Modal Regression for Fixed Effects Panel Data[*]

Aman Ullah [†]   Tao Wang [‡]   Weixin Yao [§]

November 25, 2020

Abstract

Most research on panel data focuses on mean or quantile regression while there is not much research about regression methods based on the mode. In this paper, we propose a new model named *fixed effects modal regression* for panel data in which we model how the conditional mode of the response variable depends on the covariates, and employ a kernel-based objective function to simplify the computation. The proposed modal regression can complement the mean and quantile regressions, and provide better central tendency measure and prediction performance when the data is skewed. We present a linear dummy modal regression (LDMR) method and a pseudo-demodeing two-step (PDTS) method to estimate the proposed modal regression. The computations can be easily implemented using a modified modal-expectation-maximization (MEM) algorithm. We investigate the asymptotic properties of the modal estimators under some mild regularity conditions when the number of individuals, N, and the number of time periods, T, go to infinity. The optimal bandwidths with order $(NT)^{-1/7}$ are obtained by minimizing the asymptotic weighted mean squared errors. Monte Carlo simulations and two real data analyses of a public capital productivity study and a carbon dioxide (CO_2) emissions study are presented to demonstrate the finite sample performance of the newly proposed modal regression.

Keywords: Asymptotic property, Fixed effects, MEM algorithm, Modal regression, Panel data, Pseudo-demodeing.

JEL Codes: C01, C14, C23, C51.

[†]Department of Economics at University of California, Riverside, CA 92521. E-mail: aman.ullah@ucr.edu.

[‡]Department of Economics at University of California, Riverside, CA 92521. E-mail: tao.wang@email.ucr.edu.

[§]Department of Statistics at University of California, Riverside, CA 92521. E-mail: weixin.yao@ucr.edu.

1   Introduction

Mean, median, and mode are three of most commonly and popularly used location measures, and focus on different population characteristics. Each quantity has its own merit and complements each other. Built on the ideas of mean and median, mean regression and median regression are extensively investigated and popularly used to model the relationship between a response variable Y and covariates X. However, the research about the regression model built on the idea of mode is rather limited, and has not gotten enough attention that it deserves partly due to its computational difficulty. Compared to mean, mode is a better and more stable numerical characteristic of a dataset when data has outliers or the error distribution is skewed (Chen, 2018). Tarter and Lock (1993) argued that we should not stick to a single function for summarizing the relationship among variables and should also pay attention to the conditional local modes, especially when data has high skewness. Due to the special characteristics of mode, many authors have made efforts to identify modes of population distributions for low-dimensional data based on the nonparametric kernel density estimation; see, for example, Silverman (1981), Muller and Sawitzki (1991), Friedman and Fisher (1999), Chen et al. (2016), as well as documentations of the R package "np" for nonparametric mode estimation. In the regression framework,[1] Sager and Thisted (1982) pointed out that a modal regression estimator can be derived from a nonparametric density estimate to analyze the relationship among variables. However, such modal regression method derived from multivariate kernel density estimation is difficult to apply when the dimension of the covariates is large. In the econometrics literature, the path-breaking papers of Lee (1989, 1993) started the idea of linear modal regression. However, Lee (1989, 1993) used an objective function with bounded support and assumed the tuning parameter h to be fixed. Therefore, they required the error to be symmetric to get a consistent modal line. Yao et al. (2012) proposed a kernel-based objective function to find the nonparametric modal estimator but also required the error to be symmetric. Note however when the error is symmetric, the modal regression line is the same as the mean regression line. Kemp and Santos Silva (2012) and Yao and Li (2014) found out that the above limitations can be solved if we let the tuning parameter h go to zero. They also proposed a more general kernel-based objective function and established the general consistency of the linear modal regression estimate even when the error density is skewed. Such findings greatly simplify the computations of modal regression.

More specifically, let $f(Y \mid X)$ be the conditional density function of Y given X. The traditional regression model usually uses the mean or the quantile of $f(Y \mid X)$ to model the

---

[1]Modal regression can complement mean and quantile regressions and provide some other useful information regarding the features of conditional distributions that the existing regression models might miss, especially for the skewed dataset. For example, assume Y and X satisfy $Y = X^T \beta_m + \sigma(X)\xi$, where ξ has a density with mean 0 and mode 1, β_m is a vector of coefficients, $\sigma(X) = m(X) - X^T \beta_m$ in which $m(X)$ is a nonlinear function, and X^T denotes the transpose of X. Then $\mathbb{E}(Y \mid X) = X^T \beta_m$, while $Mode(Y \mid X) = m(X)$. The mean regression is linear, but the modal regression could be nonlinear. Similarly, it is also possible that the mean regression is nonlinear, but the modal regression is linear.

relationship between Y and X. For example, the commonly used linear regression model assumes that the mean of $f(Y \mid X)$ is a linear function of X. The linear modal regression (Kemp and Santos Silva, 2012; Yao and Li, 2014) instead assumes that the *mode* of $f(Y \mid X)$, denoted by $Mode(Y \mid X)$, is a linear function of covariates X: $Mode(Y \mid X) = X^T\beta$. The regression coefficient β, which reveals the change in the mode of the response variable Y corresponding to a unit change in the covariates X, can be estimated effectively by maximizing the following kernel-based objective function

$$Q_N(\beta) = \frac{1}{Nh} \sum_{i=1}^{N} \phi\left(\frac{Y_i - X_i^T\beta}{h}\right),$$

where $\phi(.)$ denotes a smooth kernel, and h is a bandwidth. Such idea of imposing a certain model assumption directly on the conditional mode, $Mode(Y \mid X)$, along with the kernel-based objective function greatly simplifies the computation of the modal regression, and avoids the nonparametric multivariate kernel density estimation of $f(X, Y)$ that is needed for the fully nonparametric multivariate modal estimation. Due to such reasons, we have experienced much development of modal regression recently. Especially, with the distinguished characteristics of modal regression (such as robustness and better prediction performance (shorter prediction interval)), the idea of linear modal regression was subsequently extended by many researchers such as Yao and Xiang (2016), Zhou and Huang (2016), Chen et al. (2016), Krief (2017), Chen (2018), Li and Huang (2019), Ota et al. (2019), Kemp et al. (2020), among others.

Panel data models have been well developed and widely used in empirical economics due to the appealing feature that it allows in modeling time-invariant individual specific effects. Baltagi (2009) and Baltagi (2013) gave an excellent overview of the panel data analysis technique. There are a large number of estimation methods that could be used to estimate panel data models, such as the first difference approach and the profile least-squares method. Su and Ullah (2006) presented a local linear kernel estimator for a nonparametric panel data model with fixed effects, where they derived the asymptotic theorems with fixed T and large N. Gao and Li (2013) proposed using the profile least-squares method to concentrate out fixed effects and estimated the model by a kernel method. Lee et al. (2019) introduced a local linear kernel estimator for the marginal effect in fixed-effect panel data models. Canay (2011) developed a two-step estimator for panel data quantile regression models with fixed effects. For other related literature, we refer the interested readers to Su and Ullah (2011), Lin et al. (2014), among others. While the above papers collectively cover a large class of panel models, all of them are based on mean or quantile regression. This motivates us to investigate the fixed effects panel data model from the modal regression view. To the best of our knowledge, the present paper is the first work to develop theory and methodology for *fixed effects modal regression* for panel data.

Given a panel dataset $\left\{\left(Y_{it}, X_{it}^T\right) : i = 1, \cdots, N; t = 1, \cdots, T\right\}$, the linear fixed effects *mean* regression assumes $\mathbb{E}(Y_{it} \mid X_{it}, \alpha_i) = X_{it}^T\beta_m + \alpha_i$ with $X_{it} \in \mathbb{R}^q$ and $\beta_m \in \mathbb{K}_1^q$, where α_i is treated

as a fixed parameter and allowed to be correlated with X_{it}, and $\mathbb{K}_1$ is a compact set. In this paper, we propose to model how the conditional *mode* changes with covariates for the panel data by the following linear *fixed effects modal regression*

$$Mode(Y_{it} \mid X_{it}, \mu_i) = X_{it}^T \beta + \mu_i, \tag{1.1}$$

where μ_i is treated as a fixed parameter in modal regression, $\beta \in \mathbb{K}_2^q$, and $\mathbb{K}_2$ is a compact set, i.e.,

$$Y_{it} = X_{it}^T \beta + \mu_i + v_{it} \tag{1.2}$$

with $Mode(v_{it} \mid X_{it}, \mu_i) = 0$. For the idiosyncratic error term, we assume that v_{it} is independently and identically distributed (i.i.d.) over i and t, and independent of unobservable time-invariant individual effect μ_i (Henderson et al., 2008; Lin et al., 2014). For the sake of simplicity, we focus on the one-way error-component model and balanced panel dataset. However, the methodology could be easily extended to the case of unbalanced panel data or a two-way error-component model. It is straightforward to see that different from the linear modal regression for the cross-sectional data (Kemp and Santos Silva, 2012; Yao and Li, 2014), the estimation and the corresponding asymptotic properties for the linear *fixed effects modal regression* depend on the assumptions for (N, T). It is noticed that we only need $NT \to \infty$ to establish the asymptotic theorems for the proposed LDMR estimators. However, we require T grow much faster than N to simplify the deviation of the asymptotic results for the proposed PDTS estimators (see Section 3). Given these, we mainly focus on the case in which both N and T go to infinity in this paper, and also discuss other various cases under different conditions of N and T (finite or going to infinity) in the remarks.

We first propose a LDMR method to estimate (1.1) by using dummy variables for fixed effects $\{\mu_i\}_{i=1}^N$. The estimates of all parameters can be obtained by extending the modal-expectation-maximization (MEM) algorithm (Li et al., 2007; Yao, 2013) to maximize a kernel-based objective function. Consistency and asymptotic normality of the estimators are investigated under some mild conditions with both N and T going to infinity. We also give the asymptomatically optimal bandwidth and show how to obtain it in practice. Meanwhile, for large N case, which is a typical situation for panel data from an annually conducted panel survey, one problem of the LDMR method is that there might exist multiple local maxima and there is no way to ensure the found solution is the global optimum. To this end, we propose a PDTS method under certain conditions to concentrate out fixed effects through mean regression, and derive the asymptotic bias and variance as well as the asymptotic normality of the resulting estimators. The advantage of this PDTS method is that it not only incorporates the mode structure of the data, but also inherits the estimation superiorities of fixed effects mean regression. A similar approach has been adopted in a quantile regression for panel data; see Canay (2011). It is noticed that we are required to impose a more restrictive condition on N and

T, notably that at the optimal bandwidth convergence rate, $N^a/T \to 0$ for some $a > 4/3$ with $(N, T) \to \infty$, to prove asymptotic normality for the PDTS estimators. Bandwidth selection is also investigated for the PDTS method. Monte Carlo experiments and two empirical analyses are conducted to demonstrate the finite sample performance of the proposed modal regression. It is important to emphasize that the convergence rate ($\sqrt{NTh^3}$) of *fixed effects modal regression* in this paper is slower than that ($\sqrt{NTh}$) of the nonparametric univariate fixed effects mean regression, which is the cost we need to pay in order to estimate the conditional mode (Parzen, 1962). However, for finite sample performance of skewed data, *fixed effects modal regression* might still provide estimates with smaller standard errors and better prediction performance than fixed effects mean regression with least squares estimates based on our simulation results.

The rest of this paper is organized as follows. In Section 2, we propose a LDMR estimation method, where the consistency and asymptotic properties of the estimators are investigated. Also, the optimal bandwidth is reported. In Section 3, we introduce a PDTS method to remove the individual effects, establish the asymptotic normality for estimators, and provide an optimal bandwidth. Section 4 has numerical results, where Monte Carlo simulations and two real data analyses are presented to illustrate the finite sample performance of the proposed model. We conclude this paper in Section 5. The tables which summarize the simulation and empirical results are deferred to the Appendix A. All technical proofs are given in the Appendix B.

2 Linear Dummy Modal Regression Method

In this section, we investigate *fixed effects modal regression* focusing on large N and large T case, and formally establish sufficient conditions for consistency and asymptotic normality of the estimators. We also discuss other various cases under different conditions of N and T (finite or going to infinity) in the remarks.

2.1 LDMR Modal Estimators

When the individual intercept μ_i is treated as a fixed parameter, the resulting model, known as the fixed effects regression, could be viewed as a special case of the classical linear model. We can then rewrite (1.1) with a little bit abuse of notation as

$$Mode(Y_{it} \mid X_{it}, \mu) = X_{it}^T \beta + Z_{\mu,i}^T \mu, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{2.1}$$

where $Z_{\mu,i} = (Z_{\mu,i}^{(1)}, \cdots, Z_{\mu,i}^{(N)})^T$, $\mu = (\mu_1, \cdots, \mu_N)^T$, and $Z_{\mu,i}^{(j)}$ denotes a dummy variable that is 0 for all observations with $i \neq j$ and 1 for $i = j$, $j = 1, \cdots, N$. We extend the method of Yao and Li (2014) to estimate (2.1) by maximizing the following kernel-based objective function

4

$$Q_{NT}(\beta, \mu) = \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi \left(\frac{Y_{it} - X_{it}^T \beta - Z_{\mu,i}^T \mu}{h_0} \right), \tag{2.2}$$

where h_0 is a bandwidth used in this section. Let $\hat{\beta}$ and $\hat{\mu}$ be the maximizers of (2.2) and thus the proposed LDMR estimators. According to Yao et al. (2012), the choice of kernel is not very important in modal regression. In this paper, we let ϕ be a normal kernel for the simplicity of computation; see Section 2.2 for more discussions on kernel choice.

Notice that unlike fixed effects mean regression, there is no closed-form expression of the maximizers of (2.2). We extend so-called MEM algorithm (Li et al., 2007; Yao, 2013) to estimate the modal coefficients in (2.2). The detailed description is summarized in Algorithm 1, where E-step and M-step are iterated until the algorithm converges.

Algorithm 1 MEM (Li et al., 2007; Yao, 2013)

E-Step. Calculate the weight $\pi \left(i, t \mid \beta^{(g)}, \mu^{(g)} \right)$ as

$$\pi \left(i, t \mid \beta^{(g)}, \mu^{(g)} \right) = \frac{\phi \left(\frac{Y_{it} - X_{it}^T \beta^{(g)} - Z_{\mu,i}^T \mu^{(g)}}{h_0} \right)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \phi \left(\frac{Y_{it} - X_{it}^T \beta^{(g)} - Z_{\mu,i}^T \mu^{(g)}}{h_0} \right)}.$$

M-Step. Update the values of $\beta^{(g+1)}$ and $\mu^{(g+1)}$ by

$$(\beta^{(g+1)}, \mu^{(g+1)}) = \arg\max_{\beta, \mu} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \pi \left(i, t \mid \beta^{(g)}, \mu^{(g)} \right) \log \left(\frac{1}{h_0} \phi \left(\frac{Y_{it} - X_{it}^T \beta - Z_{\mu,i}^T \mu}{h_0} \right) \right) \right\},$$

where g is the iteration indicator.

Remark 2.1. *For normal kernel function, M step has a closed-form expression $(X^{*T} W_g X^*)^{-1} X^{*T} W_g Y$, where W_g is a diagonal matrix associated with the weight π and X^* is the corresponding variable matrix defined as $X^* = (X \ Z)$ in which $X = (X_1, \cdots, X_N)^T$, $Z = (Z_1, \cdots, Z_N)^T$, $X_i = (X_{i1}^T, \cdots, X_{iT}^T)^T$, and $Z_i = (Z_{\mu,i}^T, \cdots, Z_{\mu,i}^T)^T$. It is necessary to point out that as the converged value depends on the starting point and there is no guarantee that the MEM algorithm will converge to the global maximizer,[2] it is prudent to try different starting points and choose the best optimal value (Yao et al., 2012).*

2.2 Asymptotic Properties

In what follows, we derive the asymptotic properties of the LDMR estimators. We consider the consistency of $(\hat{\mu}, \hat{\beta})$, where we say that $\hat{\mu}$ is weakly consistent if $\hat{\mu}_i$ convergences to μ_{i0} uniformly over $1 \leq i \leq N$ in which μ_{i0} is a fixed true parameter. We then derive the limiting

---

[2]Theorem 2.1 in Yao and Li (2014) indicates that Algorithm 1 will monotonically non-decrease the objective function (2.2), which means MEM algorithm has the ascending property.

distribution of $\hat{\beta}$. To start with, we list the regularity sufficient conditions that will be used to establish the asymptotic results of the LDMR estimators.

C1 The random variables (Y_{it}, X_{it}^T) are i.i.d. across the i index. For each $i = 1, \cdots, N$, $\left\{ X_{it}^T : t = 1, \cdots, T \right\}$ are the realization of a strictly stationary α-mixing process with mixing coefficient $\alpha_i(j)$. It holds that $\alpha(j) = \max_i \alpha_i(j)$ for all $i = 1, \cdots, N$, where the coefficient $\alpha(j)$ decays exponentially fast to zero as $j \to \infty$.

C2 The unobservable time-invariant individual effect $\{\mu_i\}_{i=1}^n$ is an i.i.d. sequence of random variables with $Mode(\mu_i \mid X_{it}) \neq 0$. v_{it} is i.i.d. across all i and t with $Mode(v_{it} \mid X_{it}, \mu_i) = 0$.

C3 As $(N, T) \to \infty$, $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \|X_{it}\|^4 = O_p(1)$, where $\|.\|$ is Euclidean distance. Define $X_{it}^{*T} = (X_{it}^T, Z_{\mu,i}^T)$, as $(N, T) \to \infty$, $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it}^* X_{it}^{*T} f_v^{(2)}(0 \mid X_{it}, Z_{\mu,i}))$ converges in probability to a negative definite matrix, and $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it}^* X_{it}^{*T} f_v(0 \mid X_{it}, Z_{\mu,i}))$ and $(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (X_{it}^* f_v^{(3)}(0 \mid X_{it}, Z_{\mu,i}))$ converge in probability.

C4 $\phi(.)$ is bounded, symmetric about zero, and has compact support. Moreover, it has bounded continuous third derivative and finite second moment.

C5 For the conditional error density, $f_v(v \mid .)$, it follows $f_v(v \mid .) < f_v(0 \mid .)$. In addition, $f_v^{(c)}(v \mid .)$ is bounded and continuous in a neighbor of 0 for $c = 0, 1, 2, 3$, where $f_v^{(c)}(.)$ is the cth derivative of $f_v(.)$ and $f_v^{(1)}(0 \mid .) = 0$.

We give some comments on the above conditions. C1 is standard in the fixed effects panel literature. It excludes the temporal dependence and sticks to the identical distribution for all i to focus on the applicability of modal regression for fixed effects data. The mixing coefficient $\alpha_i(j)$ is defined as $\alpha_i(j) \equiv \sup_t \sup_{\left\{ A \in \mathcal{G}_{-\infty}^{i,t}, B \in \mathcal{G}_{t+k}^{i+\infty} \right\}} |P(A \cap B) - P(A)P(B)|$ in which $\mathcal{G}_{-\infty}^{i,t} = \sigma(\cdots, X_{i,t-1}, X_{it})$ and $\mathcal{G}_{t+k}^{i,+\infty} = \sigma(X_{i,t+k}, X_{i,t+k+1}, \cdots)$ are the σ-fields. The i.i.d. condition in C2 is the same as the ones used by Henderson et al. (2008) and Lin et al. (2014). It assumes the independence across the individuals and restricts the temporal dependence. To keep the exposition as clear as possible, the present paper does not extend the results to the case of serial dependence in v_{it} (see Remarks 2.2 and 2.3). C3, the standard moments condition, ensures the existence of the asymptotic mean and variance for modal estimators. C4 imposed on kernel function is in the line with Kemp and Santos Silva (2012) and Yao and Li (2014) to economize the proof. Note that the compact support condition can be relaxed if we impose certain restriction on the tail of the kernel function (Eddy, 1980). In particular, the standard normal kernel function is allowed. C5 implies certain smoothness of $f_v(v_{it} \mid X_{it}, Z_{\mu,i})$ around 0, and indicates that the conditional density of v_{it} has a well-defined global mode at 0 (Kemp and Santos Silva, 2012). All conditions on bandwidth are specified in the theorems stated below.

Remark 2.2. *C1 could be alternatively replaced by the condition that $\{X_{it}^T, t \geq 1\}$ are i.i.d for each fixed i and independent across i. However, we expect this structure to have no effect on the*

asymptotic behavior of the LDMR estimators besides a great deal of notational effort. Notice that C2 of v_{it} is restrictive, as when T is large, time series may have dependence in practice. We could replace the i.i.d. condition with the cross-sectional independence and the stationary α-mixing process to allow for serial dependence, i.e., for each $i = 1, \cdots, N$, $\{(Y_{it}, v_{it}) : t = 1, \cdots, T\}$ are the realization of a stationary α-mixing process with mixing coefficient $\alpha_i(j)$. Then the asymptotic consistency and normality results in this paper still hold under some regularity conditions, indicating the asymptotic negligence of the dependence, which is consistent with the general results in kernel literature; see Su et al. (2009).

Remark 2.3. *If we allow for serial dependence in v_{it}, we could propose a more efficient estimator by incorporating the information provided by the error autocorrelation structure. For example, one can define the serially correlated errors follow an autoregressive (AR) process with finite order d, i.e., $v_{it} = v_{i,t-1} + \cdots + v_{i,t-d} + e_{it}$ where e_{it} is i.i.d. with $Mode(e_{it} \mid X_{it}, \mu_i, v_{i,t-1}, \cdots, v_{i,t-d}) = 0$ for the purpose of identification. We can then extend the method proposed in Hidalgo (1992) to propose an adaptive estimation procedure for fixed effects modal regression. Future studies could fruitfully explore this issue further.*

When the tuning parameter h_0 goes to zero, for the general skewed error density, we have the following consistency properties of the LDMR estimators. As the primary focus of this paper is on the estimation of β, we only present the asymptotic distribution for $\hat{\beta}$.

Theorem 2.1. *Under the regularity conditions C1-C5, as $(N, T) \to \infty$, $h_0 \to 0$, and $NTh_0^5 \to \infty$, there exists consistent maximizers $(\hat{\beta}, \hat{\mu})$ of (2.2) such that with probability approaching one,*

(i) $\quad \|\hat{\beta} - \beta_0\| = O_p\left((NTh_0^3)^{-1/2} + h_0^2\right),$

(ii) $\quad \|\hat{\mu} - \mu_0\| = O_p\left((NTh_0^3)^{-1/2} + h_0^2\right),$

where β_0 and μ_0 are the true parameters of the fixed effects modal regression (1.1).

Theorem 2.2. *With $NTh_0^7 = O(1)$, under the same conditions in Theorem 2.1, the estimator $\hat{\beta}$ satisfying the consistency result in Theorem 2.1 has the following asymptotic result*

$$\sqrt{NTh_0^3}\left(\hat{\beta} - \beta_0 - \frac{h_0^2}{2}J^{-1}M + o_p(h_0^2)\right) \xrightarrow{d} \mathcal{N}\left(0, v_2 J^{-1}LJ^{-1}\right).$$

If with $NTh_0^7 \to 0$, we then have

$$\sqrt{NTh_0^3}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} \mathcal{N}\left(0, v_2 J^{-1}LJ^{-1}\right),$$

where $v_2 = \int t^2\phi^2(t)dt$, $J = \lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(X_{it}(X_{it}^T - Z_{\mu,i}^T\Phi^{-1}\Psi)f_v^{(2)}(0 \mid X_{it}, Z_{\mu,i})\right),$

$L = lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(X_{it}X_{it}^{T}f_v(0 \mid X_{it}, Z_{\mu,i})\right), M = lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}$
$\mathbb{E}\left(X_{it}f_v^{(3)}(0 \mid X_{it}, Z_{\mu,i})\right), \Phi = lim_{N\to\infty}(1/N)\sum_{i=1}^{N}\mathbb{E}\left(Z_{\mu,i}Z_{\mu,i}^{T}f_v^{(2)}(0 \mid X_{it}, Z_{\mu,i})\right), and \Psi = lim_{N,T\to\infty}$
$(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(\sum_{t=1}^{T}(Z_{\mu,i}X_{it}^{T}f_v^{(2)}(0 \mid X_{it}, Z_{\mu,i}))\right).$

The proofs of Theorems 2.1 and 2.2 are given in the Appendix B. The first term $(NTh_0^3)^{-1/2}$ in the convergence rates characterizes the magnitude of the estimation variance, while the second term h_0^2 characterizes the magnitude of the estimation bias. Comparing to the nonparametric univariate fixed effects mean regression, the convergence rate of *fixed effects modal regression* is much slower, which is the cost we need to pay in order to estimate the conditional mode (Parzen, 1962). Notice that Theorem 2.2 shows that the bias term can be successfully eliminated by allowing $NTh_0^7 \to 0$. However, the optimal bandwidth h_0 in Theorem 2.2 is $\hat{h}_0 \sim (NT)^{-1/7}$, which obviously does not satisfy the condition. This means we must undersmooth these types of estimators to remove the effect of asymptotic bias. The estimators can then converge to a normal distribution at a rate that close to $(NT)^{2/7}$. We incorporate this later when we adopt a plug-in method to choose the optimal bandwidth in practice.

Remark 2.4. *Besides undersmoothing, several other strategies (e.g., bootstrap) could be applied to remove the bias of estimators. For example, we could utilize a simple one-step bias correction based on the analytic form of the asymptotic bias. Then the one-step bias corrected estimator is defined as $\hat{\beta}^* = \hat{\beta} - h_0^2\hat{J}^{-1}\hat{M}/2$, where $\hat{J}$ and $\hat{M}$ are the estimates of J and M. Under the same conditions in Theorem 2.2, $\hat{\beta}^*$ has the limiting distribution with $\sqrt{NTh_0^3}(\hat{\beta}^* - \beta_0) \xrightarrow{d} \mathcal{N}(0, v_2J^{-1}LJ^{-1})$. Although it is interesting to compare the different bias correction methods, we would restrict our attention to the undersmoothing.*

Remark 2.5. *If we allow $T \to \infty$ with finite N, under the conditions C1-C5 ($(N,T) \to \infty$ in C3 is replaced with $T \to \infty$), as $h_0 \to 0$, $Th_0^5 \to \infty$, and $Th_0^7 = O(1)$, the consistent LDMR estimator of β has the asymptotic result $\sqrt{Th_0^3}(\hat{\beta}-\beta_0 - \frac{h_0^2}{2}J^{-1}M+o_p(h_0^2)) \xrightarrow{d} \mathcal{N}(0, v_2N^{-1}J^{-1}LJ^{-1})$.*

Remark 2.6. *Allowing $N \to \infty$ with fixed T, we could replace C1 with the condition that the random variables (Y_{it}, X_{it}^{T}) are i.i.d. over i and t. Together with the conditions C2-C5 ($(N,T) \to \infty$ in C3 is replaced with $N \to \infty$), as $h_0 \to 0$, $Nh_0^5 \to \infty$, and $Nh_0^7 = O(1)$, the consistent LDMR estimator of β has the asymptotic result $\sqrt{Nh_0^3}(\hat{\beta}-\beta_0 - \frac{h_0^2}{2}J^{-1}M+o_p(h_0^2)) \xrightarrow{d} \mathcal{N}(0, v_2T^{-1}J^{-1}LJ^{-1})$.*

2.3 Optimal Bandwidth

Theorems 2.1 and 2.2 imply that the bandwidth plays a crucial role in modal regression, which influences the estimation accuracy and governs the trade-off between bias and variance. Considering the estimator of β, it turns out that the corresponding asymptotic weighted mean squared errors $(Asy(WMSE))$ is

$$Asy(WMSE) = \mathbb{E}\left\{\left(\hat{\beta} - \beta_0\right)^T W \left(\hat{\beta} - \beta_0\right)\right\} \approx \frac{M^T J^{-1} W J^{-1} M h_0^4}{4} + \frac{v_2 \operatorname{tr}\left(J^{-1} L J^{-1} W\right)}{N T h_0^3},$$

$$(2.3)$$

where $tr(.)$ represents trace, and W is a weight matrix. The symbol "$a_{NT} \approx b_{NT}$" indicates that $a_{NT}/b_{NT} \to 1$ as $NT \to \infty$, i.e., $a_{NT} = b_{NT} + (s.o.)$, where $(s.o.)$ denotes the term that having probability order smaller than that of b_{NT}. By defining $\hat{h}_0 = \arg\min_{h_0} Asy(WMSE)$, we have the following corollary regarding the asymptotically optimal bandwidth.

Corollary 1. *Under the same conditions in Theorem 2.2, the optimal bandwidth h_0 that minimizes $Asy(WMSE)$ is*

$$\hat{h}_0 = \left(\frac{M^T J^{-1} W J^{-1} M}{3 v_2 tr(J^{-1} L J^{-1} W)}\right)^{-\frac{1}{7}} (NT)^{-\frac{1}{7}}.$$

Remark 2.7. *If $W = (J^{-1} L J^{-1})^{-1}$, which is proportional to the inverse of the asymptotic variance of $\hat{\beta}$, then $tr(J^{-1} L J^{-1} W) = q$. Additionally, it has been observed that the corresponding modal estimator requires undersmoothing to remove the effect of asymptotic bias (Kemp and Santos Silva, 2012). We then set $\hat{h}_0 = \left(\frac{M^T J^{-1} M}{3 v_2 q}\right)^{-\frac{1}{7}} (NT)^{-0.143}$, which is the default bandwidth we use in the numerical studies.*

The results in Corollary 1 or Remark 2.7 cannot be used directly due to the unknown function $f_v(.)$. To obtain the optimal bandwidth in practice, we propose a plug-in procedure for practical implementation. We follow Yao and Li (2014) to replace the unknown quantities in above corollary/remark by the corresponding estimates with the assumption of independence between v_{it} and X_{it}. Notice that similar argument has been put forward in the quantile regression literature; see Lee (2003). We then use the fixed effects mean regression to get the estimate of v_{it}, denoted by $\hat{v}_{it}$, and apply the nonparametric kernel density estimation method to obtain the mode of $\hat{v}$, say $\hat{v}_m$. We approximate $f_v^{(c)}(.)$ by

$$\hat{f}_v^{(c)}(0 \mid X_{it}, Z_{\mu,i}) \approx \frac{1}{N T h^{c+1}} \sum_{i=1}^{N} \sum_{t=1}^{T} K^{(c)} \left(\frac{\hat{v}_{it} - \hat{v}_m}{h}\right), c = 0, 2, 3, \qquad (2.4)$$

where $K(.)$ is a smooth kernel function satisfying the condition C4 in Section 2.2. Bandwidth h used in (2.4) is chosen using the method proposed by Botev et al. (2010).

3   Pseudo-Demodeing Two-Step Method

The above proposed LDMR method is very effective in practice when N is not too large. However, when N goes to infinity, there might exist multiple local maxima and it is not easy to

obtain the global optimal estimates with a large number of parameters needing to be estimated. To overcome this issue, we propose a PDTS method under certain conditions to concentrate out fixed effects through mean regression. Notice that throughout this section, the number of time periods is denoted by T_N that depends on N. In what follows, we omit the subscript N.

3.1 PDTS Modal Estimators

To motivate the proposed PDTS method and be consistent with the notations in above sections, we suppose that the fixed effects mean regression of Y_{it} given X_{it} follows

$$Y_{it} = X_{it}^T \beta_m + \alpha_i + \varepsilon_{it}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \tag{3.1}$$

where ε_{it} is i.i.d. over i and t with $\mathbb{E}(\varepsilon_{it} \mid X_{it}, \alpha_i) = 0$ but $Mode(\varepsilon_{it} \mid X_{it}, \alpha_i) \neq 0$, which is often the case in practice. In addition, we make an assumption that the individual effect from mean regression, α_i, remains a source of individual effect in modal regression; see a similar assumption in Canay (2011) for a quantile regression model for panel data. Thus, the fixed effects mean regression is $\mathbb{E}(Y_{it} - \alpha_i \mid X_{it}) = X_{it}^T \beta_m$, while the *fixed effects modal regression* considered from (1.1) or (2.1) will be $Y_{it} = X_{it}^T \beta + \mu_i + v_{it}$ with $Mode(v_{it} \mid X_{it}, \mu_i) = 0$, such that

$$Mode(Y_{it} - \alpha_i \mid X_{it}, \mu_i) = X_{it}^T \beta + \mu_i - \alpha_i = X_{it}^T \beta + \gamma_1, \tag{3.2}$$

where $\gamma_1 = \mu_i - \alpha_i$ could be interpreted as the modal adjustment factor for the individual effect from mean regression; for example, see a popular related location-scale shift model in Remark 3.2. These build the underlying mechanism of the PDTS method, where we compute a $\sqrt{T}$-consistent estimator $\hat{\alpha}_i$ from (3.1) firstly, and then estimate the following linear modal regression

$$Mode(\hat{Y}_{it} \mid X_{it}) = \tilde{X}_{it}^T \theta \tag{3.3}$$

with $\hat{Y}_{it} = Y_{it} - \hat{\alpha}_i$, $\tilde{X}_{it} = (1, X_{it}^T)^T$, and $\theta = (\gamma_1, \beta^T)^T$. The detailed description of the PDTS method is as follows.

- **Step 1**: Compute the $\sqrt{T}$-consistent estimator $\hat{\alpha}_i$ by Least Squares Dummy Variable (LSDV) approach, where

$$\hat{\alpha}_i = \bar{Y}_i - \bar{X}_i^T \hat{\beta}_m \tag{3.4}$$

in which $\hat{\beta}_m = (X^T Q_D X)^{-1} X^T Q_D Y$, $\bar{Y}_i = (1/T) \sum_{t=1}^T Y_{it}$, $\bar{X}_i = (1/T) \sum_{t=1}^T X_{it}$, $Q_D = I_{TN} - D (D^T D)^{-1} D^T$, $D = I_N \otimes l_T$, I_N is the $N \times N$ identity matrix, l_T stands for a T-vector of ones, and $\otimes$ denotes Kronecker product operation. Stacking the entire data set by individuals yields $X_i = (X_{i1}^T, \cdots, X_{iT}^T)^T$, $X = (X_1, \cdots, X_N)^T$, $Y_i = (Y_{i1}, \cdots, Y_{iT})^T$, and $Y = (Y_1^T, \cdots, Y_N^T)^T$.

10

- **Step 2**: Calculate the PDTS estimator, denoted by $\hat{\theta} = (\hat{\gamma}_1, \hat{\beta}^T)^T$, through maximizing the following kernel-based objective function

$$Q_{NT}(\theta) = \frac{1}{NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi \left(\frac{\hat{Y}_{it} - \tilde{X}_{it}^T \theta}{h_1} \right), \qquad (3.5)$$

where $\phi(.)$ is a standard normal kernel, and h_1 is a bandwidth. With the available estimate $\hat{\gamma}_1$, calculate the modal estimate of individual effect μ_i, where

$$\hat{\mu}_i = \hat{\gamma}_1 + \hat{\alpha}_i. \qquad (3.6)$$

Intuitively, the PDTS estimators work because $\hat{Y}_{it}$ weakly converges to $Y_{it}^* = Y_{it} - \alpha_{i0}$ as $T \to \infty$ in which α_{i0} is the true value of α_i. On one hand, in terms of computational simplicity or stability, the PDTS method may have an advantage. On the other hand, the linear form of fixed effects mean regression might be subject to misspecification. This issue needs to be researched further (e.g., applying nonparametric fixed effects mean regression), but beyond the scope of this paper. It is noticed that if $\gamma_1 = 0$, μ_i would be the same as α_i. In this case, the modal estimator of μ_i could be calculated as $\hat{\mu}_i = \bar{Y}_i - \bar{X}_i^T \hat{\beta}$ in the second step, which is like mean estimator but using modal estimator of $\hat{\beta}$. The maximizer of (3.5) can be computed easily by applying the modified MEM algorithm, which is shown in Algorithm 2.

Algorithm 2 MEM (Li et al., 2007; Yao, 2013)

E-Step. Calculate the weight $\pi \left(i, t \mid \theta^{(g)} \right)$ as

$$\pi \left(i, t \mid \theta^{(g)} \right) = \frac{\phi \left(\frac{\hat{Y}_{it} - \tilde{X}_{it}^T \theta^{(g)}}{h_1} \right)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \phi \left(\frac{\hat{Y}_{it} - \tilde{X}_{it}^T \theta^{(g)}}{h_1} \right)}.$$

M-Step. Update the value of $\theta^{(g+1)}$ by

$$\theta^{(g+1)} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \pi \left(i, t \mid \theta^{(g)} \right) \log \left(\frac{1}{h_1} \phi \left(\frac{\hat{Y}_{it} - \tilde{X}_{it}^T \theta}{h_1} \right) \right) \right\},$$

where g is the iteration indicator.

Remark 3.1. *For fixed effects modal regression, we do not have a general projection matrix used in fixed effects mean regression (Baltagi, 2009, 2013). We instead propose to use the estimate of α_i obtained from mean regression to concentrate out fixed effects. In addition, unlike fixed effects mean regression, we could not apply the first-difference or mean difference estimation method on fixed effects modal regression, due to the fact that the mode of error term may be changed after we implement difference transformation.*[3]

---

[3]For example, if we consider $Y_{it} = X_{it}^T \beta + \mu_i + v_{it}$ with $Mode(v_{it} \mid X_{it}, \mu_i) = 0$, applying the first-difference

Remark 3.2. *PDTS method has a main advantage compared to LDMR method in that the number of parameters to be estimated is greatly reduced, and thus can be used more effectively in practice with large N. However, the PDTS method does not intend to replace the LDMR method, but provide an alternative for researchers to simply estimate fixed effects modal regression under some practical scenarios. For example, one of the popular related models is the location-scale shift model, i.e., $Y_{it} = X_{it}^T \beta_m + \alpha_i + \varepsilon_{it}$, where $\varepsilon_{it} = (\gamma_1 + X_{it}^T \beta_{m'})\xi_{it}$. With assumptions that $Mode(\xi_{it} \mid X_{it}, \alpha_i) = 1$ and $\mathbb{E}(\xi_{it} \mid X_{it}, \alpha_i) = 0$, we have $\mathbb{E}(\varepsilon_{it} \mid X_{it}, \alpha_i) = 0$, $Mode(\varepsilon_{it} \mid X_{it}, \mu_i) = X_{it}^T \beta_{m'} + \gamma_1$, $\mathbb{E}(Y_{it} \mid X_{it}, \alpha_i) = X_{it}^T \beta_m + \alpha_i$, and $Mode(Y_{it} \mid X_{it}, \mu_i) = X_{it}^T \beta + \mu_i$, where $\beta = \beta_m + \beta_{m'}$ and $\mu_i = \alpha_i + \gamma_1$. We could then apply the PDTS method to eliminate α_i. Notice that the key identifying assumption is that fixed effects modal regression shares the common individual effect α_i with fixed effects mean regression.*

3.2 Asymptotic Properties

Similar to Theorems 2.1 and 2.2, we could obtain the asymptotic results regarding $\hat{\theta}$ with the restriction that at the optimal bandwidth convergence rate, $N^a/T \to 0$ for some $a > 4/3$ with $(N, T) \to \infty$. This is a sufficient condition to ensure the remainder term stemming from the first step estimator is negligible when we show the consistency and asymptotic normality of the PDTS estimators, which are presented as follows.

Theorem 3.1. *Under the regularity conditions C1-C5, as $(N, T) \to \infty$, $h_1 \to 0$, $NTh_1^5 \to \infty$, and $\sqrt{T}h_1^2 \to \infty$, there exists a consistent maximizer $\hat{\theta}$ of (3.5) such that with probability approaching one,*

$$\|\hat{\theta} - \theta_0\| = O_p\left(\left(NTh_1^3\right)^{-1/2} + h_1^2\right),$$

where $\theta_0 = (\gamma_{10}, \beta_0^T)^T$ is the true parameter of the fixed effects modal regression (3.3).

Theorem 3.2. *With $NTh_1^7 = O(1)$, under the same conditions in Theorem 3.1, the estimator $\hat{\theta}$ satisfying the consistency result in Theorem 3.1 has the following asymptotic result*

$$\sqrt{NTh_1^3}\left(\hat{\theta} - \theta_0 - \frac{h_1^2}{2}\tilde{J}^{-1}\tilde{M} + o_p(h_1^2)\right) \xrightarrow{d} \mathcal{N}\left(0, v_2\tilde{J}^{-1}\tilde{L}\tilde{J}^{-1}\right).$$

In addition, with a further condition that $NTh_1^7 \to 0$, we have

$$\sqrt{NTh_1^3}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, v_2\tilde{J}^{-1}\tilde{L}\tilde{J}^{-1}\right),$$

where $\tilde{J} = \lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left(\tilde{X}_{it}\tilde{X}_{it}^T f_v^{(2)}(0 \mid \tilde{X}_{it})\right)$, $\tilde{L} = \lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left(\tilde{X}_{it}\tilde{X}_{it}^T f_v(0 \mid \tilde{X}_{it})\right)$, and $\tilde{M} = \lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left(\tilde{X}_{it} f_v^{(3)}(0 \mid \tilde{X}_{it})\right)$.

---

transformation on equation yields $Y_{it} - Y_{it-1} = (X_{it}^T - X_{it-1}^T)\beta + v_{it} - v_{it-1}$ in which we cannot guarantee $Mode(v_{it} - v_{it-1} \mid X_{it}) = 0$. The same problem arises if we apply the mean difference transformation.

We outline the proofs of Theorems 3.1 and 3.2 in the Appendix B. Theorem 3.2 indicates that the asymptotic distribution of the estimator is the same as it would be if α_i were known with certainty. With $NTh_1^7 \to 0$, the proposed estimator can achieve the convergence rate of $O_p(\sqrt{NTh_1^3})$. Note that the optimal rate of h_1 is $(NT)^{-1/7}$, which obviously does not satisfy the condition $NTh_1^7 \to 0$. This means we can apply the undersmoothing method to remove the asymptotic bias in the second step; see Remark 2.4. The restriction $\sqrt{T}h_1^2 \to \infty$ implies $N^a/T \to 0$ for some $a > 4/3$ with the optimal bandwidth. It indicates that T grows much faster than N, which is used to simplify the derivation of the asymptotic results; see the similar condition used in Lamarche (2010) and Galvao (2011). The intuition behind this restriction is that T must go to infinity fast enough to guarantee the consistent estimates for the fixed effects in the first step, and then for the parameters in the second step with the optimal bandwidth. It is notice that the restriction $N^a/T \to 0$ for some $a > 4/3$ rules out the case for which N and T pass to infinity at the same rate, i.e., $N/T \to c \in (0, \infty)$, as in some empirical applications. We leave the asymptotic properties under such case for the future research. Large T is not common in the economic panel dataset, thus the Monte Carlo simulations presented in Section 4 assess the finite sample performance of the estimators and show evidence that the bias is small for moderate T.

Remark 3.3. *If we allow $T \to \infty$ with finite N, under the conditions C1-C5 ($(N, T) \to \infty$ in C3 is replaced with $T \to \infty$), as $h_1 \to 0$, $Th_1^5 \to \infty$, and $Th_1^7 = O(1)$, θ_0 could be consistently estimated with $\hat{\theta}$ such that $\sqrt{Th_1^3}(\hat{\theta} - \theta_0 - \frac{h_1^2}{2}\tilde{J}^{-1}\tilde{M}) \xrightarrow{d} \mathcal{N}(0, v_2 N^{-1}\tilde{J}^{-1}\tilde{L}\tilde{J}^{-1})$. For this case, $\sqrt{T}h_1^2 \to \infty$ is obviously satisfied with the optimal bandwidth.*

Notice that

$$|\hat{\mu}_i - \mu_{i0}| \le |\hat{\gamma}_1 - \gamma_{10}| + |\hat{\alpha}_i - \alpha_{i0}|, \tag{3.7}$$

which indicates that the bias of $\hat{\mu}_i$ has two terms. The first term $|\hat{\gamma}_1 - \gamma_{10}|$ is the standard term in modal regression, and the second term $|\hat{\alpha}_i - \alpha_{i0}|$ captures the fact that α_i is being estimated by $\hat{\alpha}_i$. The boundedness of these two terms is sufficient for the boundedness of $|\hat{\mu}_i - \mu_{i0}|$. With the condition $\sqrt{T}h_1^2 \to \infty$, one can show that the leading bias term for the estimator $\hat{\mu}_i$ streams from $|\hat{\gamma}_1 - \gamma_{10}|$, while the error term stemming from the fixed effect transformation can be ignored. We then characterize the asymptotic behavior of the estimator $\hat{\mu}$.

Theorem 3.3. *Under the same conditions in Theorem 3.1, we have*

$$\|\hat{\mu} - \mu_0\| = O_p\left(\left(NTh_1^3\right)^{-1/2} + h_1^2\right).$$

Remark 3.4. *For large T with finite N case, under the same conditions in Remark 3.3, we can also prove that $\|\hat{\mu} - \mu_0\| = O_p((Th_1^3)^{-1/2} + h_1^2)$.*

3.3   Optimal Bandwidth

Exploiting the structure of the asymptotic distribution of the PDTS estimator, we can now obtain the asymptotically optimal bandwidth by minimizing the $Asy(WMSE)$ for $\hat{\theta}$, where

$$Asy(WMSE) = \mathbb{E}\left\{\left(\hat{\theta} - \theta_0\right)^T \tilde{W}\left(\hat{\theta} - \theta_0\right)\right\} \approx \frac{\tilde{M}^T \tilde{J}^{-1} \tilde{W} \tilde{J}^{-1} \tilde{M} h_1^4}{4} + \frac{v_2 \operatorname{tr}\left(\tilde{J}^{-1} \tilde{L} \tilde{J}^{-1} \tilde{W}\right)}{NT h_1^3} \quad (3.8)$$

in which $\tilde{W}$ is a diagonal weight matrix. Similar to the results in subsection 2.3, by defining $\hat{h}_1 = argmin_{h_1} Asy(WMSE)$, we have the following corollary.

Corollary 2. *Under the same conditions in Theorem 3.2, the optimal bandwidth h_1 that minimizes $Asy(WMSE)$ is*

$$\hat{h}_1 = \left(\frac{\tilde{M}^T \tilde{J}^{-1} \tilde{W} \tilde{J}^{-1} \tilde{M}}{3 v_2 tr(\tilde{J}^{-1} \tilde{L} \tilde{J}^{-1} \tilde{W})}\right)^{-\frac{1}{7}} (NT)^{-\frac{1}{7}}.$$

Remark 3.5. *If $\tilde{W} = (\tilde{J}^{-1} \tilde{L} \tilde{J}^{-1})^{-1}$, which is proportional to the inverse of the asymptotic variance of $\hat{\theta}$, then $\hat{h}_1 = \left(\frac{\tilde{M} \tilde{J}^{-1} \tilde{M}}{3 v_2 (q+1)}\right)^{-\frac{1}{7}} (NT)^{-\frac{1}{7}}$. This result holds for the case of either $(N, T) \to \infty$ or $T \to \infty$. To obtain the optimal bandwidth suggested by data, similar to the LDMR method, we propose a simple plug-in method by replacing the unknown quantities with estimates and work with the undersmoothing assumption on the bandwidth by following Kemp and Santos Silva (2012) to set $\hat{h}_1 \sim (NT)^{-0.143}$; see Remark 2.7. The final default bandwidth used in the numerical studies for the PDTS method is $\hat{h}_1 = \left(\frac{\tilde{M} \tilde{J}^{-1} \tilde{M}}{3 v_2 (q+1)}\right)^{-\frac{1}{7}} (NT)^{-0.143}$. Notice that in practice the above bandwidth selection procedure can be iterated; see Yao and Li (2014).*

4   Numerical Examples

In this section, we first carry out simulation studies to illustrate how *fixed effects modal regression* works for finite sample, and then apply the proposed model and methods to analyze a public capital productivity dataset and a carbon dioxide (CO_2) emissions dataset. Throughout this section, we suppress the words "fixed effects" for regressions whenever no confusion is caused.

4.1   Monte Carlo Experiments

Two Monte Carlo experiments with different skewed error terms are conducted to illustrate the finite sample performance of the *fixed effects modal regression*. In what follows, we use DGP to represent the data generating process. We consider using both the LDMR method and the

PDTS method to estimate modal regression for these two experiments with $M = 200$ replicates, and compare the results with fixed effects mean regression estimated by the LSDV method. We examine the finite sample performance of estimators in terms of bias, standard error, and mean squared error (MSE), where

$$MSE(\hat{\beta}) = \frac{1}{M} \sum_{j=1}^{M} \|\hat{\beta}^{(j)} - \beta\|^2$$

in which $\hat{\beta}^{(j)}$ is the estimate in the jth replication. We also present the shape of the empirical density of the standardized modal estimate to check the asymptotic normality property, and report the coverage probabilities to evaluate the prediction performance of the proposed model.

DGP 1: We generate the following *fixed effects modal regression* with skewed error term

$$Y_{it} = 2X_{it} + \mu_i + \sigma(X_{it})v_{it}, \tag{4.1}$$

where the individual effect μ_i is drawn from $U[0,1]$. X_{it} and v_{it} are independent with $v_{it} \sim 0.5N(-1, 2.5^2) + 0.5N(1, 0.5^2)$ (Yao and Li, 2014). Note that X_{it} and μ_i are correlated with $X_{it} = 0.5\mu_i + Z_{it}$, where $Z_{it} \sim U[-1,1]$. All data are generated i.i.d. across individuals and over time. We consider three cases, where in case 1 we define $\sigma(X_{it}) = X_{it}$, in case 2 we let $\sigma(X_{it}) = 0.2 + X_{it}$, and case 3 is with $\sigma(X_{it}) = 0.2$. As $\mathbb{E}(v_{it}) = 0$ and $Mode(v_{it}) = 1$, we have the following equations

$$\text{Case 1}: \begin{cases} \text{Mean Regression:} & \mathbb{E}(Y_{it} \mid X_{it}, \mu_i) = 2X_{it} + \mu_i, \\ \text{Modal Regression:} & Mode(Y_{it} \mid X_{it}, \mu_i) = 3X_{it} + \mu_i. \end{cases}$$

$$\text{Case 2}: \begin{cases} \text{Mean Regression:} & \mathbb{E}(Y_{it} \mid X_{it}, \mu_i) = 2X_{it} + \mu_i, \\ \text{Modal Regression:} & Mode(Y_{it} \mid X_{it}, \mu_i) = 3X_{it} + \mu_i + 0.2. \end{cases}$$

$$\text{Case 3}: \begin{cases} \text{Mean Regression:} & \mathbb{E}(Y_{it} \mid X_{it}, \mu_i) = 2X_{it} + \mu_i, \\ \text{Modal Regression:} & Mode(Y_{it} \mid X_{it}, \mu_i) = 2X_{it} + \mu_i + 0.2. \end{cases}$$

These indicate that modal regression and mean regression are different when we have the skewed data. We consider the simulation of combinations of $N = 5, 10, 20, 50, 100$ and $T = 20, 40, 60, 100, 200$. The results are summarized in Tables 1-3 containing the estimates, the standard errors (in parentheses), and the $MSEs$, where the bold number represents the smallest value of MSE for each combination of N and T. It can be seen from from Tables 1-3 that both of the two proposed methods could estimate the modal regression well with the finite samples. For cases 1-2, the LDMR estimator is slightly biased for small T, but there are substantial improvements with an increase in T. The results for the PDTS estimator also show small biases for cases 1-2, which become smaller as T increases. For case 3, when the error term

15

is independent of the covariate, the proposed methods in this paper work well even with small T, and the PDTS method outperforms the LDMR method.

In addition, from Tables 1-3, it can be seen that the number of individuals N turns out not to have much effect on the performance of the modal estimates in terms of the bias when we keep T fixed. As expected, modal regression has smaller standard errors of regression coefficients than those of mean regression due to the robustness characteristic of modal regression with the skewed errors. Moreover, the simulation results indicate that no estimation technique is superior for all combinations of N and T. However, for moderate T with the heteroskedasticity errors, the LDMR estimator slightly outperforms the PDTS estimator, which suggests that the LDMR estimator would rank high as the preferred choice for applied econometricians.

To consider the asymptotic normality property of the modal estimator, we compare the shape of the empirical density of the standardized modal estimate to that of the standard normal density. Due to space limitations, we only report results for $N = 5$ or 50 in Figure 1. The results for the other sample size schemes are similar. In accordance with our theoretical findings, the performance of the asymptotic normality approximation increases when T increases.

(a) DGP 1-Case 1

(b) DGP 1-Case 2

(c) DGP 1-Case 3
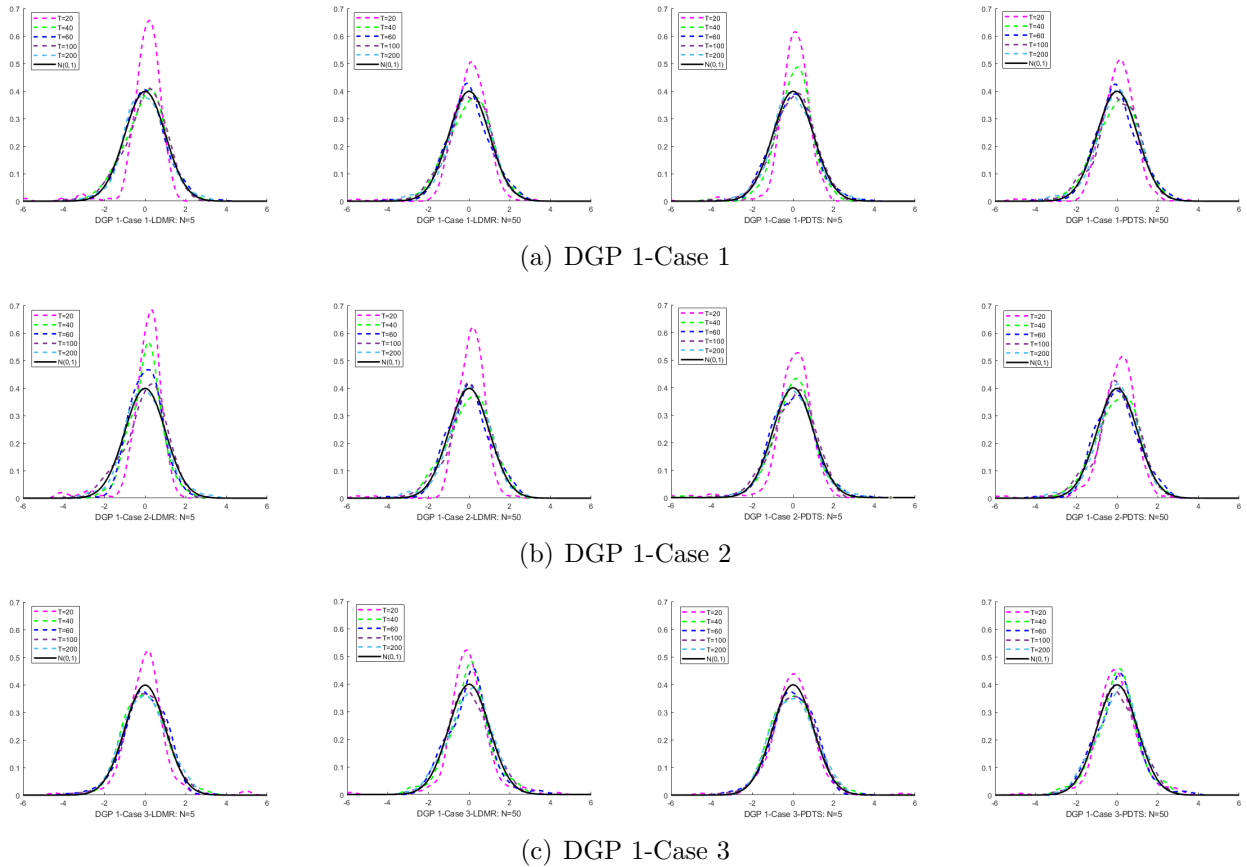
Figure 1: Empirical density of the standardized estimate: for each case, the first two columns are for the LDMR method, while the last two columns are for the PDTS method.
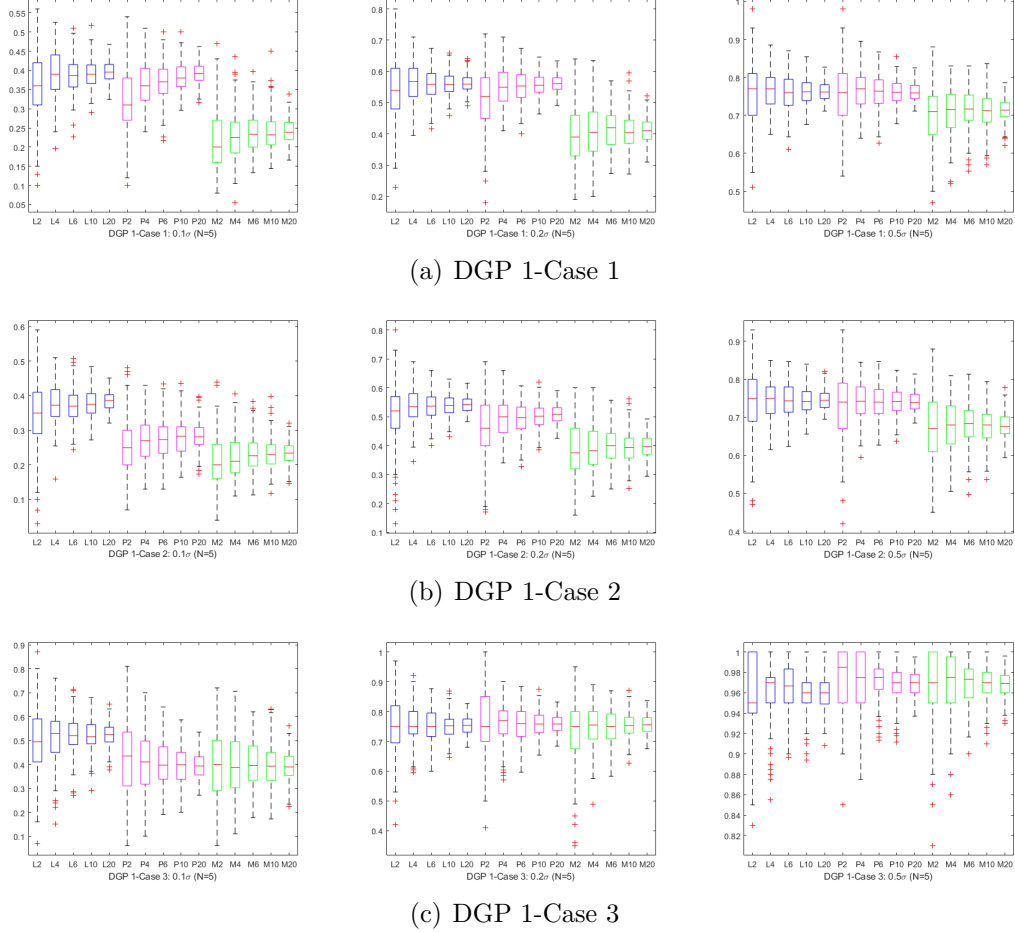
Figure 2: Boxplots of average of coverage probabilities: for each case, the plots represent the results for the interval lengths of 0.1σ, 0.2σ, and 0.5σ, respectively. For each plot, L means LDMR method, P is for PDTS method, M indicates mean regression, and the numbers 2, 4, 6, 10, 20 represent the values of $T = 20, 40, 60, 100, 200$, respectively.

It is noticed that most econometric forecasting has focused on models for the conditional mean or quantile. The proposed model in this paper offers an opportunity to expand the scope of forecasting applications based on modal regression. To assess the prediction performance of modal regression, we follow Yao and Li (2014) to report the coverage probabilities of prediction intervals of three different lengths (0.1σ, 0.2σ, 0.5σ, $\sigma = \sqrt{\text{Var}(v_{it})} \approx 2$). We follow the same DGP process as above three cases but implement the out-of-sample prediction for the additional NT data points with 200 repetitions. Specifically, we use T period data for each individual to estimate the model, and then use the resulting model to predict the value of the response variable for the additional T period points for each individual. Representative results of the coverage probabilities of the proposed model and mean regression model are reported in Figure 2, where each plot shows the coverage probabilities across different values of T for a given combination of σ and case. For space consideration, we only list the results for $N = 5$. All plots indicate that modal regression provides higher coverage probabilities compared to mean regression due

to the skewness of the error distribution. With the increase of the interval length, the modal regression and mean regression will provide closer coverage probabilities as expected.

DGP 2: In order to illustrate how LDMR method and PDTS method perform and compare to mean regression under different levels of skewness of density, we consider the following model setting with

$$Y_{it} = X_{it} + \mu_i + \sigma(X_{it})v_{it}, \tag{4.2}$$

where $\mu_i \sim U[0,1]$, $\sigma(X_{it}) = X_{it}$, and $v_{it} \sim 0.5Ga(k_1, \theta) + 0.5Ga(k_2, \theta)$ in which Ga represents the gamma distribution, $k_j \in \mathbb{N}_{>0}, j = 1, 2$, is the shape parameter that can adjust the skewness of v_{it} (coefficient of skewness$= \sqrt{4/k}$), and $\theta \in \mathbb{N}_{>0}$ is the scale parameter. To create the correlation between the regressor and the individual effect, we let $X_{it} = 0.5\mu_i + Z_{it}$, where $Z_{it} \sim U[-1,1]$. We employ two different schemes to generate the distributions of v_{it}, where $k_1 = 1$ or 7, and $k_2 = 2$. In both schemes, we set $\theta = 0.5$. These two different error distributions cover some interesting cases, where the first one is the more skewed distribution while the second one represents the less skewed distribution. Note that $\mathbb{E}(v_{it}) = 0.5(k_1 + k_2)\theta$ and $Mode(v_{it}) = 0.5(k_1 + k_2 - 1)\theta$,[4] we then have

$$\text{More Skewed}: \begin{cases} \text{Mean Regression:} & \mathbb{E}(Y_{it} \mid X_{it}, \mu_i) = 1.75X_{it} + \mu_i, \\ \text{Modal Regression:} & Mode(Y_{it} \mid X_{it}, \mu_i) = 1.5X_{it} + \mu_i. \end{cases}$$

$$\text{Less Skewed}: \begin{cases} \text{Mean Regression:} & \mathbb{E}(Y_{it} \mid X_{it}, \mu_i) = 3.25X_{it} + \mu_i, \\ \text{Modal Regression:} & Mode(Y_{it} \mid X_{it}, \mu_i) = 3X_{it} + \mu_i. \end{cases}$$

The simulation of combinations of $N = 5, 10, 20, 50, 100$ and $T = 20, 40, 60, 100, 200$ is conducted. The results of more skewed and less skewed settings are shown in Tables 4-5, respectively, containing the estimates, the standard errors (in parentheses), and the $MSEs$, where the bold number represents the smallest value of MSE for each simulation combination. It can be seen that the performances of all the modal estimates are satisfactory even for small T. Note that modal estimators have relatively large biases when T is small and improve significantly as T increases. However, similar to DGP 1, the bias of the modal estimators is not affected as N increases when T remains fixed. When the dataset is generated from the more skewed setting, modal regression could have smaller standard errors of coefficients than those of mean regression in this experiment and is a good complement to the mean regression, which are similar to the findings in DGP 1. Although the LDMR method performs slightly better than the PDTS method, there is no substantial difference between these two. This may be due to the underlying data generating process or the characteristics of the distribution. It is worth noticing that when the dataset is from the less skewed setting, modal regression provides similar

---

[4]If $X \sim Ga(\alpha, \theta)$ and $Y \sim Ga(\beta, \theta)$ are independently distributed with the same scale parameter, then $X + Y$ follows $Ga(\alpha + \beta, \theta)$ with variance $(\alpha + \beta)\theta^2$.

regression estimates to mean regression but with less accuracy.

We report the shape of the empirical density of parameter estimate and the coverage probabilities in the same way as DGP 1. Figure 3 indicates that the whole distribution of the standardized modal estimate converges to the standard normal distribution as T increases. Figure 4 shows that the modal estimator has better coverage probabilities than the mean regression estimator, which is entirely expected because of the skewed distribution of error terms.

(a) DGP 2-More Skewed

(b) DGP 2-Less Skewed
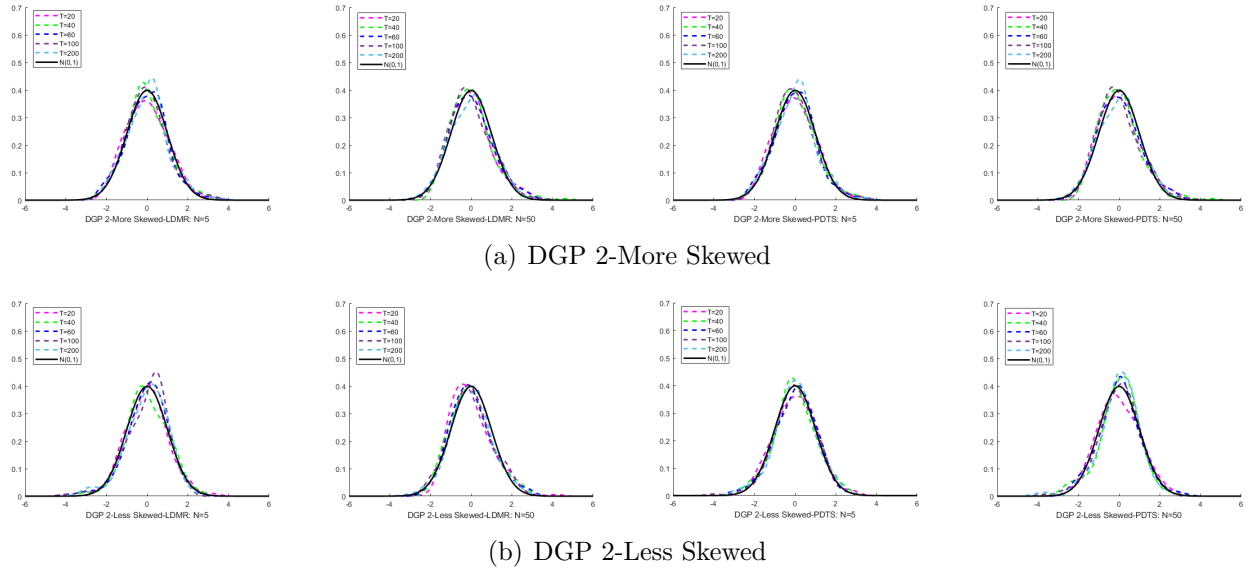
Figure 3: Empirical density of the standardized estimate: for each case, the first two columns are for the LDMR method, while the last two columns are for the PDTS method.

(a) DGP 2-More Skewed

(b) DGP 2-Less Skewed
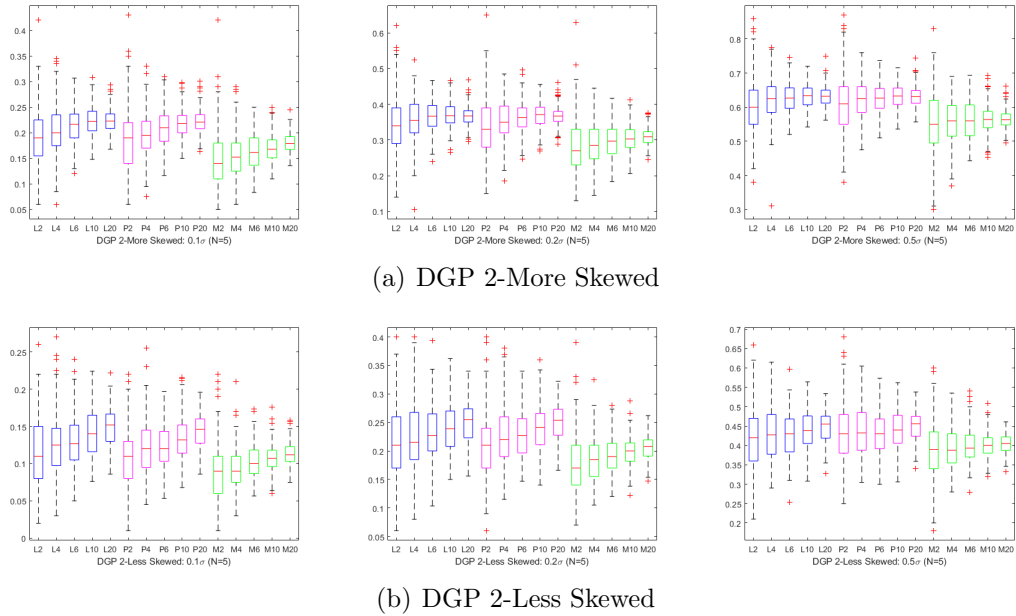
Figure 4: Boxplots of average of coverage probabilities: for each case, the plots represent the results for interval lengths of 0.1σ, 0.2σ, and 0.5σ (more skewed: $\sigma \approx 0.4330$; less skewed: $\sigma = 0.7500$), respectively. The notations in each plot are the same as those of Figure 2.

4.2 Real Data Examples

4.2.1 Public Capital Productivity Study

In order to illustrate the applicability of *fixed effects modal regression*, we follow Baltagi and Pinnoi (1995) to consider a dataset containing the record of the U.S. 48 contiguous states over the period 1970-1986 to revisit the relationship between public capital and private sector output (Munnell, 1990; Baltagi and Pinnoi, 1995; Su et al., 2013; Henderson and Ullah, 2014). The main question related to this dataset is "Is public-sector capital productive? What is the role of public-sector in affecting private economic performance ?" (Su et al., 2013). Munnell (1990) and Su et al. (2013) found that the public capital plays a positive and significant role in affecting the private-sector output, while some other researchers argued that the public capital has significant and negative effects on private productivity (Evans and Karras, 1994). Baltagi and Pinnoi (1995) claimed that public capital does not have significant effect on private productivity. However, all of these research is based on mean regression. Next, we use the following *fixed effects modal regression* to investigate this public capital dataset

$$log(Y_{it}) = \mu_i + \beta_1 log(KG_{it}) + \beta_2 log(KPR_{it}) + \beta_3 log(L_{it}) + \beta_4 UNEM_{it} + v_{it}, \qquad (4.3)$$

where $i = 1, \cdots, 48$ indicates the state and $t = 1, \cdots, 17$ indicates the year. To simplify the notation, henceforth, we drop the subscripts from the names of the covariates. Y denotes the GDP, KG represents the public capital, KPR is the private capital stock, L is the employment, $UNEM$ represents the unemployment rate used to control for business cycle effects, μ is the individual effect, and v is the disturbance term.

We estimate (4.3) via the linear fixed effects mean regression firstly. The regression coefficients with respect to KG, KPR, L, and $UNEM$ are -0.0261 (0.0281), 0.2920*** (0.0243), 0.7682*** (0.0291), and -0.0053*** (0.0010), respectively, which are consistent with the results in Baltagi and Pinnoi (1995). Note that the values in parentheses are standard errors, and asterisks indicate significance levels.[5] Then we use the proposed methods in this paper to estimate (4.3). Following Su et al. (2013), in order to obtain the standard errors of modal estimates, we use a bootstrap method to resample data 200 times across individuals while keeping the time series structure for each individual unchanged, and then calculate the bootstrap standard errors for the modal estimates; see Figure 5. The slope coefficients for the LDMR method with respect to KG, KPR, L, and $UNEM$ are -0.0276 (0.0234), 0.2421*** (0.0198), 0.8245*** (0.0246), and -0.0035*** (0.0008), respectively, and the corresponding estimates of the PDTS method are -0.0096 (0.0136), 0.2882*** (0.0162), 0.7575*** (0.0153), and -0.0039*** (0.0014), respectively. Table 6 in the Appendix A summarizes all the results.

---

[5]*: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$.

For this public capital productivity dataset, both the modal estimate and mean estimate indicate that there exists a negative relationship between public capital and GDP, but the effect is not statistically significant. In other words, the public capital has no spillover effect on average or on mode across states, which is consistent with most results obtained from parametric mean regression. For this example, it is clear that the estimated results of the modal regression are not much different from these obtained by mean regression, which could be partially explained by the almost symmetric distribution of $log(Y)$ (Figure 5).
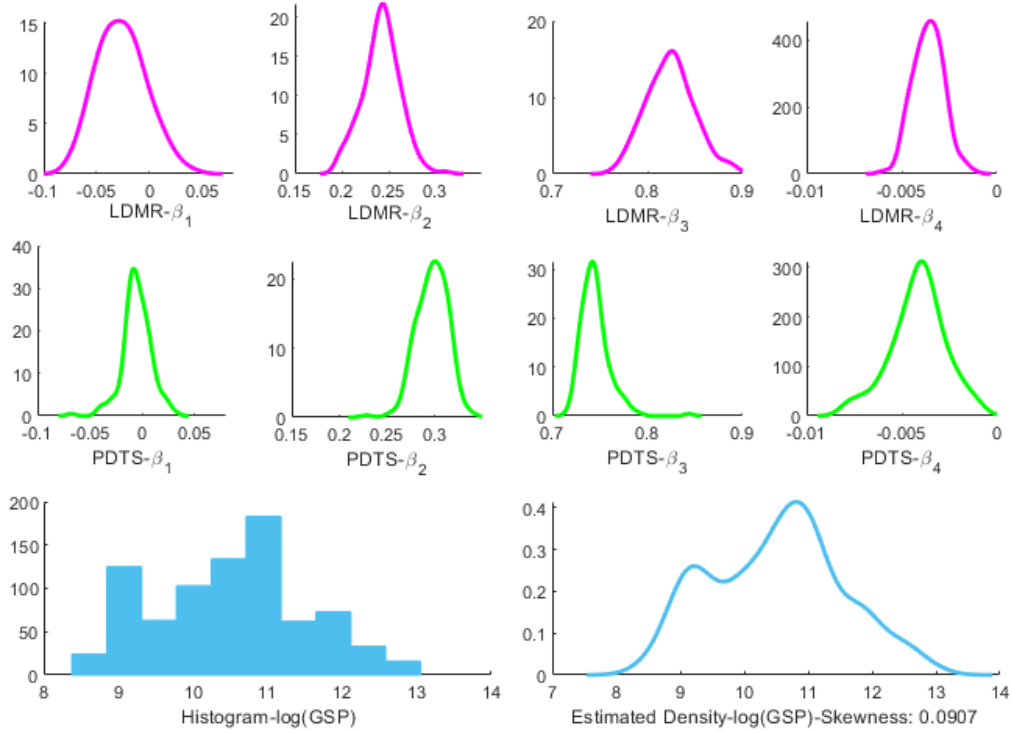
Figure 5: Bootstrap density of coefficients and estimated density of $log(Y)$

4.2.2 CO_2 Emissions Study

In this subsection, we apply the proposed methods in this paper to a U.S. fossil fuel CO_2 emissions dataset from Fragkias et al. (2013), which investigated the relationship between the scaling of CO_2 emissions and U.S. urban areas. We focus the study on 60 urban areas from the U.S. over the period 1999-2008, which is a subsample from Fragkias et al. (2013). The sample data comprises annual measures of CO_2 emissions and three inputs (population, urban population density (density), and per capita personal income (pcpi)). As is conventional in the modern urban economies literature, we follow Fragkias et al. (2013) to use the following model specification to examine the relationship between population size of cites and CO_2 emissions

$$log(CO_{2,it}) = \mu_{i/t} + \beta_1 log(Population_{it}) + \beta_2 log(Density_{it}) + \beta_3 log(PCPI_{it}) + v_{it}, \quad (4.4)$$

where $i = 1, \cdots, 60$ denotes U.S. urban area i, $t = 1, \cdots, 10$ denotes the tth year period, $\mu_{i/t}$ represents the individual effect or time effect, and v_{it} is the disturbance term.

Firstly, we control the individual effect μ_i. The results of estimating (4.4) through the linear fixed effects mean regression are 2.7498*** (1.0391), -1.7751^* (0.9932), and -0.0744 (0.0678), respectively. The coefficient of $log(Population)$ is statistically significant, which indicates that a 1% increase in population size is associated with an increase in CO_2 emissions of 274.98% on average. However, a 1% increase in the population density is associated with an average of 177.51% reduction in CO_2 emissions. It is noted that the coefficient of $log(PCPI)$ from mean regression is not statistically significant.
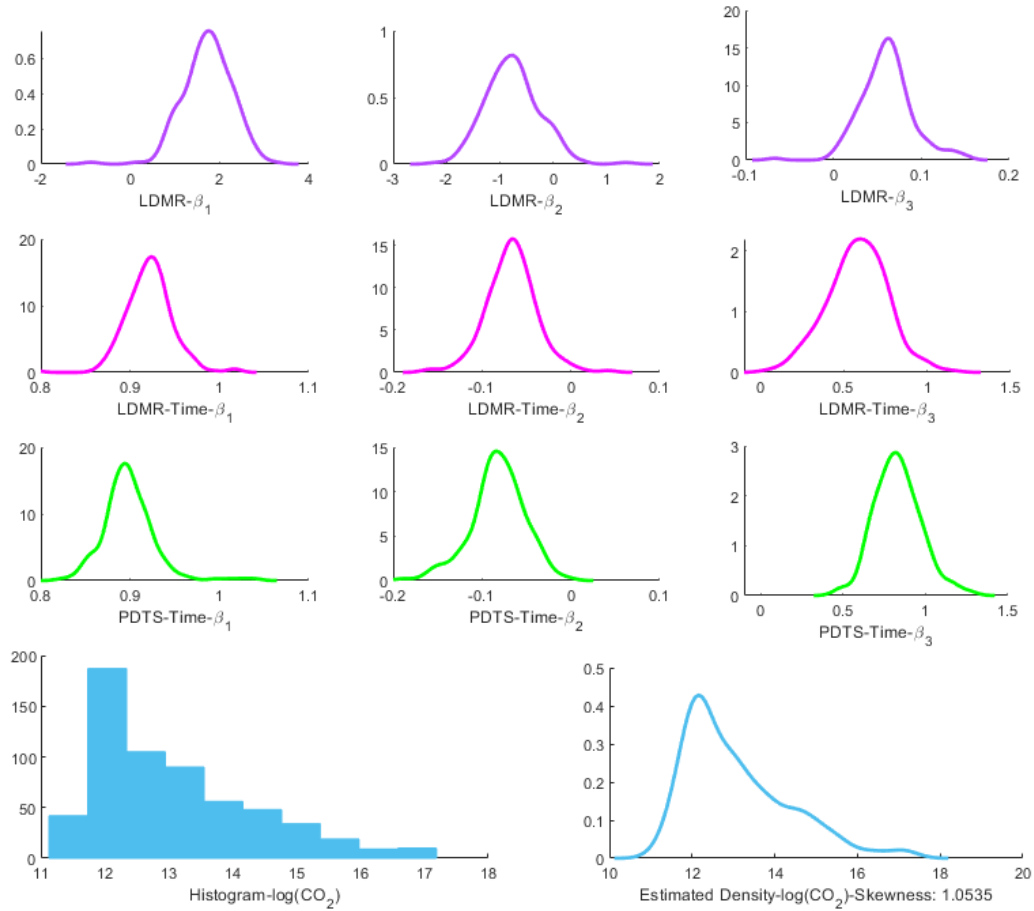
Figure 6: Bootstrap density of coefficients and estimated density of $log(CO_2)$

We then turn to the estimates from *fixed effects modal regression*. Due to the small size of T in this case, we only apply the LDMR method when we control the individual effects, where the estimates are 1.7350*** (0.5022), -0.7581 (0.4779), and 0.0609** (0.0316), respectively. The values in parentheses are the bootstrap estimates of the standard errors (Figure 6). Interestingly, compared to mean regression, the coefficient estimates for $log(Population)$ and $log(Density)$ based on modal regression are decreased by 36.90% and 57.29%, respectively.

Note that the coefficient of $log(Density)$ is not statistically significant on mode, which may be due to the "overshadowed effects of the size of the metropolitan area" (Fragkias et al., 2013). In sharp contrast to what was found with mean regression, the coefficient of $log(PCPI)$ becomes significantly positive according to the modal regression. It indicates a 1% increase in the per capital personal income is associated with a 6.09% increase in CO_2 emissions based on the mode value. The underlying mechanism is the associated positive relationship among $PCPI$, higher production, and pollution.

We control the time effect μ_t to estimate (4.4) as well. The results for the linear fixed effects mean regression are 1.0489*** (0.0281), -0.2512*** (0.0330), and 0.8526*** (0.1902), respectively. The results for the LDMR method are 0.9235*** (0.0262), -0.0675** (0.0281), and 0.5750*** (0.1788), respectively, while the results for the PDTS method are 0.9012*** (0.0292), -0.0827** (0.0322), and 0.7592*** (0.1396), respectively. All coefficients are statistically significant. The signs of all estimates form these three methods are consistent with each other, while the magnitudes of the effects at mean and mode differ considerably, especially for the variable $log(Density)$. Compared to mean regression, the estimates of $log(Density)$ on mode are decreased by 73.13% (LDMR) and 67.08% (PDTS), respectively. Tables 7-8 in the Appendix A summarize all the results. Overall, these results indicate that *fixed effects modal regression* can provide complement information to fixed effects mean regression on how the "most likely values" of the dependent variable are affected by the regressors.

5   Concluding Remarks

Mode has not received much attention for a long time. With more available datasets and powerful computation tools, it is important for (applied) econometricians to be aware of the application of modal regression, which focuses on modeling how the conditional mode of the response variable depends on the covariates since mean regression cannot reveal the whole characteristics of the dataset, especially for the skewed data. To this end, we introduce a new model named *fixed effects modal regression* for the panel data, in which the new model uses the conditional mode instead of mean to model the relationship among variables. Two estimation methods are proposed, where we call linear dummy modal regression (LDMR) method and pseudo-demodeing two-step (PDTS) method. We discuss the asymptotic properties of the modal estimators under different settings of N and T. Monte Carlo simulation results and two applications indicate that the proposed two estimation methods work well, and the newly proposed modal regression can be a good complement to mean and quantile regressions.

Similar to many kernel-based methods, how to choose the bandwidth in modal regression requires further investigation; for example, the cross-validation method may be considered. We could also follow Kemp and Santos Silva (2012) to use a sequence of bandwidths to see

different *zoom levels* of the modal regression estimation. In this paper, we assume that the conditional density of v_{it} has a well-defined global mode at 0. However, when the population is not homogeneous, our method could also be applied to the multimode setting. For this setting, our method will find multiple modal solutions if starting from multiple initial values. Each modal solution corresponds to one local modal estimator. This issue can be further studied along the lines of the mode clustering (Chen et al., 2016), which could be a very interesting topic for future study.

Compliance with Ethical Standards

References

Baltagi, B. (2009). A Companion to Econometric Analysis of Panel Data. *John Wiley and Sons, Chichester*.

Baltagi, B. (2013). Econometrics Analysis of Panel Data (5th ed.). *John Wiley and Sons, Chichester*.

Baltagi, B. and Pinnoi, N. (1995). Public Capital Stock and State Productivity Growth: Further Evidence From An Error Components Model. *Empirical Economics*, 20, 351-359.

Botev, Z. I., Grotowski, J. F., and Kroese, D. P. (2010). Kernel Density Estimation via Diffusion. *The Annals of Statistics*, 38, 2916-2957.

Canay, I. A. (2011). A Simple Approach to Quantile Regression for Panel Data. *The Econometrics Journal*, 14, 368-386.

Chen, Y. C. (2018). Modal Regression using Kernel Density Estimation: A Review. *Wiley Interdisciplinary Reviewers: Computational Statistics*, 10, e1431.

Chen, Y. C., Genovese, C. R., Tibshirani, R. J., and Wasserman, L. (2016). Nonparametric Modal Regression. *The Annals of Statistics*, 44 (2), 489-514.

Eddy, W. F. (1980). Optimum Kernel Estimators of the Mode. *The Annals of Statistics*, 8 (4), 870-882.

Evans, P. and Karras, G. (1994). Are Government Activities Productive? Evidence From A Panel of U.S. States. *Review of Economic Studies*, 76 (1), 1-11.

Fragkias, M., Lobo, J., Strumsky, D., and Seto, K. C. (2013). Does Size Matter? Scaling of CO_2 Emissions and U.S. Urban Areas. *PLOS ONE*, 8 (6), e64727.

Friedman, J. H. and Fisher, N. I. (1999). Bump Hunting in High-Dimensional Data. *Statistics and Computing*, 9, 123-143.

Galvao, A. F. (2011). Quantile Regression for Dynamic Panel Data with Fixed Effects. *Journal of Econometrics*, 164 (1), 142-157.

Gao, Y. and Li, K. (2013). Nonparametric Estimation of Fixed Effects Panel Data Models. *Journal of Nonparametric Statistics*, 25 (3), 679-693.

Henderson, D., Carroll, R., and Li, Q. (2008). Nonparametric Estimation and Testing of Fixed Effects Panel Data Models. *Journal of Econometrics*, 144 (1), 257-275.

Henderson, D. and Ullah, A. (2014). Nonparametric Estimation in A One-Way Error Component Model: A Monte Carlo Analysis. *Statistical Paradigms*, 213-237.

Hidalgo, F. J. (1992). Adaptive Semiparametric Estimation in the Presence of Autocorrelation of Unknown Form. *Journal of Time Series Analysis*, 13, 47-78.

Kemp, G. C. R., Parente, P. M. D. C., and Santos Silva, J. M. C. (2019). Dynamic Vector Mode Regression. *Journal of Business & Economic Statistics*, 38 (3), 647-661.

Kemp, G. C. R. and Santos Silva, J. M. C. (2012). Regression towards the Mode. *Journal of Econometrics*, 170 (1), 92-101.

Krief, J. M. (2017). Semi-Linear Mode Regression. *Econometrics Journal*, 20, 149-167.

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

Lee, M. (1989). Mode Regression. *Journal of Econometrics*, 42, 337-349.

Lee, M. (1993). Quadratic Model Regression. *Journal of Econometrics*, 57, 1-19.

Lee, S. (2003). Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model. *Econometric Theory*, 19 (1), 1-31.

Lee, Y., Mukherjee, D., and Ullah, A. (2019). Nonparametric Estimation of the Marginal Effect in Fixed-Effect Panel Data Models. *Journal of Multivariate Analysis*, 171, 53-67.

Li, X. and Huang, X. (2019). Linear Mode Regression with Covariate Measurement Error. *Canadian Journal of Statistics*, 47, 262-280.

Li, J., Ray, S., and Lindsay, B. G. (2007). A Nonparametric Statistical Approach to Clustering via Mode Inentification. *Journal of Machine Learning Research*, 8 (8), 1687-1723.

Lin, Z., Li, Q., and Sun, Y. (2014). A Consistent Nonparametric Test of Parametric Regression Functional Form in Fixed Effects Panel Data Models. *Journal of Econometrics*, 178, 167-179.

Muller, D. W. and Sawitzki, G. (1991). Excess Mass Estimates and Tests for Multimodality. *Journal of the American Statistical Association*, 86, 738-746.

Munnell, A. H. (1990). How Does Public Infrastructure Affect Regional Economic Performance? *New England Economic Review*, 11-33.

Ota, H., Kato, K., and Hara, S. (2019). Quantile Regression Approach to Conditional Mode Estimation. *Electronic Journal of Statistics*, 13, 3120-3160.

Parzen, M. (1962). On Estimation of a Probability Density Function and Mode. *Philos. Trans. Roy. Soc. London Ser. A*, 186, 343-414.

Sager, T. W. and Thisted, R. A. (1982). Maximum Likelihood Estimation of Isotonic Modal Regression. *The Annals of Statistics*, 10 (3), 690-707.

Silverman, B. W. (1981). Using Kernel Density Estimates to Investigate Multimodality. *Journal of Royal Statistical Society Series B*, 43, 97-99.

Su, L., Chen, Y., and Ullah, A. (2009). Functional Coefficient Estimation with Both Categorical and Continuous Data. *Advances in Econometrics*, 25, 131-167.

Su, L. and Ullah, A. (2006). Profile Likelihood Estimation of Partially Linear Panel Data Models with Fixed Effects. *Economics Letter*, 92 (1), 75-81.

Su, L. and Ullah, A. (2011). Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing, in A. Ullah and D. E. A. Giles (eds). *Handbook of Empirical Economics and Finance*, 455-497.

Su, L., Ullah, A., and Wang, Y. (2013). Nonparametric Regression Estimation with General Parametric Error Covariance: A More Efficient Two-Step Estimator. *Empirical Economics*, 1009-1024.

Tarter, M. E. and Lock, M. D. (1993). Model-Free Curve Estimation. *CRC Press, Boca Raton, FL*, 56.

Yao, W. (2013). A Note on EM Algorithm for Mixture Models. *Statistics and Probability Letters*, 83, 519-526.

Yao, W. and Li, L. (2014). A New Regression Model: Modal Linear Regression. *Scandinavian Journal of Statistics*, 41, 656-671.

Yao, W., Lindsay, B. G., and Li, R. (2012). Local Modal Regression. *Journal of Nonparametric Statistics*, 24 (3), 647-663.

Yao, W. and Xiang, S. (2016). Nonparametric and Varying Coefficient Modal Regression. *arXiv:1602.06609v1*.

Zhou, H. and Huang, X. (2016). Nonparametric Modal Regression in the Presence of Measurement Error. *Electronic Journal of Statistics*, 10, 3579-3620.

A Appendix-Tables

Table 1: Monte Carlo Results: DGP 1-Case 1

		$N=5$		$N=10$		$N=20$		$N=50$		$N=100$	
Method	T	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$
LDMR $\beta_{mode}=3$	20	2.8262 (0.4292)	0.2135	2.8040 (0.4021)	0.2402	2.8677 (0.4011)	0.1784	2.8908 (0.2855)	**0.0930**	2.9381 (0.3038)	0.0956
	40	2.9082 (0.1502)	**0.0309**	2.9331 (0.2894)	0.0878	2.8999 (0.2683)	0.0817	2.9493 (0.1774)	**0.0339**	2.9325 (0.2304)	0.0574
	60	2.9495 (0.1330)	**0.0201**	2.9622 (0.1568)	**0.0259**	2.9433 (0.1687)	**0.0315**	2.9455 (0.1379)	**0.0219**	2.9428 (0.1917)	0.0399
	100	2.9570 (0.1101)	**0.0139**	2.9499 (0.1066)	**0.0138**	2.9479 (0.1100)	**0.0148**	2.9529 (0.0992)	**0.0120**	2.9531 (0.1034)	**0.0128**
	200	2.9488 (0.0770)	**0.0085**	2.9527 (0.0729)	**0.0075**	2.9489 (0.0758)	**0.0083**	2.9445 (0.0718)	**0.0082**	2.9497 (0.0717)	**0.0076**
PDTS $\beta_{mode}=3$	20	2.7954 (0.3975)	**0.1990**	2.7873 (0.3591)	**0.1735**	2.8009 (0.3480)	**0.1601**	2.8442 (0.2736)	0.0987	2.8951 (0.2442)	**0.0749**
	40	2.9285 (0.2176)	0.0522	2.9292 (0.1827)	**0.0382**	2.8935 (0.2555)	**0.0763**	2.9200 (0.1715)	0.0357	2.9142 (0.1595)	**0.0327**
	60	2.9289 (0.1352)	0.0232	2.9388 (0.1531)	0.0271	2.9196 (0.1686)	0.0348	2.9261 (0.1312)	0.0226	2.9312 (0.1469)	**0.0262**
	100	2.9510 (0.1123)	0.0150	2.9394 (0.1074)	0.0151	2.9378 (0.1094)	0.0158	2.9422 (0.0965)	0.0126	2.9415 (0.1017)	0.0137
	200	2.9431 (0.0777)	0.0092	2.9471 (0.0738)	0.0082	2.9433 (0.0753)	0.0089	2.9390 (0.0701)	0.0086	2.9433 (0.0703)	0.0081
LSDV $\beta_{mean}=2$	20	2.0345 (0.5075)	0.2574	1.9677 (0.4855)	0.2356	2.0214 (0.4431)	0.1958	2.0280 (0.4338)	0.1881	2.0271 (0.4290)	0.1839
	40	2.0619 (0.3549)	0.1292	2.0208 (0.3813)	0.1451	2.0075 (0.3347)	0.1115	1.9897 (0.3462)	0.1194	2.0051 (0.3609)	0.1296
	60	1.9956 (0.2846)	0.0806	2.0013 (0.2710)	0.0731	2.0032 (0.2908)	0.0842	2.0134 (0.2577)	0.0663	2.0063 (0.2613)	0.0680
	100	1.9983 (0.2305)	0.0529	2.0157 (0.2228)	0.0496	2.0091 (0.2017)	0.0405	2.0197 (0.2180)	0.0476	2.0144 (0.2160)	0.0466
	200	2.0108 (0.1708)	0.0292	2.0097 (0.1499)	0.0224	1.9951 (0.1588)	0.0251	2.0143 (0.1375)	0.0190	2.0140 (0.1482)	0.0220

Table 2: Monte Carlo Results: DGP 1-Case 2

		$N=5$		$N=10$		$N=20$		$N=50$		$N=100$	
Method	T	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$
LDMR $\beta_{mode}=3$	20	2.8134 (0.4553)	0.2410	2.7818 (0.4664)	0.2641	2.8086 (0.4439)	0.2327	2.8921 (0.3587)	0.1397	2.8983 (0.2511)	**0.0731**
	40	2.9111 (0.1976)	0.0468	2.8937 (0.2450)	0.0711	2.8694 (0.2675)	0.0882	2.9008 (0.1624)	**0.0361**	2.8956 (0.1511)	**0.0336**
	60	2.9467 (0.1797)	0.0350	2.9516 (0.2037)	0.0436	2.9466 (0.1768)	**0.0339**	2.9427 (0.1514)	0.0261	2.9013 (0.1378)	**0.0286**
	100	2.9576 (0.1238)	**0.0170**	2.9494 (0.1178)	**0.0164**	2.9443 (0.1227)	**0.0181**	2.9503 (0.1080)	**0.0141**	2.9540 (0.1112)	**0.0144**
	200	2.9488 (0.0838)	**0.0096**	2.9525 (0.0772)	**0.0082**	2.9492 (0.0831)	**0.0095**	2.9435 (0.0793)	**0.0094**	2.9517 (0.0771)	**0.0082**
PDTS $\beta_{mode}=3$	20	2.8122 (0.3569)	**0.1620**	2.7859 (0.4337)	**0.2330**	2.8160 (0.4242)	0.2129	2.8802 (0.3279)	**0.1213**	2.8956 (0.2676)	0.0822
	40	2.8948 (0.1735)	**0.0410**	2.9306 (0.1951)	**0.0427**	2.9091 (0.1910)	**0.0445**	2.9208 (0.1780)	0.0378	2.9079 (0.1975)	0.0473
	60	2.9315 (0.1475)	**0.0263**	2.9348 (0.1770)	**0.0354**	2.9247 (0.1767)	0.0367	2.9255 (0.1428)	**0.0258**	2.9263 (0.1851)	0.0395
	100	2.9497 (0.1264)	0.0184	2.9372 (0.1181)	0.0178	2.9347 (0.1204)	0.0187	2.9403 (0.1043)	0.0144	2.9423 (0.1089)	0.0151
	200	2.9430 (0.0836)	0.0102	2.9475 (0.0775)	0.0087	2.9439 (0.0822)	0.0099	2.9385 (0.0771)	0.0097	2.9453 (0.0757)	0.0087
LSDV $\beta_{mean}=2$	20	2.0343 (0.5285)	0.2791	1.9626 (0.4950)	0.2452	2.0221 (0.4487)	**0.2008**	2.0284 (0.4357)	0.1897	2.0267 (0.4298)	0.1845
	40	2.0655 (0.3636)	0.1359	2.0221 (0.3880)	0.1503	2.0099 (0.3374)	0.1134	1.9894 (0.3476)	0.1203	2.0059 (0.3614)	0.1300
	60	1.9910 (0.2930)	0.0855	2.0028 (0.2765)	0.0761	2.0038 (0.2942)	0.0862	2.0136 (0.2591)	0.0670	2.0075 (0.2617)	0.0682
	100	1.9937 (0.2369)	0.0559	2.0130 (0.2282)	0.0520	2.0094 (0.2027)	0.0410	2.0174 (0.2184)	0.0478	2.0144 (0.2167)	0.0469
	200	2.0098 (0.1720)	0.0295	2.0095 (0.1527)	0.0233	1.9943 (0.1606)	0.0257	2.0137 (0.1380)	0.0191	2.0143 (0.1484)	0.0221

Table 3: Monte Carlo Results: DGP 1-Case 3

Method	T	β (N=5)	$MSE(\beta)$	β (N=10)	$MSE(\beta)$	β (N=20)	$MSE(\beta)$	β (N=50)	$MSE(\beta)$	β (N=100)	$MSE(\beta)$
LDMR $\beta_{mode}=2$	20	2.0030 (0.0696)	0.0048	1.9961 (0.0561)	0.0031	2.0016 (0.0352)	0.0012	2.0024 (0.0199)	0.0004	2.0006 (0.0182)	0.0003
	40	2.0040 (0.0310)	**0.0010**	2.0015 (0.0292)	0.0009	1.9988 (0.0167)	0.0003	1.9996 (0.0105)	**0.0001**	2.0002 (0.0070)	0.00005
	60	1.9980 (0.0230)	**0.0005**	1.9994 (0.0203)	**0.0004**	1.9989 (0.0133)	**0.0002**	1.9998 (0.0081)	0.00007	2.0004 (0.0052)	**0.00003**
	100	1.9980 (0.0177)	**0.0003**	1.9987 (0.0139)	**0.0002**	1.9992 (0.0100)	**0.0001**	1.9994 (0.0062)	**0.00004**	2.0001 (0.0042)	**0.00002**
	200	2.0010 (0.0128)	**0.0002**	1.9996 (0.0094)	**0.00009**	2.0002 (0.0064)	**0.00004**	1.9998 (0.0043)	**0.00002**	2.0000 (0.0027)	**0.000007**
PDTS $\beta_{mode}=2$	20	2.0004 (0.0557)	**0.0031**	1.9964 (0.0533)	**0.0028**	2.0020 (0.0265)	**0.0007**	2.0017 (0.0160)	**0.00026**	2.0002 (0.0110)	**0.0001**
	40	2.0040 (0.0298)	**0.0009**	2.0011 (0.0263)	**0.0007**	1.9991 (0.0157)	**0.0002**	1.9996 (0.0100)	**0.0001**	2.0003 (0.0066)	**0.00004**
	60	1.9981 (0.0224)	**0.0005**	1.9996 (0.0196)	**0.0004**	1.9991 (0.0128)	**0.0002**	1.9998 (0.0078)	**0.00006**	2.0005 (0.0050)	**0.00003**
	100	1.9979 (0.0174)	**0.0003**	1.9986 (0.0136)	**0.0002**	1.9993 (0.0099)	**0.0001**	1.9994 (0.0060)	**0.00004**	2.0001 (0.0041)	**0.00002**
	200	2.0008 (0.0126)	**0.0002**	1.9996 (0.0092)	**0.00009**	2.0002 (0.0063)	**0.00004**	1.9998 (0.0042)	**0.00002**	2.0000 (0.0027)	**0.000007**
LSDV $\beta_{mean}=2$	20	1.9998 (0.0765)	0.0058	1.9948 (0.0537)	0.0029	2.0007 (0.0354)	0.0013	2.0004 (0.0231)	0.0005	1.9996 (0.0171)	0.0003
	40	2.0036 (0.0507)	0.0026	2.0014 (0.0377)	0.0014	2.0024 (0.0232)	0.0005	1.9996 (0.0171)	0.0003	2.0008 (0.0108)	0.0001
	60	1.9954 (0.0378)	0.0014	2.0015 (0.0311)	0.0010	2.0006 (0.0210)	0.0004	2.0001 (0.0133)	0.0002	2.0011 (0.0092)	0.00009
	100	1.9954 (0.0308)	0.0010	1.9973 (0.0224)	0.0005	2.0003 (0.0160)	0.0003	1.9995 (0.0098)	0.0001	2.0000 (0.0071)	0.00005
	200	1.9990 (0.0215)	0.0005	2.0015 (0.0156)	0.0002	1.9991 (0.0113)	0.0001	1.9994 (0.0072)	0.00005	2.0003 (0.0050)	0.00002

Table 4: Monte Carlo Results: DGP 2-More Skewed

Method	T	β (N=5)	$MSE(\beta)$	β (N=10)	$MSE(\beta)$	β (N=20)	$MSE(\beta)$	β (N=50)	$MSE(\beta)$	β (N=100)	$MSE(\beta)$
LDMR $\beta_{mode}=1.5$	20	1.5342 (0.1278)	**0.0174**	1.5637 (0.1462)	**0.0253**	1.5248 (0.1373)	**0.0194**	1.5222 (0.1450)	**0.0214**	1.5460 (0.1615)	**0.0281**
	40	1.5247 (0.0976)	**0.0101**	1.5157 (0.1043)	**0.0111**	1.5218 (0.0946)	**0.0094**	1.5195 (0.1005)	**0.0104**	1.5221 (0.0987)	**0.0102**
	60	1.5198 (0.0764)	**0.0062**	1.5116 (0.0760)	**0.0059**	1.5081 (0.0854)	**0.0073**	1.5160 (0.0831)	**0.0071**	1.5143 (0.0812)	**0.0068**
	100	1.5045 (0.0564)	**0.0032**	1.5139 (0.0609)	**0.0039**	1.5078 (0.0591)	**0.0035**	1.5052 (0.0602)	**0.0036**	1.5088 (0.0543)	**0.0030**
	200	1.5019 (0.0478)	**0.0023**	1.5041 (0.0402)	**0.0016**	1.5055 (0.0420)	**0.0018**	1.5042 (0.0400)	**0.0016**	1.5039 (0.0437)	**0.0019**
PDTS $\beta_{mode}=1.5$	20	1.5440 (0.1316)	0.0192	1.5582 (0.1504)	0.0259	1.5388 (0.1353)	0.0197	1.5362 (0.1437)	0.0219	1.5592 (0.1584)	0.0285
	40	1.5287 (0.0970)	0.0102	1.5217 (0.1032)	**0.0111**	1.5285 (0.0945)	0.0097	1.5270 (0.0993)	0.0105	1.5288 (0.0980)	0.0104
	60	1.5230 (0.0767)	0.0064	1.5152 (0.0755)	**0.0059**	1.5120 (0.0851)	**0.0073**	1.5205 (0.0828)	0.0072	1.5183 (0.0807)	**0.0068**
	100	1.5058 (0.0569)	0.0033	1.5163 (0.0614)	0.0040	1.5101 (0.0592)	0.0036	1.5078 (0.0603)	0.0037	1.5112 (0.0543)	0.0031
	200	1.5024 (0.0480)	**0.0023**	1.5063 (0.0405)	0.0017	1.5063 (0.0420)	**0.0018**	1.5050 (0.0400)	**0.0016**	1.5047 (0.0436)	**0.0019**
LSDV $\beta_{mean}=1.75$	20	1.7433 (0.1677)	0.0280	1.7583 (0.1612)	0.0259	1.7484 (0.1561)	0.0242	1.7502 (0.1494)	0.0222	1.7565 (0.1604)	0.0257
	40	1.7477 (0.1100)	0.0120	1.7491 (0.1179)	0.0138	1.7534 (0.1117)	0.0124	1.7574 (0.1039)	0.0108	1.7529 (0.1145)	0.0131
	60	1.7544 (0.0981)	0.0096	1.7504 (0.0859)	0.0073	1.7507 (0.0916)	0.0083	1.7584 (0.0950)	0.0090	1.7469 (0.0839)	0.0070
	100	1.7470 (0.0700)	0.0049	1.7650 (0.0717)	0.0053	1.7555 (0.0659)	0.0044	1.7548 (0.0727)	0.0053	1.7535 (0.0657)	0.0043
	200	1.7515 (0.0519)	0.0027	1.7530 (0.0509)	0.0026	1.7546 (0.0474)	0.0023	1.7516 (0.0473)	0.0022	1.7494 (0.0484)	0.0023

Table 5: Monte Carlo Results: DGP 2-Less Skewed

		N=5		N=10		N=20		N=50		N=100	
Method	T	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$	β	$MSE(\beta)$
LDMR $\beta_{mode}=3$	20	3.1173 (0.7142)	0.5213	3.0130 (0.7652)	0.5827	3.1253 (0.7954)	0.6451	2.9638 (0.8370)	0.6983	2.9013 (0.8093)	0.6615
	40	3.0968 (0.6665)	0.4514	3.0281 (0.6743)	0.4532	3.0388 (0.6814)	0.4635	3.0421 (0.7153)	0.5109	3.0360 (0.7083)	0.5004
	60	3.0419 (0.5977)	0.3573	3.0646 (0.6076)	0.3715	3.0224 (0.6084)	0.3688	2.9849 (0.6310)	0.3964	3.0107 (0.6506)	0.4213
	100	3.0908 (0.5499)	0.3092	3.0481 (0.5579)	0.3120	3.0240 (0.5240)	0.2738	3.0805 (0.5737)	0.3340	3.0786 (0.5943)	0.3576
	200	3.0804 (0.5092)	0.2644	3.0738 (0.4934)	0.2477	3.0940 (0.5004)	0.2600	3.0265 (0.5562)	0.3085	3.0014 (0.5883)	0.3443
PDTS $\beta_{mode}=3$	20	3.1356 (0.5339)	0.3020	2.9996 (0.5018)	0.2506	3.1179 (0.5765)	0.3446	3.0430 (0.5304)	0.2818	3.0035 (0.5271)	0.2765
	40	3.0432 (0.3702)	0.1382	3.0530 (0.3878)	0.1525	3.0788 (0.4188)	0.1808	2.9780 (0.4800)	0.2298	3.0024 (0.4581)	0.2088
	60	2.9918 (0.3382)	0.1139	3.0239 (0.3372)	0.1137	2.9920 (0.3919)	0.1529	2.9883 (0.3832)	0.1463	2.9939 (0.3473)	0.1200
	100	3.0373 (0.3192)	0.1028	3.0049 (0.3208)	0.1024	2.9447 (0.3647)	0.1354	2.9883 (0.3514)	0.1230	3.0220 (0.3418)	0.1167
	200	3.0134 (0.2940)	0.0862	3.0212 (0.2380)	0.0568	3.0098 (0.2716)	0.0735	3.0050 (0.2678)	0.0714	2.9698 (0.2832)	0.0807
LSDV $\beta_{mean}=3.25$	20	3.2961 (0.4127)	**0.1716**	3.2147 (0.4126)	**0.1706**	3.3091 (0.4099)	**0.1707**	3.2545 (0.3780)	**0.1422**	3.2200 (0.3943)	**0.1556**
	40	3.2527 (0.2766)	**0.0761**	3.2594 (0.2726)	**0.0740**	3.3105 (0.2483)	**0.0650**	3.2761 (0.2542)	**0.0650**	3.2684 (0.2438)	**0.0595**
	60	3.2324 (0.2266)	**0.0514**	3.2753 (0.1957)	**0.0387**	3.2605 (0.1987)	**0.0394**	3.2414 (0.2252)	**0.0505**	3.2427 (0.1997)	**0.0397**
	100	3.2686 (0.1778)	**0.0318**	3.2535 (0.1676)	**0.0280**	3.2449 (0.1711)	**0.0291**	3.2698 (0.1649)	**0.0275**	3.2818 (0.1617)	**0.0270**
	200	3.2632 (0.1299)	**0.0170**	3.2544 (0.1139)	**0.0129**	3.2547 (0.1172)	**0.0137**	3.2574 (0.1093)	**0.0119**	3.2448 (0.1074)	**0.0115**

Table 6: The Results of Estimates of (4.3)

LDMR

Method	β_1	AL	AZ	AR	CA	CO	CT	DE	FL	GA	ID	IL	IN
LDMR	−0.0276	2.3351	2.5059	2.4022	2.6105	2.5288	2.5796	2.5480	2.4500	2.3781	2.4868	2.5081	2.3923
	(0.0234)	(0.1395)	(0.1383)	(0.1324)	(0.1558)	(0.1368)	(0.1348)	(0.1333)	(0.1424)	(0.1380)	(0.1272)	(0.1493)	(0.1395)
	β_2	IA	KS	KY	LA	ME	MD	MA	MI	MN	MS	MO	MT
	0.2421***	2.4705	2.4914	2.5363	2.6529	2.3922	2.5153	2.4616	2.5361	2.4433	2.3884	2.4386	2.5260
	(0.0198)	(0.1412)	(0.1422)	(0.1415)	(0.1527)	(0.1249)	(0.1406)	(0.1382)	(0.1464)	(0.1427)	(0.1387)	(0.1388)	(0.1372)
	β_3	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA
	0.8245***	2.4631	2.4966	2.4166	2.5528	2.6185	2.5793	2.3546	2.5202	2.4408	2.5661	2.4822	2.4038
	(0.0246)	(0.1433)	(0.1288)	(0.1243)	(0.1401)	(0.1354)	(0.1567)	(0.1355)	(0.1402)	(0.1468)	(0.1376)	(0.1355)	(0.1469)
	β_4	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY
	−0.0035***	2.4910	2.2427	2.4548	2.3510	2.5464	2.4642	2.4669	2.4978	2.5797	2.4484	2.4484	2.8373
	(0.0008)	(0.1205)	(0.1354)	(0.1345)	(0.1429)	(0.1563)	(0.1316)	(0.1278)	(0.1389)	(0.1482)	(0.1384)	(0.1416)	(0.1424)

PDTS

Method	β_1	AL	AZ	AR	CA	CO	CT	DE	FL	GA	ID	IL	IN
PDTS	−0.0096	2.1417	2.3082	2.2031	2.4405	2.3360	2.4113	2.3536	2.2733	2.1983	2.2785	2.3274	2.1995
	(0.0136)	(0.0777)	(0.0782)	(0.0779)	(0.0780)	(0.0782)	(0.0775)	(0.0771)	(0.0776)	(0.0775)	(0.0782)	(0.0779)	(0.0779)
	β_2	IA	KS	KY	LA	ME	MD	MA	MI	MN	MS	MO	MT
	0.2882***	2.2673	2.2788	2.3394	2.4548	2.2085	2.3404	2.3023	2.3571	2.2556	2.1901	2.2537	2.2883
	(0.0162)	(0.0781)	(0.0779)	(0.0784)	(0.0793)	(0.0775)	(0.0778)	(0.0766)	(0.0781)	(0.0778)	(0.0777)	(0.0776)	(0.0790)
	β_3	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA
	0.7575***	2.2514	2.2820	2.2642	2.3830	2.3945	2.4161	2.1777	2.2840	2.2627	2.3560	2.2910	2.2294
	(0.0153)	(0.0790)	(0.0784)	(0.0765)	(0.0776)	(0.0791)	(0.0779)	(0.0773)	(0.0795)	(0.0779)	(0.0782)	(0.0780)	(0.0776)
	β_4	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY
	−0.0039***	2.3285	2.0595	2.2298	2.1692	2.3338	2.2688	2.2763	2.3206	2.3869	2.2333	2.2691	2.5887
	(0.0014)	(0.0765)	(0.0773)	(0.0782)	(0.0781)	(0.0783)	(0.0779)	(0.0784)	(0.0776)	(0.0781)	(0.0779)	(0.0776)	(0.0795)

LSDV

Method	β_1	AL	AZ	AR	CA	CO	CT	DE	FL	GA	ID	IL	IN
LSDV	-0.0261	2.2016	2.3681	2.2630	2.5004	2.3959	2.4712	2.4135	2.3332	2.2582	2.3384	2.3873	2.2594
	(0.0281)	(0.1704)	(0.1696)	(0.1619)	(0.1948)	(0.1675)	(0.1666)	(0.1566)	(0.1774)	(0.1717)	(0.1546)	(0.1840)	(0.1720)
	β_2	IA	KS	KY	LA	ME	MD	MA	MI	MN	MS	MO	MT
	0.2920***	2.3272	2.3387	2.3993	2.5147	2.2684	2.4003	2.3622	2.4170	2.3155	2.2500	2.3136	2.3482
	(0.0243)	(0.1735)	(0.1719)	(0.1733)	(0.1814)	(0.1500)	(0.1738)	(0.1699)	(0.1813)	(0.1771)	(0.1664)	(0.1710)	(0.1672)
	β_3	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA
	0.7682***	2.3113	2.3419	2.3241	2.4429	2.4544	2.4760	2.2376	2.3439	2.3226	2.4159	2.3509	2.2893
	(0.0291)	(0.1647)	(0.1516)	(0.1586)	(0.1666)	(0.1646)	(0.1948)	(0.1677)	(0.1663)	(0.1818)	(0.1699)	(0.1667)	(0.1827)
	β_4	RI	SC	SD	TN	TX	UT	VT	VA	WA	WV	WI	WY
	−0.0053***	2.3884	2.1194	2.2897	2.2291	2.3937	2.3287	2.3362	2.3805	2.4468	2.2932	2.3290	2.6486
	(0.0010)	(0.1455)	(0.1607)	(0.1636)	(0.1739)	(0.1901)	(0.1614)	(0.1482)	(0.1723)	(0.1821)	(0.1661)	(0.1735)	(0.1732)

Note: Standard errors are indicated in parentheses. Asterisks indicate significance levels: * indicates $p < 0.1$; ** represents $p < 0.05$; and *** is for $p < 0.01$.

Table 7: The Results of Estimates of (4.4)-Individual Effects

Method	β_1	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
LDMR	1.7350***	-4.1292	-4.8750	-5.2397	-2.9213	-2.4056	-4.3799	-4.3764	-5.7085	-4.2332	-4.8504	-5.1363	-3.9278
	(0.5022)	(3.5255)	(3.8948)	(3.7370)	(3.9690)	(3.2564)	(3.3212)	(3.4951)	(4.3319)	(3.7601)	(3.8480)	(4.0154)	(3.0118)
	β_2	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24
	-0.7581	-3.4188	-4.2687	-6.0870	-4.5415	-3.7072	-3.5747	-4.1014	-4.1076	-2.9500	-4.0344	-2.4967	-3.8992
	(0.4779)	(3.2448)	(3.1939)	(4.5602)	(3.5779)	(3.7796)	(3.2003)	(3.3488)	(3.3656)	(3.7091)	(3.1608)	(3.1513)	(3.1811)
	β_3	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34	A35	A36
	0.0609**	-4.0378	-3.8953	-4.0056	-3.7068	-3.6251	-3.8936	-3.5716	-3.3261	-2.9153	-4.0800	-3.6426	-3.7322
	(0.0316)	(3.2982)	(4.0793)	(3.2435)	(3.1928)	(3.0159)	(3.1091)	(3.3304)	(3.5640)	(2.8781)	(3.3460)	(3.2453)	(3.4647)
		A37	A38	A39	A40	A41	A42	A43	A44	A45	A46	A47	A48
		-4.3519	-3.4850	-3.7010	-4.6837	-3.5507	-3.3613	-4.1149	-3.2822	-4.0967	-2.8925	-4.3585	-4.4434
		(3.2327)	(3.5548)	(3.3200)	(3.8350)	(3.0328)	(3.2981)	(3.3182)	(3.0043)	(3.1314)	(3.0542)	(3.3786)	(3.4947)
		A49	A50	A51	A52	A53	A54	A55	A56	A57	A58	A59	A60
		-5.3587	-3.9733	-2.9705	-4.4512	-3.8287	-4.9938	-4.1559	-3.4564	-5.4429	-3.6176	-5.1952	-4.6698
		(4.5654)	(3.2391)	(2.9636)	(3.2796)	(3.2591)	(4.0666)	(3.3955)	(3.2597)	(4.2171)	(3.1648)	(4.4958)	(4.0293)

Method	β_1	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
LSDV	2.7498***	-10.0238	-11.5268	-11.5088	-9.6848	-7.6960	-9.7058	-9.8732	-13.3148	-10.5090	-11.3895	-11.7778	-8.6687
	(1.0391)	(7.3863)	(8.1827)	(7.8573)	(8.2768)	(6.8674)	(7.0152)	(7.3969)	(9.0202)	(7.9174)	(8.0607)	(8.4581)	(6.3740)
	β_2	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24
	$-1.7751*$	-8.7634	-9.3690	-13.9775	-10.5624	-9.9873	-8.7846	-9.5678	-9.5551	-9.1307	-9.1447	-7.5831	-8.8234
	(0.9932)	(6.8473)	(6.7529)	(9.5696)	(7.5268)	(7.9470)	(6.7441)	(7.0520)	(7.1047)	(7.8419)	(6.6799)	(6.6527)	(6.7192)
	β_3	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34	A35	A36
	-0.0744	-9.5344	-10.8465	-9.2734	-8.7470	-8.3432	-8.7277	-8.8790	-9.3724	-7.3697	-9.3284	-8.8923	-9.3172
	(0.0678)	(6.9593)	(8.5741)	(6.8002)	(6.7493)	(6.3816)	(6.5802)	(7.0368)	(7.4918)	(6.0896)	(7.0818)	(6.8497)	(7.3198)
		A37	A38	A39	A40	A41	A42	A43	A44	A45	A46	A47	A48
		-9.3021	-9.3920	-9.1623	-11.0050	-8.2899	-8.6615	-9.5052	-8.1286	-9.1141	-7.8206	-9.8305	-10.1655
		(6.7589)	(7.4876)	(6.9942)	(8.0837)	(6.4110)	(6.9608)	(6.9943)	(6.3348)	(6.6196)	(6.4319)	(7.1233)	(7.3683)
		A49	A50	A51	A52	A53	A54	A55	A56	A57	A58	A59	A60
		-13.0517	-9.0266	-7.6305	-9.7121	-8.9909	-11.8796	-9.6412	-8.7209	-12.4873	-8.7951	-12.9881	-11.5617
		(9.6120)	(6.8529)	(6.2523)	(6.9229)	(6.8520)	(8.5558)	(7.1673)	(6.8733)	(8.8743)	(6.6646)	(9.4447)	(8.4730)

Note: Standard errors are indicated in parentheses. Asterisks indicate significance levels: * indicates $p < 0.1$; ** represents $p < 0.05$; and *** is for $p < 0.01$.

Table 8: The Results of Estimates of (4.4)-Time Effects

Method	β_1	β_2	β_3	Y1999	Y2000	Y2001	Y2002	Y2003	Y2004	Y2005	Y2006	Y2007	Y2008
LDMR	0.9235***	$-0.0675**$	0.5750***	-3.3322	-3.3672	-3.3852	-3.3561	-3.3621	-3.3483	-3.3099	-3.3966	-3.4903	-3.5939
	(0.0262)	(0.0281)	(0.1788)	(1.6702)	(1.7028)	(1.6959)	(1.6841)	(1.7020)	(1.6960)	(1.7129)	(1.7180)	(1.7285)	(1.7356)

Method	β_1	β_2	β_3	Y1999	Y2000	Y2001	Y2002	Y2003	Y2004	Y2005	Y2006	Y2007	Y2008
PDTS	0.9012***	$-0.0827**$	0.7592***	-4.7762	-4.8228	-4.8645	-4.8950	-4.8516	-4.8417	-4.8203	-4.9280	-5.0457	-5.1880
	(0.0292)	(0.0322)	(0.1396)	(1.3324)	(1.3339)	(1.3331)	(1.3348)	(1.3346)	(1.3341)	(1.3346)	(1.3338)	(1.3349)	(1.3346)

Method	β_1	β_2	β_3	Y1999	Y2000	Y2001	Y2002	Y2003	Y2004	Y2005	Y2006	Y2007	Y2008
LSDV	1.0489***	$-0.2512***$	0.8526***	-6.4132	-6.4598	-6.5015	-6.5320	-6.4886	-6.4787	-6.4573	-6.5650	-6.6827	-6.8250
	(0.0281)	(0.0330)	(0.1902)	(1.7791)	(1.7881)	(1.7953)	(1.7986)	(1.8048)	(1.8120)	(1.8191)	(1.8276)	(1.8366)	(1.8461)

Note: Standard errors are indicated in parentheses. Asterisks indicate significance levels: * indicates $p < 0.1$; ** represents $p < 0.05$; and *** is for $p < 0.01$.

B Appendix-Proofs

Proof of Theorem 2.1

Recall that $Y_{it} = X_{it}^T \beta + Z_{\mu,i}^T \mu + v_{it}$. We define $X_{it}^* = (X_{it}^T, Z_{\mu,i}^T)^T$, $\theta = (\beta^T, \mu^T)^T$, and $\theta_0 = (\beta_0^T, \mu_0^T)^T$, where θ_0 is the true parameter value. Let $\delta_{NT} = h_0^2 + \sqrt{(NTh_0^3)^{-1}}$, then it is sufficient to show that for any given η, there exists a large number constant a such that

$$P\left\{ \sup_{\|c\|=a} Q_{NT}\left(\theta_0 + \delta_{NT}c\right) < Q_{NT}\left(\theta_0\right) \right\} \geq 1 - \eta, \tag{A.1}$$

where $\|\cdot\|$ represents the Euclidean distance, and θ_0 is the true parameter value.

Applying the Taylor expansion, it follows

$$
\begin{aligned}
& Q_{NT}\left(\theta_0 + \delta_{NT}c\right) - Q_{NT}\left(\theta_0\right) \\
= & \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\phi\left(\frac{v_{it} - \delta_{NT}c^T X_{it}^*}{h_0}\right) - \phi\left(\frac{v_{it}}{h_0}\right) \right] \\
= & \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[-\phi^{(1)}\left(\frac{v_{it}}{h_0}\right)\left(\frac{\delta_{NT}c^T X_{it}^*}{h_0}\right) + \frac{1}{2}\phi^{(2)}\left(\frac{v_{it}}{h_0}\right)\left(\frac{\delta_{NT}c^T X_{it}^*}{h_0}\right)^2 - \frac{1}{6}\phi^{(3)}\left(\frac{v_{it}^*}{h_0}\right)\left(\frac{\delta_{NT}c^T X_{it}^*}{h_0}\right)^3 \right],
\end{aligned}
$$
$$\tag{A.2}$$

where v_{it}^* is between v_{it} and $v_{it} - \delta_{NT}c^T X_{it}^*$. Based on the result $T_{NT} = \mathbb{E}\left(T_{NT}\right) + O_p(\sqrt{\mathrm{Var}\left(T_{NT}\right)})$, we consider each part of above Taylor expansion.

For the first part, which is $I_1 = \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(-\phi^{(1)}\left(\frac{v_{it}}{h_0}\right)\left(\frac{\delta_{NT}c^T X_{it}^*}{h_0}\right)\right)$, we can calculate it directly to achieve

$$
\begin{aligned}
\mathbb{E}\left(I_1\right) = & \frac{-\delta_{NT}}{h_0^2} \iint \phi^{(1)}\left(\frac{v}{h_0}\right) f_v(v \mid X^*)\left(c^T X^*\right) dv dF(X^*) \\
= & O_p\left(\delta_{NT}ah_0^2\right).
\end{aligned}
\tag{A.3}
$$

Meanwhile, we know that

$$
\begin{aligned}
\mathbb{E}\left(I_1\right)^2 = & \frac{\delta_{NT}^2}{NTh_0^4} \iint \phi^{(1)2}\left(\frac{v}{h_0}\right) f_v(v \mid X^*)\left(c^T X^*\right)^2 dv dF(X^*) \\
= & O_p\left(\delta_{NT}^2 a^2 (NTh_0^3)^{-1}\right).
\end{aligned}
\tag{A.4}
$$

Combing (A.3) and (A.4) obtains

$$I_1 = O_p\left(\delta_{NT}ah_0^2\right) + O_p\left(\sqrt{\delta_{NT}^2 a^2 (NTh_0^3)^{-1}}\right) = O_p\left(\delta_{NT}^2 a\right). \tag{A.5}$$

For the second part, which is $I_2 = \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\frac{1}{2} \phi^{(2)} \left(\frac{v_{it}}{h_0} \right) \left(\frac{\delta_{NT} c^T X_{it}^*}{h_0} \right)^2 \right)$, we can prove

$$\mathbb{E}(I_2) = \frac{\delta_{NT}^2}{2h_0^3} \iint \phi^{(2)} \left(\frac{v}{h_0} \right) f_v(v \mid X^*) \left(c^T X^* \right)^2 dv dF(X^*)$$
$$= O_p \left(\delta_{NT}^2 a^2 \right), \tag{A.6}$$

which indicates the second part will dominate the first part when we choose a big enough.

For the third part, which is $I_3 = \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(-\frac{1}{6} \phi^{(3)} \left(\frac{v_{it}^*}{h_0} \right) \left(\frac{\delta_{NT} c^T X_{it}^*}{h_0} \right)^3 \right)$, as v_{it}^* is between v_{it} and $v_{it} - \delta_{NT} c^T X_{it}^*$, after some direct calculations we can obtain

$$\mathbb{E}(I_3) \approx \frac{-\delta_{NT}^3}{6h_0^4} \iint \phi^{(3)} \left(\frac{v}{h_0} \right) f_v(v \mid X^*) \left(c^T X^* \right)^3 dv F(X^*)$$
$$= O_p \left(\delta_{NT}^3 \right), \tag{A.7}$$

which indicates the second part dominates the third part with the assumption $NTh_0^5 \to \infty$.

Based on these, we can choose a bigger enough such that the second term dominates the other two terms with probability $1 - \eta$. Because the second term is negative, thus $P \left\{ \sup_{\|c\|=a} Q_{NT} (\theta_0 + \delta_T c) < Q_{NT} (\theta_0) \right\} \geq 1 - \eta$ holds. Hence with the probability approaching 1, there exists a local maximisers $\hat{\theta}$ such that

$$\|\hat{\theta} - \theta_0\| = O_p(h_0^2 + \sqrt{(NTh_0^3)^{-1}}). \tag{A.8}$$

$\square$

Proof of Theorem 2.2

Recall that
$$Q_{NT}(\beta, \mu) = \frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi \left(\frac{Y_{it} - X_{it}^T \beta - Z_{\mu,i}^T \mu}{h_0} \right).$$

Because $(\hat{\beta}, \hat{\mu})$ maximizes $Q_{NT}(\beta, \mu)$, we can take the derivative of $Q_{NT}(\beta, \mu)$ respect to β and μ to obtain

$$\left. \frac{\partial Q_{NT}(\beta, \mu)}{\partial \beta} \right|_{(\beta=\hat{\beta}, \mu=\hat{\mu})} = -\frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(1)} \left(\frac{Y_{it} - X_{it}^T \hat{\beta} - Z_{\mu,i}^T \hat{\mu}}{h_0} \right) \left(\frac{X_{it}}{h_0} \right) = 0. \tag{A.9}$$

$$\left. \frac{\partial Q_{NT}(\beta, \mu)}{\partial \mu} \right|_{(\beta=\hat{\beta}, \mu=\hat{\mu})} = -\frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(1)} \left(\frac{Y_{it} - X_{it}^T \hat{\beta} - Z_{\mu,i}^T \hat{\mu}}{h_0} \right) \left(\frac{Z_{\mu,i}}{h_0} \right) = 0. \tag{A.10}$$

By applying Taylor expansion for (A.9) and (A.10), we can achieve

$$
\frac{1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[-\phi^{(1)}\left(\frac{v_{it}}{h_0}\right) X_{it} + \phi^{(2)}\left(\frac{v_{it}}{h_0}\right) X_{it}\{-X_{it}^T(\hat{\beta} - \beta_0)h_0^{-1} - Z_{\mu,i}^T(\hat{\mu} - \mu_0)h_0^{-1}\} \right.
$$
$$
\left. + \frac{1}{2}\phi^{(3)}\left(\frac{v_{it}^*}{h_0}\right) X_{it}\{-X_{it}^T(\hat{\beta} - \beta_0)h_0^{-1} - Z_{\mu,i}^T(\hat{\mu} - \mu_0)h_0^{-1}\}^2 \right] + o_p(\hat{\beta} - \beta_0) = 0.
$$
(A.11)

$$
\frac{1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[-\phi^{(1)}\left(\frac{v_{it}}{h_0}\right) Z_{\mu,i} + \phi^{(2)}\left(\frac{v_{it}}{h_0}\right) Z_{\mu,i}\{-X_{it}^T(\hat{\beta} - \beta_0)h_0^{-1} - Z_{\mu,i}^T(\hat{\mu} - \mu_0)h_0^{-1}\} \right.
$$
$$
\left. + \frac{1}{2}\phi^{(3)}\left(\frac{v_{it}^*}{h_0}\right) Z_{\mu,i}\{-X_{it}^T(\hat{\beta} - \beta_0)h_0^{-1} - Z_{\mu,i}^T(\hat{\mu} - \mu_0)h_0^{-1}\}^2 \right] + o_p(\hat{\mu} - \mu_0) = 0,
$$
(A.12)

where v_{it}^* is between v_{it} and $Y_{it} - X_{it}^T\hat{\beta} - Z_{\mu,i}^T\hat{\mu}$.

We focus on (A.12) firstly. Considering $-\frac{1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(1)}\left(\frac{v_{it}}{h_0}\right) Z_{\mu,i} \right)$, we get

$$
\mathbb{E}\left(-\frac{1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(1)}\left(\frac{v_{it}}{h_0}\right) Z_{\mu,i} \right) \right) = \frac{-1}{h_0^2} \iint \phi^{(1)}\left(\frac{v}{h_0}\right) f_v(v \mid X, Z) Z \, dv dF(X)
$$
$$
= \frac{1}{h_0} \iint \phi(\tau) \tau f_v(\tau h_0 \mid X, Z) Z \, d\tau dF(X)
$$
$$
= \frac{h_0^2}{2} \mathbb{E}\left(Z f_v^{(3)}(0 \mid X, Z) \right).
$$
(A.13)

Considering $\frac{1}{NTh_0^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(2)}\left(\frac{v_{it}}{h_0}\right) \left(Z_{\mu,i} X_{it}^T \right) \right)$, we achieve

$$
\mathbb{E}\left(\frac{1}{NTh_0^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(2)}\left(\frac{v_{it}}{h_0}\right) Z_{\mu,i} X_{it}^T \right) \right) = \frac{1}{h_0^3} \iint \phi^{(2)}\left(\frac{v}{h_0}\right) f_v(v \mid X, Z) \left(ZX^T \right) dv dF(X)
$$
$$
= \frac{1}{h_0^2} \iint \phi(\tau) (\tau^2 - 1) f_v(\tau h_0 \mid X, Z) \left(ZX^T \right) d\tau dF(X)
$$
$$
= \mathbb{E}\left(ZX^T f_v^{(2)}(0 \mid X, Z) \right).
$$
(A.14)

Considering $\frac{1}{NTh_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(2)}\left(\frac{v_{it}}{h_0}\right) \left(\frac{Z_{\mu,i}}{h_0} \frac{Z_{\mu,i}^T}{h_0} \right) \right)$, we can obtain

$$
\mathbb{E}\left(\frac{1}{NTh_0^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(2)}\left(\frac{v_{it}}{h_0}\right) \left(Z_{\mu,i} Z_{\mu,i}^T \right) \right) = \frac{1}{h_0^3} \iint \phi^{(2)}\left(\frac{v}{h_0}\right) f_v(v \mid X, Z) \left(ZZ^T \right) dv dF(X)
$$
$$
= \frac{1}{h_0^2} \iint \phi(\tau) (\tau^2 - 1) f_v(\tau h_0 \mid X, Z) \left(ZZ^T \right) d\tau dF(X)
$$
$$
= \mathbb{E}\left(ZZ^T f_v^{(2)}(0 \mid X, Z) \right).
$$
(A.15)

Then it follows that

$$\hat{\mu} - \mu_0 = (\Phi + o_p(1))^{-1} \{-\Psi(\hat{\beta} - \beta_0) + o_p(1)\}, \tag{A.16}$$

where $\Phi = \lim_{N \to \infty}(1/N) \sum_{i=1}^{N} \mathbb{E}\left(Z_{\mu,i} Z_{\mu,i}^T f_v^{(2)}(0 \mid X_{it}, Z_{\mu,i})\right)$ and $\Psi = \lim_{N,T \to \infty}(1/(NT)) \sum_{i=1}^{N}$
$\sum_{t=1}^{T} \mathbb{E}\left(Z_{\mu,i} X_{it}^T f_v^{(2)}(0 \mid X_{it}, Z_{\mu,i})\right)$.

Substituting (A.16) into (A.11), we can have

$$\frac{1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[-\phi^{(1)}\left(\frac{v_{it}}{h_0}\right) X_{it} + \phi^{(2)}\left(\frac{v_{it}}{h_0}\right) X_{it}\{-X_{it}^T(\hat{\beta} - \beta_0)h_0^{-1} + Z_{\mu,i}^T(\Phi^{-1}\Psi(\hat{\beta} - \beta_0))h_0^{-1}\} \right.$$
$$\left. + \frac{1}{2}\phi^{(3)}\left(\frac{v_{it}^*}{h_0}\right) X_{it}\{-X_{it}^T(\hat{\beta} - \beta_0)h_0^{-1} + Z_{\mu,i}^T(\Phi^{-1}\Psi(\hat{\beta} - \beta_0))h_0^{-1}\}^2 \right] + o_p(\hat{\beta} - \beta_0) = 0. \tag{A.17}$$

Define

$$M_{NT} = \frac{-1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(1)}\left(\frac{v_{it}}{h_0}\right) X_{it} \right),$$

$$J_{NT} = \frac{1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\phi^{(2)}\left(\frac{v_{it}}{h_0}\right) X_{it}(X_{it}^T - Z_{\mu,i}^T \Phi^{-1}\Psi)h_0^{-1} \right),$$

we then get

$$\hat{\beta} - \beta = J_{NT}^{-1} M_{NT}(1 + o_p(1)). \tag{A.18}$$

With some calculations, we can obtain

$$\mathbb{E}\left(\frac{-1}{NTh_0^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(1)}\left(\frac{v_{it}}{h_0}\right) X_{it} \right) = \frac{1}{h_0^2} \iint \phi^{(1)}\left(\frac{v}{h_0}\right) f_v(v \mid X, Z) X \, dv dF(X)$$
$$= \frac{1}{h_0} \iint \phi(\tau) \tau f_v(\tau h_0 \mid X, Z) X \, d\tau dF(X)$$
$$= \frac{h_0^2}{2} \mathbb{E}\left(X f_v^{(3)}(0 \mid X, Z) \right). \tag{A.19}$$

$$\mathbb{E}\left(\frac{1}{NTh_0^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(2)}\left(\frac{v_{it}}{h_0}\right) \left(X_{it}(X_{it}^T - Z_{\mu,i}^T \Phi^{-1}\Psi) \right) \right)$$
$$= \frac{1}{h_0^3} \iint \phi^{(2)}\left(\frac{v}{h_0}\right) f_v(v \mid X, Z) \left(X(X^T - Z^T \Phi^{-1}\Psi) \right) dv dF(X)$$
$$= \frac{1}{h_0^2} \iint \phi(\tau)(\tau^2 - 1) f_v(\tau h_0 \mid X, Z) \left(X(X^T - Z^T \Phi^{-1}\Psi) \right) d\tau dF(X)$$
$$= \mathbb{E}\left(X(X^T - Z^T \Phi^{-1}\Psi) f_v^{(2)}(0 \mid X, Z) \right). \tag{A.20}$$

Meanwhile, based on above calculations, we have

$$\text{Cov}(M_{NT})$$

$$=\mathbb{E}\left\{\left(\frac{1}{NTh_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\phi^{(1)}\left(\frac{v_{it}}{h_0}\right)\frac{X_{it}}{h_0}\right)\right)\left(\frac{1}{NTh_0}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\phi^{(1)}\left(\frac{v_{it}}{h_0}\right)\frac{X_{it}}{h_0}\right)\right)^T\right\}\{1+o_p(1)\}$$

$$=\frac{1}{NTh_0^4}\iint\phi^{(1)2}\left(\frac{v}{h_0}\right)f_v(v\mid X,Z)XX^T dvdF(X)\{1+o_p(1)\}$$

$$=\frac{1}{NTh_0^3}v_2 L\{1+o_p(1)\}, \tag{A.21}$$

where $v_2 = \int \phi^2(\tau)\tau^2 d\tau$ and $L = \lim_{N,T\to\infty}(1/(NT))\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(X_{it}X_{it}^T f_v(0\mid X_{it}, Z_{\mu,i})\right)$.

To show Theorem 2.2, it is sufficient to show the asymptotic normality for $M_{NT}^* = \sqrt{NTh_0^3}M_{NT}$, where we prove that for any unit vector $d \in \mathbb{R}^q$,

$$\{d^T\text{Cov}(M_{NT}^*)d\}^{-1/2}\{d^T M_{NT}^* - d^T\mathbb{E}(M_{NT}^*)\} \xrightarrow{d} N(0,1). \tag{A.22}$$

Then we check Lyapunov's condition. Let $\xi_i = -1/\sqrt{NTh_0}\phi^{(1)}\left(\frac{v_{it}}{h_0}\right)d^T X_{it}$, we need to prove $NT\mathbb{E}|\xi_1|^3 \to 0$. As $\left(d^T X_{it}\right)^2 \leq \|d\|^2\|X_{it}\|^2$ and $\phi^{(1)}(.)$ is bounded, we have

$$NT\mathbb{E}|\xi_1|^3 \leq O((NT)^{-1/2}h_0^{-3/2}) \to 0. \tag{A.23}$$

Thus the asymptotic normality for M_{NT}^* holds with

$$\sqrt{NTh_0^3}\left(M_{NT} - h_0^2 M/2\right) \xrightarrow{d} N(0, v_2 L). \tag{A.24}$$

According to Slutsky's Theorem, we obtain Theorem 2.2.

$$\square$$

Proof of Theorem 3.1

The main proof steps here are similar with those of Proof of Theorem 2.1. We briefly outline the proof. Recall that

$$Q_{NT}(\gamma_1, \beta) = \frac{1}{NTh_1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi\left(\frac{\hat{Y}_{it} - \gamma_1 - X_{it}^T\beta}{h_1}\right)$$

$$= \frac{1}{NTh_1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi\left(\frac{Y_{it} - \alpha_i + \alpha_i - \hat{\alpha}_i - \tilde{X}_{it}^T\theta}{h_1}\right), \tag{A.25}$$

where $\tilde{X}_{it}^T = (1 \quad X_{it}^T)$ and $\theta = (\gamma_1 \quad \beta^T)^T$. Define $\delta_{NT} = h_1^2 + \sqrt{(NTh_1^3)^{-1}}$, it is sufficient to show that for any given η, there exists a large number constant a such that

$P\left\{\sup_{\|c\|=a} Q_{NT}(\theta_0 + \delta_{NT}c) < Q_{NT}(\theta_0)\right\} \geq 1 - \eta$, where $\|\cdot\|$ represents the Euclidean distance, and θ_0 is the true parameter value. Applying the Taylor expansion, it follows

$$
\begin{aligned}
&Q_{NT}(\theta_0 + \delta_{NT}c) - Q_{NT}(\theta_0) \\
&= \frac{1}{NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\phi\left(\frac{v_{it} + \alpha_i - \hat{\alpha}_i - \delta_{NT}c^T \tilde{X}_{it}}{h_1} \right) - \phi\left(\frac{v_{it} + \alpha_i - \hat{\alpha}_i}{h_1} \right) \right] \\
&= \frac{1}{NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[-\phi^{(1)}\left(\frac{v_{it} + \alpha_i - \hat{\alpha}_i}{h_1} \right) \left(\frac{\delta_{NT}c^T \tilde{X}_{it}}{h_1} \right) \right. \\
&\quad \left. + \frac{1}{2}\phi^{(2)}\left(\frac{v_{it} + \alpha_i - \hat{\alpha}_i}{h_1} \right) \left(\frac{\delta_{NT}c^T \tilde{X}_{it}}{h_1} \right)^2 - \frac{1}{6}\phi^{(3)}\left(\frac{v_{it}^*}{h_1} \right) \left(\frac{\delta_{NT}c^T \tilde{X}_{it}}{h_1} \right)^3 \right],
\end{aligned} \tag{A.26}
$$

where v_{it}^* is between $v_{it} + \alpha_i - \hat{\alpha}_i$ and $v_{it} + \alpha_i - \hat{\alpha}_i - \delta_{NT}c^T \tilde{X}_{it}$. Following the same steps as the Proof of Theorem 2.1, with assumption that $\sqrt{T}h_1^2 \to \infty$ (i.e., $N^a/T \to 0$ for some $a > 4/3$), one can obtain $\|\hat{\theta} - \theta_0\| \leq \delta_{NT}$.

$\square$

Proof of Theorem 3.2

Recall that $\tilde{X}_{it}^T = (1 \quad X_{it}^T)$, $\theta = (\gamma_1 \quad \beta^T)^T$, and $\hat{\theta} = [\hat{\gamma}_1 \quad \hat{\beta}^T]^T$. If $\hat{\theta}$ maximises (3.5), it will satisfy the following equation

$$
\frac{1}{NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(1)}\left(\frac{Y_{it} - \alpha_i - (\hat{\alpha}_i - \alpha_i) - \tilde{X}_{it}^T \hat{\theta}}{h_1} \right) \left(\frac{-\tilde{X}_{it}}{h_1} \right) = 0. \tag{A.27}
$$

Then we can achieve

$$
\begin{aligned}
&\frac{1}{NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(1)}\left(\frac{v_{it}}{h_1} \right) \left(\frac{-\tilde{X}_{it}}{h_1} \right) \\
&+ \frac{1}{NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(2)}\left(\frac{v_{it}}{h_1} \right) \left(\frac{\tilde{X}_{it}}{h_1} \right) \left(\frac{\alpha_i - \hat{\alpha}_i - \tilde{X}_{it}^T(\hat{\theta} - \theta)}{h_1} \right) \\
&+ \frac{1}{2NTh_1} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi^{(3)}\left(\frac{v_{it}^*}{h_1} \right) \left(\frac{\tilde{X}_{it}}{h_1} \right) \left(\frac{\alpha_i - \hat{\alpha}_i - \tilde{X}_{it}^T(\hat{\theta} - \theta)}{h_1} \right)^2 = 0, \tag{A.28}
\end{aligned}
$$

where v_{it}^* is between v_{it} and $v_{it} + \alpha_i - \hat{\alpha}_i$. It can be shown that the third term on the left-hand side of (A.28) is dominated by the second term. With assumption that $\sqrt{T}h_1^2 \to \infty$ (i.e., $N^a/T \to 0$ for some $a > 4/3$), we could then follow the same proof steps as those of Proof of Theorem 2.2 to achieve the results.

$\square$