0.939	0.926	0.940	0.940
$\rho=0.4$	0.876	0.928	0.958	0.958	0.873	0.912	0.946	0.948
$\rho=0.9$	0.535	0.896	0.949	0.942	0.500	0.880	0.948	0.944
$(N,T)=(50,20)$								
$\rho=0.0$	0.951	0.928	0.953	0.954	0.923	0.892	0.930	0.932
$\rho=0.4$	0.875	0.908	0.958	0.951	0.856	0.894	0.942	0.942
$\rho=0.9$	0.712	0.928	0.944	0.942	0.651	0.896	0.945	0.938
$(N,T)=(20,100)$								
$\rho=0.0$	0.960	0.954	0.960	0.960	0.949	0.936	0.943	0.948
$\rho=0.4$	0.872	0.937	0.954	0.952	0.858	0.919	0.940	0.948
$\rho=0.9$	0.469	0.902	0.956	0.952	0.478	0.893	0.943	0.944
$(N,T)=(100,20)$								
$\rho=0.0$	0.952	0.920	0.958	0.954	0.928	0.896	0.936	0.935
$\rho=0.4$	0.876	0.906	0.958	0.953	0.854	0.894	0.946	0.944
$\rho=0.9$	0.681	0.931	0.939	0.935	0.630	0.886	0.914	0.918

Note: Empirical coverage ratios of the 95% confidence intervals, based on 2000 simulation replications. Model 1. We use the analytical bias correction estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{BC}}(\tau)$ to estimate regression coefficients. Table shows four types of confidence intervals corresponding to different ways of calculating standard errors; (i) CI_0: the conventional standard errors in QR assuming independent errors, (ii) CI_1: cluster robust standard errors without bias-correction using $\widehat{J}_T$, (iii) CI_2: cluster robust standard errors with bias-correction using $\check{J}_T$, (iv) CI_2^s cluster robust standard errors with bias-correction using the simplified CCM estimator $\widehat{J}_T^s$.

tions of simulation runs out of 2000 replications in which confidence intervals include the true value of the parameter $\gamma(\tau) = 0$. Four confidence interval estimators are considered. CI_0 is the conventional confidence interval assuming independent errors. It is not robust to clustering. CI_1 uses $\widehat{J}_T$ to obtain the standard errors. It is robust to clustering but does not correct the bias in autocovariance estimates. CI_2 uses $\check{J}_T$ which is a cluster robust confidence interval applying bias-correction in autocovariance estimates. Finally, CI_2^s is an interval obtained from the simplified CCM estimator $\widehat{J}_T^s$. It includes bias-correction by design since it uses $\widehat{\varrho}^{(\infty)}$. We present results for $\tau = 0.5$ and 0.75.

The nonrobust confidence interval CI_0 has correct coverages when $\rho = 0$. So it is valid when errors are serially uncorrelated. But when $\rho > 0$, the coverage ratio is down to as low as 0.535 when $(N, T) = (50, 50)$. The size of undercoverage depends on the length of panel. When $T = 100$, the coverage ratio becomes as low as 0.469, about a half of the target rate 0.95. The conventional standard error and the resulting confidence interval are clearly inadequate.

The cluster robust standard errors are effective to fix this undercoverage problem. Examine CI_1. When $(N, T) = (50, 50)$ and $\rho = 0.9$, the coverage ratio improves from 0.535 to 0.896. In fact, it significantly improves the coverages ratios in all cases. But coverage ratios are still lower than the target level 0.95, often significantly so. The remaining gap is closed when we apply the bias-correction in autocovariance estimates. The interval estimator CI_2, compared to CI_1, improves the ratio to 0.949. The size of improvement due to bias-correction tends to be greater when T is larger and when $\tau = 0.75$. In fact, CI_2 has correct empirical sizes in almost all cases. The only exception may be when $(N, T) = (100, 20)$ at $\tau = 0.75$. It is an example of a relatively short panel with many nuisance parameters, representing a very challenging case. Between CI_2 and CI_2^s, the latter is comparable to the former. The simplified CCM estimator has equally good performances in every case we consider.

Table S.2 considers Model 2, a location-scale model. We continue to see undercoverages when using CI_0. The coverage of this nonrobust estimator is correct when $\rho = 0$ but it become as low as 0.473 when $(N, T) = (20, 100)$. For cluster robust confidence intervals, we see the same pattern as in Table S.1. CI_1 fix the undercoverages in CI_0 significantly, but it does not reach to the target 0.95. This gap is closed by the bias-correction. Model 2 has multiplicative heteroskedastic errors but identical autocorrelation structure, so both CI_2 and CI_2^s are all valid. Both prove to be effective to deliver correct coverage ratios.

Consider the Score and Wald tests. Tables S.3 and S.4 report the empirical sizes of tests. The nominal rate is 0.05. The definitions of three Wald tests, W_0, W_1, and W_2 remain the same as in Section 6. The same naming convention applies to the Score tests. The only difference is that the Score test uses the simplified CCM estimator.

Overall, we see similar patterns as before. The nonrobust tests have much inflated sizes. In Table S.3, the empirical size of the 5% tests can be as high as 0.566 in Model 1. The nonrobust tests are clearly inadequate. The cluster robust tests successfully cor-

TABLE S.2. Coverage ratio, Model 2.

	$\tau = 0.5$				$\tau = 0.75$			
Cases	CI_0	CI_1	CI_2	CI_2^s	CI_0	CI_1	CI_2	CI_2^s
$(N, T) = (50, 50)$								
$\rho = 0.0$	0.953	0.942	0.956	0.951	0.942	0.929	0.942	0.946
$\rho = 0.4$	0.882	0.934	0.963	0.963	0.880	0.920	0.954	0.958
$\rho = 0.9$	0.550	0.895	0.948	0.939	0.518	0.879	0.948	0.941
$(N, T) = (50, 20)$								
$\rho = 0.0$	0.966	0.949	0.966	0.967	0.942	0.922	0.956	0.959
$\rho = 0.4$	0.895	0.921	0.962	0.956	0.873	0.910	0.950	0.951
$\rho = 0.9$	0.738	0.930	0.946	0.954	0.674	0.902	0.942	0.940
$(N, T) = (20, 100)$								
$\rho = 0.0$	0.962	0.958	0.962	0.961	0.952	0.941	0.948	0.953
$\rho = 0.4$	0.880	0.938	0.956	0.956	0.868	0.926	0.944	0.952
$\rho = 0.9$	0.473	0.910	0.955	0.954	0.482	0.896	0.944	0.946
$(N, T) = (100, 20)$								
$\rho = 0.0$	0.960	0.936	0.967	0.961	0.921	0.905	0.941	0.942
$\rho = 0.4$	0.900	0.933	0.966	0.963	0.870	0.914	0.956	0.954
$\rho = 0.9$	0.690	0.934	0.944	0.940	0.640	0.888	0.926	0.924

Note: Empirical coverage ratios of the 95% confidence intervals, based on 2000 simulation replications. Model 2. See descriptions in Table S.1 for details.

rect the size problem. And the bias correction matters for good performances. Consider Model 1 when $(N, T) = (50, 50)$ and $\rho = 0.9$ and $\tau = 0.75$. The W_1 improves the empirical size from 0.500 to 0.120, then W_2 improves it further to 0.052. As a matter of fact, the size of cluster robust tests, H_2 and W_2, stay close to the nominal rate, across variations over (N, T) combinations, values of ρ, different quantile levels. The robust tests have excellent finite-sample performance.

Does this improved size hurt the power of tests? Tables S.5 and S.6 show empirical powers of the bias-corrected robust tests, S_2 and W_2. The data is generated under the alternative hypothesis in which $\gamma(\tau) \neq 0$. We set $\gamma(\tau) = c_h \cdot s_Y$ where c_h is a constant and s_Y denote the standard deviation of the dependent variable when data is generated under the null. So $c_h = 0.2$ means that the size of effect under the alternative is equal to 20% of the standard deviation of the dependent variable under the null. The power proves to be excellent across variations over (N, T) combinations, values of ρ, different quantile levels.

Between Score and Wald tests, we observe that the Score test S_2 appears to be robust to the forms of heteroskedasticity we consider in this section. It delivers good sizes in Model 2. Also the Score test is more optimistic than the Wald test because the size of S_2 tends to be bigger than the size of W_2. Overall, simulation studies show that cluster robust standard errors and tests have excellent finite-sample properties.

TABLE S.3. Size of tests, Model 1.

Cases	$\tau = 0.5$						$\tau = 0.75$					
	H_0	H_1	H_2	W_0	W_1	W_2	H_0	H_1	H_2	W_0	W_1	W_2
$(N, T) = (50, 50)$												
$\rho = 0.0$	0.055	0.067	0.058	0.048	0.062	0.047	0.050	0.054	0.048	0.061	0.074	0.060
$\rho = 0.4$	0.145	0.090	0.052	0.124	0.072	0.042	0.122	0.090	0.050	0.127	0.088	0.054
$\rho = 0.9$	0.498	0.111	0.071	0.465	0.104	0.051	0.449	0.126	0.078	0.500	0.120	0.052
$(N, T) = (50, 20)$												
$\rho = 0.0$	0.047	0.066	0.050	0.049	0.072	0.047	0.046	0.066	0.048	0.077	0.108	0.070
$\rho = 0.4$	0.144	0.111	0.067	0.125	0.092	0.042	0.109	0.090	0.054	0.144	0.106	0.058
$\rho = 0.9$	0.336	0.064	0.064	0.288	0.072	0.056	0.276	0.082	0.072	0.349	0.104	0.055
$(N, T) = (20, 100)$												
$\rho = 0.0$	0.047	0.052	0.048	0.040	0.046	0.040	0.048	0.050	0.046	0.051	0.064	0.057
$\rho = 0.4$	0.147	0.084	0.058	0.128	0.063	0.046	0.139	0.076	0.049	0.142	0.081	0.060
$\rho = 0.9$	0.566	0.124	0.050	0.531	0.098	0.044	0.516	0.112	0.052	0.522	0.108	0.057
$(N, T) = (100, 20)$												
$\rho = 0.0$	0.050	0.068	0.050	0.048	0.080	0.042	0.054	0.076	0.052	0.072	0.104	0.064
$\rho = 0.4$	0.127	0.079	0.054	0.124	0.094	0.042	0.107	0.078	0.056	0.146	0.106	0.054
$\rho = 0.9$	0.358	0.072	0.074	0.319	0.069	0.061	0.315	0.080	0.081	0.370	0.114	0.086

Note: Empirical rejection frequencies of the Score and Wald tests under the null. The nominal level is 0.05. They are based on 2000 simulation replications. We use the analytical bias correction estimator $\widehat{\boldsymbol{\theta}}_{\mathrm{BC}}(\tau)$ to estimate regression coefficients. W_0 denotes the test using conventional variance matrix estimator assuming independent errors, W_1 uses the CCM estimator without bias-correction $\widehat{J}_T$, and W_2 uses the CCM estimator with bias-correction $\check{J}_T$. The same naming convention applies to the Score tests. The only difference is that the Score test uses the simplified CCM estimator.

Table S.4. Size of tests, Model 2.

	$\tau = 0.5$						$\tau = 0.75$					
Cases	H_0	H_1	H_2	W_0	W_1	W_2	H_0	H_1	H_2	W_0	W_1	W_2
$(N, T) = (50, 50)$												
$\rho = 0.0$	0.057	0.066	0.056	0.047	0.058	0.044	0.049	0.058	0.050	0.058	0.071	0.058
$\rho = 0.4$	0.146	0.094	0.052	0.118	0.066	0.037	0.118	0.091	0.049	0.120	0.080	0.046
$\rho = 0.9$	0.494	0.106	0.075	0.450	0.105	0.052	0.445	0.126	0.073	0.482	0.121	0.052
$(N, T) = (50, 20)$												
$\rho = 0.0$	0.060	0.080	0.056	0.034	0.051	0.034	0.052	0.071	0.049	0.058	0.078	0.044
$\rho = 0.4$	0.150	0.102	0.062	0.105	0.079	0.038	0.109	0.088	0.044	0.127	0.090	0.050
$\rho = 0.9$	0.334	0.059	0.061	0.262	0.070	0.054	0.272	0.079	0.070	0.326	0.098	0.058
$(N, T) = (20, 100)$												
$\rho = 0.0$	0.044	0.047	0.042	0.038	0.042	0.038	0.046	0.049	0.046	0.048	0.059	0.052
$\rho = 0.4$	0.150	0.082	0.053	0.120	0.062	0.044	0.146	0.076	0.054	0.132	0.074	0.056
$\rho = 0.9$	0.558	0.123	0.050	0.527	0.090	0.045	0.512	0.106	0.054	0.518	0.104	0.056
$(N, T) = (100, 20)$												
$\rho = 0.0$	0.059	0.072	0.058	0.040	0.064	0.033	0.054	0.066	0.050	0.079	0.095	0.059
$\rho = 0.4$	0.134	0.078	0.052	0.100	0.067	0.034	0.092	0.066	0.046	0.130	0.086	0.044
$\rho = 0.9$	0.354	0.068	0.069	0.310	0.066	0.056	0.312	0.082	0.085	0.360	0.112	0.074

Note: Empirical rejection frequencies of the Score and Wald tests under the null. The nominal level is 0.05. They are based on 2000 simulation replications. See descriptions in Table S.3 for details.

TABLE S.5. Power of tests, Model 1.

Cases	$\tau=0.5$						$\tau=0.75$					
	H_2			W_2			H_2			W_2		
	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$
$(N,T)=(50,50)$												
$\rho=0.0$	0.976	1.000	1.000	0.966	1.000	1.000	0.956	1.000	1.000	0.947	1.000	1.000
$\rho=0.4$	0.868	1.000	1.000	0.828	1.000	1.000	0.835	1.000	1.000	0.820	1.000	1.000
$\rho=0.9$	0.440	0.940	0.999	0.362	0.900	0.998	0.390	0.895	0.994	0.337	0.856	0.992
$(N,T)=(50,20)$												
$\rho=0.0$	0.737	0.998	1.000	0.618	0.996	1.000	0.618	0.994	1.000	0.590	0.987	0.998
$\rho=0.4$	0.542	0.976	0.999	0.434	0.954	1.000	0.454	0.950	0.999	0.440	0.934	0.998
$\rho=0.9$	0.475	0.948	0.999	0.424	0.920	0.997	0.398	0.902	0.996	0.337	0.863	0.990
$(N,T)=(20,100)$												
$\rho=0.0$	0.946	1.000	1.000	0.920	1.000	1.000	0.901	1.000	1.000	0.872	1.000	1.000
$\rho=0.4$	0.756	1.000	1.000	0.708	1.000	1.000	0.716	0.998	1.000	0.699	0.999	1.000
$\rho=0.9$	0.242	0.694	0.965	0.214	0.628	0.932	0.234	0.640	0.926	0.243	0.644	0.922
$(N,T)=(100,20)$												
$\rho=0.0$	0.944	0.999	1.000	0.912	1.000	1.000	0.889	1.000	1.000	0.864	1.000	1.000
$\rho=0.4$	0.830	1.000	1.000	0.753	0.999	1.000	0.768	0.999	0.998	0.738	0.996	0.998
$\rho=0.9$	0.753	0.994	0.998	0.698	0.995	0.998	0.648	0.990	0.996	0.634	0.990	0.998

Note: Empirical rejection frequencies of the bias-corrected robust tests under the alternative. The nominal level is 0.05. They are based on 2000 simulation replications. W_2 and H_2 denote the Wald and Score tests using the bias-corrected CCM estimator, $\check{J}_T$ and $\widehat{J}^s_T$, respectively. In simulation, the regression coefficient γ is equal to $c_h \cdot s_Y$ where s_Y denote the standard deviation of Y under the null. So c_h measures the magnitude of the effect under the alternative.

TABLE S.6. Power of tests, Model 2.

Cases	$\tau=0.5$						$\tau=0.75$					
	H_2			W_2			H_2			W_2		
	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$	$c_h=0.2$	$c_h=0.4$	$c_h=0.6$
$(N,T)=(50,50)$												
$\rho=0.0$	0.919	1.000	1.000	0.885	1.000	1.000	0.872	1.000	1.000	0.854	1.000	1.000
$\rho=0.4$	0.709	0.999	1.000	0.620	0.996	1.000	0.634	0.995	1.000	0.625	0.995	1.000
$\rho=0.9$	0.297	0.809	0.988	0.237	0.716	0.966	0.258	0.718	0.973	0.205	0.652	0.945
$(N,T)=(50,20)$												
$\rho=0.0$	0.594	0.989	1.000	0.456	0.965	1.000	0.449	0.960	1.000	0.457	0.947	0.999
$\rho=0.4$	0.420	0.923	0.998	0.274	0.858	0.989	0.344	0.840	0.992	0.319	0.824	0.988
$\rho=0.9$	0.333	0.827	0.984	0.298	0.789	0.973	0.247	0.728	0.970	0.229	0.675	0.948
$(N,T)=(20,100)$												
$\rho=0.0$	0.862	0.999	1.000	0.812	1.000	1.000	0.793	0.999	1.000	0.748	0.999	1.000
$\rho=0.4$	0.595	0.990	1.000	0.541	0.987	1.000	0.534	0.984	1.000	0.528	0.982	1.000
$\rho=0.9$	0.167	0.500	0.833	0.143	0.442	0.781	0.151	0.441	0.789	0.154	0.459	0.774
$(N,T)=(100,20)$												
$\rho=0.0$	0.846	1.000	1.000	0.755	0.999	1.000	0.707	0.999	1.000	0.750	0.995	1.000
$\rho=0.4$	0.709	0.999	1.000	0.594	0.994	1.000	0.544	0.985	1.000	0.568	0.982	1.000
$\rho=0.9$	0.588	0.988	1.000	0.524	0.971	1.000	0.406	0.942	0.999	0.401	0.923	1.000

Note: Empirical rejection frequencies of the bias-corrected robust Score and Wald tests under the alternative. The nominal level is 0.05. They are based on 2000 simulation replications. See descriptions in Table S.5 for details.

S.2. Proofs

In this section, the quantile index τ will be dropped in quantities like $\alpha_i(\tau)$, $\boldsymbol{\theta}(\tau)$ or $e_{it}(\tau)$ to save space, but we want to remind readers that they are quantile specific. Recall that $\boldsymbol{x}_{it} := (\boldsymbol{w}'_{it}, z'_{it})'$, $\boldsymbol{\theta} := (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, $\widetilde{\boldsymbol{x}}_{it} := \boldsymbol{x}_{it} - \boldsymbol{c}_i$, and let $a \vee b = \max(a, b)$.

S.2.1 *Proofs for Section 4*

Proof of Theorem 4.1. It will be sufficient to show that

$$\sqrt{N_1 T}\{\widehat{\Gamma}_{j,1} - \Gamma_j + B_j/T\} = \frac{1}{\sqrt{N_1 T}} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} \big\{\big(\tau - I(e_{it} \le 0)\big)\big(\tau - I(e_{it+j} \le 0)\big)\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it+j} - \Gamma_j\big\} + o_p(1), \tag{S.4}$$

$$\sqrt{N_2 T}\{\widehat{\Gamma}_{j,2} - \Gamma_j + B_j/T\} = \frac{1}{\sqrt{N_2 T}} \sum_{i=N_1+1}^{N} \sum_{t=1}^{T-j} \big\{\big(\tau - I(e_{it} \le 0)\big)\big(\tau - I(e_{it+j} \le 0)\big)\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it+j} - \Gamma_j\big\} + o_p(1). \tag{S.5}$$

We prove (S.4) only because the same argument applies to (S.5). As a convenient intermediate step, define

$$\breve{\Gamma}_j = \frac{1}{N_1(T-j)} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} \big(\tau - I(\widehat{e}_{2,it} \le 0)\big)\big(\tau - I(\widehat{e}_{2,it+j} \le 0)\big)(\boldsymbol{x}_{it} - \boldsymbol{c}_i)(\boldsymbol{x}_{it+j} - \boldsymbol{c}_i)'.$$

This quantity uses $\boldsymbol{c}_i$ instead of $\widehat{\boldsymbol{c}}_i$. We first study the asymptotic expansion of $\breve{\Gamma}_j$. Lemma S.6 in Yoon and Galvao (2019) shows that the difference between $\breve{\Gamma}_j$ and $\widehat{\Gamma}_{j,1}$ is negligible.

Decompose $\breve{\Gamma}_j$ using an identity, $\tau - I(\widehat{e}_{it} \le 0) = \tau - I(e_{it} \le 0) - (I(\widehat{e}_{it} \le 0) - I(e_{it} \le 0))$:

$$\begin{aligned}
\breve{\Gamma}_j &= \frac{1}{N_1(T-j)} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} \big(\tau - I(e_{it} \le 0)\big)\big(\tau - I(e_{it+j} \le 0)\big)\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it+j} \\
&\quad - \frac{1}{N_1(T-j)} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} \big(\tau - I(e_{it} \le 0)\big)\big(I(\widehat{e}_{2,it+j} \le 0) - I(e_{it+j} \le 0)\big)\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it+j} \\
&\quad - \frac{1}{N_1(T-j)} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} \big(\tau - I(e_{it+j} \le 0)\big)\big(I(\widehat{e}_{2,it} \le 0) - I(e_{it} \le 0)\big)\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it+j} \\
&\quad + \frac{1}{N_1(T-j)} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} \big(I(\widehat{e}_{2,it} \le 0) - I(e_{it} \le 0)\big)\big(I(\widehat{e}_{2,it+j} \le 0) - I(e_{it+j} \le 0)\big)\widetilde{\boldsymbol{x}}_{it}\widetilde{\boldsymbol{x}}'_{it+j} \\
&= B_1 - B_2 - B_3 + B_4.
\end{aligned}$$

The first term B_1 converges to $\mathit{\Gamma}_j$ in probability by the law of large numbers. Lemma S.3 in Yoon and Galvao (2019) shows that

$$B_2 = N_1^{-1}\sum_{i=1}^{N_1} \mathrm{E}\Big[\frac{f_i(0|\boldsymbol{x}_{i1+j})}{f_i(0)}\widetilde{\boldsymbol{x}}_{i1}\widetilde{\boldsymbol{x}}'_{i1+j}\Big]V_{T,i}/T + O_p\big(N_1^{-1/2}T^{-3/4}\big),$$

$$B_3 = N_1^{-1}\sum_{i=1}^{N_1} \mathrm{E}\Big[\frac{f_i(0|\boldsymbol{x}_{i1})}{f_i(0)}\widetilde{\boldsymbol{x}}_{i1}\widetilde{\boldsymbol{x}}'_{i1+j}\Big]V_{T,i}/T + O_p\big(N_1^{-1/2}T^{-3/4}\big).$$

Lemma S.5 in Yoon and Galvao (2019) shows that

$$B_4 \to N_1^{-1}\sum_{i=1}^{N_1} \mathrm{E}\Big[\frac{f_{ij}(0,0|\boldsymbol{x}_{i1},\boldsymbol{x}_{i1+j})}{f_i(0)^2}\widetilde{\boldsymbol{x}}_{i1}\widetilde{\boldsymbol{x}}'_{i1+j}\Big]V_{T,i}/T + O_p\big(N_1^{-1/2}T^{-3/4}\big).$$

This leads to the asymptotic bias expression in Theorem 4.1. □

Proof of Theorem 4.2. Proposition 1 in Andrews (1991) does not require differentiability of the moment function and is applicable to the pseudo-estimator $\widetilde{J}_T$. The plan is to establish the asymptotic properties of $\widetilde{J}_T$ first, then study any difference between $\widehat{J}_T$ and $\widetilde{J}_T$. Rewrite

$$\widetilde{J}_T = \frac{1}{N}\sum_{i=1}^{N}\widetilde{J}_{T,i} \quad \text{where } \widetilde{J}_{T,i} = \sum_{j=-T+1}^{T-1} k\Big(\frac{j}{m_T}\Big)\widetilde{\mathit{\Gamma}}_{ij}.$$

Note that $\widetilde{J}_{T,i}$ and $\widetilde{\mathit{\Gamma}}_{ij}$ are estimators for the long-run variance and the jth order autocovariance using a single time-series. The sample splitting in $\widehat{\mathit{\Gamma}}_j$ naturally leads to a similar decomposition to $\widehat{J}_T = \frac{N_1}{N}\cdot\widehat{J}_{T,1} + \frac{N_2}{N}\cdot\widehat{J}_{T,2}$ where

$$\widehat{J}_{T,1} = \sum_{j=-T+1}^{T-1} k\Big(\frac{j}{m_T}\Big)\frac{T-|j|}{T}\widehat{\mathit{\Gamma}}_{j,1}, \quad \text{and} \quad \widehat{J}_{T,2} = \sum_{j=-T+1}^{T-1} k\Big(\frac{j}{m_T}\Big)\frac{T-|j|}{T}\widehat{\mathit{\Gamma}}_{j,2}.$$

Divide J_T accordingly $J_T = N_1/N\cdot J_{T,1} + N_2/N\cdot J_{T,2}$ where $J_{T,1}$ and $J_{T,2}$ uses $i\in I_1$ and I_2, respectively. It suffices to show that the claim for $\widehat{J}_{T,1}$. Consider first the bias:

$$\begin{aligned}\mathrm{E}[\widehat{J}_{T,1} - J_{T,1}] = &\sum_{j=-T+1}^{T-1}\Big\{k\Big(\frac{j}{m_T}\Big)-1\Big\}\frac{T-|j|}{T}\mathit{\Gamma}_j \\ &+ \sum_{j=-T+1}^{T-1} k\Big(\frac{j}{m_T}\Big)\frac{T-|j|}{T}\mathrm{E}[\widehat{\mathit{\Gamma}}_{j,1}-\mathit{\Gamma}_j] - \sum_{j=-T+1}^{T-1}\frac{|j|}{T}\mathit{\Gamma}_j.\end{aligned}$$

The first bias term comes from that we use a kernel method. Parzen (1957) showed that m_T^q times the first term in the right-hand side converges to $-k_q J^{(q)}$. So the order of the first term is $O(m_T^{-q})$. The second bias term is due to the bias in $\widehat{\Gamma}_{j,1}$. From Theorem 4.1, $\mathrm{E}[\widehat{\Gamma}_{j,1} - \Gamma_j] = -B_j/T + s_T$ where s_T denotes remainder terms of order $O(N^{-1/2}T^{-3/4})$. Since $m_T^{-1}\sum_{j=-T+1}^{T-1} k(j/m_T) \to \int k(x)\,dz$ and $m_T^{-1}\sum_{j=-T+1}^{T-1} k(j/m_T)B_j \to D$, two terms converge to objects of order $O_p(m_T/T)$ and $O_p(m_T/(N^{1/2}T^{3/4}))$. The second term is smaller than the first and can be ignored if $N^2 \gg T$. For the third term, $|\sum_{j=-T+1}^{T-1} |j|\Gamma_j| \le |J^{(q)}|$, so it is of order $O(1/T)$. In sum, the order of the bias is

$$\mathrm{E}[\widehat{J}_{T,1} - J_{T,1}] = O_p\left(\frac{1}{m_T^q}\right) + O_p\left(\frac{m_T}{T}\right) + O_p\left(\frac{m_T}{N^{1/2}T^{3/4}}\right) + O_p\left(\frac{1}{T}\right).$$

The first term dominates the second if $m_T^{1+q}/T \to 0$. If, however, $m_T^{1+q}/T \to \nu$, the two terms have the same order.

For the variance calculation, first calculate the variance of $\widetilde{J}_{T,i}$. Fix $i \in I_1$. For this, Proposition 1 in Andrews (1991) is applicable. If $m_T \to \infty$ and $m_T/T \to 0$,

$$\frac{T}{m_T}\mathrm{Var}(\widetilde{J}_{T,i}) \to 2J^2 \int k^2(z)\,dz.$$

So $\mathrm{Var}(\widetilde{J}_{ij})$ has order m_T/T. Next, we aim to show that

$$\widehat{J}_{T,i} - \widetilde{J}_{T,i} = o_p\left(\left(\frac{m_T}{T}\right)^{1/2}\right). \tag{S.6}$$

Define

$$\begin{aligned} D_T &:= \sqrt{\frac{T}{m_T}}(\widehat{J}_{T,i} - \widetilde{J}_{T,i}) = \sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right)\frac{T-|j|}{T}(\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij}) \\ &= \sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right)(\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij}) + \sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right)\frac{|j|}{T}(\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij}) \\ &= D_{T,1} + D_{T,2}. \end{aligned} \tag{S.7}$$

Consider $\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij}$ which uses a single time-series. By the stochastic equicontinuity argument, $\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij} = \mathrm{E}[\widehat{\Gamma}_{ij}] - \mathrm{E}[\widetilde{\Gamma}_{ij}] + r_T$. Kato, Galvao, and Montes-Rojas (2012) calculated that $r_T = O_p(T^{-3/4})$; see the proof of their Theorems 3.2 and 5.1. By the Taylor expansion, just as in Lemma S.5 in Yoon and Galvao (2019),

$$\mathrm{E}[\widehat{\Gamma}_{ij}] - \mathrm{E}[\widetilde{\Gamma}_{ij}] = 0 \cdot (\widehat{\alpha}_i - \alpha_{i0}) + \omega_{ij}(\widehat{\alpha}_i - \alpha_i)^2 + 0 \cdot (\widehat{\theta} - \theta_0) + O_p\big((\widehat{\alpha}_i - \alpha_i)^3\big)$$

with ω_{ij} as defined in Theorem 4.1. So $\widehat{\Gamma}_{j,i} - \widetilde{\Gamma}_{j,i} = \omega_{ij}(\widehat{\alpha}_i - \alpha_i)^2 + r_T + O_p((\widehat{\alpha}_i - \alpha_i)^3)$.

Consider $D_{T,1}$. Plugging the main terms,

$$D_{T,1} = \sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right)\big(\omega_{ij}(\widehat{\alpha}_i - \alpha_i)^2 + r_T\big) = \mathrm{E}_{T,1} + \mathrm{E}_{T,2}. \tag{S.8}$$

Denote $\bar{\omega} = \sup_i \sup_j |\omega_{ij}| < \infty$ by condition (A5). $|\mathrm{E}_{T,1}|$ is less than or equal to

$$\sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} \left| k\left(\frac{j}{m_T}\right)\right| \cdot \left|\omega_{ij}(\widehat{\alpha}_i - \alpha_i)^2\right|$$

$$\leq \bar{\omega}\sqrt{m_T} \cdot \frac{1}{m_T} \sum_{j=-T+1}^{T-1} \left| k\left(\frac{j}{m_T}\right)\right| \cdot \sqrt{T} \cdot O_p(T^{-1})$$

$$\to \bar{\omega}\sqrt{m_T} \int \left|k(z)\right| dz \cdot O_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\sqrt{\frac{m_T}{T}}\right).$$

So $\mathrm{E}_{T,1} = o_p(1)$ provided that $m_T/T \to 0$. Next, $|\mathrm{E}_{T,2}|$ is less than or equal to

$$\sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} \left| k\left(\frac{j}{m_T}\right)\right| \cdot |r_T|$$

$$= \sqrt{m_T} \cdot \frac{1}{m_T} \sum_{j=-T+1}^{T-1} \left| k\left(\frac{j}{m_T}\right)\right| \cdot \sqrt{T}|r_T|$$

$$\to \sqrt{m_T} \int \left|k(z)\right| dz \cdot \sqrt{T} \cdot O_p(r_T) = O_p\left(m_T^{1/2} \cdot T^{-1/4}\right) = O_p\left((m_T^2/T)^{1/4}\right).$$

So $\mathrm{E}_{T,2} = o_p(1)$ provided that $m_T^2/T \to 0$. Consider $D_{T,2}$. It is equal to

$$\sqrt{\frac{T}{m_T}} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right)\frac{|j|}{T}(\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij}) = \sqrt{m_T}\frac{1}{m_T} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right)\left|\frac{j}{m_T}\right|\frac{m_T}{T}(\widehat{\Gamma}_{ij} - \widetilde{\Gamma}_{ij}).$$

Since $\frac{1}{m_T}\sum_{j=-T+1}^{T-1} k(\frac{j}{m_T})|\frac{j}{m_T}| \to \int |z|k(z)\,dz$, compared to $D_{T,1}$, there is an additional factor of order T^{-1}. So it can be shown that $D_{T,2} = O_p(\frac{m_T}{T^3}) + O_p(\frac{m_T^2}{T^3})$. This proves (S.6).

Finally, observe that $\widehat{J}_{T,1} = \sum_{j=-T+1}^{T-1} k(\frac{j}{m_T})\frac{T-|j|}{T}\widehat{\Gamma}_{j,1}$ can be written as $\widehat{J}_{T,1} = \frac{1}{N_1} \times \sum_{i=1}^{N_1} R_{T,i}$ where $R_{T,i}$ is equal to

$$\sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right) \sum_{t=1}^{T-j}\left(\tau - I\left(\widehat{e}_{2,it}(\tau) \leq 0\right)\right)\left(\tau - I\left(\widehat{e}_{2,it+j}(\tau) \leq 0\right)\right)(\boldsymbol{x}_{it} - \widehat{\boldsymbol{c}}_i)(\boldsymbol{x}_{it+j} - \widehat{\boldsymbol{c}}_i)'.$$

By condition (A1) and the sample splitting, $R_{T,1}, \ldots, R_{T,N_1}$ are independent, therefore,

$$\frac{N_1 T}{m_T}\,\mathrm{Var}(\widehat{J}_{T,1}) \to 2J^2 \int k^2(z)\,dz.$$

The same arguments work for $\widehat{J}_{T,2}$. So conclude that $\frac{NT}{m_T}\,\mathrm{Var}(\widehat{J}_T) \to 2J^2 \int k^2(z)\,dz$.

Importantly, the order of the variance $\mathrm{Var}(\widehat{J}_T) = O(\frac{m_T}{NT})$. Taking averages over i does not change the order of the bias, but it affects the variance by factor of $1/N$. □

Proof of Proposition 4.2 is similar to Theorem 4.1 and omitted. Proof of Theorem 4.3 and Theorem 4.4 is similar to Theorem 4.2. Details can be found in Section S.1 of Yoon and Galvao (2019).

S.2.2 *Proofs for Section 5*

It suffices to show that $S_1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + o_p(1)$ where

$$S_1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = (N_1 T)^{-1/2} \sum_{i=1}^{N_1} \sum_{t=1}^{T} \big(\tau - I\big(y_{it} \le \widehat{\alpha}_i + \boldsymbol{w}_{it}' \widehat{\boldsymbol{\beta}}_2\big)\big) \widetilde{z}_{it}, \quad \text{and}$$

$$S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = (N_1 T)^{-1/2} \sum_{i=1}^{N_1} \sum_{t=1}^{T} \big(\tau - I\big(y_{it} \le \alpha_{i0} + \boldsymbol{w}_{it}' \boldsymbol{\beta}_0\big)\big) \widetilde{z}_{it}.$$

It is formally stated in Lemma S.3. A similar result can be shown for $S_2(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$.

LEMMA S.1. *Assume conditions in Theorem* 5.1, *then*

$$\big\{S_1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) - S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\big\} - \big\{\mathrm{E}\big[S_1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})\big] - \mathrm{E}\big[S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\big]\big\} = o_p(1).$$

PROOF. The consistency of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ under given assumptions can be found in Proposition 3.1 in GK for the two bias corrected estimators. For each $\delta > 0$ and a constant $M_1 > 0$, define $B_i(\delta) = \{(\alpha, \boldsymbol{\beta}) : |\alpha_i - \alpha_{i0}| \le \delta, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \delta, |\alpha| \le M_1, \|\boldsymbol{\beta}\| \le M_1\}$. Define

$$\psi_{it}(\alpha_i, \boldsymbol{\beta}) = I\big(e_{it} \le \big(\alpha_i - \alpha_{i0} + \boldsymbol{w}_{it}'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\big)\big) - I(e_{it} \le 0),$$

and

$$\Pi_i^*(\alpha_i, \boldsymbol{\beta}) = -T^{-1/2} \sum_{t=1}^{T} \psi_{it}(\alpha_i, \boldsymbol{\beta}) \widetilde{z}_{it}.$$

Note that $S_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) - S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = N_1^{-1/2} \sum_{i=1}^{N_1} \Pi_i^*(\alpha_i, \boldsymbol{\beta})$. When there is no confusion, we will use ψ_{it} as a short notation for $\psi_{it}(\alpha_i, \boldsymbol{\beta})$. Also without loss of generality let $\boldsymbol{\alpha}_0 = \mathbf{0}$, $\boldsymbol{\beta}_0 = \mathbf{0}$, then from $B_i(\delta)$ we may drop the subscript i and simply write it $B(\delta)$.

We aim to show that for any $(\alpha, \boldsymbol{\beta}) \in B(\delta)$, $\max_{1 \le i \le N_1} \mathrm{Var}[\Pi_i^*(\alpha_i, \boldsymbol{\beta})] = o_p(1)$ as $T \to \infty$ and $\delta = \delta_T \to 0$ a decreasing sequence of real numbers. Since $\widetilde{z}_{it}$ is bounded by condition (A2), it suffices to show the claim without $\widetilde{z}_{it}$, which is what we shall establish now. Let $\Pi_i(\alpha_i, \boldsymbol{\beta})$ be $\Pi_i^*(\alpha_i, \boldsymbol{\beta})$ without $\widetilde{z}_{it}$. By stationarity,

$$\begin{aligned}\mathrm{Var}\big[\Pi_i(\alpha_i, \boldsymbol{\beta})\big] &= \mathrm{Var}[\psi_{it}] + 2 \sum_{j=1}^{T} \left(1 - \frac{j}{T}\right) \mathrm{Cov}(\psi_{i1}, \psi_{ij+1}) \\ &\le \mathrm{Var}[\psi_{it}] + 2 \sum_{j=1}^{T} \big|\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})\big|. \end{aligned} \tag{S.9}$$

Consider the variance term

$$\begin{aligned}\mathrm{Var}[\psi_{it}] &\leq \mathrm{E}\big[\psi_{it}^2\big] = \mathrm{E}\big[|\psi_{it}|\big] \leq F_i\big(\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big) - F_i\big(-\alpha_i - \boldsymbol{w}_{it}'\boldsymbol{\beta}\big)\\ &\leq 2C_f\big(|\alpha_i| + M\|\boldsymbol{\beta}\|\big) \leq C\delta_T. \end{aligned} \tag{S.10}$$

In the above inequalities, we use the continuity of $F_i(\cdot)$, the mean value theorem, and (A5). The overall constant $C > 0$ does not depend on i. Let k_T be a sequence of increasing real numbers. We divide covariance terms into two parts $\sum_{j=1}^{k_T} |\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})| + \sum_{j=k_T+1}^{T} |\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})|$. For $\sum_{j=1}^{k_T} |\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})|$, observe that

$$\big|\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})\big| \leq \mathrm{E}\big[|\psi_{i1}\cdot\psi_{ij+1}|\big] + \mathrm{E}\big[|\psi_{i1}|\big]^2. \tag{S.11}$$

The first term in the right-hand side of (S.11) is smaller than

$$\begin{aligned}&\int\!\!\int I\big(-\big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big| \leq e_{i1} \leq \big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big|\big)\\ &\qquad \times I\big(-\big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big| \leq e_{ij+1} \leq \big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big|\big) f_{i,j}(e_{i1}, e_{ij+1})\, de_{i1}\, de_{ij+1}\\ &\leq C_f' \int\!\!\int I\big(-\big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big| \leq e_{i1} \leq \big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big|\big)\\ &\qquad \times I\big(-\big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big| \leq e_{ij+1} \leq \big|\alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}\big|\big)\, de_{i1}\, de_{ij+1}\\ &\leq 2C_f'\big(|\alpha_i| + M\|\boldsymbol{\beta}\|\big) \leq C'\delta_T.\end{aligned}$$

In the first inequality we use that $f_{i,j}(e_{i1}, e_{ij+1})$ is bounded from above. For the second term in the right-hand side of (S.11), by employing the same arguments we used for variance, we have $\mathrm{E}[|\psi_{i1}(\alpha_i, \boldsymbol{\beta})|] \leq C\delta_T$, therefore, $\mathrm{E}[|\psi_{i1}(\alpha_i, \boldsymbol{\beta})|]^2 \leq C^2\delta_T^2$. We conclude that

$$\sum_{j=1}^{k_T}\big|\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})\big| \leq C'' k_T \delta_T^2, \tag{S.12}$$

where the overall constant C'' is independent of i. We now show that $\sum_{j=k_T+1}^{T} |\mathrm{Cov}(\psi_{i1}, \psi_{ij+1})| = o_p(1)$ uniformly over i. For this, we use a covariance inequality of mixing random variables. Let $L_{it} = \alpha_i + \boldsymbol{w}_{it}'\boldsymbol{\beta}$ and observe that $\psi_{it}(\alpha_i, \boldsymbol{\beta})$ is an indicator function taking $I(0 \leq e_{it} \leq L_{it})$ if $L_{it} \geq 0$ and $-I(-L_{it} \leq e_{it} \leq 0)$ if $L_{it} < 0$. Depending on signs of L_{i1} and L_{ij+1}, we have four distinct cases. First, when $L_{i1} \geq 0$ and $L_{ij+1} \geq 0$,

$$\begin{aligned}&\big|\mathrm{Cov}\big(\psi_{i1}(\alpha_i, \boldsymbol{\beta}), \psi_{ij+1}(\alpha_i, \boldsymbol{\beta})\big)\big|\\ &\quad = \big|\mathrm{E}\big[\psi_{i1}(\alpha_i, \boldsymbol{\beta})\cdot\psi_{ij+1}(\alpha_i, \boldsymbol{\beta})\big] - \mathrm{E}\big[\psi_{i1}(\alpha_i, \boldsymbol{\beta})\big]\mathrm{E}\big[\psi_{ij+1}(\alpha_i, \boldsymbol{\beta})\big]\big|\\ &\quad = \big|\Pr(0 \leq e_{i1} \leq L_{i1}, 0 \leq e_{ij+1} \leq L_{ij+1}) - \Pr(0 \leq e_{i1} \leq L_{i1})\Pr(0 \leq e_{ij+1} \leq L_{ij+1})\big|\\ &\quad \leq 2\beta(j),\end{aligned}$$

by the definition of beta-mixing coefficient (see Bradley (2005)). The second case is when $L_{i1} \geq 0$ and $L_{ij+1} < 0$, we see that $|\operatorname{Cov}(\psi_{i1}(\alpha_i, \boldsymbol{\beta}), \psi_{ij+1}(\alpha_i, \boldsymbol{\beta}))|$ is equal to

$$\begin{aligned}
&\big|-\Pr(0 \leq e_{i1} \leq L_{i1}, 0 \leq e_{ij+1} \leq L_{ij+1}) - \Pr(0 \leq e_{i1} \leq L_{i1})\big(-\Pr(0 \leq e_{ij+1} \leq L_{ij+1})\big)\big| \\
&\quad = \big|\Pr(0 \leq e_{i1} \leq L_{i1}, 0 \leq e_{ij+1} \leq L_{ij+1}) - \Pr(0 \leq e_{i1} \leq L_{i1})\Pr(0 \leq e_{ij+1} \leq L_{ij+1})\big| \leq 2\beta(j).
\end{aligned}$$

The other two cases can be analyzed similarly, so we conclude that $|\operatorname{Cov}(\psi_{i1}, \psi_{ij+1})| \leq 2\beta(j)$. Invoke condition (A1) and conclude that for a constant $0 < a < 1$,

$$\sum_{j=k_T+1}^{T} \big|\operatorname{Cov}(\psi_{i1}, \psi_{ij+1})\big| \leq 2 \sum_{j=k_T+1}^{T} \beta(j) \leq 2B \sum_{j=k_T+1}^{T} a^j \leq 2B(1-a)^{-1} a^{k_T}, \tag{S.13}$$

where constants above are independent of i. From (S.10), (S.12), and (S.13), we have

$$\operatorname{Var}\big[\varPi_i(\alpha_i, \boldsymbol{\beta})\big] = O_p(\delta_T) + O_p(k_T \delta_T) + O_p\big(a^{k_T}\big).$$

Take $k_T = 1/\delta_T^{1/2}$. Since $a^{k_T} \ll k_T^{-1}$ for k_T sufficiently large, $\operatorname{Var}[\varPi_i(\alpha_i, \boldsymbol{\beta})] = O_p(\delta_T) + O_p(\delta_T^{1/2}) + o_p(\delta_T) = o_p(1)$ as $\delta_T \to 0$. Thus, an application of Chebyshev's inequality leads that for any $(\alpha, \boldsymbol{\beta}) \in B(\delta_T)$, uniformly for $1 \leq i \leq N$,

$$\big\|\varPi_i(\alpha, \boldsymbol{\beta}) - \mathrm{E}\big[\varPi_i(\alpha, \boldsymbol{\beta})\big]\big\| = o_p(1). \tag{S.14}$$

By using a chaining argument such as the one in Koenker and Zhao (1996) (see their Lemma A.2), we can extend (S.14) to

$$\sup_{(\alpha, \boldsymbol{\beta}) \in B(\delta_T)} \big\|\varPi_i(\alpha, \boldsymbol{\beta}) - \mathrm{E}\big[\varPi_i(\alpha, \boldsymbol{\beta})\big]\big\| = o_p(1), \tag{S.15}$$

uniformly for $1 \leq i \leq N_1$. Now we utilize the independence across i to obtain

$$\sup_{(\boldsymbol{\alpha}, \boldsymbol{\beta})} \big\|\big\{S_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) - S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\big\} - \big\{\mathrm{E}\big[S_1(\boldsymbol{\alpha}, \boldsymbol{\beta})\big] - \mathrm{E}\big[S_1(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)\big]\big\}\big\| = o_p(1),$$

where the supremum is taken over $\{(\boldsymbol{\alpha}, \boldsymbol{\beta}) : \max_{1 \leq i \leq N} |\alpha_i| \leq \delta_T, \|\boldsymbol{\beta}\| \leq \delta_T\}$. Along with the consistency of regression coefficient estimates, this proves the claim. □

Lemma S.2. *Assume conditions in Theorem* 5.1, *then* $\mathrm{E}[S_1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})] = o_p(1)$.

Proof. By Taylor expansion, $\mathrm{E}[S_1(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})]$ is equal to

$$\begin{aligned}
&(N_1 T)^{-1/2} \sum_{i=1}^{N_1} \sum_{t=1}^{T} \mathrm{E}\big[\big(\tau - F_i\big((\widehat{\alpha}_i - \alpha_{i0}) + \boldsymbol{w}_{it}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\big)\big)\widetilde{z}_{it}\big] \\
&\quad = -(N_1 T)^{-1/2} \sum_{i=1}^{N_1} \sum_{t=1}^{T} \mathrm{E}\big[f_i\big(F_i^{-1}(\tau)\big)\big((\widehat{\alpha}_i - \alpha_{i0}) + \boldsymbol{w}_{it}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\big)\widetilde{z}_{it}\big] \\
&\qquad - (N_1 T)^{-1/2} \sum_{i=1}^{N_1} \sum_{t=1}^{T} \mathrm{E}\big[f_i\big(\widetilde{F}_i^{-1}(\tau)\big)\big((\widehat{\alpha}_i - \alpha_{i0})^2 + \big(\boldsymbol{w}_{it}'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\big)^2\big)\widetilde{z}_{it}\big].
\end{aligned}$$

Due to the construction of $\widetilde{\boldsymbol{Z}}$, the first-order terms in Taylor expansion become

$$-(N_1T)^{-1/2}\sum_{i=1}^{N_1} f_i(0)\left(\mathrm{E}\left[\sum_{t=1}^{T}\widetilde{z}_{it}\right](\widehat{\alpha}_i-\alpha_{i0})+\mathrm{E}\left[\sum_{t=1}^{T}\widetilde{z}_{it}\boldsymbol{w}_{it}'\right](\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right)=0.$$

The second-order terms in Taylor expansion become, using assumption (A2),

$$-(N_1T)^{-1/2}\sum_{i=1}^{N_1} f_i(\widetilde{0})\mathrm{E}\left[\sum_{t=1}^{T}\widetilde{z}_{it}\right](\widehat{\alpha}_i-\alpha_{i0})^2-\sqrt{N_1T}O_p\big(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2\big).$$

The first term is zero. The second term is of order $\sqrt{N_1T}\cdot O_p((N_1T)^{-1})=o_p(1)$. □

Lemma S.3. *Assume conditions in Theorem* 5.1, *then*

$$\widehat{S}_1(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}}_2)=S_1(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)+o_p(1).$$

Proof. It follows from Lemmas S.1 and S.2 and the fact that $\mathrm{E}[S_1(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)]=0$. □

Proof of Theorem 5.1. The claim (i) follows from Lemma S.3, and an application of Lyapunov CLT. The Lyapunov condition can be checked following Appendix A.4 in KGM, so we omit it. A straightforward variance calculation shows that the variance of $S(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$ is J_T^s. Given the consistency of $\widehat{J}_T^s$, the claim (ii) is conventional. □

Proof of Theorem 5.2. The consistency of $\widehat{\Lambda}$ and $\check{J}_T$ follow from Kato (2012) and Theorem 4.4, respectively. Given this, the proof Theorem 5.2 is conventional so we omit the details. □

References

Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59 (3), 817–858. [11, 12]

Bertrand, M., E. Duflo, and S. Mullainathan (2004), "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*, 119, 249–275. [2]

Bradley, R. C. (2005), "Basic properties of strong mixing conditions. A survey and some open questions." *Probability Surveys*, 2, 107–144. [16]

Kato, K. (2012), "Asymptotic normality of Powell's kernel estimator." *Annals of the Institute of Statistical Mathematics*, 64, 255–273. [17]

Kato, K., A. F. Galvao, and G. Montes-Rojas (2012), "Asymptotics for panel quantile regression models with individual effects." *Journal of Econometrics*, 170, 76–91. [12]

Koenker, R. and Q. Zhao (1996), "Conditional quantile estimation and inference for arch models." *Econometric Theory*, 12, 793–813. [16]

Parzen, E. (1957), "On consistent estimates of the spectrum of a stationary time series." *The Annals of Mathematical Statistics*, 28 (2), 329–348. [11]

Yoon, J. and A. F. Galvao (2019), "Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects." Working paper, available at https://ssrn.com/abstract=3446972. [1, 2, 3, 10, 11, 12, 14]

Co-editor Frank Schorfheide handled this manuscript.

Manuscript received 4 December, 2016; final version accepted 7 October, 2019; available online 5 November, 2019.