Supplementary Document for the submission of
"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 [‡]University of Exeter

Appendix B

In this file, we provide the algorithm for the variable selection procedure and the proofs of the asymptotic results.

B.1 Algorithm

The procedure for obtaining regularized estimates is described as follows:

1. Minimize the cross-validation criterion function (2.8) in order to choose $\hat{\lambda}$.

2. Select $\tilde{\tau}$ defined in (2.15) from a sufficient large set, say $[1, \sqrt[4]{NT}]$, by using a grid search. For each choice of $\tilde{\tau}$, estimate (2.13) using a similar procedure as proposed in Hunter and Li (2005) and Wang and Xia (2009). Define

$$\hat{B}_{\tilde{\tau}}^{(n)} = (\hat{\beta}_{\tilde{\tau},1}^{(n)}, \ldots, \hat{\beta}_{\tilde{\tau},m}^{(n)})' = (\hat{b}_{\tilde{\tau},1}^{(n)}, \ldots, \hat{b}_{\tilde{\tau},q}^{(n)}) \tag{B.1}$$

to be the estimate obtained in the n^{th} iteration. Then the loss function given above can be locally approximated by

$$\begin{aligned}&\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\beta_j\right)^2 L(Z_i, z^j, \hat{\lambda}) + \sum_{s=1}^{q}\tilde{\tau}_s\frac{\|b_s\|^2}{\|\hat{b}_{\tilde{\tau},s}^{(n)}\|}\\ =&\sum_{j=1}^{m}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\beta_j\right)^2 L(Z_i, z^j, \hat{\lambda}) + \sum_{s=1}^{q}\tilde{\tau}_s\frac{\beta_{j,s}^2}{\|\hat{b}_{\tilde{\tau},s}^{(n)}\|}\right).\end{aligned} \tag{B.2}$$

The minimizer of (B.2) is given by $\hat{B}_{\tilde{\tau}}^{(n+1)} = (\hat{\beta}_{\tilde{\tau},1}^{(n+1)}, \ldots, \hat{\beta}_{\tilde{\tau},m}^{(n+1)})'$, where for $j = 1, \ldots, m$

$$\hat{\beta}_{\tilde{\tau},j}^{(n+1)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'L(Z_i, z^j, \hat{\lambda}) + D^{(n)}\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it}L(Z_i, z^j, \hat{\lambda}), \tag{B.3}$$

and $D^{(n)} = \text{diag}\left(\|\hat{b}_{\tilde{\tau},1}^{(n)}\|^{-1}\tilde{\tau}_1, \ldots, \|\hat{b}_{\tilde{\tau},q}^{(n)}\|^{-1}\tilde{\tau}_q\right)$. Repeat this procedure until $\|\hat{B}_{\tilde{\tau}}^{(n+1)} - \hat{B}_{\tilde{\tau}}^{(n)}\| < tolerance$, where *tolerance* is a sufficiently small number (say, 10^{-8}).

3. Select the optimal estimator based on the modified BIC-type criterion.

Computer code functions for implementing this procedure are available upon request and will be available at one of the authors' website.

B.2 Proofs

For notational simplicity, let $\hat{\beta}_{it} = \hat{\beta}_{-it}(Z_{it})$, $\beta_{it} = \beta(Z_{it})$, $1_{js,it} = 1(Z_{js} = Z_{it})$ and $\mathcal{Z} = \{Z_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$. Recall that we have defined $CV_0(\lambda)$, $\eta(z)$, $\Sigma_{XX}(z)$ and $\Sigma_{XX\beta}(z)$ in (A.1) of the main text. In the following proof, we will use these symbols without defining them again. Also, in this note $O(1)$ denotes some constants which may be different at each appearance.

Lemma B.1. *For two square matrices A and B with the same dimensions, suppose that A is non-singular and $\|A^{-1}B\| < 1$. 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 B.1 is straightforward and thus omitted. ■

Lemma B.2. *Under Assumption A, as $(N,T) \to (\infty, \infty)$ jointly, the following results hold uniformly in $z \in \mathcal{D}$ and $\lambda \in [0,1]^r$:*

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

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

$$\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \tilde{u}_{it}^2 = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \left(u_{it} - \frac{1}{T_{it}}\sum_{s=1}^T u_{is} L_{is,it}^p\right)\left(u_{it} - \frac{1}{T_{it}}\sum_{s=1}^T u_{is} L_{is,it}^p\right)$$
$$= \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T u_{it}^2 + \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{T_{it}}\sum_{s_1=1}^T u_{is_1} L_{is_1,it}^p \frac{1}{T_{it}}\sum_{s_2=1}^T u_{is_2} L_{is_2,it}^p$$

$$-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}u_{it}u_{is}L_{is,it}^{p}. \tag{B.4}$$

For the first term on RHS of (B.4), write

$$\begin{aligned}
&E\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}-\sigma_{u}^{2}\right|^{2}\\
&\leq \frac{1}{N^{2}T^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}c_{\delta}\left(\alpha_{ij}(|t-s|)\right)^{\delta/(4+\delta)}\left(E[u_{it}^{4+\delta}|\mathcal{X}]\cdot E[u_{js}^{4+\delta}|\mathcal{X}]\right)^{2/(4+\delta)}\\
&\leq O(1)\frac{1}{N^{2}T^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\alpha_{u,ij}(|t-s|)\right)^{\delta/(4+\delta)}=O\left(\frac{1}{NT}\right),
\end{aligned} \tag{B.5}$$

where $c_{\delta}=2^{(4+2\delta)/(4+\delta)}\cdot(4+\delta)/\delta$; 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.2.

For the third term on right hand side of (B.4),

$$\begin{aligned}
&E\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}u_{it}u_{is}L_{is,it}^{p}\right| \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left[|u_{it}|\left|\frac{T}{T_{it}}\right|\left|\frac{1}{T}\sum_{s=1}^{T}u_{is}L_{is,it}^{p}\right|\right]\\
&\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left[|u_{it}|\left|\frac{T}{T_{it}}\right|\right]^{2}E\left|\frac{1}{T}\sum_{s=1}^{T}u_{is}L_{is,it}^{p}\right|^{2}\right\}^{1/2}\\
&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\sigma_{u}^{2}E\left|\frac{T}{T_{it}}\right|^{2}E\left|\frac{1}{T}\sum_{s=1}^{T}u_{is}L_{is,it}^{p}\right|^{2}\right\}^{1/2}\\
&\leq\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma_{u}^{2}E\left|\frac{T}{T_{it}}\right|^{2}\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left|\frac{1}{T}\sum_{s=1}^{T}u_{is}L_{is,it}^{p}\right|^{2}\right\}^{1/2}\\
&\leq O(1)\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left|\frac{T}{T_{it}}\right|^{2}\right\}^{1/2}\left\{\frac{1}{NT^{2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[u_{it}u_{is}|\mathcal{X}]|\right\}^{1/2}=O\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}$$

where the second and third inequalities follow from Cauchy-Schwarz inequality; the first and second equalities follow from Assumptions A.2; the fourth inequality follows from the uniform boundedness of $L_{is,it}^{P}$. We then immediately obtain that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}u_{it}u_{is}L_{is,it}^{p}=o_{P}\left(1\right).$$

Similarly, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s_1=1}^{T}u_{is_1}L_{is_1,it}^{p}\frac{1}{T_{it}}\sum_{s_2=1}^{T}u_{is_2}L_{is_2,it}^{p} = o_P(1)$, which completes the proof of the first result of this lemma.

2) We now use Lemma A2 of Newey and Powell (2003) to verify this result hold uniformly in $z\in\mathcal{D}$ and $\lambda\in[0,1]^{r}$. Given that the cardinality of $\mathcal{D}$ is finite, it suffices to show that this result hold uniformly in $\lambda\in[0,1]^{r}$ for any given z.

Step 1: $[0,1]^r$ is a compact subset of $\mathbb{R}^r$ with Euclidean norm $\|\cdot\|$.

Step 2: Rewrite (2) of Lemma B.2 as follows:

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'L(Z_{it},z,\lambda) \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L_{is,it}^{p}\right)\left(X_{it}-\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L_{is,it}^{p}\right)'L(Z_{it},z,\lambda) \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'L(Z_{it},z,\lambda)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}L_{is_1,it}^{p}X_{is_2}'L_{is_2,it}^{p}L(Z_{it},z,\lambda) \\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L_{is,it}^{p}X_{it}'L(Z_{it},z,\lambda) \\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}'L_{is,it}^{p}L(Z_{it},z,\lambda).
\end{aligned} \tag{B.6}
$$

We now consider each term on RHS of (B.6) respectively.

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

$$
\begin{aligned}
&E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L_{is,it}^{p}X_{it}'L(Z_{it},z,\lambda)-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta_3(Z_{it},\lambda)X_{it}'L(Z_{it},z,\lambda)\right\| \\
=&E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{it}X_{it}'L(Z_{it},z,\lambda)\right\|\leq\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left\|K_{it}X_{it}'L(Z_{it},z,\lambda)\right\| \\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\left\|K_{it}\right\|^2E\left\|X_{it}L(Z_{it},z,\lambda)\right\|^2\right\}^{1/2} \\
\leq&\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left\|K_{it}\right\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left\|X_{it}L(Z_{it},z,\lambda)\right\|^2\right\}^{1/2}=o(1),
\end{aligned}
$$

where $K_{it}=\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L_{is,it}^{p}-\Delta_3(Z_{it},\lambda)$ has been defined in Assumption A.2. We then obtain that for any given $z\in\mathcal{D}$ and $\lambda\in[0,1]^r$

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L_{is,it}^{p}X_{it}'L(Z_{it},z,\lambda)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta_3(Z_{it},\lambda)\mu_X(Z_{it})'L(Z_{it},z,\lambda)+o_P(1)
$$

Similarly, for the second term on RHS of (B.6), we have

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}X_{is_1}L_{is_1,it}^{p}X_{is_2}'L_{is_2,it}^{p}L(Z_{it},z,\lambda) \\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta_3(Z_{it},\lambda)\Delta_3(Z_{it},\lambda)'L(Z_{it},z,\lambda)+o_P(1).
\end{aligned}
$$

According to the above, we can obtain

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

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \Delta_3(Z_{it},\lambda))(X_{it} - \Delta_3(Z_{it},\lambda))' L(Z_{it}, z, \lambda) + o_P(1)$$
$$:= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\Delta}_{it} + o_P(1),$$

for any given $z \in \mathcal{D}$ and $\lambda \in [0,1]^r$, where the definition of $\tilde{\Delta}_{it}$ should be obvious.

We then just need to consider

$$E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\Delta}_{it} - \Sigma_{XX}(z,\lambda)\right\|^2$$
$$= \frac{1}{N^2T^2}\sum_{h=1}^{q}\sum_{l=1}^{q}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[(\tilde{\Delta}_{it,hl} - \Sigma_{XX,hl}(z,\lambda))(\tilde{\Delta}_{js,hl} - \Sigma_{XX,js}(z,\lambda))\right]$$
$$\leq \frac{1}{N^2T^2}\sum_{h=1}^{q}\sum_{l=1}^{q}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}$$
$$\cdot c_\delta\left(\alpha_{ij}(|t-s|)\right)^{\frac{\delta}{4+\delta}}\left(E|\tilde{\Delta}_{it,hl} - \Sigma_{XX,hl}(z,\lambda)|^{2+\frac{\delta}{2}}\right)^{\frac{2}{4+\delta}}\left(E|\tilde{\Delta}_{js,hl} - \Sigma_{XX,js}(z,\lambda)|^{2+\frac{\delta}{2}}\right)^{\frac{2}{4+\delta}}$$
$$\leq O(1)\frac{1}{N^2T^2}\sum_{h=1}^{q}\sum_{l=1}^{q}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\alpha_{ij}(|t-s|)\right)^{\frac{\delta}{4+\delta}} = O\left(\frac{1}{NT}\right), \tag{B.7}$$

where $\tilde{\Delta}_{it,hl}$ and $\Sigma_{XX,hl}(z,\lambda)$ denote the $(h,l)^{th}$ elements of $\tilde{\Delta}_{it}$ and $\Sigma_{XX}(z,\lambda)$ respectively; the first inequality follows from Davydov inequality; the second inequality and the last equality follow from Assumption A.2.

Therefore, we have proved that

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

for any given $z \in \mathcal{D}$ and $\lambda \in [0,1]^r$.

Step 3: By *Step 2*, we can write

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

By expansion (2.5) of the main text, for any $\lambda_1, \lambda_2 \in [0,1]^r$, we have

$$\|\Sigma_{XX}(z,\lambda_1) - \Sigma_{XX}(z,\lambda_2)\| \leq O(1)\|\lambda_1 - \lambda_2\|,$$

which immediately implies the third condition of Lemma A2 of Newey and Powell (2003) holds.

By *Steps 1-3*, the result follows.

3). By following the procedure same as that given in 2) of this lemma, the result follows.

4)-6). These three results are special cases of 2) and 3) of this lemma.

7). Similar to the proofs of results (1) and (2) of this lemma, the result follows. 8)-9). These two results are special cases of result (7) of this lemma. ■

Note that the finite sample property of the leave-one-out estimator is different from the estimator in (2.7) provided in the main file which uses the whole sample, but they are interchangeable in the following analysis due to the assumption that both N and T are sufficiently large. Therefore, we express $\hat{\beta}_{it}$ as the estimator which uses the whole sample in what follows. A similar technique is also used in Li et al. (2013, p. 569).

$$\begin{aligned}\hat{\beta}_{it} - \beta_{it} =& \left(\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda)\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js} - \beta_{it})L(Z_{js}, Z_{it}, \lambda) \\ &+ \left(\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda)\right)^{-1} \sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js} L(Z_{js}, Z_{it}, \lambda), \end{aligned} \tag{B.8}$$

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

Proof of Lemma 2.1.1:

We use Theorem 2.1 of Newey and McFadden (1994) to verify that $\hat{\lambda} = o_P(1)$. By Assumption A.3, $CV_0(\lambda)$ is uniquely minimized at $\lambda = (\lambda_1, \ldots, \lambda_r)' = 0$. Here, λ belongs to a compact set $[0,1]^r$, and $CV_0(\lambda)$ is continuous on $[0,1]^r$. 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. For this purpose, write

$$\begin{aligned} CV(\lambda) =& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\beta_{it} - \hat{\beta}_{it}) + \gamma_{it}\right)^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\beta_{it} - \hat{\beta}_{it}) + \gamma_{it}\right)\tilde{u}_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \\ \equiv& CV_1(\lambda) + CV_2(\lambda) + CV_3, \end{aligned} \tag{B.9}$$

where $\gamma_{it} = \frac{1}{T_{it}}\sum_{s=1}^{T} X_{is}' \left(\beta(Z_{is}) - \beta(Z_{it})\right) L_{is,it}^p$.

The result (1) of Lemma B.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 proceeding further, we first investigate $\hat{\beta}_{it} - \beta_{it}$ and γ_{it}. By results (2), (3) and (7) of Lemma B.2, we can further write

$$\begin{aligned} \hat{\beta}_{it} - \beta_{it} &= \Sigma_{XX}^{-1}(Z_{it}, \lambda)\Sigma_{XX\beta}(Z_{it}, \lambda) - \beta(Z_{it}) + o_P(1) = \eta(Z_{it}, \lambda) - \beta(Z_{it}) + o_P(1), \\ \gamma_{it} &= \Delta_{3\beta}(Z_{it}, \lambda) - \Delta_3(Z_{it}, \lambda)'\beta(Z_{it}) + o_P(1). \end{aligned} \tag{B.10}$$

By (B.10), $CV_1(\lambda)$ can be rewritten as

$$\begin{aligned} CV_1(\lambda) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(&(X_{it} - \Delta_3(Z_{it}, \lambda))'(\beta(Z_{it}) - \eta(Z_{it}, \lambda)) \\ &+\Delta_{3\beta}(Z_{it}, \lambda) - \Delta_3(Z_{it}, \lambda)'\beta(Z_{it})\big)^2 + o_P(1). \end{aligned}$$

Similar to (B.7), we can obtain that $CV_1(\lambda) \to_P CV_0(\lambda)$ for any given λ. Using the procedure same as the proof of result (2) of Lemma B.2, it is easy to show $CV_1(\lambda) \to_P CV_0(\lambda)$ uniformly. Similar to the proof of result (7) of Lemma B.2, we can show that $CV_2(\lambda) = o_P(1)$ uniformly in λ.

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

Proof of Theorem 2.1.1:

In Lemma 2.1.1, we have shown $\hat{\lambda} = o_P(1)$, so it is reasonable to assume that λ, in proving this theorem, is sufficiently small and close to $0_{r\times 1}$. We now investigate the cross-validation criterion function and write

$$\begin{aligned} CV(\lambda) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\beta_{it}-\hat{\beta}_{it})\right)^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\beta_{it}-\hat{\beta}_{it})\tilde{u}_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \\ &\quad + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\beta_{it}-\hat{\beta}_{it}) + \tilde{u}_{it}\right)\gamma_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{it}^2 \\ &\equiv CV_1(\lambda) + CV_2(\lambda) + CV_3 + CV_4(\lambda) + CV_5(\lambda), \end{aligned} \tag{B.11}$$

where $\gamma_{it} = \frac{1}{T_{it}}\sum_{s=1}^{T} X_{is}'\left(\beta(Z_{it}) - \beta(Z_{is})\right) L_{is,it}^{p}$.

In (2.6), we have shown that $\gamma_{it} = O\left(\|\lambda\|^p\right)$ uniformly when λ is sufficiently small. In connection with the construction of $CV_4(\lambda)$ and $CV_5(\lambda)$, and Lemma B.2, we are able to obtain that $CV_4(\lambda) = O_P\left(\|\lambda\|^p\right)$ and $CV_5(\lambda) = O\left(\|\lambda\|^{2p}\right)$. By (1) of Lemma B.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 our analysis, we need to further consider $\hat{\beta}_{it} - \beta_{it}$. By Lemma 2.1.1, we can express the the kernel function as

$$L(Z_{js}, Z_{it}, \lambda) = 1_{js,it} + \sum_{m=1}^{r}\lambda_m 1_{m,jsit} + O(\|\lambda\|^2), \tag{B.12}$$

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

In what follows, we substitute (B.12) into each term on RHS of (B.8). Firstly,

$$\begin{aligned} &\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}' L(Z_{js}, Z_{it}, \lambda) \\ &= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}' 1_{js,it} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'\sum_{m=1}^{r}\lambda_m 1_{m,jsit} + O_P(\|\lambda\|^2) \\ &\equiv A_{1it} + A_{2it\lambda} + O_P(\|\lambda\|^2), \end{aligned} \tag{B.13}$$

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

Secondly, we have

$$\begin{aligned} &\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{it}) L(Z_{js}, Z_{it}, \lambda) \\ &= \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{it}) 1_{js,it} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'(\beta_{js}-\beta_{it})\sum_{m=1}^{r}\lambda_m 1_{m,jsit} + O_P(\|\lambda\|^2) \\ &\equiv 0 + B_{2it\lambda} + O_P(\|\lambda\|^2), \end{aligned} \tag{B.14}$$

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

Thirdly, we obtain

$$
\begin{aligned}
&\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}L(Z_{js},Z_{it},\lambda)\\
&=\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}1_{js,it}+\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}\sum_{m=1}^{r}\lambda_m 1_{m,jsit}+O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right)\\
&\equiv C_{1it}+C_{2it\lambda}+O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right), \qquad \text{(B.15)}
\end{aligned}
$$

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

For the terms on RHS of (B.13)-(B.15), by Lemma B.2, it is straightforward to obtain

$$
\begin{aligned}
&A_{1it}^{-1}=O_P(1),\quad A_{2it\lambda}=O_P\left(\|\lambda\|\right),\quad B_{2it\lambda}=O_P(\|\lambda\|),\\
&C_{1it}=O_P\left(\frac{1}{\sqrt{NT}}\right),\quad C_{2it\lambda}=O_P\left(\frac{\|\lambda\|}{\sqrt{NT}}\right). \qquad \text{(B.16)}
\end{aligned}
$$

By (B.13), using Lemma B.1 twice gives the following expression:

$$
\begin{aligned}
&\left(\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'L(Z_{js},Z_{it},\lambda)\right)^{-1}=\left(A_{1it}+A_{2it\lambda}+O_P(\|\lambda\|^2)\right)^{-1}\\
&=(A_{1it}+A_{2it\lambda})^{-1}+O_P(\|\lambda\|^2)=A_{1it}^{-1}-A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}+O_P(\|\lambda\|^2). \qquad \text{(B.17)}
\end{aligned}
$$

We then use (B.16) and (B.17) to further simplify (B.8) as follows:

$$
\hat{\beta}_{it}=\beta_{it}+\left(A_{1it}^{-1}-A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}\right)(B_{2it\lambda}+C_{1it}+C_{2it\lambda})+O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right)+O_P\left(\|\lambda\|^3\right) \qquad \text{(B.18)}
$$

We are now ready to further analyze $CV_1(\lambda)$ and $CV_2(\lambda)$ by using (B.16) and (B.18):

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

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

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

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

where the first equality follows from (9) of Lemma B.2 and (B.18); and the second equality follows from (9) of Lemma B.2 and (B.16).

Note 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 third equality follows from changing (it, js, kr) to (kr, it, js); the fourth equality follows from the definition of $1_{it,kr}$; and the fifth equality follows from the definition of A_{1it}. Note that the term on RHS of the above equation can be cancelled out by the leading term of $CV_2(\lambda)$.

Thus, we are now able to further write

$$\begin{aligned}
CV(\lambda) = &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(D_{3it}^2 - 2D_{1it}D_{2it}\right) + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}C_{1,it}\\
&-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}C_{2it\lambda} + O_P\left(\frac{\|\lambda\|^2}{\sqrt{NT}}\right) + O_P\left(\|\lambda\|^3\right)\\
&+ \text{terms independent of } \lambda.
\end{aligned} \tag{B.19}$$

Moreover, by (B.16) and some tedious algebra, we can show

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

Based on the above discussions, (B.19) can be further simplified as follows.

$$CV(\lambda) = O_P\left(\frac{\|\lambda\|}{NT}\right) + O_P(\|\lambda\|^2) + \text{ terms independent of } \lambda, \tag{B.20}$$

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

Proof of Theorem 2.1.2:

Denote

$$\check{T}_{it} = \sum_{s=1}^{T} 1(Z_{is} = Z_{it}), \quad \check{Y}_{it} = Y_{it} - \frac{1}{\check{T}_{it}} \sum_{s=1}^{T} Y_{is} 1_{is,it},$$
$$\check{X}_{it} = X_{it} - \frac{1}{\check{T}_{it}} \sum_{s=1}^{T} X_{is} 1_{is,it}, \quad \check{u}_{it} = u_{it} - \frac{1}{\check{T}_{it}} \sum_{s=1}^{T} u_{is} 1_{is,it}.$$

Note that for the large N and small T case, $\frac{1}{\check{T}_{it}} \sum_{s=1}^{T} u_{is} 1_{is,it}$ should be replaced by

$$A_{u,it} = \lim_{\lambda \to 0_{r\times 1}} \sum_{s=1}^{T} u_{is} L_{is,it}^{p} / \sum_{s=1}^{T} L_{is,it}^{p}$$

as discussed in Section 2.4. $\frac{1}{\check{T}_{it}} \sum_{s=1}^{T} Y_{is} 1_{is,it}$ and $\frac{1}{\check{T}_{it}} \sum_{s=1}^{T} X_{is} 1_{is,it}$ should be changed in a similar fashion.

Expanding the kernel functions in (2.7) by (B.12) easily leads to $\hat{\beta}(z) = \check{\beta}(z) + O_P\left(\frac{1}{NT}\right)$ by Lemma B.1 and Theorem 2.1.1, where

$$\check{\beta}(z) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \check{X}_{it}\check{X}_{it}' 1(Z_{it} = z)\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \check{X}_{it}\check{Y}_{it}' 1(Z_{it} = z).$$

Thus, it is straightforward to obtain $\sqrt{NT}(\hat{\beta}(z) - \beta(z)) = \sqrt{NT}(\check{\beta}(z) - \beta(z)) + O_P\left(\frac{1}{\sqrt{NT}}\right)$. Below we just need to focus on $\sqrt{NT}(\check{\beta}(z) - \beta(z))$, so write

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

where the second equality is due to $(\beta(Z_{it}) - \beta(z))1(Z_{it} = z) = 0$.

As with (5) of Lemma B.2, it is easy to show that

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

Therefore, we need only to focus on $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} \check{X}_{it}\check{u}_{it} 1(Z_{it} = z)$. Simple algebra shows

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

Thus, we 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 Assumptions A.2, it is easy to check $V_{T,N}(t)$ is stationary and α-mixing by following the same argument given under (A.4) on page 13 of Chen et al. (2012). 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)). For this purpose, we 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. (2012), we obtain that $E\left\|\sum_{n=1}^{k_T}\bar{V}_n\right\|^2=O\left(\frac{k_Ts_T}{T}\right)=o(1)$ and $E\left\|\hat{V}\right\|^2=O\left(\frac{T-k_Tl_T}{T}\right)=o(1)$. Thus, we just need to focus on $\sum_{n=1}^{k_T}\tilde{V}_n$ below. Using Proposition 2.6 in Fan and Yao (2003) and the condition on the α-mixing coefficient, we have

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

where C is a constant; $\alpha(\cdot)$ denotes the upper bound of the α-mixing coefficients provided in Assumption A and is achievable in the same way as Assumption A.4 of Chen et al. (2012). 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

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

which indicates the Feller condition is satisfied.

Moreover, by Cauchy-Schwarz inequality, we have

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

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

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

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

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

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

Proof of Corollary 2.1.1:

All we need to show is that $\hat{\sigma}_u^2 \to_P \sigma_u^2$. We start by writing

$$\hat{\sigma}_u^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{X}_{it}'(\beta(Z_{it})-\hat{\beta}(Z_{it}))+\tilde{u}_{it}+\gamma_{it})^2 = A_1+A_2+2A_3+2A_4+A_5,$$

where

$$A_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tilde{X}_{it}'(\beta(Z_{it})-\hat{\beta}(Z_{it})))^2, \quad A_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2,$$
$$A_3 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\beta(Z_{it})-\hat{\beta}(Z_{it}))\tilde{u}_{it}, \quad A_4 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\beta(Z_{it})-\hat{\beta}(Z_{it})+\tilde{u}_{it})\gamma_{it},$$
$$A_5 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{it}^2.$$

For A_1, we have

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

where the second inequality follows from Theorem 2.1.2. Thus, $A_1 \to_P 0$. Similarly, we can show that $A_3 \to_P 0$. By (1) of Lemma B.2, $A_2 \to_P \sigma_u^2$. Moreover, we have shown $A_4 = O_P(\|\hat{\lambda}\|^p)$ and $A_5 = O_P(\|\hat{\lambda}\|^{2p})$ in proving Theorem 2.1.1. Therefore, the result follows. ■

Note that if we replace Assumption A.3 by Assumption B, we can still show that Lemma B.2 holds by making some slight modifications to the proof. Specifically, for (2)-(3) of Lemma B.2, $\Sigma_{XX}(z)$ and $\Sigma_{XX\beta}(z)$ become $\Sigma_{XX}(\bar{z}) \cdot E[L(\tilde{Z}_{it}, \tilde{z}, \tilde{\lambda})]$ and $\Sigma_{XX\beta}(\bar{z}) \cdot E[L(\tilde{Z}_{it}, \tilde{z}, \tilde{\lambda})]$ for $\forall z \in \mathcal{D}$, respectvely; for (4)-(6) of Lemma B.2, $\Omega(z,\lambda)$, $p(z)$ and $\beta(z)$ reduce to $\Omega(\bar{z},\bar{\lambda})$, $p(\bar{z})$ and $\beta(\bar{z})$, respectively; (1) and (7)-(9) of Lemma B.2 hold without requiring any modification. Thus, when establishing asymptotic results for the irrelevant case in what follows, we will still use the basic results proved in Lemma B.2.

Proof of Lemma 2.2.1:

By Assumption B, $CV_0^*(\bar{\lambda})$ is uniquely minimized at $\bar{\lambda} = (\lambda_1, \ldots, \lambda_{r_1})' = 0$ and $\bar{\lambda}$ belongs to a compact set $[0,1]^{r_1}$. Also, $CV_0^*(\bar{\lambda})$ is continuous on $[0,1]^{r_1}$. 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. Note that λ_s for $s = r_1 + 1, \ldots, r$ associated with the irrelevant covariates get cancelled out in the asymptotic results, so they do not play any role when 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

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

where $\bar{\beta}_{it} = \beta(\bar{Z}_{it})$ and $\gamma_{it} = \frac{1}{T_{it}}\sum_{s=1}^{T} X_{is}'\left(\beta(\bar{Z}_{is}) - \beta(\bar{Z}_{it})\right)L_{is,it}^p$.

By result (1) of Lemma B.2, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \to_P \sigma_u^2$ uniformly in λ. Thus, we need only to focus on $CV_1(\lambda)$ and $CV_2(\lambda)$ below. Recall that $L(Z_{js}, z, \lambda) = L(\bar{Z}_{js}, \bar{z}, \bar{\lambda})L(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})$. As discussed before, Lemma B.2 holds if Assumption A.3 is replaced 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 $\forall z \in \mathcal{D}$,

$$
\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'L(Z_{js}, z, \lambda) \to_P \Sigma_{XX}(\bar{z}, \bar{\lambda}) \cdot E[L(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})] \tag{B.21}
$$

and

$$
\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'\beta(\bar{Z}_{js})L(Z_{js}, z, \lambda) \to_P \Sigma_{XX\beta}(\bar{z}, \bar{\lambda}) \cdot E[L(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})]. \tag{B.22}
$$

Note that $E[L(\tilde{Z}_{js}, \tilde{z}, \tilde{\lambda})]$ gets cancelled out after we substitute (B.21) and (B.22) into (B.8). We thus write

$$
\begin{aligned}
\hat{\beta}_{it} - \bar{\beta}_{it} &= \Sigma_{XX}^{-1}(\bar{Z}_{it}, \bar{\lambda})\Sigma_{XX\beta}(\bar{Z}_{it}, \bar{\lambda}) - \beta(\bar{Z}_{it}) + o_P(1) = \eta(\bar{Z}_{it}, \bar{\lambda}) - \beta(\bar{Z}_{it}) + o_P(1), \\
\gamma_{it} &= \Delta_{3\beta}(\bar{Z}_{it}, \bar{\lambda}) - \Delta_3(\bar{Z}_{it}, \bar{\lambda})'\beta(\bar{Z}_{it}) + o_P(1).
\end{aligned} \tag{B.23}
$$

By (B.23), $CV_1(\lambda)$ can be rewritten as

$$
\begin{aligned}
CV_1(\lambda) = CV_1^*(\bar{\lambda}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big((X_{it} - \Delta_3(\bar{Z}_{it}, \bar{\lambda}))'(\beta(\bar{Z}_{it}) - \eta(\bar{Z}_{it}, \bar{\lambda})) \\
+\Delta_{3\beta}(\bar{Z}_{it}, \bar{\lambda}) - \Delta_3(\bar{Z}_{it}, \bar{\lambda})'\beta(\bar{Z}_{it})\big)^2 + o_P(1).
\end{aligned}
$$

Similar to (B.7), we can obtain that $CV_1(\bar{\lambda}) \to_P CV_0^*(\bar{\lambda})$ for any given λ. Using the procedure same as the proof of result (2) of Lemma B.2, it is easy to show $CV_1(\bar{\lambda}) \to_P CV_0^*(\bar{\lambda})$ uniformly. Similar to the proof (7) of Lemma B.2, we can show that $CV_2(\lambda) = o_P(1)$ uniformly in λ.

With the above discussions, it is easy to see $CV(\lambda) \to_P CV_0^*(\bar{\lambda}) + \sigma_u^2$ uniformly in $\tilde{\lambda} \in \tilde{\mathcal{D}}$. Thus, all the conditions needed for Theorem 2.1 of Newey and McFadden (1994) are satisfied. Then the result follows. ■

Proof of Theorem 2.2.1:

1). Note that we have shown that $\hat{\lambda}_s = o_P(1)$ for $s = 1, \ldots, r_1$ in Lemma 2.2.1. Therefore, it is reasonable to assume that $\bar{\lambda}$ used in proving this theorem is sufficiently small and close to $0_{r_1 \times 1}$. For simplicity, define $\bar{1}_{itjs} = 1(\bar{Z}_{it} = \bar{Z}_{js})$ 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 $\bar{L}_{jsit,\bar{\lambda}} = L(\bar{Z}_{js}, \bar{Z}_{it}, \bar{\lambda})$ and $\tilde{L}_{jsit,\tilde{\lambda}} = L(\tilde{Z}_{js}, \tilde{Z}_{it}, \tilde{\lambda})$. Using the kernel function of Aitchison and Aitken (1976) and the expansion technique used in (B.12), we can write

$$L(Z_{js}, Z_{it}, \lambda) = \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} = \left(\bar{1}_{jsit} + \sum_{n=1}^{r_1} \lambda_n \bar{1}_{n,jsit} + O(\|\bar{\lambda}\|^2) \right) \tilde{L}_{jsit,\tilde{\lambda}}. \tag{B.24}$$

Before investigating the cross-validation criterion function, we further simplify $\hat{\beta}_{it} - \bar{\beta}_{it}$ as

$$\begin{aligned}
\hat{\beta}_{it} - \bar{\beta}_{it} &= \left(\frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \right)^{-1} \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' (\bar{\beta}_{js} - \bar{\beta}_{it}) \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \\
&\quad + \left(\frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{X}_{js}' \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \right)^{-1} \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} \tilde{X}_{js} \tilde{u}_{js} \bar{L}_{jsit,\bar{\lambda}} \tilde{L}_{jsit,\tilde{\lambda}} \\
&= \left(A_{1it} + A_{2it\lambda} + O_P(\|\bar{\lambda}\|^2) \right)^{-1} (B_{it} + C_{it}), \qquad \text{(B.25)}
\end{aligned}$$

where the term $O_P(\|\bar{\lambda}\|^2)$ in the last line follows from (B.24) and (4) of Lemma B.2; and

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

Applying a similar procedure to what has been used for proving (2) of Lemma B.2 to A_{1it} and $A_{2it\lambda}$, we obtain $A_{1it} = O_P(1)$ and $A_{2it\lambda} = O_P(\|\bar{\lambda}\|)$. Applying the same procedure as for B_{it}, 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 zero term follows from $(\bar{\beta}_{js} - \bar{\beta}_{it}) \bar{1}_{jsit} = 0$ and

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

Using a similar procedure to what has been used for proving (7) of Lemma B.2 to C_{it}, we obtain

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

where

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

Based on the above discussions, applying Lemma B.1 twice to the term on RHS of (B.25) gives

$$
\hat{\beta}_{it}-\bar{\beta}_{it} = \left(A_{1it}^{-1}-A_{1it}^{-1}A_{2it\lambda}A_{1it}^{-1}\right)\left(B_{2it\lambda}+C_{1it}+C_{2it\lambda}\right)+O_P\left(\frac{\|\bar{\lambda}\|^2}{\sqrt{NT}}\right)+O_P\left(\|\bar{\lambda}\|^3\right). \quad \text{(B.26)}
$$

Write

$$
\begin{aligned}
CV(\lambda) = {}& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\bar{\beta}_{it}-\hat{\beta}_{it})\right)^2+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}'(\bar{\beta}_{it}-\hat{\beta}_{it})\tilde{u}_{it}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}^2 \\
&+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'(\bar{\beta}_{it}-\hat{\beta}_{it})+\tilde{u}_{it}\right)\gamma_{it}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{it}^2 \\
\equiv{}& CV_1(\lambda)+CV_2(\lambda)+CV_3+CV_4(\lambda)+CV_5(\lambda),
\end{aligned}
$$

where $\gamma_{it}=\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}'\left(\beta(Z_{is})-\beta(Z_{it})\right)L_{is,it}^p$. In connection with the construction of γ_{it}, we are able to obtain that $CV_4(\lambda)=O_P\left(\|\bar{\lambda}\|^p\right)$ and $CV_5(\lambda)=O\left(\|\bar{\lambda}\|^{2p}\right)$. Replacing $\hat{\beta}_{it}-\bar{\beta}_{it}$ with (B.26) in $CV_1(\lambda)$ and $CV_2(\lambda)$ gives

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

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 $D_{3it}=\tilde{X}_{it}'A_{1it}^{-1}B_{2it\lambda}$.

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

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

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

Note 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 get cancelled out each other as in proving Theorem 2.1 in the presence of irrelevant covariates. 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

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

For simplicity, let $\Psi(\bar{Z}_{it})=p(\bar{Z}_{it})(\Sigma_X(\bar{Z}_{it})-\mu_X(\bar{Z}_{it})\mu_X(\bar{Z}_{it})')$. Firstly consider $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}D_{3it}^2$.

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

where the second equality follows from (2) of Lemma B.2, Assumption B and $B_{2it\lambda} = O_P(\|\bar{\lambda}\|^2)$; the third equality follows from a similar procedure as used for proving (2) of Lemma B.2 and Assumption B; the fifth equality follows from the first result of this theorem. Note that $E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]$ gets cancelled out. Therefore, the leading term on RHS of the above equation is unrelated with $\tilde{\lambda}$ and the remaining terms have an order of magnitude of $o_P\left(\frac{1}{NT}\right)$. 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

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

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

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 B.2, Assumption B and $C_{1it} = O_P\left(\frac{1}{\sqrt{NT}}\right)$.

Applying a similar procedure to what has been used for deriving $CV_1(\lambda)$ in Lemma 2.1.1, we can obtain

$$H_{1,NT} = \frac{1}{NT}C\cdot E\left[E[\tilde{L}_{jsit,\tilde{\lambda}}^2|\tilde{Z}_{it}]\cdot E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^{-2}\right] + o_P\left(\frac{1}{NT}\right), \tag{B.27}$$

where by the construction of $H_{1,NT}$ it is easy to know that C is a positive constant. Note that $E[\tilde{L}_{jsit,\tilde{\lambda}}^2|\tilde{Z}_{it}] \geq E[\tilde{L}_{jsit,\tilde{\lambda}}|\tilde{Z}_{it}]^2$, where the 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

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

$$= -\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}B_{2it,\lambda} - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{u}_{it}\tilde{X}_{it}'A_{1it}^{-1}C_{1it} + o_P\left(\frac{1}{NT}\right)$$
$$= H_{4,NT} + H_{5,NT} + o_P\left(\frac{1}{NT}\right).$$

Therefore,

$$CV(\lambda) = H_{1,NT} + H_{2,NT} + H_{3,NT} + H_{4,NT} + H_{5,NT} + o_P\left(\frac{1}{NT}\right), \tag{B.28}$$

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

Proof of Theorem 2.2.2:

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 $s = 1,\ldots,r_1$ is the estimate of λ_s by minimizing the CV criterion function; and λ_s for $s = r_1+1,\ldots,r$ is any arbitrary constant belonging to $[0,1]$.

Denote that $\hat{\bar{\lambda}} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_{r_1})'$, $\bar{1}_{\bar{Z}_{it},\bar{z}} = 1(\bar{Z}_{it} = \bar{z})$ 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 $n = 1,\ldots,r_1$. Thus, write

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

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

By the proof of Theorem 2.2.1, $A_{0,NT}^{-1} = O_P(1)$, $\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{u}_{js}\bar{L}_{js\bar{\lambda}}\tilde{L}_{it\tilde{\lambda}} = O_P\left(\frac{1}{\sqrt{NT}}\right)$ and $\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\gamma_{js}\bar{L}_{js\bar{\lambda}}\tilde{L}_{it\tilde{\lambda}} = O_P\left(\|\hat{\bar{\lambda}}\|^p\right)$. Thus, we need only to focus on the second term on the RHS of the above equation:

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

$$+O(\|\hat{\tilde{\lambda}}\|^2)\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\tilde{X}_{js}\tilde{X}_{js}'(\beta(\bar{Z}_{js})-\beta(\bar{z}))\tilde{L}_{it\tilde{\lambda}}$$

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

where the second equality follows from $(\beta(\bar{Z}_{js})-\beta(\bar{z}))\bar{1}_{\bar{Z}_{js},\bar{z}}=0$ and Theorem 2.2.1. The proof is then complete. ■

Proof of Theorem 2.3.1:

1). Let $\alpha_{NT}=\frac{1}{\sqrt{NT}}$ and U be an $(m\times q)$ matrix. We want to show that for any given $\epsilon>0$, there exists a large constant C such that

$$\liminf_{N}\Pr\left\{\inf_{\|U\|=C}Q_\tau\left(B_0+\alpha_{NT}U\right)>Q_\tau(B_0)\right\}\geq 1-\epsilon. \tag{B.29}$$

This implies with a probability of at least $1-\epsilon$ that there exists a local minimum in the ball $\{B_0+\alpha_{NT}U:\|U\|\leq C\}$. Hence, there exists a local minimizer such that $\|\hat{B}-B_0\|=O_P(\alpha_{NT})$. The above argument is in the same spirit as in the proofs for Theorem 1 of Fan and Li (2001) and Lemma A.1 of Wang and Xia (2009).

For notational simplicity, let U_j be the transpose of the j^{th} row of the matrix U with $j=1,\ldots,m$ and V_s be the s^{th} column of the matrix U with $s=1,\ldots,p$; and denote

$$e_j=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\left(\tilde{X}_{it}'\beta(\bar{Z}_{it})-\tilde{X}_{it}'\beta(\bar{z}^j)+\gamma_{it}+\tilde{u}_{it}\right)L(Z_{it},z^j,\hat{\lambda}),$$

where $\gamma_{it}=\frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}'\left(\beta(\bar{Z}_{is})-\beta(\bar{Z}_{it})\right)L^p_{is,it}$. By the proofs of Theorems 2.1.2 and 2.2.2, it is easy to know that $e_j=O_P(1)$ uniformly in j due to the fact that $\mathcal{D}$ is compact.

Then we write

$$\begin{aligned}
&Q_\tau(B_0+\alpha_{NT}U)-Q_\tau(B_0)\\
&=\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\beta(\bar{Z}_{it})+\gamma_{it}+\tilde{u}_{it}-\tilde{X}_{it}'\beta(\bar{z}^j)-\alpha_{NT}\tilde{X}_{it}'U_j\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&\quad+\sum_{s=1}^{q^*}\tau_s\|b_{0s}+\alpha_{NT}V_s\|+\sum_{s=q^*+1}^{q}\tau_s\|\alpha_{NT}V_s\|\\
&\quad-\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\beta(\bar{Z}_{it})+\gamma_{it}+\tilde{u}_{it}-\tilde{X}_{it}'\beta(\bar{z}^j)\right)^2L(Z_{it},z^j,\hat{\lambda})-\sum_{s=1}^{q^*}\tau_s\|b_{0s}\|\\
&=\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_{NT}\tilde{X}_{it}'U_j\right)^2L(Z_{it},z^j,\hat{\lambda})+\sum_{s=q^*+1}^{q}\tau_s\|\alpha_{NT}V_s\|+\sum_{s=1}^{q^*}\tau_s\left(\|b_{0s}+\alpha_{NT}V_s\|-\|b_{0s}\|\right)\\
&\quad-2\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{NT}U_j'\tilde{X}_{it}\left(\tilde{X}_{it}'\beta(\bar{Z}_{it})-\tilde{X}_{it}'\beta(\bar{z}^j)+\gamma_{it}+\tilde{u}_{it}\right)L(Z_{it},z^j,\hat{\lambda})\\
&\geq\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{NT}^2U_j'\tilde{X}_{it}\tilde{X}_{it}'U_jL(Z_{it},z^j,\hat{\lambda})+\sum_{s=1}^{q^*}\tau_s\left(\|b_{0s}+\alpha_NV_s\|-\|b_{0s}\|\right)
\end{aligned}$$

$$
\begin{aligned}
&-2\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\alpha_{NT}U_j'\tilde{X}_{it}\left(\tilde{X}_{it}'\beta(\bar{Z}_{it})-\tilde{X}_{it}'\beta(\bar{z}^j)+\gamma_{it}+\tilde{u}_{it}\right)L(Z_{it},z^j,\hat{\lambda})\\
\geq\ &\frac{\rho_1}{2}\sum_{j=1}^{m}\|U_j\|^2-2\sum_{j=1}^{m}U_j'e_j+\sum_{s=1}^{q^*}\tau_s\left(\|b_{0s}+\alpha_{NT}V_s\|-\|b_{0s}\|\right)\\
\geq\ &\frac{\rho_1}{2}\sum_{j=1}^{m}\|U_j\|^2-2\sum_{j=1}^{m}U_j'e_j-O(1)\sum_{s=1}^{q^*}\tau_s\frac{1}{\sqrt{NT}}\|V_s\|,
\end{aligned}
$$

where the second inequality follows from (2) of Lemma B.2 and Assumption C; and the third inequality follows from the Mean Value Theorem. Note that $\|U\|=C$, so we can further write

$$
\begin{aligned}
&Q_\tau(B_0+\alpha_{NT}U)-Q_\tau(B_0)\\
\geq\ &\frac{\rho_1}{2}\sum_{j=1}^{m}\|U_j\|^2-2\sum_{j=1}^{m}U_j'e_j-O(1)\sum_{s=1}^{q^*}\tau_s\frac{1}{\sqrt{NT}}\|V_s\|\\
\geq\ &\frac{\rho_1}{2}\sum_{j=1}^{m}\|U_j\|^2-2\left(\sum_{j=1}^{m}\|U_j\|^2\sum_{j=1}^{m}\|e_j\|^2\right)^{1/2}-O(1)\sum_{s=1}^{q^*}\tau_s\frac{1}{\sqrt{NT}}\|V_s\|\\
\geq\ &\frac{\rho_1}{2}C^2-2C\left(\sum_{j=1}^{m}\|e_j\|^2\right)^{1/2}-O(1)\frac{1}{\sqrt{NT}}\|\tau^*\|\left(\sum_{s=1}^{q^*}\|V_s\|^2\right)^{1/2}\\
=\ &\frac{\rho_1}{2}C^2-2C\left(\sum_{j=1}^{m}\|e_j\|^2\right)^{1/2}-O(1)C,
\end{aligned}
\tag{B.30}
$$

where $\frac{1}{\sqrt{NT}}\|\tau^*\|=O(1)$ by the condition given in this theorem and $\|e_j\|=O_P(1)$ uniformly in j. Note that $\frac{\rho_1}{2}C^2$ is a quadratic function in C while the remaining terms on RHS of (B.30) are linear in C. Since C can be sufficiently large, it is easy to know that RHS of (B.30) is positive with an arbitrary probability close to 1. The proof for (B.29) is now completed. ■

2). For simplicity, we show that $\Pr(\|\hat{b}_{\tau,q}\|=0)\to 1$ only. The proofs for $\hat{b}_{\tau,j}$ with $j=q^*+1,\ldots,q-1$ are the same. If $\|\hat{b}_{\tau,q}\|\neq 0$, $\hat{B}_\tau$ must satisfy the following equation

$$
0=\frac{\partial}{\partial b_q}Q_\tau(B)=A_1+A_2,\tag{B.31}
$$

where

$$
\begin{aligned}
A_1&=-\sum_{i=1}^{N}\sum_{t=1}^{T}2\tilde{X}_{it,q}\left((\tilde{Y}_{it}-\tilde{X}_{it}'\hat{\beta}_{\tau,1})L(Z_{it},z^1,\hat{\lambda}),\ldots,(\tilde{Y}_{it}-\tilde{X}_{it}'\hat{\beta}_{\tau,m})L(Z_{it},z^m,\hat{\lambda})\right)',\\
A_2&=\frac{\tau_q}{\|\hat{b}_{\tau,q}\|}\hat{b}_{\tau,q}.
\end{aligned}
$$

For $s=1,\ldots,m$, we can further write each element of A_1 as follows:

$$
\frac{1}{\sqrt{NT}}A_{1,s}=-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}2\tilde{X}_{it,q}\left(\tilde{X}_{it}'(\beta(\bar{Z}_{it})-\hat{\beta}_{\tau,s})+\gamma_{it}+\tilde{u}_{it}\right)
$$

$$
\begin{aligned}
&= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} 2\tilde{X}_{it,q}\tilde{X}_{it}'(\beta(\bar{Z}_{it}) - \hat{\beta}_{\tau,s})L(Z_{it}, z^s, \hat{\lambda}) \\
&\quad -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} 2\tilde{X}_{it,q}(\gamma_{it} + \tilde{u}_{it})L(Z_{it}, z^s, \hat{\lambda}) \\
&= -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} 2\tilde{X}_{it,q}\tilde{X}_{it}'(\beta(\bar{Z}_{it}) - \beta(\bar{z}^s))L(Z_{it}, z^s, \hat{\lambda}) \\
&\quad -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} 2\tilde{X}_{it,q}\tilde{X}_{it}'(\beta(\bar{z}^s) - \hat{\beta}_{\tau,s})L(Z_{it}, z^s, \hat{\lambda}) + O_P(1) = O_P(1),
\end{aligned}
$$

where the third equality follows from the proof of the first result of this theorem; and the fourth equality follows from Theorem 2.1.1 (or 2.2.1) and the first result of this theorem.

On the other hand, $\left\|\frac{1}{\sqrt{NT}}A_2\right\| \geq \frac{1}{\sqrt{NT}}\min_{s\in\{q^*+1,\dots,q\}}\tau_s \geq \omega_2$ by the condition given in the theorem, where ω_2 is sufficiently large. Therefore, $\Pr(\|A_1\| < \|A_2\|) \to 1$, which implies that, with a probability tending to 1, (B.31) does not hold. The above analysis implies that $\hat{b}_{\tau,q}$ must be located at a place where the objective function (2.7) is not differentiable with respect to b_q. Since equation (2.7) of the main file is not differentiable with respect to b_q only at the origin, we immediately obtain that $\Pr(\|\hat{b}_{\tau,q}\| = 0) \to 1$. In a similar fashion, we can show that $\Pr(\hat{b}_{\tau,j} = 0) \to 1$ with $j = q^*+1, \dots, q-1$. The proof is then complete. ■

Proof of Theorem 2.3.2:

By Theorem 2.3.1, we know that $\|\hat{b}_{\tau,s}\| = 0$ for $s = q^*+1, \dots, q$ with a probability tending to one. After some simple algebra, we can obtain the first derivative of $Q_\tau(B)$ with respect to β_j for $j = 1, \dots, m$. Then it is easy to know that $\hat{\beta}_{\tau,jU}$ must be the solution of the following equation:

$$
\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\left(\tilde{Y}_{it} - \tilde{X}_{itU}'\hat{\beta}_{\tau,jU}\right)L(Z_{it}, z^j, \hat{\lambda}) + \frac{1}{NT}D\hat{\beta}_{\tau,jU} = 0,
$$

where $\tilde{X}_{itU} = (\tilde{X}_{it,1}, \dots, \tilde{X}_{it,q^*})'$ and $D = \mathrm{diag}\Big(\tau_1\|\hat{b}_{\tau,1}\|^{-1}, \dots, \tau_{q^*}\|\hat{b}_{\tau,q^*}\|^{-1}\Big)$. It implies that $\hat{\beta}_{\tau,jU}$ must have the form

$$
\hat{\beta}_{\tau,jU} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{X}_{itU}'L(Z_{it}, z^j, \hat{\lambda}) + \frac{1}{2NT}D\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{Y}_{it}L(Z_{it}, z^j, \hat{\lambda}).
$$

In contrast, the oracle estimator has the following form:

$$
\left\|\hat{\beta}_{\tau,jU} - \hat{\beta}_{ora}(\bar{z}^j)\right\| \leq \left\|\Sigma_{NT}(z^j)\right\|\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{Y}_{it}L(Z_{it}, z^j, \hat{\lambda})\right\|, \tag{B.32}
$$

where

$$
\begin{aligned}
\Sigma_{NT}(z^j) &= \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{X}_{itU}'L(Z_{it}, z^j, \hat{\lambda}) + \frac{1}{2NT}D\right)^{-1} \\
&\quad - \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{X}_{itU}'L(Z_{it}, z^j, \hat{\lambda})\right)^{-1}.
\end{aligned}
$$

Since $\Sigma_{NT}(z^j)$ has finite dimensions, it is easy to know that the rate of $\left\|\Sigma_{NT}(z^j)\right\|$ converging to 0 is the same as

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{X}_{itU}'L(Z_{it},z^j,\hat{\lambda})+\frac{1}{2NT}D-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{X}_{itU}'L(Z_{it},z^j,\hat{\lambda})\right\|$$
$$=\left\|\frac{1}{2NT}D\right\|=O_P\left(\frac{\|\tau^*\|}{NT}\right).$$

Moreover, as with the proof of Theorem 2.1.1 (or 2.2.1), $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{itU}\tilde{Y}_{it}L(Z_{it},z^j,\hat{\lambda})=O_P(1)$. Therefore, for $j=1,\ldots,m$, $\left\|\hat{\beta}_{\tau,jU}-\hat{\beta}_{ora}(\bar{z}^j)\right\|=O_P\left(\frac{\|\tau^*\|}{NT}\right)$. The proof is now completed. ■

Proof of Theorem 2.3.3:

1). For an arbitrary model S, we say it is under-fitted if it misses at least one variable with a nonzero coefficient (i.e. $S\subset\mathcal{A}^c$ but $\mathcal{A}^c\neq S$); it is over fitted if S covers all relevant variables but also includes at least one redundant regressor (i.e. $\mathcal{A}^c\subset S$ but $\mathcal{A}^c\neq S$). Then, depending on whether the model S is under fitted, correctly fitted, or over fitted, we create three mutually exclusive sets $A^-=\{\tilde{\tau}\in\mathbb{R}:S\subset\mathcal{A}^c,S\neq\mathcal{A}^c\}$, $A^0=\{\tilde{\tau}\in\mathbb{R}:S=\mathcal{A}^c\}$ and $A^+=\{\tilde{\tau}\in\mathbb{R}:S\supset\mathcal{A}^c,S\neq\mathcal{A}^c\}$. Suppose that $\tilde{\beta}_j$ for $j=1,\ldots,m$ are unregularized estimates and there is a sequence $\{\hat{\tau}_{NT}\}$ that ensures (2.15) of the main file satisfies the conditions required by Theorem 2.3.1 (e.g. those used in Monte Carlo study).

Case 1: In this case, we consider under-fitted models, where $S\subset\mathcal{A}^c$ but $\mathcal{A}^c\neq S$. Without losing generality, we assume that only one variable is missing, so we assume that the first q^*-1 elements of $\hat{\beta}_{\tilde{\tau},j}$ are obtained from the under-fitted model and the remaining $q-q^*+1$ elements of $\hat{\beta}_{\tilde{\tau},j}$ are 0.

We then write

$$\begin{aligned}
RSS_{\tilde{\tau}} &= \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\hat{\beta}_{\tilde{\tau},j}\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&= \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\tilde{\beta}_j+\tilde{X}_{it}'\tilde{\beta}_j-\tilde{X}_{it}'\hat{\beta}_{\tilde{\tau},j}\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&= \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\tilde{\beta}_j\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&\quad+\frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\tilde{\beta}_j-\tilde{X}_{it}'\hat{\beta}_{\tilde{\tau},j}\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&\quad+\frac{2}{N}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\beta}_j-\hat{\beta}_{\tilde{\tau},j}\right)'\tilde{X}_{it}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\tilde{\beta}_j\right)L(Z_{it},z^j,\hat{\lambda})\\
&= \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it}-\tilde{X}_{it}'\tilde{\beta}_j\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&\quad+\frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{X}_{it}'\tilde{\beta}_j-\tilde{X}_{it}'\hat{\beta}_{\tilde{\tau},j}\right)^2L(Z_{it},z^j,\hat{\lambda})\\
&\equiv RSS^*+R_{2\tilde{\tau}},
\end{aligned}$$

where the fourth equality is due to the construction of the unregularized estimators.

We now consider $R_{2\tilde{\tau}}$ and write

$$\begin{aligned}
R_{2\tilde{\tau}} &= \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\tau},j}\right)' \tilde{X}_{it}\tilde{X}_{it}' L(Z_{it}, z^j, \hat{\lambda})\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\tau},j}\right) \\
&= \sum_{j=1}^{m}\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\tau},j}\right)' \Sigma_1(z^j)\left(\tilde{\beta}_j - \hat{\beta}_{\tilde{\tau},j}\right) + o_P(1) \\
&\geq \sum_{j=1}^{m}\rho_{\min}(\Sigma_1(z^j))\left\|\tilde{\beta}_j - \hat{\beta}_{\tilde{\tau},j}\right\|^2 + o_P(1) \\
&= O(1)\sum_{j=1}^{m}\left\|\tilde{\beta}_j - \hat{\beta}_{\tilde{\tau},j}\right\|^2 + o_P(1) \geq O(1)\sum_{j=1}^{m}\tilde{\beta}_{j,q^*}^2 + o_P(1),
\end{aligned}$$

where $\Sigma_1(z^j) = \Sigma_{XX}(\bar{z})E[\tilde{L}(\tilde{Z}_{it}, \tilde{z}, \hat{\tilde{\lambda}})|\hat{\tilde{\lambda}}]$; $\rho_{\min}(\Sigma_1(z^j))$ denotes the minimum eigenvalue of $\Sigma_1(z^j)$; $\tilde{\beta}_{j,q^*}$ denotes the q^{*th} element of $\tilde{\beta}_j$; the second equality follows from (2) of Lemma B.2 of the Appendix and Theorem 2.1.1 (or 2.2.1); and the first inequality follows from Assumption C.2.

Similarly, we can obtain that $RSS_{\hat{\tau}_{NT}} \equiv RSS^* + R_{2\hat{\tau}_{NT}}$, where

$$\begin{aligned}
R_{2\hat{\tau}_{NT}} &= \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\tau}_{NT},j}\right)' \tilde{X}_{it}\tilde{X}_{it}' L(Z_{it}, z^j, \hat{\lambda})\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\tau}_{NT},j}\right) \\
&= \sum_{j=1}^{m}\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\tau}_{NT},j}\right)' \Sigma_1(z^j)\left(\tilde{\beta}_j - \hat{\beta}_{\hat{\tau}_{NT},j}\right) + o_P(1) \\
&\leq \sum_{j=1}^{m}\rho_{\max}(\Sigma_1(z^j))\left\|\tilde{\beta}_j - \hat{\beta}_{\hat{\tau}_{NT},j}\right\|^2 + o_P(1) \\
&\leq O(1)\sum_{j=1}^{m}\left\|\tilde{\beta}_j - \hat{\beta}_{\hat{\tau}_{NT},j}\right\|^2 + o_P(1) \\
&\leq O(1)\sum_{j=1}^{m}\left\|\tilde{\beta}_j - \beta(\bar{z}^j)\right\|^2 + O(1)\sum_{j=1}^{m}\left\|\beta(\bar{z}^j) - \hat{\beta}_{\hat{\tau}_{NT},j}\right\|^2 = o_P(1),
\end{aligned}$$

where $\rho_{\max}(\Sigma_1(z^j))$ denotes the maximum eigenvalue of $\Sigma_1(z^j)$; the second equality follows from (2) of Lemma B.2 of the Appendix and Theorem 2.1.1 (or 2.2.1); the second inequality follows from Assumption C.2; and the last equality follows from Theorem 2.3.1 and the fact that both $\tilde{\beta}_j$ and $\hat{\beta}_{\hat{\tau}_{NT},j}$ are regularized estimators.

Note that by (1) of Lemma B.2 we can obtain that $RSS^* \to_P \sum_{j=1}^{m}\Pr(\bar{z}^j)\sigma_u^2$. Based on the analysis on $R_{2\tilde{\tau}}$ and $R_{2\hat{\tau}_{NT}}$, we then can further conclude that

$$\Pr\left(\inf_{\tilde{\tau}\in A^-} BIC_{\tilde{\tau}} > BIC_{\hat{\tau}_{NT}}\right) \to 1.$$

Case 2: In this case, we consider over-fitted models, where $S \supset \mathcal{A}^c$ but $\mathcal{A}^c \neq S$. Consider $\forall\tilde{\tau} \in A^+$ and recall that $\hat{B}_{\tilde{\tau}}$ determines $S_{\tilde{\tau}}$. Under such a model $S_{\tilde{\tau}}$, we can define another unpenalized estimator $\breve{B}_{\tilde{\tau}}$ as

$$\breve{B}_{\tilde{\tau}} = \underset{\beta_1,\ldots,\beta_m}{\operatorname{argmin}}\frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it} - \tilde{X}_{it}'\beta_j\right)^2 L(Z_{it}, z^j, \hat{\lambda}),$$

where, for $j = 1, \ldots, m$, $\|\beta_{j,s}\| = 0$ with $\forall s \notin S_{\tilde{\tau}}$ and $\beta_{j,s}$ denotes the s^{th} element of β_j. In other words, $\check{B}_{\tilde{\tau}} = (\check{\beta}_1, \ldots, \check{\beta}_m)'$ is the unregularized estimator under the model determined by $\hat{B}_{\tilde{\tau}}$. By definition, we obtain immediately that $RRS_{\tilde{\tau}} \geq RRS_{S_{\tilde{\tau}}}$, where

$$RRS_{S_{\tilde{\tau}}} = \frac{1}{NT}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{Y}_{it} - \tilde{X}_{it}'\check{\beta}_j\right)^2 L(Z_{it}, z^j, \hat{\lambda}).$$

It follows that

$$\begin{aligned}
&\ln RRS_{\tilde{\tau}} - \ln RRS^* \geq \ln RRS_{S_{\tilde{\tau}}} - \ln RRS^* \\
&= \ln\left\{\frac{RRS^*}{RRS^*} + \frac{1}{NT\cdot RRS^*}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\beta}_j - \check{\beta}_j\right)'\tilde{X}_{it}\tilde{X}_{it}'L(Z_{it}, z^j, \hat{\lambda})\left(\tilde{\beta}_j - \check{\beta}_j\right)\right\} \\
&\geq -\frac{O(1)}{NT\cdot RRS^*}\sum_{j=1}^{m}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\beta}_j - \check{\beta}_j\right)'\tilde{X}_{it}\tilde{X}_{it}'L(Z_{it}, z^j, \hat{\lambda})\left(\tilde{\beta}_j - \check{\beta}_j\right) \\
&\geq -\frac{O_P(1)}{RRS^*}\sum_{j=1}^{m}\rho_{max}(\Sigma_1(z^j))\left\|\tilde{\beta}_j - \check{\beta}_j\right\|^2 \\
&\geq -\frac{O_P(1)}{RRS^*}\sum_{j=1}^{m}\rho_{max}(\Sigma_1(z^j))\left\|\tilde{\beta}_j - \beta(\bar{z}^j)\right\|^2 - \frac{O_P(1)}{RRS^*}\sum_{j=1}^{m}\rho_{max}(\Sigma_1(z^j))\left\|\beta(\bar{z}^j) - \check{\beta}_j\right\|^2 \\
&\geq -\left|O_P\left(\frac{1}{NT}\right)\right|,
\end{aligned}$$

where $\tilde{\beta}_j$ for $j = 1, \ldots, m$ are unregularized estimators as those used in **Case 1**; the second inequality follows from (2) of Lemma B.2 and Theorem 2.1.1 (or 2.2.1); and the fourth inequality follows from Theorem 2.3.1.

Similarly, we can obtain that $\ln RRS_{\hat{\tau}_{NT}} - \ln RRS^* = O_P\left(\frac{1}{NT}\right)$. Thus, we obtain

$$\ln RRS_{\tilde{\tau}} - \ln RRS_{\hat{\tau}_{NT}} \geq -\left|O_P\left(\frac{1}{NT}\right)\right|.$$

We then write

$$\inf_{\tilde{\tau}\in A^+} BIC_{\tilde{\tau}} - BIC_{\hat{\tau}_{NT}} = \ln RRS_{\tilde{\tau}} - \ln RRS_{\hat{\tau}_{NT}} + (df_{\tilde{\tau}} - df_{\hat{\tau}_{NT}})\frac{\ln(NT)}{NT}.$$

By Theorem 2.3.1, we know that $\Pr(df_{\hat{\tau}_{NT}} \to q^*) = 1$. Since $\tilde{\tau} \in A^+$, we must have that $\Pr(df_{\tilde{\tau}} \geq q^* + 1) \to 1$. Then it is clear that

$$\Pr\left(\inf_{\tilde{\tau}\in A^+} BIC_{\tilde{\tau}} > BIC_{\hat{\tau}_{NT}}\right) \to 1.$$

Combining Cases 1 and 2, we obtain that $\Pr\left(\inf_{\tilde{\tau}\in A^-\cup A^+} BIC_{\tilde{\tau}} > BIC_{\hat{\tau}_{NT}}\right) \to 1$, which in turn implies $\Pr\left(S_{\hat{\tilde{\tau}}} \to \mathcal{A}^c\right) = 1$. The proof is completed.

2)-3). The second and third results of this theorem follow by noting that setting $\tilde{\tau}$ to a large constant satisfies all the conditions required by Theorem 2.3.2 and the first result of this theorem. Thus, we have

$$\hat{\beta}_{\hat{\hat{\tau}},jU} - \beta_U(\bar{z}^j) = \hat{\beta}_{ora}(\bar{z}^j) - \beta_U(\bar{z}^j) + O_P\left(\frac{1}{NT}\right).$$

Then the results follow from Theorem 2.1.2 and Theorem 2.2.2 immediately. ■

References

Aitchison, J. and Aitken, C. (1976), 'Multivariate binary discrimination by the kernel method', *Biometrika* **63**(3), 413–420.

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

Chen, J., Gao, J. and Li, D. (2012), 'A new diagnostic test for cross–section uncorrelatedness in nonparametric panel data models', *Econometric Theory* **28**(5), 1–20.

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360.

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

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

Hunter, D. R. and Li, R. (2005), 'Variable selection using mm algorithms', *The Annals of Statistics* **189**, 1617–1642.

Li, Q., Ouyang, D. and Racine, J. S. (2013), 'Categorical semiparametrics varying–coefficient models', *Journal of Applied Econometrics* **28**(4), 551–579.

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

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Su, L. and Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* **169**(1), 34–47.

Wang, H. and Xia, Y. (2009), 'Shrinkage estimation of the varying coefficient', *Journal of the American Statistical Association* **104**(486), 747–757.