Department of Econometrics and Business Statistics

Time-Varying Income Elasticities of Healthcare Expenditure for the OECD and Eurozone

Isabel Casas, Jiti Gao, Bin Peng and Shangyu Xie

December 2019

Time-Varying Income Elasticities of Healthcare Expenditure for the OECD and Eurozone

Isabel Casas[1,2], Jiti Gao[3], Bin Peng[4], and Shangyu Xie[5]

November 15, 2019

Abstract

Income elasticity dynamics of health expenditure is considered for the OECD and Eurozone over the period 1995-2014. Motivated by some modelling challenges, this paper studies a class of non-linear cointegration panel data models, controlling for cross-section dependence and certain endogeneity. Using the corresponding methods, our empirical analyses show a slight increase in the income elasticity of the healthcare expenditure over the years, but still with values under 1, meaning that healthcare is not a luxury good in the OECD and Eurozone.

Keywords: Health expenditure, Income elasticity, Nonparametric kernel smoothing, Nonstationarity
JEL Classification: C14, C23, H51

1. Department of Business and Economics, The University of Southern Denmark, Denmark. Email: icasas@sam.sdu.dk

2. Basque Centre for Applied Mathematics, Spain.

3. Department of Econometrics and Business Statistics, Monash University, Caulfield, VIC, Australia. Email: Jiti.Gao@monash.edu

4. Department of Economics, Deakin University, Burwood, VIC, Australia. Email: bin.peng@deakin.edu.au

5. RCAF and School of Banking and Finance, University of International Business and Economics, China. Email: xshyu@amss.ac.cn

1 Introduction

The income elasticity of healthcare expenditure is defined as the percentage change in healthcare expenditure in response to the percentage change in income per capita. If this elasticity is greater than one, then healthcare expenditure grows faster than income, as luxury goods do, and is driven by market forces alone (Culyer, 1988). However, significant governmental funding of health systems produces elasticities lower than one, typical of necessary goods. On this subject, there is growing literature on modelling and assessing the sustainability of health systems in some countries (e.g., Di Matteo and Di Matteo, 1998; Hartwig, 2008). Take Figure A.1 of this study as an example, in which a few facts emerge. First, there is a steady increase of average healthcare expenditure per capita during the past two decades in the OECD and Eurozone, and the rate of increase falls slightly after the Global Financial Crisis. A report of the World Health Organization on the effects of the Global Financial Crisis in European health systems (Mladovsky et al., 2012) concludes that the response has been heterogeneous among European countries: some countries have made their health system more efficient, others have extended health benefits to ensure access for low-income groups, while others have cut investment and increased patient charges. Clearly, the Global Financial Crisis has been an economic shock that has trigged the implementation of new health policies seeking to reduce the sensitivity of health systems to these economic shocks and to make them less dependent on public revenues. Second, there are clear nonstationary trends in Figure A.1, which should be accounted for by economic/econometric models. Third, certain cointegration/endogeneity should be allowed for.

Although the relevant literature has agreed that the sustainability of the healthcare cost can be more or less assessed by measuring income elasticity (e.g., Jones and Wildman, 2008; Hauck and Zhang, 2016), most earlier works use parametric models, assuming that regressors are stationary over time, and assuming exogeneity between regressors and error terms. Among the limited literature (e.g., Gerdtham and Löthgren, 2000; Okunade and Murthy, 2002; Jewel et al., 2003), few studies in this field have accounted for the nonstationarity of variables, the heterogeneous income elasticity over time, and certain endogeneity simultaneously.

That being said, we aim to capture the features associated with healthcare expenditure mentioned above by investigating some semiparametric panel data models. Panel data analyses have received considerable attention during the past two decades due to the suitability for a wide number of applied disciplines, e.g., economics, finance, and biology (Arellano, 2003; Hsiao, 2014). In addition, semi/nonparametric panel data models have

seen their popularity increase because of the flexibility of the unknown functional forms, e.g., Connor et al. (2012), Su and Jin (2012), Fan et al. (2016), just to name a few. Given that time trends are the dominant characteristic in many economic, financial, and climate datasets, it is thus worth considering some semi/nonparametric panel data models including different types of time trends. An excellent review on the challenges that arise with time trends can be seen in Phillips (2001).

In this article we study the income elasticity of healthcare expenditure specifically using time-varying coefficient panel data models with nonstationarity and endogeneity. On methodology, we first extend Phillips et al. (2017) to the panel data framework, and establish a biased corrected estimator to tackle certain endogeneity while accounting for time-varying marginal effects, and nonstationarity of the regressors. We then discuss an extension on nonstationary panel data models with interactive fixed effects. For both cases (with and without interactive fixed effects) we show that the degeneracy issue mentioned in Phillips et al. (2017) is no longer a problem when panel data get involved. Some other relevant theoretical works include, but are not limited to, Cai and Li (2008), Sun et al. (2009), Gao et al. (2019), etc. On empirical study, our results, which are based on 34 countries in the OECD and 20 counties in the Eurozone from 1995 to 2014, suggest that the income elasticities increase slightly over time. Our study shows that the country demographic structure is also related with its healthcare expenditure. The price of healthcare increases with ageing population and decreases as the rate of population under 15 years old increases. A monotonic decreasing relationship appears between the healthcare expenditure and the proportion of public funding dedicated to healthcare. This supports the positive effect of the new healthcare policies triggered by the Global Financial Crisis in the Eurozone. These policies aim at making the Eurozone health systems less dependent upon government funding and macro-economic shocks, and they seem to be working in the right direction.

The rest of this paper is organized as follows. The methodology and the associated asymptotic results are provided in Section 2. Section 3 presents the empirical study. Section 4 concludes. Appendix A further explains two technical issues, and presents the list of countries studied in this paper. To conserve space, Appendix B collects the necessary technical lemmas, mathematical proofs and some simulation studies.

Before proceeding we introduce some notations that will be used repeatedly below. $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix; $\|\cdot\|_{\mathrm{sp}}$ defines the spectral norm of a matrix; $\Longrightarrow$ denotes convergence in distribution; $\lfloor x \rfloor$ denotes the largest integer less than or equal to x; $K_h(u) = K(\frac{u}{h})$, where $K(\cdot)$ and h represent a kernel function and a bandwidth of the kernel method, respectively; i_T and I_T stand for

a $T \times 1$ vector of ones and a $T \times T$ identity matrix, respectively; $M_W = I_T - P_W$ denotes the orthogonal projection matrix generated by matrix W, where $P_W = W(W'W)^{-1}W'$, and W is a $T \times q$ matrix with rank q.

2 Methodology

We start by considering the following panel data model

$$Y_{it} = X'_{it}\beta_0(\tau_t) + \alpha_i + u_{it}, \tag{2.1}$$

where $\{(Y_{it}, X_{it}) \,|\, i = 1, \ldots, N, \, t = 1, \ldots, T\}$ are observable. Moreover, $\tau_t = t/T$, $X_{it} = X_{i,t-1} + \nu_{it}$ is a $d \times 1$ integrated process, $\beta_0(\cdot) = (\beta_{01}(\cdot), \ldots, \beta_{0d}(\cdot))'$ is the coefficient function, and α_i is the individual effect. We adopt a linear process for ν_{it} and u_{it}, i.e., $(\nu'_{it}, u_{it})' = \sum_{j=0}^{\infty}(\Phi_j, \psi_j)'\varepsilon_{i,t-j}$, where ε_{it} is a $(d+1) \times 1$ sequence of random vectors, and $(\Phi_j, \psi_j)'$ is $(d+1) \times (d+1)$ and is partitioned conformable to $(\nu'_{it}, u_{it})'$. Throughout this study, suppose $\sum_{i=1}^{N}\alpha_i = 0$ for the purpose of identification. The model (2.1) captures potential drifts in the relationship between Y_{it} and X_{it} over time, and allows for certain endogeneity through the structure of $(\nu'_{it}, u_{it})'$, which is especially useful for time series data over long horizons, because economic mechanisms are likely to evolve with institutional changes, regulatory conditions, or financial conditions.

2.1 Estimation and Asymptotic Properties

Having introduced the model (2.1), we define the estimate of $\beta_0(\delta)$ for $\forall \delta \in (0, 1)$ as follows:

$$\widehat{\beta}(\delta) = (X'W^*_\delta X)^{-1}X'W^*_\delta Y, \tag{2.2}$$

where $Y = (Y_{11}, \ldots, Y_{1T}, \ldots, Y_{N1}, \ldots, Y_{NT})'$, X is defined similar to Y, $W^*_\delta = D'_\delta W_\delta D_\delta$, $W_\delta = I_N \otimes \text{diag}\{K_h(\tau_1 - \delta), \ldots, K_h(\tau_T - \delta)\}$, $D_\delta = I_{NT} - D(D'W_\delta D)^{-1}D'W_\delta$, and $D = (-i_{N-1}, I_{N-1})' \otimes i_T$. To conserve space, the reason why (2.2) is considered is explained in Appendix A.1 of this paper. It is worth pointing out that under a panel data setting the degeneracy issue mentioned in Phillips et al. (2017) is no longer a problem for (2.2). The detailed explanation is provided in Appendix A.2.

Before stating the asymptotic properties of (2.2), we provide the necessary assumptions with discussions.

Assumption 1. *Let $K(\cdot)$ be symmetric and defined on $[-1, 1]$, and satisfy $\int_{-1}^{1} K(w)dw = 1$, $\sup_{w \in [0,1]} |K^{(1)}(w)| < \infty$, and $\int_{-1}^{1} |w|K(w)dw < \infty$.*

Assumption 2. *Suppose that each element of $\beta_0(\tau)$ is second order continuously differentiable on $[0, 1]$.*

Assumption 3. *Let $\Phi = \sum_{j=0}^{\infty} \Phi_j$ and $\psi = \sum_{j=0}^{\infty} \psi_j$. Suppose that $\max_{i \geq 1} E\|X_{i0}\|^4 < \infty$ and $\sum_{j=0}^{\infty} j\|(\Phi_j, \psi_j)\| < \infty$. Moreover, suppose that $\{\varepsilon_{it}\}$ is independent and identically distributed (i.i.d.) over i and t, $E[\varepsilon_{it}] = 0$, $E[\varepsilon_{it}\varepsilon'_{it}] = \Lambda > 0$, and $E\|\varepsilon_{it}\|^{4+\gamma_0} < \infty$ for some $\gamma_0 > 0$.*

Assumption 4. *Suppose that as $(N,T) \to (\infty, \infty)$, $h \to 0$ and $\min\{T, N\} \cdot h \to \infty$. In addition, $\lim_{(N,T)\to(\infty,\infty)} NT^2h^5 < \infty$.*

Assumption 1 imposes some standard conditions on the kernel function. Assumption 2 requires certain smoothness of the coefficient function. Both assumptions are standard in the literature of nonparametric regression (e.g., Cai and Li, 2008). Assumption 3 extends the requirements on nonstationary time series of Phillips et al. (2017) to the panel data setting. Assumption 4 restricts the rate of N and T diverging to ∞.

With these conditions in hand, we establish the first asymptotic result in the next theorem.

Theorem 2.1. *Suppose that Assumptions 1-4 are satisfied. For $\forall \delta \in (0, 1)$, as $(N, T) \to (\infty, \infty)$, $\sqrt{NT^2h} \left\{ \widehat{\beta}(\delta) - \beta_0(\delta) - \Delta_1 - O_P(h^2) \right\} \implies N(0, \frac{\mathcal{C}_*}{(1-\mathcal{C}_K)^2} \Sigma_\nu^{-1} \Sigma_{\nu,u} \Sigma_\nu^{-1})$, where Δ_1, $\mathcal{C}_*$, and $\mathcal{C}_K$ are defined in (A.2) of Appendix A, $\Sigma_\nu = \Phi'\Lambda\Phi$ and $\Sigma_{\nu,u} = \psi'\Lambda\psi\Phi'\Lambda\Phi$.*

The bias term $O_P(h^2)$ comes from the use of kernel method, and can be eliminated if one is willing to impose an assumption $N^{1/2}Th^3 \to 0$. In the following, we pay particular attention to the bias term Δ_1, which exists due to the correlation between ν_{it} and u_{it}.

We define some variables before continuing. For $\alpha \in (0, 1)$, let $\widehat{\alpha}_i(\delta)$ be yielded by plugging $\widehat{\beta}(\delta)$ in (A.4), and let $\widehat{u}_{it} = y_{it} - X'_{it}\widehat{\beta}(t/T) - \widehat{\alpha}_i$ with $\widehat{\alpha}_i = \frac{1}{\overline{\tau}-\underline{\tau}} \sum_{t=\underline{\tau}+1}^{\overline{\tau}} \widehat{\alpha}_i(\tau_t)$, where $\underline{\tau} = \lfloor \tau^*T \rfloor$, $\overline{\tau} = \lfloor (1-\tau^*)T \rfloor$, and $\tau^* \in (0, \frac{1}{2})$ is a self-selected positive constant. The use of τ^* is to avoid the boundary effects of the nonparametric kernel method. Let $\widehat{\omega}_{it} = (\nu'_{it}, \widehat{u}_{it})'$ in which the value of ν_{it} can be easily obtained by $\nu_{it} = X_{it} - X_{i,t-1}$, and construct the estimated autocovariances by $\widehat{\Gamma}_{i,\omega}(j) = \frac{1}{\overline{\tau}-\underline{\tau}} \sum_{t=\underline{\tau}+1}^{\overline{\tau}} \widehat{\omega}_{i,t-j}\widehat{\omega}'_{it}$, where $j = 0, 1, \cdots, l_T$ ($<T$), and l_T is a lag truncation number satisfying certain conditions to be specified below. These autocovariances give $\widehat{\Delta}_\omega = \frac{1}{N} \sum_{i=1}^{N} \widehat{\Delta}_{i,\omega}$ with $\widehat{\Delta}_{i,\omega} = \sum_{j=0}^{l_T} W\left(\frac{j}{l_T}\right) \widehat{\Gamma}_{i,\omega}(j)$, where $W(\cdot)$ is a lag kernel as in Andrews (1991).

We can then estimate the bias term as follows.

$$\widehat{\Delta}_{\nu u} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\Delta}_{i,\nu u} \quad \text{and} \quad \widehat{\Delta}_{\nu\bar{u}}^{t}(\delta) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\Delta}_{i,\nu\bar{u}}^{t}(\delta), \tag{2.3}$$

where $\widehat{\Delta}_{i,\nu u}$ is the first d elements of the last column of $\widehat{\Delta}_{i,\omega}$, and

$$\widehat{\Delta}_{i,\nu\bar{u}}^{t}(\delta) = \left(\sum_{s=1}^{T} K_h(\tau_s - \delta)\right)^{-1}\sum_{l=1}^{t}\sum_{s=1}^{T} K_h(\tau_s - \delta)W\left(\frac{s-l}{l_T}\right)\widehat{\Gamma}_{i,\nu u}(s-l)I(|s-l| \leq l_T),$$
$$\tag{2.4}$$

in which $\widehat{\Gamma}_{i,\nu u}(\cdot)$ is the first d elements of the last column of $\widehat{\Gamma}_{i,\omega}(\cdot)$. Finally, the bias corrected estimator of $\beta_0(\delta)$ is as follows:

$$\widehat{\beta}_{*}(\delta) = (X'W_{\delta}^{*}X)^{-1}\left(X'W_{\delta}^{*}Y - N\sum_{t=1}^{T} K_h(\tau_t - \delta)(\widehat{\Delta}_{\nu u} - \widehat{\Delta}_{\nu\bar{u}}^{t}(\delta))\right). \tag{2.5}$$

Apparently, the choice of l_T is important, so we impose the next assumption.

Assumption 5. *The lag kernel $W(\cdot)$ satisfies that (i) $W(0) = 1$ and $\int_{-1}^{1} W^2(x)dx < \infty$; (ii) $W(-x) = W(x)$; and (iii) Parzen's exponent for $q \in [0,\infty)$ such that $k_q = \lim_{x\to 0}\frac{1-W(x)}{|x|^q} < \infty$ for $q > 1/2$. Moreover, $l_T \to \infty$, $l_T h/T \to 0$, and $Nh/l_T^{2q} \to 0$.*

With the extra assumption in hand, the next theorem follows.

Theorem 2.2. *Suppose that Assumptions 1-5 are satisfied. For $\forall\delta \in (0,1)$, as $(N,T) \to (\infty,\infty)$, $\sqrt{NT^2h}\{\widehat{\beta}_{*}(\delta) - \beta_0(\delta) + O_P(h^2)\} \Longrightarrow N(0, \frac{\mathcal{C}_{K^*}}{(1-\mathcal{C}_K)^2}\Sigma_{\nu}^{-1}\Sigma_{\nu,u}\Sigma_{\nu}^{-1})$.*

The proof of Theorem 2.2 is given in the online supplementary material.

2.2 Extension — Time-Varying Panel Data Models with Interactive Fixed Effects

In this subsection, we extend (2.1) to a panel data model with interactive fixed effects. Panel data models with interactive fixed effects models have been well studied in theory and applied since Pesaran (2006) and Bai (2009). Although there are a few nonparametric based methodologies proposed and investigated in the literature of factor models (e.g., Su and Jin, 2012; Fan et al., 2016; Dong et al., 2019), majority (if not all) of these studies adopt the sieve estimation method. To the best knowledge of the authors, only Su and Wang (2017) use the kernel method to study a time-varying factor model which does not include any regressors. As nonparametric kernel regression is one of the most commonly studied methods of nonparametric regression, it is then important to extend Pesaran

(2006) and Bai (2009) to a nonparametric kernel setting, and investigate how a kernel based approach can be applied.

Specifically, we consider the next model, and employ the kernel method to recover the unknown coefficient function.

$$Y_{it} = X'_{it}\beta_0(\tau_t) + \lambda'_{0i}f_{0t} + u_{it}, \tag{2.6}$$

where u_{it} is an error term, $X_{it} = X_{i,t-1} + \nu_{it}$, ν_{it} follows a linear process $\nu_{it} = \sum_{j=0}^{\infty} \Upsilon_j \epsilon_{i,t-j}$, and ϵ_{it} stands for a $d \times 1$ random vector. Moreover, λ_{0i} and f_{0t} are $r \times 1$ unknown factor loading and factor respectively, and both r and d are fixed and known. Following our study on (2.1), we try to allow for certain correlation between X_{it} and u_{it} as well as correlation between $\lambda'_{0i}f_{0t}$ and u_{it} in the following development.

The next assumption provides some basic requirements on (2.6).

Assumption 6.

1. Let $\Upsilon = \sum_{j=0}^{\infty} \Upsilon_j$. Suppose $\max_{i \geq 1} E\|X_{i0}\|^4 < \infty$ and $\sum_{j=0}^{\infty} j\|\Upsilon_j\| < \infty$. Moreover, suppose that $\{\epsilon_{it}\}$ is i.i.d. over i and t, $E[\epsilon_{it}] = 0$, $E[\epsilon_{it}\epsilon'_{it}] = \Lambda_\epsilon > 0$, and $E\|\epsilon_{it}\|^{4+\gamma_0} < \infty$ for some $\gamma_0 > 0$.

2. For $\forall \delta \in (0,1)$, $\|L_{1NT}\| = o_P(1)$ and $\|L_{2NT}\| = O_P(1)$, where

$$L_{1NT} = \frac{1}{NT^{3/2}h} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}u_{it}K_h(\tau_t - \delta),$$

$$L_{2NT} = \frac{1}{NT^{3/2}h} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}\lambda'_{0i}f_{0t}K_h(\tau_t - \delta).$$

Assumption 6.1 is a simpler version of Assumption 3, so discussion is omitted. Assumption 6.2 permits correlation between X_{it} and u_{it}, as well as correlation between X_{it} and $\lambda'_{0i}f_{0t}$, and it can be easily verified (see Lemmas B.1 and B.2 of the supplementary file for example).

Under Assumption 6, a key fact is that for $\forall \delta \in (0,1)$, $\beta_0(\delta)$ can always be recovered consistently by a naive estimator:

$$\ddot{\beta}(\delta) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}X'_{it}K_h(\tau_t - \delta)\right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}Y_{it}K_h(\tau_t - \delta). \tag{2.7}$$

Lemma B.4 of the supplementary file shows that as long as Assumption 6 holds, $\ddot{\beta}(\delta)$ converges to $\beta_0(\delta)$ at the rate $O_P\left(\frac{1}{\sqrt{T}}\right)$ assuming $Th^4 \to 0$. We will utilize (2.7) later.

We now write (2.6) in matrix notation as follows:

$$Y_i = \phi_i[\beta_0] + F_0\lambda_{0i} + u_i, \tag{2.8}$$

where $\phi_i[\beta_0] = (X'_{i1}\beta_0(\tau_1), \ldots, X'_{iT}\beta_0(\tau_T))'$, and the other variables are defined conformably. For any given $F_0\lambda_{0i}$, when estimating $\beta_0(\delta)$, we just need to consider the OLS estimator for the following transformed model:

$$\Upsilon_\delta Y_i - \Upsilon_\delta F_0\lambda_{0i} = \Upsilon_\delta\phi_i[\beta_0] + \Upsilon_\delta u_i, \tag{2.9}$$

where $\Upsilon_\delta = \text{diag}\left\{\sqrt{\frac{1}{h}K_h(\tau_1 - \delta)}, \ldots, \sqrt{\frac{1}{h}K_h(\tau_T - \delta)}\right\}$. Since λ_{0i} is unobservable, we further remove it from (2.9) by left multiplying the projection matrix $M_{\Upsilon_\delta F_0}$:

$$M_{\Upsilon_\delta F_0}\Upsilon_\delta Y_i = M_{\Upsilon_\delta F_0}\Upsilon_\delta\phi_i[\beta_0] + M_{\Upsilon_\delta F_0}\Upsilon_\delta u_i. \tag{2.10}$$

With (2.10) in hand and accounting for the fact that F_0 is unknown as well, we write the objective function as follows:

$$Q(\beta, F) = \sum_{i=1}^N (Y_i - X_i\beta)'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta(Y_i - X_i\beta). \tag{2.11}$$

In order to avoid the singularity problem discussed in Section 4 of Bai et al. (2009), we use the information of $\ddot{\beta}(\cdot)$ to restrict β of (2.11) within $R_\delta(\beta) = \{\beta \mid \|\beta - \ddot{\beta}(\delta)\| \leq \frac{a_0}{\sqrt{T}}\}$, where a_0 is a sufficiently large constant. Moreover, in view of (2.9) and (2.10), we modify the identification condition on F of Bai (2009) by assuming $F \in D_F = \{F \mid \frac{1}{T}F'\Upsilon_\delta^2 F = I_r\}$. Finally, our estimates of $(\beta_0(\delta), F_0)$ are given below:

$$(\widetilde{\beta}(\delta), \widetilde{F}) = \operatorname*{argmin}_{\beta \in R_\delta(\beta), F \in D_F} Q(\beta, F). \tag{2.12}$$

In order to establish the consistency associated with (2.12), we impose the next assumption.

Assumption 7.

1. *Suppose* $\|\frac{1}{NTh}\sum_{i=1}^N \sum_{t=1}^T u_{it}\lambda'_{0i}f_{0t}K_h(\tau_t - \delta)\| = o_P(1)$, *and* $\|U\|_{\text{sp}} = O_P(\max\{\sqrt{N}, \sqrt{T}\})$, *where* $U = (u_1, \ldots, u_N)'$.

2. *(a) For* $\delta \in (0,1)$, $\frac{1}{Th}\sum_{t=1}^T f_{0t}f'_{0t}K_h(\tau_t - \delta) \to_P \Sigma_f(\delta)$, *and* $\max_{t \geq 1} E\|f_{0t}\|^4 < \infty$.

   (b) $\frac{1}{N}\sum_{i=1}^N \lambda_{0i}\lambda'_{0i} \to_P \Sigma_\lambda$, *and* $\max_{i \geq 1} E\|\lambda_{0i}\|^4 < \infty$.

3. Let $\inf_{F \in \mathsf{D}_F} \lambda_{\min}(\Omega_\dagger(F)) > 0$ *uniformly, where*

$$\Omega_\dagger(F) = \frac{1}{NT^2} \left\{ \Omega_1(F) - \Omega_2'(F) \left[\sum_{i=1}^N \lambda_{0i}\lambda_{0i}' \otimes I_T \right]^{-1} \Omega_2(F) \right\},$$

$$\Omega_1(F) = \sum_{i=1}^N X_i' \Upsilon_\delta M_{\Upsilon_\delta F} \Upsilon_\delta X_i, \quad \Omega_2(F) = \sum_{i=1}^N \lambda_i \otimes (M_{\Upsilon_\delta F} \Upsilon_\delta X_i).$$

Assumptions 7.1 and 7.2 are standard in the literature (Bai, 2009; Moon and Weidner, 2015). In addition, Assumption 7.1 implicitly permits certain correlation between u_{it} and $\lambda_{0i}' f_{0t}$. Assumption 7.3 is the nonparametric version of Assumption A of Bai (2009) under the nonstationary panel data setting.

We are now ready to present the first asymptotic property of (2.12).

Lemma 2.1. *Let Assumptions 1–2, 4, and 6–7 hold. For $\delta \in (0,1)$, as $(N,T) \to (\infty, \infty)$,*
(1). $\sqrt{T}(\widetilde{\beta}(\delta) - \beta_0(\delta)) = o_P(1)$; (2). $\|P_{\Upsilon_\delta \widetilde{F}} - P_{\Upsilon_\delta F_0}\| = o_P(1)$.

The first result of Lemma B.6 shows that (2.12) yields a more efficient estimator on $\beta_0(\delta)$ than the naive estimator $\ddot{\beta}(\delta)$. It is well understood that we can only recover F_0 up to a rotation matrix without more restrictive assumptions (e.g., Fan et al., 2016), so the second result is presented on the idempotent matrix associated with $\Upsilon_\delta \widetilde{F}$.

Before stating the rate of convergence, we present more restrictions on the error terms.

Assumption 8. *Let $u_t = (u_{1t}, \ldots, u_{Nt})'$ and the filtration $\mathcal{B}_{T,t} = \sigma(x_j, u_{j-1}, F_0, \Gamma_0 \,|\, j \leq t+1)$ form a martingale difference sequence such that $E[u_t \,|\, \mathcal{B}_{T,t-1}] = 0$ and $E[u_t u_t' \,|\, \mathcal{B}_{T,t-1}] = \Sigma_u = \{\sigma_{ij}\}_{N \times N}$ almost surely, where x_t is defined similar to u_t, and $\Gamma_0 = (\lambda_{01}, \ldots, \lambda_{0N})'$. In addition, let $\sigma_{ii} = \sigma_u^2$ for $i \geq 1$, $\sum_{i=1}^N \sum_{j=1, \neq i}^N |\sigma_{ij}| = O(N)$ and $N/T \to c < \infty$. Suppose also that $\max_{i,t} E[u_{it}^4 \,|\, \mathcal{B}_{T,t-1}] < \infty$ holds almost surely.*

We are now ready to establish the following theorem; its proof is given in the supplementary material.

Theorem 2.3. *Suppose that Assumptions 1–2, 4, and 6-8 hold. For $\delta \in (0,1)$, as $(N,T) \to (\infty, \infty)$, (1). $\widetilde{\beta}(\delta) - \beta_0(\delta) = O_P(\frac{1}{\sqrt{NT^2 h}} + h^2)$; (2). $\|P_{\Upsilon_\delta \widetilde{F}} - P_{\Upsilon_\delta F_0}\| = O_P\left(\sqrt[4]{T}h + \frac{1}{\sqrt{Th}} + \frac{1}{\sqrt{N}}\right)$.*

Having established Theorem 2.3, one can further investigate an asymptotic normality for $\widetilde{\beta}(\delta)$. As a matter of fact, some studies have been achieved in the supplementary file of Dong et al. (2019), where the sieve method is adopted. Under the framework of (2.6), modelling the endogeneity between X_{it} and u_{it} through a linear process requires more efforts, as one needs to account for f_{0t} in the data generating process. We leave these topics for future study.

2.3 A practical issue

Before closing this section we discuss a practical issue — dealing with mixed $I(0)/I(1)$ regressors, which is very likely to be encountered in empirical studies. Up to this point we have been assuming that X_{it} is an integrated process without including any stationary variables in either (2.1) or (2.6). Assuming that $I(0)$ variables are strictly exogenous, incorporating stationary variables is relatively easy for (2.1) due to the fact that the estimator (2.2) has a closed-form, and the asymptotic development can go through with some minor modifications. We now focus on the case with interactive fixed effects. A more realistic model is

$$Y_{it} = X'_{1it}\beta_{10}(\tau_t) + X'_{2it}\beta_{20}(\tau_t) + \lambda'_{0i}f_{0t} + u_{it}, \tag{2.13}$$

where $X_{1it} = X_{1i,t-1}+\nu_{it}$, X_{2it} is a vector including stationary variables and time invariant variables (e.g., gender and geographic variables), and the rest of the settings are identical to (2.6).

In order to estimate $\beta_{10}(\delta)$ and $\beta_{20}(\delta)$ for $\forall \delta \in (0,1)$, we can first establish a naive estimator for $\beta_{10}(\delta)$ as

$$\ddot{\beta}_1(\delta) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_{1it}X'_{1it}K_h(\tau_t - \delta)\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} X_{1it}Y_{it}K_h(\tau_t - \delta), \tag{2.14}$$

and construct $R_\delta(\beta) = \{\beta \,|\, \|\beta - \ddot{\beta}_1(\delta)\| \leq \frac{a_0}{\sqrt{Th}}\}$ under some regularity conditions. Then we define the estimators of $\beta_{10}(\delta)$ and $\beta_{20}(\delta)$ by

$$(\widetilde{\beta}_1(\delta), \widetilde{\beta}_2(\delta), \widetilde{F}) = \underset{\beta_1 \in R_\delta(\beta), \beta_2 \in R_2(\delta), F \in D_F}{\operatorname{argmin}} Q(\beta_1, \beta_2, F), \tag{2.15}$$

where $Q(\beta_1, \beta_2, F) = \sum_{i=1}^{N}(Y_i - X_{1i}\beta_1 - X_{2i}\beta_2)'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta(Y_i - X_{1i}\beta_1 - X_{2i}\beta_2)$, and $R_2(\delta)$ is the parameter space of $\beta_{20}(\delta)$. The development of (2.15) would be very similar to that given for (2.12). Once again, we leave such details for future research.

3 Application

In this section we focus on the income elasticities of healthcare expenditure for the Eurozone and OECD, where managing healthcare costs is possibly the most important concern of healthcare policy that requires a deep insight of its determinants. The consensus in the literature is that the main factor that drives healthcare expenditure is income (e.g.,

Newhouse, 1977; Gerdtham et al., 1992; Jones and Wildman, 2008; Baltagi and Moscone, 2010; Hauck and Zhang, 2016). Rising income and purchasing power enables the consumer to choose more and better healthcare products, which commonly results in an increase in the price of healthcare. Other variables have also been identified as drivers of healthcare spending. The two main questions under study are:

(1) Is the income elasticity of healthcare greater than 1 and how has it evolved over time?

(2) Has there been any change in the relationship between non-income variables and healthcare expenditure?

3.1 Data and Models

The period under study starts in 1995 and runs until 2014 for 20 countries in the Eurozone and for 34 countries in the OECD. We list the names of these countries in Appendix A.3. All variables were downloaded from the Worldbank dataset. The dependent variable in our study is the log of total healthcare expenditure per capita for country i at year t, denoted by lhe. The proxy for income is the log of the GDP per capita, denoted by $lgdp$. These variables have been adjusted for inflation and purchasing-power in U.S. dollars.

The literature has also validated some demographic variables, such as the shares of population over 65 years old and under 15 years old, as possible determinants of healthcare expenditure (see Hitiris and Posnett, 1992, among others). These are denoted in this paper by $pop65$ and $pop14$, respectively. In addition, the literature reports a positive significant relationship between the public finance share of healthcare and the total healthcare expenditure. The $public$ in our study is calculated as the percentage of government expenditure in public healthcare. In a nutshell, the dependent variable is lhe_{it} and the regressors are $lgdp_{it}$, $pop65_{it}$, $pop14_{it}$, $public_{it}$.

3.2 Brief Review on Methodology

At the starting point Newhouse (1977) shows the income elasticity of healthcare expenditure from around 1.20 to 1.50 in the OECD using a cross-sectional model,

$$\overline{lhe}_i = \beta_1 \, \overline{lgdp}_i + \beta_2 \, \overline{pop65} + \beta_3 \, \overline{pop14}_i + \beta_4 \, \overline{public}_i + \alpha + u_i, \tag{3.1}$$

where $i = 1, \ldots, N$. Variables are recorded over a number of years $t = 1, \ldots, T$ and the dependent variable, $\overline{lhe}_i$, is calculated as the average value of lhe_{it} over all years of the country i. The regressors are constructed similarly. With the rising popularity of panel

data models, Gerdtham et al. (1992) consider the individual fixed effects (FE) model,

$$lhe_{it} = \beta_1 \; lgdp_{it} + \beta_2 \; pop65_{it} + \beta_3 \; pop14_{it} + \beta_4 \; public_{it} + \alpha_i + u_{it}. \qquad (3.2)$$

Baltagi and Moscone (2010) and Moscone and Tosetti (2010) further raise the concerns on measuring heterogeneous unobserved fixed effects over time, and introduce the panel data model with interactive effects,

$$lhe_{it} = \beta_1 \; lgdp_{it} + \beta_2 \; pop65_{it} + \beta_3 \; pop14_{it} + \beta_4 \; public_{it} + \alpha_i + \lambda'_{0i} f_{0t} + u_{it}, \qquad (3.3)$$

where the first element of f_{0t} is constant 1 in order to include an individual effect in the factor structure.

Another possible source of misspecification of Models (3.1)-(3.3) is the violation of the assumption on the stationarity of the dependent and independent variables, which will lead to spurious results (Engle and Granger, 1987). Several studies have examined this assumption over the years with different datasets. McCoskey and Selden (1998) and Jewel et al. (2003) reject the unit-root hypothesis of the *lhe* and *lgdp* processes for most OECD countries in their sample. On the other hand, Gerdtham and Löthgren (2000), Okunade and Murthy (2002), and Baltagi and Moscone (2010) conclude that the series *lhe* and *lgdp* are nonstationary. In this line, Table 1 shows the percentage of countries in our sample for which the unit-root hypothesis is rejected with the Augmented Dickey-Fuller (ADF) tests at 5% significance level. The null hypothesis is not rejected for most countries. The highest rate of rejection is for variables *pop*65 and *pop*14 but still under 40%. In summary, nonstationarity should be accounted for in practice.

Table 1: Percentage of rejection of ADF tests across countries for the Eurozone and OECD. The tests are done for the alternative hypotheses of (i) random walk with drift, and (ii) random walk with drift and trend and with 1 lag.

	Eurozone		OECD	
	Drift	Drift + trend	Drift	Drift + trend
lhe	0.00%	5.00%	8.82%	5.88%
lgdp	10.00%	5.00%	5.88%	5.88%
pop65	0.00%	35.00%	5.88%	26.47%
pop14	35.00%	15.00%	26.47%	14.71%
public	10.00%	10.00%	14.71%	5.88%

In addition, Models (3.1)-(3.3) assume that the coefficients of the covariates are constant over time. Our contribution is the recommendation of two models. The first model

is a fixed effects model with time-varying coefficients estimated with $\widehat{\beta}_*(\cdot)$ defined in (2.5) as follows:

$$lhe_{it} = \beta_1(\tau_t)\, lgdp_{it} + \beta_2(\tau_t)\, pop65_{it} + \beta_3(\tau_t)\, pop14_{it} + \beta_4(\tau_t)\, public_{it} + \alpha_i + u_{it}. \quad (3.4)$$

This model allows for the covariates to be unit-root processes and the coefficients to vary over time. Additionally, $\widehat{\beta}_*(\cdot)$ corrects the bias arising from endogeneity.

The second recommendation is a model with time-varying coefficients and interactive effects that again allows for covariates to be unit-root processes. The individual fixed effects are added to the model as there is likely heterogeneity among countries. However, time fixed effects, if existing, are included in the interactive effects term, which also groups other possible determinants not included explicitly in the covariates,

$$lhe_{it} = \beta_1(\tau_t)\, lgdp_{it} + \beta_2(\tau_t)\, pop65_{it} + \beta_3(\tau_t)\, pop14_{it} + \beta_4(\tau_t)\, public_{it} + \alpha_i + \lambda'_{0i} f_{0t} + u_{it}. \quad (3.5)$$

where the first element of f_{0t} is constant 1 in order to include an individual effect in the factor structure.

3.3   Is Healthcare a Luxury in Developed Countries ?

Results in Table A.2 show the income elasticity estimates of healthcare expenditure and their 95% confidence intervals from Models (3.1)-(3.3). The computation is done using the R packages *plm* by Croissant and Millo (2008) and *phtt* by Bada and Liebl (2014). The aim of this analysis is to confirm historical results with our sample. The estimates of the coefficient of *lgdp* from Model (3.1) and Model (3.2) are over 1 for both the OECD and Eurozone, which is in concordance with earlier studies like Newhouse (1977), Leu (1986), and Gerdtham et al. (1992). On the other hand, the estimated coefficient of *lgdp* from Model (3.3) is significantly below 1 in concordance with some recent studies like Baltagi and Moscone (2010) and Murthy and Ketenci (2017). The number of interactive effects is selected with the criterion in Ahn and Horenstein (2013), implemented in the R package *phtt* by Bada and Liebl (2014) and described in Appendix B.1 in the supplementary sections. Basically, the inclusion of interactive effects in classic panel data models justifies the argument that healthcare is a necessity good. However, it does not reveal any information on possible dynamics of this elasticity over time, a matter that may be explained by the time-varying coefficients of Models (3.4)-(3.5). Thus, in the following, we focus on reporting the results associated with Models (3.3)-(3.5). Model (3.4) is estimated by the bias-corrected estimator, $\widehat{\beta}_*(\delta)$ in Equation (2.5) which is also the initial

step in Model (3.5) and from whose residuals the unobservable factors and loadings are estimated. In this way, the estimation difference from adding time-varying coefficients can be analysed by comparing Models (3.3) and (3.5), and the estimation difference from adding interactive effects can be analysed by comparing Models (3.4) and (3.5).

Figure A.2 compares the income elasticity estimates of healthcare expenditure from Models (3.3)-(3.5). Three factors for the OECD and four for the Eurozone are selected in Model (3.5) with the criterion in Ahn and Horenstein (2013). The bandwidth is selected by leave-k-out cross-validation (Chu and Marron, 1991) with $k = \lfloor T/4 \rfloor$ to ensure the independence of the subsamples. The dashed black line and light brown band correspond to the income elasticity from the parametric Model (3.3). As mentioned above and in Table A.2, the elasticity is significantly less than 1: an expected increase of 1% in income will result in an expected increase of 0.91% in healthcare spending for the Eurozone and of 0.67% for the OECD. This model hints that the fraction of GDP per capita devoted to pay healthcare is lower in the OECD countries than in the Eurozone. The opposite is estimated by Model (3.5), dashed red line with an orange band, with lower estimates for the Eurozone than the OECD with values near 0.7 and 0.79, respectively. On the other hand, the continuous black line and grey band, corresponding to the income elasticity from Model (3.4), shows similar values than Model (3.5) but with a slight variation over time. In a nutshell, all models show income elasticities less than 1 with little variation over time even for Model (3.4). This suggests that healthcare is a necessary good and that the observed increasing dynamics in healthcare expenditure per capita depend on other variables besides income.

3.4 Do Age Structure and Public Expenditure Affect the Price of Healthcare ?

A positive relationship between lhe and $pop65$, and a negative relationship between lhe and $pop14$ are expected. As reported in Fisher (1990) and Alemayehu and Warner (2004), the health bill of the young is the lowest in a health system while the largest, by far, is the health bill of the elderly. Moreover, more than one-third of people's lifetime health spending will accrue in the last years (Zweifel et al., 1999; Alemayehu and Warner, 2004). As anticipated, the estimates of the $pop65$ coefficient in Figure A.3 are positive for all models and significant for Models (3.3) and (3.5) during the whole period. The coefficient estimates of $pop65$ from Model (3.4) are time-varying and increasing over time with significant values at 5% level after 2004 for the Eurozone and after 2008 for the OECD. Estimates of the $pop14$ coefficients are negative and very similar for all models for the

Eurozone: an increase in fertility rate will result in a money saving from healthcare budgets. Similar results are found for the OECD from Models (3.4)-(3.5) however, the estimate from Model (3.3) is significantly positive, which is a counterintuitive result.

Finally, all estimators report a positive significant relationship between government public investment and healthcare expenditure (Table A.2 and third row of Figure A.3) for the OECD. This means that an increase in public funding of healthcare boosts healthcare expenditure per capita. Interestingly, Model (3.4) shows a decreasing trend of this coefficient and the 95% confidence interval of the estimates from Model (3.5) include zero. A possible explanation for this is that most countries in the Eurozone, first the Nordic countries and the rest after the Global Financial Crisis, have applied health policies to make their health systems more efficient (Mladovsky et al., 2012) and minimize the effect of higher unemployment rates and lower tax collection in healthcare expenditure per capita. The findings indicate that the latest reforms aiming to dissociate healthcare from public funding are working in the right direction. This is supported by results reported in Reeves et al. (2014) showing that austerity measures, such as those imposed in the Eurozone after the Global Financial Crisis, result in healthcare budget cuts and cheaper healthcare per capita.

In conclusion, the estimates of Model (3.5) show that they are very similar to the initial expectations and in concordance with the latest healthcare policies. Curiously, these coefficients are basically constant although they use the same bandwidth and estimation as those for Model (3.4) in addition to factors. The message is that adding interactive effects compensates for the dynamics on the relationship between healthcare expenditure and the rest of the variables because these variables are likely correlated with the unobservable factors.

3.5 Analysis of the Interactive Effects

Much research has been done to find further determinants of healthcare expenditure. It has been revealed that: (a) the deployment of medical technology increases the price of healthcare (see Newhouse, 1992; Okunade and Murthy, 2002; Murthy and Ketenci, 2017); (b) mortality rate and five-year survival rates are positively related with healthcare expenditure (Breyer et al., 2015); and (c) life expectancy at birth is also highly correlated (Murthy and Ketenci, 2017). In a more general approach, Hauck and Zhang (2016) select a list of 43 potential drivers of healthcare expenditure using Bayesian methods, and find that 16 of those are significant determinants of healthcare expenditure, which includes the unemployment rate, growth in insurance premiums, and increase in pharmaceutical

sales, among others. This is simplified by including interactive effects in the model that group similar determinants into a set of unobservable factors f_{0t} whose relationship is estimated in λ_{0i} for each country.

The algorithm in Lam and Yao (2012) and Ahn and Horenstein (2013) selects four unobservable factors from Model (3.5) for the Eurozone and three for the OECD, plotted in Figure A.4. The first factor, black line, is monotonically increasing. The second factor, in red, is a concave line with the maximum at the middle of the time period. The third factor, in green, is a sinusoidal line with the minimum near year 2005 and maxima close to years 1996 and 2010 for the Eurozone. For the Eurozone, the fourth factor, in blue, is also sinusoidal.

We do not venture into explaining the meaning of each factor, as they represent groups of variables with correlated patterns. However, the monotonic increasing pattern of the first factor follows a similar pattern as the OCDE average pharmaceutical spending and average life expectancy at birth series in Figure A.5 (two variables that have been identified before as determinants of healthcare expenditure in Okunade and Murthy (2002); Hauck and Zhang (2016); Murthy and Ketenci (2017)). This supports the conjecture that technological change is a major escalator of healthcare expenditure, as first stated in Newhouse (1992).

Although the unobservable factors are common to all countries, their loadings vary by country. Figure A.6 shows the interactive effect terms in Model (3.5) for each country in the OECD. They are plotted in geographical/political groups for ease of visualisation. Positive values of these interactive effects mean that other variables in addition to income, demographic characteristics and public expending increase the price of healthcare, while the opposite is true for negative values. We can distinguish three kind of patterns in these plots: (a) the interactive effects increase steadily over time to take a turn down after the Global Financial Crisis, for example in Finland, Iceland, the Netherlands, Portugal and the United States; (b) the interactive effects do not change much over the sample period, for example in Australia, Austria, Canada, Chile and Greece; (c) the interactive effects follow a concave curve, for example in France, Germany, Italy, Sweden and Switzerland. Many countries show a downturn in the interactive effects after the Global Financial Crisis reflecting the effects of the austerity health policies implemented after the crisis to reduce the burden of the price of health on the tax system.

4 Conclusion

In this paper the estimator for a time-varying coefficient fixed effects panel data model is derived using a nonparametric kernel smoothing technique. In our study the regressors are modelled as unit-root processes whose error term can be correlated contemporaneously. A bias corrected estimator is provided to account for the correlation between the regressor innovations and the equation error. In addition, the inclusion of a term of heterogeneous unobserved factors in the model has been proposed.

The newly developed methodology is used to analyze the income elasticity dynamics of healthcare expenditure in the OECD and Eurozone. We show that health cost per capita increases at a lower rate than does income per capita, meaning that healthcare is a necessity good. Age demographics and government public funding rate are also significantly related to the healthcare expenditure. An increase in ageing population results in an increase in the health cost, while an increase in fertility rate results in a slight drop in health cost. Furthermore, the relationship between public funding and healthcare expenditure is decreasing over the time period, although still positive.

Acknowledgements

References

Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**, 1203–1227.

Alemayehu, B. and Warner, K. E. (2004), 'The lifetime distribution of health care costs', *Health Services Research* pp. 627–642.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* (3), 817–858.

Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press.

Bada, O. and Liebl, D. (2014), 'The R package phtt: Panel data analysis with heterogeneous time trends', *Journal of Statistical Software* **59**(6), 1–34.

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bai, J., Kao, C. and Ng, S. (2009), 'Panel cointegration with global stochastic trends', *Journal of Econometrics* **149**(1), 82–99.

Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Baltagi, B. H. and Moscone, F. (2010), 'Health care expenditure and income in the oecd reconsidered: Evidence from panel data', *Economic Modelling* **27**, 804–811.

Breyer, F., Lorenz, N. and Niebel, T. (2015), 'Health care expenditures and longevity: Is there a Eubie Blake effect?', *European Journal of Health Economics* **16**, 95–112.

Cai, Z. and Li, Q. (2008), 'Nonparametric estimation of varying coefficient dynamic panel data models', *Econometric Theory* **24**(5), 1321–1342.

Chu, C. K. and Marron, J. S. (1991), 'Comparison of two bandwidth selectors with dependent errors', *Annals of Statistics* **19**, 1906–1918.

Connor, G., Hagmann, M. and Linton, O. (2012), 'Efficient semiparametric estimation of the fama-french model and extensions', *Econometrica* **80**(2), 713–754.

Croissant, Y. and Millo, G. (2008), 'Panel data econometrics in R: The plm package', *Journal of Statistical Software* **27**, 1–43.

Culyer, A. J. (1988), *Health Care Expenditures in Canada: Myth and Reality, Past and Future*, Canadian Tax Foundation. Canadian tax papers; no. 82.

Di Matteo, L. and Di Matteo, R. (1998), 'Evidence on the determinants of canadian provincial government health expenditures: 1965–1991', *Journal of Health Economics* **17**(2), 211 – 228.

Dong, C., Gao, J. and Peng, B. (2019), 'Varying-coefficient panel data models with nonstationarity and partially observed factor structure', *Available at https://ideas.repec.org/p/msh/ebswps/2018-1.html* .

Engle, R. F. and Granger, C. W. J. (1987), 'Co-integration and error correction: Representation, estimation and testing', *Econometrica* **55**(2), 251–276.

Fan, J., Liao, Y. and Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society: Series B* **75**(4), 603–680.

Fan, J., Liao, Y. and Wang, W. (2016), 'Projected principal component analysis in factor models', *Annals of Statistics* **44**(1), 219–254.

Fisher, C. R. (1990), 'Differences by age groups in health care spending', *Health Care Financing Review* **1**, 65–90.

Gao, J., Xia, K. and Zhu, H. (2019), 'Heterogeneous panel data models with cross-sectional dependence', *Forthcoming in Journal of Econometrics* p. Available at https://ideas.repec.org/p/wyi/wpaper/002482.html.

Gerdtham, U.-G. and Löthgren, M. (2000), 'On stationarity and cointegration of international health expenditure and gdp', *Journal of Health Economics* **19**, 461–475.

Gerdtham, U.-G., Soegaard, J., Andersson, F. and Jönsson, B. (1992), 'An econometric analysis of health care expenditure: A cross-section study of the OECD countries', *Journal of Health Economics* **11**, 63–84.

Hartwig, J. (2008), 'What drives health care expenditure?—baumol's model of 'unbalanced growth' revisited', *Journal of Health Economics* **27**(3), 603 – 623.

Hauck, K. and Zhang, X. (2016), 'Heterogeneity in the effect of common shocks on healthcare expenditure growth', *Health Economics* **25**(9), 1090–1103.

Hitiris, T. and Posnett, J. (1992), 'The determinants and effects of health expenditure in developed countries', *Journal of Health Economics* **11**, 173–181.

Hsiao, C. (2014), *Analysis of Panel Data, 3rd*, Cambridge University Press.

Jewel, T., Lee, J., Tieslau, M. and Strazicich, M. C. (2003), 'Stationarity of health expenditures and gdp: Evidence from panel unit root tests with heterogeneous structural breaks', *Journal of Health Economics* **22**(2), 313–323.

Jones, A. M. and Wildman, J. (2008), 'Health, income and relative deprivation: Evidence from the bhps', *Journal of Health Economics* **27**(2), 308 – 324.

Lam, C. and Yao, Q. (2012), 'Factor modeling for high-dimensional time series: Inference for the number of factors', *Annals of Statistics* **40**(2), 694–726.

Leu, R. E. (1986), The public-private mix and international health care costs, *in* T. Culyer and B. Jonsson, eds, 'Public and Private health services', Basil Blackwell, Oxford.

McCoskey, S. and Selden, T. (1998), 'Health care expenditures and GDP: Panel data unit root test results', *Journal of Health Economics* **17**, 369–376.

Mladovsky, P., Srivastava, D., Cylus, J., Karanikolos, M., Evetovits, T., Thomson, S. and McKee, M. (2012), Health policy responses to the financial crisis in Europe, Summary Report on Behalf of the European Observatory on Health Systems and Policies 2012, World Health Organization.
URL: *http://www.euro.who.int/__data/assets/pdf_file/0009/170865/e96643.pdf*

Moon, H. R. and Weidner, M. (2015), 'Linear regression for panel with unknown number of factors as interactive fixed effects', *Econometrica* **83**(4), 1543–1579.

Moscone, F. and Tosetti, E. (2010), 'Health expenditure and income in the United States', *Health Economics* **19**(12), 1385–1403.

Murthy, V. N. and Ketenci, N. (2017), 'Is technology still a major driver of health expenditure in the united states ? evidence from cointegration analysis with multiple structural breaks', *International Journal of Health Economics and Management* **17**, 29–50.

Newhouse, J. P. (1977), 'Medical care expenditure: A cross-national survey', *Journal of Human Resources* **12**(1), 115–125.

Newhouse, J. P. (1992), 'Medical care costs: How much welfare loss?', *Journal of Economic Perspectives* **6**, 3–21.

Okunade, A. A. and Murthy, V. N. (2002), 'Technology as a 'major driver' of health care costs: a cointegration analysis of the newhouse conjecture', *Journal of Health Economics* **21**(1), 147–159.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Phillips, P. C. B. (2001), 'Trending time series and macroeconomic activity: Some present and future challenges', *Journal of Econometrics* **100**(1), 21–27.

Phillips, P. C. B. and Solo, V. (1992), 'Asymptotics for linear processes', *Annals of Statistics* **20**(2), 971–1001.

Phillips, P. C., Li, D. and Gao, J. (2017), 'Estimating smooth structural change in cointegration models', *Journal of Econometrics* **196**(1), 180–195.

Reeves, A., McKee, M., Basu, S. and Stuckler, D. (2014), 'The political economy of austerity and healthcare: Cross-national analysis of expenditure changes in 27 european nations 1995-2011', *Health policy* **115**, 1–8.

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

Su, L. and Wang, X. (2017), 'On time–varying factor models: Estimation and testing', *Journal of Econometrics* **198**(1), 84–101.

Sun, Y., Carroll, R. and Li, D. (2009), 'Semiparametric estimation of fixed-effects panel data varying coefficient models', *Advances in Econometrics* **25**, 101–129.

Zweifel, P., Felder, S. and Meiers, M. (1999), 'Ageing of population and health care expenditure: A red herring?', *Health Economics* **8**, 485–496.

Appendix A

We denote some notations for simplicity. First define the next long-run covariance matrix

$$\left[\begin{array}{cc} \Delta_\nu & \Delta_{\nu u} \\ \Delta'_{\nu u} & \Delta_u \end{array} \right] = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=0}^{\infty} E \left[\begin{array}{cc} \nu_{i0}\nu'_{ij} & \nu_{i0}u_{ij} \\ \nu'_{i0}u_{ij} & u_{i0}u_{ij} \end{array} \right]. \tag{A.1}$$

In addition, we denote

$$\Delta_1 = (X'W_\delta^* X)^{-1} N \sum_{t=1}^{T} K_h(\tau_t - \delta) \left(\Delta_{\nu u} - \Delta_{\nu \bar{u}}^t(\delta) \right),$$

$$\Delta_{\nu \bar{u}}^t(\delta) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{l \le t} \left\{ \frac{1}{Z_T} \sum_{s=1}^{T} K_h(\tau_s - \delta) \mathrm{E}[\nu_{il} u_{is}] \right\},$$

$$\mathcal{C}_K = \int_{-1}^{1} \int_{-1}^{1} \min(s + 1, r + 1) K(s) K(r) ds dr,$$

$$\mathcal{C}_*(1) = \int_{-1}^{1} K^2(w) dw, \quad \mathcal{C}_*(2) = \mathcal{C}_*(1) \int_{-1}^{1} \mathsf{K}^2(s) ds + \left(\int_{-1}^{1} K(t)\mathsf{K}(t) dt \right)^2,$$

$$\mathcal{C}_*(1,2) = \int_{-1}^{1} K^2(t) \int_{-1}^{t} \mathsf{K}(s) \, ds \, dt + \int_{-1}^{1} K(t)\mathsf{K}(t)[1 - \mathsf{K}(t)] \, dt,$$

$$\mathcal{C}_* = \mathcal{C}_*(1) + \mathcal{C}_*(2) - 2\mathcal{C}_*(1,2), \tag{A.2}$$

where $Z_T = \sum_{t=1}^{T} K_h(\tau_t - \delta)$ and $\mathsf{K}(s) = \int_s^1 K(t) dt$.

A.1 The Intuition of (2.2)

We now explain the intuition of (2.2). Using the condition $\sum_{i=1}^{N} \alpha_i = 0$, we can rewrite (2.1) in a matrix format as $Y = \Psi_X(\beta_0) + DA_\alpha + u$, where $\Psi_X(\beta_0) = (X_{11}'\beta_0(\tau_1), \ldots, X_{NT}'\beta_0(\tau_T))'$ and u are defined conformably to Y, and $A_\alpha = (\alpha_2, \ldots, \alpha_N)'$. Given Assumption 2, we have $\beta_0(\tau_t) \approx \beta_0(\delta)$ for $\forall \delta \in (0, 1)$ when τ_t is in a small neighbourhood of δ. Thus, we intuitively consider the next objective function

$$L(\beta, A) = (Y - X\beta - DA)'W_\delta(Y - X\beta - DA), \tag{A.3}$$

where β and A are $d \times 1$ and $(N - 1) \times 1$ vectors respectively. Note that for $\forall \beta \in \mathbb{R}^d$, the estimate of A_α is

$$\widehat{A}(\beta) = (D'W_\delta D)^{-1} D'W_\delta(Y - X\beta), \tag{A.4}$$

Plugging $\widehat{A}(\beta)$ back into $L(\beta, A)$, we obtain the objective function of the so-called profile method (e.g., Sun et al., 2009) as follows:

$$L(\beta) = (Y - X\beta)'W_\delta^*(Y - X\beta). \tag{A.5}$$

Minimizing $L(\beta)$, (2.2) follows immediately.

A.2 Discussion on the Degeneracy Issue of Phillips et al. (2017)

Note that by some routine procedure (Phillips and Solo, 1992), we have for each i and for $t = \lfloor Tr \rfloor$ with $0 < r < 1$, $\frac{1}{\sqrt{T}} \sum_{s=1}^{t} (\nu_{is}', u_{is})' = \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor Tr \rfloor} (\nu_{is}', u_{is})' \implies \mathcal{B}_r(\Omega)$, where $\Omega =$

Figure A.1: Mean value of log-HCE (solid lines) and log-GDP (dashed lines) per capita in the Eurozone and the OECD countries during the period 1995-2014.
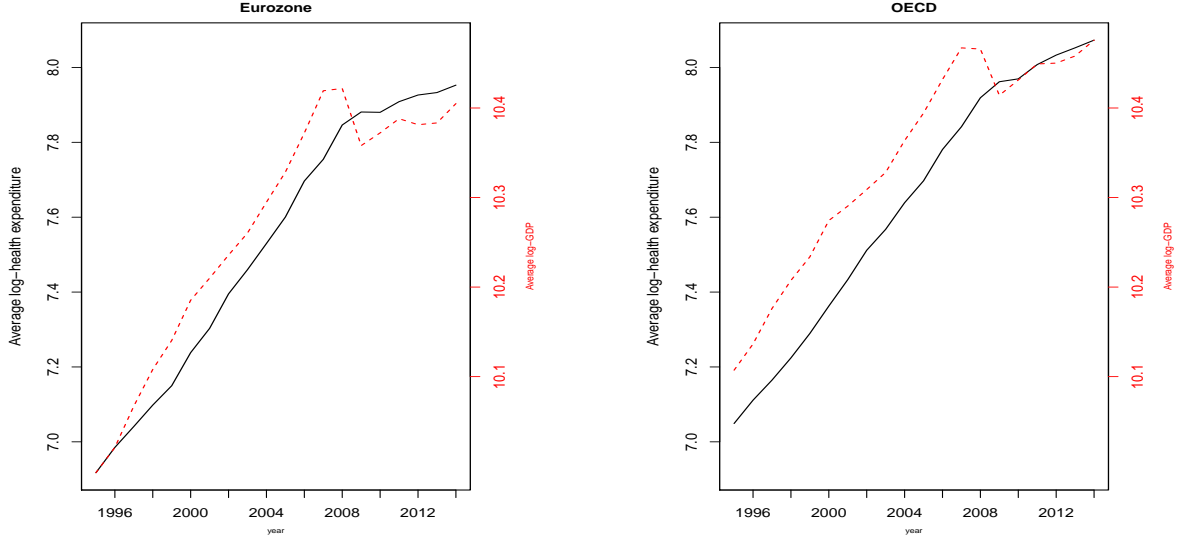
$(\Phi, \psi)' \Lambda (\Phi, \psi)$, and $\mathcal{B}_r(\Omega)$ denotes a Brownian motion with covariance matrix Ω. Moreover, by Phillips et al. (2017, Eq 2.18), for each i and for $\forall \delta \in (0,1)$, $\frac{1}{T^2 h} \sum_{t=1}^{T} X_{it} X_{it}' K_h(\tau_t - \delta) \implies \delta \Phi' \mathcal{W}_{d+1}(\Lambda) \Phi$, where $\mathcal{W}_{d+1}(\Lambda)$ is a Wishart variate having 1 degree of freedom and mean matrix Λ. Thus, when $d \geq 2$, $\Phi' \mathcal{W}_{d+1}(\Lambda) \Phi$ is singular, so it becomes problematic for nonparametric regression under the context of time series analysis as explained in Phillips et al. (2017). However, for panel data analysis, this issue is no longer a problem. An intuitive explanation is that we take the average over both i and t, so after some tedious algebra we have $\frac{1}{N} \sum_{i=1}^{N} \Phi' E[\mathcal{W}_{d+1}(\Lambda)] \Phi = \Phi' \Lambda \Phi > 0$. See Lemma B.1 of the supplementary file for details. Thus, the degeneracy issue disappears when panel data are involved.

A.3   Countries included in this study

- Eurozone: Austria, Belgium, Cyprus, Estonia, Finland, France, Germany, Greece, Ireland, Italy, Lithuania, Luxembourg, Latvia, Macedonia, Malta, Netherlands, Portugal, Slovak Republic, Slovenia, and Spain

- OECD: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Luxembourg, Latvia, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States
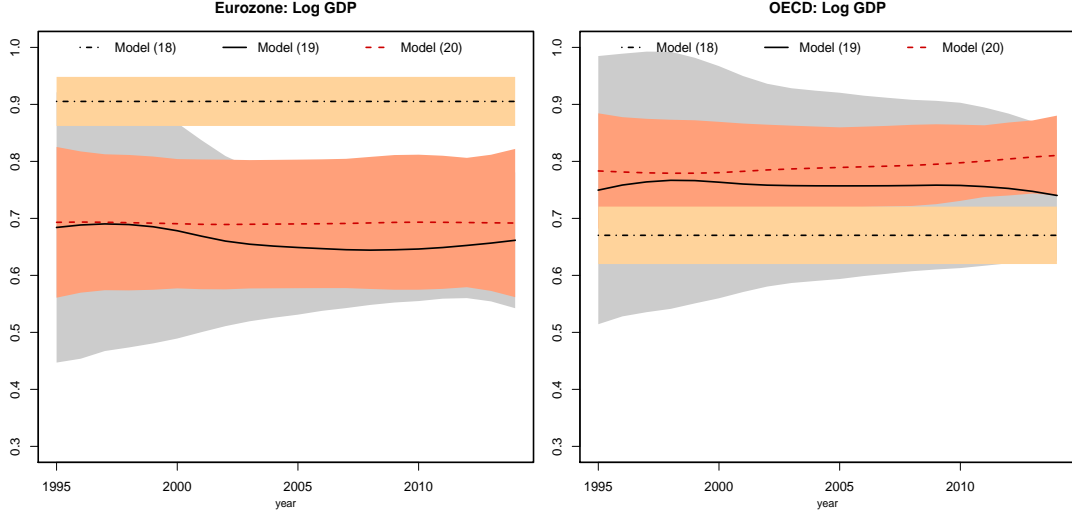
In this supplementary file, Section B.1 and Section B.2 discuss some practical issues related to estimating the model with a factor structure; Section B.3 includes some simulation studies; Section B.4 states the preliminary lemmas; Section B.5 provides all the proofs.

	Eurozone			OECD		
	Cross-sectional Model (3.1)	Fixed effects Model (3.2)	Interactive effects Model (3.3)	Cross-sectional Model (3.1)	Fixed effects Model (3.2)	Interactive effects Model (3.3)
(Intercept)	-4.771**			-5.738***		
	(-7.950, -1.592)			(-8.075, -3.405)		
$lgdp$	1.092 ***	1.043***	0.905***	1.169***	1.133***	0.670***
	(0.860,1.324)	(0.898,1.187)	(0.862,0.948)	(0.916,1.423)	(1.007,1.260)	(0.62,0.721)
$pop65$	0.017	0.085***	0.079***	0.027	0.056***	0.058***
	(-0.066,0.099)	(0.069,0.100)	(0.074,0.083)	(-0.025,0.079)	(0.045,0.067)	(0.054,0.063)
$pop14$	0.004	-0.026**	-0.047	0.012	-0.059***	0.007***
	(-0.097,0.104)	(-0.042,-0.010)	(-0.052,-0.042)	(-0.028,0.051)	(-0.073,-0.045)	(0.002,0.013)
$public$	0.058*	0.052***	0.030***	0.048*	0.053***	0.036***
	(0.013,0.103)	(0.043,0.061)	(0.028,0.033)	(0.010,0.085)	(0.047,0.06)	(0.033,0.038)

The 95% confidence intervals are displayed in parentheses.

Figure A.2: Coefficient estimates of *lgdp* during the period 1995-2014 for the Eurozone and OECD countries (left to right). The plots show the estimated values of the coefficients at each year and their 95% confidence interval (bands). TThe dashed black line with light brown bands corresponds to Model (3.3), the continuous black line with grey bands corresponds to Model (3.4), and the red line with orange bands corresponds to Model (3.5).

Appendix B

B.1 Numerical Algorithm

We now provide the numerical algorithm associated to (9) of the main text. Similar to Bai (2009), we can decompose it as follows.

$$(1). \quad \widetilde{\beta}(\delta) = \underset{\beta \in R_\delta(\beta)}{\operatorname{argmin}} \sum_{i=1}^{N} (Y_i - X_i\beta)' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta (Y_i - X_i\beta),$$

$$(2). \quad \frac{1}{NT} \sum_{i=1}^{N} \Upsilon_\delta (Y_i - X_i\widetilde{\beta}(\delta))(Y_i - X_i\widetilde{\beta}(\delta))' \Upsilon_\delta^2 \widetilde{F} = \Upsilon_\delta \widetilde{F} V_{NT},$$

where V_{NT} is a diagonal matrix with the diagonal being the r largest eigenvalues of

$$\frac{1}{NT} \sum_{i=1}^{N} \Upsilon_\delta (Y_i - X_i\widetilde{\beta}(\delta))(Y_i - X_i\widetilde{\beta}(\delta))' \Upsilon_\delta$$

arranged in descending order. One can adopt an iteration procedure as in Bai (2009) to implement the estimation method. Moreover, this procedure can be easily modified to accommodate (12). Note that the restriction set $R_\delta(\beta)$ can be easily imposed through some programs like "fmincon" of MATLAB.

Note that $\widetilde{F}$ generated in the above procedure is related to the value of δ by design. Prac-

Figure A.3: Coefficient estimates of *pop*65, *pop*14, and *public* variables (top to bottom) during the period 1995-2014 for the Eurozone and OECD countries (left to right). The plots show the estimated values of the coefficients at each year and their 95% confidence interval (bands). The dashed black line with light brown bands corresponds to Model (3.3), the continuous black line with grey bands corresponds to Model (3.4), and the red line with orange bands corresponds to Model (3.5).
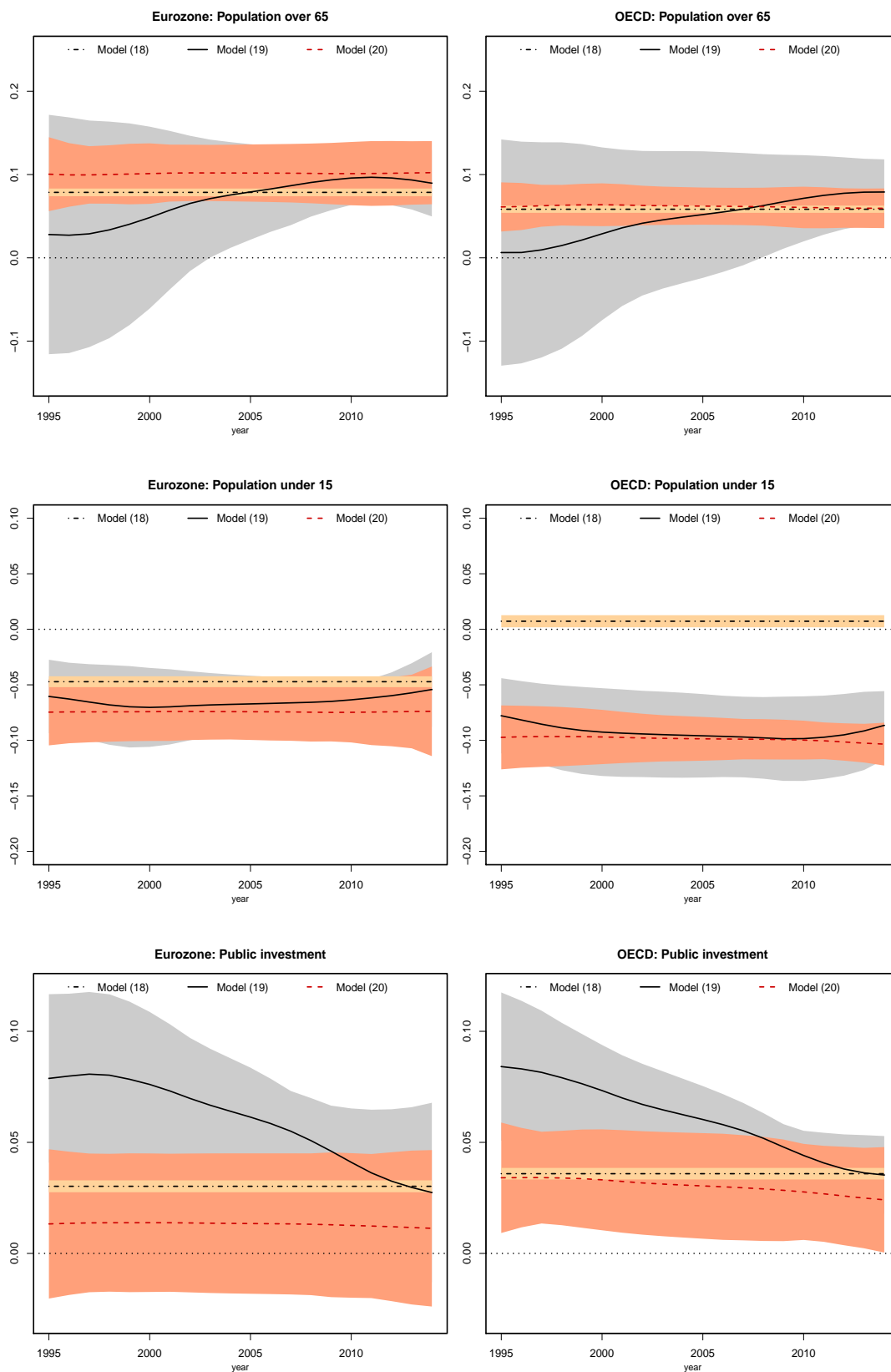
Figure A.4: Unobserved factor estimates from Model (3.5) for the Eurozone and OECD.
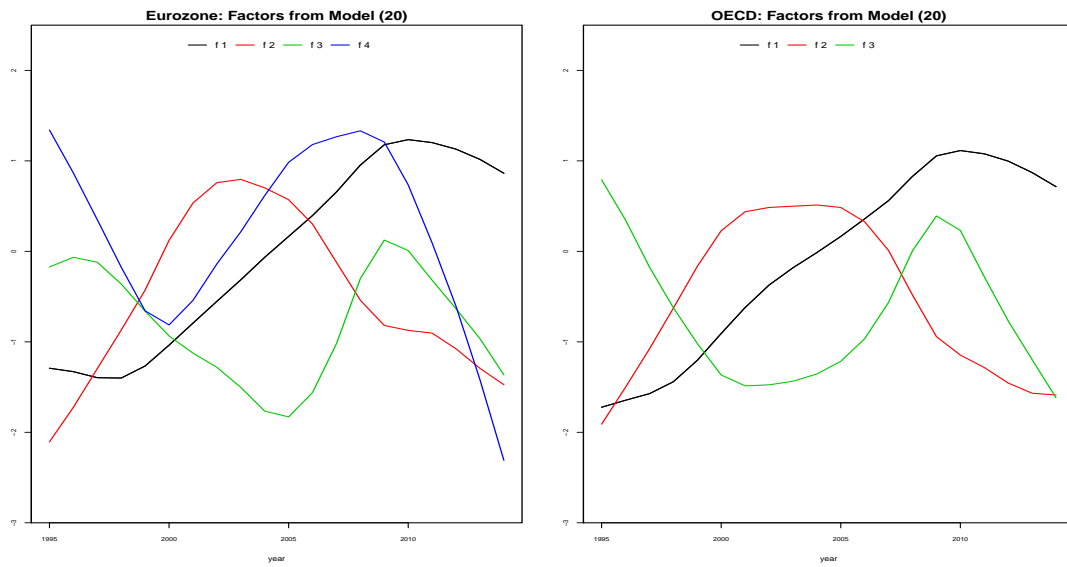

Figure A.5: Average values of the OECD of pharmaceutical price and life expectancy at birth downloaded from the OECD database at data.oecd.org.
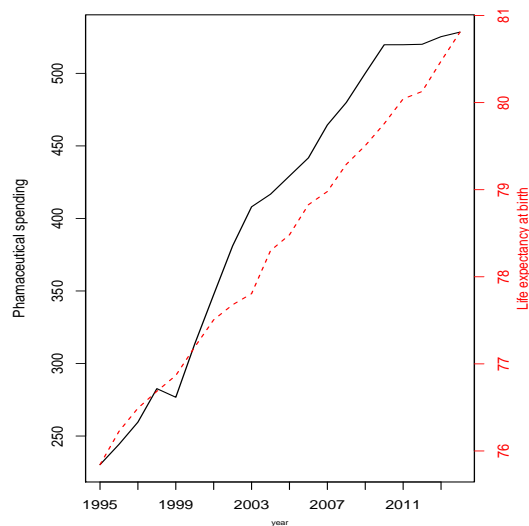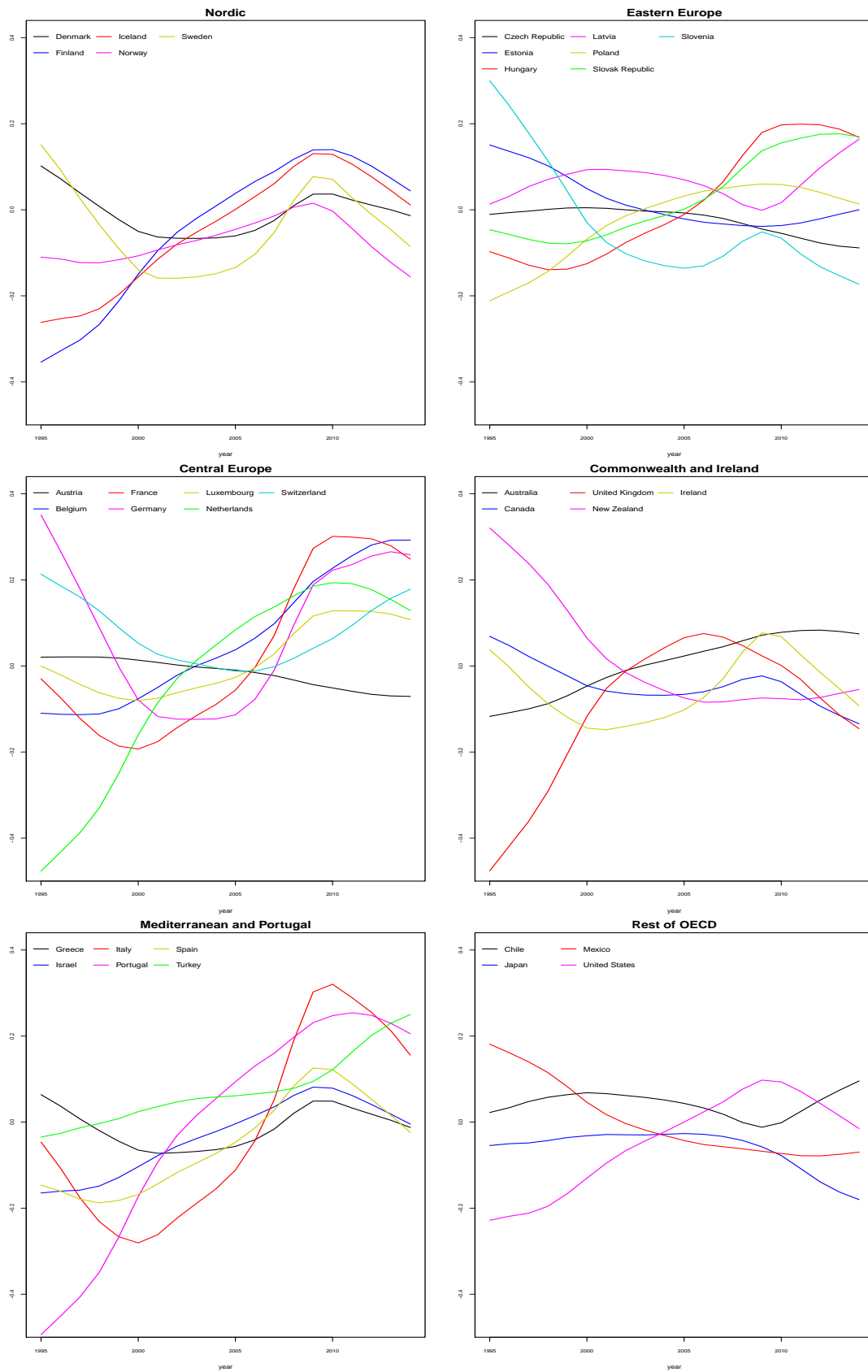

25

Figure A.6: Interactive effects obtained from Model (3.5) for the OECD.

tically, in order to obtain an estimate of F_0 which is not subject to δ, one can estimate $\beta_0(\tau_1), \ldots, \beta_0(\tau_T)$ first, and conduct PCA on $Y_{it}^* = Y_{it} - X_{it}'\widehat{\beta}(\tau_t)$, which then ensures the estimates on the factors and loadings are independent of the value of δ.

B.2 Selection of the Number of Factors

We consider the ratio criterion studied in Lam and Yao (2012) and Ahn and Horenstein (2013). Specifically, we let $\widehat{\lambda}_j$ be the jth largest eigenvalue of the estimated sample covariance matrix

$$\frac{1}{N}\sum_{i=1}^{N}\Upsilon_\delta(Y_i - X_i\widetilde{\beta}_J(\delta))(Y_i - X_i\widetilde{\beta}_J(\delta))'\Upsilon_\delta,$$

where $\widetilde{\beta}_J(\delta)$ is obtained from (9) of the main text assuming that the number of factors is a pre-specified fixed positive integer J. We then estimate the number of factors by

$$\widehat{r} = \operatorname*{argmin}_{j\in\{1,\ldots,J-1\}} \frac{\widehat{\lambda}_{j+1}}{\widehat{\lambda}_j}.$$

Note that slightly over-identifying the number of factors usually does not have any serious impact on consistency and rates of convergence of the subsequent estimation (Fan et al., 2013; Moon and Weidner, 2015). That is why $\widetilde{\beta}_J(\delta)$ is adopted above. After identifying the number of factors by $\widehat{r}$, we can update the estimate on $\beta_0(\delta)$. Alternatively, one may consider using the criterion provided in Bai and Ng (2002).

B.3 Simulation Studies

In this section, we examine the finite sample performance of the estimates studied in the main text. As well understood in the literature, the selection of an optimal bandwidth is still an open question for panel data models. In the following experiments, we consider the rule of thumb ($h = 1.06(NT)^{-1/5}$). In addition, we have experimented different choices of the bandwidth to examine the sensitivity of the bandwidth selection, which is in fact quite stable. For similarity, we omit the relevant results.

B.3.1 A Panel Data Model with Fixed Effects

We firstly consider a panel data model with fixed effects. The data generating process is of the form

$$\begin{aligned}
Y_{it} =& X_{it,1}\beta_{t,1} + X_{it,2}\beta_{t,2} + \alpha_i + u_{it}, \\
(X_{it,1}, X_{it,2}) =& (X_{i,t-1,1}, X_{i,t-1,2}) + (\nu_{it,1}, \nu_{it,2}),
\end{aligned} \tag{B.1}$$

Table B.3: Results of Model (B.1)

		(i). no correlation			(ii). with correlation		
		20	40	80	20	40	80
$\widehat{\beta}(\cdot)$	20	0.5027	0.4349	0.3595	0.5697	0.4259	0.3548
	40	0.4986	0.4006	0.3461	0.5983	0.4248	0.3538
	80	0.4740	0.3823	0.2571	0.6721	0.4641	0.3100
$\widehat{\beta}_*(\cdot)$	20	0.4844	0.4476	0.3631	0.4341	0.3739	0.3406
	40	0.4412	0.3967	0.3469	0.3982	0.3252	0.3024
	80	0.3934	0.3501	0.2041	0.3948	0.2950	0.2415

where $(\beta_{t,1}, \beta_{t,2}) = (\cos(2\pi\tau_t), \tau_t^2)$ with $\tau_t = t/T$ and $\alpha_i = \frac{1}{T^{3/2}} \sum_{t=1}^{T} X_{it,1}$ for $i = 2, \cdots, N$ and $\alpha_1 = -\sum_{i=1}^{N-1} \alpha_i$. The error terms $(u_{it}, v_{it,1}, v_{it,2})$ are generated as follows.

$$u_{it} = \rho_0 u_{i,t-1} + \xi_{it}, \quad v_{it,1} = \rho_1 v_{i,t-1,1} + \eta_{it,1}, \quad v_{it,2} = \rho_2 v_{i,t-1,2} + \eta_{it,2}.$$

Two simulation scenarios are performed: (i) the errors are generated as i.i.d. processes by setting $(\rho_0, \rho_1, \rho_2) = (0, 0, 0)$, and (ii) the errors are generated as stationary AR(1) processes by setting $(\rho_0, \rho_1, \rho_2) = (-0.5, -0.5, -0.5)$. Finally, let $(\xi_{it}, \eta_{it,1}, \eta_{it,2})'$ be i.i.d. over i and t following the next distribution

$$\begin{pmatrix} \xi_{it} \\ \eta_{it,1} \\ \eta_{it,2} \end{pmatrix} \overset{i.i.d.}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & 1 & \mu_3 \\ \mu_2 & \mu_3 & 1 \end{pmatrix} \right),$$

where $\mu_1 = \mu_2 = \mu_3 = 0.8$. Simulations are conducted for sample sizes $N = 20, 40, 80$ and $T = 20, 40, 80$ and the number of replications is $R = 500$. Following Phillips et al. (2017), we compare the performances of $\widehat{\beta}(\cdot)$ and $\widehat{\beta}_*(\cdot)$ of Section 2.1 using the absolute averages $\frac{1}{T} \sum_{t=1}^{T} \sum_{l=1}^{2} |\frac{1}{R} \sum_{r=1}^{R} (\widehat{\beta}_l^{(r)}(\tau_t) - \beta_l^0(\tau_t))|$.

As shown in Table B.3, the absolute averages converge to 0 as the sample sizes increase. $\widehat{\beta}_*(\cdot)$ has better performance compared to $\widehat{\beta}(\cdot)$, which is expected. For case (ii), the results are worse than those in case (i), which shows how biases arise with the change in (ρ_0, ρ_1, ρ_2).

B.3.2   A Panel Data Model with Interactive Fixed Effects

A second experiment investigates the estimation of a panel data model with interactive fixed effects — the performance of $\widetilde{\beta}(\cdot)$ in the main text. The data generating process is

$$Y_{it} = X_{it,1}\beta_{t,1} + X_{it,2}\beta_{t,2} + \lambda'_{0i}f_{0t} + u_{it},$$

Table B.4: Results of Model (B.2)

		(i). no correlation			(ii). with correlation		
		20	40	80	20	40	80
$\widetilde{\beta}(\cdot)$	20	0.5898	0.5530	0.5023	0.6490	0.5865	0.5243
	40	0.5054	0.4747	0.4195	0.6001	0.5225	0.4424
	80	0.4059	0.3726	0.3226	0.5245	0.4443	0.3497

$$(X_{it,1}, X_{it,2}) = (X_{i,t-1,1}, X_{i,t-1,2}) + (\nu_{it,1}, \nu_{it,2}), \tag{B.2}$$

where coefficients, regressors and error terms are the same as in Section B.3.1. Let $r = 2$, and then generate the loadings as $\lambda_{0i} = (\lambda_{0i,1}, \lambda_{0i,2})'$, where $\lambda_{0i,1} \sim \frac{X_{it,1}}{\sqrt{T}} + N(0, 1)$ and $\lambda_{0i,2} \sim \frac{X_{it,2}}{\sqrt{T}} + N(0, 1)$. Let $f_{0t} \sim N(0, I_2)$. For the factor model case, loading variables are generated as $\lambda_{0i,1} = \frac{X_{it,1}}{T^{3/2}} + N(0, 1)$, $\lambda_{0i,2} = \frac{X_{it,2}}{T^{3/2}} + N(0, 1)$. Other variables are generated in the same way as in Section B.3.1.

As shown in Table B.4, the absolute averages converge to 0 as the sample sizes increase. However, how to introduce a bias corrected estimator for $\widetilde{\beta}(\cdot)$ is beyond the scope of this study. We leave it for future work.

B.4 Preliminary Lemmas

Lemma B.1. *Suppose that Assumptions 1, 3 and 4 are satisfied. For any $\forall \delta \in (0, 1)$, as $(N, T) \to (\infty, \infty)$, $\frac{1}{NT^2h} X' W_\delta^* X \to_P (1 - \mathcal{C}_K)\Sigma_\nu$, where $\mathcal{C}_K$ and Σ_ν are defined in (16) of the main text and Theorem 1, respectively.*

Lemma B.2. *Suppose that Assumptions 1, 3 and 4 are satisfied. For any $\forall \delta \in (0, 1)$, as $(N, T) \to (\infty, \infty)$,*

$$\frac{1}{\sqrt{NT^2h}} \left\{ X' W_\delta^* u - N \sum_{t=1}^{T} K_h(\tau_t - \delta)(\Delta_{\nu u} - \Delta_{\nu \bar{u}}^t(\delta)) \right\} \Longrightarrow N(0, \mathcal{C}_* \Sigma_{\nu,u}),$$

where $\mathcal{C}_* := \mathcal{C}_*(1) + \mathcal{C}_*(2) - 2\mathcal{C}_*(1, 2)$ with $\mathcal{C}_*(1)$, $\mathcal{C}_*(2)$ and $\mathcal{C}_*(1, 2)$ being defined in (19) of the main text, and $\Sigma_{\nu,u}$ is defined in Theorem 1.

Lemma B.3. *Suppose that Assumptions 1–5 are satisfied. As $(N, T) \to (\infty, \infty)$,*

1. $\sqrt{Nh}(\widehat{\Delta}_{\nu u} - \Delta_{\nu u}) = o_P(1)$,

2. $\frac{\sqrt{Nh}}{Th} \sum_{t=1}^{T} K_h(\tau_t - \delta)(\widehat{\Delta}_{\nu \bar{u}}^t(\delta) - \Delta_{\nu \bar{u}}^t(\delta)) = o_P(1)$.

Lemma B.4. *Suppose that Assumptions 1, 2, 4 and 6 hold. For $\forall \delta \in (0, 1)$, $\ddot{\beta}(\delta) - \beta_0(\delta) = O_P\left(\frac{1}{\sqrt{Th}} + h^2\right)$.*

Lemma B.5. *Suppose that Assumptions 1–2, 4, and 6–7 hold. As $(N, T) \to (\infty, \infty)$,*

1. $\sup_{F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} u_i' \Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta u_i \right| = o(1)$;

2. $\sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i'\beta)' \Upsilon_\delta M_{\Upsilon_\delta F} \Upsilon_\delta u_i \right| = o_P(1)$;

3. $\sup_{F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \lambda_{0i}' F_0' \Upsilon_\delta M_{\Upsilon_\delta F} \Upsilon_\delta u_i \right| = o_P(1)$.

Lemma B.6. *Suppose that Assumptions 1–2, 4, and 6–8 hold. As $(N, T) \to (\infty, \infty)$,*

1. $V_{NT} \to_P V$,

2. $\frac{1}{\sqrt{T}} \|\Upsilon_\delta \widetilde{F} \Pi_{NT}^{-1} - \Upsilon_\delta F_0\| = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{\sqrt{Th}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)$,

3. $\frac{1}{T} \|F_0' \Upsilon_\delta^2 \widetilde{F} - F_0' \Upsilon_\delta^2 F_0 \Pi_{NT}\| = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O\left(\frac{1}{N}\right)$,

4. $\|P_{\Upsilon_\delta \widetilde{F}} - P_{\Upsilon_\delta F_0}\|^2 = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O\left(\frac{1}{N}\right)$,

where V is a $r \times r$ diagonal matrix consisting of the eigenvalues of $\Sigma_f(\delta)\Sigma_\lambda$, and $\Pi_{NT}^{-1} = V_{NT} \left(\frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T}\right)^{-1} \left(\frac{\Lambda_0' \Lambda_0}{N}\right)^{-1}$.

B.5 Proofs

In this subsection, we firstly present the proofs of the main results, and then provide the proofs of the preliminary lemmas.

B.5.1 Proofs of the Main Results

Proof of Theorem 1:

Observe that

$$
\begin{aligned}
\widehat{\beta}(\delta) - \beta_0(\delta) &= (X' W_\delta^* X)^{-1} X' W_\delta^* Y - \beta_0(\delta) \\
&= (X' W_\delta^* X)^{-1} X' W_\delta^* \Psi_X(\beta_0) - \beta_0(\delta) \\
&\quad + (X' W_\delta^* X)^{-1} X' W_\delta^* D A_\alpha + (X' W_\delta^* X)^{-1} X' W_\delta^* u \\
&= (X' W_\delta^* X)^{-1} X' W_\delta^* \Psi_X(\beta_0) - \beta_0(\delta) + (X' W_\delta^* X)^{-1} X' W_\delta^* u \\
&:= \Xi_{NT}(1) + \Xi_{NT}(2),
\end{aligned}
$$

where $\Psi_X(\beta_0)$ and A_α are defined in Appendix A.1; and the third equality follows from $W_\delta^* D A_\alpha = 0$ by the definition of W_δ^*.

For $\Xi_{NT}(1)$, it is easy to know that

$$
\Xi_{NT}(1) = (X' W_\delta^* X)^{-1} X' W_\delta^* \Psi_X(\beta_0) - \beta_0(\delta) = O_P(h^2)
$$

by Assumption 2 and the Taylor expansion. In connection with Lemmas B.1 and B.2, the result follows. ∎

Proof of Theorem 2:

The result follows from Lemma B.3 straightaway, so omitted. ∎

Proof of Lemma 1:

Before proceeding further, we define some variables. Let $\xi_F = \text{vec}(M_{\Upsilon_\delta F}\Upsilon_\delta F_0)$ for $\forall F \in D_F$, and let $\widetilde{M}_0 = M_{\Upsilon_\delta F_0}$ for short. In addition, let $A_1 = \sum_{i=1}^N X_i'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta X_i$, $A_2 = (\Gamma_0'\Gamma_0) \otimes I_T$, and $A_3 = \sum_{i=1}^N \lambda_{0i} \otimes (M_{\Upsilon_\delta F}\Upsilon_\delta X_i)$. We are now ready to start the proof.

Write

$$
\frac{1}{NT}Q(\beta, F) - \frac{1}{NT}Q(\beta_0(\delta), F_0)
$$

$$
= \frac{1}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - X_i\beta)'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta(\phi_i[\beta_0] - X_i\beta)
$$

$$
+ \frac{1}{NT}\sum_{i=1}^N \lambda_{0i}'F_0'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta F_0\lambda_{0i} + \frac{2}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - X_i\beta)'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta F_0\lambda_{0i}
$$

$$
+ \frac{1}{NT}\sum_{i=1}^N u_i'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta u_i + \frac{2}{NT}\sum_{i=1}^N \lambda_{0i}'F_0'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta u_i
$$

$$
+ \frac{2}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - X_i\beta)'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta u_i
$$

$$
- \frac{1}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - X_i\beta_0(\delta))'\Upsilon_\delta\widetilde{M}_0\Upsilon_\delta(\phi_i[\beta_0] - X_i\beta_0(\delta))
$$

$$
- \frac{1}{NT}\sum_{i=1}^N u_i'\Upsilon_\delta\widetilde{M}_0\Upsilon_\delta u_i - \frac{2}{NT}\sum_{i=1}^N (\phi_i[\beta_0] - X_i\beta_0(\delta))'\Upsilon_\delta\widetilde{M}_0\Upsilon_\delta u_i
$$

$$
= \frac{1}{NT}\sum_{i=1}^N (\beta_0(\delta) - \beta)'X_i'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta X_i(\beta_0(\delta) - \beta) + \frac{1}{NT}\sum_{i=1}^N \lambda_{0i}'F_0'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta F_0\lambda_{0i}
$$

$$
+ \frac{2}{NT}\sum_{i=1}^N (\beta_0(\delta) - \beta)'X_i'\Upsilon_\delta M_{\Upsilon_\delta F}\Upsilon_\delta F_0\lambda_{0i} + o_P(1)
$$

$$
= (\beta_0(\delta) - \beta)'A_1(\beta_0(\delta) - \beta) + \xi_F'A_2\xi_F + 2(\beta_0(\delta) - \beta)'A_3'\xi_F + o_P(1)
$$

$$
= \sqrt{T}(\beta_0(\delta) - \beta)'\frac{1}{NT^2}\left(A_1 - A_3'A_2^{-1}A_3\right)\sqrt{T}(\beta_0(\delta) - \beta)
$$

$$
+ \frac{1}{NT}[\xi_F' + (\beta_0(\delta) - \beta)'A_3'A_2^{-1}]A_2[\xi_F + A_2^{-1}A_3(\beta_0(\delta) - \beta)] + o_P(1)
$$

$$
= \sqrt{T}(\beta_0(\delta) - \beta)'\Omega_\dagger(F)\sqrt{T}(\beta_0(\delta) - \beta)
$$

$$
+ \frac{1}{NT}[\xi_F' + (\beta_0(\delta) - \beta)'A_3'A_2^{-1}]A_2[\xi_F + A_2^{-1}A_3(\beta_0(\delta) - \beta)] + o_P(1),
$$

where the second equality follows from Lemma B.5, and some routine procedure of the kernel method. Then by the same arguments as in Bai (2009, p. 1265), the two results follow. ∎

Proof of Theorem 3:

(1). We now start investigating the rate of convergence, and write

$$\widetilde{\beta}(\delta) - \beta_0(\delta) = \left(\sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i\right)^{-1} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta Y_i$$

$$= \left(\sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i\right)^{-1} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta \left(\phi_i[\beta_0] - X_i\beta_0(\delta)\right)$$

$$+ \left(\sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i\right)^{-1} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta F_0 \lambda_{0i}$$

$$+ \left(\sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i\right)^{-1} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta u_i$$

$$:= A_1 + A_2 + A_3.$$

By the proof of Theorem 1, it is easy to know that $A_1 = O_P(h^2)$. Below, we focus on A_2 and A_3. Let $\Xi_{NT} = \left(\frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T}\right)^{-1} \left(\frac{\Lambda_0' \Lambda_0}{N}\right)^{-1}$ for notational simplicity, which gives $\Pi_{NT}^{-1} = V_{NT}\Xi_{NT}$.

First, we investigate A_2, and write

$$\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta F_0 \lambda_{0i}$$

$$= -\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} (\Upsilon_\delta \widetilde{F} \Pi_{NT}^{-1} - \Upsilon_\delta F_0) \lambda_{0i}$$

$$= -\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \left[I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{8NT}(\widetilde{F})\right] \Xi_{NT} \lambda_{0i}$$

$$:= -J_{1NT} - \cdots - J_{8NT},$$

where the second equality follows from (B.17), and the definitions of J_{1NT} to J_{8NT} should be obvious. In view of the decomposition of J_{2NT}, it is easy to know

$$\|J_{1NT}\| = o_P(\|\widetilde{\beta}(\delta) - \beta_0(\delta)\| + h^2).$$

Thus, we start from J_{2NT} and write

$$J_{2NT} = \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \frac{1}{NT} \sum_{j=1}^{N} \Upsilon_\delta (\phi_j[\beta_0] - X_i\widetilde{\beta}(\delta))(F_0\lambda_{0j})' \Upsilon_\delta^2 \widetilde{F} \Xi_{NT} \lambda_{0i}$$

$$= \frac{1}{N^2T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_j (\beta_0(\delta) - \widetilde{\beta}(\delta)) \lambda_{0j}' \frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T} \Xi_{NT} \lambda_{0i}$$

$$+ \frac{1}{N^2T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta (\phi_j[\beta_0] - X_j\beta_0(\delta)) \lambda_{0j}' \frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T} \Xi_{NT} \lambda_{0i}$$

32

$$= -\frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_j \lambda_{0j}' \left(\frac{\Lambda_0' \Lambda_0}{N} \right)^{-1} \lambda_{0i} (\widetilde{\beta}(\delta) - \beta_0(\delta))$$

$$+ \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta (\phi_j[\beta_0] - X_j \beta_0(\delta)) \lambda_{0j}' \frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T} \Xi_{NT} \lambda_{0i}$$

$$:= -J_{2NT,1} + J_{2NT,2}.$$

By a development similar to (A.2) of Dong et al. (2019), it is easy to know that

$$\left\| \left(\sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} NT^2 J_{2NT,2} \right\| = O_P(h^2).$$

We will further study $J_{2NT,1}$ later.

For J_{3NT}, write

$$J_{3NT} = \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \frac{1}{NT} \sum_{j=1}^{N} \Upsilon_\delta F_0 \lambda_{0j} (\phi_j[\beta_0] - X_i \widetilde{\beta}(\delta))' \Upsilon_\delta^2 \widetilde{F} \Xi_{NT} \lambda_{0i}$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} (\Upsilon_\delta F_0 - \Upsilon_\delta \widetilde{F} \Pi_{NT}^{-1}) \frac{1}{NT} \sum_{j=1}^{N} \lambda_{0j} (\phi_j[\beta_0] - X_i \widetilde{\beta}(\delta))' \Upsilon_\delta^2 \widetilde{F} \Xi_{NT} \lambda_{0i}$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} J_{3NT,i},$$

where the definition of $J_{3NT,i}$ is obvious. By a development similar to (A.2) of Dong et al. (2019), we just need to focus on $\frac{1}{NT^2} \sum_{i=1}^{N} \|J_{3NT,i}\|^2$ in order to show

$$\left\| \left(\sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} NT^2 J_{3NT} \right\| = o_P(\|\widetilde{\beta}(\delta) - \beta_0(\delta)\| + h^2).$$

Then write

$$\frac{1}{NT^2} \sum_{i=1}^{N} \|J_{3NT,i}\|^2$$

$$\leq \frac{\|\Upsilon_\delta F_0 - \Upsilon_\delta \widetilde{F} \Pi_{NT}^{-1}\|^2}{NT^2} \sum_{i=1}^{N} \left\| \frac{1}{NT} \sum_{j=1}^{N} \lambda_{0j} (\phi_j[\beta_0] - X_j \widetilde{\beta}(\delta))' \Upsilon_\delta \right\|^2 \left\| \Upsilon_\delta \widetilde{F} \Xi_{NT} \lambda_{0i} \right\|^2$$

$$\leq O_P(1) \frac{\|\Upsilon_\delta F_0 - \Upsilon_\delta \widetilde{F} \Pi_{NT}^{-1}\|^2}{T} \left(\frac{1}{NT} \sum_{j=1}^{N} \|\lambda_{0j}\| \cdot \|(\phi_j[\beta_0] - X_j \widetilde{\beta}(\delta))' \Upsilon_\delta\| \right)^2$$

$$\leq O_P(1) \frac{\|\Upsilon_\delta F_0 - \Upsilon_\delta \widetilde{F} \Pi_{NT}^{-1}\|^2}{T} \left(\frac{1}{N} \sum_{j=1}^{N} \|\lambda_{0j}\|^2 \right) \left(\frac{1}{N} \sum_{j=1}^{N} \frac{1}{T^2} \|(\phi_j[\beta_0] - X_j \widetilde{\beta}(\delta))' \Upsilon_\delta\|^2 \right)$$

$$\leq O_P(1)\frac{\|\Upsilon_\delta F_0 - \Upsilon_\delta \widetilde{F}\Pi_{NT}^{-1}\|^2}{T} \cdot \frac{1}{N}\sum_{j=1}^{N}\frac{1}{T^2}(\phi_j[\beta_0] - X_i\widetilde{\beta}(\delta))'\Upsilon_\delta^2(\phi_j[\beta_0] - X_i\widetilde{\beta}(\delta))$$

$$= o_P(\|\widetilde{\beta}(\delta) - \beta_0(\delta)\| + h^2).$$

Therefore, $\left\|\left(\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\Upsilon_\delta X_i\right)^{-1}NT^2 J_{3NT}\right\| = o_P(\|\widetilde{\beta}(\delta) - \beta_0(\delta)\| + h^2)$ follows.

For J_{4NT}, write

$$
\begin{aligned}
J_{4NT} &= \frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\frac{1}{NT}\sum_{j=1}^{N}\Upsilon_\delta(\phi_j[\beta_0] - X_j\widetilde{\beta}(\delta))u_j'\Upsilon_\delta^2\widetilde{F}\Xi_{NT}\lambda_{0i} \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\frac{1}{NT}\sum_{j=1}^{N}\Upsilon_\delta(X_j\beta_0(\delta) - X_j\widetilde{\beta}(\delta))u_j'\Upsilon_\delta^2 F_0\Pi_{NT}\Xi_{NT}\lambda_{0i} \\
&\quad + \frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\frac{1}{NT}\sum_{j=1}^{N}\Upsilon_\delta(\phi_j[\beta_0] - X_j\beta_0(\delta))u_j'\Upsilon_\delta^2 F_0\Pi_{NT}\Xi_{NT}\lambda_{0i} \\
&\quad + \frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\frac{1}{NT}\sum_{j=1}^{N}\Upsilon_\delta(\phi_j[\beta_0] - X_j\widetilde{\beta}(\delta))u_j'\Upsilon_\delta^2(\widetilde{F} - F_0\Pi_{NT})\Xi_{NT}\lambda_{0i} \\
&:= J_{4NT,1} + J_{4NT,2} + J_{4NT,3}.
\end{aligned}
$$

For $J_{4NT,1}$, write

$$
\begin{aligned}
\|J_{4NT,1}\| &= \left\|\frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\frac{1}{NT}\sum_{j=1}^{N}\Upsilon_\delta(X_j\beta_0(\delta) - X_j\widetilde{\beta}(\delta))u_j'\Upsilon_\delta^2 F_0\Pi_{NT}\Xi_{NT}\lambda_{0i}\right\| \\
&\leq O_P(1)\frac{1}{NT}\sum_{i=1}^{N}\left\|\frac{X_i'}{\sqrt{T}}\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\right\| \cdot \|\lambda_{0i}\|\frac{1}{NT}\sum_{j=1}^{N}\left\|\Upsilon_\delta\frac{X_j}{\sqrt{T}}\right\| \cdot \|u_j'\Upsilon_\delta^2 F_0\| \cdot \|\beta_0(\delta) - \widetilde{\beta}(\delta)\| \\
&\leq O_P(1)\frac{1}{NT}\left\{\sum_{i=1}^{N}\left\|\frac{X_i'}{\sqrt{T}}\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\Upsilon_\delta\frac{X_i'}{\sqrt{T}}\right\|\right\}^{1/2}\left\{\sum_{i=1}^{N}\|\lambda_{0i}\|^2\right\}^{1/2} \\
&\quad \cdot \frac{1}{NT}\left\{\sum_{j=1}^{N}\left\|\frac{X_j}{\sqrt{T}}\Upsilon_\delta^2\frac{X_j}{\sqrt{T}}\right\|\right\}^{1/2}\left\{\sum_{j=1}^{N}\|u_j'\Upsilon_\delta^2 F_0\|^2\right\}^{1/2} \cdot \|\beta_0(\delta) - \widetilde{\beta}(\delta)\| \\
&\leq O_P(1)\frac{1}{\sqrt{T}}\|\beta_0(\delta) - \widetilde{\beta}(\delta)\| = o_P(\|\beta_0(\delta) - \widetilde{\beta}(\delta)\|).
\end{aligned}
$$

Thus, $J_{4NT,1}$ is negligible. Similarly, we can show both $\|J_{4NT,2}\|$ and $\|J_{4NT,3}\|$ are negligible by accounting for $\frac{1}{NT^2} = \sum_{j=1}^{N}(\phi_j[\beta_0] - X_j\beta_0(\delta))'\Upsilon_\delta^2(\phi_j[\beta_0] - X_j\beta_0(\delta)) = O_P(h^2)$ and result (2) of Lemma B.6. Analogous to the derivation of J_{3NT} and J_{4NT}, we can conclude $\|J_{5NT}\|$ is negligible.

Similarly, we can prove that

$$J_{6NT} = \frac{1}{NT^2}\sum_{i=1}^{N}X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\frac{1}{T}\Upsilon_\delta\Omega_u\Upsilon_\delta^2\widetilde{F}\Xi_{NT}\lambda_{0i} + \text{negligible terms}$$

$$= \frac{\sigma_u^2}{NT^3} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta \cdot \Upsilon_\delta^2 \widetilde{F} \Xi_{NT} \lambda_{0i} + \text{negligible terms}$$

$$:= J_{6NT,1} + \text{negligible terms}$$

and J_{7NT} is negligible, where $\Omega_u = E[u_i u_i'] = \sigma_u^2 I_T$.

Based on the above development, we have

$$\widetilde{\beta}(\delta) - \beta_0(\delta) + \left(\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} J_{2NT,1}$$

$$= -\left(\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} \left\{ \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta u_i + J_{6NT,1} + J_{8NT} \right\}$$

$$+ \text{negligible terms}$$

$$= -\left(\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} \left\{ \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta u_i \right.$$

$$\left. + \frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \frac{1}{NT} \sum_{j=1}^{N} \Upsilon_\delta u_j (F_0 \lambda_{0j})' \Upsilon_\delta^2 \widetilde{F} \Xi_{NT} \lambda_{0i} \right\}$$

$$- \left(\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} J_{6NT,1} + \text{negligible terms}$$

$$= -\left(\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} \frac{1}{NT^2} \sum_{i=1}^{N} \left\{ X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta \right.$$

$$\left. + \frac{1}{N} \sum_{j=1}^{N} X_j' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta \lambda_{0j}' (\Lambda_0' \Lambda_0 / N)^{-1} \lambda_{0i} \right\} u_i$$

$$- \left(\frac{1}{NT^2} \sum_{i=1}^{N} X_i' \Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}} \Upsilon_\delta X_i \right)^{-1} J_{6NT,1} + \text{negligible terms}$$

$$:= -B_1 - B_2 + \text{negligible terms}$$

Following a procedure similar to Lemma B.2, we can show that $B_1 = O_P(1/\sqrt{NT^2h})$. Moreover, note that $\sqrt{NT^2h} J_{6NT,1}$ can be written as

$$\left\| \sqrt{NT^2h} J_{6NT,1} \right\| = \left\| \sqrt{NT^2h} \cdot \frac{\sigma_u^2}{NT^3} \sum_{i=1}^{N} Z_i \Upsilon_\delta^2 \xi_i \right\|$$

$$= \left\| \frac{\sigma_u^2 \sqrt{N}}{\sqrt{T}} \cdot \frac{1}{NT} \sum_{i=1}^{N} \frac{\sqrt{h}}{\sqrt{T}} Z_i \Upsilon_\delta \cdot \Upsilon_\delta \xi_i \right\|$$

$$\leq O(1) \frac{\sqrt{N}}{\sqrt{T}} \left\{ \frac{1}{NT} \sum_{i=1}^{N} h \frac{Z_i'}{\sqrt{T}} \Upsilon_\delta^2 \frac{Z_i}{\sqrt{T}} \right\}^{1/2} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \xi_i' \Upsilon_\delta^2 \xi_i \right\}^{1/2}$$

$$\leq O(1) \frac{\sqrt{N}}{\sqrt{T}} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \frac{Z_i'}{\sqrt{T}} \frac{Z_i}{\sqrt{T}} \right\}^{1/2} \left\{ \frac{1}{NT} \sum_{i=1}^{N} \xi_i' \Upsilon_\delta^2 \xi_i \right\}^{1/2} = O_P \left(\frac{N}{T} \right),$$

where $Z_i = X_i'\Upsilon_\delta M_{\Upsilon_\delta \widetilde{F}}\Upsilon_\delta$ and $\xi_i = \widetilde{F}\Xi_{NT}\lambda_{0i}$. In connection with the condition $\frac{N}{T} \to c < \infty$ of Assumption 8, the first result follows.

(2). By the first result of this theorem, and (4) of Lemma B.6, the second result follows.

∎

B.5.2   Proofs of the Preliminary Lemmas

We introduce some variables which will be repeatedly used below. Let $\delta(T) = \lfloor (\delta - h)T \rfloor$. From the BN decomposition (Phillips and Solo, 1992), we have for $t \geq \delta(T)$

$$
\begin{aligned}
X_{it} &= \sum_{s=1}^{t} \nu_{is} + X_{i0} = \sum_{s=1}^{t} \bar{\nu}_{is} + \widetilde{\nu}_{i0} - \widetilde{\nu}_{it} + X_{i0} \\
&= \Big[\sum_{s=1}^{\delta(T)} \bar{\nu}_{is} + \widetilde{\nu}_{i0} - \widetilde{\nu}_{i,\delta(T)} + X_{i0} \Big] + \Big[\sum_{s=\delta(T)+1}^{t} \bar{\nu}_{is} \Big] + \Big[\widetilde{\nu}_{i,\delta(T)} - \widetilde{\nu}_{it} \Big] \\
&:= X_{i,\delta(T)} + \eta_{it} + \xi_{it},
\end{aligned}
\tag{B.3}
$$

where $\bar{\nu}_{it} = (\sum_{j=0}^{\infty} \Phi_j')\varepsilon_{it} = \Phi'\varepsilon_{it}$, and $\widetilde{\nu}_{it} = \sum_{j=0}^{\infty} \widetilde{\Phi}_j'\varepsilon_{i,t-j}$ with $\widetilde{\Phi}_j = \sum_{k=j+1}^{\infty} \Phi_k$. Note that Assumption 3 ensures $\sum_{j=0}^{\infty} \|\widetilde{\Phi}_j\| < \infty$, so that $\xi_{it} = O_P(1)$.

Proof of Lemma B.1:

We first show that

$$
X'W_\delta^* X = \sum_{i=1}^{N} \sum_{t=1}^{T} K_h(\tau_t - \delta)(\eta_{it}\eta_{it}' - \bar{\eta}_i\bar{\eta}_i') + O_P(N(Th)^{3/2} + T^2 h + NTh),
\tag{B.4}
$$

where $\bar{\eta}_i = \frac{1}{Z_T} \sum_{t=1}^{T} K_h(\tau_t - \delta)\eta_{it}$.

By the definition of W_δ^*, $X'W_\delta^* X = X'W_\delta X - X'W_\delta D(D'W_\delta D)^{-1}D'W_\delta X$, in which

$$
\begin{aligned}
&X'W_\delta D(D'W_\delta D)^{-1}D'W_\delta X \\
&= \frac{1}{Z_T} \sum_{i=1}^{N} \Big[\sum_{t=1}^{T} X_{it}K_h(\tau_t - \delta) \Big] \Big[\sum_{t=1}^{T} X_{it}'K_h(\tau_t - \delta) \Big] \\
&\quad - \frac{1}{NZ_T} \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}K_h(\tau_t - \delta) \Big] \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}'K_h(\tau_t - \delta) \Big].
\end{aligned}
$$

In connection with (B.3), we are able to write

$$
\begin{aligned}
&X'W_\delta^* X \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'K_h(\tau_t - \delta) - \frac{1}{Z_T} \sum_{i=1}^{N} \Big[\sum_{t=1}^{T} X_{it}K_h(\tau_t - \delta) \Big] \Big[\sum_{t=1}^{T} X_{it}'K_h(\tau_t - \delta) \Big] \\
&\quad + \frac{1}{NZ_T} \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}K_h(\tau_t - \delta) \Big] \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}'K_h(\tau_t - \delta) \Big]
\end{aligned}
$$

$$
= \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - X_{i,\delta(T)})(X_{it} - X_{i,\delta(T)})' K_h(\tau_t - \delta)
$$

$$
- \frac{1}{Z_T} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} (X_{it} - X_{i,\delta(T)}) K_h(\tau_t - \delta) \right] \left[\sum_{t=1}^{T} (X_{it} - X_{i,\delta(T)})' K_h(\tau_t - \delta) \right]
$$

$$
+ \frac{1}{NZ_T} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} K_h(\tau_t - \delta) \right] \left[\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}' K_h(\tau_t - \delta) \right]
$$

$$
= \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\eta_{it} \eta_{it}' - \bar{\eta}_i \bar{\eta}_i' \right) K_h(\tau_t - \delta) + 2 \sum_{i=1}^{N} \sum_{t=1}^{T} \eta_{it} \xi_{it}' K_h(\tau_t - \delta) + \sum_{i=1}^{N} \sum_{t=1}^{T} \xi_{it} \xi_{it}' K_h(\tau_t - \delta)
$$

$$
- 2 Z_T \sum_{i=1}^{N} \bar{\eta}_i \bar{\xi}_i' - Z_T \sum_{i=1}^{N} \bar{\xi}_i \bar{\xi}_i' + N Z_T \bar{X} \bar{X}'
$$

$$
\equiv \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\eta_{it} \eta_{it}' - \bar{\eta}_i \bar{\eta}_i' \right) K_h(\tau_t - \delta) + 2 R_{1,NT} + R_{2,NT} - 2 R_{3,NT} - R_{4,NT} + R_{5,NT}, \quad \text{(B.5)}
$$

where $\bar{\xi}_i = \frac{1}{Z_T} \sum_{t=1}^{T} \xi_{it} K_h(\tau_t - \delta)$ and $\bar{X} = \frac{1}{NZ_T} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} K_h(\tau_t - \delta)$.

We now look at $R_{j,NT}$ for $j = 1, \ldots, 5$. Start from $R_{1,NT}$ and $R_{3,NT}$, and write

$$
E \| R_{1,NT} \| \leq \sum_{i=1}^{N} \sum_{t=1}^{T} K_h(\tau_t - \delta) E[\|\eta_{it}\| \cdot \|\xi_{it}\|]
$$

$$
\leq \sqrt{Th} \sum_{i=1}^{N} \sum_{t=1}^{T} K_h(\tau_t - \delta) \left\{ E \left\| \frac{\eta_{it}}{\sqrt{Th}} \right\|^2 E \|\xi_{it}\|^2 \right\}^{1/2}
$$

$$
= O(1) N \sqrt{Th} \sum_{t=1}^{T} K_h(\tau_t - \delta) = O(N(Th)^{3/2}),
$$

and

$$
E \| R_{3,NT} \| \leq \sum_{i=1}^{N} Z_T E[\|\bar{\eta}_i\| \cdot \|\bar{\xi}_i\|]
$$

$$
\leq O(1)(Th)^{5/2} Z_T^{-1} \sum_{i=1}^{N} \left\{ E \left\| \frac{1}{Th} \sum_{t=1}^{T} \frac{\eta_{it}}{\sqrt{Th}} K_h(\tau_t - \delta) \right\|^2 E \left\| \frac{1}{Th} \sum_{t=1}^{T} \xi_{it} K_h(\tau_t - \delta) \right\|^2 \right\}^{1/2}
$$

$$
= O(1)(Th)^{5/2} (Z_T)^{-1} = O(N(Th)^{3/2}).
$$

Similarly, we can show that $E\|R_{2,NT}\| = O(NTh)$ and $E\|R_{4,NT}\| = O(NTh)$.

For $R_{5,NT}$, it is easy to know $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} X_{i,\delta(T)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{\delta(T)} \Phi' \varepsilon_{it} = O_P(1)$. Thus, we have

$$
\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} K_h(\tau_t - \delta) = \sum_{i=1}^{N} X_{i,\delta(T)} \sum_{t=1}^{T} K_h(\tau_t - \delta) + \sum_{i=1}^{N} \sum_{t=1}^{T} (\eta_{it} + \xi_{it}) K_h(\tau_t - \delta)
$$

$$
= O_P(Th\sqrt{NT} + N^{1/2}(Th)^{3/2}),
$$

which yields that

$$R_{5,NT} = \frac{1}{NZ_T} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} K_h(\tau_t - \delta) \right] \left[\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}' K_h(\tau_t - \delta) \right]$$

$$= O_P(1) \frac{1}{NTh} \cdot \left((Th\sqrt{NT})^2 + N(Th)^3 \right) = O_P(T^2 h). \tag{B.6}$$

Based on the analyses from (B.5) to (B.6), (B.4) follows. Moreover, $O_P(N(Th)^{3/2} + T^2 h + NTh) = o_P(NT^2 h)$.

Next, we show that as $(N,T) \to (\infty, \infty)$,

$$\frac{1}{NT^2 h} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\eta_{it} \eta_{it}' - \bar{\eta}_i \bar{\eta}_i' \right) K_h(\tau_t - \delta) \to_P \Sigma_\nu (1 - \mathcal{C}_K),$$

which can be done by considering the first and second moments. For the first moment, we have

$$E \left[\frac{1}{NT^2 h} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\eta_{it} \eta_{it}' - \bar{\eta}_i \bar{\eta}_i' \right) K_h(\tau_t - \delta) \right]$$

$$= \frac{\Phi'}{NT^2 h} \sum_{t=1}^{T} K_h(\tau_t - \delta) \sum_{i=1}^{N} \sum_{s=\delta(T)+1}^{t} E\left[\varepsilon_{is} \varepsilon_{is}' \right] \Phi$$

$$- \frac{\Phi'}{NT^2 h} \sum_{t=1}^{T} K_h(\tau_t - \delta) \sum_{i=1}^{N} \frac{1}{Z_T^2} \sum_{\ell_1, \ell_2 = 1}^{T} K_h(\tau_{\ell_1} - \delta) K_h(\tau_{\ell_2} - \delta) \sum_{l=\delta(T)+1}^{\ell_1 \wedge \ell_2} E\left[\varepsilon_{il} \varepsilon_{il}' \right] \Phi$$

$$= \frac{\Phi'}{T^2 h} \sum_{t=1}^{T} K_h(\tau_t - \delta)(t - \delta(T)) \Lambda \Phi$$

$$- \frac{\Phi'}{T^2 h} \sum_{t=1}^{T} K_h(\tau_t - \delta) \frac{1}{Z_T^2} \sum_{\ell_1, \ell_2 = 1}^{T} K_h(\tau_{\ell_1} - \delta) K_h(\tau_{\ell_2} - \delta)(\ell_1 \wedge \ell_2 - \delta(T)) \Lambda \Phi$$

$$\to (1 - \mathcal{C}_K) \Phi' \Lambda \Phi = (1 - \mathcal{C}_K) \Sigma_\nu. \tag{B.7}$$

For the second moment, it is easy to know that

$$E \left\{ \frac{1}{NT^2 h} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \left(\eta_{it} \eta_{it}' - \bar{\eta}_i \bar{\eta}_i' \right) K_h(\tau_t - \delta) \right) \right.$$

$$\left. - E \left[\frac{1}{NT^2 h} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \left(\eta_{it} \eta_{it}' - \bar{\eta}_i \bar{\eta}_i' \right) K_h(\tau_t - \delta) \right) \right] \right\}^2 = o(1)$$

in view of the cross-sectional independence of $\{\varepsilon_{it}\}$.

Thus, the result of this lemma follows. ∎

Proof of Lemma B.2:

Similar to analyses of Lemma B.1, we have

$$
\begin{aligned}
X'W_\delta^* u &= \sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-X_{i,\delta(T)})u_{it}K_h(\tau_t-\delta) \\
&\quad -\frac{1}{Z_T}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(X_{it}-X_{i,\delta(T)})K_h(\tau_t-\delta)\right]\left[\sum_{t=1}^{T}u_{it}K_h(\tau_t-\delta)\right] \\
&\quad +\frac{1}{NZ_T}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}K_h(\tau_t-\delta)\right]\left[\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}K_h(\tau_t-\delta)\right] \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\eta_{it}+\xi_{it})u_{it}K_h(\tau_t-\delta) \\
&\quad -\frac{1}{Z_T}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}(\eta_{it}+\xi_{it})K_h(\tau_t-\delta)\right]\left[\sum_{t=1}^{T}u_{it}K_h(\tau_t-\delta)\right] \\
&\quad +\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{Z_T}K_h(\tau_t-\delta)X_{it}\right]\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}K_h(\tau_t-\delta)\right] \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(\eta_{it}+\xi_{it})(u_{it}-\bar{u}_i)K_h(\tau_t-\delta)+O_P(Th^{1/2}) \\
&:= \sum_{i=1}^{N}\sum_{t=1}^{T}S_{it}(u_{it}-\bar{u}_i)K_h(\tau_t-\delta)+O_P(Th^{1/2}),
\end{aligned}
$$

where $S_{it}=\eta_{it}+\xi_{it}$ and $\bar{u}_i=\frac{1}{Z_T}\sum_{t=1}^{T}K_h(\tau_t-\delta)u_{it}$, and the third equality follows from the development similar to those given in Lemma B.1.

Next, we need only to prove that

$$
\frac{1}{\sqrt{NT^2h}}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}S_{it}(u_{it}-\bar{u}_i)K_h(\tau_t-\delta)-N\sum_{t=1}^{T}K_h(\tau_t-\delta)(\Delta_{\nu u}-\Delta_{\nu\bar{u}}^t)\right\}
$$
$$
\Longrightarrow N(0,\mathcal{C}_*\Sigma_{\nu,u}). \tag{B.8}
$$

From the BN decomposition, we have for $t\ge\delta(T)$, $u_{it}=\bar{\bar{u}}_{it}+(\widetilde{u}_{i,t-1}-\widetilde{u}_{it})$, where $\bar{\bar{u}}_{it}=(\sum_{j=0}^{\infty}\psi_j')\varepsilon_{it}=\psi'\varepsilon_{it}$, $\widetilde{u}_{it}=\sum_{j=0}^{\infty}\widetilde{\psi}_j'\varepsilon_{i,t-j}$, and $\widetilde{\psi}_j=\sum_{k=j+1}^{\infty}\psi_k$. Let $\Delta\widetilde{u}_{it}=\widetilde{u}_{it}-\widetilde{u}_{i,t-1}$. Note

$$
\begin{aligned}
&\sum_{t=1}^{T}K_h(\tau_t-\delta)S_{it}\Delta\widetilde{u}_{it} \\
&=\sum_{t=1}^{T}K_h(\tau_t-\delta)S_{it}\widetilde{u}_{it}-\sum_{t=1}^{T}K_h(\tau_t-\delta)S_{it}\widetilde{u}_{i,t-1} \\
&=\sum_{t=1}^{T}K_h(\tau_t-\delta)S_{it}\widetilde{u}_{it}-\sum_{t=1}^{T}K_h(\tau_t-\delta)\nu_{it}\widetilde{u}_{i,t-1} \\
&\quad -\left\{\sum_{t=1}^{T}K_h(\tau_{t-1}-\delta)S_{i,t-1}\widetilde{u}_{i,t-1}+\sum_{t=1}^{T}[K_h(\tau_t-\delta)-K_h(\tau_{t-1}-\delta)]S_{i,t-1}\widetilde{u}_{i,t-1}\right\}
\end{aligned}
$$

39

$$= K_h(\tau_T - \delta)S_{iT}\widetilde{u}_{iT} - \sum_{t=1}^{T} K_h(\tau_t - \delta)\nu_{it}\widetilde{u}_{i,t-1} - \sum_{t=1}^{T} [K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta)]S_{i,t-1}\widetilde{u}_{i,t-1},$$

and

$$\sum_{t=1}^{T} K_h(\tau_t - \delta)\Delta\widetilde{u}_{it}$$

$$= \sum_{t=1}^{T} K_h(\tau_t - \delta)\widetilde{u}_{it} - \sum_{t=1}^{T} \{K_h(\tau_{t-1} - \delta) + (K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta))\}\widetilde{u}_{i,t-1}$$

$$= K_h(\tau_T - \delta)\widetilde{u}_{iT} - \sum_{t=1}^{T} [K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta)]\widetilde{u}_{i,t-1}.$$

By virtue of Assumption 1, $K_h(\tau_T - \delta) = 0$ with probability 1, which indicates that

$$\sum_{t=1}^{T} K_h(\tau_t - \delta)S_{it}(-\Delta\widetilde{u}_{it})$$

$$= \sum_{t=1}^{T} K_h(\tau_t - \delta)\nu_{it}\widetilde{u}_{i,t-1} + \sum_{t=1}^{T} [K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta)]S_{i,t-1}\widetilde{u}_{i,t-1}$$

and

$$\sum_{t=1}^{T} K_h(\tau_t - \delta)(-\Delta\widetilde{u}_{it}) = \sum_{t=1}^{T} [K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta)]\widetilde{u}_{i,t-1}.$$

Furthermore, let $V_{it} = \sum_{j=\delta(T)+1}^{t} \varepsilon_{ij}$, so we can also write $S_{it} = X_{it} - X_{i,\delta(T)} = \Phi'V_{it} + \widetilde{\nu}_{i,\delta(T)} - \widetilde{\nu}_{it}$. Thus, we can further obtain that

$$\sum_{t=1}^{T} K_h(\tau_t - \delta)S_{it}u_{it} = \sum_{t=1}^{T} K_h(\tau_t - \delta)\left(S_{it}\varepsilon_{it}'\psi - S_{it}\Delta\widetilde{u}_{it}\right)$$

$$= \sum_{t=1}^{T} K_h(\tau_t - \delta)(\Phi'V_{it} + \widetilde{\nu}_{i,\delta(T)} - \widetilde{\nu}_{it})\varepsilon_{it}'\psi$$

$$+ \sum_{t=1}^{T} K_h(\tau_t - \delta)\nu_{it}\widetilde{u}_{i,t-1} + \sum_{t=1}^{T} [K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta)]S_{i,t-1}\widetilde{u}_{i,t-1}.$$

Finally, we obtain that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\left(S_{it}u_{it} - \Delta_{\nu u}\right)$$

$$:= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\Phi'V_{i,t-1}\varepsilon_{it}'\psi$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(R_{iT}(1) + R_{iT}(2) + R_{iT}(3) + R_{iT}(4) + R_{iT}(5)\right) \tag{B.9}$$

where

$$R_{iT}(1) = \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\Phi'(\varepsilon_{it}\varepsilon_{it}' - \Lambda)\psi,$$

$$R_{iT}(2) = \frac{1}{T\sqrt{h}} \sum_{t=1}^{T-1} K_h(\tau_t - \delta)\left(\widetilde{u}_{it}\nu_{i,t+1} - \sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1}\right),$$

$$R_{iT}(3) = \frac{1}{T\sqrt{h}} \sum_{t=1}^{T}[K_h(\tau_t - \delta) - K_h(\tau_{t-1} - \delta)]S_{i,t-1}\widetilde{u}_{i,t-1},$$

$$R_{iT}(4) = \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\widetilde{\nu}_{i,\delta(T)}\varepsilon_{it}'\psi,$$

$$R_{iT}(5) = -\frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\psi'(\varepsilon_{it}\widetilde{\nu}_{it} - \widetilde{\Phi}_0).$$

Below, we show that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{iT}(k) = o_P(1)$ for $k = 1,\ldots,5$. Firstly, in view of the cross-sectional independence of $\{\varepsilon_{it}\}$, it is easy to obtain that

$$E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{iT}(1)\right\|^2 = o(1)$$

by some tedious calculation. Next, we consider $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{iT}(2)$, and write

$$E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{iT}(2)\right\|^2 = \frac{1}{N}\sum_{i,\ell=1}^{N} E[R_{iT}(2)'R_{\ell T}(2)]$$

$$= \frac{1}{NT^2h}\sum_{i,\ell=1}^{N}\sum_{t,s=1}^{T-1}\left\{K_h(\tau_t - \delta)K_h(\tau_s - \delta)\right.$$

$$\left.\cdot E\left[\left(\widetilde{u}_{it}\nu_{i,t+1} - \sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1}\right)'\left(\widetilde{u}_{\ell s}\nu_{\ell,s+1} - \sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1}\right)\right]\right\}$$

$$= \frac{1}{T^2h}\sum_{t,s=1}^{T-1} K_h(\tau_t - \delta)K_h(\tau_s - \delta)E\left[\left(\widetilde{u}_{it}\nu_{i,t+1} - \sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1}\right)'\left(\widetilde{u}_{is}\nu_{i,s+1} - \sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1}\right)\right]$$

$$= h\left\{E\|R_{iT}(2)\|^2 E\|R_{iT}(2)\|^2\right\}^{1/2} = hE\|R_{iT}(2)\|^2,$$

where the third equality follows from Assumption 3. Below, we focus on $E\|R_{iT}(2)\|^2$, so write

$$E\|R_{iT}(2)\|^2 = \text{tr}\left(E[R_{iT}(2)R_{iT}(2)']\right)$$

$$= \frac{1}{T^2h}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} K_h(\tau_t - \delta)K_h(\tau_s - \delta)$$

$$
E\left(
\begin{array}{c}
(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\widetilde{\psi}_j'\varepsilon_{i,t-j}\varepsilon_{i,t+1-k}'\Phi_k)(\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\Phi_p'\varepsilon_{i,s+1-p}\varepsilon_{i,s-q}'\widetilde{\psi}_q) \\
-(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})'
\end{array}
\right)
$$

$$
= \frac{2}{T^2 h}\sum_{t=\delta(T)+1}^{\lfloor 2Th\rfloor} K_h(\tau_t-\delta)\sum_{l=0}^{\lfloor 2Th\rfloor-1} K_h(\tau_{t+1}-\delta)
$$

$$
\left(
\begin{array}{c}
\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} E[(\widetilde{\psi}_j'\varepsilon_{i,t-j}\varepsilon_{i,t+1-k}'\Phi_k)(\Phi_p'\varepsilon_{i,t+l+1-p}\varepsilon_{i,t+l-q}'\widetilde{\psi}_q)] \\
-(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})'
\end{array}
\right)
$$

$$
= \frac{2}{T^2 h}\sum_{t=\delta(T)+1}^{\lfloor 2Th\rfloor} K_h^2(\tau_t-\delta)\sum_{l=0}^{\lfloor 2Th\rfloor-1}
$$

$$
\left(
\begin{array}{c}
\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} E[(\widetilde{\psi}_j'\varepsilon_{i,t-j}\varepsilon_{i,t+1-k}'\Phi_k)(\Phi_p'\varepsilon_{i,t+l+1-p}\varepsilon_{i,t+l-q}'\widetilde{\psi}_q)] \\
-(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})'
\end{array}
\right)
$$

$$
+\frac{2}{T^2 h}\sum_{t=\delta(T)+1}^{\lfloor 2Th\rfloor} K_h^2(\tau_t-\delta)\sum_{l=0}^{\lfloor 2Th\rfloor-1} (K_h(\tau_{t+l}-\delta)-K_h(\tau_t-\delta))
$$

$$
\left(
\begin{array}{c}
\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} E[(\widetilde{\psi}_j'\varepsilon_{i,t-j}\varepsilon_{i,t+1-k}'\Phi_k)(\Phi_p'\varepsilon_{i,t+l+1-p}\varepsilon_{i,t+l-q}'\widetilde{\psi}_q)] \\
-(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})'
\end{array}
\right)
$$

where $K_h(\tau_{(t+l)}-\delta)-K_h(\tau_t-\delta)=O(1)\frac{l}{Th}$. If we show

$$
\sum_{l=0}^{\infty}\left(
\begin{array}{c}
\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} E[(\widetilde{\psi}_j'\varepsilon_{i,t-j}\varepsilon_{i,t+1-k}'\Phi_k)(\Phi_p'\varepsilon_{i,t+l+1-p}\varepsilon_{i,t+l-q}'\widetilde{\psi}_q)] \\
-(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})'
\end{array}
\right) < \infty,
$$

then it follows that $E\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{iT}(2)\|^2 \to 0$. Observe that

$$
\sum_{l=0}^{\infty}\left(
\begin{array}{c}
\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} E[(\widetilde{\psi}_j'\varepsilon_{i,t-j}\varepsilon_{i,t+1-k}'\Phi_k)(\Phi_p'\varepsilon_{i,t+l+1-p}\varepsilon_{i,t+l-q}'\widetilde{\psi}_q)] \\
-(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})(\sum_{j=0}^{\infty}\widetilde{\psi}_j'\Phi_{j+1})'
\end{array}
\right)
$$

$$
= \sum_{l=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \text{tr}\{\Phi_k'\Phi_{k+l}\otimes\widetilde{\psi}_j'\widetilde{\psi}_{j+l}\} + \sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0\vee(1-l)}^{\infty} \text{tr}\left\{(\Phi_k'\widetilde{\psi}_{k+l-1}\otimes\widetilde{\psi}_j'\Phi_{j+l+1})K_d\right\}
$$

$$
+(\nu^4-3)\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\text{tr}\left\{(\Phi_{j+1}'\otimes\widetilde{\psi}_j')(\sum_{l=1}^{d+1} e_{l,l}\otimes e_{l,l})(\Phi_{j+l+1}\otimes\widetilde{\psi}_{j+l})\right\}
$$

$$
= \sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \text{tr}\{\Phi_k'\Phi_{k+l}\}\widetilde{\psi}_j'\widetilde{\psi}_{j+l} + \sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0\vee(1-l)}^{\infty} \widetilde{\psi}_{k+l-1}'\Phi_k\Phi_{j+l+1}'\widetilde{\psi}_j
$$

$$
+(\nu^4-3)\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\text{tr}\left\{(\Phi_{j+1}'\otimes\widetilde{\psi}_j')(\sum_{l=1}^{d+1} e_{l,l}\otimes e_{l,l})(\Phi_{j+l+1}\otimes\widetilde{\psi}_{j+l})\right\}
$$

$$
:= \text{I} + \text{II} + \text{III},
$$

where $e_{l,l}$ is the $(d \times d)$ matrix where the $(l,l)^{th}$ element is one and other elements are zeros. Since $\text{tr}(A \otimes B) = \text{tr}(A)\text{tr}(B)$ and $\text{tr}(A) \leq (\text{rows}(A))^{1/2}\|A\|$, we have

$$\begin{aligned}
\text{I} &= \sum_{l=0}^{\infty} \text{tr}\{\sum_{k=0}^{\infty} \Phi_k' \Phi_{k+l}\}(\sum_{j=0}^{\infty} \widetilde{\psi}_j' \widetilde{\psi}_{j+l}) \\
&\leq \left[\sum_{l=0}^{\infty} \left|\text{tr}(\sum_{k=0}^{\infty} \Phi_k' \Phi_{k+l})\right|\right] \left[\sum_{l=0}^{\infty} \left|\sum_{j=0}^{\infty} \text{tr}(\widetilde{\psi}_j \widetilde{\psi}_{j+l}')\right|\right] \\
&\leq d \left(\sum_{k=0}^{\infty} \|\Phi_k\|\right)^2 \left(\sum_{k=0}^{\infty} \|\widetilde{\psi}_k\|\right)^2 < \infty,
\end{aligned}$$

and

$$\begin{aligned}
\text{II} &\leq \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} \|\Phi_k\| \cdot \|\widetilde{\psi}_{k-1}\| \cdot \|\widetilde{\psi}_j\| \cdot \|\Phi_{j+1}\| + \sum_{l=1}^{\infty} \left(\sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \|\Phi_k\| \cdot \|\widetilde{\psi}_{k+l-1}\| \cdot \|\widetilde{\psi}_j\| \cdot \|\Phi_{j+l+1}\|\right) \\
&\leq \left(\sum_{j=0}^{\infty} \|\Phi_j\|\right)^2 \left(\sum_{j=0}^{\infty} \|\psi_j\|\right)^2 + \left(\sum_{l=0}^{\infty} \sum_{k=0}^{\infty} \|\Phi_k\| \cdot \|\widetilde{\psi}_{k+l}\|\right) \left(\sum_{l=0}^{\infty} \sum_{j=0}^{\infty} \|\widetilde{\psi}_j\| \cdot \|\Phi_{j+l}\|\right) \\
&\leq \left(\sum_{j=0}^{\infty} \|\Phi_j\|\right)^2 \left(\sum_{j=0}^{\infty} \|\psi_j\|\right)^2 + \left(\sum_{k=0}^{\infty} \|\Phi_k\| \sum_{k=0}^{\infty} \|\widetilde{\psi}_k\|\right) \left(\sum_{j=0}^{\infty} \|\widetilde{\psi}_j\| \sum_{j=0}^{\infty} \|\Phi_j\|\right) \\
&< \infty.
\end{aligned}$$

Similarly, we can show that $\text{III} < \infty$. Thus, we have shown that $E \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{iT}(2) \right\|^2 = o(1)$. By the similar arguments, we can also show that $E \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{iT}(j) \right\|^2 = o(1)$ for $j = 3, 4, 5$.

Therefore, we conclude that

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)(S_{it}u_{it} - \Delta_{\nu u}) \\
&= \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\Phi' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} V_{i,t-1}\varepsilon_{it}'\psi + o_P(1) \\
&= \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{s<t} K_h(\tau_t - \delta)\Phi' \varepsilon_{is}\varepsilon_{it}'\psi + o_P(1) \\
&:= \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} W_{1,tN} + o_P(1).
\end{aligned} \qquad (B.10)$$

Using arguments similar to (B.9), the following result also holds.

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\left(S_{it}\bar{u}_i - \Delta_{\nu\bar{u}}^t(\delta)\right)$$

43

$$= \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} K_h(\tau_t - \delta)\Phi' \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{Z_T} \sum_{l=1}^{T} \sum_{l \neq s \leq t} K_h(\tau_l - \delta)\varepsilon_{is}\varepsilon_{il}'\psi + o_P(1)$$

$$:= \frac{1}{T\sqrt{h}} \sum_{t=1}^{T} W_{2,tN} + o_P(1), \tag{B.11}$$

where the definition of $W_{2,tN}$ should be obvious.

By (B.10) and (B.11), we have

$$\frac{1}{\sqrt{NT^2h}} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} S_{it}(u_{it} - \bar{u}_i)K_h(\tau_t - \delta) - N \sum_{t=1}^{T} K_h(\tau_t - \delta)(\Delta_{\nu u} - \Delta_{\nu,\bar{u}}^t(\delta)) \right\}$$

$$= \frac{1}{T\sqrt{h}} \sum_{t=1}^{T}(W_{1,tN} - W_{2,tN}) + o_P(1) := \frac{1}{Th} \sum_{t=1}^{T} W_{tN} + o_P(1).$$

In order to prove (B.8), we need only to prove $\frac{1}{T\sqrt{h}} \sum_{t=1}^{T} W_{tN} \Longrightarrow N(0, \mathcal{C}_*\Sigma_{\nu,u})$, so write

$$W_{tN} := W_{1,tN} - W_{2,tN} = \sum_{i=1}^{N} \left(\frac{1}{\sqrt{N}} \sum_{s<t} \mathcal{W}_{t,s}^*(\delta)\Phi'\varepsilon_{is}\varepsilon_{it}'\psi \right),$$

where

$$\mathcal{W}_{t,s}^*(\delta) = K_h(\tau_t - \delta) - \mathcal{W}_{t,s}(\delta),$$

$$\mathcal{W}_{t,s} = \frac{1}{Z_T} \left(K_h(\tau_t - \delta) \sum_{l \geq s} K_h(\tau_l - \delta) + K_h(\tau_s - \delta) \sum_{l \geq t} K_h(\tau_l - \delta) \right).$$

Note that $\{W_{tN}, \mathcal{F}_{t,N}\}$ forms a martingale difference array with mean 0, where $\mathcal{F}_{t,N} = \sigma\{\varepsilon_{is} : 1 \leq i \leq N, 1 \leq s \leq t\}$ is a σ-filed. Next we use the central limit theory for martingale difference array. For any $\epsilon > 0$, it is easy to check that as $(N,T) \to (\infty, \infty)$,

$$\frac{1}{T^2h} E\left[W_{tN}W_{tN}'I(\|W_{tN}\|^2 \geq \epsilon T\sqrt{h})|\mathcal{F}_{t-1,N}\right] \to 0.$$

Furthermore, we can prove that

$$\frac{1}{T^2h} \sum_{t=1}^{T} E\left[W_{tN}W_{tN}'|\mathcal{F}_{t-1,N}\right]$$

$$= \frac{1}{T^2h} \sum_{t=1}^{T} \sum_{s<t} \mathcal{W}_{t,s}^{*2}(\delta) \frac{1}{N} \sum_{i,j=1}^{N} E\left[\Phi'\varepsilon_{is}\varepsilon_{it}'\psi\psi'\varepsilon_{jt}\varepsilon_{js}'\Phi|\mathcal{F}_{t-1,N}\right]$$

$$= \mathcal{C}_*\Sigma_{\nu,u} + o_P(1)$$

after some tedious calculation. The proof of Lemma B.2 is now completed. ∎

Proof of Lemma B.3.

Write

$$\widehat{\Gamma}_{i,\nu u}(j) = \frac{1}{\overline{\tau} - \underline{\tau}} \sum_{t=\underline{\tau}+1}^{\overline{\tau}} \nu_{i,t-j} \widehat{u}_{it}$$

$$= \frac{1}{\overline{\tau} - \underline{\tau}} \sum_{t=\underline{\tau}+1}^{\overline{\tau}} \nu_{i,t-j} u_{it} - \frac{1}{\overline{\tau} - \underline{\tau}} \sum_{t=\underline{\tau}+1}^{\overline{\tau}} \nu_{i,t-j} \{X_{it}'(\widehat{\beta}(\tau_t) - \beta_0(\tau_t)) + \widehat{\alpha}_i - \alpha_i\}$$

$$:= \overline{\Gamma}_{i,\nu u}(j) - \widetilde{\Gamma}_{i,\nu u}(j)$$

for $j = 0, 1, \cdots, l_T$. Thus, we can further write

$$\widehat{\Delta}_{i,\nu u} = \sum_{j=0}^{l_T} W\left(\frac{j}{l_T}\right) \overline{\Gamma}_{i,\nu u}(j) - \sum_{j=0}^{l_T} W\left(\frac{j}{l_T}\right) \widetilde{\Gamma}_{i,\nu u}(j) := \overline{\Delta}_{i,\nu u} + \widetilde{\Delta}_{i,\nu u}.$$

Below, we show

$$\sqrt{Nh}(\widehat{\Delta}_{\nu u} - \Delta_{\nu u}) = \sqrt{Nh}(\overline{\Delta}_{\nu u} - \Delta_{\nu u}) + \sqrt{Nh}\overline{\widetilde{\Delta}}_{\nu u} = o_P(1), \tag{B.12}$$

where $\overline{\Delta}_{\nu u} = \frac{1}{N}\sum_{i=1}^{N} \overline{\Delta}_{i,\nu u}$ and $\overline{\widetilde{\Delta}}_{\nu u} = \frac{1}{N}\sum_{i=1}^{N} \widetilde{\Delta}_{i,\nu u}$.

First, we turn to $\sqrt{Nh}(\overline{\Delta}_{\nu u} - \Delta_{\nu u}) = o_P(1)$ in (B.12). By Proposition 1 of Andrews (1991) and Assumption 5, we have

$$E\|\sqrt{Nh}(\overline{\Delta}_{\nu u} - \Delta_{\nu u})\|^2 = E\left\|\frac{\sqrt{h}}{\sqrt{N}}\sum_{i=1}^{N}(\overline{\Delta}_{i,\nu u} - E[\overline{\Delta}_{i,\nu u}] + E[\overline{\Delta}_{i,\nu u}] - \Delta_{\nu u})\right\|^2$$

$$= hE\left\|\overline{\Delta}_{i,\nu u} - E[\overline{\Delta}_{i,\nu u}]\right\|^2 + Nh\|E[\overline{\Delta}_{i,\nu u}] - \Delta_{\nu u}\|^2$$

$$= O\left(\frac{l_T h}{\overline{\tau} - \underline{\tau}} + \frac{Nh}{l_T^{2q}}\right) = o(1). \tag{B.13}$$

Second, by the construction of $\overline{\tau}$ and $\underline{\tau}$ and a development similar to (A.31)–(A.33) of Phillips et al. (2017), we have

$$\sqrt{Nh}\overline{\widetilde{\Delta}}_{\nu u} = o_P(1). \tag{B.14}$$

Therefore, we prove (B.12) by (B.14) and (B.13).

Similarly, the second result follows. The proof is now completed. ∎

Proof of Lemma B.4:

Write

$$\ddot{\beta}(\delta) - \beta_0(\delta) = \Pi_{1NT} + \Pi_{2NT} + \Pi_{3NT} - \beta_0(\delta), \tag{B.15}$$

45

where

$$\Pi_{0NT} = \frac{1}{NT^2h} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X_{it}' K_h(\tau_t - \delta),$$

$$\Pi_{1NT} = \Pi_{0NT}^{-1} \frac{1}{NT^2h} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} u_{it} K_h(\tau_t - \delta),$$

$$\Pi_{2NT} = \Pi_{0NT}^{-1} \frac{1}{NT^2h} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} \lambda_{0i}' f_{0t} K_h(\tau_t - \delta),$$

$$\Pi_{3NT} = \Pi_{0NT}^{-1} \frac{1}{NT^2h} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X_{it}' \beta_0(\tau_t) K_h(\tau_t - \delta).$$

By Assumption 6, it is easy to know that $\Pi_{1NT} + \Pi_{2NT} = O_P\left(\frac{1}{\sqrt{T}}\right)$. By the proof of Theorem 1, $\Pi_{3NT} - \beta_0(\delta) = O_P(h^2)$. The result then follows. ∎

Proof of Lemma B.5:

(1). Write

$$\sup_{F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} u_i' \Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta u_i \right| = \sup_{F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} \text{tr}\{\Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta u_i u_i'\} \right|$$

$$= \sup_{F \in D_F} \frac{1}{NT} \left| \text{tr}\{\Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta U' U\} \right| \leq \frac{r}{NT} \|\Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta U' U\|_{\text{sp}}$$

$$\leq O(1) \frac{1}{NTh} \|U'U\| = o(1),$$

where the first inequality follows from $\text{tr}|A| \leq \text{rank}(A) \cdot \|A\|_{\text{sp}}$, the second inequality follows from the uniform boundedness of the kernel function, and the last equality follows from Assumptions 4 and 7.

(2). Write

$$\sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i\beta)' \Upsilon_\delta M_{\Upsilon_\delta F} \Upsilon_\delta u_i \right|$$

$$\leq \sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i\beta)' \Upsilon_\delta^2 u_i \right|$$

$$+ \sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i\beta)' \Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta u_i \right| := A_1 + A_2.$$

By the construction of $R_\delta(\beta)$ and Assumption 6.2, it is easy to know that $A_1 = o_P(1)$.

For A_2, write

$$\sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i\beta)' \Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta u_i \right|$$

$$= \sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \sum_{i=1}^N \text{tr}\{\Upsilon_\delta P_{\Upsilon_\delta F} \Upsilon_\delta u_i \Delta\phi'_{i,\beta}\} \right|$$

$$= \sup_{\beta \in R_\delta(\beta), F \in D_F} \left| \frac{1}{NT} \text{tr}\{P_{\Upsilon_\delta F} \Upsilon_\delta U' \Delta\phi_\beta \Upsilon_\delta\} \right|$$

$$\leq \sup_{\beta \in R_\delta(\beta)} \frac{r}{NT} \|\Upsilon_\delta U' \Delta\phi_\beta \Upsilon_\delta\|_{\text{sp}} \leq O(1)A_1 = o_P(1),$$

where $\Delta\phi_{i,\beta} = \phi_i[\beta_0] - X_i\beta$ and $\Delta\phi_\beta = (\phi_1[\beta_0], \ldots, \phi_N[\beta_0])'$.

(3) The third result follows from a procedure similar to (2) of this lemma. ∎

Proof of Lemma B.6:

(1). We now consider V_{NT} and write

$$\Upsilon_\delta \widetilde{F} V_{NT} = \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta(\phi_i[\beta_0] - X_i\widetilde{\beta}(\delta))(\phi_i[\beta_0] - X_i\widetilde{\beta}(\delta))' \Upsilon_\delta^2 \widetilde{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta(\phi_i[\beta_0] - X_i\widetilde{\beta}(\delta))(F_0\lambda_{0i})' \Upsilon_\delta^2 \widetilde{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta F_0\lambda_{0i}(\phi_i[\beta_0] - X_i\widetilde{\beta}(\delta))' \Upsilon_\delta^2 \widetilde{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta(\phi_i[\beta_0] - X_i\widetilde{\beta}(\delta))u_i' \Upsilon_\delta^2 \widetilde{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta u_i(\phi_i[\beta_0] - X_i\widetilde{\beta}(\delta))' \Upsilon_\delta^2 \widetilde{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta u_i u_i' \Upsilon_\delta^2 \widetilde{F} + \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta F_0\lambda_{0i} u_i' \Upsilon_\delta^2 \widetilde{F}$$

$$+ \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta u_i(F_0\lambda_{0i})' \Upsilon_\delta^2 \widetilde{F} + \frac{1}{NT} \sum_{i=1}^N \Upsilon_\delta(F_0\lambda_{0i})(F_0\lambda_{0i})' \Upsilon_\delta^2 \widetilde{F}$$

$$:= I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{5NT}(\widetilde{\beta}(\delta), \widetilde{F}) + I_{6NT}(\widetilde{F}) + \cdots + I_{9NT}(\widetilde{F}).$$

Note that $I_{9NT}(\widetilde{F}) = \Upsilon_\delta F_0 \cdot \frac{\Lambda_0'\Lambda_0}{N} \cdot \frac{F_0'\Upsilon_\delta^2\widetilde{F}}{T}$. Thus, we can write

$$\Upsilon_\delta \widetilde{F} V_{NT} - \Upsilon_\delta F_0 \frac{\Lambda_0'\Lambda_0}{N} \cdot \frac{F_0'\Upsilon_\delta^2\widetilde{F}}{T} = I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{8NT}(\widetilde{F}). \tag{B.16}$$

Note that the development of Theorem 3 ensures that $\frac{F_0'\Upsilon_\delta^2\widetilde{F}}{T}$ is non-singular, so rewrite (B.16) as

$$\Upsilon_\delta \widetilde{F} V_{NT} \left(\frac{F_0'\Upsilon_\delta^2\widetilde{F}}{T} \right)^{-1} \left(\frac{\Lambda_0'\Lambda_0}{N} \right)^{-1} - \Upsilon_\delta F_0$$

$$= \left[I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{8NT}(\widetilde{F}) \right] \left(\frac{F_0'\Upsilon_\delta^2\widetilde{F}}{T} \right)^{-1} \left(\frac{\Lambda_0'\Lambda_0}{N} \right)^{-1}. \tag{B.17}$$

We then examine each of the terms on the right hand side of (B.17) and show that V_{NT} is non-singular. In order to do so, we write

$$\left\| \Upsilon_\delta \widetilde{F} V_{NT} \left(\frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T} \right)^{-1} \left(\frac{\Lambda_0' \Lambda_0}{N} \right)^{-1} - \Upsilon_\delta F_0 \right\|$$
$$\leq O_P(1) \left(\| I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) \| + \cdots + \| I_{8NT}(\widetilde{F}) \| \right), \qquad (B.18)$$

and focus on each term on the right hand side of (B.18).

Note that by construction, we have $\frac{1}{\sqrt{T}} \| \Upsilon_\delta \widetilde{F} \| = O(1)$. We will repeatedly use this fact below. Start from $I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F})$ and write

$$\frac{1}{\sqrt{T}} \| I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) \| \leq O(1) \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i \widetilde{\beta}(\delta))' \Upsilon_\delta^2 (\phi_i[\beta_0] - X_i \widetilde{\beta}(\delta))$$
$$\leq O(1) \frac{1}{NT} \sum_{i=1}^{N} (\beta_0(\delta) - \widetilde{\beta}(\delta))' X_i' \Upsilon_\delta^2 X_i (\beta_0(\delta) - \widetilde{\beta}(\delta))$$
$$+ O(1) \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i \beta_0(\delta))' \Upsilon_\delta^2 (\phi_i[\beta_0] - X_i \beta_0(\delta))$$
$$= O_P(T \| \widetilde{\beta}(\delta) - \beta_0(\delta) \|^2) + O_P(T h^4),$$

where the term $O_P(T \| \widetilde{\beta}(\delta) - \beta_0(\delta) \|^2)$ follows from the development of Theorem 3, and $O_P(T h^4)$ follows from the standard argument of the kernel method.

For $I_{2NT}(\widetilde{\beta}(\delta), \widetilde{F})$, write

$$\frac{1}{\sqrt{T}} \| I_{2NT}(\widetilde{\beta}(\delta), \widetilde{F}) \| \leq O(1) \left\{ \frac{1}{NT} \sum_{i=1}^{N} (\phi_i[\beta_0] - X_i \widetilde{\beta}(\delta))' \Upsilon_\delta^2 (\phi_i[\beta_0] - X_i \widetilde{\beta}(\delta)) \right\}^{1/2}$$
$$\cdot \left\{ \frac{1}{NT} \sum_{i=1}^{N} \lambda_{0i}' F_0' \Upsilon_\delta^2 F_0 \lambda_{0i} \right\}^{1/2}$$
$$= O_P(\sqrt{T} \| \widetilde{\beta}(\delta) - \beta_0(\delta) \|) + O_P(\sqrt{T} h^2),$$

where the last line follows from the proof for $I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F})$, and $\frac{1}{NT} \sum_{i=1}^{N} \lambda_{0i}' F_0' \Upsilon_\delta^2 F_0 \lambda_{0i} = O_P(1)$ by Assumption 7.

Similar to $I_{2NT}(\widetilde{\beta}(\delta), \widetilde{F})$, we can show for $j = 3, 4, 5$,

$$\frac{1}{\sqrt{T}} \| I_{jNT}(\widetilde{\beta}(\delta), \widetilde{F}) \| = O_P(\sqrt{T} \| \widetilde{\beta}(\delta) - \beta_0(\delta) \|) + O_P(\sqrt{T} h^2).$$

We now consider $I_{6NT}(\widetilde{\beta}(\delta), \widetilde{F})$, and write

$$E \left\| \frac{1}{NT} \sum_{i=1}^{N} \Upsilon_\delta u_i u_i' \Upsilon_\delta \right\|^2$$

48

$$
= \frac{1}{N^2 T^2 h^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} E[u_{it} u_{is} u_{jt} u_{js}] K_h(\tau_t - \delta) K_h(\tau_s - \delta)
$$

$$
= \frac{1}{N^2 T^2 h^2} \sum_{t=1}^{T} \sum_{i,j=1}^{N} E[u_{it}^2 u_{jt}^2] K_h^2(\tau_t - \delta)
$$

$$
+ \frac{1}{N^2 T^2 h^2} \sum_{t \neq s} \sum_{i,j=1}^{N} E[u_{it} u_{jt}] E[u_{is} u_{js}] K_h(\tau_t - \delta) K_h(\tau_s - \delta)
$$

$$
= O(1) \frac{1}{T^2 h^2} \sum_{t=1}^{T} K_h^2(\tau_t - \delta) + \frac{1}{N^2} \sum_{i,j=1}^{N} \sigma_{ij}^2 \cdot \frac{1}{T^2 h^2} \sum_{t \neq s} K_h(\tau_t - \delta) K_h(\tau_s - \delta)
$$

$$
= O\left(\frac{1}{Th}\right) + O\left(\frac{1}{N}\right),
$$

which immediately gives

$$
\frac{1}{\sqrt{T}} \| I_{6NT}(\widetilde{\beta}(\delta), \widetilde{F}) \| = O_P\left(\frac{1}{\sqrt{Th}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right).
$$

For $I_{7NT}(\widetilde{F})$ and $I_{8NT}(\widetilde{F})$, write

$$
E \left\| \frac{1}{NT} \sum_{i=1}^{N} \Upsilon_\delta F_0 \lambda_{0i} u_i' \Upsilon_\delta \right\|^2
$$

$$
= \frac{1}{N^2 T^2 h^2} \sum_{t,s=1}^{T} \sum_{i,j=1}^{N} K_h(\tau_t - \delta) K_h(\tau_s - \delta) E[f_{0t}' \lambda_{0i} u_{is} f_{0t}' \lambda_{0j} u_{js}]
$$

$$
\leq O(1) \frac{1}{T^2 h^2} \sum_{t,s=1}^{T} K_h(\tau_t - \delta) K_h(\tau_s - \delta) \cdot \frac{1}{N^2} \sum_{i,j=1}^{N} |\sigma_{ij}| = O\left(\frac{1}{N}\right).
$$

Then we can conclude that

$$
\frac{1}{\sqrt{T}} \| I_{7NT}(\widetilde{F}) \| = O_P\left(\frac{1}{\sqrt{N}}\right) \quad \text{and} \quad \frac{1}{\sqrt{T}} \| I_{8NT}(\widetilde{F}) \| = O_P\left(\frac{1}{\sqrt{N}}\right).
$$

Based on the above derivations and by multiplying (B.17) with $\frac{F_0' \Upsilon_\delta}{T}$ from the left–hand side, we obtain that

$$
\frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T} V_{NT} = \frac{F_0' \Upsilon_\delta^2 F_0}{T} \cdot \frac{\Lambda_0' \Lambda_0}{N} \cdot \frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T} + o_P(1),
$$

which implies $\frac{F_0' \Upsilon_\delta^2 \widetilde{F}}{T}$ are the eigenvectors of the matrix $\frac{F_0' \Upsilon_\delta^2 F_0}{T} \cdot \frac{\Lambda_0' \Lambda_0}{N}$, and V_{NT} consists of the eigenvalues of the same matrix (in the limit) by noting that

$$
\frac{F_0' \Upsilon_\delta^2 F_0}{T} \cdot \frac{\Lambda_0' \Lambda_0}{N} \to_P \Sigma_f(\delta) \Sigma_\lambda
$$

after simple algebra. Thus, the result follows.

(2). Based on the above development, (B.18) can be summarized by

$$\frac{1}{\sqrt{T}}\|\Upsilon_\delta \widetilde{F}\Pi_{NT}^{-1} - \Upsilon_\delta F_0\| = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2)$$

$$+O_P\left(\frac{1}{\sqrt{T}h}\right) + O_P\left(\frac{1}{\sqrt{N}}\right).$$

(3). According to (B.17),

$$\frac{1}{T}F_0'\Upsilon_\delta^2\widetilde{F} - \frac{1}{T}F_0'\Upsilon_\delta^2 F_0\Pi_{NT} = \frac{1}{T}F_0'\Upsilon_\delta\left[I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{8NT}(\widetilde{F})\right]V_{NT}^{-1}.$$

Note that $V_{NT}^{-1} = O_P(1)$, so we focus on $\frac{1}{T}F_0'\Upsilon_\delta\left[I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{8NT}(\widetilde{F})\right]$ below. By the proof of the first result of this lemma, it is easy to know that

$$\frac{1}{T}\left\|F_0'\Upsilon_\delta\left[I_{1NT}(\widetilde{\beta}(\delta), \widetilde{F}) + \cdots + I_{5NT}(\widetilde{\beta}(\delta), \widetilde{F})\right]\right\|$$
$$= O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2).$$

We now consider $\frac{1}{T}F_0'\Upsilon_\delta I_{6NT}(\widetilde{F})$. Write

$$\frac{1}{T}F_0'\Upsilon_\delta I_{6NT}(\widetilde{F}) \le \frac{1}{T}\left\{\frac{1}{NT}\sum_{i=1}^{N}\|F_0'\Upsilon_\delta^2 u_i\|^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\|u_i'\Upsilon_\delta^2\widetilde{F}\|^2\right\}^{1/2}.$$

Simple algebra shows that $\frac{1}{NT}\sum_{i=1}^{N}\|F_0'\Upsilon_\delta^2 u_i\|^2 = O_P(1)$, so we focus on $\frac{1}{NT}\sum_{i=1}^{N}\|u_i'\Upsilon_\delta^2\widetilde{F}\|^2$ below. Write

$$\frac{1}{NT}\sum_{i=1}^{N}\|u_i'\Upsilon_\delta^2\widetilde{F}\|^2 = \frac{1}{NT}\sum_{i=1}^{N}\|u_i'\Upsilon_\delta^2(F_0\Pi_{NT} + \widetilde{F} - F_0\Pi_{NT})\|^2$$

$$\le \frac{2}{NT}\sum_{i=1}^{N}\|u_i'\Upsilon_\delta^2 F_0\Pi_{NT}\|^2 + \frac{2}{NT}\sum_{i=1}^{N}\text{tr}\left\{u_i'\Upsilon_\delta^2(\widetilde{F} - F_0\Pi_{NT})(\widetilde{F} - F_0\Pi_{NT})'\Upsilon_\delta^2 u_i\right\}$$

$$\le \frac{2}{NT}\sum_{i=1}^{N}\|u_i'\Upsilon_\delta^2 F_0\Pi_{NT}\|^2 + \frac{2}{NT}\text{tr}\left\{\Upsilon_\delta^2(\widetilde{F} - F_0\Pi_{NT})(\widetilde{F} - F_0\Pi_{NT})'\Upsilon_\delta^2 U'U\right\}$$

$$\le O_P(1) + O(1)\frac{1}{N}\|U'U\| \cdot \frac{1}{T}\|\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT}\|^2.$$

In connection with (1) of Lemma B.5 and result (2) of this lemma, it then gives that

$$\frac{1}{T}F_0'\Upsilon_\delta I_{6NT}(\widetilde{F})$$

$$= O_P\left(\frac{1}{T}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)\left\{\frac{1}{NT}\|U'U\| \cdot \frac{1}{T}\|\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT}\|^2\right\}^{1/2}$$

$$= O_P\left(\frac{1}{T}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)O_P\left(\frac{1}{\sqrt[4]{N}} + \frac{1}{\sqrt[4]{T}}\right)\left\{O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|)\right.$$

$$\left. +O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{\sqrt{T}h}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)\right\}.$$

For $\frac{1}{T}F_0'\Upsilon_\delta I_{7NT}(\widetilde{F})$, we have

$$\frac{1}{T}\|F_0'\Upsilon_\delta I_{7NT}(\widetilde{F})\| \leq \frac{1}{T}\|F_0'\Upsilon_\delta^2 F_0\| \cdot \left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N \lambda_{0i}u_i'\Upsilon_\delta\right\| \cdot \frac{1}{\sqrt{T}}\|\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT}\|$$

$$+ \frac{1}{T}\|F_0'\Upsilon_\delta^2 F_0\| \cdot \left\|\frac{1}{NT}\sum_{i=1}^N \lambda_{0i}u_i'\Upsilon_\delta^2 F_0\right\| \cdot \|\Pi_{NT}\|.$$

By Assumption 7, $\frac{1}{T}\|F_0'\Upsilon_\delta^2 F_0\| = O_P(1)$. Also, $\|\Pi_{NT}\|$ and $\frac{1}{\sqrt{T}}\|\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT}\|$ have been studied above. Thus, we focus on $\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N \lambda_{0i}u_i'\Upsilon_\delta\right\|$ and $\left\|\frac{1}{NT}\sum_{i=1}^N \lambda_{0i}u_i'\Upsilon_\delta^2 F_0\right\|$ below. Write

$$E\left\|\frac{1}{N\sqrt{T}}\sum_{i=1}^N \lambda_{0i}u_i'\Upsilon_\delta\right\|^2 = \frac{1}{N^2 Th}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T E[\lambda_{0i}'\lambda_{0j}u_{it}u_{jt}]K_h(\tau_t - \delta) = O\left(\frac{1}{N}\right).$$

Similarly, by Assumption 8, it is easy to know that

$$E\left\|\frac{1}{NT}\sum_{i=1}^N \lambda_{0i}u_i'\Upsilon_\delta^2 F_0\right\|^2 = O\left(\frac{1}{NTh}\right).$$

We then obtain that

$$\frac{1}{T}\|F_0'\Upsilon_\delta I_{7NT}(\widetilde{F})\| \leq O_P\left(\frac{1}{\sqrt{N}}\right)\left\{O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2)\right.$$

$$\left. + O_P\left(\frac{1}{\sqrt{Th}}\right) + O_P\left(\frac{1}{\sqrt{N}}\right)\right\} + O_P\left(\frac{1}{\sqrt{NTh}}\right).$$

Based on the above development, we then obtain that

$$\frac{1}{T}\|F_0'\Upsilon_\delta^2\widetilde{F} - F_0'\Upsilon_\delta^2 F_0\Pi_{NT}\|$$

$$= O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O_P\left(\frac{1}{N}\right), \qquad (\text{B.19})$$

which further indicates that

$$\frac{1}{T}\|\widetilde{F}'\Upsilon_\delta^2\widetilde{F} - \widetilde{F}'\Upsilon_\delta^2 F_0\Pi_{NT}\|$$

$$\leq \frac{1}{T}\|(\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT})'(\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT})\|$$

$$+ \frac{1}{T}\|(\Upsilon_\delta F_0\Pi_{NT})'(\Upsilon_\delta\widetilde{F} - \Upsilon_\delta F_0\Pi_{NT})\|$$

$$= O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O\left(\frac{1}{N}\right). \qquad (\text{B.20})$$

(4). Note that (B.19) and (B.20) give that

$$\frac{1}{T}\Pi_{NT}'F_0'\Upsilon_\delta^2\widetilde{F} - \frac{1}{T}\Pi_{NT}'F_0'\Upsilon_\delta^2 F_0\Pi_{NT}$$

$$= O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O_P\left(\frac{1}{N}\right)$$

and

$$I_r - \frac{1}{T}\widetilde{F}'\Upsilon_\delta^2 F_0 \Pi_{NT} = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O\left(\frac{1}{N}\right).$$

Summing up the above two equations yields that

$$I_r - \frac{1}{T}\Pi'_{NT}F_0'\Upsilon_\delta^2 F_0 \Pi_{NT} = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O_P\left(\frac{1}{N}\right).$$

Note that it is easy to know that

$$\|P_{\Upsilon_\delta \widetilde{F}} - P_{\Upsilon_\delta F_0}\|^2 = \mathrm{tr}\left\{P_{\Upsilon_\delta \widetilde{F}} + P_{\Upsilon_\delta F_0} - P_{\Upsilon_\delta \widetilde{F}}P_{\Upsilon_\delta F_0} - P_{\Upsilon_\delta F_0}P_{\Upsilon_\delta \widetilde{F}}\right\}$$
$$= 2\mathrm{tr}\left\{I_r - \frac{1}{T}\widetilde{F}'\Upsilon_\delta P_{\Upsilon_\delta F_0}\Upsilon_\delta \widetilde{F}\right\}.$$

Moreover, using (B.19), we obtain that

$$\frac{1}{T}\widetilde{F}'\Upsilon_\delta P_{\Upsilon_\delta F_0}\Upsilon_\delta \widetilde{F} = \frac{1}{T}\widetilde{F}'\Upsilon_\delta^2 F_0(F_0'\Upsilon_\delta^2 F_0)^{-1}F_0'\Upsilon_\delta^2 \widetilde{F}$$
$$= \frac{1}{T}\Pi'_{NT}F_0'\Upsilon_\delta^2 F_0(F_0'\Upsilon_\delta^2 F_0)^{-1}F_0'\Upsilon_\delta^2 F_0 \Pi_{NT}$$
$$+ O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O_P\left(\frac{1}{N}\right)$$
$$= \frac{1}{T}\Pi'_{NT}F_0'\Upsilon_\delta^2 F_0 \Pi_{NT}$$
$$+ O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O_P\left(\frac{1}{N}\right).$$

Thus, we have

$$\|P_{\Upsilon_\delta \widetilde{F}} - P_{\Upsilon_\delta F_0}\|^2 = O_P(\sqrt{T}\|\widetilde{\beta}(\delta) - \beta_0(\delta)\|) + O_P(\sqrt{T}h^2) + O_P\left(\frac{1}{Th}\right) + O_P\left(\frac{1}{N}\right).$$

The proof is now completed. ∎