ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

A Varying-Coefficient Panel Data Model with Fixed Effects: Theory and an Application to U.S. Commercial Banks

Guohua Feng, Jiti Gao, Bin Peng and Xiaohui Zhang

April 2015

Working Paper 09/15

A Varying–Coefficient Panel Data Model with Fixed Effects: Theory and an Application to U.S. Commercial Banks

Guohua Feng[§], Jiti Gao[⋆], Bin Peng[†] and Xiaohui Zhang[‡]

[§]University of North Texas, [⋆]Monash University, [†]University of Technology Sydney and [‡]Murdoch University

Abstract

In this paper, we propose a panel data semiparametric varying–coefficient model in which covariates (variables affecting the coefficients) are purely categorical. This model has two features: first, fixed effects are included to allow for correlation between individual unobserved heterogeneity and the regressors; second, it allows for cross–sectional dependence through a general spatial error dependence structure. We derive a semiparametric estimator for our model by using a modified within transformation, and then show the asymptotic and finite properties for this estimator. Finally, we illustrate our model by analyzing the effects of state–level banking regulations on the returns to scale of commercial banks in the U.S.. Our empirical results suggest that returns to scale is higher in more regulated states than in less regulated states.

Keywords: Categorical variable; estimation theory; nonlinear panel data model; returns to scale.

JEL classification: C23, C51, D24, G21

- Guohua Feng, Department of Economics, University of North Texas, Denton, TX 76201, U.S.A. Email: guohua.feng@unt.edu.
- Jiti Gao, Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia. Email: Jiti.Gao@monash.edu.

Jiti Gao acknowledges the Australian Research Council Discovery Grants Program support under Grant numbers: DP130104229 and DP150101012.

- Bin Peng, Economics Discipline Group, University of Technology Sydney, Ultimo, NSW 2007, Australia. Email: Bin.Peng@uts.edu.au.
- Xiaohui Zhang, School of Management and Governance, Murdoch University, Perth 6150, Australia. Email: Xiaohui.Zhang@murdoch.edu.au.

1 Introduction

Varying–coefficient models have attracted considerable attention in the past two decades. This is particularly true for both cross–sectional and time series varying–coefficient models. For instance, Li et al. (2002) propose a semiparametric varying–coefficient model in a cross–sectional setting, where covariates (i.e. variables affecting the coefficients) are assumed to be continuous in nature. Li and Racine (2010) extend Li et al. (2002) to a more general set–up, which admits both quantitative and qualitative covariates. More recently, Li et al. (2013) extend the cross–sectional varying coefficient model literature further by proposing a semiparametric varying–coefficient with purely categorical covariates. Similarly, considerable work has also been done on time series varying–coefficient models. For example, Gao and Phillips (2013*a*) investigate the varying–coefficient model by allowing for the existence of nonstationarity. More references along this latter line can be found in Cai et al. (2009) and Gao and Phillips (2013*b*).

However, less progress has been made with panel data varying–coefficient models, primarily because of the difficulty involved in dealing with fixed effects. For example, Cai and Li (2008) propose a varying–coefficient dynamic panel data model, where they get around this difficulty by dropping fixed effects. Sun et al. (2009) propose a panel data varying–coefficient model, where they overcome the difficulty associated with fixed effects by imposing a widely–used identification restriction such that the sum of the fixed effects is zero (c.f. Su and Ullah (2011) and Chen et al. (2013)). Rodriguez-Poo and Soberon (2014) propose to use the first difference to remove the fixed effects by allowing N to increase to ∞ with fixed T. It is worth noting that in both of the latter two studies, covariates are assumed to be purely continuous and asymptotic theories are established accordingly.

The purpose of this paper is to contribute to this literature by extending Li et al. (2013)'s cross–sectional varying coefficient model to a panel data context. To allow for unobserved individual heterogeneity, fixed effects are included in our model. As is well known, the inclusion of fixed effects has the advantage of allowing unobserved individual heterogeneity to be arbitrarily correlated with any other variables. With regards to the nature of the covariates, we follow Li et al. (2013) and only consider the case where all covariates are categorical. To remove fixed effects, we take advantage of the categorical nature of our covariates and implement the within transformation. The demeaned model can then be estimated using Li et al. (2013)'s semiparametric kernel estimation method. In addition, we establish the asymptotic properties of our estimator.

Another feature of our model is that it allows for cross–sectional dependence, an important issue that has received considerable attention in the recent panel data literature (c.f. Andrews (2005), Pesaran (2006) and Bai (2009)). There are two well–known approaches to modeling cross–sectional dependence. The first approach, due to Pesaran (2006) and Bai (2009), is to use a factor structure to capture strong correlation between individuals. The second approach is to use a spatial error structure to model weak correlation between individuals. Excellent works adopting the second approach include, but are not limited to, Pesaran and Tosetti (2011), Chen et al. (2012*a*) and Chen et al. (2012*b*). In this paper, we adopt the second approach. Specifically, as shown in Assumption A.4 in Section 2, we impose a general spatial correlation structure to link the cross–sectional dependence and stationary mixing condition together. The use of this structure enables our model to capture the types of cross–sectional dependence discussed by Chen et al. (2012*a*) and Chen et al. (2012*b*).

We apply our panel data categorical varying–coefficient model by analyzing the effects of branch banking regimes on the returns to scale of commercial banks in the U.S. over the period 1986–2005. Until the middle of the 1970's banking in the U.S. was heavily regulated at the state level: in some states banks were prohibited from branching at all (unit banking regime), in some states they were restricted to branch within a portion of the state (limited branching banking regime), and in other states they were permitted to branch statewide (statewide branching banking regime). In the mid–1980s individual states began to remove restrictions on intrastate branching. This deregulation process culminated in the passage of the Riegle–Neal Interstate Banking and Branching Efficiency Act of 1994, which permitted nationwide branching as of June 1997 (nationwide branching banking regime). Since banking regime is an important factor in determining production technology, we use it as a categorical argument (covariate) of the varying–coefficients. Specifically, we consider a categorical varying–coefficient translog cost function. Our results show that returns to scale is higher in more regulated states than in less regulated states. Our results also indicate that the majority of the banks face increasing returns to scale, a small percentage face decreasing returns to scale, and an even smaller percentage face constant returns to scale. This finding is potentially important as increasing returns to scale is often used to justify bank mergers and in policy debates on regulations limiting the size of banks.

The rest of this paper is organized as follows. Section 2 presents the varying–coefficient model in a panel data setting; Sections 2.1 and 2.2 consider the relevant and irrelevant covariate cases, respectively. In Section 3, we conduct a Monte Carlo study investigating the finite sample properties of the model. Section 4 presents the application of our methodology to the U.S. commercial bank data. Section 5 concludes this paper with some comments. The proofs of some main results are given in Appendix A. The additional proofs are provided in Appendix B of a supplementary document of this paper.

Before proceeding to Section 2, it is convenient to introduce some notations that will be used throughout this paper. $1(A)$ denotes an indicator function, i.e. $1(A) = 1$ if A is true, otherwise $1(A) = 0$; $\|\cdot\|$ denotes the Frobenius norm; $\to_P$ denotes converging in probability; $\to_D$ denotes converging in distribution; when no mis–understanding can arise, we use $\sum_{js,js\neq it}$ to represent $\sum_{\substack{1\leq j\leq N,1\leq s\leq T\\(j,s)\neq(i,t)}}$; $0_{r\times 1}$ denotes an r–dimensional zero vector; i_q denotes a q–dimensional one vector; I_q denotes a $(q\times q)$ identity matrix.

2 Model Specification

We consider the following panel data model.

$$Y_{it} = X_{it}'\beta(Z_{it}) + w_i + u_{it}, \quad i = 1,\ldots,N \text{ and } t = 1,\ldots,T, \tag{2.1}$$

where u_{it} is a random error term; $X_{it} = (X_{it,1},\ldots,X_{it,q})'$ is a q–dimensional vector of regressors; $\beta(\cdot)$ is a q–dimensional vector of unknown coefficient function; $Z_{it} = (Z_{it,1},\ldots,Z_{it,r})'$ is an r–dimensional vector of discrete covariates; w_i is a fixed effect and can be arbitrarily correlated with any other variables. To distinguish between X_{it} and Z_{it}, they are respectively referred to as regressors and covariates here and hereafter. For an r–dimensional vector z, we use z_s to denote the s^{th} component of z, and assume that z_s takes c_s different values in $\{0,1,\ldots,c_s-1\}$ and $2\leq c_s<\infty$ for $s=1,\ldots,r$. When showing the asymptotic properties of our model and estimator below, we follow Li et al. (2013) and distinguish between the case where $\beta(z)$ is not a constant function with respect to z_s for $s=1,2,\ldots,r$, and the case where some elements of z_s do not have impacts on $\beta(\cdot)$ and are independent of all other variables. The former case is referred to as "relevant case" and will be discussed in details in Section 2.1, while the latter one is referred to as "irrelevant case" and will be discussed in details in Section 2.2.

The model (2.1) extends the cross–sectional varying–coefficient model of Li et al. (2013) to a panel data setting. As in Li et al. (2013), we focus on the case where Z_{it} is purely categorical. Therefore, we also adopt the kernel function of Aitchison and Aitken (1976) for unordered covariate below:

$$l(Z_{it,s}, z_s, \lambda_s) = \begin{cases} 1, & \text{if } Z_{it,s} = z_s \\ \lambda_s, & \text{otherwise} \end{cases}, \tag{2.2}$$

where the range of λ_s is $[0,1]$ for $s=1,\ldots,r$. It is easy to see that $\lambda_s=0$ leads to an indicator function and $\lambda_s=1$ gives a uniform weight function. Note that (2.2) allows one to extend the kernel density estimation technique to multivariate discrete spaces. With (2.2), we can construct a product kernel function as follows.

$$L(Z_{it}, z, \lambda) = \prod_{s=1}^{r} l(Z_{it,s}, z_s, \lambda_s) = \prod_{s=1}^{r} \lambda_s^{1(Z_{it,s}\neq z_s)}, \tag{2.3}$$

where $\lambda = (\lambda_1,\ldots,\lambda_r)'$.

We now discuss how to deal with the fixed effects in (2.1) (i.e., w_i) before proceeding further. To remove the impacts of fixed effects, some studies assume that $\sum_{i=1}^N w_i = 0$ (c.f. Sun et al. (2009), Su and Ullah (2011) and Chen et al. (2013)); some studies propose to take the first difference (c.f. Rodriguez-Poo and Soberon (2014)); and others assume that w_i has mean 0 and is uncorrelated with any other variables (c.f. Blundell and Bond (1998)). In this paper, we take a different approach. Specifically, we take advantage of the fact that Z_{it} is purely discrete and the sample size (N, T) (especially the length of time periods T) is relatively large compared to the number of possible options for $z = (z_1, \ldots, z_r)'$ (i.e. $(N, T) \to (\infty, \infty)$, $r < \infty$ and $\max_{1 \le s \le r} c_s < \infty$), and use the within transformation to remove the fixed effects. However, we cannot follow the common practice of subtracting the average across t from both sides of (2.1), because $\beta(Z_{it})$ varies over t. To deal with this problem, we note a fact that the distance between any two observations Z_{it} and Z_{is} for $t \neq s$ (i.e. $\|Z_{it} - Z_{is}\|$) is either 0 or greater than or equal to 1. This fact indicates that we can use a modified version of the commonly used within transformation formula, where an indicator function, which is equal to one when $Z_{is} = Z_{it}$ and to zero otherwise, is used to remove the fixed effects.

Specifically let $1_{js,it} = 1(Z_{js} = Z_{it})$ for $1 \le i, j \le N$ and $1 \le t, s \le T$ and let $T_{it} = \sum_{s=1}^T 1_{is,it}$. Let $\tilde{X}_{it} = X_{it} - \frac{1}{T_{it}} \sum_{s=1}^T X_{is} 1_{is,it}$, $\tilde{Y}_{it} = Y_{it} - \frac{1}{T_{it}} \sum_{s=1}^T Y_{is} 1_{is,it}$ and $\tilde{u}_{it} = u_{it} - \frac{1}{T_{it}} \sum_{s=1}^T u_{is} 1_{is,it}$. With these notations, our modified within transformation can be written as

$$
\begin{aligned}
\tilde{Y}_{it} &= X_{it}'\beta(Z_{it}) + w_i + u_{it} - \frac{1}{T_{it}} \sum_{s=1}^T \left(X_{is}'\beta(Z_{is}) + w_i + u_{is}\right) 1_{is,it} \\
&= \tilde{X}_{it}'\beta(Z_{it}) + \tilde{u}_{it}. \qquad (2.4)
\end{aligned}
$$

The above transformation requires that the length of time periods T be relatively large compared to r and $\max_{1 \le s \le r} c_s$. In theory, this is controlled by allowing T to increase to ∞ and assuming that r and $\max_{1 \le s \le r} c_s$ are finite.

Alternatively, one can follow the set–up of Blundell and Bond (1998) and assume that w_i is uncorrelated with all the other variables and has mean 0. After some attempts, we find that this condition will make the rates of convergence provided in the following sections become slower (i.e. the rates shown in Theorems 1–4 below will become $O_P\left(\frac{1}{N}\right)$ and $O_P\left(\frac{1}{\sqrt{N}}\right)$ rather than $O_P\left(\frac{1}{NT}\right)$ and $O_P\left(\frac{1}{\sqrt{NT}}\right)$). A short explanation for the slower convergence rates is that under the simplest restriction (i.e. all variables are i.i.d. over i and t, and independent of each other) $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} u_{it} = O_P\left(\frac{1}{\sqrt{NT}}\right)$ and $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} w_i = O_P\left(\frac{1}{\sqrt{N}}\right)$. Then it is easy to see that $\frac{\sqrt{N}}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} w_i = O_P(1)$ is the driving force when we derive the asymptotic normality, because $\frac{\sqrt{N}}{NT} \sum_{i=1}^N \sum_{t=1}^T X_{it} u_{it} = O_P\left(\frac{1}{\sqrt{T}}\right)$ converges to 0 as (N, T) increase to ∞ jointly.

Note that when dealing with the fixed effects, we do not follow Sun et al. (2009), Su and Ullah (2011) and Chen et al. (2013) who assume that $\sum_{i=1}^N w_i = 0$ and expand $\beta(z)$ in a small neighborhood of z for the cases where the covariates are continuous. This is because with this assumption, the bandwidth will not play a role at all due to the fact that Z_{it} is purely discrete in this study. Furthermore, we do not take the first difference as Rodriguez-Poo and Soberon (2014), where they assume that N goes to ∞ with T being fixed.

Using our modified within transformation in (2.4), we can estimate $\beta(z)$ for $\forall z \in \mathcal{D}$ as follows.

$$\hat{\beta}(z) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{X}_{it} \tilde{X}_{it}' L(Z_{it}, z, \hat{\lambda}) \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{X}_{it} \tilde{Y}_{it} L(Z_{it}, z, \hat{\lambda}), \tag{2.5}$$

where $\hat{\lambda}$ is obtained by minimizing the following cross–validation (CV) criterion function

$$CV(\lambda) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{Y}_{it} - \tilde{X}_{it}' \hat{\beta}_{-it}(Z_{it}) \right)^2; \tag{2.6}$$

$\hat{\beta}_{-it}(Z_{it})$ is the leave–one–out estimator for $\beta(Z_{it})$

$$\hat{\beta}_{-it}(Z_{it}) = \left(\sum_{js, js \neq it} \tilde{X}_{js} \tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda) \right)^{-1} \sum_{js, js \neq it} \tilde{X}_{js} \tilde{Y}_{js} L(Z_{js}, Z_{it}, \lambda). \tag{2.7}$$

Having shown how to estimate our panel data categorical varying–coefficient model in (2.1), in what follows we will show the asymptotic properties for our estimator. As noted previously, we will first discuss the asymptotic results for the relevant covariate case in Section 2.1 and then discuss the asymptotic results for the irrelevant covariate case in Section 2.2.

2.1 The Relevant Covariate Case

We start by introducing the assumptions necessary for establishing the asymptotic results for this case.

Assumption A:

1. $\beta(z)$ is not a constant function with respect to z and uniformly bounded on the support $\mathcal{D}$ of z, i.e. $\max_{z \in \mathcal{D}} \|\beta(z)\| < \infty$. For $z = (z_1, \ldots, z_r)' \in \mathcal{D}$, z_s takes c_s different integer values in $\{0, 1, \ldots, c_s - 1\}$ and $c_s \geq 2$ for $s = 1, \ldots, r$. Moreover, r is finite and $\max_{1 \leq s \leq r} c_s < \infty$. Let $p(z) = \Pr(Z_{it} = z) > 0$ for $\forall z \in \mathcal{D}$.

2. Suppose that Z_{it} is independent and identically distributed (i.i.d.) over i and t. Moreover, $\{X_{i1}, \ldots, X_{iT}\}$ is i.i.d. across i.

3. (a) For $\forall z \in \mathcal{D}$, $i = 1, \ldots, N$ and $t = 1, \ldots, T$, $E[X_{it} | Z_{it} = z] = \mu_X(z)$, $E[X_{it} X_{it}' | Z_{it} = z] = \Sigma_X(z)$. X_{it} is independent of Z_{js} for $(i, t) \neq (j, s)$.

(b) $X_t = (X_{1t}, \ldots, X_{Nt})'$ is strictly stationary and α–mixing. Let $\alpha_i(|t-s|)$ be the α–mixing coefficient between X_{it} and X_{is}, such that for a $\delta_1 > 0$ these coefficients satisfy that $\sum_{t=1}^T \sum_{s=1}^T \max_{1\le i\le N} (\alpha_i(|t-s|))^{\frac{\delta_1}{4+\delta_1}} = O(T)$. For the same δ_1, $E\|X_{it}\|^{4+\delta_1} < \infty$ uniformly in i and t.

4. u_{it} is independent of (X_{js}, Z_{js}) for all $1 \le i,j \le N$ and $1 \le t,s \le T$. Let $u_t = (u_{1t}, \ldots, u_{Nt})'$ be strictly stationary and α–mixing. Also, $E[u_{it}] = 0$ and $E[u_{it}^2] = \sigma_u^2$ for $1 \le i \le N$ and $1 \le t \le T$. Let $\alpha_{u,ij}(|t-s|)$ denote the α–mixing coefficient between u_{it} and u_{js}, such that for a $\delta_2 > 0$, $\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T (\alpha_{u,ij}(|t-s|))^{\frac{\delta_2}{4+\delta_2}} = O(NT)$. For the same δ_2, $E|u_{it}|^{4+\delta_2} < \infty$ uniformly in i and t. On the time dimension, let $\sum_{t=1}^T \sum_{s=1}^T |E[u_{it}u_{is}]| = O(T)$ uniformly in i. In addition, let

$$\sum_{i_1=1}^N \sum_{i_2=1}^N \sum_{s_1=1}^T \sum_{s_2=1}^T \sum_{s_3=1}^T \sum_{s_4=1}^T \left| E\left[u_{i_1 s_1} u_{i_1 s_2} u_{i_2 s_3} u_{i_2 s_4}\right] \right| = O(NT^2).$$

5. $\lambda_s \in [0,1]$ for $s = 1, \ldots, r$. Denote that

$$CV_0(\lambda) = \sum_{z\in\mathcal{D}} p(z)(\eta(z) - \beta(z))'(\Sigma_X(z) - \mu_X(z)\mu_X(z)')(\eta(z) - \beta(z)),$$

where $(\Sigma_X(z) - \mu_X(z)\mu_X(z)')$ is positive definite uniformly in z. Let

$$\begin{aligned}
\Sigma_{XX}(z) &= E\left[(\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})')\, L(Z_{it}, z, \lambda)|z\right],\\
\Sigma_{XX\beta}(z) &= E\left[(\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})')\, \beta(Z_{it}) L(Z_{it}, z, \lambda)|z\right],\\
\eta(z) &= \Sigma_{XX}^{-1}(z)\Sigma_{XX\beta}(z).
\end{aligned}$$

Only $\lambda = (\lambda_1, \ldots, \lambda_r)' = 0_{r\times 1}$ makes $CV_0(\lambda) = 0$ hold. Also, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, $E\left|(\eta(Z_{it}) - \beta(Z_{it}))'(\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})')(\eta(Z_{it}) - \beta(Z_{it}))\right|^2 < \infty$.

Assumption A.1 is standard and the same as Assumption 1.1 of Li et al. (2013). It ensures that the sample size (especially the length of time series T) is relatively large compared to the number of choices on z.

Assumption A.2 is standard in the literature (c.f. Assumption A1 of Cai and Li (2008); Assumption 1 of Sun et al. (2009); Assumption A1 of Chen et al. (2013); Assumption 1.1 of Li et al. (2013) and Assumption 3.1 of Rodriguez-Poo and Soberon (2014)). Specifically, using Assumption A.2, it is easy to know that T_{it}/T converges to $p(z) = \Pr(Z_{it} = z)$ in probability one, which implies that $T_{it} = O_P(T)$ uniformly in i and t. Due to the use of conditional expectation, we are not able to impose certain weak cross–sectional dependence on X_{it} and Z_{it} as we do for u_{it}. When all elements of Z_{it} are continuous, one certainly can allow Z_{it} to be α–mixing in the same way as Sun et al. (2009) and Rodriguez-Poo and Soberon (2014). However, since Z_{it} is purely discrete in this study, we assume that Z_{it} is independent over i and t. In the literature of time series, Andrews (1984) has shown that even the process $x_{t+1} = 0.5x_t + \varepsilon_t$ is not α–mixing when ε_t has a binomial distribution. More details and relevant discussions can be found in Fan and Yao (2003). Thus, we believe Assumption 2 is reasonable. Alternatively, we can assume that Z_{it} is pre–determined. Accordingly, we need to adjust our assumptions and analysis, but the consistency and asymptotic normality remain valid. In an even more extreme case, we can assume that all Z_{it}'s are pre–determined to be z and $\lambda = 0$ (or 1). Then the model (2.1) will reduce to the classic panel data model with fixed effects. In this extreme case, the assumptions and analysis can be significantly simplified.

Assumption A.3.(a) places a restriction on the conditional moments and Assumption A.3.(b) places a further restriction on the unconditional moments. Both restrictions are standard in the literature (c.f. Assumption A1 and A4 of Chen et al. (2013) and Assumption 3.3–3.6 of Rodriguez-Poo and Soberon (2014)). Below we provide an example to demonstrate why this assumption is reasonable.

- For simplicity, suppose that all variables are scalars and consider the data generating process as $X_{it} = H_{it} + \varepsilon_{it}$ and $\varepsilon_{it} \sim N(Z_{it}, Z_{it} + 1)$, where $H_{it} = 0.5 \cdot H_{i,t-1} + v_{it}$ is an AR(1) process; $v_{it} \sim N(0, 1)$ is i.i.d. over i and t; $Z_{it} = 0$ with probability 0.4 and $Z_{it} = 1$ with probability 0.6. In this example, the requirements of Assumption A.3 are certainly satisfied. Moreover, this example particularly implies that the choice of Z_{it} affects only the value of X_{it}, but does not affect the value of X_{js} for $(j, s) \neq (i, t)$.

Assumption A.4 is in the same spirit as Assumption C of Bai (2009) and Assumption A2 and A4 of Chen et al. (2012*b*). The last equation provided in Assumption A.4 is necessary due to the within transformation. It is a simpler version of (A.18) of Chen et al. (2012*a*), where more detailed discussions are provided. Two examples are given below to demonstrate this assumption is reasonable:

- It can be easily seen that Assumption A.4 holds if u_{it} is i.i.d. over i and t.

- We now use a factor model structure as an example to show that Assumption A.4 is verifiable. Suppose that $u_{it} = \gamma_i f_t + \varepsilon_{it}$, where all variables are scalars and ε_{it} is i.i.d. over i and t with mean zero. Simple algebra shows that the coefficient $\alpha_{u,ij}(|t - s|)$ reduces to $\alpha_{ij} \cdot b(|t - s|)$, in which $\alpha_{ij} = E[\gamma_i \gamma_j]$ and $b(|t - s|)$ is the α–mixing coefficient of the factor time series $\{f_1, \ldots, f_T\}$. If f_t is a strictly stationary α–mixing process and $\hat{\gamma}_i$ is a functional coefficient which converges to 0 at a certain rate as i increases, Assumption A.4 can easily be verified. More details and useful empirical examples can be found in the context of Assumption A2 of Chen et al. (2012*b*).

Assumption A.5 is a panel data version of Assumption 2 of Li et al. (2013) and ensures that $CV_0(\lambda)$ is uniquely minimized at 0. By Theorem 2.1 of Newey and McFadden (1994), this assumption implies that $\hat{\lambda}$ obtained by minimizing (2.6) converges to $0_{r\times 1}$ as N and T go to ∞ jointly. For more details, see the proof of Lemma 1 in the Appendix. Note that $\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})'$ is the conditional covariance matrix $\text{Cov}(X_{it}|Z_{it})$, so it is reasonable to assume positive definiteness therein.

In the panel data literature, it is common to assume that $T/N \to c \in [0,\infty)$ as $(N,T) \to (\infty,\infty)$. Since we do not impose any restriction on the relationship of N and T above, our above discussion also applies to this situation.

Having discussed the assumptions, we turn now to show our asymptotic results. Our logic is as follows. Firstly, we use Theorem 2.1 of Newey and McFadden (1994) to show that minimizing the cross–validation criterion function ensures that $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_r)' = o_P(1)$ in Lemma 1. Secondly we use this property to further investigate $CV(\lambda)$ and show that the rate of convergence is $\hat{\lambda} = O_P\left(\frac{1}{NT}\right)$ in Theorem 1. Lastly, we show the asymptotic normality in Theorem 2 based on the result of Theorem 1.

Lemma 1. *Under Assumption A, as (N,T) go to (∞,∞) jointly, $\hat{\lambda} = o_P(1)$.*

This lemma indicates that $\hat{\lambda}$ converges to 0 as the sample size increases. In addition, this lemma is also very useful to derive the convergence rate of $\hat{\lambda}$. To see this, for $s = 1, \dots, r$ we can express the kernel function of Aitchison and Aitken (1976) as

$$l(Z_{it,s}, z_s, \lambda_s) = 1(Z_{it,s} = z_s) + \lambda_s 1(Z_{it,s} \neq z_s).$$

Using Lemma 1, it is reasonable to assume that λ in the derivation for Theorem 1 is sufficiently small and close to $0_{r\times 1}$. Thus, the product kernel function can be simplified as follows

$$L(Z_{js}, Z_{it}, \lambda) = 1_{js,it} + \sum_{m=1}^{r} \lambda_m 1_{m,jsit} + O(\|\lambda\|^2),$$

where $1_{m,jsit} = 1(Z_{js,m} \neq Z_{it,m}) \prod_{n=1,n\neq m}^{r} 1(Z_{js,n} = Z_{it,n})$.

Theorem 1. *Under Assumption A, as (N,T) go to (∞,∞) jointly, $\hat{\lambda} = O_P\left(\frac{1}{NT}\right)$.*

Theorem 1 shows the rate of convergence for $\hat{\lambda}$, which is consistent with the rate shown by Li et al. (2013) for the cross–sectional case. This result is helpful to establish the asymptotic normality for $\hat{\beta}(z)$, because it significantly simplifies our proof by allowing us to use the frequency estimator (i.e., $\lambda = 0_{r\times 1}$). More details are given in the Appendix.

Theorem 2. *Under Assumption A, as (N,T) go to (∞,∞) jointly,*

$$\sqrt{NT}(\hat{\beta}(z) - \beta(z)) \to_D N(0, \Xi_1(z)^{-1}\Xi_0(z)\Xi_1(z)^{-1}),$$

where

$$\Xi_0(z) = \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[u_{it}u_{js}(X_{it}-\mu_X(z))(X_{js}-\mu_X(z))'1(Z_{it}=z)1(Z_{js}=z)\right],$$

$$\Xi_1(z) = p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)'\right).$$

We now discuss how to establish the hypothesis test based on Theorem 2. (3) of Lemma A.2 (see the Appendix) has shown that

$$\hat{\Xi}_1(z) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'1(Z_{it}=z) \to_P \Xi_1(z). \tag{2.8}$$

To consistently estimate $\Xi_0(z)$, we need to impose a stronger restriction, i.e. u_{it} is i.i.d. over i and t. This restriction is in line with the spirit of Corollary 3.1.ii and Theorem 3.3 of Gao and Phillips (2013*b*). Relevant discussions can also be found in Section 2.2.2 of Fan and Yao (2003). With this restriction, $\Xi_0(z)$ reduces to $\Xi_0(z) = p(z)\sigma_u^2\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)'\right) = \sigma_u^2\Xi_1(z)$, so all we need is a consistent estimator for σ_u^2. For this purpose, we intuitively define

$$\hat{\sigma}_u^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{Y}_{it} - \tilde{X}_{it}'\hat{\beta}(Z_{it}))^2. \tag{2.9}$$

Corollary 1. *Under Assumption A, suppose further that u_{it} is i.i.d over i and t. As (N,T) go to (∞,∞) jointly,*

$$\sqrt{NT}\left(\hat{\sigma}_u^{-2}\hat{\Xi}_1(z)\right)^{1/2}(\hat{\beta}(z) - \beta(z)) \to_D N(0, I_q),$$

where $\hat{\sigma}_u^2$ and $\hat{\Xi}_1(z)$ are defined in (2.9) and (2.8) respectively.

2.2 The Irrelevant Covariate Case

In this subsection, we consider the case where some of the covariates are irrelevant in the sense that they are independent of all other variables in the model. Without losing generality, suppose the first r_1 elements of Z_{it} are relevant while the remaining $r_2 = r - r_1$ elements of Z_{it} are irrelevant. For notational simplicity, let $\bar{Z}_{it} = (Z_{it,1}, \ldots, Z_{it,r_1})'$ denote the r_1 relevant elements and let $\tilde{Z}_{it} = (Z_{it,r_1+1}, \ldots, Z_{it,r})'$ be the r_2 irrelevant elements. Conformably, we partition λ as follows $\lambda = (\bar{\lambda}', \tilde{\lambda}')'$, where $\bar{\lambda} = (\lambda_1, \ldots, \lambda_{r_1})'$ and $\tilde{\lambda} = (\lambda_{r_1+1}, \ldots, \lambda_r)'$. Let $\bar{\mathcal{D}}$ and $\tilde{\mathcal{D}}$ denote the sets that $\bar{\lambda}$ and $\tilde{\lambda}$ belong to respectively (i.e. $\mathcal{D} = \bar{\mathcal{D}} \times \tilde{\mathcal{D}}$). With these notations, it is straightforward to obtain that $p(z) = \bar{p}(\bar{z}) \cdot \tilde{p}(\tilde{z})$ and $L(Z_{it}, z, \lambda) = \bar{L}(\bar{Z}_{it}, \bar{z}, \bar{\lambda})\tilde{L}(\tilde{Z}_{it}, \tilde{z}, \tilde{\lambda})$, where $\bar{z} = (z_1, \ldots, z_{r_1})'$, $\tilde{z} = (z_{r_1+1}, \ldots, z_r)'$, $\bar{p}(\bar{z}) = \Pr(\bar{Z}_{it} = \bar{z})$, $\tilde{p}(\tilde{z}) = \Pr(\tilde{Z}_{it} = \tilde{z})$, $\bar{L}(\bar{Z}_{it}, \bar{z}, \bar{\lambda}) = \prod_{s=1}^{r_1}\lambda_s^{1(Z_{it,s}\neq z_s)}$ and $\tilde{L}(\tilde{Z}_{it}, \tilde{z}, \tilde{\lambda}) = \prod_{s=r_1+1}^{r}\lambda_s^{1(Z_{it,s}\neq z_s)}$. Also, note that $\beta(z)$, $\mu_X(z)$ and $\Sigma_X(z)$ will respectively reduce to $\beta(\bar{z})$, $\mu_X(\bar{z})$ and $\Sigma_X(\bar{z})$ with $\bar{z} \in \bar{\mathcal{D}}$ for the irrelevant covariate case.

As in the case of the relevant covariate case, we start by introducing the assumptions necessary for establishing the asymptotic results for this case.

Assumption B:

1. The irrelevant covariates $\tilde{Z}_{it}$'s for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ are independent of all the other variables.

2. $\lambda_s \in [0, 1]$ for $s = 1, \ldots, r_1, r_1 + 1, \ldots, r$. Define

$$CV_0^*(\bar{\lambda}) = \sum_{\bar{z} \in \bar{\mathcal{D}}} \bar{p}(\bar{z})(\bar{\eta}(\bar{z}) - \beta(\bar{z}))'(\Sigma_X(\bar{z}) - \mu_X(\bar{z})\mu_X(\bar{z})')(\bar{\eta}(\bar{z}) - \beta(\bar{z})),$$

where $(\Sigma_X(\bar{z}) - \mu_X(\bar{z})\mu_X(\bar{z})')$ is positive definite uniformly. Let

$$\begin{aligned}
\bar{\Sigma}_{XX}(\bar{z}) &= E[\left(\Sigma_X(\bar{Z}_{it}) - \mu_X(\bar{Z}_{it})\mu_X(\bar{Z}_{it})'\right) \bar{L}(\bar{Z}_{it}, \bar{z}, \bar{\lambda})|\bar{z}], \\
\bar{\Sigma}_{XX\beta}(\bar{z}) &= E[\left(\Sigma_X(\bar{Z}_{it}) - \mu_X(\bar{Z}_{it})\mu_X(\bar{Z}_{it})'\right) \beta(\bar{Z}_{it})\bar{L}(\bar{Z}_{it}, \bar{z}, \bar{\lambda})|\bar{z}], \\
\bar{\eta}(\bar{z}) &= \bar{\Sigma}_{XX}^{-1}(\bar{z})\bar{\Sigma}_{XX\beta}(\bar{z}).
\end{aligned}$$

Only $\bar{\lambda} = (\lambda_1, \ldots, \lambda_{r_1})' = 0_{r_1 \times 1}$ makes $CV_0^*(\bar{\lambda}) = 0$ hold. Also, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, $E\left|(\bar{\eta}(\bar{Z}_{it}) - \beta(\bar{Z}_{it}))'(\Sigma_X(\bar{Z}_{it}) - \mu_X(\bar{Z}_{it})\mu_X(\bar{Z}_{it})')(\bar{\eta}(\bar{Z}_{it}) - \beta(\bar{Z}_{it}))\right|^2 < \infty$.

Assumption B is a panel data version of Assumption 3 of Li et al. (2013). Using Assumption B, we will first show in what follows that minimizing the cross–validation criterion function ensures that $\hat{\lambda}_s = o_P(1)$ for $s = 1, \ldots, r_1$ in Lemma 2. We will then use this assumption to further investigate $CV(\lambda)$ and show that the rate of convergence is $\hat{\lambda}_s = O_P\left(\frac{1}{\sqrt{NT}}\right)$ for $s = 1, \ldots, r_1$ in Theorem 3. Lastly, we will show $\hat{\beta}(z) - \beta(\bar{z}) = O_P\left(\frac{1}{\sqrt{NT}}\right)$ in Theorem 4.

Lemma 2. *Under Assumptions A.1–A.4 and Assumption B, as (N, T) go to (∞, ∞) jointly, $\hat{\lambda}_s = o_P(1)$ for $s = 1, \ldots, r_1$.*

Like Assumption 3 of Li et al. (2013), this lemma ensures that the CV selected smoothing parameters associated with the relevant covariates will converge to 0. Lemma 2 will be used to show Theorem 3 below in the same way as Lemma 1 is used to show Theorem 1.

Theorem 3. *Under Assumptions A.1–A.4 and Assumption B, as (N, T) go to (∞, ∞) jointly,*

1. $\hat{\lambda}_s = O_P\left(\frac{1}{\sqrt{NT}}\right)$ *for* $s = 1, \ldots, r_1$;

2. $\lim_{(N,T)\to(\infty,\infty)} \Pr\left(\hat{\lambda}_{r_1+1} = 1, \ldots, \hat{\lambda}_r = 1\right) \geq \rho$ *for some* $\rho \in (0, 1)$.

Note that the rate of convergence of $\hat{\lambda}$ for the irrelevant case is much slower than in the case of the relevant case, due to the presence of irrelevant covariates. The second result of Theorem

3 reveals that the estimates of $\hat{\lambda}_s$ for $s = r_1 + 1, \ldots, r$ are not always equal to 1, which is also consistent with the simulation results provided below. Due to the cross–sectional dependence of the error terms and the weak correlation between different time periods, the possible value of ρ becomes more complicated compared to the situation in Li et al. (2013). A numerical demonstration is provided in Figure 1.

Theorem 4. *Under Assumptions A.1–A.4 and Assumption B, as* (N, T) *go to* (∞, ∞) *jointly,* $\hat{\beta}(z) - \beta(\bar{z}) = O_P\left(\frac{1}{\sqrt{NT}}\right)$.

Using Theorem 3, it is straightforward to show Theorem 4. However, we still cannot obtain the asymptotic distribution for the irrelevant covariate case. To deal with this problem, one can follow Li et al. (2013) and use bootstrapping techniques to obtain finite sample distributions for variables of interest.

3 Monte Carlo Study

In this section, we perform a Monte Carlo study to investigate the finite sample properties of our model and estimator. The data generating process (DGP) is as follows.

$$Y_{it} = X_{it}'\beta_0(Z_{it}) + w_i + u_{it} \quad \text{and} \quad X_{it} = H_{it} + V_{it}.$$

Let $Z_{it} = (Z_{1,it}, \cdots, Z_{r,it})'$, where for $\forall j = 1, \ldots, r$ $Z_{j,it}$ is i.i.d. over i and t; and Z_{it} is chosen from $\{0, 1, 2, 3\}$ with the same probability every time, i.e. $\Pr(Z_{j,it} = 0) = \Pr(Z_{j,it} = 1) = \Pr(Z_{j,it} = 2) = \Pr(Z_{j,it} = 3) = 0.25$. V_{it} is i.i.d. over i and t and follows a normal distribution $N(Z_{1,it}/2 \cdot i_q, \sqrt{Z_{1,it} + 1} \cdot I_q)$. $H_{it} = (H_{1,it}, \ldots, H_{q,it})'$. For $\forall j = 1, \ldots, q$, $H_{j,it}$ is generated as $H_{j,it} = \rho(j) H_{j,it-1} + i.i.d.N(0, 1)$ and $\rho(j) = 0.1 * \lfloor 9 \cdot U(0, 1) \rfloor$, where $U(0, 1)$ denotes the uniform distribution; $\lfloor a \rfloor$ denotes rounding the element of a to the nearest integer greater than or equal to that element, i.e. $a \leq \lfloor a \rfloor$. Thus, for $\forall j = 1, \ldots, q$, $H_{j,it}$ is independent in the cross–sectional dimension and a stationary AR(1) process in the time–series dimension with the coefficient $\rho(j)$ being randomly chosen from the set $\{0.1, 0.2, \ldots, 0.8, 0.9\}$.

To ensure the fixed effects are correlated with the regressors and covariates, we generate them as $w_i = \frac{1}{Tq} \sum_{t=1}^{T} \sum_{j=1}^{q} X_{j,it}$. To introduce some cross–sectional dependence to the error terms, let $u_t = (u_{1t}, \ldots, u_{Nt})$ and $u_t = 0.5 u_{t-1} + \varepsilon_t$, where $\varepsilon_t \sim i.i.d.N(0_{N\times 1}, \Sigma_u)$ and for $i, j = 1, \ldots, N$ the $(i, j)^{th}$ element of Σ_u is $0.5^{|i-j|}$.

For $\forall j = 1, \ldots, q$, let $\beta_{j0}(Z_{it})$ denote the j^{th} element of the coefficient function $\beta_0(z)$. For relevant and irrelevant covariate cases, they are respectively generated as follows.

- Relevant covariate case: $\beta_{j0}(Z_{it}) = j/2 \cdot \sum_{k=1}^{r} Z_{k,it} + 1$.

- Irrelevant covariate case: $\beta_{j0}(Z_{it}) = j/2 \cdot Z_{1,it} + 1$.

We estimate our panel data categorical varying–coefficient model by (2.5) for each generated data set. For notational convenience, this method is referred to as the "DMK" model, where DM stands for demeaned variables (i.e., variables formed using the modified within transformation) and K means that the estimates are obtained using the the kernel function of Aitchison and Aitken (1976). For comparison purpose, we also estimate a variant of (2.5) where the kernel function is replaced with an indicator function. This method is referred to as "DMI". In addition, we also run the above two methods without using the demeaned data (referred to as "NDMK" and "NDMI" respectively).

For each generated data set and the corresponding estimate on $\beta_0(z)$, we calculate bias and squared error (SE) as follows.

$$\text{bias} = \sum_{z\in\mathcal{D}} p(z) \left(\hat{\beta}_j(z) - \beta_{j0}(z)\right) \quad \text{and} \quad \text{SE} = \sum_{z\in\mathcal{D}} p(z) \left(\hat{\beta}_j(z) - \beta_{j0}(z)\right)^2,$$

where, for $j = 1, \ldots, q$, $\hat{\beta}_j(z)$ denotes the j^{th} element of $\hat{\beta}(z)$. We then let $q = 3$, $r = 2$ and replicate the above procedure 1000 times. Tables 1 and 2 present the mean bias (referred to as Bias) and mean squared error (MSE) for the relevant case and the irrelevant case respectively. Since the differences of the biases are down to the fifth or sixth decimal, we follow Pesaran (2006) and multiply each bias by 100 for the sake of exposition. As can be seen from Table 1, both DMK and DMI provide much more accurate estimates than NDMK and NDMI in that both of the former two have much smaller mean biases and mean squared errors. This result suggests the importance of demeaning the data using the within transformation in a panel data context when the fixed effects are included. This conclusion is further confirmed by the results for the irrelevant case, shown in Table 2, because both DMK and DMI provide much more accurate estimates too. Moreover, Table 2 shows that DMK outperforms DMI in that DMK has smaller MSE overall. In order to examine the second result of Theorem 3, we plot the distribution of $\hat{\lambda}_2$ for the irrelevant case in Figure 1. It is clear that $\hat{\lambda}_2$ is equal to 1 with high probability, but $\hat{\lambda}_2$ does not converge to 1.

4 An Application to U.S. Commercial Banks

In Section 2 we discussed the varying–coefficient model in a panel data setting and established its asymptotic properties. In this section we provide an application of this model to the analysis of the effects of geographical deregulation on the returns to scale of commercial banks in the U.S., which has attracted much attention over the past two decades. Until the middle of the 1970's banking in the U.S. was heavily regulated at the state level. Generally, there were three different types of state regulation on bank branching: "unit banking", where banks were only permitted to operate in one location (i.e., no branches); "limited branching", where the branching abilities of individual banks were limited to a portion of the state; and "statewide banking", where individual banks were permitted to branch statewide. In the mid–1980s individual states began to loosen regulations on intrastate branching, often moving from unit banking to limited branching and then to statewide banking. It is worth noting that different states changed their regulatory restrictions on expansion at different times. This deregulation process eventually culminated in the passage of the Riegle–Neal Interstate Banking and Branching Efficiency Act of 1994, which permitted nationwide branching as of June 1997[2]. In sum, commercial banks in the U.S. underwent four branch banking regimes in the 1980s and 1990s: (1) unit banking, (2) limited branching, (3) statewide banking, and (4) full interstate branching, thus offering researchers a unique opportunity to study the effects of geographical deregulation on the returns to scale of commercial banks in the U.S..

The data used in this application are obtained from the Reports of Income and Condition (Call Reports) published by the Federal Reserve Bank of Chicago. The sample covers the period 1986–2005, a period that includes the four policy regimes. We examine only continuously operating large banks with assets of at least $1 billion (in 1986 dollars) to avoid the impact of entry and exit and to focus on the performance of a core of healthy, surviving institutions. This gives a total of 466 banks over 20 years. To select the relevant variables, we follow the commonly–accepted intermediation approach (c.f. Sealey and Lindley (1977)). On the input side, we follow Wheelock and Wilson (2012) and specify three input prices: (1) price of labor; (2) price of purchased funds; and (3) price of core deposits. On the output side, three outputs are specified: (1) securities, which includes all non–loan financial assets (i.e., all financial assets minus the sum of all loans, securities, and equity); (2) consumer loans; and (3) non–consumer loans, which is composed of industrial, commercial, and real estate loans. All the quantities are constructed by following the data construction method in Berger and Mester (2003). These quantities are also deflated by the GDP deflator to the base year 1986, except for the quantity of labor. Each input price is obtained by dividing total expenditure on this input by its corresponding quantity.

4.1 The varying–coefficient translog cost function

We use a varying–coefficient translog cost function to represent the production technology of commercial banks in the U.S.. As can be seen below, this cost function has the standard form of

[2]States had the option to pass legislation to "opt in" earlier than the June 1997 federal deadline, which resulted in the Act becoming effective gradually among the U.S. states between 1994–1997. States were also allowed to pass legislation to opt out of the provisions for interstate branching, with Texas and Montana being the only states to do so. For a detailed discussion on the history of the geographical deregulation on banking in the U.S., see Jayaratne and Strahan (1997).

the varying–coefficient model described in Section 2. A primary feature of this function is that its coefficients are allowed to vary depending on the banking regime under which a bank operates, because there is considerable evidence that branch banking regime affects production technology (c.f. Watkins and West (1982); Mason (2013); Mester (2005)). Specifically, suppressing the index i and t, the varying–coefficient translog cost function is written as[3]

$$\begin{aligned}\ln C = {} & \alpha_0(Z) + \sum_{j=1}^{\bar{N}} \alpha_j(Z) \ln W_j + \sum_{m=1}^{\bar{M}} \gamma_m(Z) \ln Y_m + \tau(Z)t + \frac{1}{2}\delta(Z)t^2 \\ & + \frac{1}{2}\sum_{j=1}^{\bar{N}}\sum_{k=1}^{\bar{N}} \beta_{jk}(Z) \ln W_j \ln W_k + \frac{1}{2}\sum_{m=1}^{\bar{M}}\sum_{n=1}^{\bar{M}} \rho_{mn}(Z) \ln Y_m \ln Y_n \\ & + \sum_{j=1}^{\bar{N}}\sum_{m=1}^{\bar{M}} \psi_{jm}(Z) \ln W_j \ln Y_m + \sum_{j=1}^{\bar{N}} \phi_j(Z) t \ln W_j + \sum_{m=1}^{\bar{M}} \varphi_m(Z) t \ln Y_m, \end{aligned} \tag{4.1}$$

where C is variable cost as in Wheelock and Wilson (2012); t is a time trend; Y_m for $m = 1, \ldots, \bar{M}$ is a variable representing output; and W_j for $j = 1, \ldots, \bar{N}$ is a variable representing input price. In our particular case, $\bar{N} = \bar{M} = 3$. Z is specified to be a four–category variable indicating different branch banking regimes that existed during our sample period. Specifically, we set $Z = 1$ for banks operating in unit banking states, $Z = 2$ for banks operating in limited branching states, $Z = 3$ for banks operating in statewide banking states, and $Z = 4$ for banks operating in nationwide branching states. As previously noted, different states changed their regulatory restrictions on expansion at different times, indicating that Z varies in both the cross–sectional and time series dimensions.

The usual symmetry restrictions require $\beta_{jk}(Z) = \beta_{kj}(Z)$ for $j, k = 1, \ldots, \bar{N}$ and $\rho_{mn}(Z) = \rho_{nm}(Z)$ for $m, n = 1, \ldots, \bar{M}$. Moreover, to ensure linear homogeneity of the cost function in input prices, the following restrictions are imposed

$$\sum_{j=1}^{\bar{N}} \alpha_j(Z) = 1, \quad \sum_{j=1}^{\bar{N}} \beta_{jk}(Z) = \sum_{j=1}^{\bar{N}} \psi_{jm}(Z) = \sum_{j=1}^{\bar{N}} \phi_j(Z) = 0. \tag{4.2}$$

To impose the linear homogeneity restrictions in (4.2), we follow Griffiths et al. (2000) and normalize the cost and input prices in (4.1) by one of the input prices (say, $W_{\bar{N}}$)

$$\begin{aligned}\ln \frac{C}{W_{\bar{N}}} = {} & \alpha_0(Z) + \sum_{j=1}^{\bar{N}-1} \alpha_j(Z) \ln \frac{W_j}{W_{\bar{N}}} + \sum_{m=1}^{\bar{M}} \gamma_m(Z) \ln Y_m + \tau(Z)t + \frac{1}{2}\delta(Z)t^2 \\ & + \frac{1}{2}\sum_{j=1}^{\bar{N}-1}\sum_{k=1}^{\bar{N}-1} \beta_{jk}(Z) \ln \frac{W_j}{W_{\bar{N}}} \ln \frac{W_k}{W_{\bar{N}}} + \frac{1}{2}\sum_{m=1}^{\bar{M}}\sum_{n=1}^{\bar{M}} \rho_{mn}(Z) \ln Y_m \ln Y_n \end{aligned}$$

[3]There are two methods to estimate this cost function: one is to estimate it directly and the other is to estimate it together with its share equations. From an economic theoretical perspective, both methods are correct although the second one gives better statistical efficiency (see, for example, Feng and Serletis (2008)). However, to better illustrate our single equation panel data varying–coefficient model, we use the first method in this paper.

$$+\sum_{j=1}^{\bar{N}-1}\sum_{m=1}^{\bar{M}}\psi_{jm}(Z)\ln\frac{W_j}{W_{\bar{N}}}\ln Y_m+\sum_{j=1}^{\bar{N}-1}\phi_j(Z)t\ln\frac{W_j}{W_{\bar{N}}}+\sum_{m=1}^{\bar{M}}\varphi_m(Z)t\ln Y_m. \quad (4.3)$$

In matrix notations, the normalized varying–coefficient translog cost function in (4.3), after appending a fixed effect term and a random error term, can be written as (2.1), where the dependent variable is $\ln\frac{C}{W_{\bar{N}}}$; the regressors are a vector comprising all the variables which appear on the right hand side of (4.3); $\beta(\cdot)$ is the corresponding vector of coefficients of the translog function. Note that after the within transformation $\alpha_0(Z)$ will disappear along with the fixed effect. However, this does not affect our empirical results.

Given the estimated parameters from the varying–coefficient translog cost function (4.3), it is possible to compute returns to scale as RTS $=\left(\sum_{m=1}^{\bar{M}}\epsilon_{cY_m}\right)^{-1}$, where for $m=1,\ldots,\bar{M}$

$$\epsilon_{cY_m}=\frac{\partial\ln C}{\partial\ln Y_m}=\gamma_m(Z)+\sum_{n=1}^{\bar{M}}\rho_{mn}(Z)\ln Y_n+\sum_{j=1}^{\bar{N}}\psi_{jm}(Z)\ln W_j+\varphi_m(Z)t$$

is the cost elasticity of the j^{th} output.

4.2 The fully parametric translog cost function

For comparison purpose, we also consider a widely used fully parametric translog cost function to investigate the effects of geographical deregulation on the productivity and efficiency of commercial banks in the U.S. Three binary variables are used to control for the different branch banking regimes. Specifically, we define (i) UNIT $=1$ for banks operating in unit banking states (UNIT $=0$ otherwise); (ii) LIMITED $=1$ for banks operating in limited branching states (LIMITED $=0$ otherwise); and (iii) STATEWIDE $=1$ for banks operating in statewide banking states (STATEWIDE $=0$ otherwise). We set UNIT $=$ LIMITED $=$ STATEWIDE $=0$ to reflect the nationwide branching regime. With the definition of these binary variables, the fully parametric translog cost function is written as

$$\begin{aligned}\ln C=&\alpha_0+\sum_{j=1}^{\bar{N}}\alpha_j\ln W_j+\sum_{m=1}^{\bar{M}}\gamma_m\ln Y_m+\tau t+\frac{1}{2}\delta t^2+\frac{1}{2}\sum_{j=1}^{\bar{N}}\sum_{k=1}^{\bar{N}}\beta_{jk}\ln W_j\ln W_k\\&+\frac{1}{2}\sum_{m=1}^{\bar{M}}\sum_{n=1}^{\bar{M}}\rho_{mn}\ln Y_m\ln Y_n+\sum_{j=1}^{\bar{N}}\sum_{m=1}^{\bar{M}}\psi_{jm}\ln W_j\ln Y_m+\sum_{j=1}^{\bar{N}}\phi_j t\ln W_j\\&+\sum_{m=1}^{\bar{M}}\varphi_m t\ln Y_m+\xi_1\text{UNIT}+\xi_2\text{LIMITED}+\xi_3\text{STATEWIDE}. \quad (4.4)\end{aligned}$$

Symmetry requires $\beta_{jk}=\beta_{kj}$ for $j,k=1,\ldots,\bar{N}$ and $\rho_{mn}=\rho_{nm}$ for $m,n=1,\ldots,\bar{M}$. Moreover, linear homogeneity in inputs prices implies the following restrictions:

$$\sum_{j=1}^{\bar{N}} \alpha_j = 1, \quad \sum_{j=1}^{\bar{N}} \beta_{jk} = \sum_{j=1}^{\bar{N}} \psi_{jm} = \sum_{j=1}^{\bar{N}} \phi_j = 0. \tag{4.5}$$

To impose the linear homogeneity restrictions in (4.5), we normalize the cost and input prices in (4.4) by $W_{\bar{N}}$

$$\begin{aligned} \ln \frac{C}{W_{\bar{N}}} &= \alpha_0 + \sum_{j=1}^{\bar{N}-1} \alpha_j \ln \frac{W_j}{W_{\bar{N}}} + \sum_{m=1}^{\bar{M}} \gamma_m \ln Y_m + \tau t + \frac{1}{2}\delta t^2 + \frac{1}{2} \sum_{j=1}^{\bar{N}-1} \sum_{k=1}^{\bar{N}-1} \beta_{jk} \ln \frac{W_j}{W_{\bar{N}}} \ln \frac{W_k}{W_{\bar{N}}} \\ &+ \frac{1}{2} \sum_{m=1}^{\bar{M}} \sum_{n=1}^{\bar{M}} \rho_{mn} \ln Y_m \ln Y_n + \sum_{j=1}^{\bar{N}-1} \sum_{m=1}^{\bar{M}} \psi_{jm} \ln \frac{W_j}{W_{\bar{N}}} \ln Y_m + \sum_{j=1}^{\bar{N}-1} \phi_j t \ln \frac{W_j}{W_{\bar{N}}} \\ &+ \sum_{m=1}^{\bar{M}} \varphi_m t \ln Y_m + \xi_1 \mathrm{UNIT} + \xi_2 \mathrm{LIMITED} + \xi_3 \mathrm{STATEWIDE}, \end{aligned} \tag{4.6}$$

In matrix notations, the normalized fully parametric translog cost function in (4.6), after appending a fixed effect term and a random error term, can be written as

$$Y_{it} = X_{it}'\beta_0 + w_i + u_{it}, \tag{4.7}$$

where X_{it} is a vector comprising all the variables which appear on the right hand side of (4.6); and β_0 is the corresponding vector of coefficients of the translog function (including the intercept). As with the case of the varying coefficient translog cost function, we also perform the within transformation on (4.7) to get rid of the fixed effects.

4.3 Statistical comparison of the two competing models

We estimate the varying–coefficient translog cost function and the fully parametric translog cost function separately for the U.S. large banks. Parameter estimates and the associated standard errors for the two models are reported in Tables 5–6. To compare the performance of these two competing models, we perform a test using the procedure proposed by Li et al. (2013). For the fully parametric model, if we treat $(\alpha_0 + \xi_1\mathrm{UNIT} + \xi_2\mathrm{LIMITED} + \xi_3\mathrm{STATEWIDE})$ in (4.6) as the coefficient for the constant term, then a comparison of (4.6) with the varying–coefficient translog cost function in (4.3) reveals that the former model is a special case of the latter model. With this in mind, then, testing if the varying–coefficient translog cost function outperforms the fully parametric translog cost function is equivalent to testing if the latter model has the same specification as the former model, or more specifically, if the latter model has the same set of coefficients as the former model.

To test parameter constancy, we simply extend the bootstrap based procedure provided by Li et al. (2013) to the panel data setting. Thus, all the details and the references can be found therein. For our particular case, the test statistic is 0.4968, well above the critical value of 0.0876 at 1% level of significance. This result suggests strongly that the null hypothesis is rejected; in other words, the varying–coefficient translog cost function is preferred to the fully parametric translog cost function.

To further confirm this conclusion, we perform another test comparing the competing model by using our Corollary 1. Specifically, let $\beta(z^j)$ for $j = 1, \ldots, 4$ denote the vector of coefficients of the varying–coefficient translog cost function for branch banking regimes 1–4 respectively. Let β_0 denote the estimated coefficient (excluding ξ_1, ξ_2 and ξ_3) obtained from the fully parametric model. The null hypothesis of this latter test is then written as: $H_0 : \beta(z^j) = \beta_0$ for $j = 1, \ldots, 4$ respectively. Our results indicate that the associated p–values are zero for all the four banking regimes, which confirms that the varying–coefficient translog cost function outperforms the fully parametric translog cost function. Thus, in what follows we will focus on the empirical results obtained from the varying–coefficient translog cost function.

4.4 Empirical results from the varying–coefficient translog cost function

Table 7 presents the annual average returns to scale (RTS) estimates over the sample period, obtained by averaging over all sampled banks in each year. As can be seen from this table, the point estimates of average RTS are all greater than one, ranging from 1.0458 to 1.0688, suggesting that on average the commercial banks exhibit increasing returns to scale. This finding is consistent with those of Feng and Serletis (2010), Feng and Zhang (2012) and Wheelock and Wilson (2012). For example, Wheelock and Wilson (2012), using a non–parametric local–linear estimator to estimate the cost relationship for commercial banks in the U.S. over the period 1984–2006, find that U.S. banks operated under increasing returns to scale.

Considering that our sample period spans all the four banking regimes, it is of interest to compare the estimates of RTS across different regimes. For this purpose, we calculate the average RTS for each policy regime in each year by averaging within each regime in that year. The results are reported in Table 8, where N/A indicates that the corresponding policy regime doesn't exist or expires in that year. Looking at this table, we see that generally speaking, average RTS is higher in more regulated states than in less regulated states in a given year. Taking 1986 for example, average RTS is 1.0995 for unit banking states, as compared to 1.0407 for limited branching states and 1.0361 for statewide branching states. To give another example, in 1998 average RTS for statewide branching states is 1.0818, as compared to 1.0550 for national branching states. This result suggests that banks in more regulated states are forced to operate at scales further below their optimal scales than those in less regulated states. It is worth noting at this point that optimal scales in less regulated states are much higher than those in more regulated states. To illustrate this point, we calculate the optimal scale for each policy regime in

1986 by averaging total assets across banks under that policy regime that face constant returns to scale. Our result shows that the optimal scale for statewide branching states is \$1.177 billion, as compared to \$1 million for unit banking states and \$4 million for limited branching states. The increase in optimal scale from more regulated policy regime to less regulated policy regime suggests that geographical deregulation greatly changes banking production technology in the U.S.. Another interesting finding that emerges from Table 8 is that average RTS for statewide branching regime and that for national branching regime have increased over time. A possible explanation is that as banks grow bigger under less regulated regimes, they are more likely to afford new technologies. The adoption of new technologies further increases the banks' optimal scales over time, which results in higher RTS for given bundles of inputs.

In addition to the annual average RTS estimates, we are also interested in the estimates of RTS at individual bank level. We compute the percentage of banks facing increasing, constant, or decreasing returns to scale for each year. This computation is performed by counting the number of cases where the 95% confidence intervals[4] are strictly greater than 1.0 (indicating increasing returns to scale), strictly less than 1.0 (indicating decreasing returns to scale), or contain 1.0 (indicating constant returns to scale). The results are presented in Table 9. Two findings emerge from this table. First, on average the majority of the banks face increasing returns to scale, a small percentage face decreasing returns to scale, and an even smaller percentage face constant returns to scale. Specifically, on average 91.30% of the banks face increasing returns to scale, 5.34% face decreasing returns to scale, and 3.36% face constant returns to scale. Second, the percentage of banks facing increasing returns to scale shows a "first increase and then stabilize" pattern, the percentage of banks facing decreasing returns to scale shows a "first decrease and then stabilize" pattern, and the percentage of banks facing constant returns to scale also shows a "first decrease and then stabilize" pattern. Specifically, the percentage of banks facing increasing returns to scale increases markedly from 74.89% in 1986 to 96.78% in 2002 and then stabilizes at around that level for the rest of the sample period; the percentage of banks facing decreasing returns to scale decreases noticeably from 13.52% in 1986 to 1.93% in 2001 and then stabilizes at around that level afterwards (with the exception of the last year when the percentage goes up to 8.22%); and the percentage of banks facing constant returns to scale falls consistently from 11.59% in 1986 to 1.29% in 2002 stabilizes at around that level afterwards. This result is consistent with our previous discussion that both geographical deregulation and subsequent technological adoptions increase the bank's optimal scales over time, leaving more and more banks operating under increasing returns to scale.

[4]The confidence intervals are obtained by following the standard wild bootstrap procedure.

5 Discussion and Conclusion

We now briefly discuss how to deal with the case where the cardinality of $\mathcal{D}$ is infinite. Suppose that $r = 1$. $Z_{it} \in \{0, 1, 2, \ldots, \nu(N, T) - 1\}$ where $\nu(N, T) \to \infty$ and $\nu(N, T)/(NT) \to c$ for $0 \leq c < \infty$ as $(N, T) \to (\infty, \infty)$. In this case, the following model can be considered

$$Y_{it} = X'_{it}\beta(Z_{it}/\nu(N, T)) + w_i + u_{it}, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T \tag{5.1}$$

Here we can treat $\beta(\cdot)$ as a function with continuous covariates. (5.1) then becomes the model proposed by Sun et al. (2009). This normalization technique is similar to the one employed by Cai (2007) and Chen et al. (2012*b*) in dealing with time varying–coefficient model.

In this paper, we extend Li et al. (2013)'s cross–sectional varying–coefficient model to a panel data context, where fixed effects are included to allow for correlation between individual unobserved heterogeneity and the regressors. In dealing with the fixed effects, we do not impose any identification restriction as done in previous studies. Instead, we take advantage of the fact that our covariates are categorical, and use a modified within transformation. We show the exact asymptotic properties of our estimator for the relevant covariate case and the irrelevant covariate case. We further conduct a Monte Carlo study investigating the finite sample properties of our estimator.

Finally, we show how this estimator can be used by analyzing the effects of state–level banking regulations on the returns to scale of commercial banks in the U.S. over the period 1986–2005. Specifically, we estimate a varying–coefficient translog cost function, where branch banking regime is used as a covariate of the varying coefficient. We compare this cost function with a fully parametric cost function where branch banking regimes are treated as binary variables. Our tests reject the latter cost function in favor of the former one. Our empirical results from the varying–coefficient translog cost function show that returns to scale is higher in more regulated states than in less regulated states. Our results also indicate that the majority of the banks face increasing returns to scale, a small percentage face decreasing returns to scale, and an even smaller percentage face constant returns to scale.

Table 1: Relevant covariate case

			DMK			DMI			NDMK			NDMI		
			$N=40$	60	80	$N=40$	60	80	$N=40$	60	80	$N=40$	60	80
$\hat{\beta}_1(z)$	Bias ($\times$ 100)	$T=40$	0.0315	0.0027	-0.0289	-0.0103	-0.0243	-0.0495	8.9987	8.8838	8.7957	8.9184	8.8310	8.7558
		60	0.0201	0.0295	-0.0076	-0.0044	0.0131	-0.0198	8.2586	8.1986	8.1871	8.2068	8.1642	8.1613
		80	-0.0032	-0.0075	0.0455	-0.0210	-0.0194	0.0366	7.8970	7.9105	7.9309	7.8583	7.8851	7.9115
	MSE	$T=40$	0.0047	0.0031	0.0022	0.0047	0.0031	0.0022	0.0134	0.0118	0.0110	0.0133	0.0118	0.0110
		60	0.0026	0.0017	0.0012	0.0026	0.0017	0.0012	0.0108	0.0098	0.0093	0.0108	0.0098	0.0093
		80	0.0017	0.0012	0.0009	0.0017	0.0012	0.0009	0.0096	0.0090	0.0087	0.0095	0.0090	0.0086
$\hat{\beta}_2(z)$	Bias ($\times$ 100)	$T=40$	0.1053	0.0608	0.0426	0.0217	0.0092	0.0027	8.8522	9.0115	8.8767	8.7511	8.9489	8.8277
		60	0.0241	0.0255	0.0509	-0.0243	-0.0068	0.0264	8.2584	8.2056	8.2655	8.1954	8.1642	8.2334
		80	-0.0187	0.0074	0.0360	-0.0544	-0.0163	0.0181	7.8488	7.9398	7.9650	7.8013	7.9083	7.9409
	MSE	$T=40$	0.0048	0.0030	0.0023	0.0048	0.0030	0.0023	0.0132	0.0120	0.0112	0.0131	0.0119	0.0111
		60	0.0026	0.0017	0.0013	0.0026	0.0017	0.0013	0.0107	0.0097	0.0095	0.0106	0.0097	0.0094
		80	0.0017	0.0012	0.0009	0.0017	0.0012	0.0009	0.0095	0.0090	0.0087	0.0094	0.0090	0.0087
$\hat{\beta}_3(z)$	Bias ($\times$ 100)	$T=40$	0.1157	0.0545	0.0101	-0.0072	-0.0259	-0.0506	8.9556	8.8254	8.7616	8.8357	8.7478	8.7028
		60	0.0834	0.0885	0.0715	0.0091	0.0393	0.0352	8.2555	8.2883	8.2688	8.1791	8.2382	8.2311
		80	0.0632	0.0446	0.0075	0.0093	0.0084	-0.0192	8.0007	7.8676	7.8617	7.9438	7.8294	7.8334
	MSE	$T=40$	0.0047	0.0031	0.0023	0.0048	0.0031	0.0023	0.0133	0.0118	0.0109	0.0133	0.0117	0.0109
		60	0.0025	0.0017	0.0012	0.0025	0.0017	0.0012	0.0107	0.0100	0.0094	0.0106	0.0100	0.0094
		80	0.0018	0.0012	0.0009	0.0018	0.0012	0.0009	0.0097	0.0089	0.0085	0.0097	0.0089	0.0085

Since the differences of the biases are down to the fifth or sixth decimal, we adopt the approach in Pesaran (2006), i.e. multiply the biases by 100 and report the results in the above table.

Table 2: Irrelevant covariate case

			DMK			DMI			NDMK			NDMI		
			$N = 40$	60	80	$N = 40$	60	80	$N = 40$	60	80	$N = 40$	60	80
$\hat{\beta}_1(z)$	Bias ($\times$ 100)	$T = 40$	0.0099	-0.0145	-0.0346	-0.0212	-0.0320	-0.0475	8.9169	8.8174	8.7530	8.9047	8.8169	8.7520
		60	-0.0166	0.0239	-0.0187	-0.0194	0.0139	-0.0224	8.1980	8.1664	8.1703	8.1957	8.1621	8.1654
		80	-0.0034	-0.0181	0.0392	-0.0043	-0.0186	0.0329	7.8748	7.8810	7.9114	7.8802	7.8819	7.9066
	MSE	$T = 40$	0.0015	0.0010	0.0007	0.0047	0.0031	0.0022	0.0104	0.0098	0.0096	0.0133	0.0117	0.0110
		60	0.0008	0.0005	0.0004	0.0026	0.0017	0.0012	0.0089	0.0086	0.0084	0.0108	0.0098	0.0093
		80	0.0005	0.0004	0.0003	0.0017	0.0012	0.0009	0.0082	0.0080	0.0080	0.0096	0.0090	0.0086
$\hat{\beta}_2(z)$	Bias ($\times$ 100)	$T = 40$	0.0497	0.0354	0.0246	0.0290	0.0177	0.0048	8.7476	8.9503	8.8346	8.7455	8.9601	8.8223
		60	-0.0056	-0.0031	0.0320	-0.0292	-0.0096	0.0240	8.1997	8.1642	8.2317	8.1976	8.1617	8.2304
		80	-0.0143	-0.0104	0.0230	-0.0300	-0.0182	0.0174	7.9007	7.9091	7.9314	7.8946	7.9080	7.9312
	MSE	$T = 40$	0.0015	0.0010	0.0007	0.0048	0.0030	0.0023	0.0102	0.0101	0.0097	0.0131	0.0119	0.0111
		60	0.0008	0.0005	0.0004	0.0026	0.0017	0.0013	0.0088	0.0085	0.0085	0.0106	0.0097	0.0095
		80	0.0005	0.0003	0.0002	0.0017	0.0012	0.0009	0.0082	0.0080	0.0080	0.0096	0.0090	0.0087
$\hat{\beta}_3(z)$	Bias ($\times$ 100)	$T = 40$	0.0363	-0.0021	-0.0354	-0.0082	-0.0261	-0.0627	8.8678	8.7433	8.7117	8.8536	8.7348	8.6998
		60	0.0230	0.0361	0.0516	0.0067	0.0236	0.0397	8.1850	8.2370	8.2403	8.1818	8.2327	8.2389
		80	-0.0121	0.0158	-0.0025	-0.0256	0.0076	0.0119	7.8745	7.8306	7.8392	7.8701	7.8252	7.8388
	MSE	$T = 40$	0.0016	0.0010	0.0008	0.0047	0.0031	0.0023	0.0104	0.0098	0.0095	0.0133	0.0117	0.0109
		60	0.0008	0.0005	0.0004	0.0025	0.0017	0.0012	0.0088	0.0087	0.0085	0.0107	0.0099	0.0094
		80	0.0005	0.0003	0.0003	0.0017	0.0012	0.0009	0.0081	0.0079	0.0078	0.0095	0.0088	0.0085

Since the differences of the biases are down to the fifth or sixth decimal, we adopt the approach in Pesaran (2006), i.e. multiply the biases by 100 and report the results in the above table.

Table 3: The estimated bandwidth of the relevant covariate case

		DMK						NDMK					
		$\hat{\lambda}_1$			$\hat{\lambda}_2$			$\hat{\lambda}_1$			$\hat{\lambda}_2$		
		$N = 40$	60	80	$N = 40$	60	80	$N = 40$	60	80	$N = 40$	60	80
Mean	$T = 40$	0.00027	0.00018	0.00014	0.00034	0.00023	0.00017	0.00014	0.00009	0.00007	0.00028	0.00019	0.00014
	60	0.00017	0.00011	0.00008	0.00021	0.00014	0.00011	0.00009	0.00006	0.00005	0.00019	0.00013	0.00009
	80	0.00013	0.00008	0.00006	0.00016	0.00011	0.00008	0.00007	0.00005	0.00001	0.00014	0.00009	0.00002
Std	$T = 40$	0.00012	0.00007	0.00004	0.00014	0.00008	0.00005	0.00006	0.00004	0.00002	0.00009	0.00005	0.00003
	60	0.00006	0.00003	0.00002	0.00007	0.00004	0.00003	0.00003	0.00002	0.00001	0.00005	0.00003	0.00002
	80	0.00004	0.00002	0.00003	0.00005	0.00003	0.00007	0.00002	0.00001	0.00001	0.00003	0.00002	0.00001

Table 4: The estimated bandwidth of the irrelevant covariate case

		DMK						NDMK					
		$\hat{\lambda}_1$			$\hat{\lambda}_2$			$\hat{\lambda}_1$			$\hat{\lambda}_2$		
		$N = 40$	60	80	$N = 40$	60	80	$N = 40$	60	80	$N = 40$	60	80
Mean	$T = 40$	0.00010	0.00007	0.00005	0.44049	0.45190	0.47013	0.00004	0.00003	0.00002	0.82024	0.81796	0.82321
	60	0.00006	0.00004	0.00003	0.53770	0.55091	0.55048	0.00002	0.00002	0.00003	0.82487	0.81892	0.26469
	80	0.00004	0.00003	0.00002	0.61206	0.60602	0.61568	0.00002	0.00001	0.00001	0.81913	0.81726	0.82325
Std	$T = 40$	0.00012	0.00008	0.00006	0.24863	0.23932	0.25247	0.00003	0.00002	0.00001	0.23662	0.23412	0.23189
	60	0.00006	0.00004	0.00003	0.26223	0.27068	0.82190	0.00002	0.00001	0.00001	0.23422	0.22814	0.23399
	80	0.00004	0.00003	0.00002	0.26731	0.26951	0.27080	0.00002	0.00001	0.00001	0.23082	0.23647	0.23218

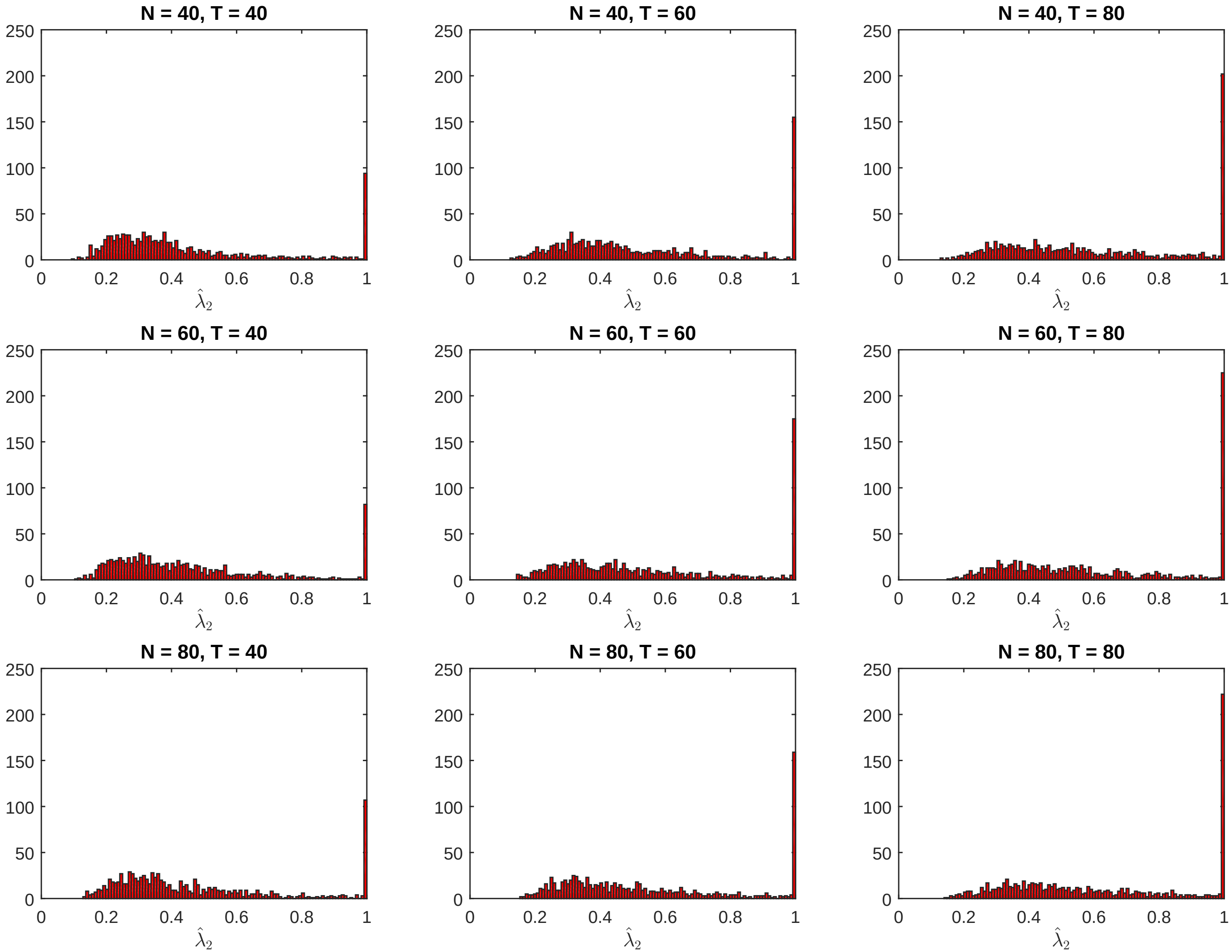

Figure 1: Distribution of $\hat{\lambda}_2$ for $N = 40, 60, 80$ and $T = 40, 60, 80$

Table 5: Parameter Estimates from the Varying–Coefficient Model

	Unit		Limited		Statewide		Nationwide	
Coefficients	Est	Std	Est	Std	Est	Std	Est	Std
α_1	0.3912	0.0295	0.3343	0.0129	0.3442	0.0055	0.4199	0.0091
α_2	0.0556	0.0121	0.0411	0.0056	0.0539	0.0031	0.0574	0.0033
α_3	0.5532	0.0371	0.6246	0.0140	0.6019	0.0059	0.5227	0.0097
γ_1	0.1844	0.0507	0.2823	0.0129	0.3079	0.0064	0.3636	0.0062
γ_2	0.0514	0.0169	0.1381	0.0088	0.1218	0.0043	0.1163	0.0048
γ_3	0.6368	0.0252	0.4759	0.0133	0.5068	0.0072	0.4877	0.0082
β_{11}	0.1055	0.0440	0.0612	0.0204	0.1758	0.0096	0.2159	0.0124
β_{12}	-0.0001	0.0071	0.0045	0.0033	-0.0276	0.0027	-0.0173	0.0029
β_{13}	-0.1054	0.0477	-0.0657	0.0213	-0.1482	0.0091	-0.1986	0.0115
β_{22}	0.0222	0.0031	0.0155	0.0010	0.0170	0.0014	0.0178	0.0011
β_{23}	-0.0221	0.0070	-0.0200	0.0033	0.0106	0.0026	-0.0005	0.0031
β_{33}	0.1274	0.0520	0.0857	0.0223	0.1376	0.0092	0.1992	0.0113
ρ_{11}	0.1126	0.0220	0.1196	0.0119	0.1464	0.0045	0.1567	0.0049
ρ_{12}	0.0170	0.0158	0.0082	0.0069	-0.0099	0.0024	-0.0095	0.0027
ρ_{13}	-0.1727	0.0108	-0.1521	0.0063	-0.1530	0.0029	-0.1493	0.0052
ρ_{22}	0.0145	0.0186	0.0601	0.0038	0.0351	0.0025	0.0190	0.0027
ρ_{23}	-0.0341	0.0110	-0.0568	0.0055	-0.0158	0.0032	0.0015	0.0034
ρ_{33}	0.2414	0.0137	0.2011	0.0060	0.1758	0.0045	0.1460	0.0076
τ	-0.0034	0.0009	-0.0035	0.0004	-0.0031	0.0002	-0.0071	0.0006
δ	-0.0002	0.0000	-0.0001	0.0000	0.0000	0.0000	0.0002	0.0000
ψ_{11}	0.0353	0.0182	0.0333	0.0137	0.0026	0.0039	0.0153	0.0040
ψ_{12}	0.0072	0.0112	-0.0169	0.0097	-0.0128	0.0030	0.0012	0.0029
ψ_{13}	-0.0244	0.0139	-0.0090	0.0082	0.0174	0.0046	-0.0146	0.0043
ψ_{21}	-0.0040	0.0043	0.0036	0.0020	0.0149	0.0016	-0.0032	0.0017
ψ_{22}	-0.0004	0.0040	-0.0016	0.0019	-0.0069	0.0014	-0.0022	0.0011
ψ_{23}	0.0068	0.0039	0.0031	0.0019	-0.0010	0.0017	0.0102	0.0019
ψ_{31}	-0.0313	0.0199	-0.0368	0.0144	-0.0175	0.0042	-0.0121	0.0043
ψ_{32}	-0.0069	0.0124	0.0185	0.0106	0.0197	0.0029	0.0010	0.0031
ψ_{33}	0.0176	0.0139	0.0059	0.0084	-0.0164	0.0048	0.0044	0.0045
ϕ_1	0.0029	0.0008	0.0018	0.0004	0.0001	0.0002	-0.0045	0.0006
ϕ_2	0.0002	0.0002	-0.0004	0.0001	0.0007	0.0001	0.0006	0.0001
ϕ_3	-0.0031	0.0008	-0.0014	0.0004	-0.0008	0.0002	0.0039	0.0006
φ_1	-0.0005	0.0004	-0.0011	0.0002	-0.0002	0.0001	-0.0017	0.0002
φ_2	-0.0008	0.0005	0.0009	0.0002	-0.0002	0.0001	-0.0002	0.0002
φ_3	0.0009	0.0005	0.0001	0.0002	-0.0002	0.0001	0.0015	0.0003

Est and Std stand for estimates and standard deviation, respectively.

Table 6: Parameter Estimates from the Fully Parametric Model

Coefficients	Est	Std
α_1	0.3554	0.0045
α_2	0.0633	0.0019
α_3	0.5814	0.0047
γ_1	0.3213	0.0034
γ_1	0.1304	0.0023
γ_3	0.5022	0.0040
β_{11}	0.1797	0.0128
β_{12}	-0.0190	0.0021
β_{31}	-0.1607	0.0120
β_{22}	0.0162	0.0007
β_{23}	0.0028	0.0021
β_{33}	0.1579	0.0115
γ_{11}	0.1463	0.0036
γ_{12}	-0.0028	0.0020
γ_{13}	-0.1527	0.0026
γ_{22}	0.0280	0.0012
γ_{23}	-0.0141	0.0021
γ_{33}	0.1669	0.0033
τ	-0.0031	0.0001
δ	0.0000	0.0000
ψ_{11}	0.0235	0.0036
ψ_{12}	-0.0059	0.0025
ψ_{13}	-0.0103	0.0036
ψ_{21}	0.0088	0.0011
ψ_{22}	-0.0066	0.0008
ψ_{23}	0.0058	0.0011
ψ_{31}	-0.0324	0.0038
ψ_{32}	0.0125	0.0025
ψ_{33}	0.0045	0.0038
ϕ_1	-0.0007	0.0003
ϕ_2	0.0004	0.0000
ϕ_3	0.0003	0.0003
φ_1	-0.0005	0.0001
φ_2	-0.0002	0.0001
φ_3	0.0003	0.0001
UNIT	-0.0102	0.0035
LIMITED	0.0003	0.0027
STATEWIDE	0.0071	0.0023

Est and Std stand for estimates and standard deviation, respectively.

Table 7: Average Returns to Scale

Year	RTS	std
1986	1.0526	0.0060
1987	1.0528	0.0059
1988	1.0458	0.0043
1989	1.0469	0.0038
1990	1.0503	0.0036
1991	1.0508	0.0038
1992	1.0531	0.0040
1993	1.0533	0.0040
1994	1.0559	0.0043
1995	1.0563	0.0042
1996	1.0616	0.0043
1997	1.0649	0.0044
1998	1.0564	0.0065
1999	1.0585	0.0064
2000	1.0590	0.0064
2001	1.0621	0.0064
2002	1.0644	0.0067
2003	1.0667	0.0067
2004	1.0682	0.0066
2005	1.0688	0.0066
Average	1.0576	0.0034

Table 8: Average Returns to Scale under Different Banking Regimes

	Unit		Limit		Statewide		Nationwide	
Year	RTS	std	RTS	std	RTS	std	RTS	std
1986	1.0995	0.0228	1.0407	0.0055	1.0361	0.0050	N/A	N/A
1987	1.0995	0.0226	1.0405	0.0055	1.0377	0.0050	N/A	N/A
1988	1.0962	0.0205	1.0410	0.0053	1.0413	0.0050	N/A	N/A
1989	1.0986	0.0198	1.0383	0.0052	1.0492	0.0050	N/A	N/A
1990	1.1022	0.0197	1.0405	0.0052	1.0522	0.0050	N/A	N/A
1991	1.0981	0.0218	1.0400	0.0053	1.0573	0.0052	N/A	N/A
1992	N/A	N/A	1.0395	0.0054	1.0594	0.0052	N/A	N/A
1993	N/A	N/A	1.0380	0.0055	1.0605	0.0053	N/A	N/A
1994	N/A	N/A	1.0332	0.0056	1.0621	0.0053	N/A	N/A
1995	N/A	N/A	1.0323	0.0054	1.0629	0.0052	N/A	N/A
1996	N/A	N/A	1.0365	0.0054	1.0685	0.0052	N/A	N/A
1997	N/A	N/A	1.0391	0.0054	1.0709	0.0052	N/A	N/A
1998	N/A	N/A	N/A	N/A	1.0818	0.0059	1.0550	0.0069
1999	N/A	N/A	N/A	N/A	1.0854	0.0059	1.0569	0.0068
2000	N/A	N/A	N/A	N/A	1.0872	0.0058	1.0577	0.0067
2001	N/A	N/A	N/A	N/A	1.0912	0.0058	1.0607	0.0067
2002	N/A	N/A	N/A	N/A	N/A	N/A	1.0644	0.0067
2003	N/A	N/A	N/A	N/A	N/A	N/A	1.0667	0.0067
2004	N/A	N/A	N/A	N/A	N/A	N/A	1.0682	0.0066
2005	N/A	N/A	N/A	N/A	N/A	N/A	1.0688	0.0066
Average	1.0995	0.0213	1.0390	0.0052	1.0605	0.0051	1.0625	0.0067

Table 9: Returns To Scale at Individual Bank Level

Year	DRS	CRS	IRS
1986	13.52%	11.59%	74.89%
1987	11.59%	12.23%	76.18%
1988	13.09%	7.94%	78.97%
1989	13.09%	5.36%	81.55%
1990	9.23%	3.86%	86.91%
1991	5.79%	3.65%	90.56%
1992	5.36%	3.00%	91.63%
1993	5.79%	2.58%	91.63%
1994	4.51%	2.15%	93.35%
1995	4.72%	1.50%	93.78%
1996	3.65%	1.50%	94.85%
1997	3.65%	0.43%	95.92%
1998	2.58%	2.58%	94.85%
1999	2.58%	2.36%	95.06%
2000	2.79%	1.50%	95.71%
2001	1.93%	2.58%	95.49%
2002	1.93%	1.29%	96.78%
2003	1.93%	0.86%	97.21%
2004	2.15%	0.86%	97.00%
2005	2.15%	1.29%	96.57%
Average	5.34%	3.36%	91.30%

Appendix A: Proofs of the main results

The proofs of Lemma A.2, Lemma 1, Lemma 2, Corollary 1 and Theorem 4 are provided in Appendix B of the supplementary document of this paper. In this file, we only provide the proofs of some main results. For notational simplicity, let $\hat{\beta}_{it} = \hat{\beta}_{-it}(Z_{it})$, $\beta_{it} = \beta(Z_{it})$ and $1_{js,it} = 1(Z_{js} = Z_{it})$. Recall that we have denoted $CV_0(\lambda)$, $\eta(z)$, $\Sigma_{XX}(z)$ and $\Sigma_{XX\beta}(z)$ in Assumption A.5. In the following proof, we will repeatly use these notations without mentioning the definitions of them again.

Lemma A.1. *For two square matrices A and B with the same dimensions, suppose that A is non-singular and $\left\|A^{-1}B\right\| < 1$. Then we have the following expansion:*

$$(A+B)^{-1} = A^{-1} - A^{-1}BA^{-1} + A^{-1}BA^{-1}BA^{-1} - A^{-1}BA^{-1}BA^{-1}BA^{-1} + \cdots$$

The proof of Lemma A.1 is straightforward and thus omitted.

Lemma A.2. *Under Assumption A, as $(N,T) \to (\infty,\infty)$ jointly*

1. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{u}_{it}^2 \to_P \sigma_u^2$;
2. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}' L(Z_{it}, z, \lambda) - \Sigma_{XX}(z) \to_P 0$;
3. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}' \beta(Z_{it}) L(Z_{it}, z, \lambda) - \Sigma_{XX\beta}(z) \to_P 0$;
4. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}' - E[\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})'] \to_P 0$;
5. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}' 1(Z_{it} = z) - p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)'\right) \to_P 0$;
6. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{X}_{it}' \beta(Z_{it}) - E[(\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})')\beta(Z_{it})] \to_P 0$;
7. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{u}_{it} 1(Z_{it} = z) = O_P\left(\frac{1}{\sqrt{NT}}\right)$;
8. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{u}_{it} L(Z_{it}, z, \lambda) = O_P\left(\frac{1}{\sqrt{NT}}\right)$;
9. $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}\tilde{u}_{it} = O_P\left(\frac{1}{\sqrt{NT}}\right)$.

Proof of Theorem 1:

In Lemma 1, we have shown $\hat{\lambda} = o_P(1)$, so it is reasonable to assume that λ in the proof of this theorem is sufficiently small and close to $0_{r\times 1}$. We now investigate the cross–validation criterion function and write

$$\begin{aligned} CV(\lambda) &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left(\tilde{X}_{it}'(\beta_{it} - \hat{\beta}_{it})\right)^2 + \frac{2}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{X}_{it}'(\beta_{it} - \hat{\beta}_{it})\tilde{u}_{it} + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{u}_{it}^2 \\ &\equiv CV_1(\lambda) + CV_2(\lambda) + CV_3. \end{aligned} \tag{A.1}$$

By (1) of Lemma A.2, $CV_3 = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \tilde{u}_{it}^2 \to_P \sigma_u^2$ and is independent of λ, so we focus on $CV_1(\lambda)$ and $CV_2(\lambda)$ below. To facilitate the analysis, we need to further consider $\hat{\beta}_{it} - \beta_{it}$. Notice that the kernel function of Aitchison and Aitken (1976) for the s^{th} covariate with $s = 1, \dots, r$ can be written as

$$l(Z_{it,s}, z_s, \lambda_s) = 1(Z_{it,s} = z_s) + \lambda_s 1(Z_{it,s} \neq z_s).$$

By Lemma 1, we can express the the kernel function (2.3) as

$$L(Z_{js}, Z_{it}, \lambda) = 1_{js,it} + \sum_{m=1}^{r} \lambda_m 1_{m,jsit} + O(\|\lambda\|^2), \tag{A.2}$$

where $1_{m,jsit} = 1(Z_{js,m} \neq Z_{it,m}) \prod_{n=1,n\neq m}^{r} 1(Z_{js,n} = Z_{it,n})$.

Let

$$\begin{aligned}\hat{\beta}_{it} - \beta_{it} &= \left(\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda)\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta_{js} - \beta_{it}) L(Z_{js}, Z_{it}, \lambda) \\ &\quad + \left(\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda)\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{u}_{js} L(Z_{js}, Z_{it}, \lambda), \end{aligned} \tag{A.3}$$

where we denote $\beta_{it} = \beta(Z_{it})$ for notational simplicity.

Below, we bring (A.2) to each term on RHS of (B.25) and investigate them respectively. Firstly,

$$\begin{aligned}&\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda) \\ &= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' 1_{js,it} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' \sum_{m=1}^{r} \lambda_m 1_{m,jsit} + O_P(\|\lambda\|^2) \\ &\equiv A_{1it} + A_{2it\lambda} + O_P(\|\lambda\|^2), \end{aligned} \tag{A.4}$$

where the first equality is due to result (4) of Lemma A.2.

Secondly,

$$\begin{aligned}&\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta_{js} - \beta_{it}) L(Z_{js}, Z_{it}, \lambda) \\ &= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta_{js} - \beta_{it}) 1_{js,it} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta_{js} - \beta_{it}) \sum_{m=1}^{r} \lambda_m 1_{m,jsit} + O_P(\|\lambda\|^2) \\ &\equiv 0 + B_{2it\lambda} + O_P(\|\lambda\|^2), \end{aligned} \tag{A.5}$$

where the first equality is due to (4) and (6) of Lemma A.2 and the uniform bound on $\beta(z)$; the term 0 of the last line is due to the construction of $(\beta_{js} - \beta_{it})1_{js,it} = 0$.

Thirdly,

$$\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{u}_{js} L(Z_{js}, Z_{it}, \lambda)$$

$$= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}1_{js,it} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}\sum_{m=1}^{r}\lambda_m 1_{m,jsit} + O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right)$$

$$\equiv C_{1it} + C_{2it\lambda} + O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right), \tag{A.6}$$

where the first equality is due to (9) of Lemma A.2.

For the terms on RHS of (A.4)–(A.6), by Lemma A.2, it is straightforward to obtain that

$$A_{1it} = O_P(1), \quad A_{2it\lambda} = O_P(\|\lambda\|), \quad B_{2it\lambda} = O_P(\|\lambda\|),$$
$$C_{1it} = O_P\left(\frac{1}{\sqrt{NT}}\right), \quad C_{2it\lambda} = O_P\left(\frac{\|\lambda\|}{\sqrt{NT}}\right). \tag{A.7}$$

By (A.4), using Lemma A.1 twice gives the following expression.

$$\left(\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}'_{js}L(Z_{js}, Z_{it}, \lambda)\right)^{-1} = \left(A_{1it} + A_{2it\lambda} + O_P(\|\lambda\|^2)\right)^{-1}$$

$$= (A_{1it} + A_{2it\lambda})^{-1} + O_P(\|\lambda\|^2) = A_{1it}^{-1} - A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1} + O_P(\|\lambda\|^2) \tag{A.8}$$

We then use (A.7) and (A.8) to further simplify (B.25) as follows.

$$\hat{\beta}_{it} = \beta_{it} + \left(A_{1it}^{-1} - A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}\right)(B_{2it\lambda} + C_{1it} + C_{2it\lambda}) + O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right) + O_P\left(\|\lambda\|^3\right) \tag{A.9}$$

We are now ready to further analyze $CV_1(\lambda)$ and $CV_2(\lambda)$ by using (A.7) and (A.9).

$$CV_1(\lambda) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}'_{it}(\beta_{it} - \hat{\beta}_{it})\right)^2$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\tilde{X}'_{it}\left(A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1} - A_{1it}^{-1}\right)(B_{2it\lambda} + C_{1it} + C_{2it\lambda})\right\}^2 + O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right) + O_P\left(\|\lambda\|^3\right)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(D_{3it}^2 - 2D_{1it}D_{2it} + 2D_{2it}D_{3it}\right)$$
$$+ O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right) + O_P\left(\|\lambda\|^3\right) + \text{ terms independent of } \lambda,$$

where $D_{1it} = \tilde{X}'_{it}A_{1it}^{-1}\left(A_{2it\lambda}A_{1it}^{-1}C_{1it} - C_{2it\lambda}\right)$, $D_{2it} = \tilde{X}'_{it}A_{1it}^{-1}C_{1it}$ and $D_{3it} = \tilde{X}'_{it}A_{1it}^{-1}B_{2it\lambda}$.

$$CV_2(\lambda) = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}'_{it}(\beta_{it} - \hat{\beta}_{it})$$

$$= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}'_{it}A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}(B_{2it\lambda} + C_{1it} + C_{2it\lambda})$$
$$- \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}'_{it}A_{1it}^{-1}(B_{2it\lambda} + C_{1it} + C_{2it\lambda}) + O_P\left(\frac{\|\lambda\|^2}{NT}\right) + O_P\left(\frac{\|\lambda\|^3}{\sqrt{NT}}\right)$$

$$= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}'_{it}A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}C_{1,it} - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}'_{it}A_{1it}^{-1}B_{2it\lambda} - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}'_{it}A_{1it}^{-1}C_{2it\lambda}$$

$$+O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right)+\text{ terms independent of }\lambda,$$

where the first equality follows from (9) of Lemma A.2 and (A.9); and the second equality follows from (9) of Lemma A.2 and (A.7).

Notice that

$$\begin{aligned}
&\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{2it}D_{3it}=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'A_{1it}^{-1}C_{1it}\tilde{X}_{it}'A_{1it}^{-1}B_{2it\lambda}\\
&=\frac{2}{N^3T^3}\sum_{m=1}^{r}\lambda_m\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{k=1}^{N}\sum_{r=1}^{T}\tilde{X}_{it}'A_{1it}^{-1}\tilde{X}_{js}\tilde{u}_{js}1_{js,it}\tilde{X}_{it}'A_{1it}^{-1}\tilde{X}_{kr}\tilde{X}_{kr}'(\beta_{kr}-\beta_{it})1_{m,krit}\\
&=\frac{2}{N^3T^3}\sum_{m=1}^{r}\lambda_m\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{k=1}^{N}\sum_{r=1}^{T}\tilde{X}_{kr}'A_{1kr}^{-1}\tilde{X}_{it}\tilde{u}_{it}1_{it,kr}\tilde{X}_{kr}'A_{1kr}^{-1}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{kr})1_{m,jskr}\\
&=\frac{2}{N^3T^3}\sum_{m=1}^{r}\lambda_m\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{k=1}^{N}\sum_{r=1}^{T}\tilde{X}_{it}'A_{1it}^{-1}\tilde{X}_{kr}\tilde{u}_{it}1_{it,kr}\tilde{X}_{kr}'A_{1it}^{-1}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{it})1_{m,jsit}\\
&=\frac{2}{N^2T^2}\sum_{m=1}^{r}\lambda_m\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{it}'A_{1it}^{-1}\tilde{u}_{it}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{it})1_{m,jsit}\\
&=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}B_{2it\lambda},
\end{aligned}$$

where the second equality follows from changing the index (it, js, kr) to (kr, it, js); the third equality follows from the definition of $1_{it,kr}$; the fourth equality follows from the definition of A_{1it}. Note that the term on RHS above can be canceled with the leading term of $CV_2(\lambda)$.

Thus, we are now able to further write

$$\begin{aligned}
CV(\lambda)=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(D_{3it}^2-2D_{1it}D_{2it}\right)+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}C_{1,it}\\
&-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}C_{2it\lambda}+O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right)+O_P\left(\|\lambda\|^3\right)\\
&+\text{ terms independent of }\lambda. \qquad (A.10)
\end{aligned}$$

Moreover, by (A.7) and some tedious algebra similar to the proof for $CV_1(\lambda)$ of Lemma 1 provided in the supplement of this paper

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{3it}=O_P\left(\|\lambda\|^2\right),\quad \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{1it}D_{2it}=O_P\left(\frac{\|\lambda\|}{NT}\right),\\
&\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}C_{1,it}=O_P\left(\frac{\|\lambda\|}{NT}\right),\\
&\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}C_{2it\lambda}=O_P\left(\frac{\|\lambda\|}{NT}\right).
\end{aligned}$$

Based on the above, (A.10) can be further simplified as follows.

$$CV(\lambda) = O_P\left(\frac{\|\lambda\|}{NT}\right) + O_P(\|\lambda\|^2) + \text{ terms independent of } \lambda, \tag{A.11}$$

which immediately implies that $\hat{\lambda} = O_P\left(\frac{1}{NT}\right)$. ■

Proof of Theorem 2:

Replacing $L(Z_{it}, z, \lambda)$ with (A.2) in (2.5) easily leads to that $\hat{\beta}(z) = \tilde{\beta}(z) + O_P\left(\frac{1}{NT}\right)$ by Lemma A.1 and Theorem 1, where $\tilde{\beta}(z)$ denotes a frequency estimator in the same manner as in $\hat{\beta}(z)$ but with $\lambda_s = 0$ for all $s = 1, \ldots, r$. Thus, it is straightforward to obtain that $\sqrt{NT}(\hat{\beta}(z) - \beta(z)) = \sqrt{NT}(\tilde{\beta}(z) - \beta(z)) + O_P\left(\frac{1}{\sqrt{NT}}\right)$. Below we just need to focus on $\sqrt{NT}(\tilde{\beta}(z) - \beta(z))$, so write

$$\begin{aligned}
&\sqrt{NT}(\tilde{\beta}(z) - \beta(z)) \\
&= \sqrt{NT}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}' 1(Z_{it} = z)\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\left(\tilde{X}_{it}'(\beta(Z_{it}) - \beta(z)) + \tilde{u}_{it}\right)1(Z_{it} = z) \\
&= \sqrt{NT}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}' 1(Z_{it} = z)\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}1(Z_{it} = z),
\end{aligned}$$

where the last line is due to the construction, i.e. $(\beta(Z_{it}) - \beta(z))1(Z_{it} = z) = 0$.

The result (5) of Lemma A.2 has shown that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}' 1(Z_{it} = z) \to_P p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)'\right) = \Xi_1(z).$$

Therefore, we need only to focus on $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}1(Z_{it} = z)$. In the proof for (7) of Lemma A.2, we have shown that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}1(Z_{it} = z) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \mu_X(z)\right)u_{it}1(Z_{it} = z) + o_P(1).$$

Thus, focus on $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}u_{it}1(Z_{it} = z)$ below. For notational simplicity, denote that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \mu_X(z)\right)u_{it}1(Z_{it} = z) = \sum_{t=1}^{T}V_{T,N}(t),$$

where $V_{T,N}(t) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left(X_{it} - \mu_X(z)\right)u_{it}1(Z_{it} = z)$. By the construction of $V_{T,N}(t)$ and Assumptions A.2–A.4, $V_{T,N}(t)$ is stationary and α–mixing. We can then apply the large–block and small–block technique to show the normality below (c.f. Theorem 2.21 in Fan and Yao (2003); Lemma A.1 in Gao (2007); Lemma A.1 in Chen et al. (2012*a*)). Partition the set $\{1, \ldots, T\}$ into $2k_T + 1$ subsets with a large block of size l_T, a small block of size s_T and the remaining set of size $T - k_T(l_T + s_T)$, where, for any $\lambda > 2$, $l_T = \lfloor T^{(\lambda-1)/\lambda}\rfloor$, $s_T = \lfloor T^{1/\lambda}\rfloor$ and $k_T = \lfloor T/(l_T + s_T)\rfloor$. Denote that for $n = 1, \ldots, k_T$

$$\tilde{V}_n = \sum_{t=(n-1)(l_T+s_T)+1}^{nl_T+(n-1)s_T} V_{T,N}(t), \quad \bar{V}_n = \sum_{t=nl_T+(n-1)s_T+1}^{n(l_T+s_T)} V_{T,N}(t) \text{ and } \hat{V} = \sum_{t=k_T(l_T+s_T)+1}^{T} V_{T,N}(t).$$

By the properties of α–mixing process and a procedure similar to A.6 and A.7 in Chen et al. (2012a), we obtain that $E\left\|\sum_{n=1}^{k_T}\bar{V}_n\right\|^2 = O\left(\frac{k_T s_T}{T}\right) = o(1)$ and $E\left\|\hat{V}\right\|^2 = O\left(\frac{T-k_T l_T}{T}\right) = o(1)$. Thus, we just need to focus on $\sum_{n=1}^{k_T}\tilde{V}_n$ below. In connection with Proposition 2.6 in Fan and Yao (2003) and the condition on the α–mixing coefficient, we have

$$\left|E\left[\exp\left\{\sum_{n=1}^{k_T}\|\tilde{V}_n\|\right\}\right] - \prod_{n=1}^{k_T} E\left[\exp\left\{\|\tilde{V}_n\|\right\}\right]\right| \leq C(k_T-1)\alpha(s_T) \to 0,$$

where C is a constant; $\alpha(\cdot)$ denotes the upper bounded of the α–mixing coefficients provided in Assumption A and is achievable in the same way as Assumption A.4 of Chen et al. (2012a). Then we obtain that $\tilde{V}_n$ for $n = 1,\ldots,k_T$ are asymptotically independent. Furthermore, as in the proof of Theorem 2.21.(ii) in Fan and Yao (2003), we have $\mathrm{Cov}\left[\tilde{V}_1\right] = \frac{l_T}{T}\Xi_0(z)(I_q + o(1))$, where

$$\Xi_0(z) = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[u_{it}u_{js}(X_{it}-\mu_X(z))(X_{js}-\mu_X(z))'1(Z_{it}=z)1(Z_{js}=z)\right].$$

It further implies that

$$\sum_{n=1}^{k_T}\mathrm{Cov}\left[\tilde{V}_n\right] = k_T\cdot\mathrm{Cov}\left[\tilde{V}_1\right] = \frac{k_T l_T}{T}\Xi_0\left(I_q + o\left(1\right)\right) \to \Xi_0.$$

Thus, the Feller condition is satisfied.

Moreover, by Cauchy–Schwarz inequality, we have

$$E\left[\left\|\tilde{V}_n\right\|^2\cdot I\left\{\|V_n\|\geq\varepsilon\right\}\right] \leq \left\{E\left\|\tilde{V}_n\right\|^3\right\}^{2/3}\cdot\left\{P\left(\left\|\tilde{V}_n\right\|\geq\varepsilon\right)\right\}^{1/3} \leq C\left\{E\left\|\tilde{V}_n\right\|^3\right\}^{2/3}\cdot\left\{E\left\|\tilde{V}_n\right\|^2\right\}^{1/3}$$

and by Lemma B.2 in Chen et al. (2012a)

$$E\left\|\tilde{V}_n\right\|^3 \leq \left(\frac{l_T}{T}\right)^{3/2}\left\{E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(X_{i1}-\mu_X(z)\right)u_{i1}1(Z_{i1}=z)\right\|^4\right\}^{3/4} < \infty.$$

Therefore, $E\left\|\tilde{V}_n\right\|^3 = O\left(\left(\frac{l_T}{T}\right)^{3/2}\right)$, which implies that

$$E\left[\left\|\tilde{V}_n\right\|^2\cdot I\left\{\|V_n\|\geq\varepsilon\right\}\right] \leq O\left(\left(\frac{l_T}{T}\right)^{4/3}\right) = o\left(\frac{l_T}{T}\right).$$

Consequently, $\sum_{n=1}^{k_T}E\left[\left\|\tilde{V}_n\right\|^2\cdot I\left\{\|V_n\|\geq\varepsilon\right\}\right] = o\left(\frac{k_T l_T}{T}\right) = o\left(1\right)$. Therefore, the Lindeberg condition is satisfied. Based on the above, $\sqrt{NT}(\tilde{\beta}(z)-\beta(z)) \to_D N(0,\Xi_1(z)^{-1}\Xi_0(z)\Xi_1(z)^{-1})$. We thus complete the proof. ■

Proof of Theorem 3:

(1) Notice that we have shown that $\hat{\lambda}_s = o_P(1)$ for $s = 1,\ldots,r_1$ in Lemma 2. Therefore, it is reasonable to assume that $\bar{\lambda}$ used in the proof of this theorem is sufficiently small and close to $0_{r_1\times 1}$.

For simplicity, denote that $\bar{1}_{itjs} = 1(\bar{Z}_{it} = \bar{Z}_{js})$ and $\bar{1}_{n,itjs} = 1(Z_{it,n} \neq Z_{js,n}) \prod_{m=1,m\neq n}^{r_1} 1(Z_{it,m} = Z_{js,m})$ for $n = 1, \ldots, r_1$. Let $\bar{L}_{jsit,\bar{\lambda}} = \bar{L}(\bar{Z}_{js}, \bar{Z}_{it}, \bar{\lambda})$ and $\tilde{L}_{jsit,\tilde{\lambda}} = \tilde{L}(\tilde{Z}_{js}, \tilde{Z}_{it}, \tilde{\lambda})$. By the definition of the kernel function of Aitchison and Aitken (1976) and the expansion same as (A.2), we can write

$$L(Z_{js}, Z_{it}, \lambda) = \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} = \left(\bar{1}_{jsit} + \sum_{n=1}^{r_1} \lambda_n \bar{1}_{n,jsit} + O(\|\bar{\lambda}\|^2) \right) \tilde{L}_{jsit,\tilde{\lambda}}. \tag{A.12}$$

Before investigate the cross–validation criterion function, we further simplify $\hat{\beta}_{it} - \bar{\beta}_{it}$. Write

$$\begin{aligned}
\hat{\beta}_{it} - \bar{\beta}_{it} &= \left(\frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \right)^{-1} \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' (\bar{\beta}_{js} - \bar{\beta}_{it}) \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \\
&\quad + \left(\frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \right)^{-1} \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{u}_{js} \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \\
&= \left(A_{1it} + A_{2it\lambda} + O_P(\|\bar{\lambda}\|^2) \right)^{-1} \left(B_{it} + C_{it} \right),
\end{aligned} \tag{A.13}$$

where the term $O_P(\|\bar{\lambda}\|^2)$ in the last line follows from (A.12) and (4) of Lemma A.2; and

$$\begin{aligned}
A_{1it} &= \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' \bar{1}_{jsit} \tilde{L}_{jsit,\tilde{\lambda}} \\
A_{2it\lambda} &= \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' \sum_{n=1}^{r_1} \lambda_n \bar{1}_{n,jsit} \tilde{L}_{jsit,\tilde{\lambda}} \\
B_{it} &= \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' (\bar{\beta}_{js} - \bar{\beta}_{it}) \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \\
C_{it} &= \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{u}_{js} \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}}.
\end{aligned}$$

For A_{1it} and $A_{2it\lambda}$, by the procedure similar to (2) of Lemma A.2, it is easy to know that $A_{1it} = O_P(1)$ and $A_{2it\lambda} = O_P(\|\bar{\lambda}\|)$. For B_{it}, by the procedure similar to (2)–(3) of Lemma A.2, we have

$$\begin{aligned}
B_{it} &= \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' (\bar{\beta}_{js} - \bar{\beta}_{it}) \left(\bar{1}_{jsit} + \sum_{n=1}^{r_1} \lambda_n \bar{1}_{n,jsit} + O(\|\bar{\lambda}\|^2) \right) \tilde{L}_{jsit,\tilde{\lambda}} \\
&= 0 + B_{2it\lambda} + O_P(\|\bar{\lambda}\|^2),
\end{aligned}$$

where the term 0 follows from the construction of $(\bar{\beta}_{js} - \bar{\beta}_{it}) \bar{1}_{jsit} = 0$ and

$$\begin{aligned}
B_{2it\lambda} &= \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' (\bar{\beta}_{js} - \bar{\beta}_{it}) \sum_{n=1}^{r_1} \lambda_n \bar{1}_{n,jsit} \tilde{L}_{jsit,\tilde{\lambda}} \\
&= \sum_{n=1}^{r_1} \lambda_n \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' (\bar{\beta}_{js} - \bar{\beta}_{it}) \bar{1}_{n,jsit} \tilde{L}_{jsit,\tilde{\lambda}} = O_P\left(\|\bar{\lambda}\|\right).
\end{aligned}$$

By the procedure same as (7) and (9) of Lemma A.2, for C_{it} we obtain that

$$C_{it} = \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{u}_{js} \left(\bar{1}_{jsit} + \sum_{n=1}^{r_1} \lambda_n \bar{1}_{n,jsit} + O(\|\bar{\lambda}\|^2) \right) \tilde{L}_{itjs,\tilde{\lambda}}$$

$$= C_{1it} + C_{2it\lambda} + O_P\left(\frac{\|\bar{\lambda}\|^2}{\sqrt{NT}}\right),$$

where

$$C_{1it} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}\bar{1}_{jsit}\tilde{L}_{jsit,\tilde{\lambda}} = O_P\left(\frac{1}{\sqrt{NT}}\right),$$
$$C_{2it\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}\sum_{n=1}^{r_1}\lambda_n\bar{1}_{n,jsit}\tilde{L}_{jsit,\tilde{\lambda}} = O_P\left(\frac{\|\bar{\lambda}\|}{\sqrt{NT}}\right).$$

Based on the above, applying Lemma A.1 twice to the term on RHS of (A.13) gives that

$$\hat{\beta}_{it} - \bar{\beta}_{it} = \left(A_{1it}^{-1} - A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}\right)\left(B_{2it\lambda} + C_{1it} + C_{2it\lambda}\right) + O_P\left(\frac{\|\bar{\lambda}\|^2}{\sqrt{NT}}\right) + O_P\left(\|\bar{\lambda}\|^3\right). \quad \text{(A.14)}$$

Recall that

$$CV(\lambda) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\beta_{it} - \hat{\beta}_{it})\right)^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\beta_{it} - \hat{\beta}_{it})\tilde{u}_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2$$
$$\equiv CV_1(\lambda) + CV_2(\lambda) + CV_3. \quad \text{(A.15)}$$

Then, replacing $\hat{\beta}_{it} - \bar{\beta}_{it}$ with (A.14) in $CV_1(\lambda)$ and $CV_2(\lambda)$ gives that

$$CV_1(\lambda) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\hat{\beta}_{it} - \bar{\beta}_{it})\right)^2$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\tilde{X}_{it}'\left(A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1} - A_{1it}^{-1}\right)\left(B_{2it\lambda} + C_{1it} + C_{2it\lambda}\right)\right\}^2 + O_P\left(\frac{\|\bar{\lambda}\|^2}{\sqrt{NT}}\right) + O_P\left(\|\bar{\lambda}\|^3\right)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(D_{3it}^2 - 2D_{1it}D_{2it} + 2D_{2it}D_{3it}\right)$$
$$+ O_P\left(\frac{\|\bar{\lambda}\|^2}{\sqrt{NT}}\right) + O_P\left(\|\bar{\lambda}\|^3\right) + \text{ terms independent of } \lambda,$$

where $D_{1it} = \tilde{X}_{it}'A_{1it}^{-1}\left(A_{2it\lambda}A_{1it}^{-1}C_{1it} - C_{2it\lambda}\right)$, $D_{2it} = \tilde{X}_{it}'A_{1it}^{-1}C_{1it}$ and $D_{3it} = \tilde{X}_{it}'A_{1it}^{-1}B_{2it\lambda}$.

$$CV_2(\lambda) = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'(\bar{\beta}_{it} - \hat{\beta}_{it})$$
$$= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}A_{2it,\lambda}A_{1it}^{-1}\left(B_{2it,\lambda} + C_{1it} + C_{2it,\lambda}\right)$$
$$-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}\left(B_{2it,\lambda} + C_{1it} + C_{2it,\lambda}\right) + O_P\left(\frac{\|\bar{\lambda}\|^2}{NT}\right) + O_P\left(\frac{\|\bar{\lambda}\|^3}{\sqrt{NT}}\right).$$

Then it is easy to know that the leading term of $CV_2(\lambda)$ is

$$-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}X_{it}'A_{1it}^{-1}B_{2it,\lambda} = O_P\left(\frac{\|\bar{\lambda}\|}{\sqrt{NT}}\right).$$

For $CV_1(\lambda)$, the leading terms are

$$\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{2it}D_{3it}=O_P\left(\frac{\|\bar{\lambda}\|}{\sqrt{NT}}\right) \quad \text{and} \quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{3it}^2=O_P\left(\|\bar{\lambda}\|^2\right).$$

Notice that the two leading terms $\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{2it}D_{3it}$ and $-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}B_{2it,\lambda}$ cannot cancel each other as the proof of Theorem 1 when the irrelevant covariates exist. Thus, the first result of this theorem follows.

(2) We now investigate the asymptotic behaviour of $\hat{\lambda}_s$ for $s=r_1+1,\ldots,r$. Based on the first result of this theorem, we know that

$$\begin{aligned}
CV_1(\lambda) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(D_{1it}-D_{2it}-D_{3it})^2+O_P\left(\frac{\|\bar{\lambda}\|^2}{\sqrt{NT}}\right)+O_P\left(\|\bar{\lambda}\|^3\right)\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(D_{1it}-D_{2it}-D_{3it})^2+o_P\left(\frac{1}{NT}\right).
\end{aligned}$$

For simplicity, let $\Psi(\bar{Z}_{it})=\bar{p}(\bar{Z}_{it})(\Sigma_X(\bar{Z}_{it})-\mu_X(\bar{Z}_{it})\mu_X(\bar{Z}_{it})')$ in the following analysis. Then we consider $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{3it}^2$ at first.

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{3it}^2=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'A_{1it}^{-1}B_{2it\lambda}\right)^2\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-1}B_{2it\lambda}\right)^2+o_P(\|\bar{\lambda}\|^2)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-1}\right.\\
&\left.\cdot\sum_{n=1}^{r_1}\lambda_nE[X_{js}X_{js}'(\beta(\bar{Z}_{js})-\beta(\bar{Z}_{it}))\bar{1}_{n,jsit}|\bar{Z}_{it}]\cdot E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]\right)^2+o_P(\|\bar{\lambda}\|^2)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})\cdot\sum_{n=1}^{r_1}\lambda_nE[X_{js}X_{js}'(\beta(\bar{Z}_{js})-\beta(\bar{Z}_{it}))\bar{1}_{n,jsit}|\bar{Z}_{it}]\right)^2+o_P(\|\bar{\lambda}\|^2)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})\cdot\sum_{n=1}^{r_1}\lambda_nE[X_{js}X_{js}'(\beta(\bar{Z}_{js})-\beta(\bar{Z}_{it}))\bar{1}_{n,jsit}|\bar{Z}_{it}]\right)^2+o_P\left(\frac{1}{NT}\right)
\end{aligned}$$

where the second equality follows from (2) of Lemma A.2, Assumption B and $B_{2it\lambda}=O_P(\|\bar{\lambda}\|^2)$; the third equality follows from the procedure similar to (3) of Lemma A.2 and Assumption B; the fifth equality follows from the first result of this theorem. Notice that $E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]$ gets canceled above. Therefore, the leading term on RHS above is unrelated with $\tilde{\lambda}$ and the remainder terms have order $o_P\left(\frac{1}{NT}\right)$. Also we know that $D_{1it}^2=o_P\left(\frac{1}{NT}\right)$, $D_{1it}D_{2it}=o_P\left(\frac{1}{NT}\right)$ and $D_{1it}D_{3it}=o_P\left(\frac{1}{NT}\right)$ due to the first result of this theorem. Then we can further write

$$CV_1(\lambda)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(2D_{2it}D_{3it}+D_{2it}^2\right)+o_P\left(\frac{1}{NT}\right)+\text{ terms unrelated to }\tilde{\lambda}.$$

Notice that both of $2D_{2it}D_{3it}$ and D_{2it}^2 have order $O_P\left(\frac{1}{NT}\right)$.

We now further investigate the leading terms of $CV_1(\lambda)$.

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{2it}^2=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'A_{1it}^{-1}C_{1it}C_{1it}'A_{1it}^{-1}\tilde{X}_{it}\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})C_{1it}C_{1it}'\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{it}E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-2}+o_P\left(\frac{1}{NT}\right)\\
&=\frac{1}{N^3T^3}\sum_{i,t}\sum_{j,s}\sum_{k,r}\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{js}\tilde{u}_{js}\bar{1}_{jsit}\tilde{L}_{jsit,\tilde{\lambda}}\tilde{X}_{kr}'\tilde{u}_{kr}\bar{1}_{krit}\tilde{L}_{krit,\tilde{\lambda}}\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{it}E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-2}\\
&\quad+o_P\left(\frac{1}{NT}\right)\\
&=o_P\left(\frac{1}{NT}\right)+\frac{1}{N^3T^3}\sum_{i,t}\sum_{j,s}\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{js}\tilde{X}_{js}'\tilde{u}_{js}^2\bar{1}_{jsit}\tilde{L}_{jsit,\tilde{\lambda}}^2\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{it}E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-2}\\
&\quad+\frac{1}{N^3T^3}\sum_{i,t}\sum_{j,s}\sum_{k,r\neq j,s}\tilde{X}_{it}'\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{js}\tilde{u}_{js}\bar{1}_{jsit}\tilde{L}_{jsit,\tilde{\lambda}}\tilde{X}_{kr}'\tilde{u}_{kr}\bar{1}_{krit}\tilde{L}_{krit,\tilde{\lambda}}\Psi^{-1}(\bar{Z}_{it})\tilde{X}_{it}E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-2}\\
&\equiv H_{1,NT}+H_{2,NT}+o_P\left(\frac{1}{NT}\right),
\end{aligned}
$$

where the second equality follows from (2) of Lemma A.2, Assumption B and $C_{1it}=O_P\left(\frac{1}{\sqrt{NT}}\right)$.

By going through the procedure similar to the proof for $CV_1(\lambda)$ in Lemma 1, we can obtain that

$$
H_{1,NT}=\frac{1}{NT}C\cdot E\left[E[\tilde{L}_{jsit,\tilde{\lambda}}^2|\tilde{Z}_{it}]\cdot E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-2}\right]+o_P\left(\frac{1}{NT}\right), \tag{A.16}
$$

where by the construction of $H_{1,NT}$ it is easy to know that C is a positive constant. Notice that $E[\tilde{L}_{jsit,\tilde{\lambda}}^2|\tilde{Z}_{it}]\geq E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^2$ and equality holds if and only if $\lambda_s=1$ for all $s=r_1+1,\ldots,r$. Hence, $H_{1,NT}$ is minimized at the upper bound values for $\lambda_s=1$ for all $s=r_1+1,\ldots,r$.

For the term $\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{2it}D_{3it}=O_P\left(\frac{1}{NT}\right)$, denote $H_{3,NT}=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{2it}D_{3it}$.

For the term CV_2, by the first result of this theorem we further write

$$
\begin{aligned}
CV_2(\lambda)&=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}A_{2it,\lambda}A_{1it}^{-1}\left(B_{2it,\lambda}+C_{1it}+C_{2it,\lambda}\right)\\
&\quad-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}\left(B_{2it,\lambda}+C_{1it}+C_{2it,\lambda}\right)+O_P\left(\frac{\|\bar{\lambda}\|^2}{NT}\right)+O_P\left(\frac{\|\bar{\lambda}\|^3}{\sqrt{NT}}\right)\\
&=-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}B_{2it,\lambda}-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}C_{1it}+o_P\left(\frac{1}{NT}\right)\\
&=H_{4,NT}+H_{5,NT}+o_P\left(\frac{1}{NT}\right).
\end{aligned}
$$

Therefore,

$$
CV(\lambda)=H_{1,NT}+H_{2,NT}+H_{3,NT}+H_{4,NT}+H_{5,NT}+o_P\left(\frac{1}{NT}\right), \tag{A.17}
$$

where $H_{1,NT}$ to $H_{5,NT}$ all include $\tilde{\lambda}$. Moreover, based on the first result of this theorem, it is easy to know that $H_{1,NT}$ to $H_{5,NT}$ all have order $O_P\left(\frac{1}{NT}\right)$ and $H_{1,NT}$ is minimized at $\lambda_s=1$ for all $s=r_1+1,\ldots,r$. By the similar argument on page 578 of Li et al. (2013), the second result of this theorem holds. ■

References

Aitchison, J. and Aitken, C. (1976), 'Multivariate binary discrimination by the kernel method', *Biometrika* pp. 413–420.

Andrews, D. W. K. (1984), 'Nonstrong mixing autoregressive processes', *Journal of Applied Probability* pp. 930–934.

Andrews, D. W. K. (2005), 'Cross–section regression with common shocks', *Econometrica* pp. 1551–1585.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* pp. 1229–1279.

Berger, A. and Mester, L. (2003), 'Explaining the dramatic changes in performance of u.s. banks: technical change, deregulation, and dynamic changes in competition', *Journal of Financial Intermediation* pp. 57–95.

Blundell, R. and Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics* pp. 115–143.

Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Springer, New York.

Cai, Z. (2007), 'Trending time-varying coefficient time series models with serially correlated errors', *Journal of Econometrics* pp. 163–188.

Cai, Z. and Li, Q. (2008), 'Nonparametric estimation of varying coefficient dynamic panel data models', *Econometric Theory* pp. 1321–1342.

Cai, Z., Li, Q. and Park, J. (2009), 'Functional–coefficient models for nonstationary time series data', *Journal of Econometrics* pp. 101–113.

Chen, J., Gao, J. and Li, D. (2012*a*), 'A new diagnostic test for cross–section uncorrelatedness in nonparametric panel data models', *Econometric Theory* pp. 1–20.

Chen, J., Gao, J. and Li, D. (2012*b*), 'Semiparametric trending panel data models with cross–sectional dependence', *Journal of Econometrics* pp. 71–85.

Chen, J., Gao, J. and Li, D. (2013), 'Estimation in partially linear single–index panel data models with fixed effects', *Journal of Business and Economic Statistics* pp. 315–330.

Fan, J. and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer-Verlag.

Feng, G. and Serletis, A. (2008), 'Productivity trends in u.s. manufacturing: Evidence from the nq and aim cost functions', *Journal of Econometrics* pp. 281–311.

Feng, G. and Serletis, A. (2010), 'Efficiency, technical change, and returns to scale in large u.s. banks: Panel data evidence from an output distance function satisfying theoretical regularity', *Journal of Banking and Finance* pp. 127–138.

Feng, G. and Zhang, X. (2012), 'Productivity and efficiency at large and community banks in the u.s.: a bayesian true random effects stochastic distance frontier analysis', *Journal of Banking and Finance* pp. 1883–1895.

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman and Hall.

Gao, J. and Phillips, P. C. B. (2013*a*), 'Functional coefficient nonstationary regression', *Cowles Foundation Discussion Paper NO. 1911* .

Gao, J. and Phillips, P. C. B. (2013*b*), 'Semiparametric estimation in triangular system equations with nonstationarity', *Journal of Econometrics* pp. 59–79.

Griffiths, W. E., O'Donnell, C. J. and Cruz, A. T. (2000), 'Imposing regularity conditions on a system of cost and cost–share equations: A bayesian approach', *Australian Journal of Agricultural and Resource Economics* pp. 107–127.

Jayaratne, J. and Strahan, P. E. (1997), 'The benefits of branching deregulation', *Economic Policy Review* pp. 13–29.

Li, Q., Huang, C. J., Li, D. and F, T. (2002), 'Semiparametric smooth coefficient models', *Journal of Business and Economic Statistics* pp. 412–422.

Li, Q., Ouyang, D. and Racine, J. S. (2013), 'Categorical semiparametrics varying–coefficient models', *Journal of Applied Econometrics* pp. 551–579.

Li, Q. and Racine, J. S. (2010), 'Smooth varying–coefficient estimation and inference for qualitative and quantitative data', *Econometric Theory* pp. 1607–1637.

Mason, J. E. (2013), *The Transformation of Commercial Banking in the United States 1956–1991*, United Kingdom: Taylor and Francis.

Mester, L. J. (2005), 'Optimal industrial structure in banking', *Working Papers 08-2, Federal Reserve Bank of Philadelphia* .

Newey, W. K. and McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of Econometrics* pp. 2113–2245.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* pp. 967–1012.

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

Rodriguez-Poo, J. M. and Soberon, A. (2014), 'Direct semi–parametric estimation of fixed effects panel data varying coefficient models', *Econometrics Journal* pp. 107–138.

Sealey, C. W. and Lindley, J. T. (1977), 'Inputs, outpus, and a theory of production and cost at depository financial institutions', *Journal of Finance* pp. 1251–1266.

Su, L. and Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* pp. 34–47.

Su, L. and Ullah, A. (2011), *Nonparametric and semiparametric panel econometric models: estimation and testing*, Handbook of Empirical Economics and Finance, New York: Taylor and Francis Group.

Sun, Y., Carroll, R. J. and Li, D. (2009), 'Semiparametric estimation of fixed effects panel data varying coefficient models', *Advances in Econometrics* pp. 101–130.

Watkins, T. G. and West, R. C. (1982), 'Bank holding companies: development and regulation', *Economic Review* pp. 3–13.

Wheelock, D. C. and Wilson, P. W. (2012), 'Do large banks have lower costs? new estimates of returns to scale for u.s. banks', *Journal of Money, Credit and Banking* pp. 171–199.

Appendix B: Additional proofs

For notational simplicity, let $\hat{\beta}_{it} = \hat{\beta}_{-it}(Z_{it})$, $\beta_{it} = \beta(Z_{it})$ and $1_{js,it} = 1(Z_{js} = Z_{it})$. Recall that we have denoted $CV_0(\lambda)$, $\eta(z)$, $\Sigma_{XX}(z)$ and $\Sigma_{XX\beta}(z)$ in Assumption A.5. In the following proofs, we will keep using these notations without mentioning the definitions of them again.

By Assumption A.1, each element of Z_{it} takes finite different integer values and the dimension of Z_{it} is finite. Therefore, when $Z_{it} \neq Z_{is}$, $\|Z_{it} - Z_{is}\| \geq 1$. It is easy to know that the following uniform convergence holds:

$$\sup_{z\in\mathcal{D}} \left| \frac{1}{T} \sum_{s=1}^{T} 1(Z_{is} = z) - p(z) \right| \to_P 0, \tag{B.1}$$

$$\sup_{z\in\mathcal{D}} \left| \frac{1}{T} \sum_{s=1}^{T} X_{is} 1(Z_{is} = z) - p(z)\mu_X(z) \right| \to_P 0 \tag{B.2}$$

where $p(z) = \Pr(Z_{is} = z) = E[1(Z_{is} = z)]$, $\mu_X(z) = E[X_{it}|Z_{it} = z]$ and $\mathcal{D}$ is a compact set by definition. In the following proofs, let $\mathcal{Z} = \{Z_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$.

Lemma A.2. *Under Assumption A, as $(N,T) \to (\infty,\infty)$ jointly*

1. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{u}_{it}^2 \to_P \sigma_u^2$;
2. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' L(Z_{it}, z, \lambda) - \Sigma_{XX}(z) \to_P 0$;
3. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' \beta(Z_{it}) L(Z_{it}, z, \lambda) - \Sigma_{XX\beta}(z) \to_P 0$;
4. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' - E[\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})'] \to_P 0$;
5. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' 1(Z_{it} = z) - p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)'\right) \to_P 0$;
6. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}' \beta(Z_{it}) - E[(\Sigma_X(Z_{it}) - \mu_X(Z_{it})\mu_X(Z_{it})')\beta(Z_{it})] \to_P 0$;
7. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it} 1(Z_{it} = z) = O_P\left(\frac{1}{\sqrt{NT}}\right)$;
8. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it} L(Z_{it}, z, \lambda) = O_P\left(\frac{1}{\sqrt{NT}}\right)$;
9. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{u}_{it} = O_P\left(\frac{1}{\sqrt{NT}}\right)$.

Proof of Lemma A.2:

1). Expand $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2$ as follows:

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)\\ =&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s_1=1}^{T}u_{is_1}1_{is_1,it}\frac{1}{T_{it}}\sum_{s_2=1}^{T}u_{is_2}1_{is_2,it}\\ &-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}u_{it}u_{is}1_{is,it}\end{aligned}\tag{B.3}$$

For the first term on RHS of (B.3), write

$$\begin{aligned}&E\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2-\sigma_u^2\right|^2\\ \leq&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{\delta_2}\left(\alpha_{u,ij}(|t-s|)\right)^{\delta_2/(4+\delta_2)}\left(E[u_{it}^2]^{2+\delta_2/2}\cdot E[u_{js}^2]^{2+\delta_2/2}\right)^{2/(4+\delta_2)}\\ \leq&O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\alpha_{u,ij}(|t-s|)\right)^{\delta_2/(4+\delta_2)}=O\left(\frac{1}{NT}\right),\end{aligned}$$

where $c_{\delta_2}=2^{(4+2\delta_2)/(4+\delta_2)}\cdot(4+\delta_2)/\delta_2$, the first inequality is due to the Davydov inequality (c.f. pages 19–20 in Bosq (1996) and the supplement of Su and Jin (2012)), and the last line follows from Assumption A.4.

For the second term on RHS of (B.3), let $U_{it}=\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}$. Below we, respectively, evaluate

$$E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}^2\Big|\mathcal{Z}\right]\quad\text{and}\quad E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}^2\right)^2\Big|\mathcal{Z}\right].$$

Write

$$\begin{aligned}&E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}U_{it}^2\Big|\mathcal{Z}\right]\\ =&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s=1}^{T}1_{is,it}E\left[u_{is}^2|\mathcal{Z}\right]+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=2}^{T}\sum_{s_2=1}^{s_1-1}1_{is_1,it}1_{is_2,it}E[u_{is_1}u_{is_2}|\mathcal{Z}]\\ =&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s=1}^{T}1_{is,it}E\left[u_{is}^2\right]+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=2}^{T}\sum_{s_2=1}^{s_1-1}1_{is_1,it}1_{is_2,it}E[u_{is_1}u_{is_2}]\\ =&A_{1,NT}+2A_{2,NT}.\end{aligned}$$

For $A_{1,NT}$, write

$$\begin{aligned}A_{1,NT}&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s=1}^{T}1_{is,it}E\left[u_{is}^2\right]=\frac{\sigma_u^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\\ &=\frac{\sigma_u^2}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{T_{it}}{T}\right)^{-1}=O_P\left(\frac{1}{T}\right),\end{aligned}$$

where the last equality follows from (B.1).

For $A_{2,NT}$, write

$$
\begin{aligned}
|A_{2,NT}| &= \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{T_{it}^2} \sum_{s_1=2}^{T} \sum_{s_2=1}^{s_1-1} 1_{is_1,it} 1_{is_2,it} E[u_{is_1} u_{is_2}] \right| \\
&= \left| \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{T^2}{T_{it}^2} \sum_{s_1=2}^{T} \sum_{s_2=1}^{s_1-1} 1_{is_1,it} 1_{is_2,it} E[u_{is_1} u_{is_2}] \right| \\
&\leq \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{T^2}{T_{it}^2} \sum_{s_1=2}^{T} \sum_{s_2=1}^{s_1-1} 1_{is_1,it} 1_{is_2,it} \left| E[u_{is_1} u_{is_2}] \right| \\
&\leq O_P(1) \frac{1}{NT^3} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s_1=2}^{T} \sum_{s_2=1}^{s_1-1} \left| E[u_{is_1} u_{is_2}] \right| \\
&= O_P(1) \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{s_1=2}^{T} \sum_{s_2=1}^{s_1-1} \left| E[u_{is_1} u_{is_2}] \right| = O_P(1) \frac{1}{T^2} \sum_{s_1=2}^{T} \sum_{s_2=1}^{s_1-1} \left| E[u_{is_1} u_{is_2}] \right| = O_P\left(\frac{1}{T}\right),
\end{aligned}
$$

where the second inequality follows from (B.1), and the last line follows from Assumption A.4.

We then consider the second moments of $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}^2$ conditional on $\mathcal{Z}$.

$$
\begin{aligned}
& E\left[\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}^2 \right)^2 \middle| \mathcal{Z} \right] = \frac{1}{N^2T^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} E\left[U_{i_1t_1}^2 U_{i_2t_2}^2 | \mathcal{Z} \right] \\
&= \frac{1}{N^2T^2} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \frac{1}{T_{i_1t_1}^2} \frac{1}{T_{i_2t_2}^2} \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} \sum_{s_3=1}^{T} \sum_{s_4=1}^{T} E\left[u_{i_1s_1} u_{i_1s_2} u_{i_2s_3} u_{i_2s_4} | \mathcal{Z} \right] \\
&= \frac{1}{N^2T^6} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \frac{T^2}{T_{i_1t_1}^2} \frac{T^2}{T_{i_2t_2}^2} \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} \sum_{s_3=1}^{T} \sum_{s_4=1}^{T} E\left[u_{i_1s_1} u_{i_1s_2} u_{i_2s_3} u_{i_2s_4} \right] \\
&\leq O_P(1) \frac{1}{N^2T^4} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \sum_{s_1=1}^{T} \sum_{s_2=1}^{T} \sum_{s_3=1}^{T} \sum_{s_4=1}^{T} \left| E\left[u_{i_1s_1} u_{i_1s_2} u_{i_2s_3} u_{i_2s_4} \right] \right| = O_P\left(\frac{1}{NT^2} \right),
\end{aligned}
$$

where the first inequality follows from (B.1), and the last equality follows from Assumption A.4.

Based on the above, we have shown that conditional on $\mathcal{Z}$

$$
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}^2 \to_P 0, \tag{B.4}
$$

which implies that $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}^2 \to_P 0$ unconditionally.

For the third term on right on of (B.3), write

$$
\begin{aligned}
E\left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it} U_{it} | \mathcal{Z} \right] &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1_{is,it}}{T_{it}} E\left[u_{it} u_{is} | \mathcal{Z} \right] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{1_{is,it}}{T_{it}} E\left[u_{it} u_{is} \right] \\
&\leq \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \frac{T}{T_{it}} \left| E\left[u_{it} u_{is} \right] \right| \leq O_P(1) \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| E\left[u_{it} u_{is} \right] \right| \\
&\leq O_P(1) \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| E\left[u_{it} u_{is} \right] \right| = O_P\left(\frac{1}{T} \right),
\end{aligned}
$$

where the second inequality follows from (B.1), and the last line follows from Assumption A.4.

We then consider the second moments of $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}u_{it}u_{is}1_{is,it}$ conditional on $\mathcal{Z}$:

$$E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}U_{it}\right)^2\Big|\mathcal{Z}\right] = \frac{1}{N^2T^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}E\left[u_{i_1t_1}U_{i_1t_1}u_{i_2t_2}U_{i_2t_2}|\mathcal{Z}\right]$$

$$= \frac{1}{N^2T^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\frac{1}{T_{i_1t_1}}\frac{1}{T_{i_2t_2}}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}E\left[u_{i_1t_1}u_{i_1s_1}u_{i_2t_2}u_{i_2s_2}|\mathcal{Z}\right]$$

$$= \frac{1}{N^2T^4}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\frac{T}{T_{i_1t_1}}\frac{T}{T_{i_2t_2}}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}E\left[u_{i_1t_1}u_{i_1s_1}u_{i_2t_2}u_{i_2s_2}\right]$$

$$\leq O_P(1)\frac{1}{N^2T^4}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\left|E\left[u_{i_1t_1}u_{i_1s_1}u_{i_2t_2}u_{i_2s_2}\right]\right| = O_P\left(\frac{1}{NT^2}\right),$$

where the first inequality follows from (B.1), and the last equality follows from Assumption A.4.

Based on the above, we have shown that conditional on $\mathcal{Z}$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}U_{it}\to_P 0, \tag{B.5}$$

which implies that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}U_{it}\to_P 0$ unconditionally. The proof is now completed. ■

2). Write

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'L(Z_{it},z,\lambda)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)'L(Z_{it},z,\lambda)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'L(Z_{it},z,\lambda)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda)$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}X_{it}'L(Z_{it},z,\lambda)$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}'1_{is,it}L(Z_{it},z,\lambda). \tag{B.6}$$

We now consider each term on RHS of (B.6) respectively. We consider the first term on RHS of (B.6) as follows.

$$E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'L(Z_{it},z,\lambda)-E[\Sigma_X(Z_{it})L(Z_{it},z,\lambda)]\right\|^2\right]$$

$$= \frac{1}{N^2T^2}\sum_{m=1}^{q}\sum_{n=1}^{q}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\Big[\left(X_{it,m}X_{it,n}L(Z_{it},z,\lambda)-E[\Sigma_{X,mn}(Z_{lk})L(Z_{lk},z,\lambda)]\right)$$

$$\cdot\left(X_{is,m}X_{is,n}L(Z_{is},z,\lambda)-E[\Sigma_{X,mn}(Z_{lk})L(Z_{lk},z,\lambda)]\right)\Big]$$

$$\leq \frac{1}{N^2T^2}\sum_{m=1}^{q}\sum_{n=1}^{q}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{E\left[|X_{it,m}X_{it,n}L(Z_{it},z,\lambda)|^2\right]E\left[|X_{is,m}X_{is,n}L(Z_{it},z,\lambda)|^2\right]\right\}^{1/2}$$
$$\leq O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{E\left[\|X_{it}\|^4\right]E\left[\|X_{is}\|^4\right]\right\}^{1/2} = O\left(\frac{1}{N}\right), \tag{B.7}$$

where $X_{it,m}$ denotes the m^{th} element of X_{it} for $m = 1,\ldots,q$, $\Sigma_{X,mn}(z)$ denotes the $(m,n)^{th}$ element of $\Sigma_X(z)$ for $m,n = 1,\ldots,q$; the first inequality follows from Cauchy–Schwarz inequality, and the second inequality follows from $L(Z_{it},z,\lambda)$ being bounded uniformly. It thus implies that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'L(Z_{it},z,\lambda) - E[\Sigma_X(Z_{it})L(Z_{it},z,\lambda)] \to_P 0.$$

For the second term on RHS of (B.6), we write

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda)$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda). \tag{B.8}$$

Notice that

$$E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda)\right\|^2 \Big|\mathcal{Z}\right]$$
$$\leq E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{T}{T_{it}}\frac{1}{T}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right\|\left\|\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}1_{is_2,it}L(Z_{it},z,\lambda)\right\|\right)^2 \Big|\mathcal{Z}\right]$$
$$= E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{p(Z_{it})}p(Z_{it})\mu(Z_{it}) + o_P(1) - \mu_X(Z_{it})\right\|\left\|\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}1_{is_2,it}L(Z_{it},z,\lambda)\right\|\right)^2 \Big|\mathcal{Z}\right]$$
$$\leq o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{T}\frac{T}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}1_{is_2,it}L(Z_{it},z,\lambda)\right\|\right)^2 \Big|\mathcal{Z}\right]$$
$$\leq o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T}\sum_{s_2=1}^{T}\|X_{is_2}\|\right)^2 \Big|\mathcal{Z}\right]$$
$$= o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s_2=1}^{T}\|X_{is_2}\|\right)^2 \Big|\mathcal{Z}\right] = o_P(1), \tag{B.9}$$

where first equality follows from (B.1) and (B.2), the third inequality follows from (B.1) and the fact that $1_{is_2,it}$ and $L(Z_{it},z,\lambda)$ being bounded uniformly, and the last equality follows from Assumption A.3. Thus, we obtain that

$$E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda)\right\|^2\right] = o_P(1),$$

which allows us to write

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}L(Z_{it},z,\lambda)+o_P(1) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\left(\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}-\mu_X(Z_{it})'\right)L(Z_{it},z,\lambda) \\
& +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\mu_X(Z_{it})'L(Z_{it},z,\lambda)+o_P(1) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\mu_X(Z_{it})'L(Z_{it},z,\lambda)+o_P(1) \\
\to_P & \; E[\mu_X(Z_{it})\mu_X(Z_{it})'L(Z_{it},z,\lambda)],
\end{aligned}
\tag{B.10}
$$

where the first equality follows from (B.8) and (B.9), the third equality follows from the procedure similar to (B.9), and the last line follows from the same procedure as (B.7).

For the last two terms on RHS of (B.6), write

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}X_{it}'L(Z_{it},z,\lambda)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})X_{it}'L(Z_{it},z,\lambda)+o_P\left(1\right) \\
& \to_P E[\mu_X(Z_{it})X_{it}'L(Z_{it},z,\lambda)]=E[\mu_X(Z_{it})\mu_X(Z_{it})'L(Z_{it},z,\lambda)],
\end{aligned}
$$

where the first equality follows from the same procedure as (B.8) and (B.9), and the last line follows from the procedure similar to (B.7). Based on the above, the result follows. ■

3). Write

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\beta(Z_{it})L(Z_{it},z,\lambda) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)'\beta(Z_{it})L(Z_{it},z,\lambda) \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda) \\
+ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda) \\
& -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda) \\
& -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}'1_{is,it}\beta(Z_{it})L(Z_{it},z,\lambda).
\end{aligned}
\tag{B.11}
$$

We now consider each term on RHS of (B.11) respectively. We consider the first term on RHS of

(B.11) as follows:

$$\begin{aligned}
&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda)-E[\Sigma_X(Z_{it})\beta(Z_{it})L(Z_{it},z,\lambda)]\right\|^2\right]\\
&=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\Big\{\left(X_{it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda)-E[\Sigma_X(Z_{it})\beta(Z_{it})L(Z_{it},z,\lambda)]\right)'\\
&\quad\cdot\left(X_{is}X_{is}'\beta(Z_{is})L(Z_{is},z,\lambda)-E[\Sigma_X(Z_{is})\beta(Z_{is})L(Z_{is},z,\lambda)|z]\right)\Big\}\\
&\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{E\left[\left\|X_{it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda)-E[\Sigma_X(Z_{it})\beta(Z_{it})L(Z_{it},z,\lambda)]\right]\right\|^2\\
&\quad\cdot E\left[\left\|X_{is}X_{is}'\beta(Z_{is})L(Z_{is},z,\lambda)-E[\Sigma_X(Z_{is})\beta(Z_{is})L(Z_{is},z,\lambda)]\right\|^2\right]\Big\}^{1/2}\\
&\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{E\left[\left\|X_{it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda)\right\|^2\right]\cdot E\left[\left\|X_{is}X_{is}'\beta(Z_{is})L(Z_{is},z,\lambda)\right\|^2\right]\Big\}^{1/2}\\
&\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{E\left\|X_{it}\right\|^4\cdot E\left\|X_{is}\right\|^4\Big\}^{1/2}=O\left(\frac{1}{N}\right)
\end{aligned}\tag{B.12}$$

where the first inequality follows from Cauchy–Schwarz inequality, the third inequality follows from $\beta(\cdot)$ and $L(Z_{it},z,\lambda)$ being bounded uniformly, and the last equality follows from Assumption A.3. It thus implies that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\beta(Z_{it})L(Z_{it},z,\lambda)-E[\Sigma_X(Z_{it})\beta(Z_{it})L(Z_{it},z,\lambda)]\to_P 0.$$

For the second term on RHS of (B.11), we write

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it}-\mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\\
&\quad+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda).
\end{aligned}\tag{B.13}$$

Notice that

$$\begin{aligned}
&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it}-\mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}'1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\right\|^2\Big|\mathcal{Z}\right]\\
&\le E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{T}{T_{it}}\frac{1}{T}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it}-\mu_X(Z_{it})\right\|\left\|\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\right\|\right)^2\Big|\mathcal{Z}\right]\\
&=E\Bigg[\Bigg(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{p(Z_{it})}p(Z_{it})\mu(Z_{it})+o_P(1)-\mu_X(Z_{it})\right\|\\
&\quad\cdot\left\|\frac{1}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\right\|\Bigg)^2\Big|\mathcal{Z}\Bigg]
\end{aligned}$$

$$\leq o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{T}\frac{T}{T_{it}}\sum_{s_2=1}^{T}X_{is_2}1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\right\|\right)^2\Big|\mathcal{Z}\right]$$
$$\leq o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T}\sum_{s_2=1}^{T}\|X_{is_2}\|\right)^2\Big|\mathcal{Z}\right] = o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s_2=1}^{T}\|X_{is_2}\|\right)^2\Big|\mathcal{Z}\right]$$
$$= o_P(1)\frac{1}{N^2T^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}E\left[\|X_{i_1t_1}\|^2\|X_{i_2t_2}\|^2\Big|\mathcal{Z}\right]$$
$$= o_P(1)\frac{1}{N^2T^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left\{E\left[\|X_{i_1t_1}\|^4\Big|Z_{i_1t_1}\right]E\left[\|X_{i_2t_2}\|^4\Big|Z_{i_2t_2}\right]\right\}^{1/2} = o_P(1), \tag{B.14}$$

where first equality follows from (B.1) and (B.2), the third inequality follows from (B.1) and the fact that $1_{is_2,it}$, $L(Z_{it},z,\lambda)$ and $\beta(\cdot)$ being bounded uniformly, and the last line follows from Assumption A.3. Thus, we obtain that

$$E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it}-\mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T}X'_{is_2}1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\right\|^2\right] = o_P(1),$$

which allows us to write

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X'_{is_2}1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)\\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\frac{1}{T_{it}}\sum_{s_2=1}^{T}X'_{is_2}1_{is_2,it}\beta(Z_{it})L(Z_{it},z,\lambda)+o_P(1)\\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\left(\frac{1}{T_{it}}\sum_{s_2=1}^{T}X'_{is_2}1_{is_2,it}-\mu_X(Z_{it})'\right)\beta(Z_{it})L(Z_{it},z,\lambda)\\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\mu_X(Z_{it})'\beta(Z_{it})L(Z_{it},z,\lambda)+o_P(1)\\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\mu_X(Z_{it})'\beta(Z_{it})L(Z_{it},z,\lambda)+o_P(1)\\
\to_P\ &E[\mu_X(Z_{it})\mu_X(Z_{it})'\beta(Z_{it})L(Z_{it},z,\lambda)],
\end{aligned} \tag{B.15}$$

where the first equality follows from (B.13) and (B.14), the third equality follows from the procedure similar to (B.14), and the last line follows from the same procedure as (B.7).

For the last two terms on RHS of (B.11), write

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}X'_{it}\beta(Z_{it})L(Z_{it},z,\lambda)\\
=\ &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})X'_{it}\beta(Z_{it})L(Z_{it},z,\lambda)+o_P(1)\\
\to_P\ &E[\mu_X(Z_{it})X'_{it}\beta(Z_{it})L(Z_{it},z,\lambda)] = E[\mu_X(Z_{it})\mu_X(Z_{it})'\beta(Z_{it})L(Z_{it},z,\lambda)],
\end{aligned}$$

where the first equality follows from the same procedure as (B.13) and (B.14), and the last line follows from the procedure similar to (B.7).

Based on the above, the result follows. ■

4). Set $L(Z_{it}, z, \lambda) = 1$ in the second result of this lemma. Then the result follows immediately. ■

5). Write

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'1(Z_{it}=z)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)'1(Z_{it}=z)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'1(Z_{it}=z)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}1(Z_{it}=z)\\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}X_{it}'1(Z_{it}=z)\\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}'1_{is,it}1(Z_{it}=z).
\end{aligned}
\tag{B.16}
$$

We now consider each term on RHS of (B.16) respectively. For the first term on RHS of (B.16), write

$$
\begin{aligned}
&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'1(Z_{it}=z)-p(z)\Sigma_X(z)\right\|^2\right]\\
&=\frac{1}{N^2T^2}\sum_{m=1}^{q}\sum_{n=1}^{q}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\Big[(X_{it,m}X_{it,n}1(Z_{it}=z)-p(z)\Sigma_{X,mn}(z))\\
&\cdot(X_{is,m}X_{is,n}1(Z_{is}=z)-p(z)\Sigma_{X,mn}(z))\Big]\\
&\le\frac{1}{N^2T^2}\sum_{m=1}^{q}\sum_{n=1}^{q}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{E\left[|X_{it,m}X_{it,n}|^2\right]E\left[|X_{is,m}X_{is,n}|^2\right]\right\}^{1/2}\\
&\le O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{E\left[\|X_{it}\|^4\right]E\left[\|X_{is}\|^4\right]\right\}^{1/2}=O\left(\frac{1}{N}\right),
\end{aligned}
\tag{B.17}
$$

where $X_{it,m}$ denotes the m^{th} element of X_{it} for $m = 1, \ldots, q$, $\Sigma_{X,mn}(z)$ denotes the $(m, n)^{th}$ element of $\Sigma_X(z)$ for $m, n = 1, \ldots, q$, the first inequality follows from Cauchy–Schwarz inequality, and and the second inequality follows from $L(Z_{it}, z, \lambda)$ being bounded uniformly.

It thus implies that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'1(Z_{it}=z)-p(z)\Sigma_X(z)\to_P 0.$$

For the second term on RHS of (B.16), we write

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}1_{is_1,it}X_{is_2}'1_{is_2,it}1(Z_{it}=z)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T} X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T} X'_{is_2}1_{is_2,it}1(Z_{it}=z)$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\frac{1}{T_{it}}\sum_{s_2=1}^{T} X'_{is_2}1_{is_2,it}1(Z_{it}=z). \tag{B.18}$$

Notice that

$$E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T} X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T} X'_{is_2}1_{is_2,it}1(Z_{it}=z)\right\|^2 \middle| \mathcal{Z}\right]$$
$$\leq E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{T}{T_{it}}\frac{1}{T}\sum_{s_1=1}^{T} X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right\|\left\|\frac{1}{T_{it}}\sum_{s_2=1}^{T} X_{is_2}1_{is_2,it}1(Z_{it}=z)\right\|\right)^2 \middle| \mathcal{Z}\right]$$
$$= E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{p(Z_{it})}p(Z_{it})\mu(Z_{it}) + o_P(1) - \mu_X(Z_{it})\right\|\left\|\frac{1}{T_{it}}\sum_{s_2=1}^{T} X_{is_2}1_{is_2,it}1(Z_{it}=z)\right\|\right)^2 \middle| \mathcal{Z}\right]$$
$$\leq o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\frac{1}{T}\frac{T}{T_{it}}\sum_{s_2=1}^{T} X_{is_2}1_{is_2,it}1(Z_{it}=z)\right\|\right)^2 \middle| \mathcal{Z}\right]$$
$$\leq o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T}\sum_{s_2=1}^{T}\|X_{is_2}\|\right)^2 \middle| \mathcal{Z}\right] = o_P(1)E\left[\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{s_2=1}^{T}\|X_{is_2}\|\right)^2 \middle| \mathcal{Z}\right]$$
$$= o_P(1)\frac{1}{N^2T^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} E\left[\|X_{i_1t_1}\|^2\|X_{i_2t_2}\|^2 \middle| \mathcal{Z}\right]$$
$$= o_P(1)\frac{1}{N^2T^2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left\{E\left[\|X_{i_1t_1}\|^4 \middle| Z_{i_1t_1}\right]E\left[\|X_{i_2t_2}\|^4 \middle| Z_{i_2t_2}\right]\right\}^{1/2} = o_P(1), \tag{B.19}$$

where the first equality follows from (B.1) and (B.2), the third inequality follows from (B.1) and the fact that $1_{is_2,it}$ and $1(Z_{it}=z)$ being bounded uniformly, and the last line follows from Assumption A.3. Thus, we obtain that

$$E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s_1=1}^{T} X_{is_1}1_{is_1,it} - \mu_X(Z_{it})\right)\frac{1}{T_{it}}\sum_{s_2=1}^{T} X'_{is_2}1_{is_2,it}1(Z_{it}=z)\right\|^2\right] = o_P(1),$$

which allows us to write

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T} X_{is_1}1_{is_1,it}X'_{is_2}1_{is_2,it}1(Z_{it}=z)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\frac{1}{T_{it}}\sum_{s_2=1}^{T} X'_{is_2}1_{is_2,it}1(Z_{it}=z) + o_P(1)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\left(\frac{1}{T_{it}}\sum_{s_2=1}^{T} X'_{is_2}1_{is_2,it} - \mu_X(Z_{it})'\right)1(Z_{it}=z)$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\mu_X(Z_{it})'1(Z_{it}=z) + o_P(1)$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})\mu_X(Z_{it})'1(Z_{it}=z) + o_P(1)$$

$$\to_P p(z)\mu_X(z)\mu_X(z)', \tag{B.20}$$

where the first equality follows from (B.18) and (B.19), the third equality follows from the procedure similar to (B.19), and the last line follows from the same procedure as (B.7).

For the last two terms on RHS of (B.16), write

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}X_{it}'1(Z_{it}=z) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(Z_{it})X_{it}'1(Z_{it}=z)+o_P(1)\\ &\to_P p(z)\mu_X(z)\mu_X(z)',\end{aligned}$$

where the first equality follows from the same procedure as (B.18) and (B.19), and the last line follows from the procedure similar to (B.7). Based on the above, the result follows. ■

6). Set $L(Z_{it},z,\lambda)=1$ in the third result of this lemma. Then the result follows immediately. ■

7). Write

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}1(Z_{it}=z)\\ &=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)1(Z_{it}=z)\\ &=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\mu_X(z))\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)1(Z_{it}=z)\\ &\quad-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}-\mu_X(z)\right)\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)1(Z_{it}=z).\end{aligned}\tag{B.21}$$

Firstly, we consider the next term on RHS of (B.21).

$$\begin{aligned}&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}-\mu_X(z)\right)u_{it}1(Z_{it}=z)\right\|^2\Big|\mathcal{Z}\right]\\ &\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}|E[u_{it_1}u_{jt_2}]|\,E\left[\left|\left(\frac{1}{T_{it_1}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it_1}-\mu_X(z)\right)'\right.\right.\\ &\quad\cdot\left.\left.\left(\frac{1}{T_{jt_2}}\sum_{s_2=1}^{T}X_{js_2}1_{js_2,jt_2}-\mu_X(z)\right)\right|1(Z_{it_1}=z)1(Z_{jt_2}=z)\Big|\mathcal{Z}\right]\\ &\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}|E[u_{it_1}u_{jt_2}]|\\ &\quad\left\{E\left[\left\|\frac{T}{T_{it_1}}\frac{1}{T}\sum_{s_1=1}^{T}X_{is_1}1(Z_{is_1}=z)-\mu_X(z)\right\|^2 1(Z_{it_1}=z)\Big|\mathcal{Z}\right]\right\}^{1/2}\\ &\quad\cdot\left\{E\left[\left\|\frac{T}{T_{jt_2}}\frac{1}{T}\sum_{s_2=1}^{T}X_{js_2}1(Z_{js_2}=z)-\mu_X(z)\right\|^2 1(Z_{jt_2}=z)\Big|\mathcal{Z}\right]\right\}^{1/2}\end{aligned}$$

$$
\begin{aligned}
&\le o_P(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T} c_{\delta_2}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)} \\
&\quad\cdot\left(E\left|u_{it_1}\right|^{2+\delta_2/2}\right)^{2/(4+\delta_2)}\left(E\left|u_{jt_2}\right|^{2+\delta_2/2}\right)^{2/(4+\delta_2)} \\
&\le o_P(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)} = o_P\left(\frac{1}{NT}\right),
\end{aligned}
$$

where $c_{\delta_2} = 2^{(4+2\delta_2)/(4+\delta_2)}\cdot(4+\delta_2)/\delta_2$, the first equality follows from Assumption A.4, the first inequality follows from Cauchy–Schwarz inequality, the second inequality follows from Davydov inequality and a procedure similar to (B.9), and the last line follows from Assumption A.4.

Similarly, we can obtain that

$$
E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}-\mu_X(z)\right)\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}1(Z_{it}=z)\right\|^2\Big|\mathcal{Z}\right] = o_P\left(\frac{1}{NT}\right)
$$

and

$$
E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu_X(z)\right)\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}1(Z_{it}=z)\right\|^2\Big|\mathcal{Z}\right] = o_P\left(\frac{1}{NT}\right).
$$

We therefore can further write

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}1(Z_{it}=z) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu_X(z)\right)u_{it}1(Z_{it}=z) + o_P\left(\frac{1}{\sqrt{NT}}\right). \quad \text{(B.22)}
$$

For the term on RHS of (B.22), write

$$
\begin{aligned}
&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu_X(z)\right)u_{it}1(Z_{it}=z)\right\|^2\right] \\
&\le \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E[(X_{it}-\mu_X(z))'\left(X_{js}-\mu_X(z)\right)u_{it}u_{js}1(Z_{it}=z)1(Z_{js}=z)]\right| \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E[u_{it}u_{js}]\right|\left|E[(X_{it}-\mu_X(z))'\left(X_{js}-\mu_X(z)\right)1(Z_{it}=z)1(Z_{js}=z)]\right| \\
&\le O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{\delta_2}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)}\left(E\left|u_{it}\right|^{2+\delta_2/2}E\left|u_{js}\right|^{2+\delta_2/2}\right)^{2/(4+\delta)} \\
&\quad\cdot\left\{E[\|X_{it}-\mu_X(z)\|^2]\cdot E[\|X_{js}-\mu_X(z)\|^2]\right\}^{1/2} \\
&\le O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)} = O\left(\frac{1}{NT}\right),
\end{aligned}
$$

where $c_{\delta_2} = 2^{(4+2\delta_2)/(4+\delta_2)}\cdot(4+\delta_2)/\delta_2$, the second equality follows from Assumption A.4, the first inequality follows from Davydov inequality and Cauchy–Schwarz inequality, and the last line follows from Assumptions A.3–A.4. Based on the above, the result follows. ■

8). Write

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}L(Z_{it},z,\lambda)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)L(Z_{it},z,\lambda)\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\mu_X(z))\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)L(Z_{it},z,\lambda)\\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}-\mu_X(z)\right)\left(u_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}\right)L(Z_{it},z,\lambda). \qquad \text{(B.23)}
\end{aligned}
$$

Firstly, we consider the next term on RHS of (B.23).

$$
\begin{aligned}
&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}-\mu_X(z)\right)u_{it}L(Z_{it},z,\lambda)\right\|^2\Big|\mathcal{Z}\right]\\
\le&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}|E[u_{it_1}u_{jt_2}]|\left|E\Bigg[\left(\frac{1}{T_{it_1}}\sum_{s_1=1}^{T}X_{is_1}1_{is_1,it_1}-\mu_X(z)\right)'\right.\\
&\cdot\left.\left(\frac{1}{T_{jt_2}}\sum_{s_2=1}^{T}X_{js_2}1_{js_2,jt_2}-\mu_X(z)\right)L(Z_{it_1},z,\lambda)L(Z_{jt_2},z,\lambda)\Big|\mathcal{Z}\Bigg]\right|\\
\le&\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}|E[u_{it_1}u_{jt_2}]|\left\{E\left[\left\|\frac{1}{T_{it_1}}\sum_{s_1=1}^{T}X_{is_1}1(Z_{is_1}=z)-\mu_X(z)\right\|^2\Big|\mathcal{Z}\right]\right\}^{1/2}\\
&\cdot\left\{E\left[\left\|\frac{1}{T_{jt_2}}\sum_{s_2=1}^{T}X_{js_2}1(Z_{js_2}=z)-\mu_X(z)\right\|^2\Big|\mathcal{Z}\right]\right\}^{1/2}\\
\le&O_P(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}c_{\delta_2}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)}\\
&\cdot\left(E\left|u_{it_1}\right|^{2+\delta_2/2}\right)^{2/(4+\delta_2)}\left(E\left|u_{jt_2}\right|^{2+\delta_2/2}\right)^{2/(4+\delta_2)}\\
\le&O_P(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)}=O_P\left(\frac{1}{NT}\right),
\end{aligned}
$$

where $c_{\delta_2}=2^{(4+2\delta_2)/(4+\delta_2)}\cdot(4+\delta_2)/\delta_2$, the first equality follows from Assumption A.4, the first inequality follows from Cauchy–Schwarz inequality and $L(Z_{it,z,\lambda})$ being bounded uniformly, the second inequality follows from Davydov inequality and a procedure similar to (B.9), and the last line follows from Assumption A.4.

Similarly, we can obtain that

$$
E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}1_{is,it}-\mu_X(z)\right)\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}L(Z_{it},z,\lambda)\right\|^2\Big|\mathcal{Z}\right]=O_P\left(\frac{1}{NT}\right)
$$

and

$$
E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\mu_X(z))\frac{1}{T_{it}}\sum_{s=1}^{T}u_{is}1_{is,it}L(Z_{it},z,\lambda)\right\|^2\Big|\mathcal{Z}\right]=O_P\left(\frac{1}{NT}\right).
$$

We therefore can further write

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}L(Z_{it},z,\lambda)\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu_X(z)\right)u_{it}L(Z_{it},z,\lambda)+O_P\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}\tag{B.24}
$$

For the term on RHS of (B.24), write

$$
\begin{aligned}
&E\left[\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu_X(z)\right)u_{it}L(Z_{it},z,\lambda)\right\|^2\right]\\
&=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E[(X_{it}-\mu_X(z))'\left(X_{js}-\mu_X(z)\right)u_{it}u_{js}L(Z_{it},z,\lambda)L(Z_{js},z,\lambda)]\\
&\le\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[u_{it}u_{js}]|\left|E[(X_{it}-\mu_X(z))'\left(X_{js}-\mu_X(z)\right)L(Z_{it},z,\lambda)L(Z_{js},z,\lambda)]\right|\\
&\le O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{\delta_2}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)}\left(E\left|u_{it}\right|^{2+\delta_2/2}E\left|u_{js}\right|^{2+\delta_2/2}\right)^{2/(4+\delta)}\\
&\quad\cdot\left\{E[\|X_{it}-\mu_X(z)\|^2]\cdot E[\|X_{js}-\mu_X(z)\|^2]\right\}^{1/2}\\
&\le O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\alpha_{u,ij}(|t_1-t_2|)\right)^{\delta_2/(4+\delta_2)}=O\left(\frac{1}{NT}\right),
\end{aligned}
$$

where $c_{\delta_2}=2^{(4+2\delta_2)/(4+\delta_2)}\cdot(4+\delta_2)/\delta_2$, the second equality follows from Assumption A.4, the first inequality follows from Davydov inequality and Cauchy–Schwarz inequality, and the last line follows from Assumptions A.3–A.4. Based on the above, the result follows. ■

9). Set $L(Z_{it},z,\lambda)=1$ in the eighth result of this lemma. Then the result follows immediately. ■

Notice that the finite sample property of the leave–one–out estimator is different from the estimator (2.5) provided in the main file which uses the whole sample, but they are interchangeable in the following analysis due to N and T being sufficiently large. Therefore, we express $\hat{\beta}_{it}$ as the estimator which uses the whole sample in the rest proof of this supplementary file. The similar approach can be seen on page 569 of Li et al. (2013).

$$
\begin{aligned}
\hat{\beta}_{it}-\beta_{it}=&\left(\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'L(Z_{js},Z_{it},\lambda)\right)^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{it})L(Z_{js},Z_{it},\lambda)\\
&+\left(\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'L(Z_{js},Z_{it},\lambda)\right)^{-1}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}L(Z_{js},Z_{it},\lambda),
\end{aligned}\tag{B.25}
$$

where we denote $\beta_{it}=\beta(Z_{it})$ for notational simplicity.

Proof of Lemma 1:

We use Theorem 2.1 of Newey and McFadden (1994) to verify that $\hat{\lambda} = o_P(1)$. By Assumption A.5, $CV_0(\lambda)$ is uniquely minimized at $\lambda = (\lambda_1, \ldots, \lambda_r)' = 0$. Each component of λ belongs to $[0,1]$, so λ belongs to a compact set $[0,1]^r$ on $\mathbb{R}^r$. Also, $CV_0(\lambda)$ is continuous on the support $[0,1]^r$ of λ. Then we need only to show that $CV(\lambda)$ converges uniformly in probability to $CV_0(\lambda) + c$ below, where c is a positive constant uniformly in λ. Thus, write

$$
\begin{aligned}
CV(\lambda) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\beta_{it}-\hat{\beta}_{it})\right)^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\beta_{it}-\hat{\beta}_{it})\tilde{u}_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \\
&\equiv CV_1(\lambda) + CV_2(\lambda) + CV_3. \qquad (B.26)
\end{aligned}
$$

The result (1) of Lemma A.2 implies $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \to_P \sigma_u^2$ uniformly in λ. Thus, we just need to focus on $CV_1(\lambda)$ and $CV_2(\lambda)$ below.

Before we proceed further, we investigate $\hat{\beta}_{it} - \beta_{it}$ firstly. We now consider each term on RHS of (B.25) respectively. By results (2), (3) and (8) of Lemma A.2, we can further write (B.25) as

$$
\hat{\beta}_{it} - \beta_{it} = \Sigma_{XX}^{-1}(Z_{it})\Sigma_{XX\beta}(Z_{it}) - \beta(Z_{it}) + o_P(1) = \eta(Z_{it}) - \beta(Z_{it}) + o_P(1). \qquad (B.27)
$$

By (B.27) and using derivations similar to result (3) of Lemma A.2, $CV_1(\lambda)$ can be rewritten as

$$
\begin{aligned}
CV_1(\lambda) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\beta_{it}-\hat{\beta}_{it})'\tilde{X}_{it}\tilde{X}_{it}'(\beta_{it}-\hat{\beta}_{it}) \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\eta(Z_{it})-\beta(Z_{it}))'\left(X_{it}-\mu_X(Z_{it})\right)\left(X_{it}-\mu_X(Z_{it})\right)'(\eta(Z_{it})-\beta(Z_{it})) + o_P(1),
\end{aligned}
$$

where the last line follows similarly from the derivation of (B.8).

For the term on RHS above, write

$$
\begin{aligned}
&E\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\eta(Z_{it})-\beta(Z_{it}))'\left(X_{it}-\mu_X(Z_{it})\right)\left(X_{it}-\mu_X(Z_{it})\right)'(\eta(Z_{it})-\beta(Z_{it})) - CV_0(\lambda)\right|^2 \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \\
&\quad E\Big[\left((\eta(Z_{it})-\beta(Z_{it}))'\left(X_{it}-\mu_X(Z_{it})\right)\left(X_{it}-\mu_X(Z_{it})\right)'(\eta(Z_{it})-\beta(Z_{it})) - CV_0(\lambda)\right) \\
&\quad \cdot\left((\eta(Z_{is})-\beta(Z_{is}))'\left(X_{is}-\mu_X(Z_{is})\right)\left(X_{is}-\mu_X(Z_{is})\right)'(\eta(Z_{is})-\beta(Z_{is})) - CV_0(\lambda)\right)\Big] \\
&\leq \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\Big\{E\left|(\eta(Z_{it})-\beta(Z_{it}))'(\Sigma_X(Z_{it})-\mu_X(Z_{it})\mu_X(Z_{it})')(\eta(Z_{it})-\beta(Z_{it}))\right|^2 \\
&\quad \cdot E\left|(\eta(Z_{is})-\beta(Z_{is}))'(\Sigma_X(Z_{is})-\mu_X(Z_{is})\mu_X(Z_{is})')(\eta(Z_{is})-\beta(Z_{is}))\right|^2\Big\}^{1/2} = O\left(\frac{1}{N}\right), \qquad (B.28)
\end{aligned}
$$

where the first equality follows from Assumption A.2, the first inequality follows from Cauchy–Schwarz inequality, the last equality follows from Assumption A.5. Therefore, we have shown that $CV_1(\lambda) \to_P CV_0(\lambda)$. Similarly, we can show that $CV_2(\lambda) = o_P(1)$.

Therefore, we have shown that $CV(\lambda) \to_P CV_0(\lambda) + \sigma_u^2$. So far we have verified all the conditions needed for Theorem 2.1 of Newey and McFadden (1994). Then the result follows. ■

Proof of Corollary 1:

All we need to show is that $\hat{\sigma}_u^2 \to_P \sigma_u^2$ in the proof of this corollary. Write

$$\hat{\sigma}_u^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{X}_{it}'(\beta(Z_{it}) - \hat{\beta}(Z_{it})) + \tilde{u}_{it})^2 = A_1 + A_2 + 2A_3,$$

where

$$A_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{X}_{it}'(\beta(Z_{it}) - \hat{\beta}(Z_{it})))^2, \quad A_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2,$$
$$A_3 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\beta(Z_{it}) - \hat{\beta}(Z_{it}))\tilde{u}_{it}.$$

For A_1, write

$$|A_1| \le \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\tilde{X}_{it}\right\|^2\left\|\beta(Z_{it}) - \hat{\beta}(Z_{it})\right\|^2 \le O_P\left(\frac{1}{NT}\right)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\tilde{X}_{it}\right\|^2 = O_P\left(\frac{1}{NT}\right),$$

where the second inequality follows from Theorem 2; the equality follows from Assumption A.3.(b). Thus, $A_1 \to_P 0$. Similarly, we can show that $A_3 \to_P 0$. By (1) of Lemma A.2, $A_2 \to_P \sigma_u^2$. Therefore, the result follows. ■

Notice that if we use Assumption B to replace Assumption A.5, Lemma A.2 still holds after some slight modification. Specifically, for (2)–(3) of Lemma A.2, $\Sigma_{XX}(z)$ and $\Sigma_{XX\beta}(z)$ become $\bar{\Sigma}_{XX}(\bar{z}) \cdot E[\tilde{L}(\tilde{Z}_{it}, \tilde{z}, \tilde{\lambda})]$ and $\bar{\Sigma}_{XX\beta}(\bar{z}) \cdot E[\tilde{L}(\tilde{Z}_{it}, \tilde{z}, \tilde{\lambda})]$ for $\forall z \in \mathcal{D}$, respectvely; for (4)–(6) of Lemma A.2, $\mu_X(z)$, $\Sigma_X(z)$ and $\beta(z)$ reduce $\mu_X(\bar{z})$, $\Sigma_X(\bar{z})$ and $\beta(\bar{z})$, respectively, (1) and (7)–(9) of Lemma A.2 hold without requiring any modification. Thus, to establish the following asymptotic results for the irrelevant case, we still can refer to the basic results proved in Lemma A.2.

Proof of Lemma 2:

By Assumption B, $CV_0^*(\bar{\lambda})$ is uniquely minimized at $\bar{\lambda} = (\lambda_1, \ldots, \lambda_{r_1})' = 0$. Each component of λ belongs to $[0, 1]$, so λ belongs to a compact set on $\mathbb{R}^r$. Also, $CV_0^*(\bar{\lambda})$ is continuous on the support $[0, 1]^{r_1}$ of $\bar{\lambda}$. Then we need only to show that $CV(\lambda)$ converges uniformly in probability to $CV_0^*(\bar{\lambda}) + c$ below, where c is a positive constant uniformly in $\bar{\lambda}$. Note that λ_s for $s = r_1 + 1, \ldots, r$ associated with the irrelevant covariates get canceled out in the asymptotic results, so they do not play a role as we minimize the cross–validation criterion function. Without loss of generality, λ_s for $s = r_1 + 1, \ldots, r$ can be considered as arbitrary constants. The following procedure holds uniformly in λ_s for $s = r_1+1, \ldots, r$.

Note also that for the irrelevant case the coefficient function reduces to $\beta(\bar{z})$. Thus, write

$$CV(\lambda) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\bar{\beta}_{it} - \hat{\beta}_{it})\right)^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\bar{\beta}_{it} - \hat{\beta}_{it})\tilde{u}_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2$$
$$\equiv CV_1(\lambda) + CV_2(\lambda) + CV_3, \tag{B.29}$$

where $\bar{\beta}_{it} = \beta(\bar{Z}_{it})$ for simplicity. The result (1) of Lemma A.2 has shown that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \to_P \sigma_u^2$

uniformly in $\bar{\lambda}$. Thus, we just need to focus on $CV_1(\lambda)$ and $CV_2(\lambda)$ below. Investigate $\hat{\beta}_{it} - \bar{\beta}_{it}$ firstly. Due to the same reason as (B.25), write

$$\hat{\beta}_{it} - \bar{\beta}_{it} = \left(\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda)\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\bar{\beta}_{js} - \bar{\beta}_{it}) L(Z_{js}, Z_{it}, \lambda)$$
$$+ \left(\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda)\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{u}_{js} L(Z_{js}, Z_{it}, \lambda). \tag{B.30}$$

Recall that $L(Z_{js}, z, \lambda) = \bar{L}(\bar{Z}_{js}, \bar{z}, \bar{\lambda})\tilde{L}(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})$, where

$$\bar{L}(\bar{Z}_{js}, \bar{z}, \bar{\lambda}) = \prod_{s=1}^{r_1} \lambda_s^{1(Z_{it,s}\neq z_s)} \quad \text{and} \quad \tilde{L}(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda}) = \prod_{s=r_1+1}^{r} \lambda_s^{1(Z_{it,s}\neq z_s)}.$$

As discussed before, Lemma A.2 holds with some slight modification after replacing Assumption A.5 by Assumption B. Thus, it is easy to know that $\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{u}_{js} L(Z_{js}, z, \lambda) \to_P 0$. Moreover, for $\forall z \in \mathcal{D}$,

$$\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, z, \lambda) \to_P \bar{\Sigma}_{XX}(\bar{z}) \cdot E[\tilde{L}(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})] \tag{B.31}$$

and

$$\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}' \beta(\bar{Z}_{js}) L(Z_{js}, z, \lambda) \to_P \bar{\Sigma}_{XX\beta}(\bar{z}) \cdot E[\tilde{L}(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})]. \tag{B.32}$$

Notice that $E[\tilde{L}(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})]$ gets canceled out after we plug (B.31) and (B.32) in (B.30). We thus write (B.30) as

$$\hat{\beta}_{it} - \bar{\beta}_{it} = \bar{\Sigma}_{XX}^{-1}(\bar{Z}_{it})\bar{\Sigma}_{XX\beta}(\bar{Z}_{it}) - \beta(\bar{Z}_{it}) + o_P(1) = \bar{\eta}(\bar{Z}_{it}) - \beta(\bar{Z}_{it}) + o_P(1). \tag{B.33}$$

By (B.33) and derivations similar to result (3) of Lemma A.2, $CV_1(\lambda)$ can be rewritten as

$$\begin{aligned} CV_1(\lambda) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} (\bar{\beta}_{it} - \hat{\beta}_{it})'\tilde{X}_{it}\tilde{X}_{it}'(\bar{\beta}_{it} - \hat{\beta}_{it}) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} (\bar{\eta}(\bar{Z}_{it}) - \beta(\bar{Z}_{it}))' \left(X_{it} - \mu_X(\bar{Z}_{it})\right)\left(X_{it} - \mu_X(\bar{Z}_{it})\right)' (\bar{\eta}(\bar{Z}_{it}) - \beta(\bar{Z}_{it})) + o_P(1) \\ &\to_P CV_0^*(\bar{\lambda}), \end{aligned}$$

where the second equality follows from a procedure similar to (B.8); the last line follows from a procedure same as (B.28). Therefore, we have shown that $CV_1(\lambda) \to_P CV_0^*(\bar{\lambda})$. Similarly, we can show that $CV_2(\lambda) = o_P(1)$. Based on the above, we have shown that $CV(\lambda) \to_P CV_0^*(\bar{\lambda}) + \sigma_u^2$ uniformly in $\tilde{\lambda} \in \tilde{\mathcal{D}}$. So far we have verified all the conditions needed for Theorem 2.1 of Newey and McFadden (1994). Then the result follows. ■

Proof of Theorem 4:

The kernel functions for the relevant and irrelevant covariates are given as follows.

$$\bar{L}_{js\bar{\lambda}} = \prod_{s=1}^{r_1} \hat{\lambda}_s^{1(Z_{it,s}\neq z_s)} \quad \text{and} \quad \tilde{L}_{js\tilde{\lambda}} = \prod_{s=r_1+1}^{r} \lambda_s^{1(Z_{it,s}\neq z_s)},$$

where $\hat{\lambda}_s$ for $s = 1, \dots, r_1$ is the estimate of λ_s by minimizing the CV criterion function; and λ_s for $s = r_1 + 1, \dots, r$ is any arbitrary constant belongs to $[0, 1]$.

Denote that $\hat{\bar{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_{r_1})'$, $\bar{1}_{\bar{Z}_{it},\bar{z}} = 1(\bar{Z}_{it} = \bar{z})$ and $\bar{1}_{n,\bar{Z}_{js},\bar{z}} = 1(Z_{it,n} \neq z_n)\prod_{m=1,m\neq n}^{r_1} 1(Z_{it,m} = z_m)$ for $n = 1, \dots, r_1$. Thus, write

$$\begin{aligned}\hat{\beta}(z) - \beta(\bar{z}) &= A_{0,NT}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\bar{L}_{js\bar{\lambda}}\tilde{L}_{it\tilde{\lambda}} \\ &\quad + A_{0,NT}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{u}_{js}\bar{L}_{js\bar{\lambda}}\tilde{L}_{it\tilde{\lambda}},\end{aligned}$$

where $A_{0,NT} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'\bar{L}_{js\bar{\lambda}}\tilde{L}_{it\tilde{\lambda}}$.

We have shown $A_{0,NT} = O_P(1)$ and $\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{u}_{js}\bar{L}_{js\bar{\lambda}}\tilde{L}_{it\tilde{\lambda}} = O_P\left(\frac{1}{\sqrt{NT}}\right)$ in the proof of Theorem 3. Thus, we need only to focus on the next term:

$$\begin{aligned}&\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\left(\bar{1}_{\bar{Z}_{js},\bar{z}} + \sum_{n=1}^{r_1}\hat{\lambda}_n\bar{1}_{n,\bar{Z}_{js},\bar{z}} + O(\|\hat{\bar{\lambda}}\|^2)\right)\tilde{L}_{it\tilde{\lambda}} \\ &= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\bar{1}_{\bar{Z}_{js},\bar{z}}\tilde{L}_{it\tilde{\lambda}} \\ &\quad + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\sum_{n=1}^{r_1}\hat{\lambda}_n\bar{1}_{n,\bar{Z}_{js},\bar{z}}\tilde{L}_{it\tilde{\lambda}} \\ &\quad + O(\|\hat{\bar{\lambda}}\|^2)\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} \tilde{X}_{js}\tilde{X}_{js}'(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\tilde{L}_{it\tilde{\lambda}} \\ &= 0 + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T} X_{js}X_{js}'(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\left(\sum_{n=1}^{r_1}\hat{\lambda}_n\bar{1}_{n,\bar{Z}_{js},\bar{z}}\right)\tilde{L}_{it\tilde{\lambda}} + O_P\left(\frac{1}{NT}\right) \\ &= O_P\left(\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

where the second equality follows from $(\beta(\bar{Z}_{js}) - \beta(\bar{z}))\bar{1}_{\bar{Z}_{js},\bar{z}} = 0$ and Theorem 3, and the last line follows from the procedure same as (5) and (6) of Lemma A.2. Based on the above, the proof is complete. ■