Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Quantile Regression Models[*]

Matthew Harding[†], Carlos Lamarche[‡], and M. Hashem Pesaran[§]

December 17, 2019

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. The new quantile regression estimator is shown to be consistent and its asymptotic distribution is derived. 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: C21, C31, C33, D12, L94

Keywords: Common Correlated Effects; Dynamic Panel; Quantile Regression; Smart Meter; Randomized Experiment

---

[*]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, University of Waterloo, Central Bank of Argentina, the 2017 International Association for Applied Econometrics conference, and the 2017 Boneyard Econometrics conference in celebration of Professor Roger Koenker.

[†]Department of Economics, University of California at 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`

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. 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, Fernández-Val, Hahn and Newey (2013), Chernozhukov, Fernández-Val, Hoderlein, Holzmann and Newey (2015), Harding and Lamarche (2014, 2017), and 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 (2019) 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 focus on estimation and inference of mean quantile coefficients. We allow for the possibility that unobserved factors and included regressors are correlated and study the conditions under which the slope coefficients are estimated consistently. The proposed estimator is generally applicable, but requires an important rank condition which is met only if the number of unobserved factors minus 1 is less than or equal to the number of exogenous variables. In practice, this condition implies that practitioners can identify and estimate a model with many factors if a sufficiently large number of variables are included in the model. Another important condition, which is similar to a condition used in Chudik and Pesaran (2015), is that a sufficient number of lagged values of cross section averages are need to approximate the unobserved factors. 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 proposed estimator can be used in practice by exploring an application of electricity pricing and smart technology. Section 5 concludes. Mathematical proofs and additional Monte Carlo results are offered in an Online Appendix.

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, $\xrightarrow{p}$ 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

As it will be clear below, the estimation of a dynamic quantile regression model with factors is challenging, because these factors are not observed by the researcher and could be correlated with other included regressors in the model. Furthermore, the autoregressive nature of the quantile model limits practitioners from employing the CCE approach of Pesaran (2006) and Chudik and Pesaran (2015). To fill this gap in the literature, this section proposes estimation strategies based on individual quantile regression models augmented by cross section averages of observable variables and their lags. While the proposed estimation method is simple to implement in practice, they require large N and T quantile models that include several time varying regressors possibly correlated with the latent factors.

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 the lagged dependent variable:

$$y_{it} = \alpha_i + \lambda_i y_{it-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \boldsymbol{\gamma}_i' \mathbf{f}_t + \xi_{it}. \tag{2.1}$$

We assume that y_{it} started a long time ago. The variable $\mathbf{x}_{it}$ is a $p_x \times 1$ vector of regressors specific to cross-sectional unit i at time t, $\boldsymbol{\beta}_i$ is the vector of associated regression coefficients, $\mathbf{f}_t$ is an $r \times 1$ vector of unobserved factors, $\boldsymbol{\gamma}_i$ is a vector of latent factor loadings, and α_i is an individual effect potentially correlated with the exogenous regressors, $\mathbf{x}_{it}$. The error term is ξ_{it}.

If a practitioner estimates the model assuming homogeneous parameters (i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for $1 \leq i \leq N$), the estimates are not consistent because the method does not correct for the bias arising from a conjunction of dynamics, heterogeneity and cross-sectional dependence. Furthermore, the asymptotic variance of the estimator is misspecified too. Pesaran and Smith (1995) provide expressions for the bias of the homogeneous fixed effects estimates for a linear heterogeneous dynamic model and found that the bias can be substantial.

The associated conditional panel quantile function is given by

$$Q_{Y_{it}}(\tau | y_{it-1}, \mathbf{x}_{it}, \boldsymbol{\theta}_i(\tau), \mathbf{f}_t) = \alpha_i(\tau) + \lambda_i(\tau) y_{it-1} + \mathbf{x}_{it}' \boldsymbol{\beta}_i(\tau) + \mathbf{f}_t' \boldsymbol{\gamma}_i(\tau), \tag{2.2}$$

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 $Q_{Y_{it}}(\tau | \mathfrak{F}_{it}) := \inf\{y : P(Y_{it} \leq y | \mathfrak{F}_{it}) \geq \tau\}$, where $\mathfrak{F}_{it} = (y_{it-1}, \mathbf{x}_{it}', \boldsymbol{\theta}_i(\tau)', \mathbf{f}_t')'$. The term $\mathbf{f}_t' \boldsymbol{\gamma}_i(\tau)$ can be interpreted as a quantile-specific function capturing unobserved heterogeneity, not adequately controlled by the inclusion of $\mathbf{x}_{it}$. The quantile model can be considered to be semi-parametric since the functional form of the conditional distribution of y_{it} given $\mathfrak{F}_{it}$ is left unspecified and no parametric assumption is imposed on the relation between the regressors and the latent variables in the model.

To relate the quantile function (2.2) to the underlying data generating process (2.1), we introduce the quantile and unit specific error term $u_{it}(\tau)$ defined by

$$u_{it}(\tau) = y_{it} - Q_{Y_{it}}(\tau | \mathfrak{F}_{it}). \tag{2.3}$$

From the definition of quantiles, it now readily follows that,

$$P(u_{it}(\tau) \leq 0 | \mathfrak{F}_{it}) = \tau. \tag{2.4}$$

Under (2.1) and using (2.3), it also follows that

$$u_{it}(\tau) = \xi_{it} + [\mu_{it} - Q_{Y_{it}}(\tau | \mathfrak{F}_{it})], \tag{2.5}$$

where $\mu_{it} = \alpha_i + \lambda_i y_{it-1} + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \boldsymbol{\gamma}_i' \mathbf{f}_t$ is the predictable component of (2.1), with respect to the information set $\mathfrak{F}_{it}$. Hence, conditional on $\mathfrak{F}_{it}$, $u_{it}(\tau)$ and ξ_{it} have the same properties over i and t. But it is clear that, in general, $E(u_{it}(\tau)) \neq 0$, even though $E(\xi_{it}) = 0$.

The $p_x \times 1$ vector of regressors is assumed to follow the general linear process

$$\mathbf{x}_{it} = \boldsymbol{\alpha}_{ix} + \boldsymbol{\Gamma}_i' \mathbf{f}_t + \boldsymbol{v}_{it}, \tag{2.6}$$

where $\boldsymbol{\alpha}_{ix}$ 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) or (2.2) can accommodate additional lags of the dependent variable, time-invariant covariates, and lags of the exogenous covariates. These variations can be incorporated at a cost of additional notational complexity.

Due to the combination of cross-sectional error dependence ($\boldsymbol{\gamma}_i \neq \mathbf{0}$), and dynamics ($\lambda_i \neq 0$) in equation (2.2), 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.

2.1. **Estimation**

We consider consistent estimation of the parameters of interest by estimating the dynamic quantile regression model with interactive effects defined by (2.2). To this end, we make the the following assumptions:

Assumption 1. *The error terms ξ_{it} for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$ in equation* (2.1) *have mean zero and are serially and cross-sectionally independent conditional on $\mathfrak{F}_{it}$.*

Remark 1. It follows from (2.3) that under Assumption 1, the quantile specific errors, $u_{it}(\tau)$, are also serially and cross-sectionally independent, but in general, $u_{it}(\tau)$ need not have mean zero.

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 ξ_{it} and $\boldsymbol{v}_{it}$ for all i and t.*

Assumption 3. *The factor loadings $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_{\gamma i}$ and $vec[\boldsymbol{\Gamma}_i] = vec[\boldsymbol{\Gamma}] + \boldsymbol{\eta}_{\Gamma i}$ are distributed independently of ξ_{jt} and $\boldsymbol{v}_{jt}$ for all i, j and t with means $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$, and bounded variances. The error terms $\boldsymbol{\eta}_{\gamma i}$ and $\boldsymbol{\eta}_{\Gamma i}$ are distributed independently 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 regressors* $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, \ldots, x_{it,p_x})' \in \mathcal{X} \subseteq R^{p_x}$ *are generated according to equation* (2.6)*, and the vector of errors* $\boldsymbol{v}_{it}$ *in* (2.6) *follows a stationary process with mean zero, finite covariance matrix, and finite fourth order cumulants, and summable autocovariances (uniformly in* i*). The innovations* $\boldsymbol{v}_{it}$ *and* ξ_{it} *are independently distributed.*

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

$$\begin{aligned} \lambda_i &= \lambda + (1 - |\lambda|)\nu_{i\lambda} \\ \boldsymbol{\beta}_i &= \boldsymbol{\beta} + \nu_{i\beta}, \end{aligned} \tag{2.7}$$

where $\|\boldsymbol{\beta}\|_1 < K$*,* $\sup_i |\nu_{i\lambda}|$ *and* $|\lambda|$ *are bounded away from 1, and*

$$\boldsymbol{\nu}_i = \begin{pmatrix} \nu_{i\lambda} \\ \nu_{i\beta} \end{pmatrix} \sim IID\left(\mathbf{0}, \boldsymbol{\Omega}_{\vartheta}\right), \tag{2.8}$$

with $\|\boldsymbol{\Omega}_{\vartheta}\| < K$*,* $\boldsymbol{\Omega}_{\vartheta}$ *is a symmetric positive definite matrix. Furthermore,*

$$E(\lambda_i^l \alpha_i | \mathcal{F}_t) = a_l,\ E(\lambda_i^l \boldsymbol{\beta}_i | \mathcal{F}_t) = \mathbf{b}_l,\ E(\lambda_i^l \boldsymbol{\gamma}_i | \mathcal{F}_t) = \mathbf{c}_l, \tag{2.9}$$

for all i *and* $l = 0, 1, 2, \ldots.$ *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*,* $\mathbf{b}_l$ *and* $\mathbf{c}_l$ *are exponentially decaying in* l*, such that* $|a_l| < K_a\rho^l$*,* $\|\mathbf{b}_l\| < K_b\rho^l$*, and* $\|\mathbf{c}_l\| < K_c\rho^l$ *for some positive* $\rho < 1$*. The parameters* λ_i *and* $\boldsymbol{\beta}_i$ *are independently distributed over* i*, and* $\boldsymbol{\nu}_i$ *is independently distributed of* $\boldsymbol{\gamma}_i$*,* $\boldsymbol{\Gamma}_i$*,* ξ_{it}*,* $\mathbf{v}_{it}'$*, and* $\mathbf{f}_t$ *for all* i *and* t*.*

Assumption 6. *The* $(p_x + 1) \times r$ *matrix* $\mathbf{C} = E(\mathbf{C}_i) = (\boldsymbol{\gamma}, \boldsymbol{\Gamma})'$ *has full column rank.*

Assumption 1 is standard in panel data models and as seen from Remark 1, implies Assumption A3 in Ando and Bai (2017). The remaining assumptions are similar to those in Pesaran (2006) and Chudik and Pesaran (2015). Also, 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 $\boldsymbol{\vartheta}_i$ with respect to $\boldsymbol{\vartheta}$ are mean-zero random variables independently distributed of other variables in the model. Specification (2.7) ensures that $\sup_i |\lambda_i| < 1$, so long as $\sup_i |\nu_{i\lambda}| < 1$, and $|\lambda| < 1$. A convenient distribution for $\nu_{i\lambda}$ is a beta distribution defined on $(0, 1)$. The moment conditions in (2.9) 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

λ_i is independently distributed of $\alpha_i, \boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$ and $E\left(\lambda_i^l\right)$ decays exponentially in l. This last condition is met, for example, if λ_i 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 implies that $p_x \geq r - 1$, and ensures the large N representation of the unobserved factors. Equation (2.1), after pre-multiplying by $(1 - \lambda_i L)^{-1}$ where L is the lag operator, can be written as,

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

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

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

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)

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

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

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

Hence, overall

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

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

Similarly, taking cross-sectional averages of equation (2.6), we obtain,

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

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

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

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

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

Pre-multiplying both sides of (2.13) by $\mathbf{C}(L)'$ under Assumption 6, we obtain the following result for $\mathbf{f}_t$:

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

where $\mathbf{f}_0 = -(\mathbf{C}(1)'\mathbf{C}(1))^{-1}\mathbf{C}(1)'\mathbf{d}$ and $\mathbf{G}(L) = [\mathbf{C}(L)'\mathbf{C}(L)]^{-1}\mathbf{C}(L)'\boldsymbol{\Lambda}(L)$ is an $r \times (p_x + 1)$ distributed lag matrix.

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 the finite 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.

Let $\boldsymbol{\delta}_i(L) = \boldsymbol{\gamma}_i'\sum_{l=0}^{\infty}\mathbf{G}_l L^l = \sum_{l=0}^{\infty}\boldsymbol{\delta}_{il}L^l$, $\boldsymbol{\delta}_{il} = (\boldsymbol{\delta}_{iy,l}', \boldsymbol{\delta}_{ix,l}')'$, $\boldsymbol{\delta}_{iy,l}$ is a reduced form coefficient for the cross-sectional average of y_{it-l}, $\boldsymbol{\delta}_{ix,l}$ 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. Finally, substituting the representation of the factors in equation (2.1), we obtain

$$y_{it} = \beta_{i0} + \lambda_i y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta}_i + \sum_{l=0}^{p_T} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il} + \xi_{it} + h_{it,N}, \tag{2.15}$$

where $\beta_{i0} = \alpha_i + \boldsymbol{\gamma}_i'\mathbf{f}_0$, and

$$h_{it,N} = \sum_{l=p_T+1}^{\infty} \bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il} + O_p(N^{-1/2}). \tag{2.16}$$

The total number of unknown parameters for the first part of (2.15) which is augmented with cross-sectional averages and their lagged values is $(2 + p_x) + (p_x + 1)(p_T + 1)$. The second part of (2.15) includes ξ_{it}, a component due to the truncation of the underlying infinite order distributed lag function $\boldsymbol{\delta}_i(L)$, and an $O_p(N^{-1/2})$ term associated with approximating $\mathbf{f}_t$ with cross-section averages. 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.

Remark 2. 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 $\beta_{0i} = \alpha_i$.

Define $\mathbf{X}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \bar{\mathbf{z}}_t', \bar{\mathbf{z}}_{t-1}', ..., \bar{\mathbf{z}}_{t-p_T}')'$, $\boldsymbol{\pi}_i = (\lambda_i, \boldsymbol{\beta}_i', \alpha_i, \boldsymbol{\delta}_i')'$, $\boldsymbol{\delta}_i = (\boldsymbol{\delta}_{i1}', \boldsymbol{\delta}_{i2}', \ldots, \boldsymbol{\delta}_{ip_T}')'$. Consider now equation (2.15), rewrite it more compactly as,

$$y_{it} = \mathbf{X}_{it}'\boldsymbol{\pi}_i + h_{it,N} + \xi_{it}, \tag{2.17}$$

and consider the following two optimization problems:

$$\hat{\boldsymbol{\theta}}_i(\tau) = \arg\min_{\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_{it} - \mathbf{W}_{it}'\boldsymbol{\theta}_i), \tag{2.18}$$

where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the standard quantile regression loss function, $\boldsymbol{\theta}_i = (\lambda_i, \boldsymbol{\beta}_i', \alpha_i, \boldsymbol{\gamma}_i')'$, $\boldsymbol{\Theta}_i$ is a compact set in $\mathbb{R}^{2+p_x+r}$, $\mathbf{W}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \mathbf{f}_t')'$ depends on the latent factor $\mathbf{f}_t$ following equation (2.1), and

$$\hat{\boldsymbol{\pi}}_i(\tau) = \arg\min_{\boldsymbol{\pi}_i \in \boldsymbol{\Pi}_i} \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(y_{it} - \mathbf{X}_{it}'\boldsymbol{\pi}_i), \tag{2.19}$$

where $\boldsymbol{\pi}_i(\tau) := (\lambda_i(\tau), \boldsymbol{\beta}_i(\tau)', \alpha_i(\tau), \boldsymbol{\delta}_i(\tau)')'$ and $\boldsymbol{\Pi}_i$ is a compact set in $\mathbb{R}^{(2+p_x)+(p_x+1)(p_T+2)}$. The first optimization problem depends on $\mathbf{f}_t$, and it is not feasible. The second optimization problem is feasible, but its use to obtain estimates of $\lambda_i(\tau)$ and $\boldsymbol{\beta}_i(\tau)$ requires justification. A formal analysis of the relationship between the two optimization problems is provided in the Online Appendix. Here, we provide an intuitive rationale by showing that under our assumptions the two optimization problems converge as N, T and $p_T \to \infty$.

With this in mind, using equations (2.1) and (2.17), first note that

$$\rho_\tau(y_{it} - \mathbf{W}_{it}'\boldsymbol{\theta}_i) - \rho_\tau(y_{it} - \mathbf{X}_{it}'\boldsymbol{\pi}_i) = \rho_\tau(\xi_{it}) - \rho_\tau(\xi_{it} + h_{it,N}), \tag{2.20}$$

and by Knight's (1998) identity, we have that

$$\left|\rho_\tau(\xi_{it}) - \rho_\tau(\xi_{it} + h_{it,N})\right| \leq 3\left|h_{it,N}\right|, \tag{2.21}$$

and hence the difference between the feasible and infeasible criteria is:

$$\left|\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\xi_{it}) - \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\xi_{it} + h_{it,N})\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\left|h_{it,N}\right|, \tag{2.22}$$

which upon using (2.16), yields

$$\left|\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\xi_{it}+h_{it,N})-\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\xi_{it})\right|\leq K\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{l=p_T+1}^{\infty}\bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}\right|+O_p\left(\frac{1}{\sqrt{N}}\right). \tag{2.23}$$

But,

$$\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{l=p_T+1}^{\infty}\bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}\right|\leq\frac{1}{T}\sum_{t=1}^{T}\sum_{l=p_T+1}^{\infty}\|\bar{\mathbf{z}}_{t-l}\|\|\boldsymbol{\delta}_{il}\|,$$

and since under Assumption 7, $\|\boldsymbol{\delta}_{il}\|<K\rho^l$ for all i and l, then

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left|\sum_{l=p_T+1}^{\infty}\bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}\right| &\leq K\rho^{p_T+1}\sum_{j=0}^{\infty}\rho^j\left(\frac{1}{T}\sum_{t=1}^{T}\|\bar{\mathbf{z}}_{t-j-p_T-1}\|\right)\\ &\leq \left(\frac{K\rho^{p_T+1}}{1-\rho}\right)\sup_j\left(\frac{1}{T}\sum_{t=1}^{T}\|\bar{\mathbf{z}}_{t-j-p_T-1}\|\right).\end{aligned}$$

Using this results in (2.23), now yields

$$\left|\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\xi_{it}+h_{it,N})-\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\xi_{it})\right|\leq\left(\frac{K\rho^{p_T+1}}{1-\rho}\right)\sup_j\left(\frac{1}{T}\sum_{t=1}^{T}\|\bar{\mathbf{z}}_{t-j-p_T-1}\|\right)+O_p\left(\frac{1}{\sqrt{N}}\right),$$

which tends to zero as N, T, and $p_T\to\infty$, since under Assumption 7, $0<\rho<1$, and under Assumptions 4 and 5, $\{\bar{\mathbf{z}}_t\}$ process is stationary and its average is bounded in p_T.

In short, having replaced the unobserved factors, $\mathbf{f}_t$, in (2.1) by the current and lagged cross-section averages $\bar{\mathbf{z}}_t$, we can now consider the approximate quantile function associated with (2.15) which we write as

$$Q_{Y_{it}}(\tau|\bar{\mathfrak{F}}_{it})=\alpha_i(\tau)+\lambda_i(\tau)y_{it-1}+\mathbf{x}_{it}'\boldsymbol{\beta}_i(\tau)+\sum_{l=0}^{p_T}\bar{\mathbf{z}}_{t-l}'\boldsymbol{\delta}_{il}(\tau), \tag{2.24}$$

where the feasible set $\bar{\mathfrak{F}}_{it}$ includes $\mathbf{x}_{it}$, y_{it-1}, and $\bar{\mathbf{z}}_{t-s}$ for all $s=0,1,2,\ldots,p_T$.

This quantile function can now be used to estimate the parameters of interest $\boldsymbol{\vartheta}_i(\tau):=(\lambda_i(\tau),\boldsymbol{\beta}_i(\tau)')'$ for each i and $0<\tau<1$, by minimizing the individual specific objective function given by (2.19).

We also propose a quantile mean group estimator for $\boldsymbol{\vartheta}(\tau):=E((\lambda_i(\tau),\boldsymbol{\beta}_i(\tau)')')$. The estimator is,

$$\hat{\boldsymbol{\vartheta}}(\tau)=\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\vartheta}}_i(\tau)=\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{\Xi}_i\circ\hat{\boldsymbol{\pi}}_i(\tau)\right), \tag{2.25}$$

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 (2.25) as quantile common correlated effects mean group estimator, or simply 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 for each i and $0 < \tau < 1$ defined in (2.25) 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. Although it is possible to consider other functionals of the random coefficients $\boldsymbol{\vartheta}_i(\tau)$, 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, $\bar{\mathbf{z}}_t$. 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 (2.19) and (2.25), respectively. Throughout this section, we write equation (2.17) as $y_{it} = \mathbf{X}_{it}'\boldsymbol{\pi}_i + e_{it}$, where $e_{it} = \xi_{it} + h_{it,N}$, $h_{it,N}$ is defined by (2.16), and recall that $\mathbf{X}_{it} = (y_{it-1}, \mathbf{x}_{it}', 1, \bar{\mathbf{z}}_t', \bar{\mathbf{z}}_{t-1}', ..., \bar{\mathbf{z}}_{t-p_T}')'$. Recall also that $h_{it,N} \to 0$ as N, T, and $p_T \to \infty$.

The following result states the weak consistency of the estimator:

Theorem 1 (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* (2.2)-(2.6)*, and Assumptions 1-7 and S.1-S.2 hold. As N, T and p_T go jointly to infinity such that $p_T^3/T \to \varkappa$, $0 < \varkappa < \infty$, and* $\log(N)/T \to 0$*, the cross-section augmented quantile regression estimator, $\hat{\boldsymbol{\pi}}_i(\tau)$, defined by* (2.19)*, is consistent uniformly over $1 \leq i \leq N$.*

A proof of Theorem 1 is in the Online Appendix. 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 $\bar{\mathbf{z}}_t, \bar{\mathbf{z}}_{t-1}, \ldots, \bar{\mathbf{z}}_{t-p_T}$ to account for the unobserved factors, $\mathbf{f}_t$. 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 $p_T^3/T \to \varkappa$. We need, however, $N \to \infty$ for consistency of our approach.

The result of Theorem 1 holds when $\log(N)/T \to 0$, as $N \to \infty$. This is the rate derived in Kato, Galvao and Montes-Rojas (2012), and the restriction that T grows at most polynomially in N is not required in Chudik and Pesaran (2015). The difference in the conditions is explained by the rates needed to eliminate asymptotically a term in the Bahadur representation of the estimator that arises in quantile problems with incidental parameters.

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 2 (Consistency of $\hat{\boldsymbol{\vartheta}}(\tau)$)**.** *Under the conditions of Theorem 1, as* (N, T, p_T) *go jointly to infinity with* $p_T^3/T \to \varkappa$*,* $0 < \varkappa < \infty$*, and* $\log(N)/T \to 0$*, the mean quantile group estimator defined by* (2.25) *for the panel data model* (2.2) - (2.6) *is weakly consistent, namely for every* $0 < \tau < 1$*,* $\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau) \xrightarrow{p} 0$*.*

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

Theorem 3 (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* (2.2)-(2.6)*, and Assumptions 1-7 and S.1-S.3 hold. As* $(N, T, p_T) \to \infty$ *with* $p_T^3/T \to \varkappa$*,* $0 < \varkappa < \infty$*, and* $N^{2/3}(\log(N))/T \to 0$*, the mean group quantile regression estimator, defined by* (2.25)*, for a model with interactive effects,* $\sqrt{N}(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_v)$*.*

Because the approximation of the factors requires $N \to \infty$ and we let N and T go jointly to infinity, the rates of Theorem 3 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 4. *Under the Assumptions of Theorem 3, as* $(N, T, p_T) \to \infty$ *with* $p_T^3/T \to \varkappa$*,* $0 < \varkappa < \infty$*, and* $N^2(\log(N))^3/T \to 0$*, the mean group quantile regression estimator, defined by* (2.25)*, for a model with interactive effects with* $\boldsymbol{\vartheta}_i(\tau) = \boldsymbol{\vartheta}(\tau)$ *for* $1 \leq i \leq N$*,* $\sqrt{NT}(\hat{\boldsymbol{\vartheta}}(\tau) - \boldsymbol{\vartheta}(\tau)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}_\psi)$*.*

The convergence of the QMG estimator in Theorem 3 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 4 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.

Remark 3. A recent paper by Galvao, Gu and Volgushev (2018)**?** finds improvements on the rates of N and T that are similar to the usual conditions in standard nonlinear panel data models. Their theoretical investigation does not include dynamic models. We expect, however, that similar improvements can be achieved in our case.

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:

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

where $\beta_{0i} = \alpha_i + \beta_0$, the error term u_{it} is distributed as F_u, 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

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

with $\boldsymbol{\theta}_i = (\alpha_i, \lambda_i, \boldsymbol{\beta}_i', \boldsymbol{\gamma}_i')'$, $\boldsymbol{\beta}_i = (\beta_{1,i}, \beta_{2,i})'$, $\boldsymbol{\gamma}_i = (\gamma_{1,i}, \gamma_{2,i})'$, $\beta_{i0}(\tau) = \alpha_i + \beta_0(\tau)$, 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,

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

with $\boldsymbol{\theta}_i(\tau) = (\alpha_i(\tau), \lambda_i, \boldsymbol{\beta}_i'(\tau), \boldsymbol{\gamma}_i')'$, $\boldsymbol{\beta}_i(\tau) = (\beta_{1,i}(\tau), \beta_{2,i})'$, $\beta_{i0}(\tau) = \alpha_i(\tau) + \beta_0$, $\alpha_i(\tau) = \alpha_i + \kappa_{0i} F_u^{-1}(\tau)$ and $\beta_{1,i}(\tau) = \beta_{1,i} + \kappa_{0i}\kappa_{1i} F_u^{-1}(\tau)$. 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 $\beta_{2,i} = 0.5$, for $1 \leq i \leq N$. Note that for S sufficiently large, we have that,

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

where $\xi_{it} = \kappa_{0i}(1 + \kappa_{1i}x_{1,it})u_{it}$. 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

$$x_{j,it} = \mu_{j,i} + \Gamma_{j,i} f_{j,t} + v_{j,it}, \tag{3.5}$$

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

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

for $j \in \{1, 2\}$, with $\mu_{1,i} = \mu_{2,i} = \mu_i \sim iid\, \mathcal{N}(0.5, 1)$, $\varepsilon_{j,it} \sim iid\, \mathcal{N}(0, 1)$, and $\varepsilon_{j,t} \sim iid\, \mathcal{N}(0, 1)$. 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), $\gamma_{1,i}$ and $\gamma_{2,i}$, and in equation (3.5), $\Gamma_{1,i}$ and $\Gamma_{2,i}$, are generated as $\gamma_{j,i} \sim iid\, \mathcal{N}(0.5, 1)$ and $\Gamma_{j,i} \sim iid\, \mathcal{N}(0.5, 1)$ 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 $\alpha_i = \bar{x}_{1i} + \gamma_{1,i}\bar{f}_1 + \gamma_{2,i}\bar{f}_2 + \bar{u}_i + a_i$, where the individual specific averages are defined as $\bar{x}_{1i} = T^{-1}\sum_{t=1}^{T} x_{1,it}$, $\bar{f}_j = T^{-1}\sum_{t=1}^{T} f_{j,t}$, $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$. The error term a_i in the equation for α_i is assumed to be distributed as $\mathcal{N}(0, 1)$.

In the simulations, we set $\lambda_i = \lambda = 0.5$ for $i = 1, 2, \ldots, N$ and we assume that the error term u_{it} in equation (3.1) is $iid\, \mathcal{N}(0, 1)$. In the Online Appendix, we consider the error term u_{it} distributed as t-student with 4 degrees of freedom (t_4), and χ^2 with 3 degrees of freedom (χ^2_3). We also consider different variations of the model considered here including the case of heterogeneous λ_i's and models without factor structure. We assume that $\beta_{1,i} = \beta_1 + \nu_{1,i}$, where $\beta_1 = 1$ and $\nu_{1,i} \sim \mathcal{U}(-0.25, 0.25)$. Thus, $\beta_{1,i}(\tau) = \beta_{1,i} + \kappa_{0i}\kappa_{1i}F_u^{-1}(\tau) = 1 + \nu_{1,i} + \kappa_{0i}\kappa_{1i}F_u^{-1}(\tau)$. In this case, $E(\beta_{1,i}(\tau)) = \beta_1(\tau) = 1 + 0.1F_u^{-1}(\tau)$.

The first columns of Table 3.1 presents the bias and root mean square error (RMSE) for the slope parameter $\beta_1(\tau)$. The table shows 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 QMG estimator with the instrumental variable quantile regression estimator for dynamic panel data model developed by Galvao (2011), using y_{it-2} as an instrument for y_{it-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

N	T		$\tau = 0.50$ quantile DQR	$\tau = 0.50$ quantile QMG	$\tau = 0.25$ quantile DQR	$\tau = 0.25$ quantile QMG	$\tau = 0.50$ quantile DQR	$\tau = 0.50$ quantile QMG	$\tau = 0.25$ quantile DQR	$\tau = 0.25$ quantile QMG
			$\text{Corr}(\beta_{1,i}, \bar{x}_{1,i}) = 0$				$\text{Corr}(\beta_{1,i}, \bar{x}_{1,i}) \neq 0$			
100	50	Bias	-0.192	0.050	-0.166	0.063	-0.260	0.040	-0.226	0.053
100	50	RMSE	0.221	0.061	0.198	0.073	0.293	0.090	0.262	0.096
100	100	Bias	-0.263	0.024	-0.240	0.031	-0.339	0.019	-0.311	0.026
100	100	RMSE	0.280	0.034	0.259	0.041	0.359	0.077	0.332	0.080
100	200	Bias	-0.291	0.002	-0.269	0.008	-0.373	-0.009	-0.344	-0.004
100	200	RMSE	0.300	0.021	0.279	0.023	0.387	0.081	0.359	0.082
200	50	Bias	-0.190	0.057	-0.164	0.070	-0.260	0.046	-0.227	0.058
200	50	RMSE	0.216	0.063	0.192	0.075	0.283	0.073	0.253	0.082
200	100	Bias	-0.256	0.030	-0.233	0.036	-0.333	0.021	-0.302	0.028
200	100	RMSE	0.272	0.034	0.250	0.040	0.348	0.059	0.319	0.061
200	200	Bias	-0.293	0.013	-0.271	0.017	-0.374	0.007	-0.345	0.011
200	200	RMSE	0.302	0.019	0.280	0.022	0.383	0.054	0.355	0.055

TABLE 3.1. *Bias and root mean square error (RMSE) of quantile regression estimators for* $\beta_1(\tau)$. *DQR denotes the instrumental variable quantile regression estimator for dynamic quantile regression, and QMG denotes the proposed mean quantile group estimator defined by* (2.25).

standard quantile estimators, $\hat{\beta}_{1,i}(\tau)$, using $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{y}_{t-1}, \bar{\mathbf{x}}_t')'$ to proxy the true unobserved factors $f_{1,t}$ and $f_{2,t}$.

As can be seen from Table 3.1, not surprisingly, the DQR estimator of β_1 is biased. Furthermore, the bias of DQR estimator tends to increase with T, and tends 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.

In the last four columns of Table 3.1, we show evidence on the performance of the QMG estimator in models where $\beta_{1,i}$ is correlated with one of the regressors, $x_{1,it}$. Specifically, the data generating process is the same as the model presented in equations (3.1) - (3.7), however, we generate $\beta_{1,i}$ as follows:

$$\beta_{1,i} = \beta_1 + \nu_{1,i} + 0.25\sqrt{T}\bar{v}_{1,i},$$

			$\tau = 0.50$ quantile				$\tau = 0.25$ quantile			
			Parameter: λ		Parameter: θ_1		Parameter: λ		Parameter: θ_1	
N	T		DQR	QMG	DQR	QMG	DQR	QMG	DQR	QMG
					$\mathrm{Corr}(\beta_{1,i}, \bar{x}_{1,i}) = 0$					
100	50	Bias	0.191	-0.061	0.664	-0.066	0.187	-0.060	0.624	-0.018
100	50	RMSE	0.201	0.063	0.731	0.097	0.197	0.062	0.686	0.080
100	100	Bias	0.226	-0.021	0.714	-0.013	0.223	-0.021	0.659	0.008
100	100	RMSE	0.231	0.024	0.747	0.050	0.229	0.024	0.689	0.052
100	200	Bias	0.240	-0.003	0.735	0.003	0.238	-0.003	0.682	0.015
100	200	RMSE	0.242	0.009	0.752	0.041	0.241	0.009	0.699	0.045
200	50	Bias	0.190	-0.065	0.657	-0.068	0.185	-0.065	0.618	-0.027
200	50	RMSE	0.199	0.066	0.721	0.085	0.195	0.067	0.674	0.060
200	100	Bias	0.222	-0.027	0.701	-0.025	0.220	-0.027	0.652	-0.006
200	100	RMSE	0.227	0.028	0.731	0.043	0.225	0.028	0.678	0.037
200	200	Bias	0.240	-0.010	0.734	-0.005	0.239	-0.010	0.679	0.007
200	200	RMSE	0.243	0.011	0.748	0.027	0.241	0.011	0.693	0.028
					$\mathrm{Corr}(\beta_{1,i}, \bar{x}_{1,i}) \neq 0$					
100	50	Bias	0.227	-0.057	0.746	-0.047	0.222	-0.057	0.704	-0.016
100	50	RMSE	0.234	0.059	0.825	0.168	0.229	0.059	0.785	0.246
100	100	Bias	0.261	-0.019	0.784	-0.003	0.259	-0.019	0.729	0.017
100	100	RMSE	0.265	0.022	0.828	0.151	0.263	0.022	0.778	0.153
100	200	Bias	0.276	0.000	0.801	-0.010	0.274	0.000	0.740	0.000
100	200	RMSE	0.278	0.008	0.840	0.162	0.276	0.008	0.783	0.164
200	50	Bias	0.224	-0.063	0.730	-0.051	0.220	-0.063	0.694	-0.013
200	50	RMSE	0.231	0.064	0.803	0.124	0.227	0.065	0.761	0.115
200	100	Bias	0.258	-0.025	0.778	-0.020	0.255	-0.025	0.731	-0.001
200	100	RMSE	0.262	0.026	0.812	0.110	0.259	0.026	0.765	0.108
200	200	Bias	0.277	-0.008	0.816	-0.005	0.275	-0.008	0.755	0.006
200	200	RMSE	0.279	0.010	0.839	0.106	0.277	0.010	0.779	0.106

TABLE 3.2. *Bias and root mean square error (RMSE) of quantile regression estimators for λ and θ_1. 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* (2.25).

where $\bar{v}_{1,i} = T^{-1}\sum_{t=1}^{T} v_{1,it}$, and $v_{1,it}$ is the idyosincratic component of $x_{1,it}$ defined by (3.5). We note that the term $\sqrt{T}\bar{v}_{1,i}$ ensures that the correlation of between $\beta_{1,i}$ and $\bar{x}_{1,i}$ is different from zero for all T, and as $T \to \infty$. As shown in the first four columns of the table, the QMG estimator offers the best finite sample performance in terms of bias and RMSE, although as to be expected, the QMG estimator performs less well when we allow for non-zero correlation between slopes and the regressors.

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.

Table 3.2 shows the bias and RMSE of the DQR and QMG estimators for the parameters of interest. 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. Overall, 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.

We also investigate the relative performance of QMG in models with and without factor structure. We use the same data generating process as in equations (3.1) - (3.7), but we generalize the parametrization of equation (3.1) as follows:

$$y_{it} = \beta_{0i} + \lambda_i y_{it-1} + \beta_{1,i}x_{1,it} + \beta_{2,i}x_{2,it} + \sigma_\gamma \sum_{j=1}^{2} \gamma_{j,i} f_{j,t} + \kappa_{0i}(1 + \kappa_{1i}x_{1,it})u_{it}.$$

Naturally, when $\sigma_\gamma = 0$, the model does not include latent factors. When we set $\sigma_\gamma = 1$, we obtain equation (3.1). We set σ_γ to take values in the interval $[0, 1]$.

Figure 3.1 presents the bias and RMSE of the estimators for $E(\lambda_i(\tau))$ and $E(\beta_{1,i}(\tau))$ when $N = 100$ and $T = 200$. 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 σ_γ is not too small (namely, when the latent factors are important).

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

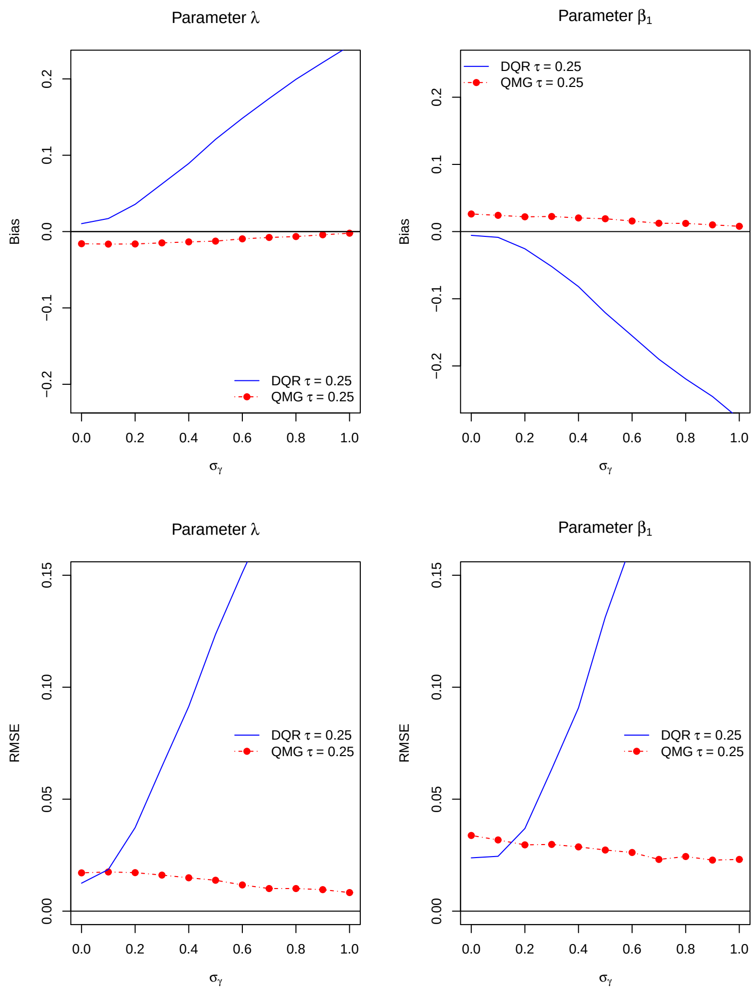

FIGURE 3.1. *Small sample performance of the DQR and QMG estimators in models with ($\sigma_\gamma \neq 0$) and without ($\sigma_\gamma = 0$) latent factors, as distinguished by the parameter σ_γ.*

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

[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.

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 (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 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.

Treated households were then further randomized by received additional enabling technologies. All treated households had access to a website ("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

[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.*

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 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:

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

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 that includes temperature and dew point, $\mathbf{x}_{ig,t} = (x_{1,ig,t}, x_{2,ig,t}, \ldots, x_{1,ig,t-4}, x_{2,ig,t-4})'$. 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

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

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 $p_T = 4$, although we offer evidence on the robustness of results in Section 4.4. 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 4.5.

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 proposed quantile regression estimator, labeled QMG.

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

	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.*

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.[4] 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 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.

4.4. **Robustness to Lagged Cross Section Averages**

This section offers additional results by evaluating the sensitivity of the main coefficient estimates of $\lambda_{ig}(\tau)$ and $\delta_{ig}(\tau)$ to the number of lagged cross-section averages used to proxy the latent factors. Using Figure 4.1, we report results by varying the number of lagged cross-section averages of electricity consumption, temperature, and drew point. The figure presents results by different quantiles and treatment groups. The number of lagged cross-section averages included in the model varies from 0 to 10. For instance, $p_T = 0$ means that we estimate model (4.2) by replacing $\mathbf{f}_t$ with $(\bar{y}_t, \bar{x}_{1,t}, \bar{x}_{2,t})$,

[4]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.

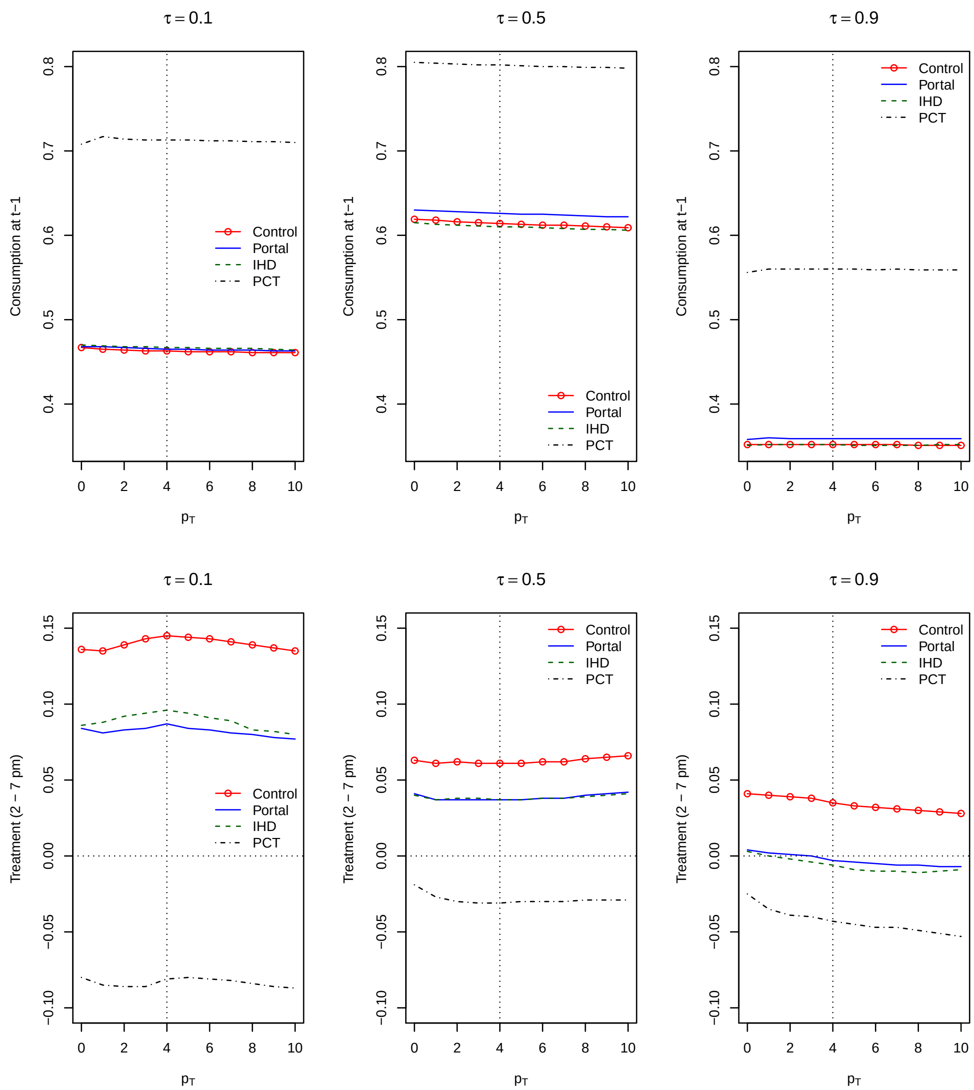

FIGURE 4.1. *Robustness to the number of lagged cross-section averages.*

$p_T = 1$ means that we estimate the model by replacing $\mathbf{f}_t$ with $(\bar{y}_t, \bar{x}_{1,t}, \bar{x}_{2,t}, \bar{y}_{t-1}, \bar{x}_{1,t-1}, \bar{x}_{2,t-1})$, etc. The vertical dotted line indicates the number of lagged cross-section averages considered in Table 4.2 ($p_T = 4$).

The evidence illustrates that the results are robust to the number of lagged cross-section averages used to proxy $\mathbf{f}_t$. In most cases, we find that the most significant changes in the estimates are obtained when p_T is increased from 0 to a positive number, which is largely consistent with the idea that the approximation of factors in dynamic settings relies heavily on lag values of cross-sectional averages.

4.5. **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 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

		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.*

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.

4.6. **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

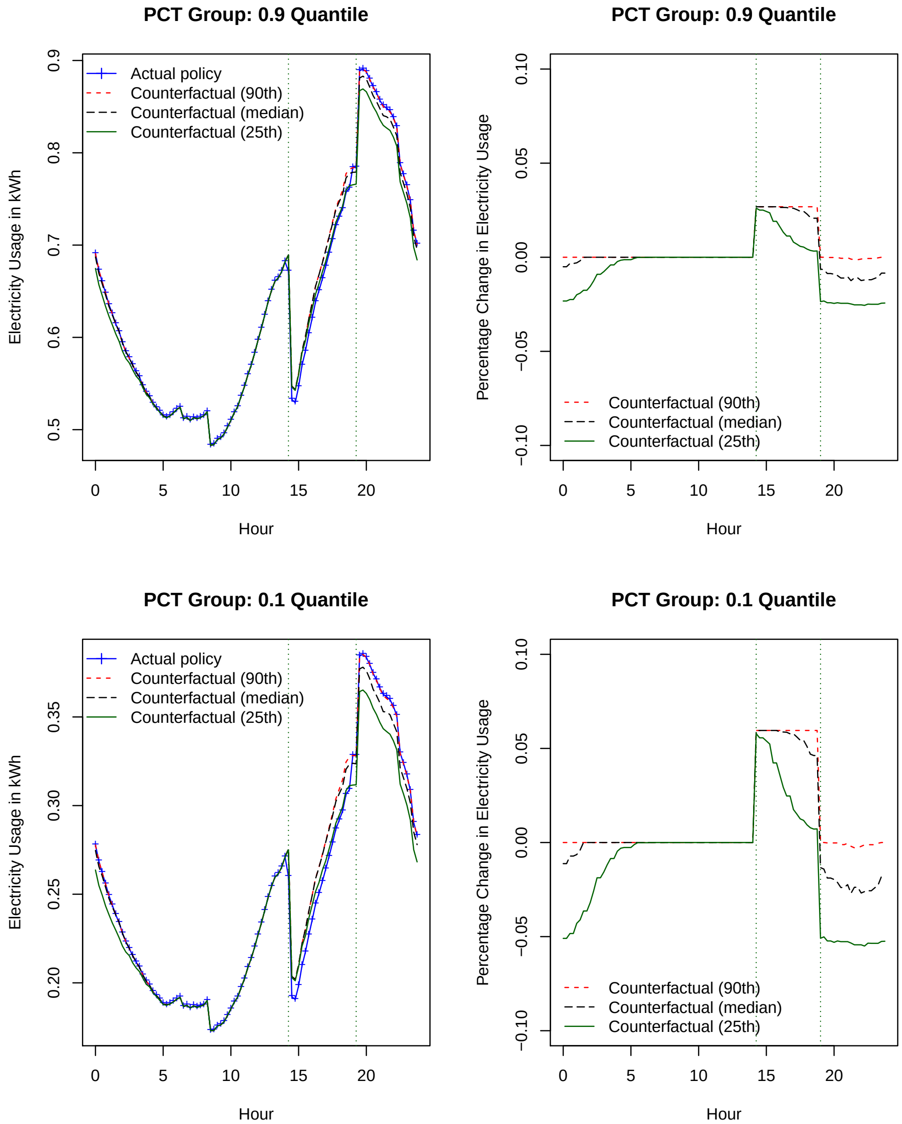

FIGURE 4.2. *Counterfactual policies for customers with a PCT. The right panels show the percentage change in electricity usage with respect to the actual policy.*

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 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

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 (2019): "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.

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., J. Gu, and S. Volgushev (2018): "On the Unbiased Asymptotic Normality of Quantile Regression with Fixed Effects," arXiv:1807.11863 [econ.EM].

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.

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.

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.

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.

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.