

# Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects

JUNGMO YOON

College of Economics and Finance, Hanyang University

ANTONIO F. GALVAO

Department of Economics, University of Arizona

This study develops cluster robust inference methods for panel quantile regression (QR) models with individual fixed effects, allowing for temporal correlation within each individual. The conventional QR standard errors can seriously underestimate the uncertainty of estimators and, therefore, overestimate the significance of effects, when outcomes are serially correlated. Thus, we propose a clustered covariance matrix (CCM) estimator to solve this problem. The CCM estimator is an extension of the heteroskedasticity and autocorrelation consistent covariance matrix estimator for QR models with fixed effects. The autocovariance element in the CCM estimator can be substantially biased, due to the incidental parameter problem. Thus, we develop a bias-correction method for the CCM estimator. We derive an optimal bandwidth formula that minimizes the asymptotic mean squared errors, and propose a data-driven bandwidth selection rule. We also propose two cluster robust tests, and establish their asymptotic properties. We then illustrate the practical usefulness of the proposed methods using an empirical application.

**KEYWORDS.** Cluster robust standard errors, quantile regression, panel data, heteroskedasticity and autocorrelation consistent covariance matrix estimation.

**JEL CLASSIFICATION.** C21, C23.

## 1. INTRODUCTION

This paper proposes cluster robust inference methods for panel quantile regression (QR) models under two conditions: first, we allow for temporal correlation within each individual (or cross-sectional group); second, we use individual fixed effects (FE) to control

---

Jungmo Yoon: [jmyoon@hanyang.ac.kr](mailto:jmyoon@hanyang.ac.kr)

Antonio F. Galvao: [agalvao@email.arizona.edu](mailto:agalvao@email.arizona.edu)

An earlier version of the paper was circulated under the title “Robust Inference for Panel Quantile Regression Models with Individual Fixed Effects and Serial Correlation.” The authors are grateful to the three anonymous referees for their constructive comments and suggestions. In addition, we thank Ryo Okui and the participants at the 2013 SETA meeting at Sungkyunkwan University, the 2013 California Econometrics Conference at UCLA, and the 2014 KEA-KAEA International Conference at Yonsei University for their helpful discussions and comments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2018S1A5A8027022).

for unobserved heterogeneity across individuals. The potential pitfall of ignoring within-group correlation in panel data mean regression models is illustrated by [Bertrand, Duflo, and Mullainathan \(2004\)](#). By generating a random fictitious law in a simulation, they found that if one ignores serial correlation, a test at the 5% level may reject the null hypothesis of no effect as often as 40% of the time, thus incorrectly categorizing an ineffective policy as effective too often. This tendency of overrejection persists, even if one includes individual FE.

[Arellano \(1987\)](#) suggested a robust covariance matrix estimator for a panel data model in the ordinary least squares (OLS) case. The estimator allows for FE and arbitrary within-cluster correlation structure. This clustered standard error is the method preferred by [Bertrand, Duflo, and Mullainathan \(2004\)](#) to correct the size distortion. The cluster robust method has been extended by [Kézdi \(2004\)](#), [Donald and Lang \(2007\)](#), and [Hansen \(2007\)](#), and found a wide audience in applied economics and finance research. [Cameron and Miller \(2015\)](#) provided a comprehensive survey of recent advancements in this literature. We show numerically that the conventional QR estimation for panel data models with FE suffers from similar size distortions to those seen in the least squares case. However, although it is increasingly common in many fields of empirical research to use a QR model to analyze panel data, to the best of our knowledge, there is no QR equivalent of Arellano's procedure. This study aims to fill this gap in the literature.

To solve the size distortion problem in QR panel models, we develop a cluster robust covariance matrix estimator, a quantile analogue to [Arellano \(1987\)](#). The proposed estimator, hereafter referred to as the clustered covariance matrix (CCM) estimator, is robust to heteroskedasticity and serial correlation of arbitrary forms. It is a panel data extension of the heteroskedasticity and autocorrelation consistent (HAC) estimation of the covariance matrix of the parameter estimators in a QR. The HAC covariance estimation for a QR and its asymptotic properties have not been actively studied, even for a single time series data. The main challenge is that the moment function that defines the autocovariance matrices is not differentiable. We extend the results of [Linton and Whang \(2007\)](#) and [Okui \(2010\)](#) to develop new methods for panel data QRs. The former work proposes a quantilogram in time series QRs and develops properties for the sample autocovariance. The latter work establishes asymptotic properties for the sample autocovariance in panel data for conditional mean models.

The CCM estimator for the quantile panel model has the standard sandwich form. The middle term, the autocovariance element, suffers from similar bias effects to those seen in the least squares case of the panel data models with FE, as observed in [Okui \(2010\)](#). We propose to use a bias-corrected estimator for the autocovariance terms. We then establish the asymptotic properties of the bias-corrected CCM estimator. This correction significantly improves the finite-sample performance of the CCM estimator.

To implement the proposed estimator in practice, we derive optimal bandwidth formulae that minimize the asymptotic mean squared errors. We also discuss a data driven bandwidth selection procedure. Establishing optimal bandwidths is important to making correct inferences in applications. Therefore, we present a data-driven bandwidth selection procedure. To the best of our knowledge, the asymptotic mean squared error (MSE) calculation of the long-run variance matrix estimator in a QR is new.

Note that to derive the results for the proposed CCM estimator for the panel QR, we use large panel asymptotics, where the cross-sectional dimension,  $N$ , is comparable to the time-series dimension,  $T$ .<sup>1</sup> In addition, the CCM requires restrictions on the time-series dependence through a mixing condition; as well as nuisance parameters, such as a bandwidth. This is in contrast to the cluster robust standard errors for panel data mean regression models, which do not require a mixing condition or a bandwidth. The mixing condition is one of the requirements to obtain consistent regression coefficient estimators in FE QR models.<sup>2</sup> The choice of a tuning parameter can be viewed as a shortcoming of the proposed method, although our simulation study shows that the proposed bandwidth selection rule exhibits excellent finite-sample performance. The existence of a consistent CCM estimator without truncation (that gives all observations equal weights) for FE QR models remains an open question.

We propose two cluster robust tests. The first test is a variant of the quantile rank-score test. When error terms share a common distribution, a rank-score test is distribution-free, meaning that its limiting distribution does not depend on conditional density functions (Gutenbrunner, Jureckova, Koenker, and Portnoy (1993)). We extend this idea to FE models; thus, the resulting Score test is distribution-free and easy to implement. The second test is an extension of the conventional Wald test for linear restrictions. Under regularity conditions, Score and Wald tests converge to a chi-squared distribution with degrees of freedom equal to the number of restrictions tested; therefore, the critical values for a given level of significance are widely available.

Finally, we provide an empirical example to illustrate the proposed methods. We examine the effects of randomly generated fictitious laws on state unemployment rates; the data cover 51 states over 37 years, and illustrate the main setup in which both  $N$  and  $T$  are large and comparable to each other. The results show that the conventional QR standard errors can seriously underestimate the uncertainty of estimators, but that the cluster robust standard errors are effective in correcting the size issue. We also examine, through simulation and a second real-data example, the performance of the CCM estimator when  $T$  is much larger than  $N$ , or vice versa, including a wide range of cases from  $(N, T) = (20, 100)$  to  $(N, T) = (400, 30)$ . Monte Carlo simulations can be found in the Appendix of the Supplemental Material (Yoon and Galvao (2020)).

Next, we briefly review the related literature. There is a well-established body of literature on QRs with dependent outcomes. For time-series data, refer to Koenker and Zhao (1996), Koenker and Xiao (2006), and Komunjer and Vuong (2010). For panel data, Kato, Galvao, and Montes-Rojas (2012) showed the consistency of QR estimators, and provided a Bahadur representation for the common regression coefficients under a mixing condition. However, to justify the Score test, we need a sharper convergence rate than those available for the dependent case. To achieve this, we adapt the methods of Masry and Fan (1997) and El Ghouch and van Keilegom (2009). However, Kato, Galvao,

---

<sup>1</sup>It is known that the FE QR estimator suffers from incidental parameter problems. Thus, it has become standard in the panel QR literature to employ large panel asymptotics, where the cross-sectional dimension is comparable to the time series dimension, together with bias correction for the estimation of the slope coefficients of interest.

<sup>2</sup>The current proof strategy also relies heavily on the mixing condition.

and Montes-Rojas (2012) did not provide any inference methods for the dependent outcomes. Thus, we complement their work by providing feasible and practical inference procedures.

There is also a growing body of literature on panel data QRs; see, for example, Abrevaya and Dahl (2008), Canay (2011), Galvao, Lamarche, and Lima (2013), Chernozhukov, Fernández-Val, Hahn, and Newey (2013), Arellano and Bonhomme (2015), Galvao and Wang (2015), Chetverikov, Larsen, and Palmer (2016), and Graham, Hahn, Poirier, and Powell (2018). This study adds to this literature by proposing cluster robust inference methods for panel QR models.

For cross-sectional data, we contribute to a growing body of literature on QR with correlated errors. Jung (1996) studied a potential efficiency improvement when errors are correlated in a single-equation model. Jun and Pinkse (2009) extended this approach to models with multiple equations, where errors across equations are correlated. They study a cross-sectional regression and focus on the efficiency of the QR estimators. Our work is also closely related to the QR equicorrelated errors model in Wang and He (2007) and Chen, Wei, and Parzen (2004), where a group-specific random effect is viewed as part of the overall error.<sup>3</sup> In this setup, Parente and Santos Silva (2016) proposed a consistent estimator of the covariance matrix when the error terms are correlated within clusters, but independent across clusters. Hagemann (2017) developed a wild bootstrap procedure for cluster-robust inferences in QR models. Our work contributes to this literature by considering a QR panel data model with individual specific FE.

The rest of the paper is organized as follows. Section 2 introduces the FE panel QR model, and considers asymptotically unbiased estimators. Section 3 introduces the cluster covariance matrix, and proposes its estimators. Section 4 studies the asymptotic properties of the CCM estimator and the bias-correction method. Section 5 proposes the cluster robust tests. Section 6 presents a real-data application. Lastly, Section 7 concludes the paper. An Appendix in the Supplemental Material provides all mathematical proofs and numerical simulation results.

## 2. THE ECONOMETRIC MODEL

We consider a FE panel data QR model

$$y_{it} = \alpha_{i0}(\tau) + \mathbf{x}'_{it}\boldsymbol{\theta}_0(\tau) + e_{it}(\tau), \quad i = 1, \dots, N, t = 1, \dots, T, \quad (1)$$

where  $y_{it}$  is the scalar response for the  $i$ -th individual at  $t$ th time period,  $\mathbf{x}_{it}$  is a  $d \times 1$  vector of covariates.  $\alpha_{i0}(\tau)$  is the specific FE for the  $i$ th individual, and the quantile-specific error  $e_{it}(\tau)$  measures the distance between the response  $y_{it}$  and its  $\tau$ th conditional quantile. For stationary processes  $\{(y_{it}, \mathbf{x}_{it}), t \geq 1\}$ , we allow time dependence in a given individual, but assume independence across individuals. We assume that the  $\tau$ th quantile of the error  $e_{it}(\tau)$  is equal to zero. Under this normalization, we rewrite equation (1) as

<sup>3</sup>The work of Wang and He (2007) is highly suitable for a separate problem, different to ours, known as the Moulton problem, a term popularized by the influential papers of Moulton (1986, 1990).

the following conditional quantile model:

$$Q_{y_{it}}(\tau | \mathbf{x}_{it}, \alpha_{i0}(\tau)) = \alpha_{i0}(\tau) + \mathbf{x}'_{it} \boldsymbol{\theta}_0(\tau), \tag{2}$$

where  $Q_{y_{it}}(\tau | \mathbf{x}_{it}, \alpha_{i0}(\tau))$  is the conditional  $\tau$ -quantile of  $y_{it}$ , given  $(\mathbf{x}_{it}, \alpha_{i0}(\tau))$ . We use the vector notation  $\boldsymbol{\alpha}(\tau) = (\alpha_1(\tau), \dots, \alpha_N(\tau))$  to denote a set of FE.

The main contribution of this study is to propose a cluster robust covariance matrix estimator for the parameters of interest in model (2), and to establish its consistency. The estimator is an extension of the HAC covariance estimator for QR panel data, where we correct the bias of the CCM sandwich estimator. Furthermore, we propose Score- and Wald-type tests for general linear hypotheses on the vector  $\boldsymbol{\theta}_0(\tau)$ . It is possible to formulate a wide variety of tests using variants of the Score and Wald tests, from simple tests on a single QR coefficient to joint tests involving many covariates and distinct quantiles.

### 2.1 Estimators for $\boldsymbol{\theta}_0(\tau)$

To implement the proposed methods, the parameters of interest  $\boldsymbol{\theta}_0(\tau)$  in (2) need to be estimated. This section discusses how to obtain estimators, such that

$$\sqrt{NT}(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)) \xrightarrow{d} N(0, \Sigma), \tag{3}$$

where  $\Sigma = \Lambda^{-1} J \Lambda^{-1}$  is the sandwich formula for the variance-covariance matrix. In general,  $\Sigma$ ,  $\Lambda$ , and  $J$  are quantile-specific, but we drop  $\tau$  in the notation to save space.

We employ bias-corrected smoothed FE QR estimators. Galvao and Kato (2016) (GK hereafter) suggested an estimator defined by a minimizer of the QR objective check function that uses kernels to smoothly approximate the nondifferentiable indicator function in the objective function. The FE smoothed quantile regression (FE-SQR) estimator is defined as follows. Let  $K(\cdot)$  be a kernel function and  $G(\cdot)$  be the survival function of  $K(\cdot)$ ;

$$\int_{-\infty}^{\infty} K(u) du = 1, \quad G(u) := \int_u^{\infty} K(v) dv,$$

where  $K(\cdot)$  is not required to be nonnegative (we use higher-order kernels). Let  $\{b_N\}$  be a sequence of positive numbers (bandwidths), such that  $b_N \rightarrow 0$  as  $N \rightarrow \infty$ , and write  $G_{b_N}(\cdot) = G(\cdot/b_N)$ . Note that  $G_{b_N}(y_{it} - \alpha_i - \mathbf{x}'_{it} \boldsymbol{\theta})$  is a smooth approximation of  $I(y_{it} \leq \alpha_i + \mathbf{x}'_{it} \boldsymbol{\theta})$ .<sup>4</sup> The FE-SQR estimator of  $\boldsymbol{\theta}_0(\tau)$ ,  $\widehat{\boldsymbol{\theta}}_S(\tau)$ , is given by

$$\begin{aligned} & (\widehat{\boldsymbol{\alpha}}_S(\tau), \widehat{\boldsymbol{\theta}}_S(\tau)) \\ & := \arg \min_{(\boldsymbol{\alpha}, \boldsymbol{\theta}) \in \mathcal{A}^N \times \mathcal{B}} \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \alpha_i - \mathbf{x}'_{it} \boldsymbol{\theta}) \{ \tau - G_{b_N}(y_{it} - \alpha_i - \mathbf{x}'_{it} \boldsymbol{\theta}) \} \right], \tag{4} \end{aligned}$$

<sup>4</sup>GK regarded  $T = T_N$  as a function of  $N$  and assumed that  $T_N \rightarrow \infty$  as  $N \rightarrow \infty$ ; the bandwidth is chosen as a function of  $N$  and  $T$ , but because  $T$  is a function of  $N$ , it is denoted simply as a function of  $N$ . Because the dependence of  $T$  on  $N$  is arbitrary, this assumption covers the case where  $N$  and  $T$  jointly go to infinity.

where  $\mathcal{A}$  is a compact subset of  $\mathbb{R}$ ,  $\mathcal{A}^N$  is the product of  $N$  copies of  $\mathcal{A}$ , and  $\mathcal{B}$  is a compact subset of  $\mathbb{R}^d$ . The compactness of the set  $\mathcal{A}^N \times \mathcal{B}$  is required to ensure the existence of  $(\widehat{\alpha}(\tau), \widehat{\theta}(\tau))$ .<sup>5</sup>

The FE-SQR estimator in (4) is asymptotically biased when  $N$  and  $T$  grow at the same rate. In particular, the FE-SQR estimator is consistent and has a limiting normal distribution with a bias in the mean. Importantly, by using the smoothing technique, GK are able to derive an analytic expression of the incidental parameter bias. Assuming that  $(N, T) \rightarrow \infty$ , with  $N/T \rightarrow \lambda$  for some  $0 < \lambda < \infty$ , and under some regularity conditions, GK show that

$$\sqrt{NT}(\widehat{\theta}_S(\tau) - \theta_0(\tau)) \xrightarrow{d} N(\sqrt{\lambda}b, \Lambda^{-1}J\Lambda^{-1}), \quad (5)$$

where  $b$  is a factor of the bias term; see GK for the exact expression.<sup>6</sup> Thus, the FE-SQR estimator in (4) has an asymptotic bias term  $\sqrt{\lambda}b$ . Nevertheless, this bias term can be corrected and successfully eliminated using various methods. GK considered two such methods: the analytical method of bias correction, and the half-panel jackknife method (Dhaene and Jochmans (2015)). We examine the same two bias-corrected estimators here.

The first bias-correction method is based on the analytic form of the asymptotic bias. GK identified  $b$  in population terms, and suggested estimating its components using their sample analogs. We refer the reader to GK for details. Define the resulting estimator  $\widehat{b}$ . The bias-corrected estimator is then

$$\widehat{\theta}_{BC}(\tau) := \widehat{\theta}_S(\tau) - \widehat{b}/T. \quad (6)$$

Under some suitable conditions, the bias-corrected estimator,  $\widehat{\theta}_{BC}(\tau)$ , has the limiting normal distribution in (3). Note that the asymptotic variance-covariance matrix of the bias-corrected estimator is equal to that of the FE-SQR estimator in (5).

The second method is the jackknife bias-corrected smoothed panel QR estimator. Assume, for simplicity, that  $T$  is even, and divide the time index  $\{1, \dots, T\}$  into two subsets,  $S_1 = \{1, \dots, T/2\}$  and  $S_2 = \{T/2 + 1, \dots, T\}$ . Let  $\widehat{\theta}_{S_1}(\tau)$  be the estimator given in (4) using data with time index  $t \in S_1$ . Define  $\widehat{\theta}_{S_2}(\tau)$  similarly. Then the half-panel jackknife estimator is

$$\widehat{\theta}_{JK}(\tau) = 2\widehat{\theta}_S(\tau) - \frac{1}{2}(\widehat{\theta}_{S_1}(\tau) + \widehat{\theta}_{S_2}(\tau)). \quad (7)$$

Importantly for our purpose, it is known that  $\widehat{\theta}_{JK}(\tau)$  is approximated by the normal distribution in (3). As in the previous case, the variance-covariance matrix of the half-panel jackknife estimator is the same sandwich formula one obtains with the FE-SQR estimator in (5). In the remainder of the paper, we use either  $\widehat{\theta}_{BC}(\tau)$  or  $\widehat{\theta}_{JK}(\tau)$  to estimate  $\theta_0(\tau)$ .

<sup>5</sup>The FE-SQR estimator works under relatively mild conditions on  $N$  and  $T$ , so is applicable to many empirically relevant cases. However, this advantage comes at the cost of introducing an additional tuning parameter  $b_N$ .

<sup>6</sup>This paper imposes the same assumptions as those in GK, thus the bias  $b$  in equation (5) has the same form as that in GK.

REMARK 2.1. The asymptotic results derived here assume that the length of time,  $T$ , and the number of individuals (groups),  $N$ , grow simultaneously. Hansen (2007) used the same joint asymptotics for the panel mean regression. A more conventional choice is that  $N$  is large, but  $T$  is small (fixed); see Arellano (1987), and Kézdi (2004). Theoretically, we are compelled to use the joint asymptotics because we have only  $T$  observations to estimate each of the FE. Because we are not allowed to difference out the FE, we need to let  $T$  grow along with  $N$ . For a QR, Wang and He (2007) assumed fixed  $T$ , but their model has no individual FE. For an OLS, one can remove FE by data transformation, and thus fix  $T$ .

### 3. THE CLUSTER ROBUST COVARIANCE MATRIX

This section discusses the CCM for the panel FE QR model and its estimation. The statistical inference on  $\theta_0(\tau)$  may be implemented by estimating  $\Sigma = \Lambda^{-1}J\Lambda^{-1}$ . With suitable consistent estimators  $\widehat{\Lambda}$  and  $\widehat{J}$ , such that  $\widehat{\Lambda} \xrightarrow{P} \Lambda$  and  $\widehat{J} \xrightarrow{P} J$ , the CCM estimator is given by

$$\widehat{\Sigma} = \widehat{\Lambda}^{-1}\widehat{J}\widehat{\Lambda}^{-1},$$

and is consistent,  $\widehat{\Sigma} \xrightarrow{P} \Sigma$ .

#### 3.1 Clustered covariance matrix

Let  $f_i(e|\mathbf{x})$  be the conditional density of  $e_{it}$ , given  $\mathbf{x}_{it} = \mathbf{x}$ , and let  $f_i(e) = E[f_i(e|\mathbf{x}_{i1})]$  be the marginal density of  $e_{it}$ . Define  $\mathbf{c}_i = E[f_i(0|\mathbf{x}_{i1})\mathbf{x}_{i1}]/f_i(0)$  and<sup>7</sup>

$$\Lambda := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E[f_i(0|\mathbf{x}_{i1})\mathbf{x}_{i1}(\mathbf{x}_{i1} - \mathbf{c}_i)'],$$

$$J := \lim_{N, T \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N J_{T,i},$$

where  $J_{T,i}$  denotes the covariance matrix of  $\frac{1}{\sqrt{T}} \sum_{t=1}^T (\tau - I(e_{it}(\tau) \leq 0))(\mathbf{x}_{it} - \mathbf{c}_i)$ . The components in  $\Lambda$  and  $J$  are allowed to vary across  $i$ , as in Hahn and Kuersteiner (2011), to accommodate cross-sectional heterogeneity. The crucial term in the covariance matrix  $J_{T,i}$  is the autocovariance of  $h_{it} = \tau - I(e_{it}(\tau) \leq 0)$  between periods  $t$  and  $s$ :

$$\varrho_{its} := E[(\tau - I(e_{it}(\tau) \leq 0))(\tau - I(e_{is}(\tau) \leq 0))|\mathbf{x}_{it}, \mathbf{x}_{is}, \alpha_i].$$

In the absence of temporal dependence,  $\varrho_{its}$  is zero, and  $J$  becomes the variance matrix for *iid* data. This is because, given the conditioning variables,  $\varrho_{its}$  is  $P(e_{it}(\tau) \leq 0, e_{is}(\tau) \leq 0) - \tau^2$ . Without temporal dependence, for  $t \neq s$ ,  $\varrho_{its} = P(e_{it}(\tau) \leq 0) \cdot P(e_{is}(\tau) \leq 0) - \tau^2 =$

<sup>7</sup>We use the notation  $f_i(0)$  and  $f_i(0|\mathbf{x}_{i1})$  for simplicity to refer to the density functions evaluated at the true quantile as  $f_i(Q_y(\tau|\mathbf{x}_{i1}))$  and  $f_i(Q_y(\tau|\mathbf{x}_{i1})|\mathbf{x}_{i1})$ , respectively.

$\tau \cdot \tau - \tau^2 = 0$ . Under serial dependence,  $\varrho_{its}$  may not be equal to zero, and thus, may contribute to the variance matrix of the estimator.

Let  $j = t - s$  denote the time lag. Then  $J_T = \frac{1}{N} \sum_{i=1}^N J_{T,i}$  can be written as

$$J_T = \sum_{j=-T+1}^{T-1} \Gamma_j, \quad \text{where} \tag{8}$$

$$\Gamma_j = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-j} E[(\tau - I(e_{it}(\tau) \leq 0))(\tau - I(e_{it+j}(\tau) \leq 0))(\mathbf{x}_{it} - \mathbf{c}_i)(\mathbf{x}_{it+j} - \mathbf{c}_i)'],$$

for  $j \geq 0$ , and  $\Gamma_{-j} = \Gamma_j'$ . The  $j$ th order autocovariance matrix  $\Gamma_j$  reduces to the variance matrix when  $j = 0$ ; that is,  $\Gamma_0 = N^{-1} \sum_{i=1}^N \tau(1 - \tau)E[(\mathbf{x}_{i1} - \mathbf{c}_i)(\mathbf{x}_{i1} - \mathbf{c}_i)']$ , with  $J$  and  $\Lambda$  given above. The resulting covariance matrix  $\Sigma$  allows for the heteroskedasticity and serial correlation of arbitrary form.

Note that in QR models, the two terms in the sandwich formula have distinct roles;  $\Lambda$  accounts for the heteroskedasticity, and  $J$  manages the serial correlation. The term  $\Lambda$  includes conditional densities that allow the conditional heteroskedasticity. The term  $J$  includes  $\varrho_{ij} = \varrho_{its}$ , for  $j = t - s$ , which measures the degree of serial correlation between periods in the  $i$ th group. In contrast, in an OLS, all actions take place in the middle term, corresponding to  $J$  here.

It is also interesting to note how the CCM works in a QR. The sign of  $e_{it}(\tau)$  is critical. Consider, for instance, a location-scale model with errors  $e_{it}(\tau) = \sigma(\mathbf{x}'_{it} \delta_0)u_{it}(\tau)$ , where  $\sigma(\cdot)$  is a strictly positive function of a linear index  $\mathbf{x}'_{it} \delta_0$ , and  $u_{it}(\tau)$  is the common error, with the  $\tau$ th quantile equal to zero. The overall error is defined as the product of  $\sigma(\cdot)$  and  $u_{it}(\tau)$ . Because  $\sigma(\cdot)$  does not alter the sign of the overall error, the structure of  $J$  is determined solely by the sign of  $u_{it}(\tau)$ . However, for  $\Lambda$ , the conditional heteroskedasticity induced by  $\sigma(\cdot)$  is still in effect. Thus, it is possible to have a model with heteroskedastic errors, but with a time dependence structure that is identical over individuals.

This consideration motivates one important special case of the general CCM. If  $\varrho_{ij} = \varrho_j$  for all  $i$ , then  $J_T$  simplifies to  $J_T^s = \sum_{j=-T+1}^{T-1} \Gamma_j^s$ , where

$$\Gamma_j^s = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-j} \varrho_j E[(\mathbf{x}_{it} - \mathbf{c}_i)(\mathbf{x}_{it+j} - \mathbf{c}_i)'], \tag{9}$$

for  $j \geq 0$ , and  $\Gamma_{-j}^s = (\Gamma_j^s)'$ . In this case, the autocorrelation structure can be as flexible as possible over time, but remains identical over individuals. This alternative, simplified covariance matrix is useful for the Score test below.

The estimation of  $\Lambda$  in the CCM,  $\Sigma$ , is relatively straightforward. The Powell (1991) estimator for  $\Lambda$  is known to be consistent and asymptotically normal under weakly dependent data; see Kato (2012) for details. The estimation of the term  $J$  under dependent data is more involved; therefore, we concentrate on that here.



### 3.2 Estimating the CCM

Let  $\widehat{e}_{it}(\tau) = y_{it} - \widehat{\alpha}_i(\tau) - \mathbf{x}'_{it}\widehat{\boldsymbol{\theta}}(\tau)$  be the residual from model (1), using one of the bias-corrected estimators in Section 2. The Powell estimator for  $\Lambda$  is

$$\widehat{\Lambda} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \frac{1}{2b_1} I(|\widehat{e}_{it}(\tau)| \leq b_1) \mathbf{x}_{it}(\mathbf{x}_{it} - \widehat{\mathbf{c}}_i)', \tag{10}$$

where  $b_1$  is a bandwidth. This estimator is valid under the heteroskedasticity of general form. To estimate  $\mathbf{c}_i$ , let  $K_2(\cdot)$  and  $b_2$  be a kernel function and another bandwidth sequence, respectively. Then  $\widehat{\mathbf{c}}_i = \frac{1}{T} \sum_{t=1}^T \frac{1}{b_2} K_2(\frac{\widehat{e}_{it}-0}{b_2}) \mathbf{x}_{it} / \widehat{f}_i(0)$ , and  $\widehat{f}_i(0) = \frac{1}{T} \sum_{t=1}^T \frac{1}{b_2} K_2(\frac{\widehat{e}_{it}-0}{b_2})$ . The conditions on  $K_2$ ,  $b_1$ , and  $b_2$  can be found in Section 4.1.

Consider  $J$  in the CCM estimator. From (8),  $J = \sum_{j=-\infty}^{\infty} \Gamma_j$  is a long-run variance matrix. As is well known in the time series literature, simply adding autocovariance estimates does not lead to a consistent estimator. The estimation uncertainty from the higher-order autocovariances must be weighed down. Following Andrews (1991) and Newey and West (1994), we consider a class of kernel estimators of the form

$$\widehat{J}_T = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right) \frac{T-|j|}{T} \widehat{\Gamma}_j, \tag{11}$$

where  $k(\cdot)$  is a kernel function, and  $m_T$  is a sequence of increasing numbers known as the lag truncation or bandwidth parameter, and  $\widehat{\Gamma}_j$  is a consistent estimator of  $\Gamma_j$ .

Assume that kernel functions belong to the class:

$$\mathcal{K}_1 = \left\{ k(\cdot) : R \rightarrow [-1, 1] \mid k(0) = 1, k(-z) = k(z), \forall z \in R, \int_{-\infty}^{\infty} k^2(z) dz < \infty, k(\cdot) \text{ is continuous almost everywhere and at } 0 \right\}.$$

The class of kernels  $\mathcal{K}_1$  contains the Bartlett, Parzen, and quadratic spectrum (QS) kernels. They generate positive semidefinite estimators in finite sample. The QS kernel is an example of a kernel that gives nonzero weights for all lags  $j = 1, \dots, T - 1$ .

For the kernel smoothing estimator, the choice of the bandwidth is known to be important. We propose an optimal bandwidth  $m_T^*$  that minimizes the asymptotic MSE of an estimator of  $J_T$ . When calculating the MSE, existing arguments in Andrews (1991) or in Newey and West (1994), are not directly applicable, because they depend crucially on the differentiability of the moment function. We develop an alternative approach utilizing the methods in Linton and Whang (2007) and Okui (2010). To the best of our knowledge, this is the first time the optimal bandwidth selection rule has been established for a long-run variance estimation in the QR literature.

The computation of (11) requires a consistent estimator of  $\Gamma_j$ . We use a sample splitting device.<sup>8</sup> We use half the cross-sectional units to estimate  $\boldsymbol{\theta}(\tau)$ , and use the remain-

<sup>8</sup>Technically, the sample splitting allows us to extend the cross-sectional independence of  $e_{it}(\tau)$  over  $1 \leq i \leq N$  to that of  $\widehat{e}_{it}(\tau)$ . See Section S.2 in the Supplemental Appendix for further details.

ing half to estimate the  $j$ th autocovariance. See [Bickel \(1982\)](#) and [Schick \(1986\)](#) for applications. Let  $[z]$  denote the integer part of  $z$ . Define  $N_1 = [N/2]$  and  $N_2 = N - N_1$ , and define sets  $I_1 = \{1, \dots, N_1\}$  and  $I_2 = \{N_1 + 1, \dots, N\}$ . Let  $\hat{\theta}_1(\tau)$  and  $\hat{\theta}_2(\tau)$  be the regression coefficient estimates using subsamples  $I_1$  and  $I_2$ , respectively. For  $i \in I_1$ , let  $\hat{e}_{2,it}(\tau) = y_{it} - \hat{\alpha}_i(\tau) - \mathbf{x}'_{it}\hat{\theta}_2(\tau)$ , and define

$$\hat{\Gamma}_{j,1} = \frac{1}{N_1(T-j)} \sum_{i=1}^{N_1} \sum_{t=1}^{T-j} (\tau - I(\hat{e}_{2,it}(\tau) \leq 0))(\tau - I(\hat{e}_{2,it+j}(\tau) \leq 0))(\mathbf{x}_{it} - \hat{\mathbf{c}}_{2,i})(\mathbf{x}_{it+j} - \hat{\mathbf{c}}_{2,i})'.$$

The sample splitting is applied to the construction of  $\hat{\mathbf{c}}_i$  as well. For  $i \in I_1$ ,  $\hat{\mathbf{c}}_{2,i}$  uses  $\hat{e}_{2,it}(\tau)$ . Similarly, for  $i \in I_2$ , let  $\hat{e}_{1,it}(\tau) = y_{it} - \hat{\alpha}_i(\tau) - \mathbf{x}'_{it}\hat{\theta}_1(\tau)$ , and define

$$\begin{aligned} \hat{\Gamma}_{j,2} &= \frac{1}{N_2(T-j)} \\ &\times \sum_{i=N_1+1}^N \sum_{t=1}^{T-j} (\tau - I(\hat{e}_{1,it}(\tau) \leq 0))(\tau - I(\hat{e}_{1,it+j}(\tau) \leq 0))(\mathbf{x}_{it} - \hat{\mathbf{c}}_{1,i})(\mathbf{x}_{it+j} - \hat{\mathbf{c}}_{1,i})'. \end{aligned}$$

The consistent  $j$ th order autocovariance matrix estimate is then given by

$$\hat{\Gamma}_j = \frac{N_1}{N} \cdot \hat{\Gamma}_{j,1} + \frac{N_2}{N} \cdot \hat{\Gamma}_{j,2}. \tag{12}$$

The regression slope estimate used for inference is  $\hat{\theta}(\tau) = N_1/N \cdot \hat{\theta}_1(\tau) + N_2/N \cdot \hat{\theta}_2(\tau)$ . Observe that we use the entire sample to construct  $\hat{\Gamma}_j$  and  $\hat{\theta}(\tau)$ , unlike the name ‘‘sample splitting’’ may suggest. The scalar autocovariance (without  $\mathbf{x}_{it} - \mathbf{c}_i$ ) can be estimated by

$$\hat{\varrho}_j = \frac{N_1}{N} \hat{\varrho}_{j,1} + \frac{N_2}{N} \hat{\varrho}_{j,2}, \tag{13}$$

where  $\hat{\varrho}_{j,k}$  ( $k = 1, 2$ ) is simply  $\hat{\Gamma}_{j,k}$  without the  $(\mathbf{x}_{it} - \hat{\mathbf{c}}_{k,i})(\mathbf{x}_{it+j} - \hat{\mathbf{c}}_{k,i})'$  term. To study the asymptotic properties of  $\hat{J}_T$ , it is useful to consider a pseudo-estimator that is identical to  $\hat{J}_T$ , but that uses true errors, as follows:

$$\begin{aligned} \tilde{J}_T &= \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m}\right) \frac{T-|j|}{T} \tilde{\Gamma}_j, \quad \text{where} \\ \tilde{\Gamma}_j &= \frac{1}{N(T-j)} \sum_{i=1}^N \sum_{t=1}^{T-j} (\tau - I(e_{it} \leq 0))(\tau - I(e_{it+j} \leq 0))(\mathbf{x}_{it} - \mathbf{c}_i)(\mathbf{x}_{it+j} - \mathbf{c}_i)'. \end{aligned}$$

There is one remaining complication. As in the standard FE conditional mean models, it turns out that the autocovariance for the QR estimator,  $\hat{\Gamma}_j$ , is asymptotically biased, with the bias of order  $1/T$ . We formally establish this result in the next section. This implies that the long-run variance estimator  $\hat{J}_T$  will be asymptotically biased as well. This is owing to the incidental parameter problem (see the following remarks). Given that

the regression coefficient estimator  $\widehat{\boldsymbol{\theta}}(\tau)$  has order  $1/\sqrt{NT}$ , the bias of size  $1/T$  cannot be regarded as small if  $N$  and  $T$  have the same order. For linear panel data models, this bias property of the autocovariance estimator has been the subject of careful analysis by Okui (2010). The asymptotic bias in the estimator negatively affects finite-sample performance of the CCM estimator. The bias correction in Section 4 significantly improves the finite-sample performance.

REMARK 3.1. The asymptotic bias exists because of the difficulty in estimating  $\alpha_i$  precisely. If  $T$  is finite, it would be impossible to estimate  $\alpha_i$  consistently; thus,  $\widehat{\Gamma}_j$  would be inconsistent. If one assumes that  $T \rightarrow \infty$ , then  $\widehat{\Gamma}_j$  will be consistent, because its bias is of order  $T^{-1}$ . However, this does not mean that the bias will be small in a finite sample.

REMARK 3.2. Hansen (2007) does not require truncation or kernel smoothing to show the consistency of the CCM for the OLS case. The difference between this and the QR case is the result of an interesting property of Arellano’s variance matrix estimator considered in Hansen (2007); that is, its consistency is not affected by the consistent estimation of the FE; see Remark 9 in Stock and Watson (2008) for further details.

#### 4. ASYMPTOTIC PROPERTIES OF THE CCM ESTIMATOR

This section derives the asymptotic properties of the CCM estimator, develops an alternative bias-corrected estimator, and establishes its limiting properties. We first introduce our assumptions and the notation in Section 4.1. Section 4.2 derives the asymptotic bias expression for the covariance matrix estimator  $\widehat{\Gamma}_j$ , and the asymptotic MSE of the long-run variance matrix estimator  $\widehat{J}_T$ . Section 4.3 develops a bias-correction method. Section 4.4 presents the bias-corrected long-run variance estimator  $\check{J}_T$ , and the resulting CCM estimator. Section 4.5 develops a data-driven bandwidth selection rule.

##### 4.1 Assumptions

This section provides a list of conditions under which the estimators for  $\boldsymbol{\theta}(\tau)$ ,  $\Gamma_j$ , and  $J_T$  have desirable properties.

(A1)  $\{(y_{it}, \mathbf{x}_{it}), t \geq 1\}$  is stationary and  $\beta$ -mixing for each fixed  $i$ , and independent across  $i$ . Let  $\beta_i(j)$  denote the  $\beta$ -mixing coefficients of  $\{(y_{it}, \mathbf{x}_{it}), t \geq 1\}$ . There exist constants  $a \in (0, 1)$  and  $B > 0$ , such that  $\sup_{i \geq 1} \beta_i(j) \leq Ba^j$ , for all  $j \geq 1$ .

(A2) There exists a constant  $M > 0$ , such that  $\sup_{i \geq 1, t \geq 1} \|\mathbf{x}_{it}\| \leq M$  (a.s.).

(A3) For each  $\delta > 0$ ,

$$\epsilon_\delta := \inf_{i \geq 1} \inf_{\|\boldsymbol{\theta}\|_1 = \delta} \mathbb{E} \left[ \int_0^{\alpha + \mathbf{x}'_{it} \boldsymbol{\theta}} \{F_i(s|\mathbf{x}_{i1}) - \tau\} ds \right] > 0, \tag{14}$$

where  $\|\cdot\|_1$  denotes the  $\ell_1$  norm.<sup>9</sup>

(A4) Matrixes  $\Lambda$  and  $J$  exist and are nonsingular.

<sup>9</sup>There is no significant role in the  $\ell_1$  norm, because any norm on a fixed dimensional Euclidean space is equivalent. The  $\ell_1$  norm is used simply to avoid notation such as  $\|(\alpha_i - \alpha_{i0}, \boldsymbol{\theta}' - \boldsymbol{\theta}'_0)'\|$ .

Assumption (A1) is a mixing condition. KGM used the same  $\beta$ -mixing condition, whereas Hahn and Kuersteiner (2011) assume  $\alpha$ -mixing.<sup>10</sup> We assume exponentially decaying mixing coefficients. The same condition was used in KGM and in Hahn and Kuersteiner (2011). Assumption (A2), taken from Koenker (2004), imposes a restriction on the covariates. Without FE, a usual assumption with dependent data is  $E\|x_{it}\|^{2+\delta} < \infty$ , for some  $\delta > 0$ , uniformly in  $i$  and  $t$  (see Koenker and Zhao (1996)). Thus, (A2) is rather strict. It is used to ensure the asymptotic expansion of the first-order conditions. See the proof of Proposition 3.1 in GK and Theorem 5.1 in the Appendix of the Supplemental Material. Assumption (A3) is an identification assumption of QR coefficients. See KGM for further descriptions. Assumption (A4) ensures that the CCM is well-defined.

Correct model specification in equation (2) and assumption (A1) constitute the model we consider here. The conditional quantile zero restriction in (2) concerns only contemporaneous covariate values. Thus, the model allows  $x_{it}$  to be predetermined with respect to time-varying errors. We do not require strict exogeneity, but do use large panel data asymptotics.<sup>11</sup>

Recall that  $f_i(e|\mathbf{x})$  denotes the conditional density of  $e_{it}$ , given  $\mathbf{x}_{it} = \mathbf{x}$ , and  $f_i(e)$  denotes the marginal density of  $e_{it}$ . Let  $f_{i,j}(e_1, e_{1+j}) = f_{ij}(e_1, e_{1+j}|\mathbf{x}_1, \mathbf{x}_{1+j})$  denote a joint density of  $(e_{i1}, e_{i1+j})$ , given  $(\mathbf{x}_{i1}, \mathbf{x}_{i1+j}) = (\mathbf{x}_1, \mathbf{x}_{1+j})$ . Define  $f_i^{(k)}(e|\mathbf{x}) = (\partial/\partial e)^k f_i(e|\mathbf{x})$ , for  $k = 0, 1, \dots, r$ , where  $f_i^{(0)}(e|\mathbf{x})$  means  $f_i(e|\mathbf{x})$ . In addition, define  $f_{i,j}^{(k,l)}(e_{i1}, e_{i1+j}) := \partial^{k+l} f_{i,j}(e_{i1}, e_{i1+j})/\partial e_{i1}^k \partial e_{i1+j}^l$ , where  $f_{i,j}^{(0,0)}(e_{i1}, e_{i1+j})$  stands for  $f_{i,j}(e_1, e_{1+j})$ .

(A5) Assume the following: (a) Let  $r \geq 4$  be an even integer; then,  $f_i(e)$  is  $r$ -times continuously differentiable. (b) For all  $k = 0, 1, \dots, r$ ,  $|f_i^{(k)}(e|\mathbf{x})| \leq C_f$  uniformly over  $e$  and  $i$ , and  $c_f \leq f_i(0|\mathbf{x})$ , for all  $i$ . (c)  $f_{i,j}(e_{i1}, e_{i1+j})$  is  $r$ -times continuously differentiable with respect to  $(e_{i1}, e_{i1+j})$ . (d) There exist a constant  $C'_f > 0$  and an integrable function  $D_f$ , such that  $|f_{i,j}^{(k,0)}(e_{i1}, e_{i1+j})| \leq C'_f$  and  $|f_{i,j}^{(0,l)}(e_{i1}, e_{i1+j})| \leq D_f(e_{i1+j})$  uniformly over  $(e_{i1}, e_{i1+j})$  and  $i$ , for all  $k = 0, 1, \dots, r$  and  $l = 0, 1, \dots, r$ .

Assumptions (A5)(a)–(b) are restrictions on the marginal and conditional densities. They are stronger than those common in the literature, which require that the given conditions hold for  $r = 1$ . We need stronger conditions because the bias correction requires higher-order asymptotics. Assumptions (A5)(c)–(d) are restrictions on a joint error density. They are used to obtain a useful upper bound for covariances across error terms. For kernel functions and bandwidths, we have the following conditions:

(A6) (a) The kernel function  $K(\cdot)$  for the smoothed QR is symmetric and three-times differentiable, and has bounded support. It is an  $r$ th order kernel, with  $r$  as defined in (A5), satisfying

$$\int_{-\infty}^{\infty} K(u) du = 1, \quad \int_{-\infty}^{\infty} u^j K(u) du = 0, \quad j = 1, \dots, r - 1, \quad \int_{-\infty}^{\infty} u^r K(u) du \neq 0.$$

<sup>10</sup>See Bradley (2005) for the definitions of several strong mixing conditions and their relationship.

<sup>11</sup>In standard conditional mean models, if the strict exogeneity assumption is violated, then the FE estimator is not fixed- $T$  consistent, however, as  $T \rightarrow \infty$  with  $N$ , the FE estimator becomes consistent for the coefficient vector on  $x_{it}$ , provided that the specification is correct; see, for example, Hahn and Kuersteiner (2002).

- (b) The bandwidth for the smoothed QR has order  $b_N = O(T^{-c})$ , where  $1/r < c < 1/3$ .
- (c) The kernel function  $K_2(\cdot)$  is continuous and bounded, and of bounded variation.
- (d) The bandwidth sequences  $b_j, j = 1, 2$ , satisfy that  $b_j \rightarrow 0$  and  $(\log N)/(Tb_j) \rightarrow 0$ .

Assumptions (A6)(a)–(b) are conditions needed to obtain the smoothed FE-QR estimator. We use them for the bias-corrected regression coefficient estimators. Assumption (c) is a mild condition on the kernel function, satisfied by a Gaussian or an Epanechnikov kernel, and (d) is a common condition on bandwidths.

For the rest of the paper, we assume there exist regression coefficient estimates satisfying  $\hat{\alpha}_i(\tau) - \alpha_{i0}(\tau) = O_p(T^{-1/2})$ , for each  $1 \leq i \leq N$  and  $\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau) = O_p((NT)^{-1/2})$ . Under conditions (A1)–(A6), the two estimators discussed in Section 2 are known to satisfy these conditions. Thus far, conditions (A1)–(A6), taken from GK, are used to construct asymptotically unbiased regression coefficient estimates. For the variance matrix estimation, some additional conditions are necessary.

Two types of autocovariances are important for the CCM estimator, corresponding to two moment functions. The first is related to

$$h_{it}(\alpha_i, \boldsymbol{\theta}) = \tau - I(e_{it} \leq \alpha_i - \alpha_{i0} + \mathbf{x}'_{it}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)).$$

Use  $h_{it}$  as a short notation for  $h_{it}(\alpha_{i0}, \boldsymbol{\theta}_0) = (\tau - I(e_{it} \leq 0))$ . The  $j$ th order autocovariance of  $h_{it}$  is  $q_{ij} = E[h_{it}h_{it+j}] = E[(\tau - I(e_{it} \leq 0))(\tau - I(e_{it+j} \leq 0))]$ , and the variance of  $T^{-1} \sum_{t=1}^T h_{it}$  (for the  $i$ -th group) is

$$V_{T,i} = q_{i0} + 2 \sum_{j=1}^{T-1} \frac{T-j}{T} q_{ij}.$$

Let  $V_T = N^{-1} \sum_{i=1}^N V_{T,i}$ . The quantity  $V_T$  is equal to  $J_T$ , without  $(\mathbf{x}_{it} - \mathbf{c}_i)(\mathbf{x}_{it+j} - \mathbf{c}_i)'$ . The second moment function concerns

$$H_{it}(\alpha_i, \boldsymbol{\theta}) = h_{it}(\alpha_i, \boldsymbol{\theta})(\mathbf{x}_{it} - \mathbf{c}_i).$$

Let  $\tilde{\mathbf{x}}_{it} = (\mathbf{x}_{it} - \mathbf{c}_i)$  and  $H_{it}$  denote  $H_{it}(\alpha_{i0}, \boldsymbol{\theta}_0)$ . With this notation,  $\Gamma_{ij} = E[H_{it}H_{it+j}]$  is the  $j$ th order autocovariance for  $H_{it}$ , and  $\Gamma_j = N^{-1} \sum_{i=1}^N \Gamma_{ij}$ . Let  $J_{T,i}$  be the variance of  $T^{-1} \sum_{t=1}^T H_{it}$ ; then,  $J_T = N^{-1} \sum_{i=1}^N J_{T,i}$ . We have the following conditions on  $h_{it}$  and  $H_{it}$ . Let  $\|\cdot\|$  denote the Euclidean norm of a vector or matrix.

(A7) For each  $1 \leq i \leq N$ ,  $\sum_{j=-\infty}^{\infty} |q_{ij}| < \infty$  and  $\sum_{j=-\infty}^{\infty} \|\Gamma_{ij}\| < \infty$ .

This assumption restricts the autocovariances to be absolutely summable. It is also required that  $E[h_{it}h_{ik}h_{il}h_{im}] < C$  and  $E[H_{it}H_{ik}H_{il}H_{im}] < C$ , for a constant  $C < \infty$  and for any  $t, k, l, m$ . From (A2) and  $|h_{it}| \leq 1$ , the finite fourth moment condition is satisfied. Another crucial condition in this regard is the summability of the cumulants. Following Andrews (1991), define  $\kappa^i(t, t + j, t + l, t + m)$  as the fourth-order cumulant of  $(h_{it}, h_{it+j}, h_{it+l}, h_{it+m})$ . Similarly, let  $\kappa^i_{abcg}(t, t + j, t + l, t + m)$  denote the fourth-order cumulant of  $(H_{a,it}, H_{b,it+j}, H_{c,it+l}, H_{g,it+m})$ , where  $H_{a,it}$  denotes the  $a$ th element of  $H_{it}$ .

(A8) For each  $1 \leq i \leq N$ ,  $h_{it}$  and  $H_{it}$  are mean zero, fourth-order stationary sequences of random variables, and  $\sum_{j,l,m=-\infty}^{\infty} \kappa^i(t, t+j, t+l, t+m) < \infty$  and  $\sum_{j,l,m=-\infty}^{\infty} \kappa_{abcg}^i(t, t+j, t+l, t+m) < \infty$ , for all  $a, b, c, g \in \{1, 2, \dots, d\}$ .

Hagemann (2013) showed that the geometric moment contracting property, a weaker condition than the summability of cumulants, would be sufficient to achieve the same goal. This direction is promising but is left to future research.

#### 4.2 Asymptotic properties of estimators for $\Gamma_j$ and $J_T$

As is the common regression coefficient estimate  $\hat{\theta}(\tau)$ , the autocovariance estimator  $\hat{\Gamma}_j$  is asymptotically biased. The bias is of order  $1/T$ . The next theorem states the result.

THEOREM 4.1. Assume conditions (A1)–(A6) hold. Fix  $j$ . As  $N, T \rightarrow \infty$ ,

$$\begin{aligned} & \sqrt{NT} \left\{ \hat{\Gamma}_j - \Gamma_j + \frac{B_j}{T} \right\} \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^{T-j} \{ (\tau - I(e_{it} \leq 0)) (\tau - I(e_{it+j} \leq 0)) \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}'_{it+j} - \Gamma_j \} + o_p(1), \end{aligned} \quad (15)$$

where

$$\begin{aligned} B_j &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \omega_{ij} \cdot V_{T,i} \quad \text{and} \\ \omega_{ij} &= \mathbb{E} \left[ \left( 2 \frac{f_i(0|x_{i1})}{f_i(0)} - \frac{f_{ij}(0, 0|x_{i1}, x_{i1+j})}{f_i(0)^2} \right) \tilde{\mathbf{x}}_{i1} \tilde{\mathbf{x}}'_{i1+j} \right]. \end{aligned}$$

In particular, if  $N/T \rightarrow \lambda$ , for a constant  $0 < \lambda < \infty$ , then

$$\hat{\Gamma}_j - \tilde{\Gamma}_j = -\frac{B_j}{T} + o_p(T^{-1}).$$

It is shown in the Supplemental Appendix that  $\hat{\Gamma}_j - \tilde{\Gamma}_j = -B_j/T + o_p((NT)^{-1/2})$ . This implies the first part of Theorem 4.1. The second part states that when  $N$  and  $T$  are comparable in size, the remainder terms in the bias become of order  $o_p(T^{-1})$ , and can be ignored in the subsequent analysis. In fact, the necessary condition for simplification is that  $N^2/T \rightarrow \infty$ , which holds under the assumption  $N/T \rightarrow \lambda$ . To keep the presentation simple, we use this implication throughout the remainder of the paper.

In the special case where  $e_{it}$  and  $e_{it+j}$  are independent and the model does not include any covariates  $x_{it}$ , the asymptotic bias simplifies to  $-V_T/T$ . This is because  $f_i(0|x_{i1}) = f_i(0)$  and  $f_{ij}(0, 0|x_{i1}, x_{i1+j}) = f_i^2(0)$ , which make  $\omega_{ij} = 1$ . Interestingly, the bias in this special case does not depend on  $j$ . This is the quantile analogue of the bias expression in Okui (2010), who studied a linear panel model without covariates. However, in general, unlike Okui (2010), the bias in the autocovariance estimator for a QR depends on  $j$ .

The sign of the bias  $B_j$  can be positive or negative, in general. Of the two terms that constitute the bias,  $V_{T,i}$  is the long-run variance of  $h_{it}$ , and so is always positive. The sign of the bias depends on  $\omega_{ij}$ . To motivate the bias correction in the autocovariance estimator, we consider the special case where  $(e_{it}, e_{it+j})$  follow a bivariate normal distribution, with unit variances and correlation coefficient  $\rho$ , and the model has no covariates. Then  $\omega_{ij}$  reduces to  $2 - \phi_j(0, 0)/\phi(0)^2$ , where  $\phi_j(\cdot, \cdot)$  denotes a bivariate normal density with mean  $(-\Phi^{-1}(\tau), -\Phi^{-1}(\tau))$ , variances  $(1, 1)$ , and correlation coefficient  $\rho$ , and  $\Phi^{-1}(\cdot)$  is the quantile function of the standard normal random variable.<sup>12</sup> In the special case of  $\rho = 0$ ,  $\omega_{ij}$  is equal to one. As  $\rho$  increases toward one,  $\omega_{ij}$  stays positive but reduces to zero. The sign of  $\omega_{ij}$  eventually becomes negative. When  $\tau = 0.5$ , the turning point comes when  $\delta = 0.867$ , after which  $\omega_{ij}$  becomes negative.<sup>13</sup> In this example, unless the autocorrelation coefficient is not too close to 1,  $\omega_{ij}$  will be positive, which means that the bias will be downward; that is,  $\widehat{\Gamma}_j < \Gamma_j$ . Hence, without the bias correction, the bias is not properly adjusted, and the resulting cluster robust variance estimator will underestimate the true uncertainty in the regression coefficient estimates.

To see how the bias correction works, suppose for a moment that  $B_j$  can be consistently estimated by  $\widehat{B}_j$ . Then it is possible to define a bias-corrected autocovariance estimator as

$$\widehat{\Gamma}_j^{(1)} = \widehat{\Gamma}_j + \frac{1}{T} \widehat{B}_j. \tag{16}$$

The following proposition provides an asymptotic representation for  $\widehat{\Gamma}_j^{(1)}$ .

**PROPOSITION 4.1.** *Suppose that  $N/T \rightarrow \lambda$  and  $\widehat{B}_j = B_j + o_p(1)$ . Then*

$$\begin{aligned} & \sqrt{NT} \{ \widehat{\Gamma}_j^{(1)} - \Gamma_j \} \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^{T-j} \{ (\tau - I(e_{it} \leq 0)) (\tau - I(e_{it+j} \leq 0)) \tilde{x}_{it} \tilde{x}'_{it+j} - \Gamma_j \} + o_p(1). \end{aligned} \tag{17}$$

The crucial step in constructing a feasible bias correction is to estimate  $B_j$ . This is the main topic in Section 4.3.

**REMARK 4.1.** The bias in the  $j$ th autocovariance matrix estimator  $\widehat{\Gamma}_j$  has order  $1/T$ . It is plausible to expect that the effect of the bias correction will be larger for a relatively short panel. However, the same reasoning may not be true for the long-run variance estimator  $\widehat{J}_T$ , because it is a weighted sum of many autocovariances over  $j$ .

We now study the asymptotic properties of  $\widehat{J}_T$  in (11), the bias-uncorrected middle term in the CCM estimator. We examine the bias-corrected estimator in the following sections. Following Parzen (1957), define  $k_{\bar{q}} = \lim_{x \rightarrow 0} (1 - k(x))/|x|^{\bar{q}}$ , for  $0 \leq \bar{q} < \infty$ . A larger value of  $\bar{q}$  for which  $k_{\bar{q}}$  is finite means that the kernel is smoother at zero. Let  $q$

<sup>12</sup>Here,  $\phi(\cdot)$  denotes a univariate normal density with mean  $-\Phi^{-1}(\tau)$  and variance one.

<sup>13</sup>The value of  $\omega_{ij}$  is quantile dependent. When  $\tau = 0.25$ , the turning points becomes  $\delta = 0.792$ .

be the largest, positive  $\bar{q}$ , such that  $0 < k_{\bar{q}} < \infty$ . For the truncated kernel,  $q = 0$  because  $k_{\bar{q}} = 0$  for all  $\bar{q} < \infty$ . For the Bartlett kernel,  $q = 1$  and  $k_1 = 1$ , and for the QS kernel,  $q = 2$  and  $k_2 = 1.4212$ . Define  $J^{(q)} = \sum_{j=-\infty}^{\infty} |j|^q \Gamma_j$ .

Let  $W$  be a  $d^2 \times d^2$  weighting matrix, and  $\text{vec}(\cdot)$  be the column by column vectorization function. Let  $\text{tr}(\cdot)$  denote the trace function, and  $\otimes$  be the Kronecker product operator. Let  $K_{dd}$  denote the  $d^2 \times d^2$  communication matrix that transforms  $\text{vec}(A)$  into  $\text{vec}(A')$ , as in Andrews (1991). That is,  $K_{dd} = \sum_{i=1}^d \sum_{j=1}^d e_i e_j' \otimes e_j e_i'$ , where  $e_i$  is the  $i$ -th elementary  $d \times 1$  vector. Define  $\text{MSE}(\hat{J}_T) = E[\text{vec}(\hat{J}_T - J_T)' W \text{vec}(\hat{J}_T - J_T)]$ . Consider kernels that allows a nonzero  $k_q$  for  $0 < q < \infty$ . These include the Bartlett kernel in Newey and West (1994) and the QS kernel in Andrews (1991).

**THEOREM 4.2.** *Assume conditions (A1)–(A8) hold. Let  $k(\cdot) \in \mathcal{K}_1$  such that  $\int |k(z)| dz < \infty$  and  $\int |z|k(z) dz < \infty$ . (i) Assuming that  $m_T \rightarrow \infty$ ,  $m_T/T \rightarrow 0$ , and  $m_T^2/(NT^{3/2}) \rightarrow 0$ , we have*

$$\hat{J}_T \xrightarrow{P} J_T.$$

(ii) *In addition, assume that  $m_T^2/T \rightarrow 0$  and, for some  $q \in (0, \infty)$ ,  $k_q$  and  $|J^{(q)}|$  are finite. Assuming that  $m_T^{q+1}/T \rightarrow 0$  and  $m_T^{2q+1}/(NT) \rightarrow \nu$ , for some  $0 < \nu < \infty$ , we have*

$$\lim_{N, T \rightarrow \infty} \frac{NT}{m_T} \text{MSE}(\hat{J}_T) = k_q^2 (\text{vec}(J^{(q)})' W \text{vec}(J^{(q)})) \nu^{-1} + \int k^2(z) dz \cdot \text{tr}(W(I + K_{dd})J \otimes J).$$

*On the other hand, assuming that  $m_T^{q+1}/T \rightarrow \nu$ , for some  $0 < \nu < \infty$ ,  $m_T^{2q+1}/(NT) \rightarrow 0$ , and  $D := \lim_{T \rightarrow \infty} m_T^{-1} \sum_{j=-T+1}^{T-1} k(\frac{j}{m_T}) B_j$  exists, we have*

$$\lim_{N, T \rightarrow \infty} \frac{T^2}{m_T^2} \text{MSE}(\hat{J}_T) = (\text{vec}(-k_q J^{(q)} \nu^{-1} - D)' W \text{vec}(-k_q J^{(q)} \nu^{-1} - D)).$$

For the bias, variance, and MSE of  $\hat{J}_T$ , some comments are in order. First, the bias has two leading terms, of order  $1/m_T^q$  and  $m_T/T$ , respectively. The first bias term exists because we use a kernel estimator; the second bias term is due to the bias in autocovariance estimator. Second, the variance has two leading terms, of order  $m_T/(NT)$  and  $m_T^2/(NT^{3/2})$ , respectively. The second term exists because the regression coefficients are unknown and have to be estimated. If  $m_T^2/T \rightarrow 0$ , the first variance term dominates the second and, therefore, the estimation uncertainty for the regression coefficients can be ignored. Third, we consider the MSE expressions. If  $m_T^{q+1}/T \rightarrow 0$ , the first bias term dominates the second, and if  $m_T^{2q+1}/(NT) \rightarrow \nu$ , the order of the squared bias is comparable to the variance. This leads to the first MSE of  $\hat{J}_T$ . If one uses an optimal bandwidth of order  $(NT)^{1/(2q+1)}$ , the given conditions imply that  $N^{q+1}/T^q \rightarrow 0$ . If  $m_T^{q+1}/T \rightarrow \nu$ , the two leading bias terms are comparable, and if  $m_T^{2q+1}/(NT) \rightarrow 0$ , the squared bias dominates the variance. This leads to the second MSE expression. Note that if we use an optimal bandwidth of order  $T^{1/(q+1)}$ , the conditions mean that  $N^{q+1}/T^q \rightarrow \infty$ .



### 4.3 Bias correction in the autocovariance matrix estimator

This subsection discusses how to estimate  $B_j$  and construct bias-corrected estimators for  $\Gamma_j$  and  $J_T$ . The bias of  $\widehat{\Gamma}_j$  is  $B_j/T$ , where  $B_j = N^{-1} \sum_{i=1}^N w_{ij} V_{T,i}$ . For each  $i$ , we can estimate  $\omega_{ij}$  and  $V_{T,i}$ . Let  $i \in I_1$ . Let  $\widehat{\omega}_{ij} = 2\widehat{\omega}_{1,ij} - \widehat{\omega}_{2,ij}$ , where

$$\widehat{\omega}_{1,ij} = \frac{1}{T} \sum_{t=1}^T \frac{1}{b_2} K_2\left(\frac{\widehat{e}_{2,it}}{b_2}\right) (\mathbf{x}_{it} - \widehat{\mathbf{c}}_{2,i})(\mathbf{x}_{it+j} - \widehat{\mathbf{c}}_{2,i})' / \widehat{f}_i(0),$$

$$\widehat{\omega}_{2,ij} = \frac{1}{T} \sum_{t=1}^{T-j} \frac{1}{b_3^2} K_3\left(\frac{\widehat{e}_{2,it}}{b_3}, \frac{\widehat{e}_{2,it+j}}{b_3}\right) (\mathbf{x}_{it} - \widehat{\mathbf{c}}_{2,i})(\mathbf{x}_{it+j} - \widehat{\mathbf{c}}_{2,i})' / \widehat{f}_i(0)^2,$$

where  $b_3$  is another sequence of bandwidths, and  $K_3(\cdot, \cdot)$  is a bivariate kernel function. Assume the following conditions:

(A9) (a)  $K_3(\cdot, \cdot)$  is a bivariate density function that is continuous and bounded, and of bounded variation. (b) The bandwidth sequence  $b_3$  satisfies that  $b_3 \rightarrow 0$  and  $(\log N)^2 / (Tb_3) \rightarrow 0$ .

Because  $V_{T,i}$  is the long-run variance of  $h_{it}$ , we use a kernel method. For each  $i$ , let  $n_{T,i} \rightarrow \infty$  as  $T \rightarrow \infty$ , and define  $\widehat{V}_{T,i} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{n_{T,i}}\right) \frac{T-|j|}{T} \widehat{\varrho}_{ij}$  and  $\widehat{\varrho}_{ij} = \frac{1}{T-j} \sum_{t=1}^{T-j} (\tau - I(\widehat{e}_{2,it} \leq 0))(\tau - I(\widehat{e}_{2,it+j} \leq 0))$ . Then the bias in Theorem 4.1 can be estimated as

$$\widehat{B}_j = \frac{1}{N} \sum_{i=1}^N \widehat{w}_{ij} \widehat{V}_{T,i}. \tag{18}$$

The working paper version of this study, [Yoon and Galvao \(2019\)](#), shows that  $\widehat{B}_j$  is consistent.

Although the estimation for each  $i$  is feasible, we find that using a simpler alternative often leads to better finite-sample performance. To keep the discussion simple, we focus on the case where  $V_{T,i} = V_T$  for all  $i$ . This amounts to the case that  $\varrho_{ij} = \varrho_j$  for all  $i$ . Note that this simplifying assumption applies only to the bias correction, and not to the estimation of the CCM estimator itself.

Under this simplification,  $B_j = \omega_j \cdot V_T$ , where  $\omega_j = N^{-1} \sum_{i=1}^N \omega_{ij}$ . Write  $\omega_j = 2\omega_{1,j} - \omega_{2,j}$ . Each term can be estimated by

$$\widehat{\omega}_{1,j} = \frac{1}{N} \sum_{i=1}^N \widehat{\omega}_{1,ij}, \quad \text{and} \quad \widehat{\omega}_{2,j} = \frac{1}{N} \sum_{i=1}^N \widehat{\omega}_{2,ij}.$$

Let  $n_T$  be a bandwidth sequence, possibly different from  $m_T$ . Then

$$\widehat{V}_T = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{n_T}\right) \frac{T-|j|}{T} \widehat{\varrho}_j,$$

with  $\widehat{\varrho}_j$  as defined as in (13). The results show that, as in the case of  $\widehat{\Gamma}_j$ , the scalar autocovariance estimator  $\widehat{\varrho}_j$  has asymptotic bias of order  $1/T$ . The following proposition states the result.

PROPOSITION 4.2. Assume conditions (A1)–(A6) hold. Let  $q_{ij} = q_j$ , for all  $1 \leq i \leq N$ . Define  $f_{ij}(0, 0) = E[f_{ij}(0, 0|x_{i1}, x_{i1+j})]$ , and  $s_j = N^{-1} \sum_{i=1}^N (2 - \frac{f_{ij}(0,0)}{f_i^2(0)})$ . As  $N, T \rightarrow \infty$ , we have

$$\begin{aligned} & \sqrt{NT} \left\{ \widehat{q}_j - q_j + \frac{s_j V_T}{T} \right\} \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^{T-j} \{ (\tau - I(e_{it} \leq 0)) (\tau - I(e_{it+j} \leq 0)) - q_j \} + o_p(1). \end{aligned} \tag{19}$$

REMARK 4.2. As before, if  $e_{it}$  and  $e_{it+j}$  are independent in the time-series dimension,  $s_j = 1$ , because  $f_{ij}(0, 0) = f_i^2(0)$ . In this special case, the asymptotic bias becomes  $-V_T/T$ , and does not depend on  $j$ .

To obtain good finite-sample performance of the CCM estimator, we apply the bias-correction procedure for  $\widehat{q}_j$ . If  $V_T$  and  $s_j$  can be consistently estimated by  $\widehat{V}_T$  and  $\widehat{s}_j$ , respectively, it becomes possible to define a bias-corrected autocovariance estimator  $\widehat{q}_j^{(1)} = \widehat{q}_j + \frac{1}{T} \widehat{s}_j \widehat{V}_T$ . Analogous to Proposition 4.1, if  $\widehat{s}_j \widehat{V}_T$  is a consistent estimator of  $s_j V_T$ , one can obtain

$$\sqrt{NT} \{ \widehat{q}_j^{(1)} - q_j \} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^{T-j} \{ (\tau - I(e_{it} \leq 0)) (\tau - I(e_{it+j} \leq 0)) - q_j \} + o_p(1).$$

There is a subtle point to make. It would be desirable to use a bias-corrected autocovariance estimator  $\widehat{q}_j^{(1)}$ , instead of the biased estimator  $\widehat{q}_j$ , to obtain  $\widehat{V}_T$ . However, this is not feasible because the bias in  $\widehat{q}_j$  depends on the long-run variance itself. For this reason, we consider an iterative procedure.

If the autocovariance estimator underestimates the true value,  $\widehat{q}_j < q_j$ ,  $\widehat{V}_T$  will underestimate  $V_T$ . In order to correct this, we apply the iterative method of Okui (2010). To keep the presentation simple, consider a special case where  $s_j$  is known to be one. In general, the term  $s_j$  in the bias expression can be estimated in the same way as  $\omega_j$ . The general case can be found in Yoon and Galvao (2019). After the  $i$ th iteration, the next long-run variance is estimated by

$$\widehat{V}_T^{(i+1)} = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{n_T}\right) \frac{T-|j|}{T} \widehat{q}_j^{(i)},$$

and the autocovariance estimators are updated by

$$\widehat{q}_j^{(i+1)} = \widehat{q}_j + \frac{1}{T} \widehat{V}_T^{(i+1)}, \quad j = 1, \dots, T-1.$$

Note that  $\widehat{q}_j$  is the original autocovariance estimator, and is fixed in iteration. This is repeated until convergence, with the resulting estimators denoted as  $\widehat{q}_j^{(\infty)}$  and  $\widehat{V}_T^{(\infty)}$ . To keep the exposition simple, we use the same bandwidth  $n_T$  throughout the iteration. The iteration procedure can be characterized as follows. Let  $\widehat{q}^{(i)} = (\widehat{q}_0^{(i)}, \widehat{q}_1^{(i)}, \dots, \widehat{q}_{T-1}^{(i)})'$

and  $\widehat{\varrho} = (\widehat{\varrho}_0, \widehat{\varrho}_1, \dots, \widehat{\varrho}_{T-1})'$ . Let  $\iota_T$  be a  $T \times 1$  vector of ones, and let  $I_T$  be the  $T \times T$  identity matrix. Define

$$K_T = \left( k(0), 2\frac{T-1}{T}k\left(\frac{1}{n_T}\right), 2\frac{T-2}{T}k\left(\frac{2}{n_T}\right), \dots, 2\frac{1}{T}k\left(\frac{T-1}{n_T}\right) \right)'$$

The limit of the iterative procedure can be written as

$$\widehat{\varrho}^{(\infty)} = \left( I_T - \frac{1}{T}\iota_T K_T' \right)^{-1} \widehat{\varrho},$$

and

$$\widehat{V}_T^{(\infty)} = K_T' \widehat{\varrho}^{(\infty)} = \left( 1 + \frac{\iota_T' K_T}{T - \iota_T' K_T} \right) \widehat{V}_T.$$

The iteration converges when  $k(z) < 1$ , for  $z \neq 0$ , which is satisfied by commonly used kernels, such as the Bartlett and the QS kernel. See Okui (2010) for further discussions.

**THEOREM 4.3.** *Assume conditions (A1)–(A9) hold, and that  $\varrho_{ij} = \varrho_j$ . Let  $k(\cdot) \in \mathcal{K}_1$ , such that  $\int |k(z)| dz < \infty$  and  $\int |z|k(z) dz < \infty$ . Assuming that  $n_T \rightarrow \infty$ ,  $n_T/T \rightarrow 0$ , and  $n_T^2/(NT^{3/2}) \rightarrow 0$ , we have*

$$\widehat{V}_T^{(\infty)} \xrightarrow{p} V_T.$$

Moreover, assume that  $n_T^2/T \rightarrow 0$  and, for some  $q \in (0, \infty)$ ,  $k_q$  and  $|V^{(q)}|$  are finite. Assuming that  $n_T^{2q+1}/(NT) \rightarrow \nu$ , for some  $0 < \nu < \infty$ , we have

$$\lim_{N, T \rightarrow \infty} \frac{NT}{n_T} \text{MSE}(\widehat{V}_T^{(\infty)}) = 2V^2 \int k(z)^2 dz + (k_q V^{(q)})^2 \nu^{-1}.$$

A comparison of the results of Theorems 4.2 and 4.3 shows that the MSE expression becomes simpler. This is because the bias correction helps simplify the bias of  $\widehat{V}_T^{(\infty)}$ .

To summarize, the bias in  $\widehat{F}_j$  can be estimated by  $T^{-1}\widehat{B}_j = T^{-1}\widehat{\omega}_j \cdot \widehat{V}_T^{(\infty)}$ . The bandwidth  $n_T$  used to calculate  $\widehat{V}_T^{(\infty)}$  is obtained by minimizing the MSE in Theorem 4.3.

#### 4.4 The bias-corrected CCM estimator

When the bias-corrected autocovariance estimator  $\widehat{F}_j^{(1)} = \widehat{F}_j + \frac{1}{T}\widehat{B}_j$  is available, we can define the following bias-corrected estimator for  $J_T$ :

$$\check{J}_T = \sum_{j=-T+1}^{T-1} k\left(\frac{j}{m_T}\right) \frac{T-|j|}{T} \widehat{F}_j^{(1)}. \tag{20}$$

The next result presents the limiting properties of  $\check{J}_T$ .

**THEOREM 4.4.** *Assume conditions (A1)–(A9) hold. Let  $k(\cdot) \in \mathcal{K}_1$ , such that  $\int |k(z)| dz < \infty$  and  $\int |z|k(z) dz < \infty$ . Assuming that  $m_T \rightarrow \infty$ ,  $m_T/T \rightarrow 0$ , and  $m_T^2/(NT^{3/2}) \rightarrow 0$ , we have*

$$\check{J}_T \xrightarrow{P} J_T.$$

*In addition, assume that  $m_T^2/T \rightarrow 0$  and, for some  $q \in (0, \infty)$ ,  $k_q$  and  $|J^{(q)}|$  are finite. Assuming that  $m_T^{q+1}/(N^{1/2}T) \rightarrow 0$  and  $m_T^{2q+1}/(NT) \rightarrow \nu$ , for some  $0 < \nu < \infty$ , we have*

$$\lim_{N, T \rightarrow \infty} \frac{NT}{m_T} \text{MSE}(\check{J}_T) = k_q^2 (\text{vec}(J^{(q)})' W \text{vec}(J^{(q)})) \nu^{-1} + \int k^2(z) dz \cdot \text{tr}(W(I + K_{dd})J \otimes J).$$

Compared with those in Theorem 4.2, the results in Theorem 4.4 are simpler. There is only one MSE expression, and the condition  $m_T^{q+1}/T \rightarrow 0$  in Theorem 4.2 is replaced by the weaker condition  $m_T^{q+1}/(N^{1/2}T) \rightarrow 0$ . This is owing to the bias correction. The bias-corrected CCM estimator is then given by

$$\widehat{\Sigma} = \widehat{\Lambda}^{-1} \check{J}_T \widehat{\Lambda}^{-1}. \tag{21}$$

It is straightforward to see that the bias-corrected CCM estimator  $\widehat{\Sigma}$  in (21) consistently estimates the asymptotic covariance matrix of  $\widehat{\theta}$ , that is,  $\Lambda^{-1}J\Lambda^{-1}$ .

In some cases, the simplified long-run variance matrix  $J_T^s$  is suitable. For example, the Score test in Section 5 assumes that  $f_i(e_{it}|x) = f_i(e_{it})$  and, under this conditional homoskedasticity assumption,  $\varrho_{ij} = \varrho_j$  for all  $i$ . A natural estimator for this case,  $\widehat{J}_T^s$ , is considered in Yoon and Galvao (2019), who show that  $m_T/T \rightarrow 0$  instead of  $m_T^2/T \rightarrow 0$  is sufficient to obtain the MSE of  $\widehat{J}_T^s$ .

#### 4.5 Optimal bandwidth and data-driven bandwidth selection

The optimal bandwidth for the bias-corrected estimator  $\check{J}_T$  in (20) can be obtained as follows. By differentiation, obtain the MSE optimal bandwidth

$$m_T^* = \left( \frac{qk_q^2\phi(q)}{\int k^2(z) dz} \right)^{1/(2q+1)} (NT)^{1/(2q+1)}, \quad \text{where} \tag{22}$$

$$\phi(q) = \frac{2 \text{vec}(J^{(q)})' W \text{vec}(J^{(q)})}{\text{tr}(W(I + K_{dd})J \otimes J)}.$$

Once a kernel function  $k(\cdot)$  is fixed, all elements are determined in (22), except  $\phi(q)$ . When the number of covariates is one, this quantity reduces to  $\phi(q) = (J^{(q)})^2$ . For the Bartlett and QS kernels, the values of  $q$  are one and two, respectively, and we have

$$m_T^* = 1.1447(\phi(1)NT)^{1/3}, \quad \text{for Bartlett kernel;}$$

$$m_T^* = 1.3221(\phi(2)NT)^{1/5}, \quad \text{for QS kernel.}$$

The value of  $\phi(q)$  is unknown, and thus has to be estimated. Let  $H_{it,a}$  be the  $a$ th element of  $H_{it}$ . To estimate  $\phi(q)$ , we follow Andrews (1991) and use the parametric approximation that is valid if each  $H_{it,a}$  follows an AR(1) process (with different parameters for each  $a$ ). We employ a  $d$  univariate approximation with a simple weighting matrix  $W$ . For this purpose, let  $W$  assign weights only to the diagonal elements of  $\check{J}_T$ . A natural choice would be an equal weight of one, for  $a = 1, \dots, d$ . We use this case, and obtain

$$\phi(q) = \frac{\sum_{a=1}^d (J_{aa}^{(q)})^2}{\sum_{a=1}^d J_{aa}^2},$$

where  $J_{aa}^{(q)}$  and  $J_{aa}$  are the  $a$ -th diagonal elements of  $J^{(q)}$  and  $J$ , respectively. Consider an AR(1) model for  $H_{it,a}$ , with  $(\delta_a, \sigma_a)$  as the autoregressive and innovation variance parameters, respectively, for  $a = 1, \dots, d$ . Thus, we have that

$$\begin{aligned} \hat{\phi}(1) &= \sum_{a=1}^d \frac{4\hat{\delta}_a^2 \hat{\sigma}_a^4}{(1 - \hat{\delta}_a)^6 (1 + \hat{\delta}_a)^2} \bigg/ \sum_{a=1}^d \frac{\hat{\sigma}_a^4}{(1 - \hat{\delta}_a)^4} \quad \text{and} \\ \hat{\phi}(2) &= \sum_{a=1}^d \frac{4\hat{\delta}_a^2 \hat{\sigma}_a^4}{(1 - \hat{\delta}_a)^8} \bigg/ \sum_{a=1}^d \frac{\hat{\sigma}_a^4}{(1 - \hat{\delta}_a)^4}. \end{aligned}$$

With these estimates, the optimal bandwidth selector is given by

$$\hat{m}_T^* = \left( \frac{qk_q^2 \hat{\phi}(q)}{\int k^2(z) dz} \right)^{1/(2q+1)} (NT)^{1/(2q+1)}. \tag{23}$$

Examples of the bandwidth selection rule are

$$\begin{aligned} \hat{m}_T^* &= 1.1447(\hat{\phi}(1)NT)^{1/3}, \quad \text{for Bartlett kernel;} \\ \hat{m}_T^* &= 1.3221(\hat{\phi}(2)NT)^{1/5} \quad \text{for QS kernel.} \end{aligned}$$

Similarly, the optimal bandwidth for estimating  $\hat{V}_T^{(\infty)}$  is

$$n_T^* = \left( \frac{qk_q^2 \phi_1(q)}{\int k^2(z) dz} \right)^{1/(2q+1)} (NT)^{1/(2q+1)}, \tag{24}$$

where  $\phi_1(q) = (\frac{V^{(q)}}{V})^2$ . When  $h_{it}$  follows an AR(1) process with the autoregressive coefficient  $\delta$ , then

$$\phi_1(1) = \left( \frac{2\delta}{1 - \delta^2} \right)^2 \quad \text{and} \quad \phi_1(2) = \left( \frac{2\delta}{(1 - \delta)^2} \right)^2.$$

Let  $\widehat{\phi}_1(1)$  and  $\widehat{\phi}_1(2)$  be estimates of  $\phi_1(1)$  and  $\phi_1(2)$ , respectively, using  $\widehat{\delta}$ . Then the feasible bandwidth selection rule is  $\widehat{n}_T^* = 1.1447(\widehat{\phi}_1(1)NT)^{1/3}$  for the Bartlett kernel, and  $\widehat{n}_T^* = 1.3221(\widehat{\phi}_1(2)NT)^{1/5}$  for the QS kernel.

### 5. CLUSTER ROBUST INFERENCE METHODS

This section develops two tests based on the CCM and establishes their limit distributions. Let  $\boldsymbol{\theta}_0(\tau) = (\boldsymbol{\beta}_0(\tau)', \boldsymbol{\gamma}_0(\tau)')'$  and  $\mathbf{x}_{it} = (\mathbf{w}'_{it}, \mathbf{z}'_{it})'$ , where  $\mathbf{w}_{it}$  and  $\mathbf{z}_{it}$  are  $d_w \times 1$  and  $d_z \times 1$  vectors, respectively.

#### 5.1 Score test

For the Score test, we estimate model (1) under the null hypothesis  $H_0 : \boldsymbol{\gamma}_0(\tau) = 0$ . By an abuse of notation, let  $(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau))$  denote one of the bias-corrected estimators in Section 2, obtained under the null hypothesis.

Let  $\mathbf{W}^*$  and  $\mathbf{Z}^*$  be  $NT \times d_w$  and  $NT \times d_z$  matrices, with elements  $\mathbf{w}_{it}$  and  $\mathbf{z}_{it}$ , respectively. Let  $\mathbf{d}_i$  be a dummy variable for the  $i$ th individual, and  $\mathbf{W} = (\mathbf{d}_1, \dots, \mathbf{d}_{N-1}, \mathbf{W}^*)$ . Define  $\tilde{\mathbf{Z}}$  as the projection error of  $\mathbf{Z}^*$  on the space spanned by the columns of the matrix  $\mathbf{W}$ ; that is,  $\tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}')\mathbf{Z}^*$ . Let  $\tilde{z}_{it}$  be an entity in  $\tilde{\mathbf{Z}}$  that corresponds to the  $t$ th period of the  $i$ th individual. Observe that  $\sum_{t=1}^T \tilde{z}_{it} = 0$  and  $\sum_{t=1}^T \tilde{z}_{it}\mathbf{w}'_{it} = 0$ , for a given  $i$ . Let  $\widehat{\boldsymbol{\beta}}_1(\tau)$  and  $\widehat{\boldsymbol{\beta}}_2(\tau)$  be the slope estimates using samples  $I_1$  and  $I_2$ , respectively. The test statistic is based on a subgradient

$$S(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) = \frac{N_1}{N} \cdot S_1(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) + \frac{N_2}{N} \cdot S_2(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)), \tag{25}$$

where

$$S_1(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) = \frac{1}{\sqrt{N_1 T}} \sum_{i=1}^{N_1} \sum_{t=1}^T (\tau - I(y_{it} \leq \widehat{\alpha}_i(\tau) + \mathbf{w}'_{it}\widehat{\boldsymbol{\beta}}_2(\tau)))\tilde{z}_{it},$$

$$S_2(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) = \frac{1}{\sqrt{N_2 T}} \sum_{i=N_1+1}^N \sum_{t=1}^T (\tau - I(y_{it} \leq \widehat{\alpha}_i(\tau) + \mathbf{w}'_{it}\widehat{\boldsymbol{\beta}}_1(\tau)))\tilde{z}_{it}.$$

For the Score test, in this section,  $J_T^s$  uses  $\tilde{z}_{it}$  in place of  $\mathbf{x}_{it}$ . By Lemma S.3 in the Supplemental Appendix, and a straightforward variance calculation, one can easily show that  $J_T^s$  is the CCM for  $S(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau))$ . This CCM can be estimated by  $\widehat{J}_T^s$  (using  $\tilde{z}_{it}$  in place of  $\mathbf{x}_{it}$ ). The Score statistic  $H(\tau)$  is then defined by

$$H(\tau) := S(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau))(\widehat{J}_T^s)^{-1}S(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)). \tag{26}$$

The asymptotic properties of  $S(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau))$ , and  $H(\tau)$  are stated below.

**THEOREM 5.1.** *Assume conditions (A1)–(A9) hold, and  $f_i(e|\mathbf{x}) = f_i(e)$ , for any  $\mathbf{x}$ . Under the null hypothesis  $H_0 : \boldsymbol{\gamma}_0(\tau) = 0$ ,*

- (i)  $S(\widehat{\boldsymbol{\alpha}}(\tau), \widehat{\boldsymbol{\beta}}(\tau)) \xrightarrow{d} N(0, J_T^s),$
- (ii)  $H(\tau) \xrightarrow{d} \chi_{d_z}^2.$

Theorem 5.1 states that the Score statistic  $H(\tau)$  is asymptotically  $\chi_{d_z}^2$  distributed. Therefore, the critical values for a given level of significance are easy to find.

In general, Score tests in QR work under the location-shift model. The Score test we propose for the FE-QR is no exception, in that it does not allow heteroskedastic errors. Given this drawback, the main advantage of the Score test lies in its simplicity, because it is free of the  $\Lambda$  term, which includes conditional densities. When it is appropriate to allow heteroskedastic errors, the Wald test in the next section is useful.

### 5.2 Wald test

Consider a general linear hypothesis  $H_0 : R\boldsymbol{\theta}_0(\tau) = r$  versus  $H_1 : R\boldsymbol{\theta}_0(\tau) \neq r$ . Estimate all coefficients in model (1) and let  $\widehat{\boldsymbol{\theta}}(\tau)$  denote one of the bias-corrected estimators. Define the Wald statistic as follows:

$$W(\tau) = NT(R\widehat{\boldsymbol{\theta}}(\tau) - r)'(R'\widehat{\Lambda}^{-1}J_T\widehat{\Lambda}^{-1}R)^{-1}(R\widehat{\boldsymbol{\theta}}(\tau) - r). \tag{27}$$

The limiting null distribution of the Wald statistic is stated in the next lemma.

**THEOREM 5.2.** *Assume conditions (A1)–(A9) hold. Then (i)  $\widehat{\Lambda} \xrightarrow{P} \Lambda$ . (ii) Under the null hypothesis,  $W(\tau) \xrightarrow{d} \chi_{d_z}^2.$*

The Wald statistic  $W(\tau)$  is valid under heteroskedasticity of arbitrary form. The practical implementation of the test is simple. For a Wald test with a  $100(1 - \alpha)\%$  significance level, we reject the null hypothesis of no effect when the value of  $W(\tau)$  from our sample is greater than the  $100(1 - \alpha)$ -th percentile of the chi-square distribution with  $d_z$  degrees of freedom.

In our simulation, we concentrate on the special case  $H_0 : \boldsymbol{\gamma}_0(\tau) = 0$  vs.  $H_1 : \boldsymbol{\gamma}_0(\tau) \neq 0$  to compare the Wald and Score tests. In this special case, the Wald test relies on the size of the difference  $\widehat{\boldsymbol{\gamma}}(\tau) - \boldsymbol{\gamma}_0(\tau)$ . Divide

$$\Lambda = \begin{pmatrix} \Lambda_{ww} & \Lambda_{wz} \\ \Lambda_{zw} & \Lambda_{zz} \end{pmatrix}, \quad \mathbf{c}_i = \begin{pmatrix} \mathbf{c}_{wi} \\ \mathbf{c}_{zi} \end{pmatrix}.$$

Define  $\tilde{\mathbf{w}}_{it} = \mathbf{z}_{it} - \mathbf{c}_{zi} - \Lambda_{zx}\Lambda_{xx}^{-1}(\mathbf{w}_{it} - \mathbf{c}_{wi})$ . Let  $J_T^w$  be equal to  $J_T$ , but using  $\tilde{\mathbf{w}}_{it}$  in place of  $\mathbf{x}_{it}$ . Define  $\Omega = (\Lambda_{zz} - \Lambda_{zw}\Lambda_{ww}^{-1}\Lambda_{wz})^{-1}J_T^w(\Lambda_{zz} - \Lambda_{zw}\Lambda_{ww}^{-1}\Lambda_{wz})^{-1}$ . Let  $\widehat{\Omega}$  be its estimate using the CCM. Then the Wald statistic becomes

$$W(\tau) = NT(\widehat{\boldsymbol{\gamma}}(\tau) - \boldsymbol{\gamma}_0(\tau))\widehat{\Omega}^{-1}(\widehat{\boldsymbol{\gamma}}(\tau) - \boldsymbol{\gamma}_0(\tau)). \tag{28}$$

The limit distribution of  $W(\tau)$  is the chi-squared distribution with  $d_z$  degrees of freedom.

## 6. REAL-DATA APPLICATION

To illustrate the use of the CCM estimator, we consider the movement of state-level unemployment rates. This exercise closely follows the spirit of [Bertrand, Duflo, and Mullainathan \(2004\)](#), who analyzed state-level female wages.<sup>14</sup>

The sample includes the yearly unemployment rates of 51 states for a period of 37 years, ranging from 1976 to 2012; thus, there are  $51 \times 37 = 1887$  observations. The unemployment rates and other state and national level variables were extracted from the Bureau of Labor Statistics's unemployment database.<sup>15</sup> To remove the common trend in the movement of the unemployment rate, we subtract the national unemployment rate from each state's unemployment rate. The dependent variable measures state-specific shocks in the labor market in a given year, not shared by other states. This detrended state unemployment rate is still highly serially correlated; see [Table 1](#). The independent variables include the logarithm of the total number of employed in a given state, the state FE, and a randomly generated placebo law. We estimate model (2) with these variables.

To investigate how some commonly used testing methods perform in the presence of serial correlation, we examine their empirical sizes under randomly introduced laws. We first draw a year at random, and choose half of the 51 states at random. The selected states after the chosen year are considered to be affected by the passage of the law. Because the law is chosen at random, it is independent of the outcome. Therefore, we expect that a 5% significance test will reject the null hypothesis of no treatment effect 5% of the time.

In each simulation run, we repeat the random draws of the treated years and states. We then estimate the model and count how often the absolute value of the t-statistic on the spurious treatment variable exceeds 1.97. We try two different standard errors for QR, and one for the OLS: "QR-1" uses a standard error assuming *iid* errors, and "QR-2" uses a standard error that allows non-*iid* errors.<sup>16</sup> Finally, we report "OLS" which uses the conventional OLS standard error, assuming *iid* errors, for the sake of comparison.

[Table 1\(a\)](#) illustrates the problem. It reports the rejection ratios of the tests when the nominal rate is 5%. We find that the conventional QR standard errors have high rejection ratios, much larger than the nominal rate of 5%. This is a clear indication that we need to correct the dependence in the errors. It also indicates that the degree of the size inflation may depend on the quantile levels. The rejection ratio of the uncorrected OLS standard error, about 35%, is close to the ratios reported in [Bertrand, Duflo, and Mullainathan](#) (see their [Table 3](#)). The results confirm that a QR estimation of a panel data model with FE suffers from similar size distortion effects to those seen in the OLS case.

We then examine how our proposed solutions fare in this exercise. The bias-corrected CCM estimator,  $\hat{\Sigma} = \hat{A}^{-1} \check{J}_T \hat{A}^{-1}$ , is obtained as follows:

<sup>14</sup>A second alternative illustration, which considers a capital structure model, can be found in the working paper of this study, [Yoon and Galvao \(2019\)](#).

<sup>15</sup>available from the BLS website <http://www.bls.gov/data/#unemployment>

<sup>16</sup>We use the R function `rqr()` for implementation. For QR-1, we use the option `se='iid'`, which estimates the model assuming *iid* errors. For QR-2, we use an option `se='ker'`, which use Powell's kernel estimate of the sandwich variance matrix formula.



TABLE 1. Size of tests. State unemployment rate regression.

(a) Conventional t-tests in QR and OLS.

Quantile Index	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	QR-1	QR-2	OLS
$\tau = 0.50$	0.892 (0.091)	0.758 (0.159)	0.624 (0.219)	0.258 (0.001)	0.216 (0.001)	0.351 (0.011)
$\tau = 0.75$				0.382 (0.011)	0.298 (0.010)	

(b) Robust Score and Wald tests.

Quantile Index	Score			Wald		
	$S_0$	$S_1$	$S_2$	$W_0$	$W_1$	$W_2$
$\tau = 0.50$	0.269 (0.009)	0.054 (0.005)	0.049 (0.005)	0.215 (0.009)	0.067 (0.006)	0.052 (0.005)
$\tau = 0.75$	0.295 (0.010)	0.065 (0.006)	0.060 (0.005)	0.313 (0.010)	0.100 (0.006)	0.061 (0.005)

*Note:* The tables report the fractions of simulation runs in which we reject the null hypothesis of no treatment effect. The nominal rate is 0.05, and the number of simulation replications is 2000. The dependent variable is the state unemployment rate. The first three columns in Table (a) show the average values of three autocorrelation coefficients of the dependent variable. Columns 4–6 show the rejection ratios of the conventional t-statistics of the spurious treatment variable in the QR and OLS. Table (b) shows the empirical sizes of the Score and Wald tests.  $W_0$  uses the conventional variance matrix estimator assuming independent errors,  $W_1$  uses the CCM estimator without bias correction  $\hat{J}_T$ , and  $W_2$  uses the CCM estimator with bias correction  $\hat{J}_T$ . The same naming convention applies to the Score tests.

(1) Under the QR model in (1), use the FE-SQR estimator to obtain a bias-corrected estimate,  $\hat{\theta}_{BC}$  or  $\hat{\theta}_{JK}$  (in (6) or (7), resp.). Then  $\hat{e}_{it}(\tau)$  is obtained. Following GK, we use the fourth-order kernel (as defined in their Section 4) and the bandwidth rule  $b_N = \hat{s}_e \cdot (NT)^{-1/7}$ , where  $\hat{s}_e$  is the standard deviation of  $\hat{e}_{it}$  obtained by the FE-SQR estimator. To estimate  $\Lambda$ , we use equation (10), with the uniform kernel with bandwidth  $b_1 = \min(s, IQR/1.34) \cdot b_{HS}$ . Here,  $s$  and IQR are the standard deviation and the interquartile range of the residuals  $\hat{e}_{it}$ , and  $b_{HS}$  is the Hall and Sheather (1988) bandwidth using  $\hat{e}_{it}$ , as implemented in Koenker and Machado (1999).<sup>17</sup> To estimate  $f_i$  and  $\mathbf{c}_i$ , we use the Gaussian kernel for  $K_2(\cdot)$  with the bandwidths  $b_2$  selected by Silverman’s rule of thumb using  $\hat{e}_{it}$ .

(2) To compute  $\check{J}_T$ , we need to calculate the bias  $\hat{B}_j = \hat{\omega}_j \cdot \hat{V}_T^{(\infty)}$  and  $\hat{\Gamma}_j$  as in (12), and then construct the bias-corrected autocovariance matrix  $\hat{\Gamma}_j^{(1)}$  as in (16). The bias correction  $\hat{B}_j$  requires that we estimate the scalar long-run variance  $V_T$ , and the quantities  $s_j$  and  $\omega_j$ . The former can be estimated by the iterative bias correction in Section 4.3, and the latter can be estimated by a nonparametric estimator with the following choice of tuning parameter: the kernel  $K_3(\cdot, \cdot)$  is a product of Gaussian kernels, and the bandwidth is chosen by a bivariate rule of thumb,  $b_3 = \hat{s}_e \cdot (NT)^{-1/6}$ . For simplicity, we use the same bandwidth for each coordinate. In our experience, a simple alternative, taking  $s_j = 1$  and  $\omega_j = 1$  as a simple approximation of nonparametric estimates worked well, too.

<sup>17</sup>The Hall and Sheather bandwidth is calculated using the R function `bandwidth.rq`.

(3) Now, we estimate the middle term of the CCM using the bias-corrected  $\check{J}_T$  in (20). Here, note that a one-time bias correction is sufficient. We use the optimal bandwidth estimate  $\hat{m}_T^*$  in (23) to estimate  $J_T$ . A bandwidth cannot exceed  $T - 1$  by design. Therefore, we use  $\min(\hat{m}_T^*, T - 1)$  as the bandwidth value in actual implementation.

(4) Finally, the CCM estimator is constructed from (21), from which cluster robust standard errors can be obtained. The sample splitting is a useful theoretical device, but it is known to make little difference in actual implementation. See Schick (1986) for related discussions. Therefore, we do not use this in our simulation exercise.

We now examine the results. Table 1(b) reports the rejection ratios of the cluster robust Score and Wald tests. Because we focus on a single regression coefficient, robust t-tests employing CCM estimators will lead to the same results as those of the Wald tests here.  $W_0$  uses the conventional variance matrix estimator, assuming independent errors,  $W_1$  uses the CCM estimator without bias correction  $\hat{J}_T$ , and  $W_2$  uses the CCM estimator with bias correction  $\check{J}_T$ . The same naming convention applies to the Score tests. Both  $\hat{J}_T$  and  $\check{J}_T$  use the same bandwidth; thus, if two estimators lead to different results, this is due solely to the bias correction.

A nonrobust Wald test  $W_0$  has size distortions of similar magnitude to those in a conventional t-test. Empirical rejection ratios of a 5% test can be as high as 31.3%, and using CCM estimators can greatly improve this problem. With  $W_1$ , the rejection ratios change from 0.215 to 0.067 ( $\tau = 0.5$ ) and from 0.313 to 0.100 ( $\tau = 0.75$ ). The remaining gap is further reduced by the bias correction. With  $W_2$ , the rejection ratios become 0.052 and 0.061 (when  $\tau = 0.5$  and 0.75, resp.) and are close to the nominal rate. The Supplemental Appendix, Section S.1, provides an extensive simulation exercise that examines several issues, the Score versus Wald tests, the performance of confidence interval estimators, several  $N$  and  $T$ , and choices of regression coefficients estimators.

## 7. CONCLUSION

It has become common practice among applied researchers using OLS to correct the clustering problem in an inference when there is serial correlation in the errors. This study shows that it is equally important to correct standard errors in FE panel QR models. We show by simulation that conventional QR standard errors have a serious size distortion problem. We then show by simulation exercises and real data examples that the proposed CCM estimator can be an effective solution. To ensure good finite-sample performance of the CCM estimator, we propose a bias-correction method and an MSE optimal bandwidth selection rule. Our results indicate that the proposed methods are effective. Many issues remain to be investigated. An extension to a two-way panel model accounting for time fixed effects is a topic in future work. Moreover, we have suggested a bias corrected estimator for the asymptotic CCM matrix in which the inside term of the sandwich formula is corrected. However, the estimator of the outside terms of the matrix depend nonlinearly on the FE estimator and might benefit from the bias-correction. Such extension will be pursued in further research.

## REFERENCES

- Abrevaya, J. and C. M. Dahl (2008), “The effects of birth inputs on birthweight: Evidence from quantile estimation on panel data.” *Journal of Business and Economic Statistics*, 26, 379–397. [582]
- Andrews, D. W. K. (1991), “Heteroskedasticity and autocorrelation consistent covariance matrix estimation.” *Econometrica*, 59 (3), 817–858. [587, 591, 594, 599]
- Arellano, M. (1987), “Computing robust standard errors for within-groups estimators.” *Oxford Bulletin of Economics and Statistics*, 49, 431–434. [580, 585]
- Arellano, M. and S. Bonhomme (2016), “Nonlinear panel data estimation via quantile regressions.” *Econometrics Journal*, 19, 61–94. [582]
- Bertrand, M., E. Duflo, and S. Mullainathan (2004), “How much should we trust differences-in-differences estimates?” *Quarterly Journal of Economics*, 119, 249–275. [580, 602]
- Bickel, P. J. (1982), “On adaptive estimation.” *The Annals of Statistics*, 10 (3), 647–671. [588]
- Bradley, R. C. (2005), “Basic properties of strong mixing conditions. A survey and some open questions.” *Probability Surveys*, 2, 107–144. [590]
- Cameron, A. C. and D. L. Miller (2015), “A practitioner’s guide to cluster-robust inference.” *Journal of Human Resources*, 50, 317–372. [580]
- Canay, I. A. (2011), “A simple approach to quantile regression for panel data.” *Econometrics Journal*, 14, 368–386. [582]
- Chen, L., L.-J. Wei, and M. I. Parzen (2004), “Quantile regression for correlated observations.” In *Proceedings of the Second Seattle Symposium in Biostatistics*. Lect. Notes Stat., Vol. 179, 51–69, Springer. [582]
- Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013), “Average and quantile effects in nonseparable panel models.” *Econometrica*, 81, 535–580. [582]
- Chetverikov, D., B. Larsen, and C. Palmer (2016), “IV quantile regression for group-level treatments, with an application to the distributional effects of trade.” *Econometrica*, 84, 809–833. [582]
- Dhaene, G. and K. Jochmans (2015), “Split-panel jackknife estimation of fixed-effect models.” *The Review of Economic Studies*, 82, 991–1030. [584]
- Donald, S. G. and K. Lang (2007), “Inference with difference-in-differences and other panel data.” *Review of Economics and Statistics*, 89 (2), 221–233. [580]
- El Ghouch, A. and I. van Keilegom (2009), “Local linear quantile regression with dependent censored data.” *Statistica Sinica*, 19, 1621–1640. [581]
- Galvao, A. F. and K. Kato (2016), “Smoothed quantile regression for panel data.” *Journal of Econometrics*, 193, 92–112. [583]

- Galvao, A. F., C. Lamarche, and L. Lima (2013), "Estimation of censored quantile regression for panel data with fixed effects." *Journal of the American Statistical Association*, 108 (503), 1075–1089. [582]
- Galvao, A. F. and L. Wang (2015), "Efficient minimum distance estimator for quantile regression fixed effects panel data." *Journal of Multivariate Analysis*, 133, 1–26. [582]
- Graham, B. S., J. Hahn, A. Poirier, and J. L. Powell (2018), "A quantile correlated random coefficients panel data model." *Journal of Econometrics*, 206, 305–335. [582]
- Gutenbrunner, C., J. Jureckova, R. Koenker, and S. Portnoy (1993), "Tests of linear hypotheses based on regression rank scores." *Journal of Nonparametric Statistics*, 2, 307–331. [581]
- Hagemann, A. (2013), "Robust spectral analysis." Working paper, available at <https://arxiv.org/pdf/1111.1965.pdf>. [592]
- Hagemann, A. (2017), "Cluster-robust bootstrap inference in quantile regression models." *Journal of the American Statistical Association*, 112 (517), 446–456. [582]
- Hahn, J. and G. M. Kuersteiner (2002), "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both  $n$  and  $T$  are large." *Econometrica*, 70, 1639–1657. [590]
- Hahn, J. and G. M. Kuersteiner (2011), "Bias reduction for dynamic nonlinear panel models with fixed effects." *Econometric Theory*, 27, 1152–1191. [585, 590]
- Hall, P. and S. J. Sheather (1988), "On the distribution of a studentized quantile." *Journal of the Royal Statistical Society. Ser. B*, 50 (3), 381–391. [603]
- Hansen, C. B. (2007), "Asymptotic properties of a robust variance matrix estimator for panel data when  $T$  is large." *Journal of Econometrics*, 141, 597–620. [580, 585, 589]
- Jun, S. J. and J. Pinkse (2009), "Efficient semiparametric seemingly unrelated quantile regression estimation." *Econometric Theory*, 1392–1414. [582]
- Jung, S.-H. (1996), "Quasi-likelihood for median regression models." *Journal of the American Statistical Association*, 91 (433), 251–257. [582]
- Kato, K. (2012), "Asymptotic normality of Powell's kernel estimator." *Annals of the Institute of Statistical Mathematics*, 64, 255–273. [586]
- 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. [581, 582]
- Kézdi, G. (2004), "Robust standard error estimation in fixed-effects panel models." *Hungarian Statistical Review*, 9, 95–116. [580, 585]
- Koenker, R. (2004), "Quantile regression for longitudinal data." *Journal of Multivariate Analysis*, 91, 74–89. [590]
- Koenker, R. and J. A. F. Machado (1999), "Goodness of fit and related inference processes for quantile regression." *Journal of the American Statistical Association*, 94, 1296–1310. [603]

- Koenker, R. and Z. Xiao (2006), "Quantile autoregression." *Journal of the American Statistical Association*, 101, 980–990. [581]
- Koenker, R. and Q. Zhao (1996), "Conditional quantile estimation and inference for arch models." *Econometric Theory*, 12, 793–813. [581, 590]
- Komunjer, I. and Q. Vuong (2010), "Efficient estimation in dynamic conditional quantile models." *Journal of Econometrics*, 157, 272–285. [581]
- Linton, O. and Y.-J. Whang (2007), "The quantilogram: With an application to evaluating directional predictability." *Journal of Econometrics*, 141 (1), 250–282. [580, 587]
- Masry, E. and J. Fan (1997), "Local polynomial estimation of regression functions for mixing processes." *Scandinavian Journal of Statistics*, 24 (2), 165–179. [581]
- Moulton, B. R. (1986), "Random group effects and the precision of regression estimates." *Journal of Econometrics*, 32 (3), 385–397. [582]
- Moulton, B. R. (1990), "An illustration of a pitfall in estimating the effects of aggregate variables on micro units." *The Review of Economics and Statistics*, 72 (2), 334–338. [582]
- Newey, W. K. and K. D. West (1994), "Automatic lag selection in covariance matrix estimation." *The Review of Economic Studies*, 61 (4), 631–653. [587, 594]
- Okui, R. (2010), "Asymptotically unbiased estimation of autocovariances and autocorrelations with long panel data." *Econometric Theory*, 26 (5), 1263–1304. [580, 587, 589, 592, 596, 597]
- Parente, P. and J. Santos Silva (2016), "Quantile regression with clustered data." *Journal of Econometric Methods*, 5 (1), 1–15. [582]
- Parzen, E. (1957), "On consistent estimates of the spectrum of a stationary time series." *The Annals of Mathematical Statistics*, 28 (2), 329–348. [593]
- Powell, J. L. (1991), "Estimation of monotonic regression models under quantile restrictions." In *Nonparametric and Semiparametric Methods in Econometrics* (W. Barnett, J. Powell, and G. Tauchen, eds.), Cambridge University Press, Cambridge. [586]
- Schick, A. (1986), "On asymptotically efficient estimation in semiparametric models." *The Annals of Statistics*, 14 (3), 1139–1151. [588, 604]
- Stock, J. H. and M. W. Watson (2008), "Heteroskedasticity-robust standard errors for fixed effects panel data regression." *Econometrica*, 76 (1), 155–174. [589]
- Wang, H. and X. He (2007), "Detecting differential expressions in GeneChip microarray studies: A quantile approach." *Journal of the American Statistical Association*, 102, 104–112. [582, 585]
- 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>. [595, 596, 598, 602]

Yoon, J. and A. F. Galvao (2020), “Supplement to ‘Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects.’” *Quantitative Economics Supplemental Material*, 11, <https://doi.org/10.3982/QE802>. [581]

---

Co-editor Frank Schorfheide handled this manuscript.

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