CESifo WORKING PAPERS

7211
2018

August 2018

Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Quantile Regression Models

Matthew Harding, Carlos Lamarche, M. Hashem Pesaran

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7211*
Category 12: Empirical and Theoretical Methods

Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Quantile Regression Models

Abstract

This paper proposes a quantile regression estimator for a heterogeneous panel model with lagged dependent variables and interactive effects. The paper adopts the Common Correlated Effects (CCE) approach proposed by Pesaran (2006) and Chudik and Pesaran (2015) and demonstrates that the extension to the estimation of dynamic quantile regression models is feasible under similar conditions to the ones used in the literature. We establish consistency and derive the asymptotic distribution of the new quantile regression estimator. Monte Carlo studies are carried out to study the small sample behavior of the proposed approach. The evidence shows that the estimator can significantly improve on the performance of existing estimators as long as the time series dimension of the panel is large. We present an application to the evaluation of Time-of-Use pricing using a large randomized control trial.

JEL-Codes: C210, C310, C330, D120, L940.

Keywords: common correlated effects, dynamic panel, quantile regression, smart meter, randomized experiment.

Matthew Harding
Department of Economics
University of California at Irvine
USA – Irvine, CA 9269
Harding1@uci.edu

Carlos Lamarche
Department of Economics
University of Kentucky
USA – Lexington, KY 40506-0034
clamarche@uky.edu

M. Hashem Pesaran
Department of Economics
University of Southern California & Trinity College
USA – Los Angeles, CA 90089-0253
pesaran@usc.edu

August 14, 2018
We would like to thank Alexander Chudik, Ron Smith, and Jeff Wooldridge for helpful comments and suggestions as well as seminar participants at Michigan State University, University of California at Irvine, University of Glasgow, University of Miami, Universidad Di Tella, University of Oklahoma, Universidad de San Andres, Central Bank of Argentina, the 2017 International Association for Applied Econometrics conference, and the 2017 Boneyard Econometrics conference in celebration of Professor Roger Koenker.

1. **Introduction**

In the last decade, the literature on linear panel data models has made significant progress on the estimation of models with multi-factor error structure. Recent papers have focused on the estimation of models with a fixed number of unobserved factors (see e.g. Pesaran (2006), Bai (2009), Pesaran and Chudik (2014), Moon and Weidner (2015, 2017), Chudik and Pesaran (2015)). The Common Correlated Effects (CCE) approach of Pesaran (2006) is robust to cross-sectional dependence and slope heterogeneity, and it has been further developed to allow for possible unit roots in factors and spatial forms of weak cross-sectional dependence (see e.g., Kapetanios, Pesaran, and Yagamata (2011), Pesaran and Tosetti (2011) and Pesaran, Smith and Yagamata (2013)). The estimation of dynamic panel data models is investigated in Chudik and Pesaran (2015) and Moon and Weidner (2015, 2017). Moon and Weidner develop estimation approaches for models with lagged dependent variables and cross-sectional dependence, but they assume homogeneous coefficients. In an important paper, Chudik and Pesaran (2015) extend the approach developed by Pesaran (2006) to dynamic panel data models with heterogeneous slopes, for situations where the cross-sectional dimension (N) and the time-series dimension (T) are relatively large. This method however does not offer the possibility of estimating heterogeneous distributional effects, which is an important consideration for practice. For instance, the effect of a policy can be heterogeneous throughout the conditional distribution of the response variable, and therefore, it might not be well summarized by the average treatment effect.

Quantile regression, as introduced in the seminal work by Koenker and Bassett (1978), provides a convenient way to estimate distributional effects of policy variables, although in general these type of heterogeneous treatment effects are identified and estimated under the assumption that the slope coefficients are the same over all cross-sectional units. This condition is used in a number of different approaches that have been recently developed for the estimation of panel quantile regression models. The recent literature include work by Koenker (2004), Lamarche (2010), Galvao (2011), Rosen (2012), Galvao, Lamarche and Lima (2013), Chernozhukov, Fernandez-Val, Hahn and Newey (2013) and Chernozhukov, Fernandez-Val, Hoderlein, Holzmann and Newey (2015), Harding and Lamarche (2014, 2017), Arellano and Bonhomme (2016), among others. Slope heterogeneity in quantile regression is investigated in Galvao and Wang (2015). In related work, Ando and Bai (2017) and Chen, Dolado and Gonzalo (2017) investigate quantile factor models. With the exception of Galvao (2011) and Arellano and Bonhomme (2016), the literature has focused on estimating static models. Moreover, the panel quantile regression literature does not address cross-sectional dependence with the exception of Harding and Lamarche (2014) that adopt the approach proposed by Pesaran (2006) to estimate a static model with interactive effects. This paper extends the panel quantile literature to dynamic models with heterogeneous slopes and multi-factor error structure when both T and N are large.

We adopt a CCE approach and focus on estimation and inference of mean quantile coefficients. We allow for the possibility that unobserved factors and included regressors are correlated and we study the conditions under which the slope coefficients are estimated consistently. An important condition is that one plus the number of cross-sectional averages must be larger than the number of unobserved factors. Another important condition, which is similar to a condition used in Chudik and Pesaran (2015), is that a large number of lags of cross section averages used to approximate the factors needs to be included in the individual-specific equations of the panel. Under standard regularity conditions including T tending to infinity at a faster rate than N as in Kato, Galvao and Montes-Rojas (2012), we show that the average quantile estimator is consistent and asymptotically Gaussian. Moreover, we investigate the finite sample performance of the proposed approach in comparison with the method for dynamic models developed by Galvao (2011). Using a comprehensive set of Monte Carlo experiments, we find that the proposed estimator has a satisfactory performance under different dynamic specifications when T is relatively large.

We apply the method to estimate how consumers respond to time-of-use (TOU) electricity pricing and different type of technologies that allow communication between customers and utility companies. The use of a quantile-specific demand equation allows us to estimate the short and long run impacts of different enabling technologies, while including three key features of the problem: dynamics, slope heterogeneity and cross-sectional dependence. We use a data set of more than 6.5 million observations obtained from a large randomized control trial which includes $N = 779$ customers observed over $T = 8639$ time intervals.

Our findings suggest that smart thermostats are particularly effective relative to other enabling technologies and the differential effects are more pronounced at the lower tail of the conditional distribution of energy consumption. Smart thermostats, in addition of providing real time information on consumption and pricing, allow households to respond to price changes in advance by programming temperature settings for different times of the day. We also find that treated households appear to reduce overall consumption as a result of these technologies relative to the control group, but the average response does not truly summarize the distributional effect of the technologies. We also investigate the long-run effect of a change in energy price for different enabling technologies across different age and income groups.

The paper is organized as follows. The next section introduces the model and the proposed estimator. It also establishes the asymptotic properties of the estimator. Section 3 provides simulation experiments to investigate the small sample performance of the proposed estimator. Section 4 demonstrates how the estimator can be used in practice by exploring an application of electricity pricing and smart technology. Section 5 concludes. Mathematical proofs are provided in the Appendix and additional Monte Carlo results are offered in an online Supplement.

Notations: Generic positive finite constants are denoted by $K_a, K_b, \ldots$, and can take different values at different instances and are bounded in N and T (the panel dimensions). The largest and the smallest eigenvalues of the $N \times N$ real symmetric matrix $\mathbf{A} = (a_{ij})$ are denoted by $\zeta_{\max}(\mathbf{A})$ and $\zeta_{\min}(\mathbf{A})$, respectively, and its spectral (or operator) norm by $\|\mathbf{A}\| = \zeta_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$. $\xrightarrow{a.s.}$ denotes almost sure convergence, $\xrightarrow{\ell_1}$ convergence in the ℓ_1 norm, $\xrightarrow{p}$ convergence in probability, and $\xrightarrow{d}$ convergence in distribution. We denote $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$ as the ℓ_1 norm of vector $\mathbf{x}$. All asymptotics are carried out under N and $T \to \infty$, jointly.

2. Model and assumptions

We consider a dynamic panel data model for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where $y_{it} \in \mathbb{R}$ is the response variable for cross-sectional unit i at time t and y_{it-1} denotes a lagged dependent variable. Consider the following conditional panel quantile function:

$$Q_{Y_{it}}(\tau | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \boldsymbol{f}_t) = \alpha_i(\tau) + \lambda_i(\tau) y_{it-1} + \mathbf{x}_{it}' \boldsymbol{\beta}_i(\tau) + \mathbf{f}_t' \boldsymbol{\gamma}_i(\tau), \tag{2.1}$$

where τ is a quantile in the interval $(0,1)$, $\boldsymbol{\theta}_i(\tau) = (\alpha_i(\tau), \lambda_i(\tau), \boldsymbol{\beta}_i'(\tau), \boldsymbol{\gamma}_i'(\tau))'$ and the conditional quantile function is defined as $Q_{Y_{it}}(\tau | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \mathbf{f}_t) := \inf\{y : P(Y_{it} \leq y | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \mathbf{f}_t) \geq \tau\}$. The variable $\mathbf{x}_{it}$ is a $p_x \times 1$ vector of regressors specific to cross-sectional unit i at time t, $\boldsymbol{\beta}_i(\tau)$ is the associated regression coefficients, $\mathbf{f}_t$ is an $r \times 1$ vector of unobserved factors, $\boldsymbol{\gamma}_i(\tau)$ is a vector of latent factor loadings, and $\alpha_i(\tau)$ is an individual effect potentially correlated with the regressor variables, $\mathbf{x}_{it}$. The term $\mathbf{f}_t' \boldsymbol{\gamma}_i(\tau)$ can be interpreted as a quantile-specific function capturing unobserved heterogeneity that was not adequately controlled by the inclusion of $\mathbf{x}_{it}$.

The model can be considered to be semi-parametric since the functional form of the conditional distribution of Y_{it} given $(y_{it-1}, \mathbf{x}_{it}', \boldsymbol{\theta}_i(\tau), \mathbf{f}_t')'$ is left unspecified and no parametric assumption is imposed on the relation between the regressors and the latent variables in the model. The $p_x \times 1$ vector of regressors is assumed to follow the general linear process

$$\mathbf{x}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\Gamma}_i' \mathbf{f}_t + \boldsymbol{v}_{it}, \tag{2.2}$$

where $\boldsymbol{\mu}_i$ is an individual effect, $\boldsymbol{\Gamma}_i$ is a $r \times p_x$ matrix of factor loadings in the $\mathbf{x}_{it}$ equation, and $\boldsymbol{v}_{it}$ is a p_x-dimensional vector assumed to follow a stationary process independently distributed of other variables in the model.

Naturally, model (2.1) can accommodate additional lags of the dependent variable, deterministic trends, time-invariant covariates, and lags of the exogenous covariates. These variations can be incorporated at a cost of additional notational complexity. The conditional panel quantile model (2.1) is fairly general and includes several recent panel data models as special cases:

Example 1. Let $u_{it} := y_{it} - \alpha_i - \lambda_i y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i - \mathbf{f}'_t\boldsymbol{\gamma}_i$ be an identically independently distributed (i.i.d.) Gaussian random variable with the cumulative distribution function denoted by Φ_u. Then, if $\tau = 0.5$, the conditional median function in (2.1) becomes,

$$Q_{Y_{it}}(0.5|y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \boldsymbol{f}_t) = E(y_{it}|y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i, \boldsymbol{f}_t), \tag{2.3}$$

since $\alpha_i(\tau) = \alpha_i + Q_u(\tau) = \alpha_i + \Phi_u^{-1}(0.5) = \alpha_i$, $\lambda_i(\tau) = \lambda_i$, $\boldsymbol{\beta}_i(\tau) = (\beta_{i1}, \ldots, \beta_{ip_x})' = \boldsymbol{\beta}_i$, and $\boldsymbol{\gamma}_i(\tau) = \boldsymbol{\gamma}_i$. Estimation of the conditional mean model (2.3) is discussed in a series of recent papers by Chudik and Pesaran (2015) and Chudik, Mohaddes, Pesaran and Raissi (2017). The conditional mean model $E(y_{it}|\mathbf{x}_{it}, \boldsymbol{\theta}_i, \boldsymbol{f}_t)$ is investigated in Pesaran (2006) and Bai (2009).

Example 2. Galvao (2011) proposes an instrumental variable approach for estimation of a dynamic quantile regression model when both $\boldsymbol{\gamma}_i(\tau) = \mathbf{0}$ and $\boldsymbol{\Gamma}_i = \mathbf{0}$. If $\boldsymbol{\gamma}_i(\tau) = \mathbf{0}$, $\boldsymbol{\beta}_i(\tau) = \boldsymbol{\beta}(\tau)$, and $\boldsymbol{\lambda}_i(\tau) = \boldsymbol{\lambda}(\tau)$ for $1 \leq i \leq N$, model (2.1) becomes the panel data model studied by Galvao (2011):

$$Q_{Y_{it}}(\tau|y_{it-1}, \mathbf{x}_{it}, \alpha_i(\tau)) = \alpha_i(\tau) + \lambda(\tau)y_{it-1} + \mathbf{x}'_{it}\boldsymbol{\beta}(\tau). \tag{2.4}$$

Example 3. Harding and Lamarche (2014) propose a quantile regression approach for the estimation of model (2.1) with homogeneous slope coefficients in the static case where $\lambda_i(\tau) = 0$ for $1 \leq i \leq N$. If $\lambda_i(\tau) = 0$ and $\boldsymbol{\beta}_i(\tau) = \boldsymbol{\beta}(\tau)$ for $1 \leq i \leq N$, then equation (2.1) becomes the panel quantile function studied by Harding and Lamarche (2014):

$$Q_{Y_{it}}(\tau|\mathbf{x}_{it}, \tilde{\boldsymbol{\gamma}}_i(\tau), \tilde{\boldsymbol{f}}_t) = \mathbf{x}'_{it}\boldsymbol{\beta}(\tau) + \tilde{\mathbf{f}}'_t\tilde{\boldsymbol{\gamma}}_i(\tau), \tag{2.5}$$

where $\tilde{\mathbf{f}}_t = (1, \mathbf{f}'_t)'$ and $\tilde{\boldsymbol{\gamma}}_i(\tau) = (\alpha_i(\tau), \boldsymbol{\gamma}_i(\tau)')'$. Moreover, if $\mathbf{f}_t = 1$ for $1 \leq t \leq T$, then equation (2.5) becomes the model studied by Koenker (2004) and Lamarche (2010), where $Q_{Y_{it}}(\tau|\mathbf{x}_{it}, a_i) = a_i + \mathbf{x}'_{it}\boldsymbol{\beta}(\tau)$, and $a_i = \alpha_i + \gamma_i$.

Example 4. Consider for simplicity a panel version of the quantile autoregressive model introduced by Koenker and Xiao (2006) with one lagged dependent variable: $y_{it} = \theta_{i0}(\upsilon_{it}) + \theta_i(\upsilon_{it})y_{it-1}$, where

υ_{it} is a standard uniform random variable. This last equation leads to the same conditional quantile equation (2.1) when the coefficients of $\mathbf{x}_{it}$ are set to zero, and $\boldsymbol{\gamma}_i(\tau) = \mathbf{0}$ for all $1 \leq i \leq N$.

Due to the combination of cross-sectional error dependence ($\boldsymbol{\gamma}_i \neq \mathbf{0}$), and dynamics ($\lambda_i \neq 0$) in equation (2.1), existing panel quantile regression approaches are inconsistent for the estimation of $(\lambda_i, \boldsymbol{\beta}_i')'$ for $i = 1, \ldots, N$. In this paper, we are interested in estimating the contemporaneous effect of a change in $\mathbf{x}_{it}$ on the quantiles of the conditional distribution of the response variable as well as its long run effect. For instance, in Section 4, we estimate an autoregressive panel quantile model for energy consumption with interactive effects. Our primary focus is to identify and estimate the effect of different technologies that enable households to respond to time-of-use pricing on energy consumption, focusing on the distributional effect of the assigned technologies.

2.1. **Estimation**

We consider consistent estimation of the parameters of interest by estimating the dynamic quantile regression model with interactive effects defined by (2.1). To this end, we make the the following assumptions:

Assumption 1. *For all $0 < \tau < 1$, the conditional quantile function in equation (2.1) satisfies $P(u_{it}(\tau) \leq 0 | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \boldsymbol{f}_t) = \tau$, where $u_{it}(\tau) := y_{it} - Q_{Y_{it}}(\tau | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \boldsymbol{f}_t)$ is identically and independently distributed over i and identically distributed over t, conditional on $(y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \boldsymbol{f}_t)$.*

Assumption 2. *The $r \times 1$ vector of common factors $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{rt})'$ is a covariance stationary process with absolute summable autocovariances, distributed independently of $u_{it}(\tau)$ and $\boldsymbol{v}_{it}$ for all i, t, and τ.*

Assumption 3. *The factor loadings $\boldsymbol{\gamma}_i(\tau) = \boldsymbol{\gamma}(\tau) + \boldsymbol{\eta}_{\gamma i}$ and $vec[\boldsymbol{\Gamma}_i] = vec[\boldsymbol{\Gamma}] + \boldsymbol{\eta}_{\Gamma i}$ are distributed independently of $u_{jt}(\tau)$ and $\boldsymbol{v}_{jt}$ for all i and j with means $\boldsymbol{\gamma}(\tau)$ and $\boldsymbol{\Gamma}(\tau)$, and bounded variances. The error terms $\boldsymbol{\eta}_{\gamma i}$ and $\boldsymbol{\eta}_{\Gamma i}$ are independent of each other. Moreover, these random variables are independently and identically distributed over i with zero means and covariances $\boldsymbol{\Omega}_\gamma$ and $\boldsymbol{\Omega}_\Gamma$, respectively, with $\|\boldsymbol{\Omega}_\gamma\| < K$ and $\|\boldsymbol{\Omega}_\Gamma\| < K$.*

Assumption 4. *The variables $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, \ldots, x_{it,p_x})' \in \mathcal{X} \subseteq R^{p_x}$ and $u_{it}(\tau)$ are independently distributed. The regressors $\mathbf{x}_{it}$ are generated according to equation (2.2), and the vector of errors $\boldsymbol{v}_{it}$ in (2.2) follows a stationary process with mean zero, finite covariance matrix, and finite fourth order cumulants, and summable autocovariances (uniformly in i).*

Assumption 5. *The p_x+1-dimensional vector of slope coefficients $\boldsymbol{\vartheta}_i(\tau)=[\lambda_i(\tau),\boldsymbol{\beta}_i'(\tau)]'$ follows the random coefficient representation:*

$$\begin{aligned}\lambda_i(\tau) &= \lambda(\tau)+(1-|\lambda(\tau)|)\nu_{i\lambda} \\ \boldsymbol{\beta}_i(\tau) &= \boldsymbol{\beta}(\tau)+\nu_{i\beta},\end{aligned} \tag{2.6}$$

where $\boldsymbol{\beta}(\tau)<K$, $\sup_i|\nu_{i\lambda}|<1$, and $|\lambda(\tau)|<1$ for all $\tau\in(0,1)$, and

$$\boldsymbol{\nu}_i=\begin{pmatrix}\nu_{i\lambda}\\ \nu_{i\beta}\end{pmatrix}\sim IID\left(\mathbf{0},\boldsymbol{\Omega}_\vartheta\right), \tag{2.7}$$

with $\|\boldsymbol{\Omega}_\vartheta\|<K$, $\boldsymbol{\Omega}_\vartheta$ is a symmetric positive definite matrix. Furthermore, for each $\tau\in(0,1)$,

$$E(\lambda_i^l(\tau)\alpha_i(\tau)|\mathcal{F}_t)=a_l(\tau),\ E(\lambda_i^l(\tau)\boldsymbol{\beta}_i(\tau)|\mathcal{F}_t)=\mathbf{b}_l(\tau),\ E(\lambda_i^l(\tau)\boldsymbol{\gamma}_i(\tau)|\mathcal{F}_t)=\mathbf{c}_l(\tau), \tag{2.8}$$

for $l=0,1,2,....$ where $\mathcal{F}_t=(\mathbf{f}_t,\mathbf{f}_{t-1},\ldots;\mathbf{x}_{it},\mathbf{x}_{it-1},\ldots,i=1,2,\ldots,N)$, and $a_l(\tau)$, $\mathbf{b}_l(\tau)$ and $\mathbf{c}_l(\tau)$ are exponentially decaying in l, such that $|a_l(\tau)|<K_a\rho^l$, $\|\mathbf{b}_l(\tau)\|<K_b\rho^l$, and $\|\mathbf{c}_l(\tau)\|<K_c\rho^l$ for some positive $\rho<1$. The parameters $\lambda_i(\tau)$ and $\boldsymbol{\beta}_i(\tau)$ are independently distributed over i, and $\boldsymbol{\nu}_i$ is independently distributed of $\boldsymbol{\gamma}_i(\tau)$, $\boldsymbol{\Gamma}_i(\tau)$, $u_{it}(\tau)$, $\boldsymbol{v}_{it}'$, and $\mathbf{f}_t$ for all i, t and τ.

Assumption 6. *Let $\mathbf{C}(\tau)=E(\mathbf{C}_i(\tau))=(\boldsymbol{\gamma}(\tau),\boldsymbol{\Gamma})'$, and suppose that $p_x\geq r-1$, and the $(p_x+1)\times r$ dimensional matrix $\mathbf{C}(\tau)$ has full column rank, for all values of $0<\tau<1$.*

Assumption 1 is similar to Assumption A3 in Ando and Bai (2017) and Assumption 4.iii in Chen, Dolado and Gonzalo (2017). The assumption is slightly weaker than other assumptions in the literature since it can allow for forms of serial dependence, as explicitly stated later in Assumptions 9 and 11. In general, the other assumptions are similar to those in Pesaran (2006) and Chudik and Pesaran (2015). One key difference is that we require certain conditions on the quantile coefficients. Another difference is that it is common to assume that there exists an N-dimensional vector of non-stochastic weights that satisfy granularity conditions, namely that they are of order N^{-1}. Such effects are important in small samples but do not affect the asymptotic results established below in Section 2.2. Therefore, without loss of generality, we consider the case of equal weights $1/N$. Assumption 5 introduces heterogeneous slope coefficients assuming that deviations of $\boldsymbol{\vartheta}_i(\tau)$ with respect to $\boldsymbol{\vartheta}(\tau)$ are mean-zero random variables independently distributed of other variables in the model. Specification (2.6) ensures that $\sup_{i,\tau}|\lambda_i(\tau)|<1$, so long as $\sup_i|\nu_{i\lambda}|<1$, and $|\lambda(\tau)|<1$ for all $\tau\in(0,1)$. A convenient distribution for $\nu_{i\lambda}$ is a beta distribution defined on $(0,1)$. The moment conditions in (2.8) are required for consistent estimation of $\mathbf{f}_t$ (up to a non-singular $r\times r$ transformation) from cross section averages of $\mathbf{z}_{it}=(y_{it},\mathbf{x}_{it}')'$ and their lagged values. These conditions are met when $\lambda_i(\tau)$ is independently distributed of $\alpha_i(\tau),\boldsymbol{\beta}_i(\tau)$ and $\boldsymbol{\gamma}_i(\tau)$ and $E\left(\lambda_i^l(\tau)\right)$

decays exponentially in l. This last condition is met, for example, if $\lambda_i(\tau)$ is distributed over i uniformly on $[-b, b]$ for any b in $0 < b < 1$.

Moreover, it is worth mentioning that the full rank Assumption 6 ensures the large N representation of the unobserved factors. Under Assumption 1, we write,

$$y_{it} = \alpha_i(\tau) + \lambda_i(\tau) y_{it-1} + \mathbf{x}'_{it}\boldsymbol{\beta}_i(\tau) + \mathbf{f}'_t\boldsymbol{\gamma}_i(\tau) + u_{it}(\tau), \tag{2.9}$$

where $u_{it}(\tau)$ is a random variable whose τ-th conditional quantile is equal to zero, and it is assumed that y_{it} has started a long time in the past. Note that by Assumption 5, $(1 - \lambda_i L)$ is invertible for all $i = 1, 2, \dots, N$, where L is the lag operator. Then equation (2.9), after pre-multiplying by $(1 - \lambda_i L)^{-1}$, can be written as,

$$y_{it} = \sum_{l=0}^{\infty} \lambda_i^l(\tau)\alpha_i(\tau) + \sum_{l=0}^{\infty} \lambda_i^l(\tau)\boldsymbol{\beta}'_i(\tau)\mathbf{x}_{it-l} + \sum_{l=0}^{\infty} \lambda_i^l(\tau)\boldsymbol{\gamma}'_i(\tau)\mathbf{f}_{t-l} + \sum_{l=0}^{\infty} \lambda_i^l(\tau) u_{it-l}(\tau). \tag{2.10}$$

We now derive a large N representation for a linear combination of the latent factors following Pesaran (2006) and Chudik and Pesaran (2015). Denote the last term of the above equation by $\xi_{it}(\tau)$, and note that it can be written as $\xi_{it}(\tau) = \lambda_i(\tau)\xi_{it-1}(\tau) + u_{it}(\tau)$, which is a stationary AR(1) process for all $1 \leq i \leq N$, since by Assumption 5 $\sup_{i,\tau} |\lambda_i(\tau)| \leq \rho < 1$. Also, since for each t and τ, the errors, $u_{it}(\tau)$, and $\lambda_i(\tau)$ are assumed to be cross-sectionally independent, it then readily follows that (see Pesaran (2006))

$$\bar{\xi}_t(\tau) = N^{-1} \sum_{i=1}^{N} \xi_{it}(\tau) = O_p(N^{-1/2}).$$

Similarly, consider the cross section averages of the other terms of (2.10), and note that under Assumption 5, for the first term we have (recall that by Assumption 5 $\{a_l\}$ is absolute summable)

$$\sum_{l=0}^{\infty} \left[N^{-1} \sum_{i=1}^{N} \lambda_i^l(\tau)\alpha_i(\tau) \right] = \sum_{l=0}^{\infty} a_l(\tau) + O_p(N^{-1/2}).$$

Similarly, conditional on $\mathcal{F}_t$ we have (noting that by Assumption 5 $\mathbf{b}_l(\tau)$ and $\mathbf{c}_l(\tau)$ are absolute summable)

$$\begin{aligned}
\sum_{l=0}^{\infty} \left[N^{-1} \sum_{i=1}^{N} \lambda_i^l(\tau)\boldsymbol{\beta}'_i(\tau)\mathbf{x}_{it-l} \right] &= \sum_{l=0}^{\infty} \mathbf{b}'_l(\tau)\bar{\mathbf{x}}_{t-l} + O_p(N^{-1/2}), \\
\sum_{l=0}^{\infty} \left[N^{-1} \sum_{i=1}^{N} \lambda_i^l(\tau)\boldsymbol{\gamma}'_i(\tau) \right] \mathbf{f}_{t-l} &= \sum_{l=0}^{\infty} \mathbf{c}'_l(\tau)\mathbf{f}_{t-l} + O_p(N^{-1/2}).
\end{aligned}$$

Hence, overall

$$\bar{y}_t = a(1,\tau) + \mathbf{b}(L,\tau)'\bar{\mathbf{x}}_t + \mathbf{c}(L,\tau)'\mathbf{f}_t + O_p(N^{-1/2}), \tag{2.11}$$

where $a(1,\tau) = \sum_{l=0}^{\infty} a_l(\tau), \mathbf{b}(L,\tau) = \sum_{l=0}^{\infty} \mathbf{b}_l(\tau)L^l$, and $\mathbf{c}(L,\tau) = \sum_{l=0}^{\infty} \mathbf{c}_l(\tau)L^l$.

Similarly, taking cross-sectional averages of equation (2.2), we obtain,

$$\bar{\mathbf{x}}_t = \bar{\boldsymbol{\mu}} + \boldsymbol{\Gamma}'\mathbf{f}_t + O_p(N^{-1/2}), \tag{2.12}$$

where $\bar{\boldsymbol{\mu}} = N^{-1}\sum_{i=1}^{N} \boldsymbol{\mu}_i$ and $\boldsymbol{\Gamma} = E(\boldsymbol{\Gamma}_i)$. See also Assumptions 5 and 6. Combining (2.11) and (2.12), we have

$$\mathbf{C}(L,\tau)\mathbf{f}_t = \boldsymbol{\Lambda}(L,\tau)\bar{\mathbf{z}}_t - \mathbf{d}(\tau) + O_p(N^{-1/2}), \tag{2.13}$$

where $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{\mathbf{x}}_t')'$,

$$\mathbf{d}(\tau) = \begin{pmatrix} a(1,\tau) \\ \bar{\boldsymbol{\mu}} \end{pmatrix}, \mathbf{C}(L,\tau) = \begin{pmatrix} \mathbf{c}(L,\tau)' \\ \boldsymbol{\Gamma}' \end{pmatrix}, \boldsymbol{\Lambda}(L,\tau) = \begin{pmatrix} 1 & -\mathbf{b}(L,\tau)' \\ 0 & \mathbf{I}_{p_x} \end{pmatrix}.$$

Pre-multiplying both sides of (2.13) by $\mathbf{C}(L,\tau)'$ and assuming that rank of $\mathbf{C}(L,\tau)$ is equal to the number of factors r we obtain the following result for $\mathbf{f}_t$:

$$\mathbf{f}_t = \mathbf{f}_0(\tau) + \mathbf{G}(L,\tau)\bar{\mathbf{z}}_t + O_p(N^{-1/2}), \tag{2.14}$$

where $\mathbf{f}_0(\tau) = -(\mathbf{C}(1,\tau)'\mathbf{C}(1,\tau))^{-1}\mathbf{C}(1,\tau)'\mathbf{d}(\tau)$ and $\mathbf{G}(L,\tau) = (\mathbf{C}(L,\tau)'\mathbf{C}(L,\tau))^{-1}\mathbf{C}(L,\tau)'\boldsymbol{\Lambda}(L,\tau)$ is an $r \times (p_x + 1)$ distributed lag matrix. Integrating out τ on the right hand side of (2.14) and defining $\mathbf{f}_0 = \int_0^1 \mathbf{f}_0(\tau)d\tau$ and $\mathbf{G}(L) = \int_0^1 \mathbf{G}(L,\tau)d\tau$, we have that

$$\mathbf{f}_t = \mathbf{f}_0 + \mathbf{G}(L)\bar{\mathbf{z}}_t + O_p(N^{-1/2}). \tag{2.15}$$

Finally, substituting the representation of the factors in equation (2.9), we obtain

$$y_{it} = \beta_{0i}(\tau) + \lambda_i(\tau)y_{it-1} + \boldsymbol{\beta}_i'(\tau)\mathbf{x}_{it} + \boldsymbol{\delta}_i(L,\tau)'\bar{\mathbf{z}}_t + u_{it}(\tau) + O_p(N^{-1/2}), \tag{2.16}$$

where $\beta_{0i}(\tau) = \alpha_i(\tau) + \boldsymbol{\gamma}_i'(\tau)\mathbf{f}_0$, $\boldsymbol{\delta}_i(L,\tau) = \boldsymbol{\gamma}_i'(\tau)\sum_{l=0}^{\infty}\mathbf{G}_l L^l = \sum_{l=0}^{\infty}\boldsymbol{\delta}_{il}(\tau)L^l$, $\boldsymbol{\delta}_{il}(\tau) = (\boldsymbol{\delta}_{iy,l}'(\tau), \boldsymbol{\delta}_{ix,l}'(\tau))'$, $\boldsymbol{\delta}_{iy,l}(\tau)$ is a reduced form coefficient for the cross-sectional average of y_{it-l}, $\boldsymbol{\delta}_{ix,l}(\tau)$ is a reduced form coefficient for the cross-sectional average of $\mathbf{x}_{it-l}$, and $\bar{\mathbf{z}}_{t-l} = (\bar{y}_{t-l}, \bar{\mathbf{x}}_{t-l}')'$ is a $(p_x + 1) \times 1$ dimensional vector.

Remark 1. Since $\mathbf{f}_0$ is not identified and its value can be absorbed in the intercept term of equations (2.11) and (2.12), in what follows, and without loss of generality, we set $\mathbf{f}_0 = \mathbf{0}$, and note that under this normalization $\beta_{0i}(\tau) = \alpha_i(\tau)$.

Assumption 7. *The infinite order distributed lag matrix function* $\mathbf{G}(L) = \mathbf{G}_0 + \mathbf{G}_1 L + \ldots = \sum_{l=0}^{\infty}\mathbf{G}_l L^l$, *where* $\|\mathbf{G}_l\| < K\rho^l$ *for all* l *and some positive* $\rho < 1$ *and constant* $K > 0$.

Assumption 7 follows from the exponential decay condition stated in Assumption 5 (see Lemma A.1 in Chudik and Pesaran (2013)). Recall that $\mathbf{G}(L)$ is an infinite order distributed lag matrix function with exponentially decaying coefficients and hence can be suitably truncated as

$$y_{it} = \alpha_i(\tau) + \lambda_i(\tau)y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta}_i(\tau) + \sum_{l=0}^{p_T} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau) + e_{it}(\tau), \tag{2.17}$$

where

$$e_{it}(\tau) = u_{it}(\tau) + \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau) + O_p(N^{-1/2}). \tag{2.18}$$

Note that by Assumption 7, $\|\boldsymbol{\delta}_{il}(\tau)\| < K\rho^l$ with $0 < \rho < 1$, because $\|\boldsymbol{\gamma}_i(\tau)\| < K$ as implied by Assumption 3. It follows that,

$$\left\| \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau) \right\| \leq K\rho^{p_T} \sum_{l=1}^{\infty} \|\bar{\mathbf{z}}_{t-l}\| \rho^l,$$

and by Lemma A.4. (result (A.18)) in Chudik and Pesaran (2015), this remainder term becomes asymptotically negligible as $N, T \to \infty$.

The total number of parameters for the augmented part of (2.17) is $(p_x + 1)(p_T + 1)$. The error term $e_{it}(\tau)$ includes $u_{it}(\tau)$, a term $O_p(N^{-1/2})$ associated with approximating $\mathbf{f}_t$ with cross-section averages, and an error component due to the truncation of the underlying infinite order distributed lag function $\boldsymbol{\delta}_i(\tau, L)$. Moreover, the number of lags is denoted by p_T and it is assumed that $p_{T_i} = p_T$ for all i for the simplicity of exposition. It is also assumed that the number of lags to approximate the factors is known and that $E(\lambda_i^l)$ decays exponentially which is satisfied by Assumption 5, and $\boldsymbol{\beta}_i$ and λ_i are independently distributed, although this is not required.

Equation (2.16) leads to a conditional quantile function that is naturally different than equation (2.1) since $\mathbf{f}_t$ is unknown and we use a large N representation for $\mathbf{f}_t$. The following condition is needed for identification of the parameter of interest $(\lambda_i(\tau), \boldsymbol{\beta}_i'(\tau))'$ in equations (2.9) and (2.17).

Assumption 8. *Consider* (2.9) *and its approximate version* (2.17)*, let* $\mathbf{W}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \mathbf{f}_t')'$ *and* $\mathbf{X}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \bar{\mathbf{z}}_t', \bar{\mathbf{z}}_{t-1}', \ldots, \bar{\mathbf{z}}_{t-p_T}')'$*, and define* $\boldsymbol{\delta}_i(\tau) := (\boldsymbol{\delta}_{i1}(\tau)', \boldsymbol{\delta}_{i2}(\tau)', \ldots, \boldsymbol{\delta}_{ip_T}(\tau)')'$*. Then for all* $\tau \in (0,1)$*,* $(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau)) \in$ *int* $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{G}$*, a compact and convex set, and* $(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)) \in$ *int* $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$*, which is compact and convex. Also define,*

$$\begin{aligned}
\boldsymbol{\Pi}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau)) &:= E(\mathbf{W}_{it}\psi_\tau(y_{it} - \lambda_i(\tau)y_{it-1} - \mathbf{x}_{it}'\boldsymbol{\beta}_i(\tau) - \alpha_i(\tau) - \mathbf{f}_t'\boldsymbol{\gamma}_i(\tau))), \\
\mathbf{J}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau)) &:= \frac{\partial}{\partial(\lambda_i(\tau), \boldsymbol{\beta}_i'(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i'(\tau))}\boldsymbol{\Pi}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau)),
\end{aligned}$$

where $\psi_\tau(u) = \tau - I(u \leq 0)$ *is the quantile influence function, and*

$$\begin{aligned}\tilde{\mathbf{\Pi}}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)) \quad &:= \quad E\left(\mathbf{X}_{it}\psi_\tau(y_{it} - \lambda_i(\tau)y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i(\tau) - \alpha_i(\tau) - \sum_{l=0}^{p_T} \bar{\mathbf{z}}'_{t-l}\boldsymbol{\delta}_{il}(\tau))\right), \\ \tilde{\mathbf{J}}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)) \quad &= \quad \frac{\partial}{\partial(\lambda_i(\tau), \boldsymbol{\beta}'_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}'_i(\tau))}\tilde{\mathbf{\Pi}}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)).\end{aligned}$$

The Jacobian $\mathbf{J}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau))$ *is continuous and full rank uniformly over* $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{G}$*, and the Jacobian* $\tilde{\mathbf{J}}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau))$ *is continuous and full rank uniformly over* $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$*. The image of the parameter spaces* $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{G}$ *and* $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$ *are simply connected under the mappings* $(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau)) \mapsto \mathbf{\Pi}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\gamma}_i(\tau))$ *and* $(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)) \mapsto \tilde{\mathbf{\Pi}}(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau), \alpha_i(\tau), \boldsymbol{\delta}_i(\tau))$*.*

The first part of Assumption 8 imposes compactness over the parameter space and it can be relaxed since the quantile objective functions corresponding to equations (2.9) and its approximate version (2.17) are convex in parameters. The second part is an identification condition that requires full rank and continuity as in Chernozhukov and Hansen (2006) and Harding and Lamarche (2014). It implies global identification of the parameters $(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau)')'$ for all $\tau \in (0, 1)$. It differs from those conditions in Chernozhukov and Hansen (2006) and Harding and Lamarche (2014) in that they reflect the specific nature of the identification problem in a panel quantile model with latent factors. The last part of the assumption requires that the image of the parameter space $\Lambda\times\mathcal{B}\times\mathcal{A}\times\mathcal{G}$ and the image of the parameter space $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$ are homotopic to a point, ruling out the possibility of holes in the image of the sets.

The following theorem describes identification of the parameters $(\lambda_i(\tau), \boldsymbol{\beta}_i(\tau)')'$ in a quantile regression model augmented with cross-sectional averages. The proof is presented in Appendix A.

Theorem 1 (Identification of $\boldsymbol{\vartheta}_i(\tau)$)**.** *Under Assumptions 1-8, the parameter of interest* $\boldsymbol{\vartheta}_i(\tau) = (\lambda_i(\tau), \boldsymbol{\beta}_i(\tau)')'$ *is identified in equation* (2.9) *and equation* (2.17) *for each* τ*.*

We now present an approach that can be used to estimate a dynamic quantile regression model with interactive effects. The quantile regression procedure is similar in spirit to Pesaran (2006), Harding and Lamarche (2014) and Chudik and Pesaran (2015). Define the parameter $\boldsymbol{\pi}_i(\tau) := (\lambda_i(\tau), \boldsymbol{\beta}_i(\tau)', \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)')'$ with $\boldsymbol{\delta}_i(\tau) = (\boldsymbol{\delta}_{i1}(\tau)', \boldsymbol{\delta}_{i2}(\tau)', \ldots, \boldsymbol{\delta}_{ip_T}(\tau)')'$, and

$$C_{it}(\tau, \boldsymbol{\pi}_i) = \rho_\tau\left(y_{it} - \lambda_i(\tau)y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i(\tau) - \alpha_i(\tau) - \sum_{l=0}^{p_T} \bar{\mathbf{z}}'_{t-l}\boldsymbol{\delta}_{il}(\tau)\right), \tag{2.19}$$

where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the standard quantile regression loss function. First, we minimize the individual specific objective function (2.19) for $\boldsymbol{\pi}_i(\tau)$,

$$\hat{\boldsymbol{\pi}}_i(\tau) = \arg \min_{\boldsymbol{\pi}_i \in \boldsymbol{\Pi}_i} \sum_{t=p_T+1}^{T} C_{it}(\tau, \boldsymbol{\pi}_i), \tag{2.20}$$

where $\boldsymbol{\Pi}_i$ is compact set in $\mathbb{R}^{(p_x+1)(p_T+2)}$. Therefore, the quantile regression estimator for heterogeneous effects in a dynamic panel quantile with interactive effects, $\hat{\boldsymbol{\pi}}_i(\tau)$, is based on the cross-sectionally augmented regression (2.17). We also propose a quantile mean group estimator for $\boldsymbol{\vartheta}(\tau) := E((\lambda_i(\tau), \boldsymbol{\beta}_i(\tau)')')$. The estimator is,

$$\hat{\boldsymbol{\vartheta}}(\tau) = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\vartheta}}_i(\tau) = \frac{1}{N} \sum_{i=1}^{N} \left(\boldsymbol{\Xi}_i \circ \hat{\boldsymbol{\pi}}_i(\tau) \right), \tag{2.21}$$

where $\circ$ denotes Hadamard product, $\boldsymbol{\Xi}_i = (\boldsymbol{\iota}_i', \mathbf{0}_i')'$ with $\boldsymbol{\iota}_i$ denoting a $p_x + 1$ dimensional vector of ones and $\mathbf{0}_i$ a $(p_x + 1)(p_T + 1)$ dimensional vector of zeros. We denote the estimator defined in (2.21) as quantile common correlated effects mean group estimator, QCCEMG. In what follows, for convenience, we shorten the label simply to QMG. One could also consider a pooled version, the common correlated effects pooled estimator proposed in Pesaran (2006). We can consider a weighted average of the individual estimates with weights defined by the covariance matrix of $\hat{\boldsymbol{\pi}}_i(\tau)$.

The interpretation of the estimator defined in (2.21) is associated with heterogeneous coefficients modeled as $\boldsymbol{\vartheta}_i(\tau) = \boldsymbol{\vartheta}(\tau) + \boldsymbol{\nu}_i$, where $\boldsymbol{\nu}_i$ is a mean-zero error term independent of the regressors. We are interested in $\boldsymbol{\vartheta}(\tau)$, which motivates the average. Large N helps to understand the average restriction and recover the parameter of interest. Furthermore, note that we need a panel with large T, because of the short T bias involved in estimating quantile regressions with lagged dependent variables, and the fact that we are approximating $\mathbf{f}_t$ by current and past values of cross section averages, $\bar{\mathbf{z}}_t$, and we need both N and T to be large for this purpose.

2.2. Asymptotic Theory

This section investigates the large sample properties of the proposed quantile estimator and its mean group counterpart defined by equations (2.20) and (2.21), respectively. Throughout this section, we set $\mathbf{X}_{it} = (y_{it-1}, \mathbf{x}_{it}', \ddot{\mathbf{z}}_t')'$, where $\ddot{\mathbf{z}}_t' = (1, \bar{\mathbf{z}}_t', \bar{\mathbf{z}}_{t-1}', ..., \bar{\mathbf{z}}_{t-p_T}')'$, and write equation (2.17) as $y_{it} = \mathbf{X}_{it}'\boldsymbol{\pi}_i + e_{it}$. Also $\|\cdot\|_1$ stands for the ℓ_1-norm.

We consider the following regularity conditions for the consistency of the proposed estimators:

Assumption 9. *The vector* $\{(\mathbf{X}_{it}', \mathbf{f}_t', y_{it}) : t = 1, 2, \ldots\}$ *is stationary and independent across* i*, and* β*-mixing time series with* β*-mixing coefficients* $\beta_i(j)$*. Then, there exists constants* $a \in (0, 1)$ *and* $B > 0$ *such that* $\sup_{1 \leq i \leq N} \beta_i(j) \leq Ba^j$ *for all* $j \geq 1$*.*

Assumption 10. *There exist a series of constants independent of* i *and* τ *such that* $\sup_{i,\tau} \|\boldsymbol{\gamma}_i(\tau)\| < K_{\boldsymbol{\gamma}}$*,* $\sup_i \|\boldsymbol{\Gamma}_i\| < K_{\boldsymbol{\Gamma}}$*,* $\sup_i \|y_{i0}\| < K_y$*, and* $\sup_i \|u_{i1}\| < K_u$*, and additionally a constant* M_x *such that* $\sup_i \|\boldsymbol{X}_{it}\| \leq M_x$ *(a.s.). In addition,* $\inf_{i,\tau} \zeta_{\min} E\left[\boldsymbol{\gamma}_i(\tau)\boldsymbol{\gamma}_i'(\tau)\right] > 0$*, and* $\inf_{i \geq 1} \zeta_{\min}(E(\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i')) > 0$*.*

Assumption 11. *For each* $\eta > 0$*,*

$$\epsilon_\eta := \inf_i \inf_{\|\boldsymbol{\pi}\|_1 = \eta} E\left[\int_0^{\boldsymbol{X}_{i1}'\boldsymbol{\pi}} \left(G_i(s|\boldsymbol{X}_{i1}) - \tau\right) ds\right],$$

where G_i *is defined as a conditional distribution of* u_{it} *and the conditional densities* g_i *is continuous, uniformly bounded away from 0 and* ∞*, with continuous derivatives everywhere. Moreover, the joint distribution of* $(u_{i,1}, u_{i,1+j})$*,* $g_{i,j}(u_{i,1}, u_{i,1+j}|\mathbf{X}_{i,1}, \mathbf{X}_{i,1+j}) \leq C_f$ *with* $C_f > 0$*, uniformly over* $(u_{i,1}, u_{i,1+j}, \mathbf{X}_{i,1}, \mathbf{X}_{i,1+j})$ *for all* $i \geq 1$ *and* $j \geq 1$*.*

Assumption 12. *Let* $\mathbf{S}_{iT} = T^{-1}\sum_{t=1}^{T} \mathbf{X}_{it}\mathbf{X}_{it}'$ *and assume that there exists* T_0 *such that for all* $T > T_0$*,* $\inf_i \zeta_{\min}(\mathbf{S}_{iT}) > 0$*, and* $\sup_i \zeta_{\min}(\mathbf{S}_{iT}) > K$*, and* $\mathbf{S}_{iT} \xrightarrow{p} \mathbf{S}_i = E(\mathbf{X}_{it}\mathbf{X}_{it}')$*, such that* $\inf_i(\zeta_{\min}(\mathbf{S}_i)) > 0$*.*

Similar conditions are used in the literature. For instance, a version of Assumption 9 has been used in Hahn and Kuersteiner (2011), Kato, Galvao and Montes-Rojas (2012), and Galvao, Lamarche and Lima (2013). The condition allows for dependence across time, implying that we need to apply a Bernstein type inequality for β-mixing sequences (Corollary C.1. in Kato, Galvao and Montes-Rojas (2012)) rather than a Hoeffding's inequality to show weak consistency. See Lemma 2 in Appendix A. Assumption 9 is a high-level assumption since $\mathbf{X}_{it}$ includes $\bar{y}_{t-1}, \bar{y}_{t-2}, \ldots$, and it is not easy to verify from the basic Assumptions 1-6. Assumption 10 is needed for the consistency of the estimator and for obtaining a well-defined limiting distribution. It requires that the regressors are strictly bounded, with the implication that the support of the error distributions is bounded and all coefficients, including the factor loadings, are bounded too. In the case of homogeneous coefficients $\boldsymbol{\vartheta}_i(\tau) = \boldsymbol{\vartheta}(\tau)$ for all $1 \leq i \leq N$, it can be changed to $\sup_{i \geq 1} E[\|\mathbf{X}_{i1}\|] < \infty$ as indicated in Kato et al. (2012) or $\max_{it} \|y_{it}\| = O_p(\sqrt{NT})$ and $\max_{it} \|\mathbf{x}_{it}\| = O_p(\sqrt{NT})$ as in Galvao (2011), but new results on stochastic inequalities for non-i.i.d. cases are needed. Note that the last part of Assumption 10 implies, by Assumption 3, that $E(\boldsymbol{\gamma}_i(\tau)\boldsymbol{\gamma}_i'(\tau))$ and $E(\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i')$ are non-singular matrices that do not depend on i. Assumption 11 is an identification condition and is similar to Assumptions (A3) and (D2) in Kato, Galvao and Montes-Rojas (2012). The assumption also imposes conditions on the joint distributions because the data can be non-independently distributed. Lastly, Assumption 12 is standard in the quantile regression literature and it is analogous to Assumption 7.b and 7.c in Chudik and Pesaran (2015). It guarantees that the inverse of $E[g_i(0|\mathbf{X}_{it})\mathbf{X}_{it}\mathbf{X}_{it}']$ exists, and jointly with Assumption 11, it implies that these inverses are uniformly bounded across i.

The following result states the weak consistency of the estimator:

Theorem 2 (Uniform consistency of $\hat{\boldsymbol{\pi}}_i(\tau)$)**.** *Suppose the τ-th conditional quantile function of y_{it} for $i = 1, ..., N$ and $t = 1, ..., T$ is given by the panel data model* (2.1)-(2.2) *and Assumptions 1-12 hold. As N, T and p_T go jointly to infinity with $p_T^3/T \to 0$ and $(\log(N))^2/T \to 0$, the cross-section augmented quantile regression estimator, $\hat{\boldsymbol{\pi}}_i(\tau)$, defined by* (2.20)*, is consistent uniformly over $1 \leq i \leq N$.*

As suspected, different conditions lead to changes in Theorem 2. Under less general conditions in Assumption 9, (i.e., not allowing for time series dependence), an application of Hoeffding's inequality leads to a bound in Theorem 2 that is $O(\exp(-T)) = o(N^{-1})$ which is satisfied when $\log(N)/T \to 0$.

It is perhaps worth noting that $\boldsymbol{\pi}_i(\tau)$ is estimated by quantile regressions for each unit i separately, but we augment such quantile regressions with $\bar{\mathbf{z}}_t, \bar{\mathbf{z}}_{t-1}, \ldots, \bar{\mathbf{z}}_{t-p_T}$. For N sufficiently large, the consistency of quantile estimators for each unit i can be justified using standard (non-panel) results for quantile regressions. Thus, if N is fixed, then $\sqrt{T}(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau))$ converges in distribution to a mean zero random variable with covariance $\mathcal{V}$, under $T \to \infty$ and $p_T^3/T \to 0$. The form of the covariance matrix $\mathcal{V}$ depends on Assumption 9 and under i.i.d. conditions, the asymptotic covariance matrix is similar to the ones obtained in Koenker (2005). We need, however, $N \to \infty$ for consistency of our approach.

As discussed in Chudik and Pesaran (2015), the consistency of individual coefficients is not always necessary for the consistency of the mean group estimator. Our next result establishes the consistency of the QMG estimator.

Theorem 3 (Consistency of $\hat{\boldsymbol{\vartheta}}(\tau)$)**.** *Under the conditions of Theorem 2, as (N, p_T, T) go jointly to infinity with $p_T^3/T \to 0$ and $(\log(N))^2/T \to 0$, the mean quantile group estimator defined by* (2.21) *for a model with interactive effects is weakly consistent, namely for every $0 < \tau < 1$, $\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau) \xrightarrow{p} 0$.*

We now turn our attention to the asymptotic distribution of the proposed estimator. We consider the following additional regularity condition:

Assumption 13. *Let* $\mathbf{J}_i := E[g_i(0|\mathbf{X}_{it})\mathbf{X}_{it}\mathbf{X}_{it}']$, $\mathbf{D}_i := \frac{1}{\sqrt{T-p_T}}\sum_{t=1+p_T}^{T}\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_0)\mathbf{X}_{it}$, $\dot{\mathbf{D}}_i := \frac{1}{T-p_T}\sum_{t=1+p_T}^{T}\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_{i0})\mathbf{X}_{it}$, $\dot{\boldsymbol{\Xi}}_i := \boldsymbol{\Xi}_i\boldsymbol{\Xi}_i'$, $\mathbf{V}_i := Var(\mathbf{D}_i)$, $\dot{\mathbf{V}}_i := Var(\dot{\mathbf{D}}_i)$, *and* $\boldsymbol{\Omega}_\vartheta := Var(\boldsymbol{\vartheta}_i(\tau))$. *The following conditions hold:*

(a) *Let* $\mathbf{J}_N = N^{-1}\sum_{i=1}^{N}\dot{\boldsymbol{\Xi}}_i \circ \mathbf{J}_i$ *and* $\mathbf{V}_N = N^{-1}\sum_{i=1}^{N}\dot{\boldsymbol{\Xi}}_i \circ \mathbf{V}_i$. *The limit* $\mathbf{J} := \lim_{N\to\infty}\mathbf{J}_N$, $\mathbf{V} := \lim_{N\to\infty}\mathbf{V}_N$, *and* $\mathbf{V}_\psi := \lim_{N\to\infty}\mathbf{J}_N^{-1}\mathbf{V}_N\mathbf{J}_N^{-1}$ *exist and are non-singular.*
(b) *Let* $\dot{\mathbf{V}}_N = N^{-1}\sum_{i=1}^{N}\dot{\boldsymbol{\Xi}}_i \circ \dot{\mathbf{V}}_i$. *The limiting matrices* $\boldsymbol{\Omega}_\psi = \lim_{N\to\infty}\mathbf{J}_N^{-1}\dot{\mathbf{V}}_N\mathbf{J}_N^{-1}$ *and* $\mathbf{V}_v = \boldsymbol{\Omega}_\psi + \boldsymbol{\Omega}_\vartheta$ *exist and are non-singular.*

Assumption 13 has two parts which correspond to the case of heterogeneous and homogeneous coefficients. The first part is standard in the panel quantile literature for models with homogeneous coefficients and it is needed for the existence of limiting forms of positive definite matrices and to invoke a Central Limit Theorem. The second part relates to slope heterogeneity in a quantile framework. Assumption 13.b allows a general form of slope heterogeneity while guaranteeing that the covariance matrix of the QMG estimator is well defined.

The following theorem establishes the asymptotic distribution of the quantile mean group estimator.

Theorem 4 (Asymptotic Distribution of $\hat{\boldsymbol{\vartheta}}(\tau)$)**.** *Suppose the τ-th conditional quantile function of y_{it} for $i = 1, ..., N$ and $t = 1, ..., T$ is given by the panel data model* (2.1)-(2.2) *and Assumptions 1-13 hold. As $(N, p_T, T) \to \infty$ with $p_T^3/T \to 0$ and $N^{2/3}(\log(N))/T \to 0$, the mean group quantile regression estimator, defined by* (2.21)*, for a model with interactive effects,* $\sqrt{N}(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_v)$.

It should be noted that for fixed N, $\sqrt{T}(\hat{\boldsymbol{\vartheta}}(\tau)-\boldsymbol{\vartheta}(\tau))$ is asymptotically a Gaussian random variable. However, because the approximation of the factors requires $N \to \infty$ and we let N and T go jointly to infinity, the rates of Theorem 4 suggest that T has to be larger than N in finite samples to eliminate biases from incidental parameters and truncation of possibly infinite lag polynomials.

The following theorem establishes the asymptotic distribution of the quantile mean group estimator when the $\boldsymbol{\vartheta}_i(\tau)$'s are homogeneous.

Theorem 5. *Under the Assumptions of Theorem 4, as $(N, p_T, T) \to \infty$ with $p_T^3/T \to 0$ and $N^2(\log(N))^3/T \to 0$, the mean group quantile regression estimator, defined by* (2.21)*, for a model with interactive effects with $\boldsymbol{\vartheta}_i(\tau) = \boldsymbol{\vartheta}(\tau)$ for $1 \le i \le N$,* $\sqrt{NT}(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_\psi)$.

The convergence of the QMG estimator in Theorem 4 is $\sqrt{N}$ due to the heterogeneity of the parameter of interest, $\boldsymbol{\vartheta}_i(\tau)$. The standard $\sqrt{NT}$ convergence is obtained in Theorem 5 when the coefficients are not heterogeneous. These results appear to be comparable to standard convergence results for panel data estimators of conditional mean models with interactive effects (e.g., Pesaran (2006) and Chudik and Pesaran (2015)), but it is important to point out the difference in terms of the restrictions on T relative to N, due mainly to the estimation of individual parameters and the non-linearity of the quantile function.

2.3. **Inference**

The asymptotic covariance matrix can be consistently estimated using existing estimators. For large N and T, we define $\hat{u}_{it}(\tau) := y_{it} - \mathbf{X}_{it}'\hat{\boldsymbol{\pi}}_i(\tau)$, h_N to be a sequence of bandwidths such that $h_N \to 0$ as $N \to \infty$, and $K_{h_N}(u) = h_N^{-1}K(u/h_N)$ be a Kernel estimator. Then we can use the following estimators to consistently estimate $\mathbf{V}_\psi$,

$$\hat{\mathbf{J}}_i := \frac{1}{T-p_T}\sum_{t=p_T+1}^{T} K(\hat{u}_{it}(\tau))\mathbf{X}_{it}\mathbf{X}_{it}', \quad \hat{\mathbf{D}}_i := \frac{1}{T-p_T}\sum_{t=p_T+1}^{T} \hat{\sigma}_\psi^2(q)\mathbf{X}_{it}\mathbf{X}_{it}', \tag{2.22}$$

where, by the derivations in Section S.1 in the Supplement to the manuscript,

$$\hat{\sigma}_\psi^2(q) := \tau(1-\tau) + 2\sum_{j=1}^{q-1}\left(1-\frac{j}{q}\right)\left[I(Y_{i1} \le \mathbf{X}_{i1}'\hat{\boldsymbol{\pi}}_i(\tau), Y_{i1+j} \le \mathbf{X}_{i1+j}'\hat{\boldsymbol{\pi}}_i(\tau)) - \tau^2\right]. \tag{2.23}$$

for a positive integer $q \ge 1$. Note that $q = 1$ gives $\sigma_\psi^2(1) = \tau(1-\tau)$, the variance of $\psi_\tau = \tau - I(u < 0)$ in the case of independent observations (see, e.g., Koenker 2005). The matrix $\mathbf{J}_i$ can also be estimated by the method for non i.i.d. observations proposed by Hendricks and Koenker (1992). On the other hand, the matrix $\mathbf{V}_v$ can be estimated using $(N-1)^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\vartheta}}_i(\tau) - \hat{\boldsymbol{\vartheta}}(\tau))(\hat{\boldsymbol{\vartheta}}_i(\tau) - \hat{\boldsymbol{\vartheta}}(\tau))'$. Notice that if $v_i(\tau) = v_i$ for all i, the estimator $\hat{\mathbf{V}}_v$ is quantile invariant and therefore a consistent estimator of $\mathbf{V}_v$ can be defined as in equation (32) in Chudik and Pesaran (2015).

3. Monte Carlo

This section reports results of several simulation exercises designed to evaluate the small sample performance of the proposed estimator. Observations on y_{it} for $i = 1, 2, \ldots, N$ and $t = -S + 1, -S + 2, .., 0, 1, ..., T$ are generated according to the following model with two factors:

$$y_{it} = \beta_{0i} + \lambda_i y_{i,t-1} + \beta_{1i}x_{1,it} + \beta_{2i}x_{2,it} + \gamma_{1i}f_{1t} + \gamma_{2i}f_{2t} + \kappa_{0i}(1 + \kappa_{1i}x_{1,it})u_{it}, \tag{3.1}$$

where $\beta_{0i} = \alpha_i + \beta_0$, the error term u_{it} is distributed as F, κ_{0i} is an i.i.d. random variable distributed as uniform $\mathcal{U}(0.9, 1.1)$, and κ_{1i} is an i.i.d. random variable distributed as uniform $\mathcal{U}(0, 0.2)$. Depending on the values of κ_{0i} and κ_{1i}, we have two conditional quantile functions. (a) When $\kappa_{0i} = 1$ and $\kappa_{1i} = 0$ for all $1 \leq i \leq N$, we have

$$Q_{Y_{it}}(\tau | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i, \mathbf{f}_t) = \beta_{0i}(\tau) + \lambda_i y_{i,t-1} + \beta_{1i} x_{1,it} + \beta_{2i} x_{2,it} + \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t}, \tag{3.2}$$

with $\boldsymbol{\theta}_i = (\alpha_i, \lambda_i, \boldsymbol{\beta}_i', \boldsymbol{\gamma}_i')'$, $\boldsymbol{\beta}_i = (\beta_{1i}, \beta_{2i})'$, $\boldsymbol{\gamma}_i = (\gamma_{1i}, \gamma_{2i})'$, $\beta_{i0}(\tau) = \alpha_i + \beta_0(\tau)$, and $\beta_0(\tau) = \beta_0 + F_u^{-1}(\tau)$. (b) When $\kappa_{0i} \neq 1$ and $\kappa_{1i} \neq 0$ for all $1 \leq i \leq N$, the conditional quantile function of (3.1) becomes,

$$Q_{Y_{it}}(\tau | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \mathbf{f}_t) = \beta_{0i}(\tau) + \lambda_i y_{i,t-1} + \beta_{1i}(\tau) x_{1,it} + \beta_{2i} x_{2,it} + \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t}, \tag{3.3}$$

with $\boldsymbol{\theta}_i(\tau) = (\alpha_i(\tau), \lambda_i, \boldsymbol{\beta}_i'(\tau), \boldsymbol{\gamma}_i')'$, $\boldsymbol{\beta}_i(\tau) = (\beta_{1i}(\tau), \beta_{2i})'$, $\beta_{i0}(\tau) = \alpha_i(\tau) + \beta_0$, $\alpha_i(\tau) = \alpha_i + \kappa_{0i} F_u^{-1}(\tau)$ and $\beta_{1i}(\tau) = \beta_{1i} + \kappa_{0i}\kappa_{1i} F_u^{-1}(\tau)$. For each i, models (3.2) and (3.3) are typically referred to in the literature as location shift and location-scale shift models, respectively (see, e.g., Koenker (2005)). In all experiments, to simplify the exposition and without loss of generality, we set $\beta_0 = 0$ and $\beta_{2i} = 0.5$, for $1 \leq i \leq N$. Note that for S sufficiently large, we have that (with $\beta_0 = 0$),

$$y_{i0} \approx \frac{\alpha_i}{1 - \lambda_i} + \beta_{1i} \sum_{j=0}^{S-1} \lambda_i^j x_{1i,-j} + \beta_{2i} \sum_{j=0}^{S-1} \lambda_i^j x_{2i,-j} + \sum_{j=0}^{S-1} \lambda_i^j \xi_{i,-j}, \tag{3.4}$$

where $\xi_{it} = \gamma_{1i} f_{1t} + \gamma_{2i} f_{2t} + \kappa_{0i}(1 + \kappa_{1i} x_{1,it}) u_{it}$. In all the variants of the model considered in the simulations, we set $S = 200$ to minimize the effects of the initial values on the outcomes. The regressors, $x_{j,it}$, are generated as

$$x_{ji,t} = \mu_i + \Gamma_{ji} f_{jt} + v_{ji,t}, \tag{3.5}$$

$$v_{ji,t} = \rho_x v_{j,it-1} + \sqrt{1 - \rho_x^2}\, \varepsilon_{ji,t}, \tag{3.6}$$

$$f_{jt} = \rho_f f_{j,t-1} + \sqrt{1 - \rho_f^2}\, \varepsilon_{jt}, \tag{3.7}$$

for $j \in \{1, 2\}$, with $\mu_i \sim iid\mathcal{N}(0.5, 1)$, $\varepsilon_{ji,t} \sim iid\mathcal{N}(0, 1)$, and $\varepsilon_{jt} \sim iid\mathcal{N}(0, 1)$. We consider the case of relatively persistent regressors by setting $\rho_x = 0.8$ and $\rho_f = 0.9$. Moreover, without loss of generality we set $x_{ji,-S} = 0$ and $f_{j,-S} = 0$.

The factor loadings in equation (3.1), γ_{1i} and γ_{2i}, and in equation (3.5), Γ_{1i} and Γ_{2i}, are generated as $\gamma_{ji} \sim iid\mathcal{N}(0.5, 1)$ and $\Gamma_{ji} \sim iid\mathcal{N}(0.5, 1)$ for $j \in \{1, 2\}$. These factor loadings ensure that the rank condition in Assumption 6 is met. Finally, the fixed effects, α_i, are allowed to be correlated with the errors by generating them as $\alpha_i = \bar{x}_{1i} + \gamma_{1i}\bar{f}_1 + \gamma_{2i}\bar{f}_2 + \bar{u}_i + a_i$, where the individual specific averages are defined as $\bar{x}_{1i} = T^{-1}\sum_{t=1}^{T} x_{1,it}$, $\bar{f}_j = T^{-1}\sum_{t=1}^{T} f_{jt}$, $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$. The error term a_i in the equation for α_i is assumed to be distributed as $\mathcal{N}(0,1)$.

Initially, we set $\lambda_i = \lambda$ for $i = 1,2,\ldots,N$ and consider three values of $\lambda = \{0.25, 0.50, 0.75\}$. Later in Figure 3.1, we investigate the performance of the QMG estimator with heterogeneous λ_i's. Moreover, in addition to the experiments presented in this section, we also considered static panel data experiments (i.e., when $\lambda_i = 0$, for all i) and compare the performance of QMG estimator with a number of existing panel quantile regression estimators. For relatively large T, the performance of the proposed estimator was similar in both the static panel data model and dynamic panel data model. Thus, we present results for the dynamic model only to save space.

In the simulations, we assume that the error term u_{it} in equation (3.1) is an i.i.d. random variable distributed as Standard Normal, t-student with 4 degrees of freedom (t_4), and χ^2 with 3 degrees of freedom (χ^2_3). We consider the following four variations of the model (with $\lambda_i = \lambda$):

Design 1: (Location shift model with homogeneous slopes). We consider $\beta_1 = 1$ in a location shift model with $\kappa_{1i} = 0$ for all $1 \le i \le N$.
Design 2: (Location shift model with heterogeneous slopes). We consider heterogeneous slope parameters $\beta_{1i} = \beta_1 + \nu_{1i}$ in a location shift model, where $\kappa_{1i} = 0$ for all $1 \le i \le N$, $\beta_1 = 1$ and $\nu_{1i} \sim \mathcal{U}(-0.25, 0.25)$. The parameter $\beta_{1i}(\tau) = \beta_{1i}$ for all i and τ.
Design 3: (Location-scale shift model with homogeneous slopes). We consider homogenous slope parameters $\beta_1 = 1$ in a location-scale shift model with $\kappa_{1i} \sim \mathcal{U}(0, 0.2)$. In this case, the slope parameter $\beta_{1i}(\tau) = \beta_1 + \kappa_{0i}\kappa_{1i}F_u^{-1}(\tau)$ and $E(\beta_{1i}(\tau)) = \beta_1 + 0.1F_u^{-1}(\tau)$.
Design 4: (Location-scale shift model with heterogeneous slopes). We consider heterogeneous slope parameters as in Design 2, $\beta_{1i} = \beta_1 + \nu_{1i}$, in a location-scale shift model with $\kappa_{1i} \sim \mathcal{U}(0, 0.2)$. We assume $\beta_1 = 1$ and $\nu_{1i} \sim \mathcal{U}(-0.25, 0.25)$ which implies that $\beta_{1i}(\tau) = \beta_{1i} + \kappa_{0i}\kappa_{1i}F_u^{-1}(\tau) = 1 + \nu_{1i} + \kappa_{0i}\kappa_{1i}F_u^{-1}(\tau)$. In this case, $E(\beta_{1i}(\tau)) = \beta_1(\tau) = 1 + 0.1F_u^{-1}(\tau)$.

Tables 3.1 to Table 3.2 present the bias and root mean square error (RMSE) for the slope parameter $\beta_1(\tau)$ in the location shift model with $\lambda = 0.5$. The summary results for other choices of λ are provided in the online supplement. We focus on $\lambda = 0.5$ here, since we obtain similar estimates in the empirical application to be discussed in Section 4. While Table 3.1 presents results for Designs 1 and 2, Table 3.2 presents results for Designs 3 and 4. The tables show results for quantile regression estimators at two quantiles, $\tau \in \{0.25, 0.50\}$, based on sample sizes of $N \in \{100, 200\}$ and $T \in \{50, 100, 200\}$.

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Design 1: Location shift with homogeneous slopes											
100	50	Bias	0.188	-0.059	0.186	-0.059	0.177	-0.062	0.175	-0.070	0.135	-0.091	0.134	-0.051
100	50	RMSE	0.199	0.061	0.197	0.062	0.187	0.064	0.186	0.072	0.149	0.092	0.147	0.053
100	100	Bias	0.221	-0.019	0.219	-0.020	0.208	-0.023	0.209	-0.027	0.172	-0.038	0.165	-0.018
100	100	RMSE	0.226	0.022	0.224	0.023	0.214	0.025	0.214	0.029	0.179	0.040	0.171	0.019
100	200	Bias	0.240	-0.002	0.239	-0.002	0.231	-0.004	0.230	-0.006	0.195	-0.016	0.185	-0.005
100	200	RMSE	0.243	0.008	0.242	0.009	0.233	0.008	0.233	0.010	0.198	0.017	0.189	0.007
200	50	Bias	0.194	-0.063	0.192	-0.063	0.177	-0.067	0.176	-0.075	0.140	-0.092	0.136	-0.052
200	50	RMSE	0.204	0.064	0.202	0.064	0.187	0.068	0.186	0.075	0.152	0.093	0.147	0.053
200	100	Bias	0.231	-0.026	0.229	-0.026	0.209	-0.027	0.209	-0.031	0.173	-0.043	0.166	-0.021
200	100	RMSE	0.235	0.027	0.234	0.027	0.214	0.028	0.214	0.032	0.179	0.043	0.172	0.021
200	200	Bias	0.242	-0.009	0.241	-0.009	0.232	-0.009	0.232	-0.011	0.191	-0.019	0.182	-0.008
200	200	RMSE	0.244	0.010	0.243	0.010	0.234	0.010	0.235	0.012	0.194	0.020	0.185	0.009
N	T		Design 2: Location shift with heterogeneous slopes											
100	50	Bias	0.193	-0.059	0.191	-0.057	0.174	-0.062	0.173	-0.069	0.139	-0.090	0.137	-0.051
100	50	RMSE	0.203	0.061	0.201	0.059	0.186	0.064	0.185	0.072	0.152	0.091	0.149	0.052
100	100	Bias	0.229	-0.020	0.228	-0.020	0.211	-0.022	0.210	-0.026	0.178	-0.039	0.171	-0.018
100	100	RMSE	0.234	0.023	0.232	0.023	0.217	0.024	0.216	0.028	0.184	0.040	0.176	0.020
100	200	Bias	0.242	-0.002	0.241	-0.002	0.231	-0.004	0.232	-0.006	0.193	-0.016	0.185	-0.005
100	200	RMSE	0.244	0.008	0.244	0.009	0.234	0.009	0.235	0.011	0.196	0.017	0.188	0.007
200	50	Bias	0.194	-0.064	0.193	-0.063	0.178	-0.066	0.176	-0.074	0.138	-0.091	0.135	-0.051
200	50	RMSE	0.205	0.065	0.204	0.064	0.187	0.067	0.185	0.075	0.150	0.092	0.146	0.052
200	100	Bias	0.223	-0.026	0.222	-0.026	0.215	-0.027	0.215	-0.031	0.176	-0.042	0.170	-0.020
200	100	RMSE	0.228	0.027	0.227	0.027	0.220	0.028	0.220	0.032	0.182	0.042	0.175	0.021
200	200	Bias	0.244	-0.009	0.243	-0.009	0.232	-0.010	0.232	-0.012	0.195	-0.019	0.186	-0.008
200	200	RMSE	0.246	0.010	0.245	0.010	0.234	0.011	0.234	0.013	0.198	0.019	0.189	0.008

TABLE 3.1. *Bias and root mean square error (RMSE) of quantile regression estimators for $\beta_1(\tau)$ in Designs 1 and 2. In all the variations of the model, $\lambda = 0.5$. DQR denotes the instrumental variable quantile regression estimator for dynamic quantile regression, and QMG denotes the proposed mean quantile group estimator defined by* (2.21).

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Design 3: Location-scale shift with homogeneous slopes											
100	50	Bias	0.187	-0.061	0.183	-0.062	0.174	-0.063	0.171	-0.071	0.117	-0.084	0.114	-0.048
100	50	RMSE	0.198	0.063	0.194	0.064	0.184	0.065	0.182	0.073	0.130	0.085	0.126	0.049
100	100	Bias	0.219	-0.020	0.216	-0.021	0.206	-0.024	0.204	-0.028	0.150	-0.035	0.141	-0.016
100	100	RMSE	0.224	0.023	0.221	0.024	0.211	0.025	0.209	0.030	0.156	0.036	0.147	0.018
100	200	Bias	0.239	-0.002	0.237	-0.003	0.228	-0.004	0.225	-0.007	0.170	-0.014	0.159	-0.004
100	200	RMSE	0.241	0.009	0.240	0.009	0.231	0.009	0.228	0.010	0.173	0.015	0.162	0.007
200	50	Bias	0.192	-0.066	0.189	-0.065	0.174	-0.069	0.171	-0.076	0.120	-0.086	0.115	-0.048
200	50	RMSE	0.202	0.066	0.199	0.066	0.185	0.069	0.181	0.077	0.132	0.086	0.125	0.049
200	100	Bias	0.229	-0.027	0.226	-0.027	0.206	-0.027	0.204	-0.032	0.151	-0.039	0.142	-0.019
200	100	RMSE	0.233	0.028	0.231	0.028	0.211	0.028	0.210	0.032	0.156	0.039	0.148	0.019
200	200	Bias	0.240	-0.009	0.238	-0.009	0.229	-0.009	0.228	-0.012	0.167	-0.017	0.156	-0.007
200	200	RMSE	0.242	0.011	0.240	0.011	0.232	0.010	0.230	0.013	0.170	0.018	0.159	0.008
N	T		Design 4: Location-scale shift with heterogeneous slopes											
100	50	Bias	0.191	-0.061	0.188	-0.059	0.171	-0.064	0.169	-0.071	0.121	-0.083	0.116	-0.047
100	50	RMSE	0.202	0.063	0.199	0.062	0.183	0.066	0.181	0.074	0.133	0.084	0.127	0.048
100	100	Bias	0.227	-0.021	0.225	-0.021	0.208	-0.023	0.206	-0.027	0.155	-0.036	0.146	-0.017
100	100	RMSE	0.232	0.024	0.230	0.024	0.214	0.025	0.212	0.029	0.160	0.037	0.151	0.018
100	200	Bias	0.240	-0.002	0.238	-0.002	0.228	-0.005	0.228	-0.007	0.168	-0.014	0.158	-0.004
100	200	RMSE	0.243	0.008	0.241	0.009	0.231	0.009	0.230	0.011	0.172	0.015	0.161	0.007
200	50	Bias	0.193	-0.066	0.190	-0.065	0.175	-0.067	0.171	-0.075	0.119	-0.084	0.114	-0.047
200	50	RMSE	0.203	0.067	0.201	0.066	0.184	0.068	0.180	0.076	0.130	0.085	0.124	0.048
200	100	Bias	0.221	-0.027	0.219	-0.027	0.212	-0.027	0.210	-0.031	0.153	-0.038	0.145	-0.019
200	100	RMSE	0.226	0.028	0.224	0.028	0.217	0.028	0.216	0.032	0.159	0.039	0.150	0.019
200	200	Bias	0.242	-0.010	0.240	-0.009	0.229	-0.010	0.227	-0.012	0.170	-0.017	0.159	-0.007
200	200	RMSE	0.244	0.011	0.242	0.011	0.231	0.011	0.229	0.013	0.173	0.017	0.162	0.008

TABLE 3.2. *Bias and root mean square error (RMSE) of quantile regression estimators for $\beta_1(\tau)$ in Designs 3 and 4. In all the variations of the model, $\lambda = 0.5$. Also, see notes to Table 3.1.*

We compare the performance of the QMG estimator with the instrumental variable quantile regression estimator for dynamic panel data model developed by Galvao (2011), using $y_{i,t-2}$ as an instrument for $y_{i,t-1}$. This estimator is denoted by DQR. However, it is important to bear in mind that Galvao's model does not allow for the interactive term, $\lambda_i f_t$, and could generate biases that cannot be eliminated by use of instrumental variables. The QMG, is computed as the simple cross sectional average of standard quantile estimators, $\hat{\beta}_{1i}(\tau)$, using $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{y}_{t-1}, \bar{\mathbf{x}}_t')'$ to proxy the true unobserved factors f_{1t} and f_{2t}. We do not consider other existing quantile estimators, such as the classical quantile regression estimator, the fixed effects minimum distance quantile regression estimator by Galvao and Wang (2015), and the penalized quantile regression estimator, since all these estimators are biased when the model includes a lagged dependent variable. Therefore, we restrict our comparison to DQR, which is the only estimator in the literature proposed for dynamic panel quantile regression models.

3.1. **Bias and Root Mean Square Error**

As can be seen from Table 3.1, not surprisingly, the DQR estimator of β_1 is biased and that its bias tends to be slightly larger in the case where the slopes are heterogeneous. Furthermore, the bias of DQR estimator tends to increase with T, and tend to be similar for both 0.5 and 0.25 quantiles. On the other hand, the performance of the QMG estimator is excellent, with biases in general lower than 10% for $T = 50$, and decreasing rapidly to 1% when $T = 200$. In all the variations of the model considered in the table, the QMG estimator performs much better than DQR in terms of RMSE, as well.

Table 3.2 presents results for the location-scale shift model where $\beta_1(\tau)$ changes by quantile. We continue to see that the DQR estimator is biased and performs poorly in terms of RMSE. The performance of the QMG estimator in these variations of the model is similar to the results reported for the baseline model in Table 3.1, with low biases and small RMSE. For values of T larger than 50, the bias of the proposed estimator is always negative and ranges between 0.7% and 4%, and its RMSE is substantially below that of the DQR estimator. The RMSE of QMG relative to DQR is around 30 percent for $N = 100, T = 50$, and falls to around 0.05 for $N = T = 200$. The relative efficiency of the QMG estimator is similar across all the four designs.

We expanded the simulation evidence for the slope parameter β_1 to consider different values of λ. In the online supplement we present results for $\lambda \in \{0.25, 0.75\}$ considering the same designs as in Tables 3.1 and 3.2, with $N = 100$ and $T = 200$. We considered a moderate N and large T panel

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ_1		Parameter: λ		Parameter: θ_1	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.191	0.053	0.644	-0.058	-0.187	0.055	0.639	-0.043
100	50	RMSE	0.221	0.061	0.702	0.085	0.218	0.064	0.700	0.080
100	100	Bias	-0.253	0.022	0.703	-0.011	-0.249	0.023	0.703	-0.007
100	100	RMSE	0.270	0.029	0.736	0.044	0.266	0.032	0.737	0.046
100	200	Bias	-0.293	0.003	0.738	0.007	-0.291	0.003	0.736	0.008
100	200	RMSE	0.303	0.015	0.758	0.029	0.301	0.016	0.758	0.030
200	50	Bias	-0.198	0.056	0.666	-0.069	-0.195	0.057	0.665	-0.054
200	50	RMSE	0.225	0.060	0.722	0.083	0.221	0.061	0.724	0.071
200	100	Bias	-0.271	0.028	0.734	-0.027	-0.267	0.028	0.731	-0.023
200	100	RMSE	0.286	0.031	0.760	0.038	0.283	0.031	0.759	0.036
200	200	Bias	-0.294	0.011	0.744	-0.006	-0.292	0.011	0.740	-0.004
200	200	RMSE	0.303	0.015	0.757	0.017	0.301	0.015	0.754	0.018
N	T		t_4 distribution							
100	50	Bias	-0.171	0.057	0.607	-0.066	-0.166	0.063	0.614	-0.060
100	50	RMSE	0.200	0.065	0.677	0.092	0.196	0.073	0.684	0.096
100	100	Bias	-0.230	0.027	0.664	-0.019	-0.230	0.032	0.669	-0.018
100	100	RMSE	0.247	0.034	0.698	0.044	0.247	0.040	0.704	0.049
100	200	Bias	-0.277	0.005	0.699	0.002	-0.275	0.007	0.702	0.000
100	200	RMSE	0.287	0.015	0.723	0.028	0.285	0.018	0.725	0.030
200	50	Bias	-0.176	0.059	0.593	-0.083	-0.170	0.066	0.599	-0.073
200	50	RMSE	0.203	0.063	0.654	0.093	0.197	0.072	0.659	0.107
200	100	Bias	-0.235	0.031	0.651	-0.028	-0.233	0.035	0.658	-0.027
200	100	RMSE	0.252	0.034	0.683	0.038	0.250	0.038	0.690	0.041
200	200	Bias	-0.281	0.011	0.695	-0.007	-0.280	0.014	0.701	-0.008
200	200	RMSE	0.290	0.015	0.710	0.018	0.289	0.018	0.717	0.022
N	T		χ^2_3 distribution							
100	50	Bias	-0.109	0.080	0.473	-0.102	-0.109	0.046	0.466	-0.058
100	50	RMSE	0.152	0.097	0.529	0.146	0.148	0.063	0.519	0.104
100	100	Bias	-0.173	0.040	0.543	-0.043	-0.164	0.018	0.515	-0.018
100	100	RMSE	0.194	0.053	0.579	0.080	0.184	0.031	0.551	0.052
100	200	Bias	-0.217	0.020	0.576	-0.011	-0.204	0.005	0.540	-0.003
100	200	RMSE	0.229	0.031	0.597	0.046	0.216	0.018	0.560	0.033
200	50	Bias	-0.116	0.085	0.480	-0.098	-0.115	0.048	0.455	-0.056
200	50	RMSE	0.153	0.094	0.522	0.124	0.149	0.057	0.492	0.079
200	100	Bias	-0.174	0.047	0.542	-0.044	-0.167	0.023	0.511	-0.020
200	100	RMSE	0.192	0.053	0.565	0.063	0.185	0.029	0.533	0.040
200	200	Bias	-0.208	0.023	0.576	-0.018	-0.198	0.010	0.534	-0.006
200	200	RMSE	0.219	0.028	0.590	0.036	0.208	0.015	0.549	0.023

TABLE 3.3. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ_1 in Design 1. In all the variations of the model, $\lambda = 0.5$. Also, see notes to Table 3.1.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ_1		Parameter: λ		Parameter: θ_1	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.195	0.049	0.668	-0.063	-0.190	0.049	0.670	-0.038
100	50	RMSE	0.224	0.058	0.735	0.091	0.220	0.060	0.739	0.110
100	100	Bias	-0.267	0.021	0.732	-0.014	-0.265	0.022	0.729	-0.008
100	100	RMSE	0.283	0.033	0.770	0.054	0.280	0.035	0.770	0.055
100	200	Bias	-0.296	0.001	0.740	0.005	-0.294	0.001	0.741	0.006
100	200	RMSE	0.307	0.021	0.759	0.040	0.305	0.022	0.760	0.042
200	50	Bias	-0.200	0.055	0.674	-0.072	-0.196	0.054	0.678	-0.058
200	50	RMSE	0.226	0.060	0.741	0.087	0.223	0.060	0.752	0.078
200	100	Bias	-0.258	0.028	0.711	-0.025	-0.255	0.029	0.712	-0.020
200	100	RMSE	0.272	0.033	0.744	0.040	0.270	0.034	0.748	0.039
200	200	Bias	-0.301	0.011	0.742	-0.006	-0.299	0.010	0.739	-0.006
200	200	RMSE	0.309	0.018	0.758	0.028	0.307	0.018	0.756	0.028
N	T		t_4 distribution							
100	50	Bias	-0.170	0.054	0.591	-0.070	-0.165	0.061	0.600	-0.061
100	50	RMSE	0.206	0.065	0.661	0.102	0.201	0.074	0.669	0.106
100	100	Bias	-0.237	0.024	0.669	-0.019	-0.235	0.028	0.667	-0.019
100	100	RMSE	0.255	0.035	0.702	0.053	0.253	0.040	0.703	0.059
100	200	Bias	-0.279	0.006	0.696	0.002	-0.280	0.008	0.700	0.003
100	200	RMSE	0.290	0.021	0.716	0.041	0.290	0.023	0.722	0.042
200	50	Bias	-0.175	0.058	0.598	-0.078	-0.169	0.065	0.597	-0.072
200	50	RMSE	0.199	0.064	0.642	0.092	0.194	0.071	0.645	0.091
200	100	Bias	-0.245	0.029	0.678	-0.030	-0.242	0.032	0.685	-0.030
200	100	RMSE	0.260	0.033	0.708	0.045	0.257	0.038	0.716	0.047
200	200	Bias	-0.277	0.012	0.705	-0.006	-0.277	0.014	0.706	-0.009
200	200	RMSE	0.285	0.019	0.719	0.029	0.285	0.021	0.722	0.031
N	T		χ^2_3 distribution							
100	50	Bias	-0.114	0.081	0.486	-0.096	-0.115	0.045	0.463	-0.056
100	50	RMSE	0.158	0.099	0.540	0.142	0.156	0.063	0.513	0.100
100	100	Bias	-0.179	0.042	0.568	-0.041	-0.171	0.020	0.535	-0.017
100	100	RMSE	0.199	0.058	0.601	0.083	0.190	0.036	0.567	0.060
100	200	Bias	-0.212	0.021	0.577	-0.009	-0.201	0.007	0.542	0.002
100	200	RMSE	0.226	0.034	0.600	0.052	0.215	0.022	0.564	0.042
200	50	Bias	-0.114	0.082	0.479	-0.099	-0.110	0.047	0.462	-0.056
200	50	RMSE	0.148	0.093	0.528	0.130	0.142	0.058	0.510	0.086
200	100	Bias	-0.181	0.046	0.546	-0.044	-0.174	0.024	0.517	-0.019
200	100	RMSE	0.200	0.052	0.571	0.065	0.192	0.030	0.541	0.042
200	200	Bias	-0.214	0.021	0.584	-0.020	-0.203	0.009	0.546	-0.007
200	200	RMSE	0.225	0.028	0.598	0.039	0.214	0.017	0.560	0.028

TABLE 3.4. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ_1 in Design 2. In all the variations of the model, $\lambda = 0.5$. Also, see notes to Table 3.1.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ_1		Parameter: λ		Parameter: θ_1	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.188	0.055	0.637	-0.060	-0.165	0.070	0.596	-0.015
100	50	RMSE	0.219	0.063	0.694	0.089	0.198	0.078	0.648	0.072
100	100	Bias	-0.250	0.023	0.695	-0.011	-0.225	0.032	0.649	0.009
100	100	RMSE	0.267	0.031	0.729	0.045	0.244	0.039	0.678	0.046
100	200	Bias	-0.291	0.003	0.729	0.007	-0.268	0.009	0.676	0.018
100	200	RMSE	0.300	0.015	0.749	0.029	0.278	0.019	0.695	0.035
200	50	Bias	-0.195	0.057	0.661	-0.073	-0.172	0.071	0.618	-0.024
200	50	RMSE	0.222	0.061	0.717	0.083	0.202	0.075	0.666	0.053
200	100	Bias	-0.268	0.029	0.724	-0.028	-0.243	0.036	0.673	-0.008
200	100	RMSE	0.283	0.032	0.750	0.039	0.260	0.038	0.696	0.030
200	200	Bias	-0.291	0.011	0.735	-0.006	-0.268	0.016	0.678	0.005
200	200	RMSE	0.300	0.015	0.748	0.018	0.277	0.019	0.690	0.018
N	T		t_4 distribution							
100	50	Bias	-0.168	0.057	0.595	-0.067	-0.145	0.077	0.556	-0.027
100	50	RMSE	0.197	0.067	0.664	0.094	0.179	0.086	0.614	0.078
100	100	Bias	-0.226	0.027	0.653	-0.020	-0.207	0.042	0.598	0.003
100	100	RMSE	0.243	0.035	0.686	0.046	0.226	0.049	0.628	0.048
100	200	Bias	-0.272	0.005	0.689	0.001	-0.251	0.014	0.619	0.013
100	200	RMSE	0.282	0.015	0.712	0.028	0.262	0.022	0.640	0.034
200	50	Bias	-0.171	0.060	0.583	-0.083	-0.148	0.081	0.544	-0.035
200	50	RMSE	0.199	0.065	0.640	0.094	0.178	0.086	0.594	0.069
200	100	Bias	-0.231	0.031	0.639	-0.028	-0.210	0.043	0.589	-0.008
200	100	RMSE	0.247	0.034	0.670	0.039	0.228	0.046	0.615	0.033
200	200	Bias	-0.276	0.012	0.684	-0.007	-0.256	0.020	0.619	0.005
200	200	RMSE	0.285	0.015	0.699	0.018	0.265	0.023	0.632	0.022
N	T		χ^2_3 distribution							
100	50	Bias	-0.131	0.098	0.433	-0.109	-0.102	0.075	0.413	-0.013
100	50	RMSE	0.178	0.116	0.488	0.161	0.150	0.089	0.457	0.093
100	100	Bias	-0.203	0.044	0.495	-0.051	-0.163	0.030	0.437	0.001
100	100	RMSE	0.227	0.059	0.531	0.093	0.187	0.040	0.466	0.053
100	200	Bias	-0.251	0.020	0.520	-0.016	-0.205	0.010	0.445	0.006
100	200	RMSE	0.265	0.034	0.541	0.055	0.220	0.022	0.465	0.036
200	50	Bias	-0.139	0.101	0.439	-0.108	-0.107	0.076	0.404	-0.015
200	50	RMSE	0.177	0.110	0.479	0.138	0.147	0.082	0.432	0.062
200	100	Bias	-0.204	0.053	0.496	-0.052	-0.165	0.034	0.437	-0.002
200	100	RMSE	0.224	0.059	0.520	0.073	0.187	0.040	0.455	0.038
200	200	Bias	-0.242	0.025	0.520	-0.022	-0.199	0.015	0.443	0.003
200	200	RMSE	0.254	0.032	0.535	0.045	0.212	0.020	0.457	0.025

TABLE 3.5. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ_1 in Design 3. In all the variations of the model, $\lambda = 0.5$. Also, see notes to Table 3.1.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ_1		Parameter: λ		Parameter: θ_1	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.194	0.052	0.662	-0.064	-0.167	0.064	0.631	-0.013
100	50	RMSE	0.223	0.061	0.729	0.094	0.201	0.074	0.692	0.084
100	100	Bias	-0.264	0.022	0.724	-0.016	-0.241	0.030	0.674	0.007
100	100	RMSE	0.280	0.034	0.763	0.056	0.258	0.042	0.710	0.055
100	200	Bias	-0.292	0.002	0.733	0.005	-0.270	0.007	0.679	0.015
100	200	RMSE	0.304	0.022	0.752	0.041	0.282	0.023	0.697	0.045
200	50	Bias	-0.198	0.057	0.664	-0.074	-0.174	0.068	0.635	-0.030
200	50	RMSE	0.225	0.062	0.731	0.090	0.203	0.073	0.695	0.061
200	100	Bias	-0.254	0.029	0.705	-0.026	-0.231	0.036	0.659	-0.005
200	100	RMSE	0.269	0.034	0.738	0.041	0.247	0.041	0.689	0.035
200	200	Bias	-0.298	0.011	0.734	-0.006	-0.274	0.015	0.677	0.003
200	200	RMSE	0.306	0.019	0.749	0.028	0.283	0.022	0.690	0.029
N	T		t_4 distribution							
100	50	Bias	-0.166	0.055	0.580	-0.071	-0.143	0.077	0.545	-0.029
100	50	RMSE	0.202	0.067	0.651	0.104	0.184	0.088	0.603	0.095
100	100	Bias	-0.232	0.025	0.657	-0.021	-0.211	0.039	0.598	0.002
100	100	RMSE	0.250	0.036	0.689	0.055	0.231	0.048	0.628	0.056
100	200	Bias	-0.275	0.006	0.685	0.001	-0.256	0.015	0.619	0.015
100	200	RMSE	0.285	0.021	0.705	0.041	0.267	0.027	0.639	0.045
200	50	Bias	-0.170	0.060	0.586	-0.079	-0.147	0.079	0.545	-0.039
200	50	RMSE	0.195	0.065	0.629	0.093	0.173	0.085	0.585	0.069
200	100	Bias	-0.241	0.029	0.664	-0.031	-0.219	0.040	0.609	-0.012
200	100	RMSE	0.257	0.034	0.693	0.046	0.236	0.045	0.634	0.038
200	200	Bias	-0.272	0.013	0.692	-0.006	-0.253	0.019	0.623	0.003
200	200	RMSE	0.281	0.019	0.707	0.029	0.262	0.025	0.638	0.030
N	T		χ^2_3 distribution							
100	50	Bias	-0.139	0.096	0.443	-0.107	-0.108	0.073	0.408	-0.015
100	50	RMSE	0.187	0.114	0.498	0.156	0.159	0.087	0.449	0.088
100	100	Bias	-0.210	0.047	0.517	-0.050	-0.171	0.031	0.451	0.002
100	100	RMSE	0.231	0.064	0.551	0.094	0.194	0.045	0.479	0.061
100	200	Bias	-0.246	0.020	0.521	-0.016	-0.202	0.012	0.449	0.010
100	200	RMSE	0.262	0.036	0.544	0.059	0.219	0.026	0.470	0.045
200	50	Bias	-0.136	0.098	0.439	-0.108	-0.103	0.074	0.409	-0.015
200	50	RMSE	0.173	0.110	0.486	0.143	0.142	0.083	0.443	0.072
200	100	Bias	-0.212	0.050	0.495	-0.053	-0.172	0.034	0.438	-0.002
200	100	RMSE	0.233	0.057	0.519	0.075	0.194	0.039	0.456	0.038
200	200	Bias	-0.249	0.022	0.529	-0.025	-0.204	0.014	0.453	0.001
200	200	RMSE	0.261	0.031	0.544	0.047	0.217	0.021	0.466	0.030

TABLE 3.6. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ_1 in Design 4. In all the variations of the model, $\lambda = 0.5$. Also, see notes to Table 3.1.*

because our application in Section 4 employs a data set with 779 households and 8639 time-series observations. We see that the QMG estimator continues to perform better than the DQR estimator. We also find that the performance of the QMG estimator is invariant to the choice of λ, at least in the simulations considered thus far. We do investigate the performance of the QMG estimator in the heterogeneous case when $\lambda_i \in [0.025, 0.925]$, below.

We now turn our attention to the estimators of $\lambda(\tau)$ and $\theta_1(\tau) = \beta_1(\tau)/(1 - \lambda(\tau))$. The estimator for $\theta_1(\tau)$ is defined as $\hat{\beta}_1(\tau)/(1 - \hat{\lambda}(\tau))$ and is computed by plugging in the quantile estimates corresponding to $\lambda(\tau)$ and $\beta_1(\tau)$. We employ this method for both the DQR and QMG estimators.

Tables 3.3, 3.4, 3.5 and 3.6 show the bias and RMSE of the DQR and QMG estimators for the parameters of interest. These four tables show results for the four different designs we consider in this section. Each table presents, in columns, the performance of the estimators at $\tau \in \{0.25, 0.50\}$ and in rows the different samples sizes and distributions for the error term. The upper block presents results when u_{it} is distributed as $\mathcal{N}(0, 1)$, the middle panel shows results when $u_{it} \sim t_4$ and the lower block presents results when $u_{it} \sim \chi^2_3$.

As before, the results indicate that the bias of the DQR estimator can be large, in particular for the long run coefficient θ_1. The QMG estimator offers nearly zero biases for large N and T. The DQR estimator is biased and its performance is not satisfactory in terms of both bias and RMSE. The location-scale shift case, presented in Tables 3.5 and 3.6, reveals similar findings. Overall, when $\lambda = 0.5$, the QMG estimator offers the best performance in terms of bias and RMSE in the class of estimators for the dynamic quantile panel data models considered in this section.

Figure 3.1 offers a visual display of the small sample performance of the QMG estimator as λ increases. The figure shows the bias and RMSE of the QMG estimator at $\tau \in \{0.25, 0.50\}$ for λ, β_1 and θ_1 for different true values of λ. We considered Design 1 with $N = 100$ and $T = 200$. Recall that when λ increases, θ_1 increases too. For instance, while $\lambda = 0$ gives $\theta_1 = \beta_1 = 1$, $\lambda = 0.9$ gives $\theta = 10$ in our simulation experiment. Consistent with our previous evidence, we see that the performance of QMG estimator does not depend on λ when the interest is in estimating β_1. The bias tends to increase slightly, but it is never larger than 1% for values of λ close to unity. We also find that the RMSE of the estimator of β_1 does not change with λ. On the other hand, we observe that the absolute value of the bias of the QMG estimator for θ_1 increases rapidly $\lambda \to 1$. The figure shows that the bias, in absolute value, is negligible for $\lambda < 0.75$, and it increases rapidly when $\lambda > 0.8$. Note however that the bias in relative terms is always less than 10%. We also find

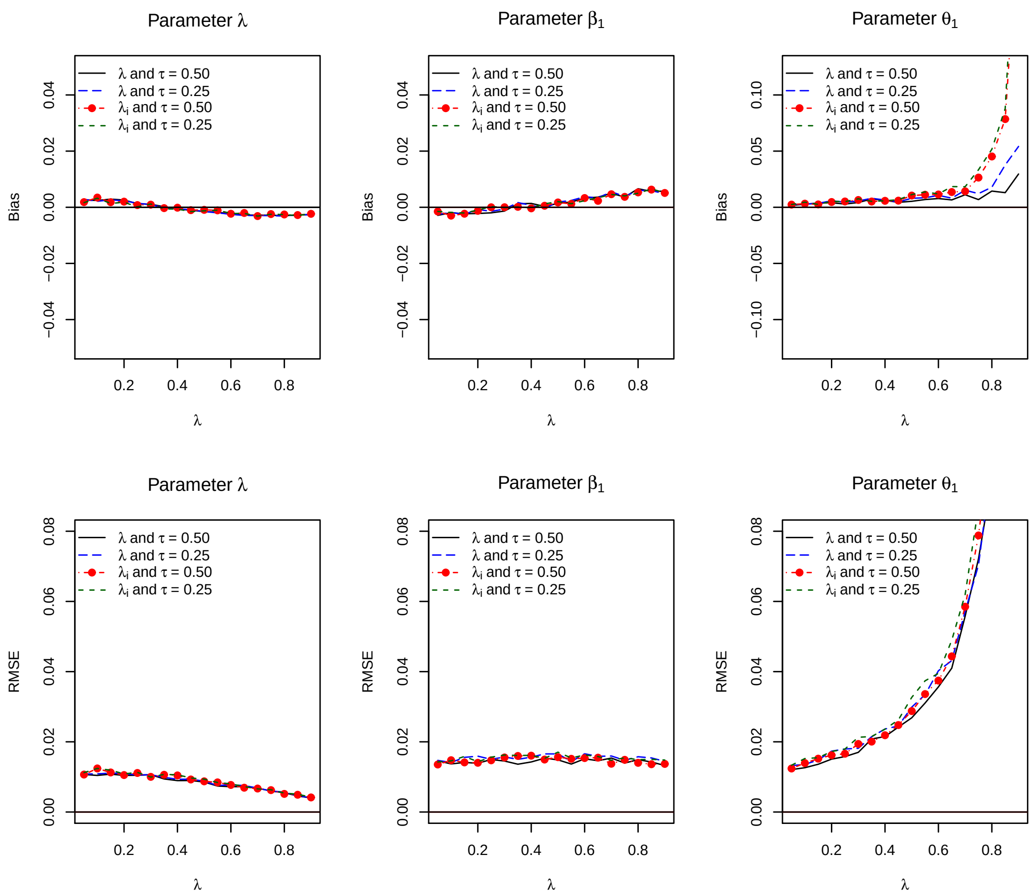

FIGURE 3.1. *Small sample performance of the QMG estimator for different values of* λ. *The figure present Bias and RMSE of the QMG estimator for* $E(\lambda(\tau))$, $E(\beta_1(\tau))$ *and* $E(\theta_1(\tau))$ *at the 0.25 and 0.50 quantiles.*

that the RMSE increases with λ and that the RMSE of the QMG estimator at $\tau = 0.25$ is larger than the QMG estimator at $\tau = 0.50$, as to be expected.

Figure 3.1 also shows the bias and RMSE of the QMG estimator when $\lambda_i = \lambda + \omega_i$, where $\omega_i \sim \mathcal{U}[-0.025, 0.025]$ and λ takes values in the interval $\lambda \in [0.05, 0.90]$. The parametrization guarantees that θ_1 exists for all values of λ_i for $i = 1, \ldots, N$. We generate data using Design 1 with $N = 100$ and $T = 200$. Consistent with our expectations, the bias and RMSE of the estimator tends to be similar to the case of homogeneous λ's, although the performance deteriorates for large values of $\lambda = E(\lambda_i)$. We see an increase in the variance of the estimator, but the bias for θ_1 remains, in absolute value,

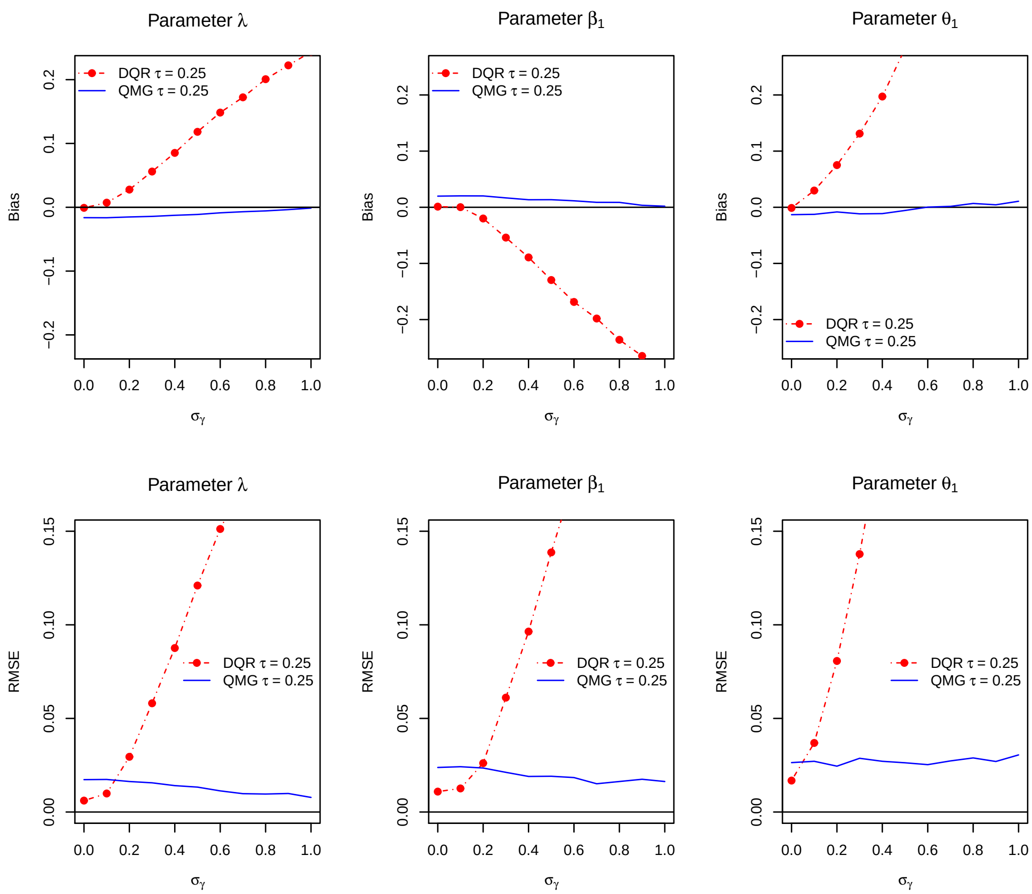

FIGURE 3.2. *Small sample performance of the DQR and QMG estimators in models with and without latent factors.*

small for $E(\lambda_i) < 0.65$. As can be seen from Figure 3.1, the parameter vector $(E(\lambda_i), \beta_1)$ can be estimated with small bias and excellent RMSE performance in the case of heterogeneous λ_i's, so long as N and T are sufficiently large, and $E(\lambda_i)$ is not too close to unity.

Finally, we investigate the relative performance of DQR and QMG in models with and without factor structure, i.e. $\sum_{j=1}^{2} \sigma_\gamma \gamma_{ji} f_{jt}$ in equation (3.1). As in Figure 3.1, we generate data using Design 1 with $N = 100$ and $T = 200$. In contrast with the previous design, we generate $\gamma_{1i} \sim iid\mathcal{N}(0.5, 1)$ and $\gamma_{2i} \sim iid\mathcal{N}(0.5, 1)$, and we set σ_γ to take values in the interval $[0, 1]$. Naturally, when $\sigma_\gamma = 0$, the model does not include latent factors. Figure 3.2 presents the bias and RMSE of the estimators for λ, β_1 and θ_1. Consistent again with expectations, when equation (3.1) does not include factors, the DQR estimator offers the best finite sample performance. However, as shown in the figure, the QMG performs reasonably well even when $\sigma_\gamma = 0$ and it offers the best performance in terms of bias and RMSE when the degree of parameter heterogeneity is not too small.

3.2. **Inference**

We now turn our attention to the standard error of the QMG estimator for $\lambda(\tau)$ and $\beta_1(\tau)$. Table 3.7 reports the average estimated standard errors obtained by the procedure outlined in Sections 2.2 (see Theorems 4 and 5) and 2.3 (equations (2.22) and (2.23)). When estimating $\mathbf{V}_v$ and $\mathbf{V}_\psi$, the asymptotic variance of the QMG estimators, we set $q = 3$ to minimize potential biases in the estimation of the standard errors. While the upper panels of Table 3.7 show the standard error of the QMG estimator in Designs 1 and 2, the lower panels show the standard error in Designs 3 and 4. We also report the standard deviation of the estimator based on 400 Monte Carlo repetitions. Because T relative to N is important for inference, we report estimates with $N = 100$ and $T \in \{100, 200, 400\}$

The results show that the estimated standard errors approximate very closely to the standard deviation of the estimator when T is larger than N. This result is expected by the rates of convergence needed to establish the consistency of the QMG estimator. The approximation is excellent in the case of the Normal and t_4 distributions. The evidence when $u_{it} \sim \chi^2_3$ suggests that a larger T relative to N is needed for the standard error to be well approximated.

Table 3.8 provides empirical coverage probabilities for a nominal 95% confidence interval. The probabilities are calculated based on asymptotic Gaussian confidence intervals consistent with Theorem 4. We see different finite sample performances of the estimator for λ and β_1. If we examine the results across the different distributions, the QMG estimator in some cases does not perform well for λ when $T/N < 4$. On the other hand, the coverage probabilities for β_1 approximate closely 0.95 with the exception of the case when $T = N = 100$. Lastly, we investigate the performance of the QMG estimator in terms of power. The results are shown in the lower panel of Table 3.8. We compute the power for testing $H_0 : \lambda = 0.5$ with the alternative hypothesis $H_a : \lambda = 0.55$ and $H_0 : \beta_1 = 1$ with the alternative hypothesis $H_a : \beta_1 = 1.1$. The condition on the rate of convergence plays an important role in ensuring that the estimator has good power. In particular, the power is high for values of $T > 100$, although how quickly the power approaches 1 depends on the distribution of the error term and the number of cross-sectional units, N.

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			λ		β_1		λ		β_1		λ		β_1	
			0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
N	T		Design 1: Location shift with homogeneous slopes											
100	100	Std error	0.004	0.004	0.016	0.016	0.004	0.005	0.017	0.018	0.007	0.004	0.029	0.018
100	100	Std dev	0.006	0.007	0.017	0.019	0.009	0.011	0.020	0.024	0.010	0.007	0.036	0.025
100	200	Std error	0.004	0.004	0.013	0.014	0.004	0.004	0.014	0.016	0.006	0.004	0.025	0.016
100	200	Std dev	0.004	0.004	0.010	0.012	0.008	0.009	0.016	0.018	0.007	0.005	0.022	0.016
100	400	Std error	0.004	0.004	0.012	0.013	0.003	0.003	0.010	0.012	0.005	0.003	0.019	0.012
100	400	Std dev	0.003	0.003	0.007	0.008	0.006	0.006	0.010	0.012	0.005	0.004	0.015	0.012
N	T		Design 2: Location shift with heterogeneous slopes											
100	100	Std error	0.009	0.009	0.028	0.029	0.009	0.010	0.030	0.033	0.011	0.008	0.047	0.034
100	100	Std dev	0.006	0.006	0.021	0.023	0.006	0.007	0.022	0.025	0.008	0.006	0.036	0.025
100	200	Std error	0.006	0.007	0.022	0.023	0.006	0.007	0.023	0.025	0.008	0.006	0.035	0.026
100	200	Std dev	0.004	0.004	0.019	0.019	0.003	0.005	0.018	0.019	0.006	0.004	0.026	0.020
100	400	Std error	0.005	0.006	0.021	0.022	0.005	0.006	0.022	0.023	0.008	0.005	0.032	0.024
100	400	Std dev	0.003	0.003	0.017	0.018	0.003	0.003	0.016	0.018	0.004	0.003	0.021	0.018
N	T		Design 3: Location-scale shift with homogeneous slopes											
100	100	Std error	0.004	0.004	0.016	0.016	0.005	0.005	0.017	0.019	0.007	0.004	0.031	0.020
100	100	Std dev	0.006	0.007	0.018	0.020	0.009	0.011	0.020	0.025	0.009	0.007	0.039	0.028
100	200	Std error	0.004	0.004	0.012	0.013	0.004	0.004	0.014	0.016	0.006	0.004	0.027	0.017
100	200	Std dev	0.004	0.004	0.011	0.013	0.008	0.008	0.016	0.019	0.007	0.005	0.027	0.019
100	400	Std error	0.003	0.003	0.011	0.012	0.004	0.004	0.013	0.015	0.006	0.004	0.026	0.017
100	400	Std dev	0.003	0.003	0.008	0.009	0.006	0.006	0.011	0.012	0.005	0.004	0.021	0.014
N	T		Design 4: Location-scale shift with heterogeneous slopes											
100	100	Std error	0.009	0.010	0.028	0.030	0.009	0.010	0.030	0.034	0.011	0.007	0.052	0.037
100	100	Std dev	0.006	0.007	0.022	0.024	0.006	0.007	0.023	0.025	0.007	0.006	0.039	0.026
100	200	Std error	0.006	0.007	0.022	0.023	0.006	0.007	0.023	0.026	0.008	0.005	0.039	0.028
100	200	Std dev	0.004	0.004	0.019	0.020	0.004	0.005	0.018	0.020	0.005	0.004	0.031	0.022
100	400	Std error	0.005	0.006	0.021	0.022	0.005	0.006	0.022	0.024	0.007	0.005	0.036	0.026
100	400	Std dev	0.003	0.003	0.017	0.018	0.003	0.003	0.017	0.019	0.004	0.003	0.025	0.019

TABLE 3.7. *Standard error (Std error) of the QMG estimator for λ and β_1. Std dev denotes standard deviation and $\lambda = 0.5$ in the simulations.*

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			λ		β_1		λ		β_1		λ		β_1	
			0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
N	T	Design	Coverage probability for a nominal 95% confidence interval											
100	100	1	0.548	0.558	0.872	0.830	0.675	0.632	0.890	0.838	0.845	0.782	0.885	0.860
100	200	1	0.642	0.682	0.932	0.932	0.740	0.722	0.918	0.905	0.945	0.890	0.978	0.938
100	400	1	0.712	0.750	0.968	0.978	0.718	0.785	0.950	0.938	0.925	0.862	0.990	0.952
100	100	2	0.878	0.902	0.970	0.978	0.922	0.952	0.978	0.985	0.985	0.965	0.990	0.988
100	200	2	0.842	0.885	0.958	0.960	0.912	0.900	0.970	0.982	0.988	0.965	0.990	0.978
100	400	2	0.862	0.870	0.968	0.958	0.912	0.925	0.988	0.985	0.985	0.970	0.992	0.980
100	100	3	0.592	0.585	0.848	0.805	0.652	0.632	0.890	0.862	0.830	0.752	0.875	0.845
100	200	3	0.565	0.578	0.868	0.812	0.745	0.735	0.918	0.895	0.922	0.882	0.938	0.910
100	400	3	0.745	0.748	0.965	0.968	0.845	0.888	0.978	0.982	0.970	0.940	0.978	0.975
100	100	4	0.892	0.892	0.975	0.982	0.935	0.955	0.978	0.975	0.990	0.965	0.988	0.980
100	200	4	0.840	0.882	0.958	0.955	0.920	0.922	0.972	0.988	0.980	0.950	0.970	0.968
100	400	4	0.868	0.885	0.952	0.960	0.910	0.930	0.985	0.980	0.990	0.990	0.995	0.985
N	T	Design	Power performance											
100	100	1	1.000	1.000	0.995	0.992	1.000	1.000	0.988	0.938	1.000	1.000	0.498	0.972
100	200	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.930	1.000
100	400	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	100	2	1.000	1.000	0.822	0.788	1.000	1.000	0.758	0.580	1.000	1.000	0.172	0.692
100	200	2	1.000	1.000	0.995	0.988	1.000	1.000	0.998	0.982	1.000	1.000	0.735	0.985
100	400	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000	0.925	0.992
100	100	3	1.000	1.000	0.988	0.952	1.000	1.000	0.975	0.848	1.000	1.000	0.368	0.838
100	200	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.872	1.000
100	400	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	100	4	1.000	1.000	0.795	0.640	1.000	1.000	0.745	0.432	1.000	1.000	0.105	0.432
100	200	4	1.000	1.000	0.995	0.978	1.000	1.000	0.995	0.955	1.000	1.000	0.565	0.942
100	400	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.992	1.000	1.000	0.835	0.995

TABLE 3.8. *Empirical coverage probability for a nominal 95 percent confidence interval and Power performance of the QMG estimator. We compute the power for testing* $H_0 : \lambda = 0.5$ *and* $H_0 : \beta_1 = 1$, *against alternatives* $H_a : \lambda = 0.55$ *and* $H_a : \beta_1 = 1.1$.

4. **Time-of-Use Pricing, Smart Technology and Energy Savings**

In recent years electric utilities around the country have installed a vast number of smart meters in homes and businesses. This new digital technology replaces the outdated electric meters used in previous decades and allows two-way communication between devices inside the home and the utility. This has lead to a renewed interest in the roll-out of various Time-of-Use (TOU) electricity pricing strategies[1] since utilities now have the ability to communicate prices to the consumers in real time. While economists have explored this topic in earlier decades, especially after the 1970s energy crisis, the technology enabling customers to respond to these novel electric rates was largely not available.

Technological advances referred to as “smart technologies” remove however the limitations of earlier decades and can meaningfully allow customers to take advantage of time varying electric rates to respond to peak demand prices or conserve electricity more broadly. Thus, it appears that substantial peak load reductions can in fact be achieved from TOU pricing (Jessoe and Rapson (2014), Ito (2014)). The literature however documents just how important the different types of enabling technology are on consumer responsiveness. Harding and Lamarche (2016) estimate the impact of TOU pricing using a randomized controlled trial of over 11 million observations on 15-minute interval electricity consumption in the US and show that smart devices with automation features achieve the highest peak demand savings and monetary incentives alone are not sufficient by themselves to motivate consumers to respond to time varying prices in an economically significant fashion.

In this section, we consider data from a similar randomized controlled trial, to study effectiveness of three major enabling technologies (web portal, in-home display and smart thermostat) within the context of TOU pricing. By allowing for interactive effects in the quantile regressions, we also take account of possible differences in unobserved common effects on households with differing characteristics.

We apply our quantile regression approach to estimate an autoregressive panel quantile regression model for energy consumption with interactive effects. We then compare the effect of different technologies on energy consumption, focusing on the distributional effects of these randomly assigned technologies. We find that smart thermostats are particularly effective relative to other technologies at enabling households to respond to TOU pricing. The differential effects are more pronounced at

[1] In addition to TOU rates a variety of dynamic pricing strategies are currently explored. See Harding and Sexton (2017) for a comprehensive review of recent developments.

the lower tail of the conditional distribution of energy consumption. While households appear to reduce overall consumption as a result of these technologies relative to the control group, the average response fails to capture the distributional effects of the technologies across households. Since utilities face a heterogenous customer base, understanding the distributional impact of the policies has important regulatory consequences. Lastly, we investigate the long-run effect of a change in energy price for different enabling technologies, focusing on the differential effects for different age and income groups.

4.1. **Data**

We employ data from a large scale randomized controlled trial (RCT) of TOU pricing for residential electricity consumers in a South Central US state. The data used in this paper includes 779 customers who were randomly assigned to a time-of-use pricing structure and received three different enabling technologies. All households had previously installed smart meters recording electricity consumption at 15 minute intervals.[2]

The random allocation of a large sample of households into three treatment groups and one control group, and the availability of electricity readings measured over 15-minute intervals make the application of our QMG estimator particularly well suited to answer questions about the distributional effect of enabling technologies.

The experiment was conducted during four months from June 1st to September 30th of 2011. After households signed up for the program, they were randomly assigned into three different treatment groups and a control group. Consumers randomized to the control group were informed they were not eligible for the program at that time but might be allowed to join next year. These households were kept on standard residential tariff and did not receive any enabling technology. On the other hand, customers who were selected to the treatment groups were assigned a time-of-use pricing rate which varied over two daily time periods. During the off-peak part of the day consisting of all hours except 2pm to 7pm, the rate charged for electricity consumption was \$0.042 kWh. During the on-peak part of the day, which was the period from 2pm to 7pm, the rate charged was \$0.23 kWh. Weekends were considered to be off-peak throughout.

[2]While many utilities consider data such as the one collected from this experiment to be proprietary, similar data is publicly available. For example the CER data from Ireland is commonly used as a test data set for the evaluation of a pricing experiment using high-frequency smart meter data.

	Control		Portal		IHD		PCT	
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
Kilowatt-hours	0.61	0.51	0.62	0.52	0.59	0.48	0.59	0.48
Treatment	0.14	0.35	0.14	0.35	0.14	0.35	0.14	0.35
High Income ($> \$75,000$)	0.38	0.49	0.58	0.49	0.51	0.50	0.49	0.50
Medium Income	0.31	0.46	0.22	0.41	0.31	0.46	0.28	0.45
Low Income ($< \$30,000$)	0.31	0.46	0.21	0.40	0.18	0.39	0.23	0.42
Mature (65 or older)	0.20	0.40	0.26	0.44	0.28	0.45	0.31	0.46
Family Life	0.49	0.50	0.42	0.49	0.45	0.50	0.37	0.48
Young (45 or younger)	0.31	0.46	0.32	0.47	0.27	0.44	0.33	0.47
Temperature (°F)	84.88	12.85	84.85	12.85	84.89	12.85	84.95	12.85
Dew Point (°F)	58.51	7.91	58.53	7.93	58.50	7.91	58.43	7.88
Number of households	242		189		152		196	
Number of periods	8639		8639		8639		8639	
Number of observations	2090638		1632771		1313128		1693244	

Table 4.1. *Descriptive Statistics for the Smart Meter Data. The control group include households that have no access to the enabling technologies. Portal means that the households have access to a website, IHD denotes in-home display and PCT denotes "smart" programmable communicating thermostat. Households in the IHD and PCT groups also had access to a website.*

Treated households were then further randomized by received additional enabling technologies. All treated households had access to a website ("web portal") which exhibited information on their electricity consumption and prices in real time. Our sample includes a group of 189 households who were limited to the website as the only enabling technology.

The other households in the treatment group were randomly assigned to receive one of these two *additional* enabling technologies: an in-home display (IHD) or a "smart" programmable communicating thermostat (PCT). An IHD is a small wireless tablet which displays information on electricity usage and cost in real time and is typically placed in a highly visible place in the house, e.g. kitchen. The PCT provides an interface that allows the customer to program and control the air conditioning system and respond to future and current price events. It also offers the same price and consumption information as displayed on the IHD screen. While a group of 152 households received in-home displays, another group of 196 customers received "smart" programmable communicating thermostats.

The large scale RCT has a high degree of compliance among treated participants. Only a small proportion participants (less than 4%) were switched to alternative treatments, largely due to problems installing the required technology. We restrict the sample to households who did not change treatment status and whose electricity readings measured over 15-minute were consistently recorded in the period between June and September. As shown in Table 4.1, we this leads to a balanced panel of 6,729,781 observations with $N = 779$ and $T = 8,639$. Since the majority of the households had central AC, we focus only on these households in the analysis.

Only a limited number of observed covariates is available for the analysis. This is common in this industry since utilities have very little information on the customers themselves. Demographic information was collected from the Nielsen's PRIZM® segmentation system.[3] and allows us to partition our sample by life stages and income. In Table 4.1, "young years" is designed to capture younger households, under 45 years of age with no children. The "family life" segment captures middle aged families with children. Households were also clustered by income into three groups: low, middle and high. The high group includes households with income above \$75,000 and the middle income group captures households with income between \$30,000 and \$75,000. These types of customer segmentations are rather insufficient to capture treatment heterogeneity and further highlight the attractiveness of econometric approaches such as the one proposed in this paper to overcome data limitations.

Due to confidentiality reasons we don't have access to exact address information for these households. We do however know the zip codes in which the households reside and are thus able to further augment our sample with zip-code specific temperature and humidity data from collected from Weather Underground.

4.2. Model

Recall that each household was randomly assigned to either a treatment group or the control group. Let $g \in \{0, 1, 2, 3\}$ denote the groups, $g = 0$ denoting the control group, and $g \in \{1, 2, 3\}$ denoting households assigned to either Portal, IHD or PCT. Designate the households by $i = 1, 2, \ldots, N_g$ and 15-minute intervals by $t = 1, 2, \ldots, T$. Recall that only households with a continuous record of electricity consumption over 96 (15 minutes) intervals per day and over roughly 90 days are included. To explore the importance of heterogeneity of treatment effects, we consider the following dynamic panel data model:

$$y_{igt} = \alpha_{ig} + \lambda_{ig} y_{it-1} + \delta_{ig} d_t(g) + \mathbf{x}'_{ig,t} \boldsymbol{\beta}_{ig} + \mathbf{f}'_t \boldsymbol{\gamma}_{ig} + u_{igt}, \tag{4.1}$$

[3]PRIZM partitions the U.S. population into 66 types, or segments, aligned along two major dimensions, life stages and income.

where y_{igt} is the natural logarithm of electricity usage for household i in group $g \in \{0, 1, 2, 3\}$ during the 15-minute interval t, and the associated vector of weather measurements $\mathbf{x}_{ig,t} = (x_{ig,t,1}, x_{ig,t,2})'$ includes temperature and dew point. We note that $\mathbf{x}_{ig,t}$ is the same for all individuals in the same location, irrespective of their group assignment. But the inclusion of fixed effects in the model allows assignment of the treatment to depend on location-specific variables, $\mathbf{x}_{ig,t}$. The variable $d_t(g)$ indicates the treatment assignment g and it takes the value 1 if t is between 2 pm and 7 pm during weekdays, and 0 otherwise. Our quantile treatment coefficients are identified by the time variation associated with TOU pricing:

$$Q_{Y_{igt}}(\tau|\cdot) = \alpha_{ig}(\tau) + \lambda_{ig}(\tau)y_{it-1} + \delta_{ig}(\tau)d_t(g) + \mathbf{x}'_{ig,t}\boldsymbol{\beta}_{ig,t}(\tau) + \mathbf{f}'_t\boldsymbol{\gamma}_{ig}(\tau), \tag{4.2}$$

where $Q_{Y_{igt}}(\tau|\cdot)$ is the τ-th conditional quantile function and $\delta_{ig}(\tau)$ is the quantile treatment effect (QTE) of interest.

We estimate the model using our QMG estimator for each quantile τ and group g separately. The estimator is implemented considering cross-sectional averages of the logarithm of electricity usage, $(\bar{y}_t, \bar{y}_{t-1}, \ldots, \bar{y}_{t-p_T})$, as well as cross-sectional averages of temperature and dew point, $(\bar{x}_{1,t}, \ldots, \bar{x}_{1,t-p_T}, \bar{x}_{2,t}, \ldots, \bar{x}_{2,t-p_T})$. Note that $\bar{y}_t = N^{-1}\sum_{i,g} y_{ig,t}$, and $\bar{x}_{j,t} = N^{-1}\sum_{i,g} x_{j,ig,t}$, and $N = \sum_{g=0}^{3} N_g$. We follow the recommendations of the theory in Section 2 and set $p_T = 4$.[4] The standard errors are estimated by the procedure described in Section 2.3 using $q = 3$, which allows incorporating possible dependence across time in the estimation of the asymptotic covariance matrix of the QMG estimator. We do not include controls for demographics in the main results shown in the next section, but we explore heterogeneity of effects among consumers with different observable characteristics (i.e., high vs. low income) in Section 4.4.

4.3. **Main Empirical Results**

Table 4.2 reports results for the coefficient $\lambda_g(\tau) = E(\lambda_{ig}(\tau))$ and the QTE, $\delta_g(\tau) = E(\delta_{ig}(\tau))$, for the four groups: control group, portal, in-home display (IHD), and programmable communicating thermostats (PCT). The last two columns present results obtained by using fixed effects (FE) estimators which produces inconsistent results in dynamic heterogeneous panels, and the CCE mean group (CCEMG) estimator as in Chudik and Pesaran (2015) that allows for heterogeneity and interactive effects. The first five columns show the quantile regression version of the CCEMG estimator, labeled QMG.

[4]We examine the sensitivity of results to the choice of p_T in the online Supplement. See Section S.4.

	QMG					FE	CCEMG
	0.10	0.25	0.50	0.75	0.90		
				Control Group			
Consumption at	0.464	0.573	0.616	0.477	0.353	0.623	0.474
$t-1$ (in logs)	(0.020)	(0.021)	(0.021)	(0.020)	(0.015)	(0.001)	(0.009)
Treatment	0.135	0.102	0.059	0.044	0.037	0.145	0.086
(2pm - 7pm)	(0.009)	(0.008)	(0.006)	(0.006)	(0.006)	(0.001)	(0.020)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	242	242	242	242	242	242	242
$N \times T$	2090638	2090638	2090638	2090638	2090638	2090638	2090638
				Portal			
Consumption at	0.468	0.586	0.628	0.484	0.360	0.622	0.487
$t-1$ (in logs)	(0.021)	(0.023)	(0.024)	(0.022)	(0.015)	(0.001)	(0.009)
Treatment	0.081	0.060	0.037	0.019	0.000	0.102	0.043
(2pm - 7pm)	(0.012)	(0.011)	(0.009)	(0.010)	(0.013)	(0.001)	(0.019)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	189	189	189	189	189	189	189
$N \times T$	1632771	1632771	1632771	1632771	1632771	1632771	1632771
				IHD			
Consumption at	0.469	0.578	0.612	0.473	0.352	0.627	0.478
$t-1$ (in logs)	(0.022)	(0.025)	(0.027)	(0.025)	(0.018)	(0.001)	(0.009)
Treatment	0.087	0.064	0.037	0.022	-0.004	0.089	0.040
(2pm - 7pm)	(0.015)	(0.012)	(0.010)	(0.008)	(0.011)	(0.002)	(0.019)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	152	152	152	152	152	152	152
$N \times T$	1313128	1313128	1313128	1313128	1313128	1313128	1313128
				PCT			
Consumption at	0.716	0.783	0.804	0.692	0.561	0.771	0.680
$t-1$ (in logs)	(0.024)	(0.020)	(0.019)	(0.022)	(0.021)	(0.000)	(0.007)
Treatment	-0.081	-0.052	-0.027	-0.030	-0.037	-0.010	-0.067
(2pm - 7pm)	(0.020)	(0.015)	(0.010)	(0.010)	(0.014)	(0.001)	(0.016)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	196	196	196	196	196	196	196
$N \times T$	1693244	1693244	1693244	1693244	1693244	1693244	1693244

TABLE 4.2. *Quantile Mean Group estimator results for the control group and different technologies. FE denotes fixed effects and CCEMG denotes the Common Correlated Mean Group estimator due to Chudik and Pesaran (2015). IHD denotes in-home display and PCT is programmable communicating thermostats. Standard errors are in parentheses.*

It is important to note, that in absence of a rich set of covariates specific to the consumers, it is important that the panel regressions contains unobserved effects that are different from time dummies. For instance, homes can have different levels of insulation that lead to different electricity usage when weather conditions experience sharp changes. We allow for consumer-specific common effects by the availability of the data and the use of CCE type estimators.

The FE results tend to overestimate the effect of the lagged dependent variable and the treatment effect, which is in line with the theoretical results obtained by Pesaran and Smith (1995) on the inconsistency of the FE estimators for dynamic heterogenous panels even for N and T large panels that we are considering here. Because these results are likely to be biased, we concentrate our attention on the CCEMG estimates. The positive and significant coefficient for the control group indicates that consumption *increases* by 9.0% from 2 pm to 7 pm when temperature is likely to be high.[5] However, TOU pricing scheme seem to reduce energy consumption since the other treatment effects are smaller than 0.086. The table shows, however, that the technology adopted by households crucially determines whether the households engage in some saving behavior. The coefficient for Portal and IHD are positive and significant, and they suggest a smaller (relative to the control group) 4% *increase* in energy use (although the differences might not be statistically significantly different from zero). However, the effect for the households using PCT are negative and significant relative to the other groups. The estimates show that smart thermostats are particularly effective in enabling consumers to respond to TOU pricing. Households provided with a PCT achieve a *reduction* of 6.5% when energy prices are high.

Households response, however, is not homogeneous across the quantiles of the conditional distribution of electricity consumption. Among consumers with a PCT technology, we find the largest energy saving in the lower tail of the conditional distribution, while the effect of TOU pricing is weakly significant at the upper conditional quantile. When we examine the distributional effect across households with Portal and IHD technologies, we find a similar pattern. The QTE decreases in absolute value as we go across quantiles, changing from a significant effect at the 0.1 quantile to an effect not significantly different than zero at the 0.9 quantile. The effect of using PCT continues to be negative at the lower tail, and the effect of IHD is positive, although smaller than the estimate for the control group. This is an interesting finding that has policy implications as it suggests that consumers react to the price changes, but the IHD is substantially less effective than the PCT in

[5]The mean maximum daily temperature was 99°F and the median was 103°F. The months of July and August were very similar and September was substantially cooler with mean temperatures of 88.6°F.

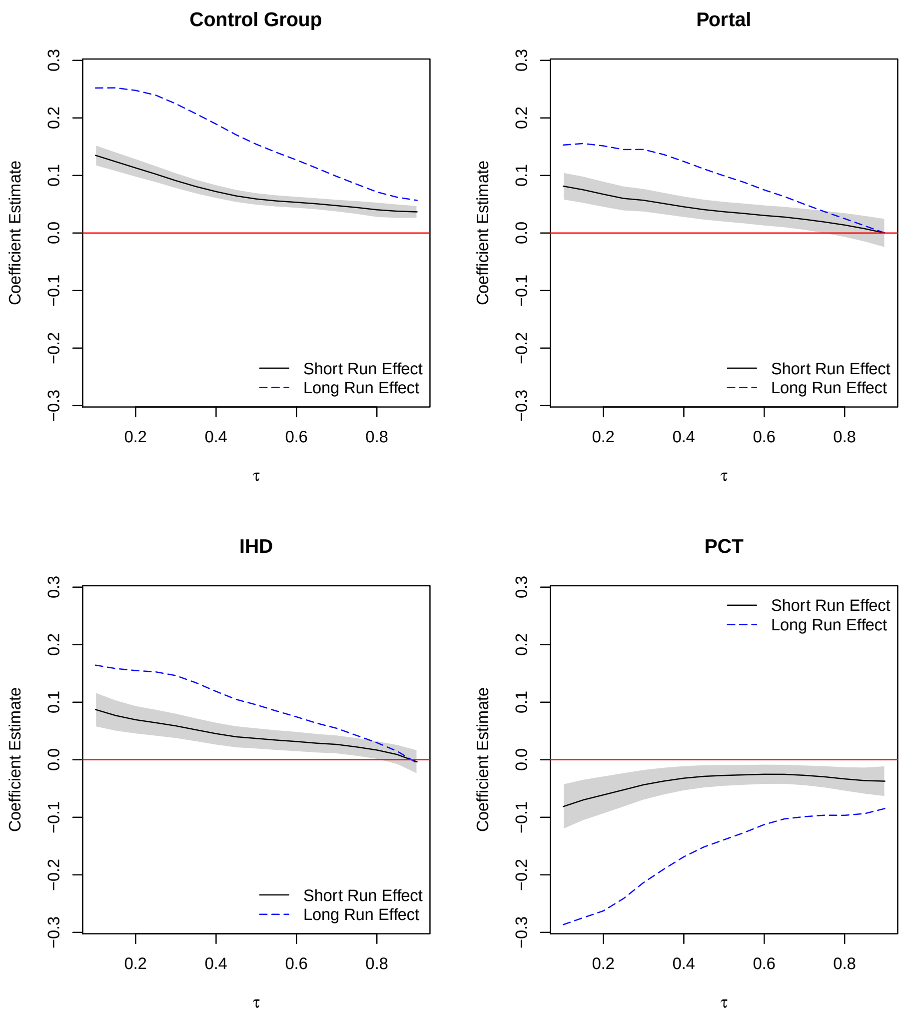

FIGURE 4.1. *Short and Long Run Quantile Regression Results. The figure shows the QTE coefficient $\delta_g(\tau)$ for the control group, portal group, in-home-device (IHD), and programmable communicating thermostats (PCT). The grey area denotes a 95 percent point-wise confidence interval.*

terms of energy savings. This might explain why in spite of the huge initial popularity of IHD technologies they have failed to be adopted at scale.

Figure 4.1 offers a clear view of the main findings. The figure presents estimates of the QTE as a function of the quantile τ of the conditional distribution of electricity consumption. We present estimates for the short run and long run effects for the three treatment and the control groups. The continuous line show the QMG estimates of $\delta_g(\tau)$ (impact effect) and the dashed line show QMG estimates for $\theta_g(\tau) = \delta_g(\tau)/(1 - \lambda_g(\tau))$ (long-run effect) which is estimated as discussed above. The gray areas denote 95% point-wise confidence intervals.

We find that the estimates for the control group are smaller as compared to the estimates for the treatment groups. This can be interpreted as suggesting that Portal and IHD reduce consumption during periods of high electricity cost but these technologies do not seem to achieve a significant energy reduction at any quantile. On the other hand, the profiles of QTE for the control group and PCT group are clearly different, suggesting that smart thermostats are effective in allowing households to respond to the increase in the price of electricity between 2 pm and 7 pm. Moreover, it is interesting to see that the largest savings differentials in the short-run and long-run are estimated at the lower tail of the conditional distribution, while these differentials are small at the 0.9 quantile. The evidence indicates that households provided with a PCT can engage in considerable energy savings in the long run and the impact of the enabling technology is greater at the 0.1 quantile of the distribution of electricity consumption.

4.4. **Responsiveness across Demographics**

It is often important for policymakers to understand how the responsiveness to TOU pricing and enabling technologies changes with household demographics. This section addresses this question offering evidence on how consumers with different characteristics respond to TOU pricing. The household characteristics are limited to age and income of the family.

We first turn our attention to estimating the QTE across different income levels. Table 4.3 is similar to Table 4.2 although it shows separate results for high- and low-income families. As discussed previously in Section 4.1, the high income group includes households with income above $75,000 and we combine the low and middle income groups to form a group of households with income below $75,000. As expected, high income households in the control group consume more electricity between 2 pm and 7 pm than low income households in the control group. The differential is fairly constant across quantiles. It is very interesting to discover that the results for the other groups are exactly the opposite: the coefficient estimates for high income consumers are smaller than the coefficient estimates for low income consumers. This suggests that high income customers are more successful in taking advantage of the existing information about price and consumption, and consequently, engage in larger electricity savings. This may not be a pure behavioral effect and may

		QMG					CCEMG
		0.10	0.25	0.50	0.75	0.90	
				Control Group			
High Income	Consumption at	0.468	0.580	0.621	0.490	0.367	0.482
	$t-1$ (in logs)	(0.032)	(0.035)	(0.036)	(0.034)	(0.025)	(0.009)
	Treatment	0.143	0.118	0.069	0.054	0.042	0.097
		(0.015)	(0.013)	(0.009)	(0.008)	(0.008)	(0.017)
Low Income	Consumption at	0.462	0.568	0.613	0.469	0.344	0.469
	$t-1$ (in logs)	(0.025)	(0.026)	(0.026)	(0.024)	(0.018)	(0.009)
	Treatment	0.130	0.093	0.053	0.038	0.033	0.079
		(0.012)	(0.009)	(0.007)	(0.009)	(0.008)	(0.021)
				Portal			
High Income	Consumption at	0.463	0.579	0.616	0.479	0.351	0.481
	$t-1$ (in logs)	(0.029)	(0.031)	(0.033)	(0.030)	(0.020)	(0.009)
	Treatment	0.071	0.053	0.033	0.011	-0.010	0.033
		(0.017)	(0.017)	(0.014)	(0.017)	(0.021)	(0.019)
Low Income	Consumption at	0.474	0.594	0.645	0.492	0.371	0.495
	$t-1$ (in logs)	(0.030)	(0.033)	(0.035)	(0.032)	(0.024)	(0.009)
	Treatment	0.096	0.069	0.042	0.030	0.014	0.057
		(0.020)	(0.014)	(0.011)	(0.009)	(0.011)	(0.019)
				IHD			
High Income	Consumption at	0.469	0.581	0.607	0.479	0.364	0.486
	$t-1$ (in logs)	(0.032)	(0.036)	(0.038)	(0.037)	(0.025)	(0.009)
	Treatment	0.074	0.050	0.021	0.006	-0.026	0.018
		(0.019)	(0.019)	(0.016)	(0.014)	(0.019)	(0.022)
Low Income	Consumption at	0.470	0.576	0.618	0.467	0.339	0.471
	$t-1$ (in logs)	(0.031)	(0.035)	(0.038)	(0.034)	(0.027)	(0.009)
	Treatment	0.101	0.079	0.053	0.039	0.019	0.063
		(0.022)	(0.015)	(0.011)	(0.010)	(0.011)	(0.019)
				PCT			
High Income	Consumption at	0.708	0.785	0.808	0.685	0.536	0.675
	$t-1$ (in logs)	(0.033)	(0.027)	(0.024)	(0.028)	(0.026)	(0.008)
	Treatment	-0.096	-0.064	-0.036	-0.042	-0.052	-0.080
		(0.030)	(0.023)	(0.016)	(0.017)	(0.022)	(0.015)
Low Income	Consumption at	0.724	0.781	0.800	0.700	0.585	0.685
	$t-1$ (in logs)	(0.036)	(0.031)	(0.029)	(0.033)	(0.032)	(0.007)
	Treatment	-0.067	-0.041	-0.019	-0.017	-0.023	-0.055
		(0.028)	(0.021)	(0.012)	(0.011)	(0.016)	(0.016)

TABLE 4.3. *Quantile Mean Group estimator results by Income Levels. CCEMG denotes the Common Correlated Mean Group estimator due to Chudik and Pesaran (2015). IHD denotes in-home display and PCT is programmable communicating thermostats. Standard errors are in parentheses.*

		QMG					CCEMG
		0.10	0.25	0.50	0.75	0.90	
				Control Group			
Family years	Consumption at	0.492	0.616	0.663	0.513	0.369	0.505
	$t-1$ (in logs)	(0.028)	(0.029)	(0.030)	(0.028)	(0.021)	(0.009)
	Treatment	0.146	0.107	0.059	0.047	0.032	0.091
		(0.013)	(0.010)	(0.007)	(0.007)	(0.008)	(0.018)
Young years	Consumption at	0.438	0.531	0.571	0.442	0.337	0.444
	$t-1$ (in logs)	(0.029)	(0.029)	(0.030)	(0.027)	(0.020)	(0.009)
	Treatment	0.124	0.098	0.060	0.042	0.041	0.081
		(0.013)	(0.011)	(0.009)	(0.010)	(0.008)	(0.021)
				Portal			
Family years	Consumption at	0.470	0.599	0.651	0.505	0.363	0.500
	$t-1$ (in logs)	(0.033)	(0.035)	(0.036)	(0.034)	(0.024)	(0.009)
	Treatment	0.093	0.071	0.047	0.035	0.023	0.056
		(0.019)	(0.015)	(0.011)	(0.012)	(0.013)	(0.018)
Young years	Consumption at	0.465	0.576	0.612	0.470	0.357	0.477
	$t-1$ (in logs)	(0.027)	(0.030)	(0.032)	(0.029)	(0.020)	(0.009)
	Treatment	0.073	0.052	0.030	0.008	-0.016	0.034
		(0.016)	(0.016)	(0.014)	(0.016)	(0.020)	(0.020)
				IHD			
Family years	Consumption at	0.507	0.644	0.693	0.538	0.384	0.534
	$t-1$ (in logs)	(0.029)	(0.030)	(0.031)	(0.032)	(0.025)	(0.009)
	Treatment	0.068	0.049	0.026	0.012	-0.008	0.027
		(0.023)	(0.018)	(0.012)	(0.010)	(0.016)	(0.018)
Young years	Consumption at	0.439	0.525	0.547	0.420	0.326	0.433
	$t-1$ (in logs)	(0.032)	(0.037)	(0.040)	(0.037)	(0.026)	(0.009)
	Treatment	0.103	0.077	0.046	0.030	0.000	0.051
		(0.021)	(0.016)	(0.014)	(0.013)	(0.015)	(0.020)
				PCT			
Family years	Consumption at	0.718	0.778	0.799	0.688	0.558	0.678
	$t-1$ (in logs)	(0.037)	(0.030)	(0.026)	(0.032)	(0.032)	(0.008)
	Treatment	-0.072	-0.045	-0.020	-0.020	-0.012	-0.052
		(0.033)	(0.024)	(0.014)	(0.016)	(0.022)	(0.015)
Young years	Consumption at	0.715	0.786	0.807	0.695	0.563	0.682
	$t-1$ (in logs)	(0.032)	(0.029)	(0.027)	(0.030)	(0.028)	(0.007)
	Treatment	-0.087	-0.057	-0.032	-0.035	-0.052	-0.076
		(0.025)	(0.020)	(0.013)	(0.013)	(0.018)	(0.016)

TABLE 4.4. *Quantile Mean Group estimator results by Family Stages. CCEMG denotes the CCEMG denotes the Common Correlated Mean Group estimator due to Chudik and Pesaran (2015). IHD denotes in-home display and PCT is programmable communicating thermostats. Standard errors are in parentheses.*

come from the fact that high income consumers have not only larger cooling systems but perhaps also more sophisticated ones which can achieve higher savings. This is true for all quantiles and groups. When we compare the evidence in Table 4.3 with the evidence presented in Table 4.2, we find that the effect of PCT continues to be negative but it is now significant at the 0.9 quantile for high income households and insignificant for low income households. Thus, high-income customers who are conditionally consuming high levels of electricity reduce consumption by 5.1% relative to other times of the day and by roughly 9.3% relative to the control group in the period 2 pm to 7 pm.

Lastly, we investigate how households at different life stages respond to TOU pricing and the different technologies. In Table 4.4, the group called "family life" includes middle aged families with children, while "other years" refers to younger households under 45 years of age and no children and customers typically over 65 years of age. Again, as in the previous table, we see considerable response heterogeneity by group demographics. For instance, we find larger energy savings among families with no children who were provided a PCT, with the gains ranging from 3.1% at the 0.5 quantile to 8.3% at the 0.10 quantile. However, PCT does not seem to be an effective technology for middle aged families at the upper quantiles of the conditional distribution of electricity consumption.

4.5. **A Counterfactual Exercise**

In practice, regulators and electric utility managers must balance several concerns when implementing dynamic pricing strategies. Considerations range from the peak price level, the variability of prices over the course of the day, and the determination of days when the utility ought to increase prices to critical peak levels (often several times the baseline off-peak price) in order to prevent blackouts. These decisions are complex and it is important to base their conclusions on sound data driven counterfactual simulations.

Models such as the one developed in this paper can play an important role in evaluating relevant counterfactuals and allowing decision makers to choose optimal data driven strategies. While it is beyond the scope of our paper to provide an in-depth exploration of the menu of strategies available to a utility, we will briefly exemplify the process by evaluating a scenario where the utility decides to execute the peak pricing option only if temperature exceeds a certain threshold. This is usually coupled with further prediction models which may indicate that on days where the temperature is high the risk of a blackout also increases substantially. Thus, while utilities have to avoid this very costly scenario, they also have to balance their responsibilities towards their consumers. Daily peak prices may avoid blackouts, but will also cost consumers extra money and can lead to unhappy customers, when the rationale for higher prices is decoupled from the risk of a blackout. Many utilities have in fact opted to employ similar strategies in recent years which are commonly labeled as "variable peak pricing" rates.

Using our model, we can explore a series of counterfactuals. We create a decision rule that deviated from the actual policy, by only switching on the counterfactual policy if temperature exceeds a certain threshold defined as percentiles of the temperature distribution. In this simplified example, we consider actual temperature, though in the real world this strategy would be implemented using a secondary prediction algorithms for the temperature a few days ahead. Thus, we contrast counterfactual policies which are turned on if the temperature exceeds the 90th, 50th, and 25th percentiles, respectively. To understand the rationale, we can imagine that reasoning behind turning on the peak prices if temperature exceeds the 90th percentile is a way of explaining to consumers that they will be subjected to higher prices only on very hot days where the risk of a blackout is significantly greater than on a regular day.

For simplicity, we compare the baseline policy and counterfactual policies for customers with a PCT and investigate the response heterogeneity by considering households at both the top 90th quantile and bottom 10th quantile of the conditional usage distribution (Figure 4.2). Since in practice it is often required to display results in terms of kWh load curves over the course of the day we do so in the figures below for each policy while also reporting the percent change in electricity usage relative to the actual baseline policy.

We see that the counterfactual policies reduce savings during the peak hours as a function of the threshold at which they are implemented. The reductions are, however, relatively minor indicating that there may be a gain in efficiency from targeting only the hottest days (which is consistent with current practice by many utilities). Less strict counterfactuals also result in lower levels of off-peak load shifting during the evening and night hours.

5. Conclusions

In this paper, we extend the Common Correlated Effects (CCE) approach of Pesaran (2006) and Chudik and Pesaran (2015) to the estimation and inference of dynamic panel quantile regression models with interactive effects. We propose a new quantile estimator and show that it is consistent and asymptotically normal under standard regularity conditions in the quantile and dynamic linear

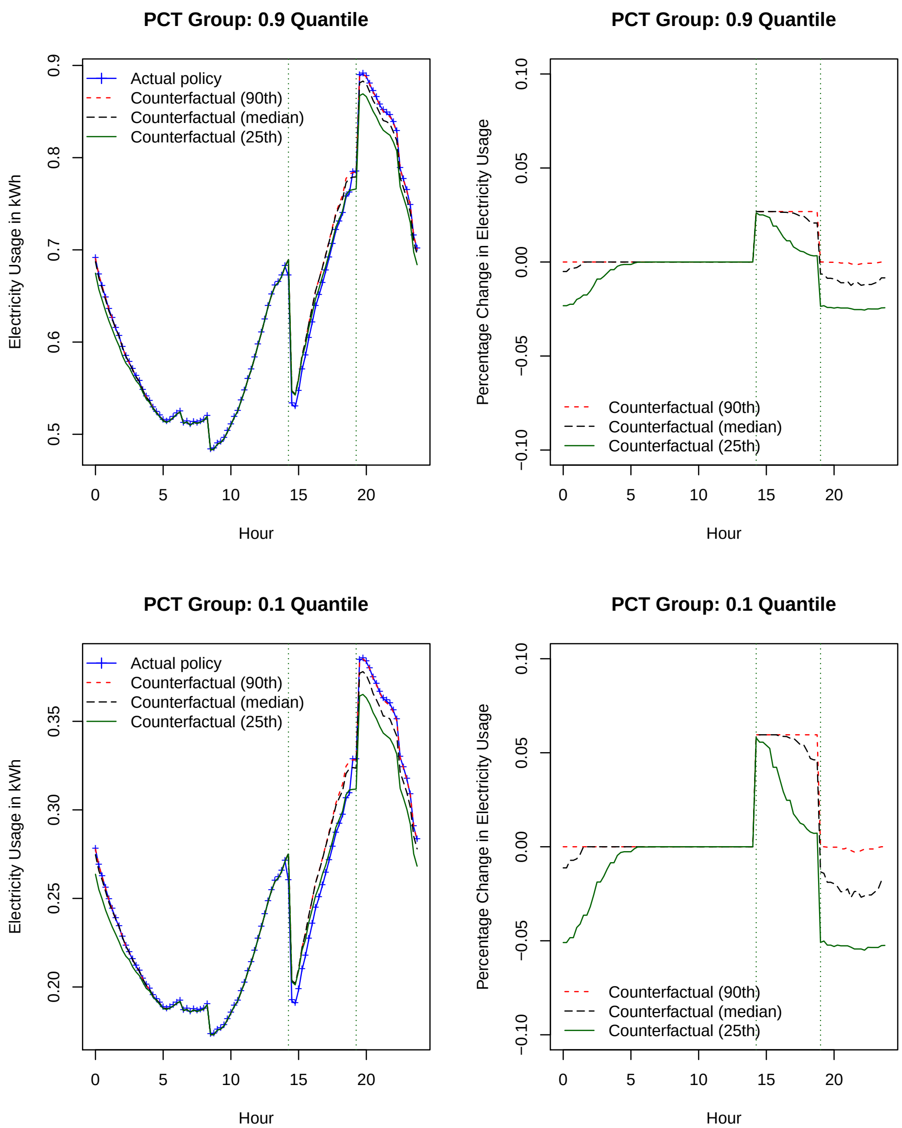

FIGURE 4.2. *Counterfactual policies for customers with a PCT. The right panels show the percentage change in electricity usage with respect to the actual policy.*

panel literatures. We require, however, a larger T/N for inference as compared to the standard CCEMG estimators developed for linear panel data models. An important condition is that the individual models need to be augmented by a sufficiently large number of lagged cross section averages that proxy the unobserved common effects. We also show that the approach offers good finite sample performance in the class of dynamic quantile regression estimators, as long as the time series dimension of the panel is large. Lastly, we demonstrate how the approach can be used in practice by documenting how the use of different technologies that allow consumers to be informed about electricity prices and consumption are associated with energy savings. Using data from a large scale randomized experiment that contains more than 6 million observations, we semiparametrically estimate a dynamic equation for electricity consumption with slope heterogeneity and cross-sectional dependence. The results offer several new insights useful for policy, while illustrating that the average effect does not summarize the distributional effect of the technologies.

Several directions remain to be investigated. Inference procedures are proposed but they require a detailed investigation in the case of long run effects. Moreover, although T is relatively large in our empirical application, offering an estimation approach that helps to reduce potential biases in short T applications seems of fundamental importance. A bias-corrected mean quantile group estimator is being investigated for the case of heterogeneous quantile coefficients following closely the approach developed in Chudik and Pesaran (2015).

References

Ambrosetti, A., and G. Prodi (1993): *A Primer of Nonlinear Analysis*, vol. 34. Cambridge Studies in Advanced Mathematics.

Ando, T., and J. Bai (2017): "Quantile Co-Movement in Financial Markets; a Panel Quantile Model with Unobserved Heterogeneity," 30th Australasian Finance and Banking Conference. Available at SSRN: https://ssrn.com/abstract=2953039.

Arellano, M., and S. Bonhomme (2016): "Nonlinear panel data estimation via quantile regressions," *The Econometrics Journal*, 19(3), C61–C94.

Bai, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77(4), 1229–1279.

Chen, L., J. J. Dolado, and J. Gonzalo (2017): "Quantile Factor Models," UC3M Working Papers, Universidad Carlos III de Madrid.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

Chernozhukov, V., I. Fernández-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015): "Nonparametric identification in panels using quantiles," *Journal of Econometrics*, 188(2), 378 – 392.

Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491 – 525.

Chudik, A., K. Mohaddes, M. Pesaran, and M. Raissi (2017): "Is There a Debt-Threshold Effect on Output Growth?," *Review of Economics and Statistics*, 99(1), 135–150.

Chudik, A., and M. H. Pesaran (2013): "Econometric Analysis of High Dimensional VARs Featuring a Dominant Unit," *Econometric Reviews*, 32(5-6), 592–649.

——— (2015): "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors," *Journal of Econometrics*, 188(2), 393 – 420.

Galvao, A. F. (2011): "Quantile regression for dynamic panel data with fixed effects," *Journal of Econometrics*, 164(1), 142 – 157.

Galvao, A. F., C. Lamarche, and L. R. Lima (2013): "Estimation of Censored Quantile Regression for Panel Data With Fixed Effects," *Journal of the American Statistical Association*, 108(503), 1075–1089.

Galvao, A. F., and L. Wang (2015): "Efficient minimum distance estimator for quantile regression fixed effects panel data," *Journal of Multivariate Analysis*, 133(C), 1–26.

Hahn, J., and G. Kuersteiner (2011): "Bias Reduction For Dynamic Nonlinear Panel Models With Fixed Effects," *Econometric Theory*, 27(06), 1152–1191.

Harding, M., and C. Lamarche (2014): "Estimating and testing a quantile regression model with interactive effects," *Journal of Econometrics*, pp. 101 – 113.

——— (2016): "Empowering Consumers Through Data and Smart Technology: Experimental Evidence on the Consequences of Time-of-Use Electricity Pricing Policies," *Journal of Policy Analysis and Management*, 35(4), 906–931.

——— (2017): "Penalized Quantile Regression with Semiparametric Correlated Effects: An Application with Heterogeneous Preferences," *Journal of Applied Econometrics*, 32(2), 342–358.

Harding, M., and S. Sexton (2017): "Household Response to Time-Varying Electricity Prices," *Annual Review of Resource Economics*, 9, 337–59.

Hendricks, W., and R. Koenker (1992): "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association*, 87(417), 58–68.

Ito, K. (2014): "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing," *American Economic Review*, 104(2), 537–563.

Jessoe, K., and D. Rapson (2014): "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use," *American Economic Review*, 4(104), 1417–1438.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011): "Panels with non-stationary multifactor error structures," *Journal of Econometrics*, 160(2), 326 – 348.

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.

Knight, K. (1998): "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *Annals of Statistics*, 26, 755–770.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R., and Z. Xiao (2006): "Quantile Autoregression," *Journal of the American Statistical Association*, 101(475), 980–990.

Lamarche, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408.

Moon, H. R., and M. Weidner (2015): "Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83(4), 1543–1579.

——— (2017): "Dynamic Linear Panel Regression Models with Interactive Effects," *Econometric Theory*, pp. 158–195.

Newey, W. K., and D. McFadden (1994): "Chapter 36 Large sample estimation and hypothesis testing," vol. 4 of *Handbook of Econometrics*, pp. 2111 – 2245. Elsevier.

Pesaran, M., and R. Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels," *Journal of Econometrics*, 68(1), 79 – 113.

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

Pesaran, M. H., and A. Chudik (2014): "Aggregation in large dynamic panels," *Journal of Econometrics*, 178, Part 2(0), 273 – 285.

Pesaran, M. H., L. V. Smith, and T. Yamagata (2013): "Panel unit root tests in the presence of a multifactor error structure," *Journal of Econometrics*, 175(2), 94 – 115.

Pesaran, M. H., and E. Tosetti (2011): "Large panels with common factors and spatial correlation," *Journal of Econometrics*, 161(2), 182 – 202.

Rosen, A. M. (2012): "Set identification via quantile restrictions in short panels," *Journal of Econometrics*, 166(1), 127 – 137.

Appendix A. Mathematical Proofs

A.1. Notations and Definitions

The proofs make use of Knight's (1998) identity: $\rho_\tau(u-v)-\rho_\tau(u) = -v\psi_\tau + \int_0^v (I(v \le s) - I(v \le 0))ds$, where $\rho_\tau = u(\tau - I(u \le 0))$ is the quantile regression check function and $\psi_\tau(u) = \tau - I(u \le 0)$ is the quantile influence function.

Throughout this appendix, we omit, at times, τ in $\boldsymbol{\pi}_i(\tau)$ for notational simplicity. Recall that $\boldsymbol{\pi}_i := (\lambda_i, \boldsymbol{\beta}_i', \alpha_i, \boldsymbol{\delta}_i')'$ where $\boldsymbol{\delta}_i := (\boldsymbol{\delta}_{i1}', \boldsymbol{\delta}_{i2}', \ldots, \boldsymbol{\delta}_{ip_T}')'$. We denote the dimension of the vector $\boldsymbol{\pi}_i$ by $p = (2+p_x) + (1+p_T)(1+p_x)$. Also $\mathbf{X}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \bar{\mathbf{z}}_t', \bar{\mathbf{z}}_{t-1}', ..., \bar{\mathbf{z}}_{t-p_T}')'$ and $\bar{\mathbf{z}}_t := (\bar{y}_t, \bar{\mathbf{x}}_t')'$. We define $\Delta_i(\boldsymbol{\pi}_i) = \mathbb{M}_i(\boldsymbol{\pi}_i) - \mathbb{M}_i(\boldsymbol{\pi}_{i0})$ and

$$\begin{aligned}
\mathbb{M}_i(\boldsymbol{\pi}_i) &:= \frac{1}{T-p_T}\sum_{t=1+p_T}^{T} \rho_\tau\left(y_{it} - \mathbf{X}_{it}'\boldsymbol{\pi}_i\right), \quad \mathbb{M}_i(\boldsymbol{\pi}_{i0}) := \frac{1}{T-p_T}\sum_{t=1+p_T}^{T} \rho_\tau\left(y_{it} - \mathbf{X}_{it}'\boldsymbol{\pi}_{i0}\right), \\
\mathbb{H}_i(\boldsymbol{\pi}_i) &:= \frac{1}{T-p_T}\sum_{t=1+p_T}^{T} \psi_\tau\left(y_{it} - \mathbf{X}_{it}'\boldsymbol{\pi}_i\right)\mathbf{X}_{it}, \quad H_i(\boldsymbol{\pi}_i) := E(\mathbb{H}_i(\boldsymbol{\pi}_i)) = E\left[\tau - G_i(\cdot|\mathbf{X}_{it})\right],
\end{aligned}$$

$$\mathbf{J}_i \quad := \quad \frac{\partial H_i(\boldsymbol{\pi}_i)}{\partial \boldsymbol{\pi}_i} = E\left[g_i(0|\mathbf{X}_{it})\mathbf{X}_{it}\mathbf{X}_{it}'\right].$$

We define the two conditional quantile functions:

$$y_{it} = \alpha_i(\tau) + \lambda_i(\tau)y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta}_i(\tau) + \mathbf{f}_t'\boldsymbol{\gamma}_i(\tau) + u_{it}(\tau) \tag{A.1}$$

$$y_{it} = \alpha_i(\tau) + \lambda_i(\tau)y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta}_i(\tau) + \sum_{l=0}^{p_T} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau) + e_{it}(\tau), \tag{A.2}$$

where $e_{it}(\tau) = u_{it}(\tau) + \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau) + O_p(N^{-1/2})$. See (2.17) and (2.18). Let $\mathbf{W}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \mathbf{f}_t')'$ and recall that $\mathbf{X}_{it} = (y_{it-1}, \mathbf{x}_{it}', \ddot{\mathbf{z}}_t')'$, where $\ddot{\mathbf{z}}_t = (1, \bar{\mathbf{z}}_t', \bar{\mathbf{z}}_{t-1}', ..., \bar{\mathbf{z}}_{t-p_T}')'$ with p_T sufficiently large. For any value y, y_{-1}, $\mathbf{x}$, $\mathbf{f}$, and α, we define,

$$Q(\mathbf{W}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}, \tau) := E(\rho_\tau(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \mathbf{f}'\boldsymbol{\gamma})\mathbf{W}) \tag{A.3}$$

$$Q(\mathbf{X}, \lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}, \tau) := E\left[\rho_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \sum_{l=0}^{p_T} \bar{\mathbf{z}}_{-l}'\boldsymbol{\delta}_l\right)\mathbf{X}\right]. \tag{A.4}$$

A.2. **Proofs**

Proof of Theorem 1. We first show that $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}')$ uniquely solve the quantile regression problem for all τ in the limit (i.e., as $N, T \to \infty$), which implies that the quantile score function can be set to zero. For that, we define:

$$\begin{aligned}
\boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}) &:= E\left[\mathbf{W}\psi_\tau(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \mathbf{f}_t'\boldsymbol{\gamma})\right] \\
\mathbf{J}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}) &:= \frac{\partial}{\partial(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}')}\boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}).
\end{aligned}$$

Our argument proceeds as follows. We first show that $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}')$ is the unique solution of the quantile regression minimization problem that includes the true factors. We then show that $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\delta}')$ is the unique solution of the quantile regression minimization problem that includes cross-sectional averages to approximate the true factors. Finally, we show that $(\lambda, \boldsymbol{\beta}')$ is the unique solution of both minimization problems.

The uniqueness of $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}')$ can be shown using similar arguments as in Chernozhukov and Hansen (2006, Theorem 2). Under Assumption 8, $\mathbf{J}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma})$ is continuous in $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}')$ and has full rank. The image of the space that includes all possible solutions of $\boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}) = \mathbf{0}$, contained in the interior of $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{G}$, is assumed to be simply connected under the mapping $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}') \mapsto \boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma})$. By the Monodromy Theorem 1.8 in Ambrosetti and Prodi (1993, p. 47), the mapping $\boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma})$ is one-to-one (homeomorphism) from $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{G}$ to $\boldsymbol{\Pi}(\Lambda, \mathcal{B}, \mathcal{A}, \mathcal{G})$.

Therefore, the true parameters $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\gamma}')$ uniquely solve $\boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}) = \mathbf{0}$. Denote the unique solution of $\boldsymbol{\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}) = \mathbf{0}$ by $(\lambda^*, \boldsymbol{\beta}^*, \alpha^*, \boldsymbol{\gamma}^*)$.

We now show the uniqueness of $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\delta}')$. Let

$$\begin{aligned}
\tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}) &:= E\left[\mathbf{X}\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \sum_{l=0}^{p_T} \bar{\mathbf{z}}'_{t-l}\boldsymbol{\delta}_{il}\right)\right] \\
\tilde{\mathbf{J}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}) &= \frac{\partial}{\partial(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\delta}')}\tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}).
\end{aligned}$$

The matrix $\tilde{\mathbf{J}}$ is continuous on $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\delta}')$ and has full rank under Assumption 8. The image of $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$ under the mapping $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\delta}') \mapsto \tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta})$ is assumed to be simply connected. By Theorem 1.8 in Ambrosetti and Prodi (1993), the mapping $\tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta})$ is one-to-one between $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$ and $\tilde{\boldsymbol{\Pi}}(\Lambda, \mathcal{B}, \mathcal{A}, \mathcal{D})$, the image of $\Lambda \times \mathcal{B} \times \mathcal{A} \times \mathcal{D}$ under $\tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta})$. Therefore, the parameter $(\lambda, \boldsymbol{\beta}', \alpha, \boldsymbol{\delta}')$ is the unique solution of $\tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}) = \mathbf{0}$. Denote the unique solution of $\tilde{\boldsymbol{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}) = \mathbf{0}$ by $(\lambda^\dagger, \boldsymbol{\beta}^\dagger, \alpha^\dagger, \boldsymbol{\delta}^\dagger)$. Note that the number of parameters depends on p_T, but we supress this dependence for simplicity.

We now show that $(\lambda, \boldsymbol{\beta}')$ in equation (A.1) can be identified from equation (A.2). Because $\mathbf{f} = \mathbf{G}(L)\bar{\mathbf{z}} + O_p(N^{-1/2})$ (recall that we have set $\mathbf{f}_0 = \mathbf{0}$; see Remark 1 in Section 2.1), we can express the latent factors as (see equation (2.15)),

$$\begin{aligned}
\mathbf{f} &= \sum_{l=0}^{\infty} \mathbf{G}'_l\bar{\mathbf{z}}_{-l} + O_p(N^{-1/2}) = \mathbf{G}'_0\bar{\mathbf{z}} + \mathbf{G}'_1\bar{\mathbf{z}}_{-1} + \ldots + O_p(N^{-1/2}) \\
&= \mathbf{H}\dot{\mathbf{z}} + \boldsymbol{\eta} + O_p(N^{-1/2}),
\end{aligned}$$

where $\mathbf{H}$ is a $r \times (1 + p_x)(p_T + 1)$ matrix of reduced form coefficients, $\dot{\mathbf{z}} = (\bar{\mathbf{z}}', \bar{\mathbf{z}}'_{-1}, \ldots, \bar{\mathbf{z}}'_{-p_T})'$ is a $(1 + p_x)(p_T + 1) \times 1$ vector of cross-sectional averages, and $\boldsymbol{\eta} = \sum_{l=p_T+1}^{\infty} \mathbf{G}'_l\bar{\mathbf{z}}_{-l}$. Solving for $\dot{\mathbf{z}}$ and letting $\dot{\mathbf{H}} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$, we obtain, $\dot{\mathbf{z}} = \dot{\mathbf{H}}(\mathbf{f} - \boldsymbol{\eta}) + O_p(N^{-1/2})$. Without loss of generality, we set p_T to be sufficiently large such that, under Assumption 5, we have that:

$$\left\|\sum_{l=p_T+1}^{\infty} \mathbf{G}'_l\bar{\mathbf{z}}_{-l}\right\| \leq K\rho^{p_T}\sum_{l=1}^{\infty}\left\|\rho^l\bar{\mathbf{z}}_{-(p_T+l)}\right\|.$$

Note also that for each $1 \leq i \leq N$, as in equation (A.18) in Chudik and Pesaran (2015), $\boldsymbol{\eta}_i = \{\sum_{l=p_T+1}^{\infty} \boldsymbol{\delta}'_{il}\bar{\mathbf{z}}_{p_T+1-l}, \sum_{l=p_T+1}^{\infty} \boldsymbol{\delta}'_{il}\bar{\mathbf{z}}_{p_T+2-l}, \ldots\}$. By Lemma A.4 in Chudik and Pesaran (2015), $\dot{\mathbf{H}}\boldsymbol{\eta}$

becomes asymptotically negligible as $N, T \to \infty$. We can therefore write,

$$\begin{pmatrix} 1 & \mathbf{0} & 0 & \mathbf{0} \\ 0 & \mathbf{I}_{p_x} & 0 & \mathbf{0} \\ 0 & \mathbf{0} & 1 & \mathbf{0} \\ 0 & \mathbf{0} & 0 & \dot{\mathbf{H}} \end{pmatrix} \begin{pmatrix} y_{-1} \\ \mathbf{x} \\ 1 \\ \mathbf{f} \end{pmatrix} = \begin{pmatrix} y_{-1} \\ \mathbf{x} \\ 1 \\ \dot{\mathbf{z}} \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{0} \\ 0 \\ \mathbf{f} - \mathbf{G}(L)\bar{\mathbf{z}} \end{pmatrix},$$

Moreover, by (A.19) in Lemma A.4 in Chudik and Pesaran (2015), the second term on the right hand side is asymptotically negligible. Therefore, for sufficiently large N and T, we can write $\mathbf{\Psi W} = \mathbf{X}$, where $\mathbf{W} = (y_{-1}, \mathbf{x}', 1, \mathbf{f}')'$, $\mathbf{X} = (y_{-1}, \mathbf{x}', 1, \dot{\mathbf{z}}')'$, and

$$\mathbf{\Psi} = \begin{pmatrix} 1 & \mathbf{0} & 0 & \mathbf{0} \\ 0 & \mathbf{I}_{p_x} & 0 & \mathbf{0} \\ 0 & \mathbf{0} & 1 & \mathbf{0} \\ 0 & \mathbf{0} & 0 & \dot{\mathbf{H}} \end{pmatrix}.$$

The matrix $\mathbf{\Psi}$ is invertible by Assumptions 5 and 6 (exponentially decaying coefficients and full rank condition of the matrix $\mathbf{C}$).

Also, under Assumption 5 for sufficiently large p_T, it follows that,

$$E\left[\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \sum_{l=0}^{p_T} \boldsymbol{\delta}_l' \bar{\mathbf{z}}_{-l}\right)\right] = E\left(\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \boldsymbol{\gamma}'\mathbf{G}(L)\bar{\mathbf{z}}\right)\right).$$

Therefore, the quantile regression problem can be written as

$$\begin{aligned} \tilde{\mathbf{\Pi}}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\delta}) &= E\left(\mathbf{X}\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \sum_{l=0}^{p_T} \bar{\mathbf{z}}_{t-l}' \boldsymbol{\delta}_{il}\right)\right) \\ &= E\left(\mathbf{X}\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \boldsymbol{\gamma}'\mathbf{G}(L)\bar{\mathbf{z}}\right)\right) \\ &= E\left(\mathbf{\Psi W}\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \mathbf{f}'\boldsymbol{\gamma}\right)\right) \\ &= \mathbf{\Psi} E\left(\mathbf{W}\psi_\tau\left(y - \lambda y_{-1} - \mathbf{x}'\boldsymbol{\beta} - \alpha - \mathbf{f}'\boldsymbol{\gamma}\right)\right) = \mathbf{\Psi\Pi}(\lambda, \boldsymbol{\beta}, \alpha, \boldsymbol{\gamma}) = \mathbf{0}. \end{aligned}$$

It follows that $\tilde{\mathbf{\Pi}}(\lambda^*, \boldsymbol{\beta}^*, \alpha^*, \boldsymbol{\delta}^*) = \mathbf{\Psi\Pi}(\lambda^\dagger, \boldsymbol{\beta}^\dagger, \alpha^\dagger, \boldsymbol{\gamma}^\dagger) = \mathbf{\Psi\Pi}(\lambda^*, \boldsymbol{\beta}^*, \alpha^\dagger, \boldsymbol{\gamma}^\dagger)$ because $\lambda^* = \lambda^\dagger$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}^\dagger$. We show this result by contradiction. Suppose now that there are not equal, i.e. $\lambda^* \neq \lambda^\dagger$ and $\boldsymbol{\beta}^* \neq \boldsymbol{\beta}^\dagger$. We now have,

$$\tilde{\mathbf{\Pi}}(\lambda^*, \boldsymbol{\beta}^*, \alpha^*, \boldsymbol{\delta}^*) = \mathbf{0} = \mathbf{\Psi\Pi}(\lambda^*, \boldsymbol{\beta}^*, \alpha^\dagger, \boldsymbol{\gamma}^\dagger).$$

Therefore, $\mathbf{\Pi}(\lambda^*, \boldsymbol{\beta}^*, \alpha^\dagger, \boldsymbol{\gamma}^\dagger) = \mathbf{\Psi}^{-1}\mathbf{0} = \mathbf{0}$, which is a contradiction to the uniqueness of $(\lambda^\dagger, \boldsymbol{\beta}^\dagger, \alpha^\dagger, \boldsymbol{\gamma}^\dagger)$. Therefore, $\lambda^* = \lambda^\dagger$ and $\boldsymbol{\beta}^* = \boldsymbol{\beta}^\dagger$, and the parameter of interest $\boldsymbol{\vartheta}(\tau) = (\lambda(\tau), \boldsymbol{\beta}(\tau)')'$ is uniquely identified for all τ. □

Proof of Theorem 2. The proof is divided in two parts. First, we show that the estimated factors are uniformly consistent for $i = 1, \ldots, N$ and the limiting problem corresponding to the model with augmented cross-sectional averages is $E(\rho_\tau(y_{it} - \alpha_i - \lambda_i y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i - \mathbf{f}'_t\gamma_i))$. Given the consistency of the approach, the second part of the proof establishes consistency of the reduced form coefficients, $\hat{\boldsymbol{\pi}}_i(\tau)$, for all τ.

[*Part 1: Consistency of estimated factors*] To show the uniform consistency of the estimator of the factor $\mathbf{f}_t$ in a quantile regression setting, we first note that

$$\begin{aligned} y_{it} &= \alpha_i(\tau) + \lambda_i(\tau) y_{it-1} + \mathbf{x}'_{it}\boldsymbol{\beta}_i(\tau) + \sum_{l=0}^{p_T} \bar{\mathbf{z}}'_{t-l}\boldsymbol{\delta}_{il}(\tau) + e_{it}(\tau) \\ &= \alpha_i(\tau) + \lambda_i(\tau) y_{it-1} + \mathbf{x}'_{it}\boldsymbol{\beta}_i(\tau) + \sum_{l=0}^{p_T} \bar{\mathbf{z}}'_{t-l}\boldsymbol{\delta}_{il}(\tau) + \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}'_{t-l}\boldsymbol{\delta}_{il}(\tau) + O_p(N^{-1/2}) + u_{it}(\tau). \end{aligned}$$

Let $\sum_{l=0}^{\infty} \bar{\mathbf{z}}'_{t-l}\bar{\boldsymbol{\delta}}_{il}(\tau) = \boldsymbol{\gamma}'_i(\tau)\bar{\mathbf{G}}(L)\bar{\mathbf{z}}_t$ which is obtained by a sample of size N. Similarly, $\sum_{l=0}^{p_T} \bar{\mathbf{z}}'_{t-l}\bar{\boldsymbol{\delta}}_{il}(\tau)$ is constructed with a sample of size N. Recall that $\boldsymbol{\gamma}'_i(\tau)(\mathbf{f}_t - \mathbf{G}(L)\bar{\mathbf{z}}_t) = O_p(N^{-1/2})$, and, under Assumptions 5 and 9:

$$\left\| \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}'_{-l}\boldsymbol{\delta}_{il}(\tau) \right\| \le K\rho^{p_T} \sum_{l=1}^{\infty} \left\| \rho^l \bar{\mathbf{z}}_{-(p_T+l)} \right\|,$$

which is asymptotically negligible under $(N, T, p_T) \to \infty$. Therefore, we can write the last equation as

$$y_{it} = \alpha_i(\tau) + \lambda_i(\tau) y_{it-1} + \mathbf{x}'_{it}\boldsymbol{\beta}_i(\tau) + \mathbf{f}'_t\gamma_i(\tau) + \phi_{it}(\tau) + u_{it}(\tau).$$

where $\phi_{it}(\tau) = \boldsymbol{\gamma}'_i(\tau)(\bar{\mathbf{G}}(L) - \mathbf{G}(L))\bar{\mathbf{z}}_t$. Define for each i,

$$\begin{aligned} Q_\infty(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= E(\rho_\tau(y_{it} - \alpha_i - \lambda_i y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i - \mathbf{f}'_t\boldsymbol{\gamma}_i)) \\ Q_T(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_{it} - \alpha_i - \lambda_i y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i - \mathbf{f}'_t\gamma_i) \\ \hat{Q}_T(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma}) &= \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_{it} - \alpha_i - \lambda_i y_{it-1} - \mathbf{x}'_{it}\boldsymbol{\beta}_i - \mathbf{f}'_t\gamma_i - \phi_{it}). \end{aligned}$$

We show the result in two steps. First, we prove that $Q_T(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ converges uniformly to $Q_\infty(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in $(\alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and τ. Second, we show that $\hat{Q}_T(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ converges uniformly to $Q_T(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ in $(\alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ and τ. It follows then that $\arg\min \hat{Q}_T(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$ converges to $\arg\min \hat{Q}_\infty(\tau, \alpha, \lambda, \boldsymbol{\beta}, \boldsymbol{\gamma})$.

[*Part 1: Step 1*] The first step is to show that

$$\sup_{\tau,\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma}} |Q_T(\tau,\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma}) - Q_\infty(\tau,\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma})| = o_p(1). \tag{A.5}$$

Note that $(\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma}) \mapsto \rho_\tau(y-\alpha-\lambda y_{-1}-\boldsymbol{\beta}'\mathbf{x}-\boldsymbol{\gamma}'\mathbf{f})$ is continuous for y, $\mathbf{x}$ and $\mathbf{f}$. Moreover, the dominating function corresponding to the quantile regression check function exists under Assumptions 9 and 10 and it is equal to $\rho_\tau(y-\alpha-\lambda y_{-1}-\boldsymbol{\beta}'\mathbf{x}-\boldsymbol{\gamma}'\mathbf{f}) \leq K(|\alpha|+|\lambda||y_{-1}|+\|\mathbf{x}\|\|\boldsymbol{\beta}\|+\|\boldsymbol{\gamma}\|\|\mathbf{f}\|)$. Then, using an extended version of Lemma 2.4 in Newey and McFadden (1994) for stationary processes, we can conclude that (A.5) holds (see footnote 18 to Lemma 2.4 of Newey and McFadden (1994)).

[*Part 1: Step 2*] The second part of the proof uses a version of Knight's (1998) inequality: $|\rho_\tau(u-v)-\rho_\tau(u)| \leq 3|v|$. Letting $u := y-\alpha-\lambda y_{-1}-\mathbf{x}'\boldsymbol{\beta}-\mathbf{f}'\boldsymbol{\gamma}$ and $v := \phi = \boldsymbol{\gamma}'(\tau)(\bar{\mathbf{G}}(L)-\mathbf{G}(L))\bar{\mathbf{z}}$, we have

$$\left|\hat{Q}_T(\tau,\cdot)-Q_T(\tau,\cdot)\right| = \left|\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(u_{it}-\phi_{it})-\rho_\tau(u_{it})\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\left|\boldsymbol{\gamma}_i'(\tau)(\bar{\mathbf{G}}(L)-\mathbf{G}(L))\bar{\mathbf{z}}_t\right|.$$

Since $\mathbf{G}(L) = \sum_{l=0}^{\infty}\mathbf{G}_l L^l$ and, $\bar{\mathbf{G}}(L) = \sum_{l=0}^{\infty}\bar{\mathbf{G}}_l L^l$, then

$$\frac{1}{T}\sum_{t=1}^{T}|\boldsymbol{\gamma}_i'(\tau)(\bar{\mathbf{G}}(L)-\mathbf{G}(L))\bar{\mathbf{z}}_t| \leq \sup_{i,t,\tau}\|\boldsymbol{\gamma}_i'(\tau)(\bar{\mathbf{G}}(L)-\mathbf{G}(L))\bar{\mathbf{z}}_t\| = o_p(1),$$

by Lemma 1 as long as $p_T^3/T \to 0$, and $(\log(N))^2/T \to 0$, as N and $T \to \infty$, jointly. Therefore, overall we have,

$$\sup_{(\tau,\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma})}\left|\hat{Q}_T(\tau,\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma})-Q_T(\tau,\alpha,\lambda,\boldsymbol{\beta},\boldsymbol{\gamma})\right| = o_p(1),$$

leading to the desired result for each τ.

[*Part 2: Consistency of reduced form coefficients*] For each $\eta > 0$, define the ball $\mathcal{B}_i(\eta) := \{\boldsymbol{\pi}_i : \|\boldsymbol{\pi}_i-\boldsymbol{\pi}_{i0}\|_1 \leq \eta\}$ and the boundary $\partial\mathcal{B}_i(\eta) := \{\boldsymbol{\pi}_i : \|\boldsymbol{\pi}_i-\boldsymbol{\pi}_{i0}\|_1 = \eta\}$. For each $\boldsymbol{\pi}_i \notin \mathcal{B}_i(\eta)$, define $\bar{\boldsymbol{\pi}}_i = r_i\boldsymbol{\pi}_i+(1-r_i)\boldsymbol{\pi}_{i0}$, where $r_i = \eta/\|\boldsymbol{\pi}_i-\boldsymbol{\pi}_{i0}\|$. Note that $\bar{\boldsymbol{\pi}}_i$ is in the boundary $\partial\mathcal{B}_i(\eta)$. Because the objective function is convex,

$$r_i\left(\mathbb{M}_i(\boldsymbol{\pi}_i)-\mathbb{M}_i(\boldsymbol{\pi}_{i0})\right) \geq \mathbb{M}_i(\bar{\boldsymbol{\pi}}_i)-\mathbb{M}_i(\boldsymbol{\pi}_{i0}) = \mathbb{M}_i(\bar{\boldsymbol{\pi}}_i) = E(\Delta_i(\bar{\boldsymbol{\pi}}_i))+\left(\mathbb{M}_i(\bar{\boldsymbol{\pi}}_i)-E(\Delta_i(\bar{\boldsymbol{\pi}}_i))\right), \tag{A.6}$$

Note that $\mathbb{M}_i(\boldsymbol{\pi}_{i0})$ is naturally equal to zero by definition and that $E(\Delta_i(\bar{\boldsymbol{\pi}}_i)) \geq \epsilon_\eta$ for all $1 \leq i \leq N$.

Consider now $\|\hat{\boldsymbol{\pi}}_i-\boldsymbol{\pi}_{i0}\|_1 > \eta$ which implies that $\hat{\boldsymbol{\pi}}_i \notin \mathcal{B}_i(\eta)$ for all $1 \leq i \leq N$. It follows that $\mathbb{M}_i(\hat{\boldsymbol{\pi}}_i) \leq \mathbb{M}_i(\boldsymbol{\pi}_{i0})$ for some $1 \leq i \leq N$ by definition of $\hat{\boldsymbol{\pi}}_i = \arg\min\{\mathbb{M}_i(\boldsymbol{\pi}_i)\}$, which is equivalent to (2.20).

Note that $\hat{\boldsymbol{\pi}}_i \notin \mathcal{B}_i(\eta)$ implies $\mathbb{M}(\hat{\boldsymbol{\pi}}_i) \leq 0$ by definition. Thus, by equation (A.6), the following inclusion relationships are true:

$$\left\{\max_{1\leq i\leq N} \|\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_{i0}\|_1 > \eta\right\} \subseteq \{\mathbb{M}(\boldsymbol{\pi}_i) \leq 0, \exists \boldsymbol{\pi}_i \notin \mathcal{B}_i(\eta)\} \subseteq \bigcup_{i=1}^{N} \left\{\sup_{\boldsymbol{\pi}_i \in \mathcal{B}_i(\eta)} |\Delta_i(\boldsymbol{\pi}_i) - E(\Delta_i(\boldsymbol{\pi}_i))| \geq \epsilon_\eta\right\}.$$

It follows that,

$$P\left\{\max_{1\leq i\leq N} \|\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_{i0}\|_1 > \eta\right\} \leq N \max_{1\leq i\leq N} P\left\{\sup_{\boldsymbol{\pi}_i \in \mathcal{B}_i(\eta)} |\Delta_i(\boldsymbol{\pi}_i) - E(\Delta_i(\boldsymbol{\pi}_i))| \geq \epsilon_\eta\right\}.$$

We therefore need to show that

$$\max_{1\leq i\leq N} P\left\{\sup_{\boldsymbol{\pi}_i \in \mathcal{B}_i(\eta)} |\Delta_i(\boldsymbol{\pi}_i) - E(\Delta_i(\boldsymbol{\pi}_i))| \geq \epsilon_\eta\right\} = o(N^{-1}), \tag{A.7}$$

which is similar to equation (A.3) in Kato, Galvao and Montes-Rojas (2012) and equation (15) in Galvao and Wang (2015). Recall that as $N \to \infty$, automatically $T \to \infty$ too.

Without loss of generality, we restrict all the balls $\mathcal{B}_i(\eta)$ to be equal to $\mathcal{B}(\eta)$ by setting $\boldsymbol{\pi}_{i0} = 0$. Thus, $\mathcal{B}_i(\eta) = \mathcal{B}(\eta)$ for all $1 \leq i \leq N$. We then suppress the subscript i for simplicity. Let $g_{\boldsymbol{\pi}}(u, \mathbf{X}) = \rho_\tau(u - \mathbf{X}'\boldsymbol{\pi}) - \rho_\tau(u)$. We observe that $|g_{\boldsymbol{\pi}}(u, \mathbf{X}) - g_{\bar{\boldsymbol{\pi}}}(u, \mathbf{X})| \leq C(1 + M)(\|\boldsymbol{\pi} - \bar{\boldsymbol{\pi}}\|_1)$, for some universal constant C. Since $\mathcal{B}(\eta)$ is a compact subset in $\mathbb{R}^p$, $\exists K$ ℓ_1 balls with center $\boldsymbol{\pi}^{(j)}$ and radius $\epsilon/3\kappa$ where $\kappa := C(1 + M)$.

For each $\boldsymbol{\pi} \in \mathcal{B}(\eta)$, there is $j \in \{1, ..., K\}$ such that,

$$|\Delta(\boldsymbol{\pi}) - E(\Delta(\boldsymbol{\pi}))| \leq |\Delta(\boldsymbol{\pi}^{(j)}) - E(\Delta(\boldsymbol{\pi}^{(j)}))| + \frac{2\epsilon}{3}. \tag{A.8}$$

The last inequality follows by a property of $g_{\boldsymbol{\pi}}(u, \mathbf{X})$. Notice that,

$$\begin{aligned}
|\Delta(\boldsymbol{\pi}) - E(\Delta(\boldsymbol{\pi}))| - |\Delta(\boldsymbol{\pi}^{(j)}) - E(\Delta(\boldsymbol{\pi}^{(j)}))| &\leq |\Delta(\boldsymbol{\pi}) - E(\Delta(\boldsymbol{\pi})) - \Delta(\boldsymbol{\pi}^{(j)}) + E(\Delta(\boldsymbol{\pi}^{(j)}))| \\
&\leq |\Delta(\boldsymbol{\pi}) - E(\Delta(\boldsymbol{\pi}))| + |\Delta(\boldsymbol{\pi}^{(j)}) + E(\Delta(\boldsymbol{\pi}^{(j)}))| \\
&\leq C(1 + M)\frac{\epsilon}{3\kappa} + C(1 + M)\frac{\epsilon}{3\kappa} = \frac{2}{3}\epsilon.
\end{aligned}$$

Therefore, following (A.8), we write,

$$\begin{aligned}
P\left(\sup_{\boldsymbol{\pi} \in \mathcal{B}(\eta)} |\Delta(\boldsymbol{\pi}) - E(\Delta(\boldsymbol{\pi}))| > \epsilon\right) &\leq \sum_{j=1}^{K} P\left(|\Delta(\boldsymbol{\pi}^{(j)}) - E(\Delta(\boldsymbol{\pi}^{(j)}))| + \frac{2}{3}\epsilon > \epsilon\right) \\
&= \sum_{j=1}^{K} P\left(|\Delta(\boldsymbol{\pi}^{(j)}) - E(\Delta(\boldsymbol{\pi}^{(j)}))| > \frac{\epsilon}{3}\right).
\end{aligned}$$

For each $j \in \{1, ..., K\}$, we note that $|\Delta(\boldsymbol{\pi}^{(j)}) - E(\Delta(\boldsymbol{\pi}^{(j)}))|$ satisfies the conditions of Lemma 2 in Section A.3. Taking $s = 2\log(N)$ and $q = [\sqrt{T}]$ and using the fact that $\log(N)/\sqrt{T} \to 0$,

$$P\left(|\Delta(\boldsymbol{\pi}^{(j)}) - E(\Delta(\boldsymbol{\pi}^{(j)}))|\right) \leq const \times \left(\exp\{-2\log(N)\} + \sqrt{T}Ba^{[\sqrt{T}]}\right),$$

where the constants $a \in (0, 1)$ and $B > 0$ (Assumption 9). Therefore,

$$P\left\{\sup_{\boldsymbol{\pi}_i \in \mathcal{B}_i(\eta)} |\Delta_i(\boldsymbol{\pi}_i) - E(\Delta_i(\boldsymbol{\pi}_i))| > \epsilon_\eta\right\} \leq K \times const \times \left(\exp\{-2\log(N)\} + \sqrt{T}Ba^{[\sqrt{T}]}\right). \quad \text{(A.9)}$$

Because by assumption $(\log(N))^2/T \to 0$, the right hand side of equation (A.9) is $o(N^{-1})$ which leads to the weak consistency result. □

Proof of Theorem 3. Under the stated assumptions, the results follows directly from Theorem 2. By definition, $\hat{\boldsymbol{\vartheta}}(\tau) = N^{-1}\sum_{i=1}^N \hat{\boldsymbol{\vartheta}}_i(\tau)$. Thus,

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau) &= \frac{1}{N}\sum_{i=1}^N \hat{\boldsymbol{\vartheta}}_i(\tau) - \boldsymbol{\vartheta}(\tau) = \frac{1}{N}\sum_{i=1}^N \left(\hat{\boldsymbol{\vartheta}}_i(\tau) - \boldsymbol{\vartheta}(\tau)\right) \\
&= \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Xi}_i \circ (\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau)) + \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Xi}_i \circ (\boldsymbol{\pi}_i(\tau) - \boldsymbol{\pi}(\tau)) = o_p(1),
\end{aligned}$$

The first term converges in probability to zero as established in Theorem 2 and the last equality follows by Assumption 5. □

Proof of Theorem 4. By definition, as in Theorem 3, we have

$$\begin{aligned}
\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau) &= \frac{1}{N}\sum_{i=1}^N (\hat{\boldsymbol{\vartheta}}_i(\tau) - \boldsymbol{\vartheta}(\tau)) = \frac{1}{N}\sum_{i=1}^N ((\hat{\boldsymbol{\vartheta}}_i(\tau) - \boldsymbol{\vartheta}_i(\tau)) + (\boldsymbol{\vartheta}_i(\tau) - \boldsymbol{\vartheta}(\tau))) \\
&= \frac{1}{N}\sum_{i=1}^N \boldsymbol{\Xi}_i \circ ((\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau)) + (\boldsymbol{\pi}_i(\tau) - \boldsymbol{\pi}(\tau)))
\end{aligned}$$

It follows that,

$$\sqrt{N}\left(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)\right) = \frac{\sqrt{N}}{N}\sum_{i=1}^N \boldsymbol{\Xi}_i \circ (\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau)) + \frac{\sqrt{N}}{N}\sum_{i=1}^N \boldsymbol{\Xi}_i \circ (\boldsymbol{\pi}_i(\tau) - \boldsymbol{\pi}(\tau)). \quad \text{(A.10)}$$

We now obtain the asymptotic representation of $\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau)$ following closely Galvao and Wang (2015). We use an expansion of $H_i(\hat{\boldsymbol{\pi}}_i) = E(\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i))$ around $\boldsymbol{\pi}_{i0}$ to obtain,

$$H_i(\hat{\boldsymbol{\pi}}_i) = H_i(\boldsymbol{\pi}_{i0}) + \mathbf{J}_i(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau)) + O_p\left[(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau))^2\right],$$

where $\mathbf{J}_i := \partial H_i(\boldsymbol{\pi}_i)/\partial \boldsymbol{\pi}_{i0} = E(g_i(0|\mathbf{X}_{it})\mathbf{X}_{it}\mathbf{X}_{it}'))$. Basic manipulations lead to:

$$\begin{aligned}
\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau) &= \mathbf{J}_i^{-1}\left(H_i(\hat{\boldsymbol{\pi}}_i) - H_i(\boldsymbol{\pi}_{i0}) + O_p\left[(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau))^2\right]\right. \\
&= -\mathbf{J}_i^{-1}\mathbb{H}_i(\boldsymbol{\pi}_{i0}) - \mathbf{J}_i^{-1}\left(\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i) - \mathbb{H}_i(\boldsymbol{\pi}_{i0})\right) - \mathbf{J}_i^{-1}\left(H_i(\hat{\boldsymbol{\pi}}_i) - H_i(\boldsymbol{\pi}_{i0})\right) \\
&\quad + \mathbf{J}_i^{-1}\left(\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i)\right) + \mathbf{J}_i^{-1}O_p\left[(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau))^2\right] \\
&= -\mathbf{J}_i^{-1}\mathbb{H}_i(\boldsymbol{\pi}_{i0}) - \mathbf{J}_i^{-1}\left[\left(\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i) - \mathbb{H}_i(\boldsymbol{\pi}_{i0})\right) - \left(H_i(\hat{\boldsymbol{\pi}}_i) - H_i(\boldsymbol{\pi}_{i0})\right)\right] \\
&\quad + \mathbf{J}_i^{-1}\left(\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i)\right) + \mathbf{J}_i^{-1}O_p\left[(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau))^2\right].
\end{aligned}$$

For fixed N, the second term in the last expression is $o_p(1)$. In the case of panel data with individual parameters, we need to find the order of $\max_{1\le i\le N}\left[\left(\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i) - \mathbb{H}_i(\boldsymbol{\pi}_{i0})\right) - \left(H_i(\hat{\boldsymbol{\pi}}_i) - H_i(\boldsymbol{\pi}_{i0})\right)\right]$. Lemma 3 in Section A.3 establishes that order. Moreover, by the computational property of quantile regression, $\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i(\tau)) = O_p(T^{-1})$. Therefore, for each $1 \le i \le N$, we have

$$\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau) = -\mathbf{J}_i^{-1}\mathbb{H}_i(\boldsymbol{\pi}_{i0}) + O_p(d_N) + O_p(T^{-1}) + \mathbf{J}_i^{-1}O_p\left[(\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_{i0}(\tau))^2\right], \tag{A.11}$$

where $d_N := T^{-(1-c)}\log(N) \vee T^{-1/2}\delta_N^{1/2}(\log(N))^{1/2}$ and $\delta_N = \sqrt{\log(N)/T}$ when $|\log(\delta_N)| \asymp \log(N)$ as in Theorem 5.1 in Kato, Galvao, and Montes-Rojas (2012). Substituting equation (A.11) in equation (A.10), after basic simplifications, we obtain

$$\begin{aligned}
\sqrt{N}\left(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i \circ \mathbf{J}_i^{-1}\left(\frac{1}{T-p_T}\sum_{t=1+p_T}^{T}\psi_\tau(y_{it} - \mathbf{X}_{it}'\boldsymbol{\pi}_{i0})\mathbf{X}_{it}\right) + \sqrt{N}O(d_N) \\
&\quad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\Xi}_i \circ (\boldsymbol{\pi}_i(\tau) - \boldsymbol{\pi}(\tau)).
\end{aligned} \tag{A.12}$$

The second term is $O_p(N^{1/2}d_N)$. Using the d_N rate implied by Lemma 3 for a sufficiently small c provided that $p_T^3/T \to 0$, we have that $N^{1/2}d_N = N^{1/2}\log(N)^{1/4}T^{-1/4}\log(N)^{1/2}$. Therefore, if $N^{2/3}\log(N)/T \to 0$, the second term is asymptotically negligible.

By standard arguments, as N and T tends to infinity under the conditions on p_T and the relative rate of N and T, $\sqrt{N}\left(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_v)$. $\square$

Proof of Theorem 5. If $\boldsymbol{\vartheta}_i(\tau) = \boldsymbol{\vartheta}(\tau)$ for $1 \le i \le N$, equation (A.10) can be written as

$$\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Xi}_i \circ (\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}(\tau)). \tag{A.13}$$

As in Theorem 4, for each $1 \le i \le N$, we have

$$\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}(\tau) = -\mathbf{J}_i^{-1}\mathbb{H}_i(\boldsymbol{\pi}_0) + O(d_N) + O(T^{-1}) + \mathbf{J}_i^{-1}O_p\left((\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau))^2\right), \tag{A.14}$$

Substituting, after basic simplifications, we obtain

$$\sqrt{NT}\left(\hat{\boldsymbol{\vartheta}}(\tau)-\boldsymbol{\vartheta}(\tau)\right)=\frac{\sqrt{NT}}{N(T-p_T)}\sum_{i=1}^{N}\boldsymbol{\Xi}_i\circ\mathbf{J}_i^{-1}\sum_{t=1+p_T}^{T}\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_0)\mathbf{X}_{it}+\sqrt{TN}O(d_N). \quad \text{(A.15)}$$

The second term is $O_p((NT)^{1/2}d_N)$. Using the where $d_N := T^{-(1-c)}\log(N)\vee T^{-1/2}\delta_N^{1/2}(\log(N))^{1/2}$ and $\delta_N=\sqrt{\log(N)/T}$ when $|\log(\delta_N)|\asymp\log(N)$ implied by Lemma 2, for a sufficiently small c provided that $p_T^3/T\to 0$, we have that $(NT)^{1/2}d_N=N^{1/2}\log(N)^{1/4}T^{-1/4}\log(N)^{1/2}$. Therefore, if $N^2(\log(N))^3/T\to 0$, $\|\hat{\boldsymbol{\vartheta}}(\tau)-\boldsymbol{\vartheta}(\tau)\|=O_p((NT)^{-1/2})$ if the last term is of the same other than the first term in the expression. Therefore, as N and T tends to infinity under the conditions on p_T and the relative rate of N and T, $\sqrt{NT}\left(\hat{\boldsymbol{\vartheta}}(\tau)-\boldsymbol{\vartheta}(\tau)\right)\xrightarrow{d}\mathcal{N}(\mathbf{0},\mathbf{V}_\psi)$.

□

A.3. **Lemmas**

In recent years, there has been considerable progress on establishing the rate of the remainder terms of the Bahadur representation of the quantile regression estimator. The next three lemmas are used in the proofs of Theorems 2, 3, and 4.

Lemma 1. *Under Assumptions 1-11, as N, T and p_T go jointly to infinity with $p_T^3/T\to 0$ and $(\log(N))^2/T\to 0$, $\sup_{i,t,\tau}\|\boldsymbol{\gamma}_i'(\tau)(\bar{\mathbf{G}}(L)-\mathbf{G}(L))\bar{\mathbf{z}}_t\|\xrightarrow{p}0$.*

Proof. The proof is similar to Lemma A.7 in Chudik and Pesaran (2013). As in part 1 of Theorem 2, we define $\sum_{l=0}^{\infty}\bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau)=\boldsymbol{\gamma}_i'(\tau)\mathbf{G}(L)\bar{\mathbf{z}}_t$ and $\sum_{l=0}^{\infty}\bar{\mathbf{z}}_{t-l}'\bar{\boldsymbol{\delta}}_{il}(\tau)=\boldsymbol{\gamma}_i'(\tau)\bar{\mathbf{G}}(L)\bar{\mathbf{z}}_t$, which is obtained with a sample of size N. It follows that,

$$\begin{aligned}
\left\|\boldsymbol{\gamma}_i'(\tau)(\bar{\mathbf{G}}(L)-\mathbf{G}(L))\bar{\mathbf{z}}_t\right\|_\infty &\le \sum_{l=0}^{\infty}\left\|(\bar{\boldsymbol{\delta}}_{il}(\tau)-\boldsymbol{\delta}_{il}(\tau))'\bar{\mathbf{z}}_{t-l}\right\|_\infty=\sum_{l=0}^{p_T}\left\|(\bar{\boldsymbol{\delta}}_{il}(\tau)-\boldsymbol{\delta}_{il}(\tau))'\bar{\mathbf{z}}_{t-l}\right\|_\infty \\
&\quad+\sum_{l=p_T+1}^{\infty}\left\|(\bar{\boldsymbol{\delta}}_{il}(\tau)-\boldsymbol{\delta}_{il}(\tau))'\bar{\mathbf{z}}_{t-l}\right\|_\infty \\
&\le \sum_{l=0}^{p_T}\left\|(\bar{\boldsymbol{\delta}}_{il}(\tau)-\boldsymbol{\delta}_{il}(\tau))'\bar{\mathbf{z}}_{t-l}\right\|_\infty+\sum_{l=p_T+1}^{\infty}\left(\|\bar{\boldsymbol{\delta}}_{il}'(\tau)\bar{\mathbf{z}}_{t-l}\|+\|\boldsymbol{\delta}_{il}'(\tau)\bar{\mathbf{z}}_{t-l}\|\right) \\
&\le \sum_{l=0}^{p_T}\left\|(\bar{\boldsymbol{\delta}}_{il}(\tau)-\boldsymbol{\delta}_{il}(\tau))'\bar{\mathbf{z}}_{t-l}\right\|_\infty+2K\rho^{p_T}\sum_{l=p_T+1}^{\infty}\|\bar{\mathbf{z}}_{t-l}\rho^l\|. \qquad \text{(A.16)}
\end{aligned}$$

The last inequality holds by Assumption 5, and

$$\left\| \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}'_{t-l} \boldsymbol{\delta}_{il}(\tau) \right\| \leq K \rho^{p_T} \sum_{l=1}^{\infty} \left\| \bar{\mathbf{z}}_{-(p_T+l)} \rho^l \right\|,$$

implied by Lemma A.4. (result (A.18)) in Chudik and Pesaran (2015). The first term on the right hand side of equation (A.16),

$$\begin{aligned} \sum_{l=0}^{p_T} \left\| (\bar{\boldsymbol{\delta}}_{il}(\tau) - \boldsymbol{\delta}_{il}(\tau))' \bar{\mathbf{z}}_{t-l} \right\|_\infty &\leq \sum_{l=0}^{p_T} \| \bar{\boldsymbol{\delta}}_{il}(\tau) - \boldsymbol{\delta}_{il}(\tau) \|_\infty \, \| \bar{\mathbf{z}}_{t-l} \|_\infty \leq \sum_{l=0}^{p_T} \| \bar{\boldsymbol{\delta}}_{il}(\tau) - \boldsymbol{\delta}_{il}(\tau) \|_\infty \, \| \bar{\mathbf{z}}_{t-l} \| \\ &= \sum_{l=0}^{p_T} \max_{1 \leq j \leq p_x+1} |\bar{\delta}_{ij,l}(\tau) - \delta_{ij,l}(\tau)| \, \| \bar{\mathbf{z}}_{t-l} \| \\ &\leq K p_T \max_{1 \leq j \leq p_x+1} |\bar{\delta}_{ij,l}(\tau) - \delta_{ij,l}(\tau)|, \end{aligned}$$

which holds by Assumption 10. Note that $\bar{\mathbf{z}}_t$ is bounded because the cross-sectional average of y_{it} and $\mathbf{x}_{it}$ are bounded. On the other hand, the coefficients $\delta_{ij,l}$ decay exponentially in l, which follows by Assumption 5 and Lemma A.1 in Chudik and Pesaran (2013).

For fixed N, the quantile regression estimator is $\sqrt{T}$-consistent. In panel data, by Lemma 8 in Galvao and Wang (2015), we have that $\max_j |\bar{\delta}_{ij,l}(\tau) - \delta_{ij,l}(\tau)| = O_p(\sqrt{\log(N)/T})$ (see also Kato, Galvao and Montes Rojas (2012) and Lemma 3 below). Therefore,

$$\left\| \boldsymbol{\gamma}'_i(\tau)(\bar{\mathbf{G}}(L) - \mathbf{G}(L)) \bar{\mathbf{z}}_t \right\|_\infty \leq K \frac{p_T}{\sqrt{T}} \sqrt{\log(N)} + 2K\rho^{p_T} \sum_{l=p_T+1}^{\infty} \| \bar{\mathbf{z}}_{t-l} \rho^l \|$$

If $p_T^4/T \to c$ for a constant $0 < c < \infty$ and $\log(N)^2/T \to 0$, we have that $\|\boldsymbol{\gamma}'_i(\tau)(\bar{\mathbf{G}}(L) - \mathbf{G}(L))\bar{\mathbf{z}}_t\|_\infty$ converges to zero in L_1 norm which implies convergence in probability to zero. Because T is required to grow faster in panel quantiles than in linear models, we require that $p_T^3/T \to 0$, which is a sufficient condition for convergence in probability to zero and is similar to the condition in Chudik and Pesaran (2013) and Chudik and Pesaran (2015), which requires $p_T^3/T \to c$ where $0 < c < \infty$. □

Lemma 2 (Corollary C.1, Kato et al. (2012)). *Let f be a function on $\mathcal{S}$, a Polish space, and let $\{\xi_t : t \geq 1\}$ be a stationary process taking values in a measurable space $(\mathcal{S}, \mathcal{K})$. Assume that $\mathcal{K}$ is a Borel σ-field. Let $sup_{\xi \in \mathcal{S}} |f(\xi_t)| \leq U$ for some constant U and $E(f(\xi_t)) = 0$. Take $q \in [1, T/2]$. Then,*

$$P\left(|\sum_{t=1}^{T} f(\xi_t)| \geq const \times \left\{ \sqrt{(s \vee 1)T} \sigma_q(f) + sqU \right\} \right) \leq 2e^{-s} + 2r\beta(q),$$

where $r := [T/2q]$, $\beta(\cdot)$ denote mixing coefficients of the $\{\xi_t\}$ process, and

$$\sigma_q(f) := Var(f(\xi_t)) + 2\sum_{j=1}^{q-1}(1-j/q)Cov(f(\xi_1), f(\xi_{1+j})).$$

Lemma 3. *Under regularity conditions 9-12, for any $c \in (0,1)$ and δ_N such that $|\log(\delta_N)| \asymp \log(N)$,*

$$\max_{1\leq i\leq N}\{[\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i) - \mathbb{H}_i(\boldsymbol{\pi}_{i0})] - [H_i(\hat{\boldsymbol{\pi}}_i) - H_i(\boldsymbol{\pi}_{i0})]\} = O_p(T^{-(1-c)}\log(N) \vee T^{-1/2}\delta_N^{1/2}(log(N))^{1/2}),$$

and $\max_{1\leq i\leq N}\|\hat{\boldsymbol{\pi}}_i(\tau) - \boldsymbol{\pi}_i(\tau)\| = O_p(\sqrt{\log(N)/T})$.

Proof. Without loss of generality, we assume that $\boldsymbol{\pi}_{i0} = 0$ for all $1 \leq i \leq N$. We then write $\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i) - \mathbb{H}_i(\boldsymbol{\pi}_{i0})$ as

$$\begin{aligned}
\mathbb{H}_i(\hat{\boldsymbol{\pi}}_i) - \mathbb{H}_i(\mathbf{0}) &= \frac{1}{T-p_T}\sum_{t=1+p_T}^{T}\big(\psi_\tau(y_{it} - \mathbf{X}_{it}'\hat{\boldsymbol{\pi}}_i)\mathbf{X}_{it} - \psi_\tau(y_{it})\mathbf{X}_{it}\big) \\
&= -\frac{1}{T-p_T}\sum_{t=1+p_T}^{T}\big(I(y_{it} \leq \mathbf{X}_{it}'\hat{\boldsymbol{\pi}}_i) - I(y_{it} \leq 0)\big)\mathbf{X}_{it} \\
&= -\frac{1}{T-p_T}\sum_{t=1+p_T}^{T}\big(I(u_{it} \leq \mathbf{X}_{it}'\hat{\boldsymbol{\pi}}_i) - I(u_{it} \leq 0)\big)\mathbf{X}_{it} = \frac{1}{T-p_T}\sum_{t=1+p_T}^{T}\mathfrak{G}_{\hat{\boldsymbol{\pi}}_i}(\mathbf{X}_{it}^*),
\end{aligned}$$

where $\mathfrak{G}_{\boldsymbol{\pi}_i}(\mathbf{X}_{it}^*) = \mathfrak{G}_{\boldsymbol{\pi}_i}((u_{it}, \mathbf{X}_{it}) = (I(u_{it} \leq \mathbf{X}_{it}'\boldsymbol{\pi}_i) - I(u_{it} \leq 0))\mathbf{X}_{it}$. The third equality follows because $\boldsymbol{\pi}_{i0} = \mathbf{0}$, which implies that $\boldsymbol{\gamma}_{i0} = \mathbf{0}$ for $1 \leq i \leq N$, and consequently, $y_{it} = \mathbf{X}_{it}'\boldsymbol{\pi}_{i0} + e_{it} = u_{it} + \boldsymbol{\gamma}_{i0}'(\mathbf{f}_t - \mathbf{G}(L)\bar{\mathbf{z}}_t) + \sum_{l=p_T+1}^{\infty}\bar{\mathbf{z}}_t'\boldsymbol{\delta}_{il} = u_{it}$.

Pick a $c \in (0,1)$ and let $d_N := T^{-1+c}|\log(\delta_N)| \vee T^{-1/2}\delta_N^{1/2}|\log(\delta_N)|^{1/2}$. We need to show that

$$\begin{aligned}
\max_{1\leq i\leq N}\left\{\sum_{t=1}^{T}\left(\mathfrak{G}_{\hat{\boldsymbol{\pi}}_i}(\mathbf{X}_{it}^*) - E\left(\mathfrak{G}_{\hat{\boldsymbol{\pi}}_i}(\mathbf{X}_{it}^*)\right)\right)\right\} &= O_p((T-p_T)d_N) \\
&= O_p((T-p_T)(T^{-1+c}\log(N) \vee T^{-1/2}\delta_N^{1/2}\log(N)^{1/2})).
\end{aligned}$$

Details of the proof are omitted as the proof is an application of Proposition C.1. in Kato et al. (2012) as shown in Galvao and Wang (2015). The result follows by verifying that Assumptions 9, 11, and 12 are similar to the Assumptions A2-A5 and B1-B3 in Lemmas 7 and 8 in Galvao and Wang (2015). To avoid repetition, we point out the modification. Fixing $1 \leq i \leq N$, we apply Proposition C.1. in Kato et al. (2012) to a class of functions $\tilde{\mathfrak{G}}_{i\delta_N} = \{\mathfrak{G}_{\boldsymbol{\pi}} - E\mathfrak{G}_{\boldsymbol{\pi}} : \|\boldsymbol{\pi}\| \leq \delta_N\}$. Conditions (i) and (iii) in Proposition C.1. are satisfied as in Galvao and Wang (2015) but Condition (ii) needs some attention since the condition on uniformly bounded regressors is different in our paper. Under Assumption 10, the class of functions $\tilde{\mathfrak{G}}_{i\delta_N}$ is bounded, and therefore, Condition (ii) is satisfied.

Thus, the conclusion of the first result is obtained following Lemma 7 in Galvao and Wang (2015). The second result follows directly by Lemma 8 in Galvao and Wang (2015). $\square$

Online Supplement:

Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Quantile Regression Models

Matthew Harding*, Carlos Lamarche†, and M. Hashem Pesaran‡

August 14, 2018

In this Supplement, we first offer a derivation of the variance of $\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i)$ when we allow for dependence across time. Section S.2 presents simulation results for a one factor model, which complements the evidence on a two factor model presented in Section 3. In Section S.3, we extend the simulation results presented in Section S.2 by offering results on the infeasible QMG estimator which uses the latent factor, f_t. The comparison of the infeasible quantile estimator with the QMG estimator illustrates the effect of cross-section augmentation in dealing with the estimation of f_t. This section also presents additional results on the power of the QMG estimator. Lastly, in Section S.4, we investigate the sensitivity of the results to the choice of p_T in the empirical application, by offering additional empirical results.

S.1. On the Derivation of the Asymptotic Covariance Matrix

Let $\bar{\xi}_i(\tau) := T^{-1/2}\sum_{t=1}^T \psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i)$. By definition, we have that

$$\mathrm{Var}(\bar{\xi}_i(\tau)) = \frac{1}{T}\sum_{t=1}^T \mathrm{Var}[\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i)] + 2\sum_{t\neq t'}^T \mathrm{Cov}[\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i), \psi_\tau(Y_{it'}-\mathbf{X}_{it'}'\boldsymbol{\pi}_i))].$$

Note that,

$$\mathrm{Cov}[\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i), \psi_\tau(Y_{it'}-\mathbf{X}_{it'}'\boldsymbol{\pi}_i)] \;=\; \mathrm{Cov}(\tau - I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i), \tau - I(Y_{it'}<\mathbf{X}_{it'}'\boldsymbol{\pi}_i))$$

*Department of Economics, University of California Irvine, SSPB 3207, Irvine, CA 9269; Email: harding1@uci.edu

†Department of Economics, University of Kentucky, 223G Gatton College of Business and Economics, Lexington, KY 40506-0034; Phone: (859) 257 3371; Email: clamarche@uky.edu

‡Department of Economics, University of Southern California, and Trinity College, Cambridge; Kaprielian Hall 300, Los Angeles, CA 90089-0253; Email: pesaran@usc.edu

$$\begin{aligned}\mathrm{Cov}[\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i),\psi_\tau(Y_{it'}-\mathbf{X}_{it'}'\boldsymbol{\pi}_i)] &= E((\tau-I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i))(\tau-I(Y_{it'}<\mathbf{X}_{it'}'\boldsymbol{\pi}_i)))\\ &\quad -E(\tau-I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i))E(\tau-I(Y_{it'}<\mathbf{X}_{it'}'\boldsymbol{\pi}_i))\\ &= E((\tau-I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i))(\tau-I(Y_{it'}<\mathbf{X}_{it'}'\boldsymbol{\pi}_i))),\end{aligned}$$

because $\tau = G(\mathbf{X}_{it}'\boldsymbol{\pi}_i) = G(\mathbf{X}_{it'}'\boldsymbol{\pi}_i)$ under Assumption 11. It follows that,

$$\begin{aligned}\mathrm{Cov}[\psi_\tau(y_{it}-\mathbf{X}_{it}'\boldsymbol{\pi}_i),\psi_\tau(Y_{it'}-\mathbf{X}_{it'}'\boldsymbol{\pi}_i)] &= \tau^2-\tau E(I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i))-\tau E(I(Y_{it'}<\mathbf{X}_{it'}'\boldsymbol{\pi}_i))\\ &\quad +E(I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i)I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i))\\ &= E(I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i)I(y_{it}<\mathbf{X}_{it}'\boldsymbol{\pi}_i))-\tau^2.\end{aligned}$$

Thus, we have that

$$\begin{aligned}\sigma_\psi^2(q) := \mathrm{Var}(\bar{\xi}_i(\tau)) &= \tau(1-\tau)+2\sum_{j=1}^{q-1}\left(1-\frac{j}{q}\right)\mathrm{Cov}(\psi_\tau(Y_{i1}-\mathbf{X}_{i1}'\boldsymbol{\pi}_i),\psi_\tau(Y_{i1+j}-\mathbf{X}_{i1+j}'\boldsymbol{\pi}_i))\\ &= \tau(1-\tau)+2\sum_{j=1}^{q-1}\left(1-\frac{j}{q}\right)\left[E(I(Y_{i1}<\mathbf{X}_{i1}'\boldsymbol{\pi}_i)I(Y_{i1+j}<\mathbf{X}_{i1+j}'\boldsymbol{\pi}_i))-\tau^2\right]\\ &= \tau(1-\tau)+2\sum_{j=1}^{q-1}\left(1-\frac{j}{q}\right)\left[E(I(Y_{i1}<\mathbf{X}_{i1}'\boldsymbol{\pi}_i,Y_{i1+j}<\mathbf{X}_{i1+j}'\boldsymbol{\pi}_i))-\tau^2\right].\end{aligned}$$

This parameter is estimated in Section 2.3.

S.2. **Simulation Evidence: One Factor Models**

This section reports results of several simulation exercises designed to evaluate the small sample performance of the proposed estimator. Observations on y_{it} for $i = 1, 2, \ldots, N$ and $t = -S+1, -S+2, .., 0, 1, ..., T$ are generated according to the following ARX(1) model with one factor:

$$y_{it} = \alpha_i + \lambda_i y_{i,t-1} + \beta_0 + \beta_i x_{it} + \gamma_i f_t + (1+\delta x_{it})u_{it}, \tag{S.2.1}$$

where the error term u_{it} is distributed as $F(0,\sigma_i^2)$, σ_i^2 generated as $0.5(1+\chi_i^2)$ and χ_i^2 denotes an i.i.d. random variable distributed as χ^2 distribution with 1 degree of freedom. Depending on the value of δ, we have two conditional quantile functions. When $\delta = 0$, we have

$$Q_{Y_{it}}(\tau|\alpha_i, y_{it-1}, x_{it}, f_t) = \alpha_i + \lambda_i y_{i,t-1} + \beta_0(\tau) + \beta_i x_{it} + \gamma_i f_t, \tag{S.2.2}$$

with $\beta_0(\tau) = \beta_0 + F_u^{-1}(\tau)$. On the hand, when $\delta \neq 0$, the conditional quantile function of (S.2.1) becomes,

$$Q_{Y_{it}}(\tau|\alpha_i, y_{it-1}, x_{it}, f_t) = \alpha_i + \lambda_i y_{i,t-1} + \beta_0(\tau) + \beta_i(\tau)x_{it} + \gamma_i f_t, \tag{S.2.3}$$

with $\beta_0(\tau) = \beta_0 + F_u^{-1}(\tau)$ and $\beta_i(\tau) = \beta_i + \delta F_u^{-1}(\tau)$. Models (S.2.2) and (S.2.3) are typically refereed to as location shift and location-scale shift models in the literature (see, e.g., Koenker (2005)). Quantile regression models are estimated with an overall intercept β_0 which is assumed to be zero in the simulations. Note that for S sufficiently large, we have that

$$y_{i0} \approx \frac{1}{1-\lambda}\alpha_i + \beta \sum_{j=0}^{S-1} \lambda^j x_{i,-j} + \sum_{j=0}^{S-1} \lambda^j \xi_{i,-j}, \tag{S.2.4}$$

where $\xi_{it} = \gamma_i f_t + (1 + \delta x_{it})u_{it}$, $\lambda = E(\lambda_i)$ and $\beta = E(\beta_i)$. In all the variants of the model considered in the simulations, we set $S = 200$. The regressor, x_{it}, is generated as

$$\begin{aligned} x_{it} &= \mu_i + \Gamma_i f_t + v_{it}, & \text{(S.2.5)} \\ v_{it} &= \rho_x z_{i,t-1} + \sqrt{1-\rho_x^2}\,\varepsilon_{it}, & \text{(S.2.6)} \\ f_t &= \rho_f f_{t-1} + \sqrt{1-\rho_f^2}\,\varepsilon_{ft}, & \text{(S.2.7)} \end{aligned}$$

where the i.i.d. variables $\mu_i \sim \mathcal{N}(0.5, 1)$, $\varepsilon_{it} \sim \mathcal{N}(0, 1)$, and $\varepsilon_{ft} \sim \mathcal{N}(0, 1)$. We consider the case of relatively persistent regressors by setting $\rho_x = 0.8$ and $\rho_f = 0.9$. Moreover, without loss of generality we use $x_{i,-S} = 0$ and $f_{-S} = 0$.

The factor loadings in equation (S.2.1), γ_i, and in equation (S.2.5), Γ_i, are generated as $\gamma_i \sim iid\mathcal{N}(0.5, 1)$ and $\Gamma_i \sim iid\mathcal{N}(0.5, 1)$. Finally, the fixed effects, α_i, are allowed to be correlated with the errors by generating them as $\alpha_i = \bar{x}_i + \gamma_i \bar{f} + \bar{u}_i + a_i$, where the individual specific averages are defined as $\bar{x}_i = T^{-1}\sum_{t=1}^T x_{it}$, $\bar{f} = T^{-1}\sum_{t=1}^T f_t$, $\bar{u}_i = T^{-1}\sum_{t=1}^T u_{it}$. The error term a_i in the equation for α_i is assumed to be distributed as $\mathcal{N}(0, 1)$.

Initially, we set $\lambda_i = \lambda$ for $i = 1, 2, \ldots, N$ and consider three values of $\lambda = \{0.25, 0.50, 0.75\}$. Later in Figure S.1, we investigate the performance of the estimator with heterogeneous λ_i's. Moreover, in addition to the experiments presented in this section, we also considered static panel data experiments (i.e., $\lambda = 0$) and compare the performance of the proposed approach with existing panel quantile regression approaches. For relatively large T, the performance of the proposed estimator was similar in both the static panel data model and dynamic panel data model. Thus, we present results for the dynamic model to save space.

In the simulations, we assume that the error term u_{it} in equation (S.2.1) is an i.i.d. random variable distributed as Standard Normal, t-student with 4 degrees of freedom (t_4), and χ^2 with 3 degrees of freedom (χ^2_3). We consider the following four variations of the model:

Design 1: (Location shift model with homogeneous slopes). We consider $\beta = 1$ in a location shift model with $\delta = 0$.
Design 2: (Location shift model with heterogeneous slopes). We consider heterogeneous slope parameters $\beta_i = \beta + \nu_i$ in a location shift model, where $\delta = 0$, $\beta = 1$ and $\nu_i \sim \mathcal{U}(-0.25, 0.25)$. The parameter $\beta_i(\tau) = \beta_i$ for all i and τ.
Design 3: (Location-scale shift model with homogeneous slopes). We consider homogenous slope parameters $\beta = 1$ in a location-scale shift model with $\delta = 0.1$. In this case, the slope parameter $\beta(\tau) = \beta + 0.1F_u^{-1}(\tau)$.
Design 4: (Location-scale shift model with heterogeneous slopes). We consider heterogeneous slope parameters as in Design 2, $\beta_i = \beta + \nu_i$, in a location-scale shift model with $\delta = 0.1$. We assume $\beta = 1$ and $\nu_i \sim \mathcal{U}(-0.25, 0.25)$ which implies that $\beta_i(\tau) = 1 + \nu_i + 0.1F_u^{-1}(\tau)$. In this case, $E(\beta_i(\tau)) = \beta(\tau) = 1 + 0.1F_u^{-1}(\tau)$.

Tables S.1 to Table S.2 present the bias and root mean square error (RMSE) for the slope parameter β in the location shift model with $\lambda = 0.5$. The finite sample performance for the slope parameter when the model (S.2.1) include a different value for λ is considered in Table S.9. While Table S.1 presents results for Designs 1 and 2, Table S.2 presents results for Designs 3 and 4. The tables show results for quantile regression estimators at two quantiles, $\tau \in \{0.25, 0.50\}$, based on sample sizes of $N \in \{100, 200\}$ and $T \in \{50, 100, 200\}$.

We compare the performance of the following quantile regression estimators: (i) the existing instrumental variable quantile regression estimator for a dynamic panel data model developed by Galvao (2011), labeled DQR, and (ii) the quantile mean group (QMG) estimator for a model with interactive effects. The DQR estimator uses y_{it-2} as an instrument for y_{it-1}. It should be noted that Galvao's model does not include the term $\lambda_i f_t$, which can generate biases that cannot be eliminated by the use of instrumental variables. The proposed quantile mean group estimator, QMG, is obtained as the simple cross sectional average of $\hat{\beta}_i(\tau)$ using $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{y}_{t-1}, \bar{x}_t)'$.

The tables do not provide the finite sample performance of other existing quantile estimators. The classical quantile regression estimator is biased because the individual specific effects α_i and the factor f_t are correlated with the regressor x_{it}. Also the fixed effects estimator, the recently proposed minimum distance quantile regression estimator (Galvao and Wang (2015)), and the penalized quantile regression estimator are biased when the model includes a lagged dependent variable. Therefore, we restrict our comparison to DQR, which is the only estimator in the literature proposed for dynamic panel quantile regression models.

S.2.1. **Bias and Root Mean Square Error**

In Table S.1, it might not be surprising to find out that the DQR method is biased and that its bias tends to be slightly larger in the case of heterogeneous slopes. The bias of this estimator for the slope β tends to increase as T increases and it does not seem to change at the 0.5 and 0.25 quantiles. On the other hand, the performance of the QMG estimator is excellent, with biases in general lower than 5% for $T = 50$ and biases decreasing rapidly to 1% when $T = 200$. In all the variations of the model considered in the table, the quantile estimator QMG performs better than DQR in terms of RMSE too.

Table S.2 presents results for the location-scale shift model where $\beta(\tau)$ changes by quantile. For instance, $\beta(0.5) = 1$ and $\beta(0.25) = 0.93$ in the case where the error term $u_{it} \sim \mathcal{N}(0, 1)$, and $\beta(0.5) = 1.24$ and $\beta(0.25) = 1.12$ when $u_{it} \sim \chi^2_3$. We continue to see that the DQR estimator is biased and has poor RMSE properties. The performance of the QMG estimator in these variations of the model is similar to Table S.1, with low biases and small RMSE. For values of T larger than 50, the bias of the proposed estimator is always negative and it ranges between 0.6% and 3%.

We expanded the simulation evidence for the slope parameter β to consider different values of λ. Table S.9 presents results for $\lambda \in \{0.25, 0.75\}$ considering the same designs as in Tables S.1 and S.2 and $N = 100$ and $T = 200$. We see that the QMG estimator continues to perform better than the DQR estimator. We also find that the performance of the QMG estimator is invariant to the choice of λ, at least in the simulations considered thus far. We do investigate the performance of the QMG estimator when $\lambda_i \in [0.025, 0.925]$ below.

We now turn our attention to the estimator for $\lambda(\tau)$ and $\theta(\tau) = \beta(\tau)/(1 - \lambda(\tau))$. The estimator for $\theta(\tau)$ is defined as $\hat{\beta}(\tau)/(1-\hat{\lambda}(\tau))$ and it is obtained by plugging in the quantile estimates corresponding to $\lambda(\tau)$ and $\beta(\tau)$. We employ this method for the DQR and QMG estimators.

Tables S.3, S.4, S.5 and S.6 show the bias and RMSE of the DQR and QMG estimators for the parameters of interest. These four tables show results for the four different designs we consider in this section. Each table presents, in columns, the performance of the estimators at $\tau \in \{0.25, 0.50\}$ and in rows the different samples sizes and distributions for the error term. The upper block present results when u_{it} is distributed as $\mathcal{N}(0,1)$, the middle panel shows results when $u_{it} \sim t_4$ and the lower block presents results when $u_{it} \sim \chi^2_3$. While $\lambda = 0.5$ does not change in these tables, the parameter of interest θ does change in the tables. For instance, $\theta(0.5) = 2 = \theta(0.25)$ in the Gaussian case in Table S.3, $\theta(0.5) = 2.48$ and $\theta(0.25) = 2.24$ when $u_{it} \sim \chi^2_3$ in Table S.5.

As before, the results indicate that the bias of the DQR estimator can be large, in particular for the long run coefficient θ. The bias of the QMG estimator is small in the tables and it tends to zero as T increases, as expected. For $T = 50$, however, we see that the bias of the DQR estimator is smaller than the bias of the QMG estimator in the case of χ^2_3 for the parameter λ (i.e., columns (1) and (2)). We also find that the QMG estimator has smaller variance than the DQR estimator which might be expected since the QMG estimator does not employ instrumental variables. Even in the few cases where the bias of the DQR estimator is smaller than the bias of the QMG estimator, the QMG estimator offers the best performance in terms of RMSE.

A comparison between the results for the long-run effect in the location shift model reveals that estimating heterogeneous effects is more demanding than estimating homogeneous effects, as expected. However, the QMG estimator offers nearly zero biases for large N and large T. The DQR estimator is biased and its performance is not satisfactory in terms of both bias and RMSE. The location-shift case, presented in Tables S.5 and S.6, reveals similar findings. Overall, when $\lambda = 0.5$, the QMG estimator offers the best performance in terms of bias and RMSE in the class of estimators for a dynamic quantile panel data model.

Figure S.1 offers a clear summary of the small sample performance of the QMG estimator as λ increases. The figure shows the bias and RMSE of the QMG estimator at $\tau \in \{0.25, 0.50\}$ for λ, β and θ in terms of λ. We considered Design 1 with $N = 100$ and $T = 200$. Recall that when λ increases, θ increases too. For instance, while $\lambda = 0$ gives $\theta = \beta = 1$, $\lambda = 0.9$ gives $\theta = 10$ in our simulation experiment. Consistent with our previous evidence, we see that the performance of QMG estimator does not depend on λ when the interest is in estimating β. The bias tends to slightly increase but it is never larger than 1% for large values of λ. We also find that the RMSE of the estimator of β does not change with λ. On the other hand, we find that the absolute value of the bias of the QMG estimator for θ increases exponentially when $\lambda \to 1$. The figure shows that the bias, in absolute value, is negligible for $\lambda < 0.75$, and it increases rapidly when $\lambda > 0.8$. Note however that the bias in relative terms is always less than 10%. We also find that the RMSE increases with λ and that the RMSE of the QMG estimator at $\tau = 0.25$ is larger than the QMG estimator at $\tau = 0.50$, as expected.

Figure S.1 also shows the bias and RMSE of the QMG estimator when $\lambda_i = \lambda + \omega_i$, where $\omega_i \sim \mathcal{U}[-0.025, 0.025]$ and λ takes values in the interval $\lambda \in [0.05, 0.90]$. The parametrization guarantees that θ exists for all values of λ_i for $i = 1, \ldots, N$. We generate data using Design 1 with $N = 100$ and $T = 200$. Consistent with our expectations, the bias and RMSE of the estimator tends to be similar to the case of homogeneous λ's, although the performance deteriorates for large values of $\lambda = E(\lambda_i)$. We see an increase in the variance of the estimator, but the bias for θ remains, in absolute value, small for $E(\lambda_i) < 0.65$. As it can be seen in Figure S.1, the parameter $(E(\lambda_i), \beta)$ can be estimated with small bias and excellent RMSE performance in the case of heterogeneous λ_i's.

S.2.2. **Inference**

We now turn our attention to the standard error of the QMG estimator for $\lambda(\tau)$ and $\beta(\tau)$. Table S.7 reports the average estimated standard error obtained by the procedure outlined in Section 2.2. We select $q = 3$ to minimize potential biases in the estimation of the standard errors. While the upper panels of Table S.7 show the standard error of the QMG estimator in Designs 1 and 2, the lower panels show the standard error in Designs 3 and 4. We also report the standard deviation of the estimator based on 400 Monte Carlo repetitions. Because T relative to N is important for inference, we included results with $N = 100$ and $T \in \{100, 200, 400\}$.

The results show that the estimated standard errors approximate very closely to the standard deviation of the estimator when T is larger than N. This result is expected by the rates of convergence needed to establish the consistency of the QMG estimator. The approximation is excellent in the case of the Normal and t_4 distributions. The evidence when $u_{it} \sim \chi^2_3$ suggests that a larger T relative to N is needed for the standard error to be well approximated.

Table S.8 provides empirical coverage probabilities for a nominal 95% confidence interval. The probabilities are calculated based on asymptotic Gaussian confidence intervals. We see different finite sample performances of the estimator for λ and β. If we examine the results across the different distributions, the QMG estimator in some cases does not perform well for λ when $T/N < 4$. On the other hand, the coverage probabilities for β approximate closely 0.95 with the exceptions when $T = N = 100$. Lastly, we investigate the performance of the

QMG estimator in terms of power. The results are shown in the lower panel of Table S.8. We compute the power for the estimation of λ with the alternative hypothesis $H_a : \lambda = 0.55$ and β with the alternative hypothesis $H_a : \beta = 1.1$. The condition on the rate of convergence plays an important role in ensuring that the estimator has good power. In particular, the power is high for values of $T > 100$, although how quickly approaches 1 depends on the distribution of the error term and the number of cross-sectional units, N.

As we can see in Figure S.2, the power function of the test constructed based on the QMG estimator at the 0.25 quantile tends to be symmetric and have the expected shape. As an illustration, Figure S.2 reports results based on Design 1 for the Gaussian and χ^2_3 case when $N = 100$ and $T = 400$. Consistent with the theory, a larger T for a given number of cross-sectional units leads to a better approximation of the function. The evidence shows that for this sample size and quantile, the QMG estimator seem to perform reasonably well for different distributions and parameters.

S.3. **Additional Simulation Evidence**

This section offers additional Monte Carlo evidence on the finite sample performance of the proposed estimator for different values of λ. It then compares the performance of the feasible QMG estimator with an unfeasible version of it that uses unknown factors $\mathbf{f}_t$.

S.3.1. **Autoregressive Models**

Tables S.9 presents the bias and root mean square error (RMSE) for the slope parameter β in the location shift model with $\lambda \in \{0.25, 0.75\}$, which are different to the value $\lambda = 0.5$ used in the first tables of Section 3. The table presents results for Designs 1-4, showing results for quantile regression estimators at two quantiles, $\tau \in \{0.25, 0.50\}$. The table compares the performance of the quantile regression estimator for a dynamic panel data model (Galvao 2011) denoted by DQR and the quantile mean group (QMG) estimator for a model with interactive effects. The DQR estimator uses y_{it-2} as an instrument for y_{it-1}. The proposed quantile mean group estimator, QMG, is obtained as the cross sectional average of $\hat{\beta}_i(\tau)$ using $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{y}_{t-1}, \bar{x}_t)'$. The sample size is based on $N = 100$ and $T = 200$. Table S.9 shows that the QMG method for β performs extremely well with negligible biases and low RMSE for all values of λ.

Tables S.10 and S.11 report results for λ and θ. While Table S.10 presents simulation results for the case of $\lambda = 0.25$, Table S.11 presents simulation results for the case of $\lambda = 0.75$. In terms of relative performance between DQR and QMG estimators, the tables do not offer new insights. The evidence continues to suggest that the QMG estimator performs better in small samples than the DQR estimator with considerable gains in terms of bias and RMSE. We find however that the absolute bias of the QMG estimator tends to increase as λ increases. The results however are not presented in terms of percentage bias since θ increase as λ increases.

S.3.2. **Estimation of Models with Known $\mathbf{f}_t$**

This section compares the results of the estimator QMG defined in Section 2.2 with the results obtained by employing an unfeasible version of the estimator. The infeasible estimator is defined as:

$$\tilde{\boldsymbol{\pi}}_i(\tau) = \arg \min_{\boldsymbol{\pi}_i \in \boldsymbol{\Pi}_i} \sum_{t=1}^{T} \rho_\tau \left(y_{it} - \lambda_i y_{it-1} - \boldsymbol{\beta}_i' \mathbf{x}_{it} - \boldsymbol{\gamma}_i' \mathbf{f}_t \right)$$

Therefore, $\tilde{\boldsymbol{\pi}}_i(\tau)$ is an estimator based on quantile regression with latent factors, $\mathbf{f}_t$. Moreover, we define $\tilde{\boldsymbol{\vartheta}}(\tau) = \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{\vartheta}}_i(\tau) = \frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{\Xi}_i \circ \tilde{\boldsymbol{\pi}}_i(\tau)\right)$, where $\circ$ denotes Hadamard product and $\boldsymbol{\Xi}_i = (\boldsymbol{\iota}_i', \mathbf{0}_i')'$ with ι_i denoting a vector of ones. In what follows, we denote this estimator as IQMG.

The data is generated as in Section S.2 (see equations (S.2.1)-(S.2.7)). Tables S.12 to Table S.13 present the bias and root mean square error (RMSE) of the QMG and IQMG estimators for the slope parameter β in the location shift model with $\lambda = 0.5$. Table S.12 presents results for Designs 1 and 2 and Table S.13 presents results for Designs 3 and 4. The tables show results for quantile regression estimators at two quantiles, $\tau \in \{0.25, 0.50\}$, based on sample sizes of $N \in \{100, 200\}$ and $T \in \{50, 100, 200\}$. As to be expected, the IQMG estimator yields smaller bias and smaller RMSE than its feasible counterpart QMG. We also observe that these differences tend to disappear as long as both N and T increase.

In the next four tables, Tables S.14 to Table S.17, we present results the bias and root mean square error (RMSE) of the QMG and IQMG estimators for the parameters λ and θ. The results continues to indicate that the infeasible version improves the performance of the feasible version, although again as in the case of the slope parameter β, the finite sample performance of the QMG estimator approximates very closely to the performance of the IQMG estimator when $T > 100$.

S.3.3. **Power**

This section reports additional simulation results for the power of the QMG estimator. In light of the theoretical results, we limit our investigation to document the shape of the power function as the time series dimension of the panel, T, increases. We generate data using Design 1 considering $N = 100$ and $T \in \{100, 200, 400\}$ for the case of Gaussian and χ^2_3 error term. The evidence is presented in Figures S.3 and S.4.

As shown in Figures S.3 and S.4, the power function of the test constructed based on the QMG is not symmetric when $N \approx T$, but it tends to have the expected shape as long as T increases. This is true for different distributions (i.e., Gaussian and χ^2_3) and different quantiles (i.e., $\tau \in \{0.25, 0.50\}$). The evidence shows that the QMG estimator performs well for different distributions and parameters, as long as the number of time series observations is significantly larger than the number of cross-sectional units.

S.4. **Time-of-Use Pricing, Smart Technology and Energy Savings: A Sensitivity Analysis**

In this section, we re-estimate the conditional quantile function for electricity consumption (equation (4.2) in Section 4.2) to evaluate the sensitivity of results to a change in the number of cross-section averages used to approximate the unknown factors, $\mathbf{f}_t$. As in Section 4, we estimate the model using the QMG procedure for each quantile τ and group g separately. However, we deviate from the previous analysis by using cross-section averages of the logarithm of electricity usage at t and $t-1$, $(\bar{y}_t, \bar{y}_{t-1})$, as well as cross-sectional averages of temperature and drew-point, $(\bar{x}_{1,t}, \bar{x}_{2,t})$. The evidence is reported in Tables S.18, S.19, S.20, and Figures S.5 and S.6.

As to be expected, the evidence shows consistently smaller (in absolute value) treatment effects estimates when we compare the results reported in Section 4.3 with the results shown in Table S.18. The CCEMG (see Chudik and Pesaran (2015)) estimate for the control group is reduced by 1% point in Table 4.2 compared to Table S.18, while the MG estimate for the PCT group is increased from -7% to -6.5%. The evidence across quantiles is similar, although we observe the most significant changes in the results at the 0.1 and 0.9 quantiles of the conditional distribution of electricity usage. Lastly, we obtain, in general, similar empirical results when we estimate the responsiveness to TOU pricing across demographics (see Tables S.19 and S.20) and we perform a counterfactual exercise on how policies can increase savings (see Figure S.6). The exceptions are, again, the 0.1 and 0.9 quantiles of the PCT group, where we observe interesting changes. Overall, the results appear to be robust to the choice of the number of lagged cross-section averages used to proxy $\mathbf{f}_t$. Nevertheless, we would like to emphasize the importance of setting p_T to be relatively large in applications to avoid inconsistent results.

References

CHUDIK, A., AND M. H. PESARAN (2015): "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors," *Journal of Econometrics*, 188(2), 393 – 420.

GALVAO, A. F. (2011): "Quantile regression for dynamic panel data with fixed effects," *Journal of Econometrics*, 164(1), 142 – 157.

GALVAO, A. F., AND L. WANG (2015): "Efficient minimum distance estimator for quantile regression fixed effects panel data," *Journal of Multivariate Analysis*, 133(C), 1–26.

KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press.

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Design 1: Location shift with homogeneous slopes											
100	50	Bias	0.126	-0.052	0.125	-0.053	0.111	-0.050	0.111	-0.056	0.080	-0.066	0.080	-0.035
100	50	RMSE	0.145	0.054	0.143	0.055	0.127	0.051	0.128	0.057	0.102	0.068	0.098	0.037
100	100	Bias	0.163	-0.024	0.161	-0.024	0.150	-0.022	0.149	-0.025	0.113	-0.029	0.106	-0.014
100	100	RMSE	0.173	0.025	0.171	0.025	0.161	0.024	0.161	0.027	0.125	0.030	0.117	0.015
100	200	Bias	0.177	-0.008	0.177	-0.008	0.169	-0.007	0.169	-0.009	0.134	-0.013	0.123	-0.004
100	200	RMSE	0.182	0.010	0.182	0.010	0.174	0.009	0.174	0.011	0.141	0.014	0.129	0.006
200	50	Bias	0.127	-0.057	0.125	-0.056	0.120	-0.053	0.119	-0.060	0.082	-0.068	0.081	-0.037
200	50	RMSE	0.144	0.058	0.142	0.057	0.140	0.054	0.139	0.061	0.099	0.069	0.096	0.037
200	100	Bias	0.160	-0.026	0.159	-0.025	0.139	-0.023	0.139	-0.026	0.120	-0.032	0.111	-0.015
200	100	RMSE	0.171	0.026	0.169	0.026	0.148	0.024	0.148	0.027	0.130	0.032	0.120	0.016
200	200	Bias	0.187	-0.011	0.186	-0.011	0.166	-0.010	0.168	-0.011	0.137	-0.015	0.126	-0.006
200	200	RMSE	0.192	0.012	0.191	0.011	0.171	0.011	0.173	0.012	0.143	0.015	0.131	0.007
N	T		Design 2: Location shift with heterogeneous slopes											
100	50	Bias	0.131	-0.053	0.129	-0.053	0.112	-0.049	0.113	-0.055	0.083	-0.066	0.083	-0.035
100	50	RMSE	0.151	0.054	0.148	0.055	0.129	0.051	0.130	0.057	0.106	0.068	0.103	0.036
100	100	Bias	0.173	-0.023	0.170	-0.023	0.152	-0.022	0.151	-0.025	0.115	-0.029	0.109	-0.014
100	100	RMSE	0.181	0.025	0.180	0.024	0.163	0.023	0.163	0.026	0.127	0.030	0.119	0.015
100	200	Bias	0.185	-0.009	0.184	-0.009	0.174	-0.007	0.175	-0.009	0.138	-0.012	0.128	-0.004
100	200	RMSE	0.190	0.010	0.189	0.011	0.180	0.009	0.180	0.011	0.144	0.014	0.133	0.006
200	50	Bias	0.130	-0.057	0.128	-0.056	0.124	-0.054	0.122	-0.061	0.094	-0.068	0.092	-0.036
200	50	RMSE	0.147	0.058	0.145	0.057	0.143	0.054	0.142	0.061	0.110	0.069	0.106	0.037
200	100	Bias	0.160	-0.025	0.159	-0.025	0.144	-0.024	0.144	-0.027	0.120	-0.032	0.113	-0.015
200	100	RMSE	0.171	0.026	0.169	0.026	0.152	0.024	0.152	0.027	0.129	0.032	0.122	0.016
200	200	Bias	0.186	-0.011	0.185	-0.010	0.172	-0.010	0.174	-0.011	0.139	-0.015	0.128	-0.007
200	200	RMSE	0.190	0.012	0.189	0.011	0.178	0.011	0.179	0.012	0.144	0.015	0.132	0.007

TABLE S.1. *Bias and root mean square error (RMSE) of quantile regression estimators for β in Designs 1 and 2. In all the variations of the model, $\lambda = 0.5$. DQR denotes the instrumental variable quantile regression estimator for dynamic quantile regression and QMG denotes the mean quantile group estimator.*

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Design 3: Location-scale shift with homogeneous slopes											
100	50	Bias	0.128	-0.059	0.125	-0.059	0.111	-0.056	0.110	-0.063	0.060	-0.062	0.060	-0.033
100	50	RMSE	0.146	0.061	0.143	0.061	0.131	0.058	0.130	0.065	0.080	0.064	0.078	0.035
100	100	Bias	0.170	-0.026	0.167	-0.026	0.143	-0.024	0.140	-0.027	0.093	-0.030	0.088	-0.013
100	100	RMSE	0.180	0.027	0.178	0.027	0.151	0.025	0.149	0.029	0.103	0.031	0.097	0.015
100	200	Bias	0.185	-0.010	0.183	-0.010	0.165	-0.009	0.164	-0.011	0.112	-0.013	0.102	-0.005
100	200	RMSE	0.190	0.012	0.188	0.012	0.171	0.010	0.171	0.013	0.117	0.014	0.107	0.007
200	50	Bias	0.120	-0.063	0.116	-0.063	0.113	-0.058	0.111	-0.065	0.061	-0.064	0.063	-0.033
200	50	RMSE	0.143	0.064	0.138	0.064	0.130	0.059	0.127	0.066	0.082	0.065	0.080	0.034
200	100	Bias	0.158	-0.029	0.156	-0.028	0.145	-0.026	0.143	-0.029	0.092	-0.029	0.085	-0.014
200	100	RMSE	0.168	0.030	0.166	0.029	0.154	0.027	0.152	0.030	0.101	0.030	0.093	0.014
200	200	Bias	0.185	-0.012	0.183	-0.012	0.163	-0.011	0.161	-0.013	0.109	-0.014	0.099	-0.006
200	200	RMSE	0.191	0.013	0.188	0.013	0.168	0.012	0.166	0.013	0.114	0.014	0.104	0.007
N	T		Design 4: Location-scale shift with heterogeneous slopes											
100	50	Bias	0.131	-0.062	0.127	-0.060	0.119	-0.057	0.117	-0.064	0.061	-0.065	0.062	-0.034
100	50	RMSE	0.148	0.063	0.145	0.062	0.140	0.059	0.139	0.066	0.080	0.067	0.077	0.036
100	100	Bias	0.158	-0.026	0.156	-0.026	0.153	-0.023	0.152	-0.028	0.090	-0.029	0.085	-0.013
100	100	RMSE	0.169	0.027	0.167	0.028	0.162	0.025	0.162	0.029	0.099	0.030	0.093	0.015
100	200	Bias	0.181	-0.010	0.178	-0.010	0.168	-0.009	0.167	-0.011	0.111	-0.012	0.101	-0.004
100	200	RMSE	0.185	0.012	0.182	0.012	0.174	0.010	0.173	0.012	0.116	0.014	0.106	0.006
200	50	Bias	0.133	-0.062	0.131	-0.063	0.123	-0.059	0.119	-0.066	0.064	-0.065	0.065	-0.034
200	50	RMSE	0.151	0.063	0.149	0.064	0.140	0.060	0.136	0.067	0.086	0.065	0.082	0.035
200	100	Bias	0.164	-0.029	0.161	-0.029	0.150	-0.026	0.147	-0.029	0.094	-0.030	0.087	-0.014
200	100	RMSE	0.172	0.029	0.169	0.029	0.159	0.026	0.156	0.029	0.103	0.030	0.096	0.015
200	200	Bias	0.191	-0.012	0.189	-0.013	0.170	-0.011	0.170	-0.013	0.110	-0.014	0.101	-0.006
200	200	RMSE	0.196	0.013	0.194	0.013	0.175	0.012	0.175	0.014	0.115	0.015	0.106	0.007

TABLE S.2. *Bias and root mean square error (RMSE) of quantile regression estimators for β in Designs 3 and 4. In all the variations of the model, $\lambda = 0.5$. Also, see notes to Table S.1.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.110	0.051	0.411	-0.043	-0.107	0.053	0.412	-0.028
100	50	RMSE	0.164	0.058	0.461	0.069	0.158	0.062	0.463	0.068
100	100	Bias	-0.168	0.029	0.493	-0.017	-0.163	0.028	0.495	-0.013
100	100	RMSE	0.197	0.033	0.520	0.036	0.193	0.034	0.526	0.036
100	200	Bias	-0.196	0.009	0.503	-0.005	-0.195	0.009	0.502	-0.005
100	200	RMSE	0.213	0.017	0.521	0.022	0.212	0.016	0.521	0.024
200	50	Bias	-0.109	0.058	0.416	-0.044	-0.105	0.058	0.416	-0.030
200	50	RMSE	0.156	0.062	0.456	0.060	0.152	0.061	0.464	0.053
200	100	Bias	-0.167	0.030	0.476	-0.021	-0.164	0.030	0.474	-0.014
200	100	RMSE	0.198	0.032	0.499	0.029	0.194	0.033	0.498	0.028
200	200	Bias	-0.213	0.014	0.525	-0.008	-0.212	0.013	0.526	-0.005
200	200	RMSE	0.227	0.016	0.541	0.017	0.226	0.016	0.542	0.018
N	T		t_4 distribution							
100	50	Bias	-0.089	0.049	0.362	-0.042	-0.086	0.055	0.371	-0.031
100	50	RMSE	0.139	0.058	0.401	0.075	0.140	0.067	0.416	0.073
100	100	Bias	-0.149	0.027	0.454	-0.017	-0.146	0.029	0.460	-0.018
100	100	RMSE	0.183	0.032	0.489	0.039	0.180	0.036	0.498	0.042
100	200	Bias	-0.179	0.010	0.489	-0.003	-0.178	0.012	0.496	-0.002
100	200	RMSE	0.198	0.015	0.510	0.023	0.197	0.019	0.517	0.027
200	50	Bias	-0.103	0.054	0.390	-0.046	-0.099	0.061	0.400	-0.036
200	50	RMSE	0.159	0.058	0.429	0.064	0.155	0.065	0.443	0.082
200	100	Bias	-0.132	0.028	0.419	-0.020	-0.131	0.029	0.422	-0.021
200	100	RMSE	0.161	0.031	0.438	0.032	0.159	0.033	0.442	0.037
200	200	Bias	-0.179	0.012	0.469	-0.008	-0.180	0.014	0.479	-0.007
200	200	RMSE	0.196	0.015	0.485	0.019	0.197	0.017	0.495	0.022
N	T		χ^2_3 distribution							
100	50	Bias	-0.035	0.075	0.320	-0.035	-0.037	0.041	0.309	-0.015
100	50	RMSE	0.109	0.092	0.367	0.111	0.104	0.058	0.351	0.084
100	100	Bias	-0.086	0.037	0.378	-0.015	-0.079	0.017	0.351	-0.007
100	100	RMSE	0.126	0.049	0.419	0.059	0.115	0.027	0.388	0.040
100	200	Bias	-0.123	0.018	0.409	-0.003	-0.109	0.006	0.373	0.000
100	200	RMSE	0.147	0.030	0.431	0.044	0.131	0.016	0.393	0.027
200	50	Bias	-0.035	0.072	0.321	-0.050	-0.038	0.039	0.310	-0.026
200	50	RMSE	0.098	0.082	0.354	0.088	0.090	0.049	0.343	0.061
200	100	Bias	-0.093	0.038	0.400	-0.022	-0.083	0.019	0.368	-0.008
200	100	RMSE	0.127	0.044	0.423	0.049	0.114	0.026	0.388	0.034
200	200	Bias	-0.122	0.020	0.430	-0.007	-0.108	0.009	0.391	-0.002
200	200	RMSE	0.141	0.026	0.445	0.030	0.127	0.014	0.405	0.021

TABLE S.3. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ in Design 1. In all the variations of the model, $\lambda = 0.5$.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.112	0.065	0.442	-0.015	-0.107	0.063	0.439	-0.004
100	50	RMSE	0.170	0.071	0.491	0.057	0.166	0.070	0.487	0.066
100	100	Bias	-0.179	0.035	0.526	0.002	-0.176	0.034	0.519	0.007
100	100	RMSE	0.210	0.041	0.554	0.033	0.208	0.040	0.547	0.036
100	200	Bias	-0.220	-0.006	0.488	-0.036	-0.220	-0.006	0.480	-0.035
100	200	RMSE	0.235	0.014	0.510	0.042	0.235	0.015	0.502	0.042
200	50	Bias	-0.110	0.065	0.438	-0.030	-0.107	0.064	0.428	-0.018
200	50	RMSE	0.158	0.068	0.476	0.051	0.154	0.067	0.471	0.046
200	100	Bias	-0.157	0.039	0.502	0.001	-0.155	0.038	0.500	0.005
200	100	RMSE	0.188	0.041	0.525	0.022	0.186	0.041	0.525	0.027
200	200	Bias	-0.214	0.013	0.510	-0.007	-0.212	0.012	0.507	-0.006
200	200	RMSE	0.229	0.015	0.522	0.017	0.227	0.015	0.520	0.017
N	T		t_4 distribution							
100	50	Bias	-0.059	0.083	0.436	0.027	-0.057	0.089	0.446	0.038
100	50	RMSE	0.123	0.089	0.468	0.068	0.123	0.096	0.484	0.077
100	100	Bias	-0.119	0.062	0.532	0.057	-0.116	0.065	0.541	0.055
100	100	RMSE	0.161	0.065	0.561	0.066	0.159	0.068	0.574	0.067
100	200	Bias	-0.178	0.017	0.511	0.013	-0.177	0.020	0.519	0.014
100	200	RMSE	0.197	0.021	0.530	0.026	0.196	0.024	0.540	0.030
200	50	Bias	-0.092	0.070	0.428	-0.012	-0.085	0.077	0.440	-0.024
200	50	RMSE	0.151	0.073	0.465	0.045	0.145	0.081	0.482	0.299
200	100	Bias	-0.118	0.048	0.472	0.023	-0.117	0.050	0.475	0.021
200	100	RMSE	0.148	0.050	0.490	0.034	0.147	0.052	0.493	0.037
200	200	Bias	-0.183	0.012	0.481	-0.008	-0.185	0.014	0.488	-0.007
200	200	RMSE	0.200	0.014	0.497	0.018	0.201	0.017	0.505	0.022
N	T		χ^2_3 distribution							
100	50	Bias	-0.051	0.066	0.308	-0.051	-0.053	0.032	0.299	-0.034
100	50	RMSE	0.123	0.085	0.362	0.133	0.116	0.051	0.347	0.088
100	100	Bias	-0.113	0.013	0.335	-0.062	-0.106	-0.006	0.310	-0.054
100	100	RMSE	0.146	0.034	0.382	0.084	0.135	0.023	0.352	0.068
100	200	Bias	-0.155	-0.008	0.362	-0.053	-0.142	-0.020	0.327	-0.051
100	200	RMSE	0.175	0.025	0.389	0.069	0.160	0.025	0.350	0.058
200	50	Bias	-0.060	0.055	0.341	-0.078	-0.064	0.024	0.321	-0.054
200	50	RMSE	0.109	0.066	0.376	0.129	0.103	0.037	0.355	0.076
200	100	Bias	-0.111	0.020	0.365	-0.060	-0.105	0.001	0.337	-0.045
200	100	RMSE	0.143	0.031	0.385	0.074	0.134	0.015	0.358	0.053
200	200	Bias	-0.133	0.018	0.424	-0.012	-0.119	0.007	0.388	-0.007
200	200	RMSE	0.150	0.024	0.439	0.032	0.134	0.013	0.402	0.022

TABLE S.4. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ in Design 2. In all the variations of the model, $\lambda = 0.5$.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.110	0.058	0.420	-0.048	-0.091	0.066	0.407	-0.011
100	50	RMSE	0.161	0.066	0.467	0.076	0.145	0.073	0.449	0.067
100	100	Bias	-0.183	0.032	0.502	-0.015	-0.166	0.035	0.460	0.000
100	100	RMSE	0.214	0.037	0.532	0.036	0.198	0.041	0.487	0.039
100	200	Bias	-0.208	0.013	0.525	-0.005	-0.190	0.016	0.485	0.005
100	200	RMSE	0.226	0.018	0.546	0.022	0.209	0.021	0.507	0.024
200	50	Bias	-0.100	0.065	0.409	-0.050	-0.082	0.074	0.379	-0.012
200	50	RMSE	0.158	0.068	0.462	0.063	0.142	0.077	0.421	0.047
200	100	Bias	-0.163	0.033	0.471	-0.024	-0.146	0.038	0.440	-0.006
200	100	RMSE	0.191	0.036	0.495	0.038	0.176	0.040	0.462	0.027
200	200	Bias	-0.212	0.014	0.520	-0.010	-0.194	0.017	0.477	0.000
200	200	RMSE	0.228	0.017	0.537	0.019	0.210	0.020	0.492	0.016
N	T		t_4 distribution							
100	50	Bias	-0.087	0.054	0.377	-0.056	-0.073	0.069	0.357	-0.005
100	50	RMSE	0.144	0.063	0.425	0.083	0.137	0.079	0.395	0.223
100	100	Bias	-0.129	0.029	0.452	-0.017	-0.114	0.040	0.418	0.002
100	100	RMSE	0.157	0.035	0.480	0.042	0.146	0.046	0.444	0.041
100	200	Bias	-0.173	0.012	0.481	-0.005	-0.158	0.019	0.446	0.008
100	200	RMSE	0.194	0.018	0.501	0.024	0.181	0.024	0.464	0.025
200	50	Bias	-0.087	0.058	0.387	-0.051	-0.069	0.074	0.372	-0.014
200	50	RMSE	0.130	0.063	0.426	0.068	0.115	0.079	0.407	0.080
200	100	Bias	-0.139	0.030	0.439	-0.025	-0.124	0.039	0.405	-0.008
200	100	RMSE	0.168	0.033	0.461	0.035	0.154	0.042	0.424	0.029
200	200	Bias	-0.173	0.014	0.462	-0.009	-0.157	0.019	0.424	0.000
200	200	RMSE	0.188	0.016	0.475	0.020	0.173	0.022	0.436	0.020
N	T		χ^2_3 distribution							
100	50	Bias	-0.033	0.082	0.274	-0.055	-0.025	0.060	0.260	0.008
100	50	RMSE	0.119	0.100	0.321	0.126	0.106	0.072	0.294	0.080
100	100	Bias	-0.091	0.044	0.354	-0.026	-0.067	0.025	0.326	0.003
100	100	RMSE	0.133	0.055	0.385	0.069	0.110	0.036	0.352	0.050
100	200	Bias	-0.136	0.017	0.367	-0.015	-0.102	0.010	0.320	0.002
100	200	RMSE	0.159	0.028	0.392	0.046	0.127	0.018	0.343	0.028
200	50	Bias	-0.037	0.082	0.275	-0.063	-0.025	0.057	0.274	0.001
200	50	RMSE	0.114	0.089	0.318	0.090	0.096	0.064	0.304	0.052
200	100	Bias	-0.091	0.041	0.343	-0.032	-0.064	0.024	0.312	-0.002
200	100	RMSE	0.127	0.047	0.364	0.054	0.100	0.029	0.329	0.033
200	200	Bias	-0.130	0.020	0.361	-0.015	-0.095	0.012	0.321	0.002
200	200	RMSE	0.152	0.026	0.377	0.035	0.117	0.017	0.335	0.022

TABLE S.5. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ in Design 3. In all the variations of the model, $\lambda = 0.5$.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
N	T		Normal Distribution							
100	50	Bias	-0.106	0.068	0.453	-0.038	-0.086	0.077	0.433	0.003
100	50	RMSE	0.156	0.072	0.504	0.064	0.142	0.082	0.480	0.061
100	100	Bias	-0.164	0.027	0.469	-0.021	-0.146	0.034	0.442	-0.003
100	100	RMSE	0.195	0.033	0.498	0.041	0.179	0.039	0.468	0.036
100	200	Bias	-0.201	0.011	0.514	-0.006	-0.181	0.015	0.477	0.005
100	200	RMSE	0.216	0.017	0.532	0.024	0.198	0.021	0.494	0.024
200	50	Bias	-0.116	0.069	0.443	-0.036	-0.097	0.078	0.428	-0.004
200	50	RMSE	0.169	0.072	0.484	0.055	0.152	0.081	0.464	0.044
200	100	Bias	-0.174	0.026	0.476	-0.035	-0.155	0.030	0.446	-0.019
200	100	RMSE	0.195	0.029	0.499	0.042	0.177	0.032	0.466	0.032
200	200	Bias	-0.215	0.014	0.554	-0.010	-0.196	0.018	0.516	0.001
200	200	RMSE	0.230	0.017	0.567	0.019	0.212	0.021	0.529	0.016
N	T		t_4 distribution							
100	50	Bias	-0.092	0.066	0.411	-0.034	-0.076	0.086	0.393	0.007
100	50	RMSE	0.160	0.073	0.457	0.071	0.149	0.094	0.433	0.076
100	100	Bias	-0.153	0.021	0.462	-0.029	-0.138	0.033	0.432	-0.012
100	100	RMSE	0.182	0.029	0.494	0.046	0.169	0.040	0.462	0.042
100	200	Bias	-0.173	0.019	0.505	0.010	-0.155	0.026	0.470	0.023
100	200	RMSE	0.195	0.023	0.525	0.025	0.180	0.030	0.488	0.032
200	50	Bias	-0.109	0.050	0.390	-0.068	-0.090	0.069	0.364	-0.031
200	50	RMSE	0.153	0.055	0.428	0.080	0.138	0.073	0.397	0.058
200	100	Bias	-0.145	0.033	0.457	-0.015	-0.128	0.040	0.424	-0.001
200	100	RMSE	0.172	0.036	0.480	0.030	0.156	0.043	0.445	0.031
200	200	Bias	-0.182	0.015	0.488	-0.008	-0.167	0.019	0.455	0.000
200	200	RMSE	0.199	0.018	0.502	0.019	0.185	0.022	0.469	0.018
N	T		χ^2_3 distribution							
100	50	Bias	-0.031	0.092	0.288	-0.057	-0.021	0.064	0.281	0.010
100	50	RMSE	0.108	0.107	0.344	0.111	0.097	0.076	0.320	0.073
100	100	Bias	-0.060	0.074	0.410	0.035	-0.034	0.056	0.385	0.064
100	100	RMSE	0.112	0.082	0.438	0.072	0.093	0.062	0.409	0.078
100	200	Bias	-0.106	0.051	0.445	0.052	-0.070	0.042	0.402	0.068
100	200	RMSE	0.134	0.056	0.467	0.067	0.102	0.045	0.422	0.074
200	50	Bias	-0.044	0.083	0.284	-0.064	-0.032	0.055	0.272	-0.004
200	50	RMSE	0.121	0.092	0.323	0.098	0.105	0.063	0.299	0.059
200	100	Bias	-0.101	0.041	0.334	-0.034	-0.073	0.023	0.308	-0.005
200	100	RMSE	0.137	0.047	0.356	0.055	0.108	0.028	0.325	0.031
200	200	Bias	-0.136	0.021	0.358	-0.014	-0.102	0.013	0.320	0.003
200	200	RMSE	0.157	0.026	0.374	0.031	0.123	0.017	0.334	0.020

TABLE S.6. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ in Design 4. In all the variations of the model, $\lambda = 0.5$.*

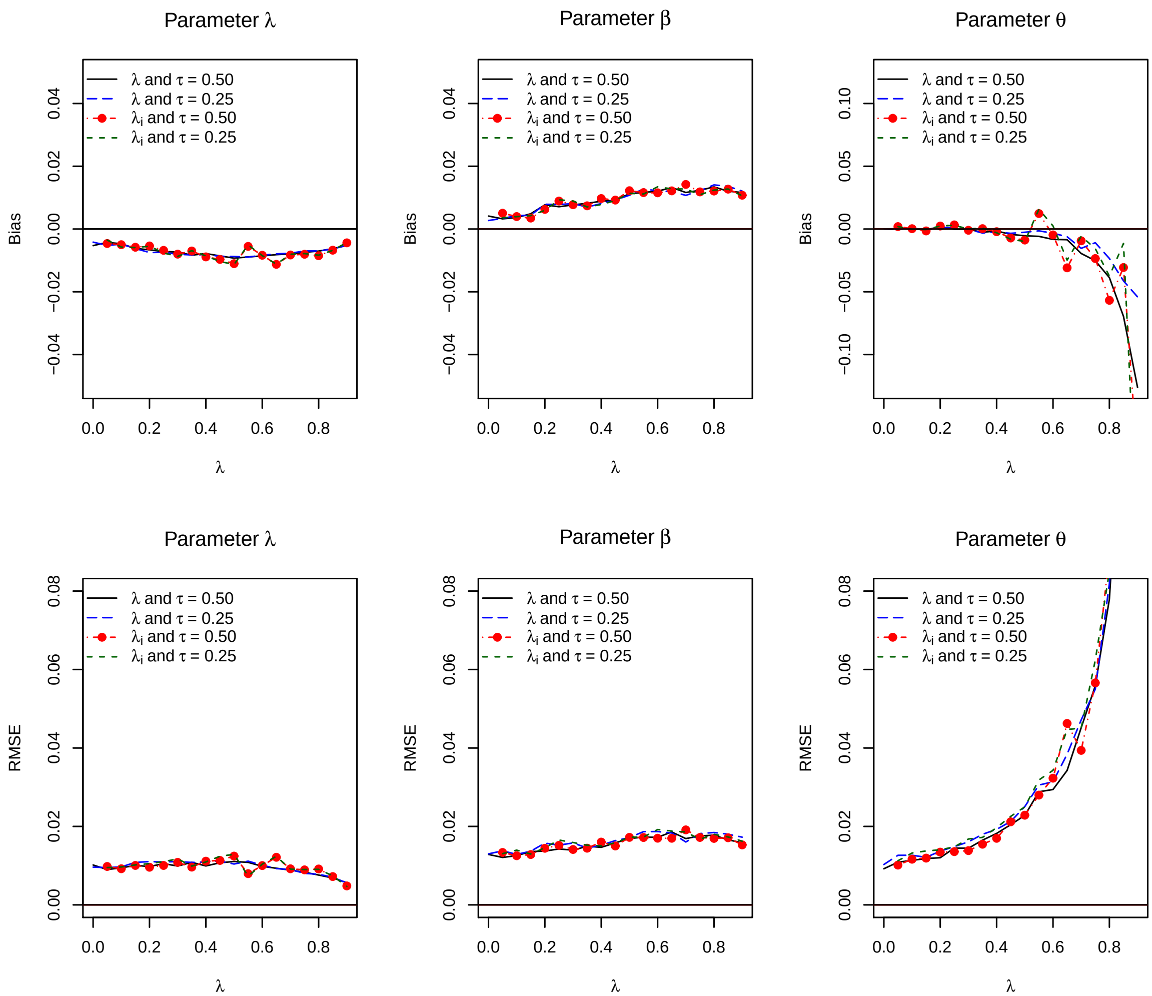

FIGURE S.1. *Small sample performance of the QMG estimator for different values of* λ. *The figure present Bias and RMSE of the QMG estimator for* $E(\lambda(\tau))$, $E(\beta(\tau))$ *and* $E(\theta(\tau))$ *at the 0.25 and 0.50 quantiles.*

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			λ		β		λ		β		λ		β	
			0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
N	T		Design 1: Location shift with homogeneous slopes											
100	100	Std error	0.009	0.010	0.024	0.024	0.009	0.010	0.026	0.027	0.011	0.009	0.044	0.033
100	100	Std dev	0.009	0.010	0.018	0.020	0.008	0.009	0.018	0.022	0.009	0.007	0.034	0.024
100	200	Std error	0.006	0.007	0.016	0.016	0.006	0.007	0.017	0.018	0.008	0.006	0.029	0.021
100	200	Std dev	0.006	0.006	0.013	0.014	0.005	0.007	0.013	0.015	0.007	0.005	0.022	0.015
100	400	Std error	0.004	0.005	0.011	0.011	0.004	0.005	0.012	0.012	0.005	0.004	0.019	0.014
100	400	Std dev	0.004	0.005	0.009	0.010	0.004	0.004	0.009	0.011	0.005	0.003	0.015	0.010
N	T		Design 2: Location shift with heterogeneous slopes											
100	100	Std error	0.009	0.010	0.028	0.027	0.009	0.010	0.030	0.030	0.011	0.009	0.046	0.036
100	100	Std dev	0.008	0.010	0.022	0.023	0.008	0.009	0.024	0.026	0.009	0.007	0.034	0.027
100	200	Std error	0.006	0.007	0.021	0.021	0.006	0.007	0.022	0.023	0.008	0.006	0.032	0.026
100	200	Std dev	0.006	0.006	0.018	0.019	0.006	0.006	0.020	0.022	0.006	0.004	0.026	0.021
100	400	Std error	0.004	0.005	0.018	0.018	0.004	0.005	0.018	0.019	0.005	0.004	0.024	0.020
100	400	Std dev	0.005	0.005	0.018	0.019	0.004	0.004	0.018	0.018	0.005	0.003	0.021	0.019
N	T		Design 3: Location-scale shift with homogeneous slopes											
100	100	Std error	0.010	0.010	0.026	0.025	0.010	0.010	0.028	0.029	0.012	0.009	0.051	0.039
100	100	Std dev	0.010	0.010	0.019	0.020	0.008	0.009	0.020	0.023	0.009	0.007	0.036	0.025
100	200	Std error	0.007	0.007	0.017	0.017	0.006	0.007	0.018	0.019	0.008	0.006	0.033	0.025
100	200	Std dev	0.006	0.006	0.013	0.014	0.006	0.006	0.013	0.015	0.006	0.004	0.023	0.017
100	400	Std error	0.005	0.005	0.012	0.012	0.004	0.005	0.012	0.013	0.006	0.004	0.022	0.016
100	400	Std dev	0.005	0.005	0.009	0.010	0.004	0.005	0.010	0.011	0.004	0.003	0.016	0.012
N	T		Design 4: Location-scale shift with heterogeneous slopes											
100	100	Std error	0.010	0.010	0.030	0.029	0.010	0.010	0.032	0.032	0.012	0.009	0.052	0.041
100	100	Std dev	0.009	0.010	0.024	0.025	0.008	0.009	0.025	0.027	0.009	0.006	0.035	0.027
100	200	Std error	0.007	0.007	0.022	0.022	0.006	0.007	0.023	0.024	0.008	0.006	0.036	0.028
100	200	Std dev	0.006	0.006	0.019	0.020	0.006	0.007	0.018	0.021	0.006	0.005	0.029	0.023
100	400	Std error	0.005	0.005	0.019	0.018	0.004	0.005	0.019	0.019	0.006	0.004	0.026	0.022
100	400	Std dev	0.004	0.005	0.016	0.017	0.004	0.005	0.017	0.018	0.004	0.003	0.022	0.019

TABLE S.7. *Standard error (Std error) of the QMG estimator for λ and β. Std dev denotes standard deviation and $\lambda = 0.5$ in the simulations.*

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			λ		β		λ		β		λ		β	
			0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25
N	T	Design	Coverage probability for a nominal 95% confidence interval											
100	100	1	0.320	0.385	0.885	0.865	0.323	0.293	0.933	0.858	0.238	0.663	0.963	0.983
100	200	1	0.775	0.778	0.933	0.938	0.775	0.743	0.968	0.938	0.648	0.923	0.970	0.995
100	400	1	0.928	0.925	0.965	0.958	0.950	0.945	0.980	0.973	0.883	0.980	0.985	0.988
100	100	2	0.298	0.343	0.915	0.898	0.330	0.265	0.910	0.865	0.228	0.695	0.953	0.985
100	200	2	0.743	0.763	0.953	0.948	0.840	0.755	0.953	0.930	0.628	0.920	0.973	0.978
100	400	2	0.930	0.933	0.945	0.945	0.958	0.953	0.945	0.968	0.888	0.968	0.968	0.975
100	100	3	0.253	0.265	0.860	0.753	0.283	0.255	0.923	0.795	0.295	0.818	0.950	0.975
100	200	3	0.698	0.715	0.950	0.888	0.743	0.678	0.978	0.893	0.713	0.948	0.968	0.985
100	400	3	0.910	0.930	0.985	0.948	0.945	0.905	0.983	0.945	0.930	0.988	0.983	0.998
100	100	4	0.220	0.265	0.890	0.788	0.240	0.220	0.925	0.820	0.328	0.828	0.965	0.978
100	200	4	0.708	0.700	0.958	0.913	0.745	0.710	0.980	0.920	0.728	0.943	0.963	0.960
100	400	4	0.948	0.923	0.975	0.950	0.948	0.938	0.960	0.940	0.900	0.993	0.978	0.978
N	T	Design	Power performance											
100	100	1	1.000	1.000	0.930	0.918	1.000	1.000	0.905	0.783	1.000	1.000	0.303	0.805
100	200	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.918	1.000
100	400	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	100	2	1.000	1.000	0.808	0.790	1.000	1.000	0.740	0.635	1.000	1.000	0.263	0.708
100	200	2	1.000	1.000	0.993	0.990	1.000	1.000	0.995	0.985	1.000	1.000	0.793	0.985
100	400	2	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.985	1.000
100	100	3	1.000	1.000	0.840	0.745	1.000	1.000	0.805	0.583	1.000	1.000	0.123	0.528
100	200	3	1.000	1.000	1.000	0.998	1.000	1.000	1.000	0.998	1.000	1.000	0.768	0.995
100	400	3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	100	4	1.000	1.000	0.680	0.575	1.000	1.000	0.670	0.530	1.000	1.000	0.098	0.435
100	200	4	1.000	1.000	0.988	0.988	1.000	1.000	0.990	0.963	1.000	1.000	0.668	0.923
100	400	4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.968	0.998

TABLE S.8. *Empirical coverage probability for a nominal 95 percent confidence interval and Power performance of the QMG estimator. We compute the power for the estimation of* λ *and* β *with the alternatives* $H_a : \lambda = 0.55$ *and* $H_a : \beta = 1.1$.

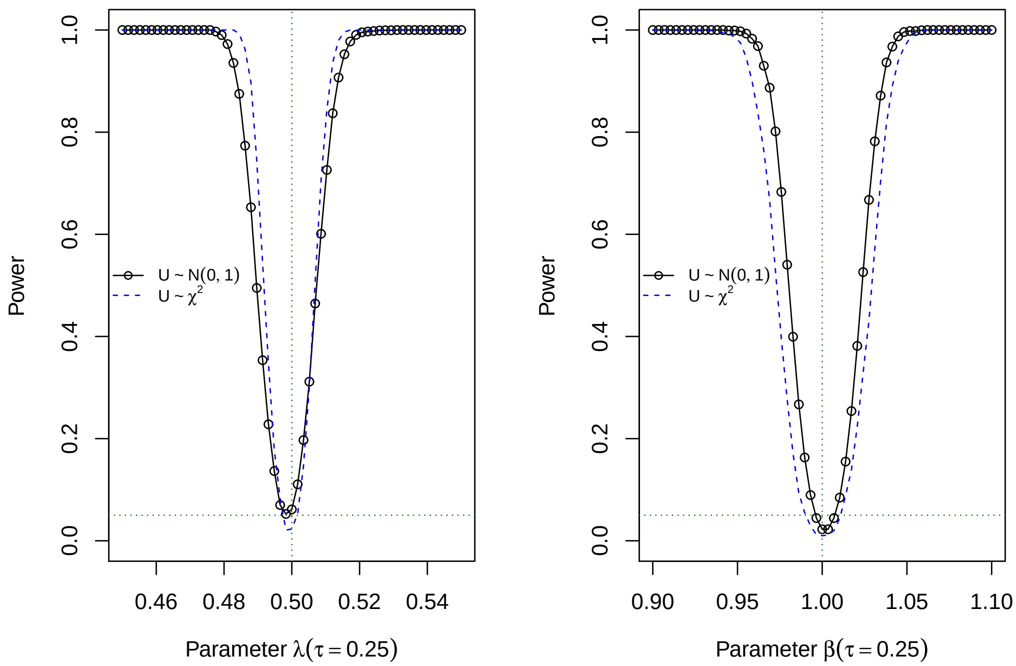

FIGURE S.2. *Power of the QMG estimator against different alternatives.*

λ	Design		Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
0.25	1	Bias	0.302	-0.007	0.302	-0.007	0.289	-0.006	0.289	-0.007	0.231	-0.008	0.215	-0.002
0.25	1	RMSE	0.310	0.010	0.309	0.010	0.297	0.008	0.297	0.010	0.240	0.011	0.223	0.005
0.25	2	Bias	0.315	-0.007	0.313	-0.007	0.298	-0.005	0.298	-0.007	0.238	-0.008	0.222	-0.002
0.25	2	RMSE	0.322	0.010	0.321	0.011	0.305	0.008	0.306	0.010	0.247	0.010	0.230	0.005
0.25	3	Bias	0.314	-0.008	0.310	-0.008	0.282	-0.007	0.280	-0.008	0.193	-0.009	0.177	-0.003
0.25	3	RMSE	0.321	0.011	0.318	0.011	0.290	0.010	0.289	0.011	0.200	0.012	0.185	0.005
0.25	4	Bias	0.309	-0.008	0.304	-0.008	0.286	-0.007	0.284	-0.008	0.194	-0.009	0.178	-0.002
0.25	4	RMSE	0.315	0.012	0.310	0.012	0.294	0.009	0.293	0.011	0.201	0.012	0.185	0.005
0.75	1	Bias	0.077	-0.007	0.077	-0.007	0.075	-0.007	0.075	-0.009	0.064	-0.014	0.059	-0.007
0.75	1	RMSE	0.079	0.008	0.079	0.008	0.078	0.008	0.078	0.009	0.068	0.015	0.063	0.007
0.75	2	Bias	0.081	-0.008	0.080	-0.008	0.078	-0.007	0.078	-0.009	0.066	-0.015	0.062	-0.007
0.75	2	RMSE	0.084	0.009	0.083	0.009	0.080	0.008	0.081	0.010	0.070	0.015	0.065	0.007
0.75	3	Bias	0.081	-0.009	0.080	-0.009	0.075	-0.009	0.075	-0.010	0.052	-0.014	0.048	-0.007
0.75	3	RMSE	0.084	0.010	0.083	0.010	0.078	0.009	0.078	0.011	0.055	0.015	0.051	0.007
0.75	4	Bias	0.076	-0.009	0.076	-0.010	0.079	-0.009	0.078	-0.008	0.052	-0.013	0.047	-0.006
0.75	4	RMSE	0.079	0.009	0.079	0.011	0.081	0.009	0.080	0.009	0.054	0.014	0.050	0.007

TABLE S.9. *Bias and root mean square error (RMSE) of quantile regression estimators for β when $\lambda \in \{0.25, 0.75\}$, $N = 100$ and $T = 200$.*

λ	Design		$\tau=0.50$ quantile				$\tau=0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
					Normal Distribution					
0.25	1	Bias	-0.267	0.006	0.313	-0.001	-0.266	0.005	0.313	-0.002
0.25	1	RMSE	0.281	0.015	0.324	0.014	0.281	0.014	0.325	0.015
0.25	2	Bias	-0.292	-0.010	0.300	-0.022	-0.291	-0.010	0.298	-0.022
0.25	2	RMSE	0.305	0.016	0.313	0.026	0.304	0.017	0.313	0.027
0.25	3	Bias	-0.280	0.009	0.326	0.000	-0.264	0.011	0.285	0.004
0.25	3	RMSE	0.295	0.015	0.339	0.014	0.279	0.017	0.298	0.016
0.25	4	Bias	-0.273	0.007	0.321	-0.002	-0.255	0.011	0.280	0.004
0.25	4	RMSE	0.286	0.015	0.332	0.015	0.268	0.018	0.292	0.016
					t_4 distribution					
0.25	1	Bias	-0.247	0.006	0.308	0.000	-0.246	0.008	0.310	0.001
0.25	1	RMSE	0.264	0.013	0.321	0.014	0.263	0.016	0.324	0.017
0.25	2	Bias	-0.247	0.013	0.321	0.010	-0.246	0.015	0.325	0.011
0.25	2	RMSE	0.263	0.017	0.334	0.018	0.262	0.021	0.338	0.020
0.25	3	Bias	-0.239	0.007	0.300	-0.001	-0.225	0.014	0.262	0.007
0.25	3	RMSE	0.257	0.015	0.313	0.016	0.246	0.019	0.274	0.017
0.25	4	Bias	-0.239	0.014	0.314	0.009	-0.225	0.021	0.275	0.017
0.25	4	RMSE	0.258	0.019	0.327	0.018	0.246	0.025	0.287	0.023
					χ^2_3 distribution					
0.25	1	Bias	-0.178	0.011	0.257	0.003	-0.163	0.002	0.235	0.001
0.25	1	RMSE	0.198	0.026	0.271	0.029	0.182	0.015	0.248	0.019
0.25	2	Bias	-0.212	-0.015	0.224	-0.032	-0.196	-0.024	0.204	-0.033
0.25	2	RMSE	0.229	0.028	0.242	0.043	0.213	0.028	0.220	0.038
0.25	3	Bias	-0.194	0.008	0.225	-0.005	-0.159	0.006	0.187	0.003
0.25	3	RMSE	0.214	0.024	0.241	0.029	0.180	0.015	0.202	0.018
0.25	4	Bias	-0.170	0.043	0.276	0.040	-0.131	0.037	0.243	0.046
0.25	4	RMSE	0.193	0.048	0.290	0.048	0.156	0.040	0.256	0.050

Table S.10. *Bias and root mean square error (RMSE) of quantile regression estimators when $\lambda = 0.25$, $N = 100$ and $T = 200$.*

λ	Design		$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
					Normal Distribution					
0.75	1	Bias	-0.107	0.012	1.180	-0.029	-0.106	0.012	1.184	-0.019
0.75	1	RMSE	0.130	0.018	1.223	0.057	0.129	0.017	1.229	0.055
0.75	2	Bias	-0.130	-0.003	1.170	-0.085	-0.129	-0.003	1.150	-0.079
0.75	2	RMSE	0.150	0.012	1.221	0.101	0.148	0.013	1.200	0.097
0.75	3	Bias	-0.118	0.016	1.249	-0.029	-0.100	0.019	1.205	-0.005
0.75	3	RMSE	0.141	0.020	1.300	0.061	0.125	0.023	1.260	0.057
0.75	4	Bias	-0.111	0.014	1.225	-0.030	-0.091	0.018	1.177	0.000
0.75	4	RMSE	0.132	0.018	1.267	0.062	0.116	0.022	1.217	0.056
					t_4 distribution					
0.75	1	Bias	-0.099	0.014	1.177	-0.024	-0.098	0.016	1.192	-0.020
0.75	1	RMSE	0.124	0.018	1.224	0.058	0.124	0.022	1.244	0.067
0.75	2	Bias	-0.096	0.021	1.229	0.008	-0.096	0.023	1.245	0.010
0.75	2	RMSE	0.123	0.024	1.274	0.054	0.123	0.028	1.294	0.064
0.75	3	Bias	-0.100	0.016	1.175	-0.032	-0.083	0.024	1.138	-0.003
0.75	3	RMSE	0.126	0.021	1.224	0.061	0.113	0.028	1.184	0.056
0.75	4	Bias	-0.097	0.023	1.229	-0.004	-0.079	0.031	1.193	0.026
0.75	4	RMSE	0.125	0.027	1.276	0.053	0.112	0.035	1.236	0.060
					χ^2_3 distribution					
0.75	1	Bias	-0.071	0.029	1.028	-0.046	-0.064	0.013	0.935	-0.020
0.75	1	RMSE	0.106	0.037	1.081	0.105	0.096	0.020	0.983	0.063
0.75	2	Bias	-0.102	0.002	0.941	-0.149	-0.095	-0.014	0.853	-0.123
0.75	2	RMSE	0.128	0.023	1.003	0.177	0.119	0.021	0.908	0.136
0.75	3	Bias	-0.079	0.028	0.930	-0.081	-0.051	0.016	0.832	-0.020
0.75	3	RMSE	0.109	0.036	0.985	0.129	0.084	0.022	0.879	0.068
0.75	4	Bias	-0.045	0.062	1.085	0.059	-0.014	0.048	0.995	0.115
0.75	4	RMSE	0.088	0.067	1.134	0.105	0.070	0.050	1.038	0.130

TABLE S.11. *Bias and root mean square error (RMSE) of quantile regression estimators when* $\lambda = 0.75$, $N = 100$ *and* $T = 200$.

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG
Design 1: Location shift with homogeneous slopes														
100	50	Bias	-0.031	-0.052	-0.031	-0.053	-0.031	-0.050	-0.035	-0.056	-0.049	-0.066	-0.025	-0.035
100	50	RMSE	0.032	0.054	0.033	0.055	0.032	0.051	0.037	0.057	0.050	0.068	0.026	0.037
100	100	Bias	-0.015	-0.024	-0.015	-0.024	-0.015	-0.022	-0.017	-0.025	-0.023	-0.029	-0.011	-0.014
100	100	RMSE	0.017	0.025	0.016	0.025	0.017	0.024	0.019	0.027	0.025	0.030	0.013	0.015
100	200	Bias	-0.008	-0.008	-0.007	-0.008	-0.007	-0.007	-0.009	-0.009	-0.012	-0.013	-0.005	-0.004
100	200	RMSE	0.009	0.010	0.008	0.010	0.008	0.009	0.010	0.011	0.013	0.014	0.007	0.006
200	50	Bias	-0.032	-0.057	-0.031	-0.056	-0.032	-0.053	-0.037	-0.060	-0.048	-0.068	-0.025	-0.037
200	50	RMSE	0.032	0.058	0.032	0.057	0.033	0.054	0.037	0.061	0.049	0.069	0.026	0.037
200	100	Bias	-0.015	-0.026	-0.015	-0.025	-0.015	-0.023	-0.017	-0.026	-0.024	-0.032	-0.012	-0.015
200	100	RMSE	0.016	0.026	0.016	0.026	0.016	0.024	0.018	0.027	0.024	0.032	0.012	0.016
200	200	Bias	-0.008	-0.011	-0.007	-0.011	-0.007	-0.010	-0.008	-0.011	-0.012	-0.015	-0.006	-0.006
200	200	RMSE	0.008	0.012	0.008	0.011	0.008	0.011	0.009	0.012	0.012	0.015	0.006	0.007
Design 2: Location shift with heterogeneous slopes														
100	50	Bias	-0.031	-0.053	-0.031	-0.053	-0.031	-0.049	-0.035	-0.055	-0.049	-0.066	-0.025	-0.035
100	50	RMSE	0.033	0.054	0.033	0.055	0.032	0.051	0.037	0.057	0.050	0.068	0.026	0.036
100	100	Bias	-0.015	-0.023	-0.014	-0.023	-0.015	-0.022	-0.017	-0.025	-0.023	-0.029	-0.011	-0.014
100	100	RMSE	0.017	0.025	0.016	0.024	0.017	0.023	0.019	0.026	0.025	0.030	0.013	0.015
100	200	Bias	-0.008	-0.009	-0.008	-0.009	-0.007	-0.007	-0.009	-0.009	-0.012	-0.012	-0.005	-0.004
100	200	RMSE	0.009	0.010	0.009	0.011	0.008	0.009	0.010	0.011	0.013	0.014	0.007	0.006
200	50	Bias	-0.032	-0.057	-0.032	-0.056	-0.033	-0.054	-0.037	-0.061	-0.048	-0.068	-0.024	-0.036
200	50	RMSE	0.033	0.058	0.032	0.057	0.033	0.054	0.038	0.061	0.049	0.069	0.025	0.037
200	100	Bias	-0.015	-0.025	-0.015	-0.025	-0.016	-0.024	-0.017	-0.027	-0.023	-0.032	-0.011	-0.015
200	100	RMSE	0.016	0.026	0.015	0.026	0.016	0.024	0.018	0.027	0.024	0.032	0.012	0.016
200	200	Bias	-0.007	-0.011	-0.007	-0.010	-0.008	-0.010	-0.009	-0.011	-0.012	-0.015	-0.006	-0.007
200	200	RMSE	0.008	0.012	0.008	0.011	0.008	0.011	0.009	0.012	0.012	0.015	0.006	0.007

TABLE S.12. *Bias and root mean square error (RMSE) of the unfeasible QMG estimator and the feasible QMG estimator for β in Designs 1 and 2. In all the variations of the model, $\lambda = 0.5$. IQMG denotes the unfeasible version of the estimator.*

			Normal Distribution				t_4 distribution				χ^2_3 distribution			
			$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.25$	
			IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG
Design 3: Location-scale shift with homogeneous slopes														
100	50	Bias	-0.036	-0.059	-0.035	-0.059	-0.036	-0.056	-0.040	-0.063	-0.045	-0.062	-0.022	-0.033
100	50	RMSE	0.038	0.061	0.037	0.061	0.037	0.058	0.042	0.065	0.046	0.064	0.024	0.035
100	100	Bias	-0.017	-0.026	-0.017	-0.026	-0.017	-0.024	-0.019	-0.027	-0.023	-0.030	-0.011	-0.013
100	100	RMSE	0.019	0.027	0.018	0.027	0.018	0.025	0.021	0.029	0.024	0.031	0.012	0.015
100	200	Bias	-0.008	-0.010	-0.008	-0.010	-0.008	-0.009	-0.009	-0.011	-0.011	-0.013	-0.005	-0.005
100	200	RMSE	0.010	0.012	0.010	0.012	0.009	0.010	0.010	0.013	0.012	0.014	0.006	0.007
200	50	Bias	-0.037	-0.063	-0.036	-0.063	-0.036	-0.058	-0.041	-0.065	-0.045	-0.064	-0.022	-0.033
200	50	RMSE	0.037	0.064	0.037	0.064	0.037	0.059	0.042	0.066	0.046	0.065	0.023	0.034
200	100	Bias	-0.018	-0.029	-0.017	-0.028	-0.017	-0.026	-0.019	-0.029	-0.022	-0.029	-0.010	-0.014
200	100	RMSE	0.018	0.030	0.018	0.029	0.018	0.027	0.020	0.030	0.022	0.030	0.011	0.014
200	200	Bias	-0.008	-0.012	-0.008	-0.012	-0.008	-0.011	-0.009	-0.013	-0.011	-0.014	-0.005	-0.006
200	200	RMSE	0.009	0.013	0.009	0.013	0.009	0.012	0.010	0.013	0.012	0.014	0.006	0.007
Design 4: Location-scale shift with heterogeneous slopes														
100	50	Bias	-0.037	-0.062	-0.036	-0.060	-0.035	-0.057	-0.040	-0.064	-0.047	-0.065	-0.023	-0.034
100	50	RMSE	0.038	0.063	0.038	0.062	0.037	0.059	0.043	0.066	0.048	0.067	0.024	0.036
100	100	Bias	-0.017	-0.026	-0.017	-0.026	-0.017	-0.023	-0.020	-0.028	-0.022	-0.029	-0.011	-0.013
100	100	RMSE	0.019	0.027	0.019	0.028	0.018	0.025	0.021	0.029	0.024	0.030	0.012	0.015
100	200	Bias	-0.008	-0.010	-0.008	-0.010	-0.008	-0.009	-0.009	-0.011	-0.010	-0.012	-0.005	-0.004
100	200	RMSE	0.009	0.012	0.010	0.012	0.009	0.010	0.010	0.012	0.012	0.014	0.006	0.006
200	50	Bias	-0.035	-0.062	-0.036	-0.063	-0.037	-0.059	-0.041	-0.066	-0.045	-0.065	-0.022	-0.034
200	50	RMSE	0.036	0.063	0.037	0.064	0.038	0.060	0.042	0.067	0.046	0.065	0.023	0.035
200	100	Bias	-0.018	-0.029	-0.017	-0.029	-0.017	-0.026	-0.019	-0.029	-0.022	-0.030	-0.010	-0.014
200	100	RMSE	0.018	0.029	0.018	0.029	0.018	0.026	0.019	0.029	0.022	0.030	0.011	0.015
200	200	Bias	-0.008	-0.012	-0.008	-0.013	-0.008	-0.011	-0.009	-0.013	-0.011	-0.014	-0.005	-0.006
200	200	RMSE	0.009	0.013	0.009	0.013	0.009	0.012	0.010	0.014	0.012	0.015	0.006	0.007

TABLE S.13. *Bias and root mean square error (RMSE) of the unfeasible QMG estimator and the feasible QMG estimator for β in Designs 3 and 4. In all the variations of the model, $\lambda = 0.5$. IQMG denotes the unfeasible version of the estimator.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG
			Normal Distribution							
100	50	Bias	0.029	0.051	-0.028	-0.043	0.032	0.053	-0.017	-0.028
100	50	RMSE	0.040	0.058	0.059	0.069	0.044	0.062	0.057	0.068
100	100	Bias	0.019	0.029	-0.011	-0.017	0.017	0.028	-0.009	-0.013
100	100	RMSE	0.024	0.033	0.031	0.036	0.024	0.034	0.034	0.036
100	200	Bias	0.009	0.009	-0.007	-0.005	0.008	0.009	-0.007	-0.005
100	200	RMSE	0.015	0.017	0.022	0.022	0.014	0.016	0.023	0.024
200	50	Bias	0.034	0.058	-0.021	-0.044	0.035	0.058	-0.010	-0.030
200	50	RMSE	0.038	0.062	0.043	0.060	0.040	0.061	0.040	0.053
200	100	Bias	0.018	0.030	-0.012	-0.021	0.017	0.030	-0.008	-0.014
200	100	RMSE	0.021	0.032	0.023	0.029	0.022	0.033	0.024	0.028
200	200	Bias	0.010	0.014	-0.004	-0.008	0.009	0.013	-0.003	-0.005
200	200	RMSE	0.013	0.016	0.016	0.017	0.013	0.016	0.016	0.018
			t_4 distribution							
100	50	Bias	0.032	0.049	-0.025	-0.042	0.037	0.055	-0.017	-0.031
100	50	RMSE	0.044	0.058	0.061	0.075	0.051	0.067	0.063	0.073
100	100	Bias	0.019	0.027	-0.013	-0.017	0.019	0.029	-0.015	-0.018
100	100	RMSE	0.025	0.032	0.033	0.039	0.028	0.036	0.039	0.042
100	200	Bias	0.010	0.010	-0.004	-0.003	0.012	0.012	-0.005	-0.002
100	200	RMSE	0.015	0.015	0.020	0.023	0.018	0.019	0.025	0.027
200	50	Bias	0.035	0.054	-0.025	-0.046	0.039	0.061	-0.021	-0.036
200	50	RMSE	0.040	0.058	0.046	0.064	0.046	0.065	0.051	0.082
200	100	Bias	0.018	0.028	-0.014	-0.020	0.019	0.029	-0.013	-0.021
200	100	RMSE	0.021	0.031	0.027	0.032	0.024	0.033	0.031	0.037
200	200	Bias	0.009	0.012	-0.007	-0.008	0.011	0.014	-0.005	-0.007
200	200	RMSE	0.012	0.015	0.018	0.019	0.015	0.017	0.020	0.022
			χ^2_3 distribution							
100	50	Bias	0.055	0.075	-0.023	-0.035	0.029	0.041	-0.011	-0.015
100	50	RMSE	0.076	0.092	0.102	0.111	0.047	0.058	0.074	0.084
100	100	Bias	0.031	0.037	-0.009	-0.015	0.018	0.017	0.001	-0.007
100	100	RMSE	0.045	0.049	0.061	0.059	0.028	0.027	0.039	0.040
100	200	Bias	0.017	0.018	-0.003	-0.003	0.009	0.006	-0.001	0.000
100	200	RMSE	0.030	0.030	0.046	0.044	0.017	0.016	0.029	0.027
200	50	Bias	0.051	0.072	-0.034	-0.050	0.028	0.039	-0.015	-0.026
200	50	RMSE	0.063	0.082	0.077	0.088	0.038	0.049	0.053	0.061
200	100	Bias	0.029	0.038	-0.015	-0.022	0.015	0.019	-0.006	-0.008
200	100	RMSE	0.038	0.044	0.046	0.049	0.022	0.026	0.031	0.034
200	200	Bias	0.017	0.020	-0.005	-0.007	0.009	0.009	-0.001	-0.002
200	200	RMSE	0.023	0.026	0.029	0.030	0.014	0.014	0.020	0.021

TABLE S.14. *Bias and root mean square error (RMSE) of the unfeasible QMG estimator and the feasible QMG estimator for λ and θ in Design 1. In all the variations of the model, $\lambda = 0.5$. IQMG denotes the unfeasible version of the estimator.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG
			Normal Distribution							
100	50	Bias	0.041	0.063	-0.003	-0.018	0.040	0.061	0.003	-0.007
100	50	RMSE	0.049	0.069	0.048	0.058	0.049	0.069	0.055	0.066
100	100	Bias	0.026	0.034	0.005	-0.001	0.024	0.033	0.008	0.004
100	100	RMSE	0.031	0.039	0.031	0.033	0.030	0.039	0.032	0.036
100	200	Bias	-0.008	-0.007	-0.041	-0.039	-0.008	-0.007	-0.040	-0.038
100	200	RMSE	0.014	0.015	0.046	0.045	0.014	0.015	0.046	0.044
200	50	Bias	0.040	0.063	-0.009	-0.034	0.040	0.063	-0.001	-0.021
200	50	RMSE	0.044	0.066	0.039	0.053	0.044	0.066	0.037	0.048
200	100	Bias	0.026	0.038	0.006	-0.003	0.025	0.037	0.009	0.002
200	100	RMSE	0.028	0.039	0.021	0.023	0.028	0.039	0.026	0.026
200	200	Bias	0.008	0.012	-0.007	-0.011	0.007	0.010	-0.007	-0.009
200	200	RMSE	0.011	0.014	0.016	0.019	0.011	0.014	0.017	0.019
			t_4 distribution							
100	50	Bias	0.045	0.060	-0.001	-0.019	0.049	0.066	0.007	-0.008
100	50	RMSE	0.053	0.068	0.054	0.065	0.060	0.076	0.061	0.067
100	100	Bias	0.032	0.039	0.014	0.011	0.032	0.042	0.013	0.009
100	100	RMSE	0.036	0.043	0.033	0.036	0.038	0.047	0.038	0.039
100	200	Bias	-0.006	-0.006	-0.035	-0.033	-0.004	-0.003	-0.035	-0.032
100	200	RMSE	0.012	0.013	0.040	0.040	0.014	0.015	0.044	0.042
200	50	Bias	0.029	0.047	-0.037	-0.058	0.033	0.054	-0.034	-0.070
200	50	RMSE	0.035	0.052	0.053	0.072	0.041	0.059	0.058	0.306
200	100	Bias	0.016	0.025	-0.018	-0.023	0.017	0.027	-0.017	-0.025
200	100	RMSE	0.020	0.029	0.029	0.034	0.023	0.031	0.033	0.039
200	200	Bias	-0.014	-0.011	-0.052	-0.053	-0.012	-0.009	-0.051	-0.053
200	200	RMSE	0.016	0.014	0.055	0.056	0.016	0.014	0.054	0.056
			χ^2_3 distribution							
100	50	Bias	0.067	0.087	-0.003	-0.009	0.041	0.053	0.013	0.008
100	50	RMSE	0.084	0.102	0.101	0.123	0.055	0.066	0.073	0.082
100	100	Bias	0.028	0.034	-0.015	-0.020	0.015	0.015	-0.005	-0.012
100	100	RMSE	0.043	0.046	0.062	0.061	0.027	0.026	0.040	0.042
100	200	Bias	0.012	0.013	-0.012	-0.012	0.003	0.001	-0.011	-0.010
100	200	RMSE	0.027	0.027	0.048	0.046	0.016	0.015	0.031	0.029
200	50	Bias	0.059	0.076	-0.020	-0.036	0.035	0.045	0.000	-0.012
200	50	RMSE	0.068	0.084	0.070	0.109	0.042	0.053	0.046	0.055
200	100	Bias	0.031	0.041	-0.009	-0.018	0.017	0.021	0.000	-0.003
200	100	RMSE	0.039	0.047	0.047	0.048	0.022	0.026	0.027	0.028
200	200	Bias	0.034	0.039	0.032	0.030	0.027	0.028	0.036	0.035
200	200	RMSE	0.037	0.042	0.043	0.043	0.029	0.030	0.041	0.040

TABLE S.15. *Bias and root mean square error (RMSE) of the unfeasible QMG estimator and the feasible QMG estimator for λ and θ in Design 2. In all the variations of the model, $\lambda = 0.5$. IQMG denotes the unfeasible version of the estimator.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG
						Normal Distribution				
100	50	Bias	0.039	0.058	-0.023	-0.048	0.042	0.066	-0.004	-0.011
100	50	RMSE	0.048	0.066	0.060	0.076	0.051	0.073	0.062	0.067
100	100	Bias	0.023	0.032	-0.007	-0.015	0.024	0.035	0.001	0.000
100	100	RMSE	0.028	0.037	0.032	0.036	0.030	0.041	0.034	0.039
100	200	Bias	0.011	0.013	-0.004	-0.005	0.011	0.016	-0.002	0.005
100	200	RMSE	0.017	0.018	0.022	0.022	0.018	0.021	0.023	0.024
200	50	Bias	0.039	0.065	-0.026	-0.050	0.044	0.074	-0.003	-0.012
200	50	RMSE	0.044	0.068	0.044	0.063	0.049	0.077	0.040	0.047
200	100	Bias	0.021	0.033	-0.013	-0.024	0.023	0.038	-0.003	-0.006
200	100	RMSE	0.024	0.036	0.027	0.038	0.027	0.040	0.026	0.027
200	200	Bias	0.010	0.014	-0.007	-0.010	0.011	0.017	-0.001	0.000
200	200	RMSE	0.013	0.017	0.017	0.019	0.014	0.020	0.016	0.016
						t_4 distribution				
100	50	Bias	0.037	0.054	-0.032	-0.056	0.044	0.069	-0.012	-0.005
100	50	RMSE	0.049	0.063	0.067	0.083	0.057	0.079	0.069	0.223
100	100	Bias	0.022	0.029	-0.010	-0.017	0.028	0.040	-0.001	0.002
100	100	RMSE	0.028	0.035	0.036	0.042	0.034	0.046	0.035	0.041
100	200	Bias	0.012	0.012	-0.004	-0.005	0.015	0.019	0.002	0.008
100	200	RMSE	0.017	0.018	0.023	0.024	0.019	0.024	0.022	0.025
200	50	Bias	0.036	0.058	-0.033	-0.051	0.048	0.074	-0.010	-0.014
200	50	RMSE	0.041	0.063	0.051	0.068	0.054	0.079	0.049	0.080
200	100	Bias	0.019	0.030	-0.017	-0.025	0.025	0.039	-0.006	-0.008
200	100	RMSE	0.023	0.033	0.028	0.035	0.029	0.042	0.028	0.029
200	200	Bias	0.010	0.014	-0.007	-0.009	0.014	0.019	0.000	0.000
200	200	RMSE	0.013	0.016	0.018	0.020	0.017	0.022	0.018	0.020
						χ^2_3 distribution				
100	50	Bias	0.060	0.082	-0.036	-0.055	0.041	0.060	0.010	0.008
100	50	RMSE	0.082	0.100	0.116	0.126	0.055	0.072	0.076	0.080
100	100	Bias	0.037	0.044	-0.016	-0.026	0.018	0.025	-0.002	0.003
100	100	RMSE	0.049	0.055	0.067	0.069	0.030	0.036	0.045	0.050
100	200	Bias	0.016	0.017	-0.010	-0.015	0.009	0.010	-0.002	0.002
100	200	RMSE	0.028	0.028	0.044	0.046	0.017	0.018	0.028	0.028
200	50	Bias	0.059	0.082	-0.041	-0.063	0.038	0.057	0.002	0.001
200	50	RMSE	0.068	0.089	0.076	0.090	0.046	0.064	0.047	0.052
200	100	Bias	0.031	0.041	-0.022	-0.032	0.017	0.024	-0.002	-0.002
200	100	RMSE	0.038	0.047	0.048	0.054	0.023	0.029	0.031	0.033
200	200	Bias	0.016	0.020	-0.011	-0.015	0.010	0.012	0.000	0.002
200	200	RMSE	0.024	0.026	0.034	0.035	0.015	0.017	0.021	0.022

TABLE S.16. *Bias and root mean square error (RMSE) of the unfeasible QMG estimator and the feasible QMG estimator for λ and θ in Design 3. In all the variations of the model, $\lambda = 0.5$. IQMG denotes the unfeasible version of the estimator.*

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ		Parameter: λ		Parameter: θ	
			IQMG	QMG	IQMG	QMG	IQMG	QMG	IQMG	QMG
			Normal Distribution							
100	50	Bias	0.045	0.067	-0.012	-0.040	0.049	0.076	0.009	0.001
100	50	RMSE	0.052	0.071	0.052	0.065	0.056	0.081	0.053	0.061
100	100	Bias	0.016	0.026	-0.020	-0.023	0.021	0.033	-0.006	-0.005
100	100	RMSE	0.023	0.032	0.036	0.042	0.027	0.039	0.036	0.036
100	200	Bias	0.007	0.010	-0.009	-0.008	0.010	0.014	-0.004	0.003
100	200	RMSE	0.013	0.016	0.022	0.024	0.016	0.020	0.023	0.024
200	50	Bias	0.043	0.068	-0.013	-0.038	0.049	0.077	0.007	-0.005
200	50	RMSE	0.047	0.071	0.038	0.057	0.052	0.080	0.036	0.044
200	100	Bias	0.013	0.025	-0.027	-0.037	0.015	0.029	-0.017	-0.021
200	100	RMSE	0.017	0.028	0.035	0.043	0.020	0.032	0.030	0.033
200	200	Bias	0.009	0.013	-0.009	-0.012	0.011	0.017	-0.003	-0.001
200	200	RMSE	0.012	0.016	0.018	0.020	0.014	0.020	0.016	0.016
			t_4 distribution							
100	50	Bias	0.048	0.066	-0.009	-0.033	0.061	0.087	0.016	0.007
100	50	RMSE	0.055	0.074	0.055	0.071	0.069	0.094	0.070	0.076
100	100	Bias	0.014	0.022	-0.023	-0.028	0.020	0.033	-0.015	-0.011
100	100	RMSE	0.023	0.029	0.041	0.045	0.028	0.040	0.038	0.041
100	200	Bias	0.019	0.019	0.012	0.011	0.022	0.027	0.017	0.024
100	200	RMSE	0.022	0.023	0.025	0.026	0.026	0.030	0.028	0.033
200	50	Bias	0.030	0.050	-0.046	-0.067	0.043	0.070	-0.017	-0.030
200	50	RMSE	0.036	0.055	0.060	0.079	0.048	0.073	0.050	0.057
200	100	Bias	0.023	0.034	-0.007	-0.014	0.027	0.040	0.002	0.000
200	100	RMSE	0.027	0.037	0.025	0.030	0.032	0.044	0.029	0.031
200	200	Bias	0.011	0.015	-0.005	-0.007	0.014	0.020	-0.001	0.001
200	200	RMSE	0.014	0.018	0.016	0.019	0.017	0.023	0.018	0.018
			χ^2_3 distribution							
100	50	Bias	0.073	0.100	-0.023	-0.041	0.050	0.072	0.022	0.026
100	50	RMSE	0.089	0.114	0.098	0.104	0.062	0.083	0.071	0.077
100	100	Bias	0.074	0.082	0.060	0.051	0.058	0.064	0.076	0.080
100	100	RMSE	0.083	0.089	0.089	0.081	0.063	0.069	0.088	0.092
100	200	Bias	0.058	0.059	0.072	0.068	0.048	0.050	0.079	0.084
100	200	RMSE	0.062	0.063	0.084	0.080	0.051	0.053	0.084	0.089
200	50	Bias	0.066	0.091	-0.024	-0.049	0.045	0.063	0.015	0.012
200	50	RMSE	0.077	0.099	0.085	0.089	0.053	0.070	0.055	0.060
200	100	Bias	0.038	0.049	-0.008	-0.018	0.024	0.031	0.012	0.011
200	100	RMSE	0.043	0.054	0.042	0.046	0.029	0.035	0.033	0.033
200	200	Bias	0.025	0.029	0.007	0.002	0.018	0.021	0.017	0.019
200	200	RMSE	0.029	0.033	0.028	0.028	0.021	0.023	0.026	0.027

TABLE S.17. *Bias and root mean square error (RMSE) of the unfeasible QMG estimator and the feasible QMG estimator for λ and θ in Design 4. In all the variations of the model, $\lambda = 0.5$. IQMG denotes the unfeasible version of the estimator.*

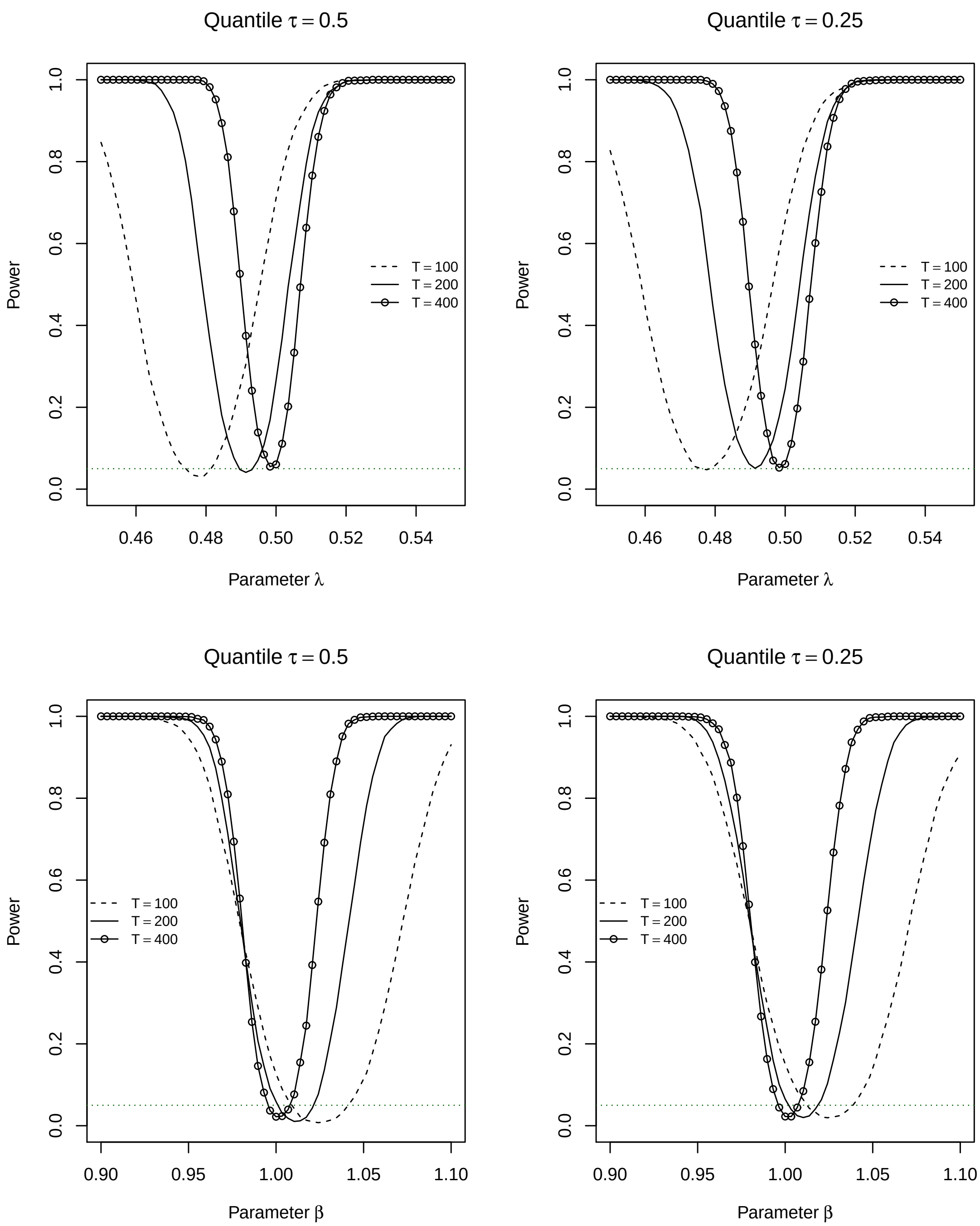

Figure S.3. *Power of the QMG estimator against different alternatives. The figures shows results when* $u \sim \mathcal{N}(0, 1)$.

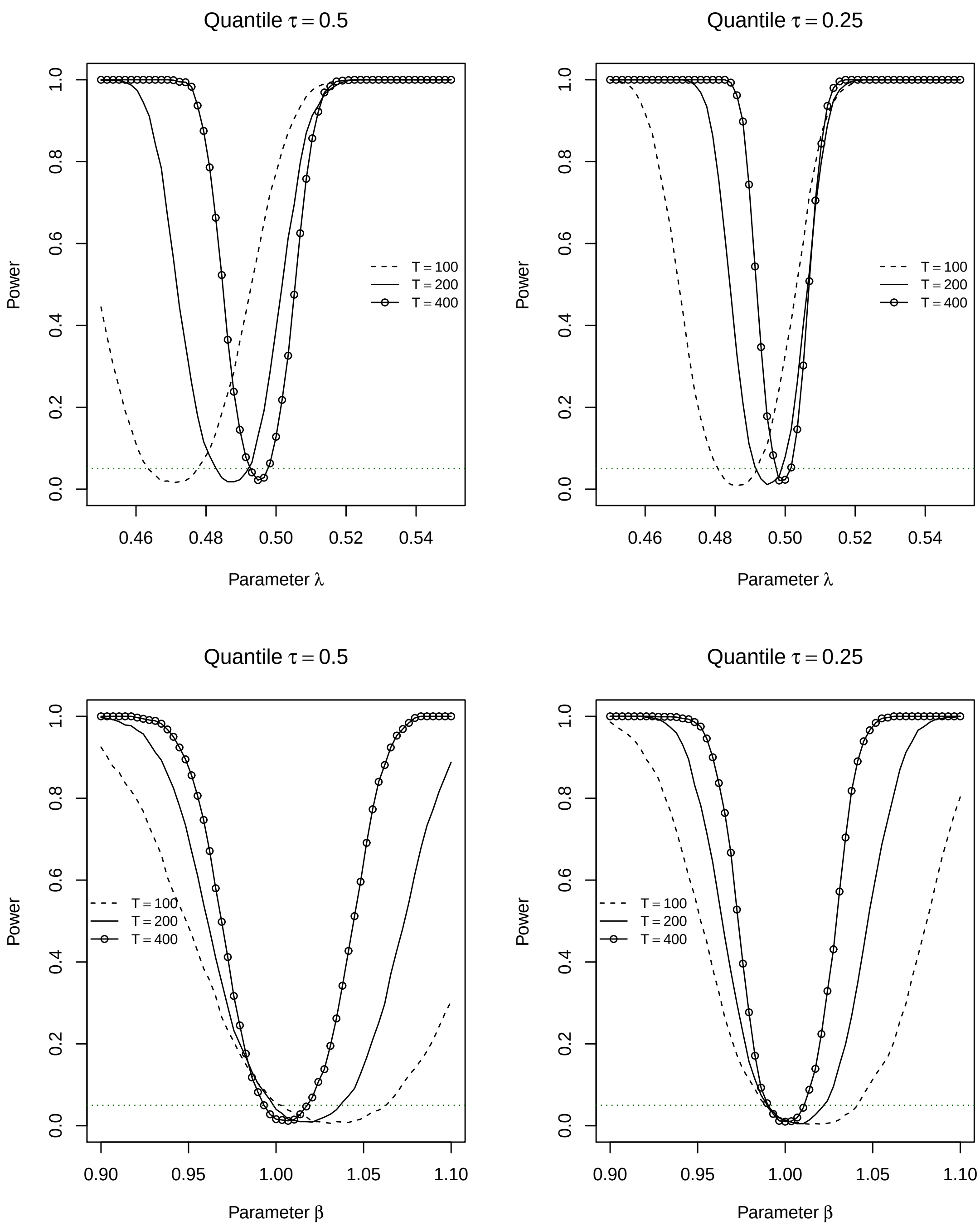

FIGURE S.4. *Power of the QMG estimator against different alternatives. The figures shows results when* $u \sim \chi^2_3$.

	QMG					FE	CCEMG
	0.10	0.25	0.50	0.75	0.90		
			Control Group				
Consumption at	0.466	0.576	0.620	0.479	0.353	0.623	0.476
$t-1$ (in logs)	(0.020)	(0.021)	(0.021)	(0.020)	(0.015)	(0.001)	(0.009)
Treatment	0.148	0.112	0.063	0.049	0.048	0.145	0.095
(2pm - 7pm)	(0.011)	(0.009)	(0.006)	(0.007)	(0.006)	(0.001)	(0.018)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	242	242	242	242	242	242	242
$N \times T$	2090638	2090638	2090638	2090638	2090638	2090638	2090638
			Portal				
Consumption at	0.469	0.588	0.631	0.485	0.360	0.622	0.489
$t-1$ (in logs)	(0.021)	(0.023)	(0.024)	(0.022)	(0.015)	(0.001)	(0.009)
Treatment	0.089	0.065	0.038	0.023	0.007	0.102	0.048
(2pm - 7pm)	(0.013)	(0.012)	(0.010)	(0.011)	(0.014)	(0.001)	(0.017)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	189	189	189	189	189	189	189
$N \times T$	1632771	1632771	1632771	1632771	1632771	1632771	1632771
			IHD				
Consumption at	0.470	0.581	0.615	0.475	0.352	0.627	0.480
$t-1$ (in logs)	(0.022)	(0.025)	(0.027)	(0.025)	(0.018)	(0.001)	(0.009)
Treatment	0.102	0.072	0.038	0.025	0.006	0.089	0.046
(2pm - 7pm)	(0.017)	(0.013)	(0.010)	(0.009)	(0.011)	(0.002)	(0.017)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	152	152	152	152	152	152	152
$N \times T$	1313128	1313128	1313128	1313128	1313128	1313128	1313128
			PCT				
Consumption at	0.718	0.785	0.806	0.693	0.560	0.771	0.682
$t-1$ (in logs)	(0.024)	(0.020)	(0.019)	(0.022)	(0.021)	(0.000)	(0.007)
Treatment	-0.092	-0.054	-0.029	-0.030	-0.029	-0.010	-0.073
(2pm - 7pm)	(0.022)	(0.016)	(0.010)	(0.010)	(0.014)	(0.001)	(0.014)
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	196	196	196	196	196	196	196
$N \times T$	1693244	1693244	1693244	1693244	1693244	1693244	1693244

TABLE S.18. *Quantile Mean Group estimator results for the control group and different technologies. FE denotes fixed effects and CCEMG denotes the Common Correlated Mean Group estimator due to Chudik and Pesaran (2015). IHD denotes in-home display and PCT is programmable communicating thermostats. Standard errors are in parentheses.*

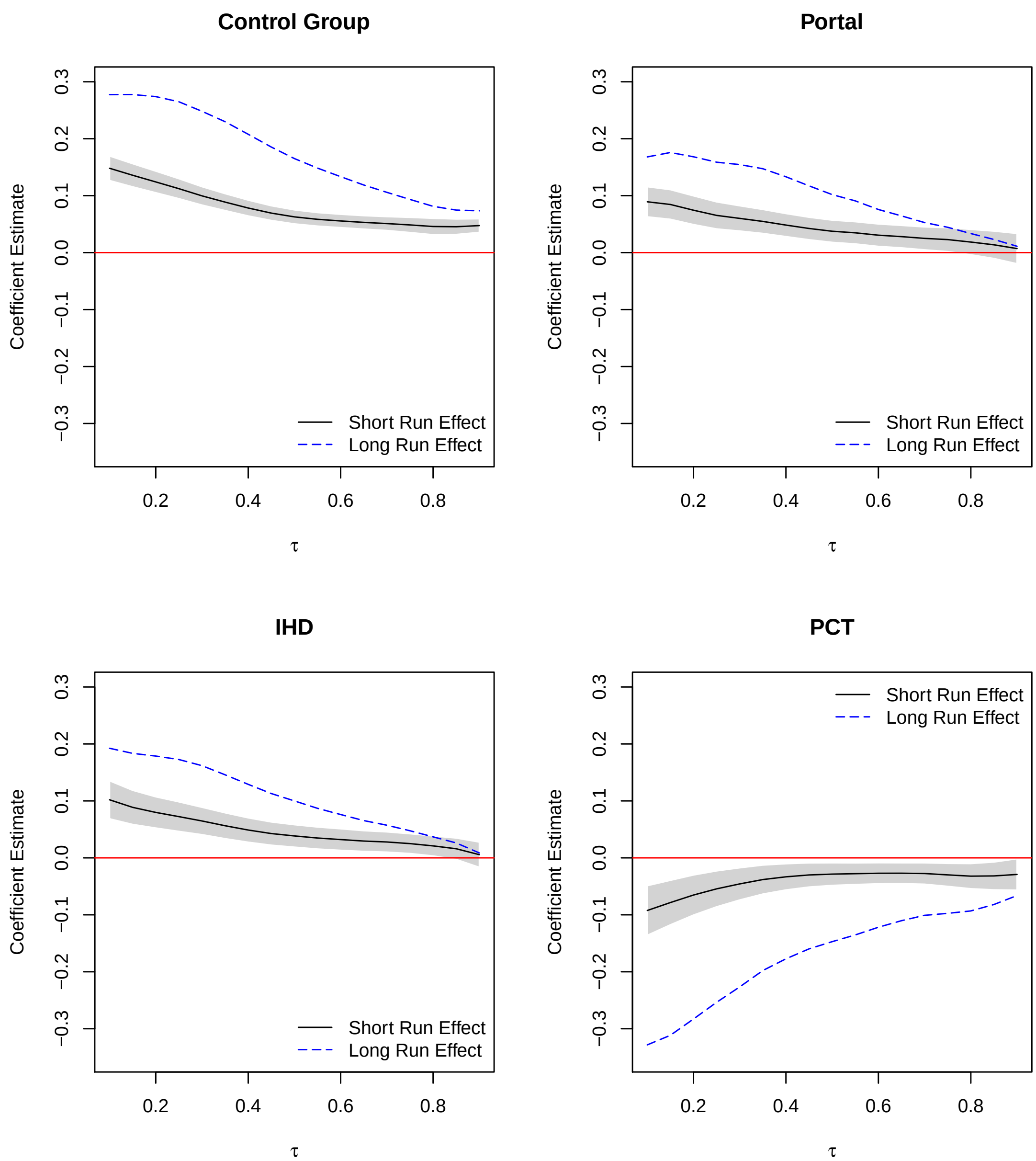

FIGURE S.5. *Short and Long Run Quantile Regression Results. The figure shows the QTE coefficient $\delta_g(\tau)$ for the control group, portal group, in-home-device (IHD), and programmable communicating thermostats (PCT). The grey area denotes a 95 percent point-wise confidence interval.*

		QMG					CCEMG
		0.10	0.25	0.50	0.75	0.90	
				Control Group			
High Income	Consumption at	0.470	0.583	0.624	0.492	0.368	0.484
	$t-1$ (in logs)	(0.032)	(0.035)	(0.036)	(0.034)	(0.025)	(0.009)
	Treatment	0.164	0.132	0.075	0.061	0.058	0.110
	(2pm - 7pm)	(0.017)	(0.015)	(0.010)	(0.009)	(0.009)	(0.016)
Low Income	Consumption at	0.463	0.572	0.617	0.472	0.344	0.472
	$t-1$ (in logs)	(0.025)	(0.026)	(0.026)	(0.024)	(0.018)	(0.009)
	Treatment	0.139	0.100	0.055	0.041	0.041	0.085
	(2pm - 7pm)	(0.014)	(0.011)	(0.008)	(0.009)	(0.009)	(0.019)
				Portal			
High Income	Consumption at	0.463	0.582	0.619	0.480	0.351	0.483
	$t-1$ (in logs)	(0.029)	(0.031)	(0.033)	(0.030)	(0.020)	(0.009)
	Treatment	0.083	0.061	0.035	0.016	-0.003	0.040
	(2pm - 7pm)	(0.018)	(0.018)	(0.015)	(0.017)	(0.022)	(0.017)
Low Income	Consumption at	0.476	0.597	0.648	0.492	0.371	0.496
	$t-1$ (in logs)	(0.030)	(0.033)	(0.035)	(0.032)	(0.024)	(0.009)
	Treatment	0.098	0.072	0.042	0.032	0.021	0.060
	(2pm - 7pm)	(0.020)	(0.016)	(0.012)	(0.009)	(0.012)	(0.018)
				IHD			
High Income	Consumption at	0.470	0.582	0.609	0.481	0.364	0.488
	$t-1$ (in logs)	(0.032)	(0.036)	(0.038)	(0.037)	(0.025)	(0.009)
	Treatment	0.092	0.062	0.025	0.011	-0.015	0.028
	(2pm - 7pm)	(0.023)	(0.021)	(0.017)	(0.015)	(0.020)	(0.017)
Low Income	Consumption at	0.471	0.579	0.621	0.469	0.340	0.472
	$t-1$ (in logs)	(0.031)	(0.035)	(0.038)	(0.034)	(0.027)	(0.009)
	Treatment	0.112	0.083	0.052	0.039	0.026	0.065
	(2pm - 7pm)	(0.025)	(0.017)	(0.011)	(0.011)	(0.011)	(0.017)
				PCT			
High Income	Consumption at	0.709	0.787	0.810	0.685	0.535	0.676
	$t-1$ (in logs)	(0.033)	(0.027)	(0.024)	(0.028)	(0.026)	(0.008)
	Treatment	-0.103	-0.064	-0.036	-0.044	-0.044	-0.085
	(2pm - 7pm)	(0.033)	(0.025)	(0.016)	(0.018)	(0.023)	(0.014)
Low Income	Consumption at	0.727	0.784	0.802	0.701	0.584	0.687
	$t-1$ (in logs)	(0.036)	(0.030)	(0.029)	(0.033)	(0.032)	(0.007)
	Treatment	-0.082	-0.045	-0.021	-0.016	-0.014	-0.061
	(2pm - 7pm)	(0.031)	(0.022)	(0.013)	(0.012)	(0.017)	(0.015)

TABLE S.19. *Quantile Mean Group estimator results by Income Levels. CCEMG denotes the Common Correlated Mean Group estimator due to Chudik and Pesaran (2015). IHD denotes in-home display and PCT is programmable communicating thermostats. Standard errors are in parentheses.*

		QMG					CEEMG
		0.10	0.25	0.50	0.75	0.90	
				Control Group			
Family years	Consumption at	0.493	0.618	0.666	0.515	0.369	0.507
	$t-1$ (in logs)	(0.027)	(0.029)	(0.029)	(0.028)	(0.021)	(0.009)
	Treatment	0.161	0.119	0.062	0.050	0.041	0.099
	(2pm - 7pm)	(0.015)	(0.012)	(0.008)	(0.008)	(0.008)	(0.017)
Young years	Consumption at	0.439	0.535	0.575	0.444	0.337	0.447
	$t-1$ (in logs)	(0.029)	(0.029)	(0.030)	(0.027)	(0.020)	(0.009)
	Treatment	0.135	0.106	0.063	0.047	0.053	0.090
	(2pm - 7pm)	(0.016)	(0.013)	(0.010)	(0.011)	(0.009)	(0.019)
				Portal			
Family years	Consumption at	0.471	0.602	0.654	0.505	0.364	0.502
	$t-1$ (in logs)	(0.027)	(0.030)	(0.032)	(0.029)	(0.020)	(0.009)
	Treatment	0.109	0.086	0.053	0.043	0.032	0.067
	(2pm - 7pm)	(0.017)	(0.017)	(0.015)	(0.016)	(0.021)	(0.017)
Young years	Consumption at	0.467	0.578	0.615	0.471	0.357	0.479
	$t-1$ (in logs)	(0.017)	(0.014)	(0.013)	(0.014)	(0.013)	(0.009)
	Treatment	0.075	0.051	0.026	0.009	-0.011	0.035
	(2pm - 7pm)	(0.028)	(0.019)	(0.014)	(0.015)	(0.020)	(0.018)
				IHD			
Family years	Consumption at	0.508	0.647	0.696	0.540	0.384	0.535
	$t-1$ (in logs)	(0.029)	(0.030)	(0.031)	(0.032)	(0.025)	(0.009)
	Treatment	0.084	0.056	0.028	0.014	0.001	0.033
	(2pm - 7pm)	(0.025)	(0.020)	(0.013)	(0.011)	(0.017)	(0.016)
Young years	Consumption at	0.440	0.527	0.550	0.422	0.327	0.435
	$t-1$ (in logs)	(0.032)	(0.037)	(0.040)	(0.037)	(0.026)	(0.009)
	Treatment	0.117	0.086	0.047	0.034	0.009	0.058
	(2pm - 7pm)	(0.024)	(0.018)	(0.015)	(0.013)	(0.015)	(0.018)
				PCT			
Family years	Consumption at	0.719	0.780	0.801	0.689	0.556	0.679
	$t-1$ (in logs)	(0.037)	(0.030)	(0.026)	(0.032)	(0.032)	(0.007)
	Treatment	-0.078	-0.042	-0.020	-0.018	-0.003	-0.055
	(2pm - 7pm)	(0.036)	(0.025)	(0.015)	(0.016)	(0.022)	(0.014)
Young years	Consumption at	0.718	0.789	0.809	0.696	0.562	0.683
	$t-1$ (in logs)	(0.032)	(0.029)	(0.027)	(0.030)	(0.027)	(0.007)
	Treatment	-0.101	-0.062	-0.034	-0.037	-0.044	-0.083
	(2pm - 7pm)	(0.028)	(0.021)	(0.013)	(0.013)	(0.018)	(0.015)

TABLE S.20. *Quantile Mean Group estimator results by Income Levels. CCEMG denotes the Common Correlated Mean Group estimator due to Chudik and Pesaran (2015). IHD denotes in-home display and PCT is programmable communicating thermostats. Standard errors are in parentheses.*

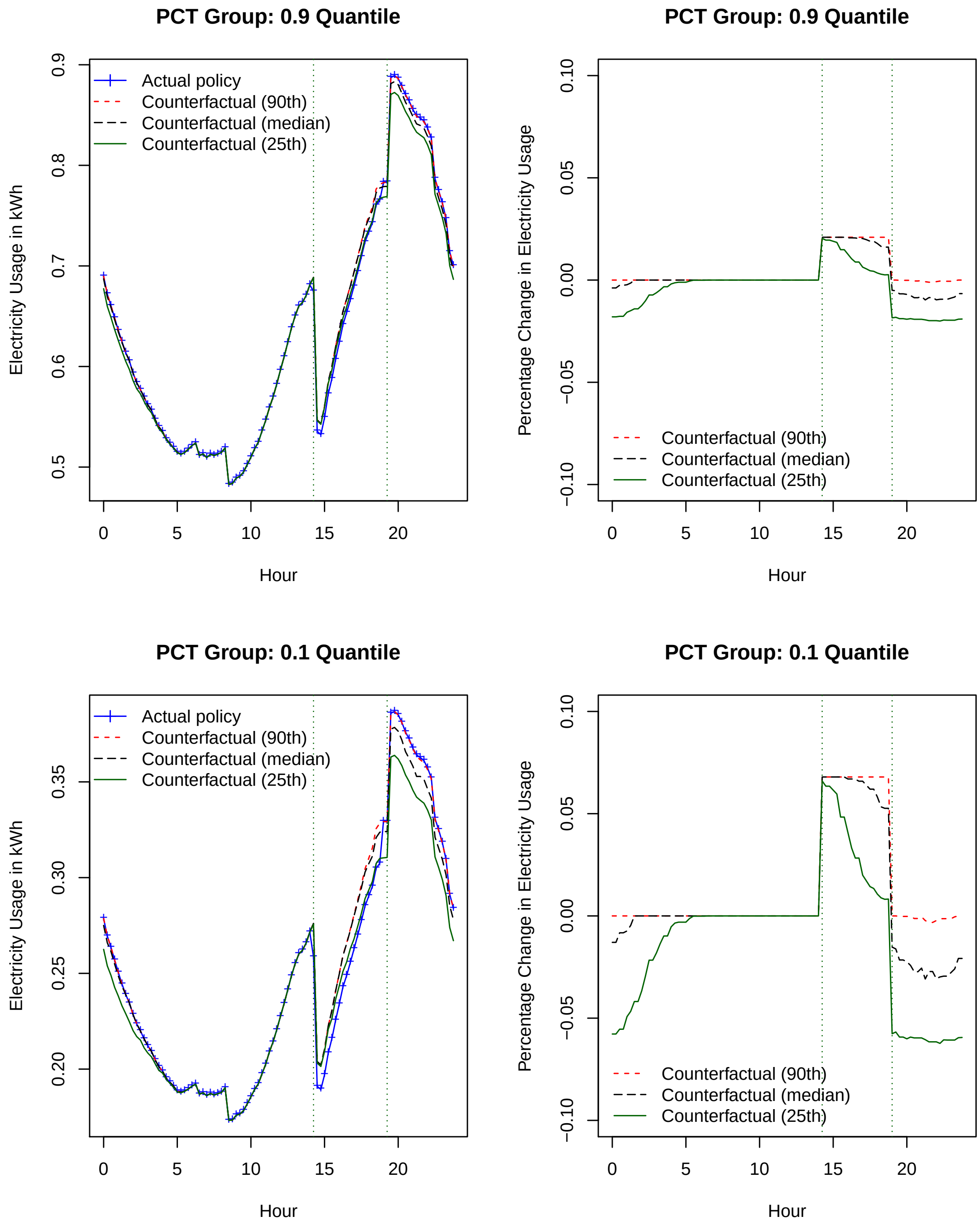

FIGURE S.6. *Counterfactual policies for customers with a PCT. The right panels show the percentage change in electricity usage with respect to the actual policy.*