DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Modelling Time-Varying Income Elasticities of Health Care Expenditure for the OECD

Isabel Casas, Jiti Gao and Shangyu Xie

CREATES Research Paper 2018-29

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Modelling Time–Varying Income Elasticities of Health Care Expenditure for the OECD

Isabel Casas* Jiti Gao† Shangyu Xie‡

November 21, 2018

Abstract

Income elasticity dynamics of health expenditure is considered for the OECD and the Eurozone over the period 1995-2014. This paper studies a novel non-linear cointegration model with fixed effects, controlling for cross-section dependence and unobserved heterogeneity. Most importantly, its coefficients can vary over time and its variables can be non-stationary. The resulting asymptotic theory is fundamentally different with a faster rate of convergence to similar kernel smoothing methodologies. A fully modified kernel regression method is also proposed to reduce the asymptotic bias. Results show a steep increase in the income elasticity for the OECD and a small increase for the Eurozone.

Keywords: Cross-sectional dependence, Health expenditure, Income elasticity, Nonparametric kernel smoothing, Non-stationarity, Super-consistency.
JEL Classification: C14, C23, G13, H51

1. Introduction

Panel data analysis has received a growing attention during the last two decades due to its suitability for a wide number of applied disciplines, such as economics, finance and biology. There exists a wealth of literature on parametric linear and non-linear panel data models

*Department of Business and Economics, The University of Southern Denmark, Denmark. icasas@sam.sdu.org

†Department of Econometrics and Business Statistics, Monash University, Australia. jiti.gao@monash.edu

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

(see Baltagi, 1995; Arellano, 2003; Hsiao, 2003). However, it is well known that parametric panel data models may be easily misspecified with inconsistent estimations due to cross-section dependence, non-stationarity and unobserved heterogeneity. Many of these issues can be addressed using nonparametric methods as proposed by Fan and Li (2004); Hjellvik et al. (2004); Cai and Li (2008); Zhang et al. (2009).

Meanwhile, trending econometric modelling of non-stationary processes has also gained a great deal of attention in recent years. Trends are the dominant characteristic in most economic, financial and climate data, and therefore cointegration models are now one of the most commonly used frameworks to capture long term relationships among trending macroeconomic time series. Thus, Phillips (2001) provide a review on the development and challenges on trends modelling which is often impossible to be explained by parametric models. In this regard, extensive literature focuses on time-varying coefficient trending models using nonparametric and semi-parametric estimation methods. The latter include linear models with coefficients that change as a function of a time scale. First, Robinson (1989) studies linear regression models with time-varying coefficients for stationary processes, which is generalized to non-stationary processes and correlated errors by Chang and Martinez-Chombo (2003) and Cai (2007a) amongst others. Gao and Hawthorne (2006) propose using a semi-parametric time series specification to model the trend in global and hemispheric temperature series while at the same time allowing for the inclusion of some explanatory variables in a parametric component. In the oceanography literature, Reikard (2009) has used these trends modelling in oceanic energy production. Chen et al. (2017) have recently applied it to an autoregressive model of the realized volatility of S&P 500 index returns. Kristensen (2012); Orbe et al. (2005); Phillips et al. (2017); Casas et al. (2017) study multi-equations cases. Finally, Phillips et al. (2017) study non-linear cointegration models in which the structural coefficients may evolve smoothly over time, giving two different limit distributions with different convergence rates in the different directions of the functional parametric space. Both rates are faster than the usual root-nh rate.

As far as we know, little work in this vast literature has been done on estimating the time-varying trend function in a panel data model. This paper extends the work of Phillips et al. (2017) to panel data and it proposes a non-linear cointegration model with time-varying coefficients and fixed effects. The aim is to describe the non-linear trending phenomenon in panel data analysis, likewise allowing for cross-section dependence on both the regressors and the residuals, as well as non-stationarity of the regressors. More specifically, we consider the following model,

$$\begin{aligned} Y_{i,t} =& \mathbf{X}'_{i,t}\beta_t + \alpha_i + u_{i,t}, \\ \mathbf{X}_{i,t} =& \mathbf{X}_{i,t-1} + \nu_{i,t}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T, \end{aligned} \tag{1}$$

where β_t are unknown functions of t/T and α_i reflects the unobservable individual effects. In summary, model (1) captures potential drifts in the relationship between Y_t and X_t over time. Such a modelling structure is especially useful for time series data over long horizons where economic mechanism are likely to evolve and be subjected to institutional change or regulatory conditions. A clear example is the evolution of the cost of health care in developed countries whose alarming increase in the past decades can risk its sustainability. This concern has escalated after the global financial crisis (GFC), especially in the Eurozone where many health care systems are funded by taxes. That funding has dropped alongside salaries and employment rate. We answer two main questions based on our results: (1) Whether the price of health care has changed over the last two decades in the OECD and Eurozone; and (2) whether the post-GFC health policies in the Eurozone have achieved their objective of creating more efficient health care systems in the Eurozone.

Model (1) is a fixed effects model where α_i is allowed to be correlated with $\mathbf{X}_{i,t}$ with an unknown correlation structure. As fixed effects are involved in panel data models, the developed nonparametric and semi-parametric procedures eliminate the influence of these fixed effects by treating them as nuisance parameters to obtain unbiased estimates of the model coefficients.

The objective of the theoretical part of this paper is to construct estimates for the time-varying functional coefficient vector β_t and establishing its asymptotic properties. As in Phillips et al. (2017), a pooled kernel-weighted estimation method into the panel data framework is proposed to eliminate the bias arising from the correlation between the regressor innovations and the error term.

We consider both large T and N with $Nh \to \infty$, $N/Th \to 0$ and h, the model bandwidth to establish the asymptotic theory. The latter condition indicates that the limit theory is mostly useful for moderate values of N and large values of T. Generally speaking, such a joint limit theory requires stronger conditions to establish the sequential convergence or diagonal path convergence. As both the time series length T and the cross-section size N tend to infinity, the resulting estimator is asymptotically normal with root-(N^2Th) converge rate which is faster than the usual root-NTh rate for non-linear models with smoothly changing coefficients and local stationarity variables.

Classical fixed effect (FE) models help to control for unobserved heterogeneity, by assuming that this heterogeneity is constant over time. This assumption might not be reasonable for a large time series. Thus, Pesaran (2006), Bai (2009) and Kneip et al. (2012) developed panel data models with a heterogeneous factor structure in the error term. Recently, Baltagi and Moscone (2010) use this new methodology to estimate the long-run (constant) income elasticity of health care expenditure in the OECD. Their findings show that the income elasticity decreases from around 0.9 to around 0.7 when heterogenous unobserved factors are added to the FE model, implying that health care is a necessity good. The income elasticity decreases to around 0.45 when variables measuring the population age structure and the rate of public funding are included. In this paper, heterogenous unobserved factors are added to the nonlinear cointegration model with time-varying coefficients. Empirical results of this model are not far from those of Baltagi and Moscone (2010) for the OECD, but they show smaller income elasticities in the Eurozone during the period 1990-2014. The income elasticities dynamics of the time-varying coefficient methods are slightly decreasing in the last decades. This suggests that health care is not becoming a luxury good in the developed countries. The demographic structure of countries is also related to health care expenditure. The price of health care increases as the population over 65 years old increases and decreases as the population under 15 years old increases. A concave relationship appears between the health care expenditure and the rate of government funding dedicated to health care, and it has been descending during the last decades. This supports the positive effect of the new health care policies triggered by the GFC. These policies aim at making the Eurozone health care system less dependent on government funding and macro-economic shocks and they seem to be working in the right direction.

The reminder of this paper is organized as follows. The estimation methodology, inference properties and the asymptotic normality of the proposed estimators are established in Section 2. An extension to include heterogenous factors in the model is shown in Section 3. This methodology is applied to study the time-varying income elasticity of health care expenditure in Section 4. Section 5 concludes the paper. The technical proofs of the main theoretical results are relegated to Appendix.

2. Inference

A word on notation. We denote $\Omega^{1/2}$ to be any matrix such that $\Omega = (\Omega^{1/2})(\Omega^{1/2})'$. We use $||A||$ to denote $\{tr(A'A)\}^{1/2}$, $\Rightarrow$ to denote weak convergence, $\lfloor x \rfloor$ to denote the largest integer $\leq x$, and $\mathcal{B}(\Omega)$ to denote Brownian motion with the covariance matrix Ω. We use K_{mn} to denote the commutation matrix of order $mn \times mn$, i.e. the matrix for which $\text{vec}A^\top = K_{mn}\text{vec}A$ where A is any $m \times n$ matrix (K_{mn} is unique).

Set $\tau = \lfloor T\delta \rfloor$ where $\lfloor \cdot \rfloor$ denotes integer part and $\delta \in (0,1)$ is the sample fraction corresponding to observation t.

2.1. Time-varying coefficient panel data models with fixed effects

Consider a panel data model of the form

$$
\begin{aligned}
Y_{i,t} =& \sum_{j=1}^{d} X_{i,t,j}\beta_{t,j} + \alpha_i + u_{i,t} \\
=& \mathbf{X}_{i,t}'\beta_t + \alpha_i + u_{i,t}, \quad i = 1, \cdots, N, \quad t = 1, \cdots, T,
\end{aligned} \tag{2}
$$

where $\mathbf{X}_{i,t} = (X_{i,t,1}, \cdots, X_{i,t,d})'$, $\boldsymbol{\beta}_t = (\beta_{t,1}, \cdots, \beta_{t,d})'$ and all β_t are unknown functions, α_i reflects the unobservable individual effect, T is the time series length and N is the cross-section size. Note that α_i is allowed to be correlated with $\mathbf{X}_{i,t}$ through some unknown structure, and hence is a sequence of fixed effects. For the purpose of identifiability, we assume $\sum_{i=1}^{N} \alpha_i = 0$ throughout the paper. We assume $\mathbf{X}_{i,t}$ is a unit-root process (thus it is non-stationary) with generating mechanism such as

$$
\mathbf{X}_{i,t} = \mathbf{X}_{i,t-1} + \nu_{i,t}, \quad t = 1, \cdots, T; \; i = 1, \cdots, N, \tag{3}
$$

with a common initialization at $t = 0$ satisfying (Phillips and Moon (1999)) $E||X_{i,0}||^4 < \infty$

$$
\omega_{i,t} = (\nu_{i,t}', u_{i,t})' = C(\mathcal{L})\varepsilon_{i,t} = \sum_{j=0}^{\infty} C_j \varepsilon_{i,t-j}, \tag{4}
$$

where (i) $C(\mathcal{L}) = \sum_{j=0}^{\infty} C_j \mathcal{L}^j$, C_j is a sequence of fixed $(d+1) \times d$ matrices across j, $\mathcal{L}$ is the lag operator; (ii) $\varepsilon_{i,t}$ is a d-dimensional sequence of random vectors across i and over t with $\text{E}(\varepsilon_{i,t}) = 0, \text{E}(\varepsilon_{i,t}\varepsilon_{i,t}') = \Lambda_i$, $\text{E}(\varepsilon_{i,t}\varepsilon_{j,t}') = \Lambda_{i,j}$ for $i \neq j$, $\text{E}(\varepsilon_{i,t}\varepsilon_{j,s}') = 0$ for any i, j and $t \neq s$, and, letting $\varepsilon_{a,i,t}$ be the ath element of $\varepsilon_{i,t}$ with $\text{E}(\varepsilon_{a,i,t}^4) = \kappa^4$ for all i and t.

Partition $C_j = [\Phi_j, \psi_j]'$ so that

$$\nu_{i,t} = \sum_{j=0}^{\infty} \Phi_j' \varepsilon_{i,t-j}, \quad \text{and} \qquad u_{i,t} = \sum_{j=0}^{\infty} \psi_j' \varepsilon_{i,t-j}. \tag{5}$$

According to the functional limit theory for a standardized linear process (Phillips and Solo (1992)), we have for $t = \lfloor T\delta_0 \rfloor$ and $0 < \delta_0 \leq 1$, for any i

$$\frac{x_{i,t}}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{s=1}^{t} \nu_{i,s} + \frac{1}{\sqrt{T}} x_{i,0} = \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor T\delta_0 \rfloor} \nu_{i,s} + o_p(1) \Longrightarrow \mathcal{B}_{d,\delta_0}(\Omega_{\nu_i}),$$

$$\frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor T\delta_0 \rfloor} u_{i,s} \Longrightarrow \mathcal{B}_{\varepsilon,\delta_0}(\Omega_{u_i}), \quad \frac{1}{\sqrt{T}} \sum_{s=1}^{\lfloor T\delta_0 \rfloor} \varepsilon_{i,s} \Longrightarrow \mathcal{B}_{\varepsilon,\delta_0}(\Lambda_i)$$

and let $\delta_0(T) = \lfloor (\delta_0 - h)T \rfloor$, we have (Phillips and Hansen (1990))

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^{T} x_{i,t} x_{i,t}' =& \frac{1}{T} \sum_{t=1}^{T} \frac{x_{i,t}}{\sqrt{T}} \frac{x_{i,t}'}{\sqrt{T}} \Longrightarrow \int_0^1 \mathcal{B}_{d,r}(\Omega_{\nu_i}) \mathcal{B}_{d,r}'(\Omega_{\nu_i}) dr > 0, \\ \frac{1}{T^2 h} \sum_{t=1}^{T} x_{i,t} x_{i,t}' K_{th}(\delta_0) \cong& \frac{x_{i,\delta(T)}}{\sqrt{T}} \frac{x_{i,\delta(T)}'}{\sqrt{T}} \left(\frac{1}{Th} \sum_{t=1}^{T} K(\frac{t - T\delta_0}{Th}) \right) \\ \Longrightarrow& \delta_0 \Phi(1)' \mathcal{W}_d(\Lambda_i) \Phi(1) = \delta_0 \mathcal{W}_d(\Omega_{\nu_i}), \end{aligned}$$

where $K(\cdot)$ is a kernel function, $K_{th}(\delta_0) = K(\frac{t-T\delta_0}{Th})$, $\mathcal{B}_{d+1,\delta}(\Omega_i) = (\mathcal{B}_{d,\delta}(\Omega_{\nu_i})', \mathcal{B}_{\delta}(\Omega_{u_i}))'$ is $(d+1)$-dimensional Brownian motion (BM) with variance matrix Ω_{ω_i}, $\mathcal{B}_{\varepsilon,\delta}(\mathrm{I}_d)$ is d-dimensional BM with variance matrix I_d, $\mathcal{W}_d(\Omega_{\nu_i}) = \mathcal{B}_{\varepsilon,\delta}(\Omega_{\nu_i}) \mathcal{B}_{\varepsilon,\delta}(\Omega_{\nu_i})'$ is a Wishart variate with d degree of freedom and mean matrix Ω_{ν_i}, and

$$\Omega_{\omega_i} = C(1)' \Lambda_i C(1) = \left[\begin{array}{cc} \Phi(1)' \Lambda_i \Phi(1) & \Phi(1)' \Lambda_i \psi(1) \\ \psi(1)' \Lambda_i \Phi(1) & \psi(1)' \Lambda_i \psi(1) \end{array} \right] = \left[\begin{array}{cc} \Omega_{\nu_i} & \Omega_{\nu u_i} \\ \Omega_{\nu_i u_i} & \Omega_{u_i} \end{array} \right],$$

$i = 0, 1, ..., N$, with $C(1) = \sum_{j=0}^{\infty} C_j$, $\Phi(1) = \sum_{j=0}^{\infty} \Phi_j$, and $\psi(1) = \sum_{j=0}^{\infty} \psi_j$. Here Ω_{ω_i} is the partitioned long run variance matrix of $\omega_{i,t} = (\nu_{i,t}', u_{i,t})'$, and denotes that

$$\Omega_{\omega} = \left[\begin{array}{cc} \Phi(1)' \Lambda_0 \Phi(1) & \Phi(1)' \Lambda_0 \psi(1) \\ \psi(1)' \Lambda_0 \Phi(1) & \psi(1)' \Lambda_0 \psi(1) \end{array} \right] \equiv \left[\begin{array}{cc} \Omega_{\nu} & \Omega_{\nu u} \\ \Omega_{\nu u} & \Omega_{u} \end{array} \right], \tag{6}$$

where $\Lambda_0 = \lim_{N \to \infty} \Lambda_i$. The limit theory also involves the partitioned components of the one-sided long run variance matrix

$$\Delta_{\omega_i} = \left[\begin{array}{cc} \Delta_{\nu_i} & \Delta_{\nu_i u_i} \\ \Delta_{\nu_i u_i}' & \Delta_{u_i} \end{array} \right] = \sum_{j=0}^{\infty} \mathrm{E}[\omega_{i,0} \omega_{i,j}'], \tag{7}$$

and assumes that $\Delta_\omega > 0$ exists such that as $N \to \infty$,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \Delta_{\omega_i} = \Delta_\omega \equiv \left[\begin{array}{cc} \Delta_\nu & \Delta_{\nu u} \\ \Delta_{\nu u} & \Delta_u \end{array}\right]. \tag{8}$$

The aim of this paper is to construct consistent estimates for the time-varying coefficient vector β_t before establishing the asymptotic properties of the estimators. As in Robinson (2012) and Cai (2007b), we propose that the coefficient vector β_t satisfies

$$\beta_{t,j} = \beta_j(\frac{t}{T}), \quad t = 1, \cdots, T, \tag{9}$$

where all β_j's are unknown smooth functions.

Two classes of nonparametric methods are developed to estimate the coefficient functions $\boldsymbol{\beta}$ without taking the first difference to remove the fixed effects. In this paper, we propose using a pooled dummy variable approach to estimate β_t, which is more efficient than the averaged method (see for example Chen et al., 2012, for detailed discussion).

Before presenting this method, we introduce the following notation:

$$\begin{aligned}
&\widetilde{\mathbf{Y}} = (Y_1', Y_2', \cdots, Y_N')', \quad Y_i' = (Y_{i,1}, Y_{i,2}, \cdots, Y_{i,T})' \\
&\widetilde{\mathbf{X}} = (X_{1,1}, \cdots, X_{1,T}, X_{2,1}, \cdots, X_{2,T}, \cdots, X_{N,1}, \cdots, X_{N,T})', \\
&\widetilde{B}(\mathbf{X}, \boldsymbol{\beta}) = (X_{1,1}'\boldsymbol{\beta}_1, \cdots, X_{1,T}'\boldsymbol{\beta}_T, X_{2,1}'\boldsymbol{\beta}_1, \cdots, X_{2,T}'\boldsymbol{\beta}_T, \cdots, X_{N,1}'\boldsymbol{\beta}_1, \cdots, X_{N,T}'\boldsymbol{\beta}_T)', \\
&\alpha_0 = (\alpha_1, \alpha_2, \cdots, \alpha_N)', \quad D_0 = \mathrm{I}_N \otimes i_T, \\
&\alpha = (\alpha_2, \cdots, \alpha_N)', \quad D = (-i_{N-1}, \mathrm{I}_{N-1})' \otimes i_T, \\
&\tilde{u} = (u_1', \cdots, u_N')', \quad u_i = (u_{i,1}, \cdots, u_{i,T})',
\end{aligned} \tag{10}$$

where $\otimes$ denotes the Kronecker product, i_k is the $k \times 1$ vector of ones and I_k is the $k \times k$ identity matrix. Rewriting model (2) in a matrix format yields

$$\widetilde{\mathbf{Y}} = \widetilde{B}(\mathbf{X}, \boldsymbol{\beta}) + D_0\alpha_0 + \tilde{u}. \tag{11}$$

As $\sum_{i=1}^{N} \alpha_i = 0$ as per the identification, model (2) can be rewritten in matrix form as

$$\widetilde{\mathbf{Y}} = \widetilde{B}(\mathbf{X}, \boldsymbol{\beta}) + D\alpha + \tilde{u}. \tag{12}$$

We adopt the Nadaraya-Watson (NW) type local level regression estimator (or local constant) to estimate time-varying coefficients

$$\boldsymbol{\beta}(\cdot) = [\beta_1(\cdot), \cdots, \beta_d(\cdot)]'.$$

Under certain smoothness conditions on $\boldsymbol{\beta}$ for some fixed $\delta_0 \in (0,1)$, we have

$$\boldsymbol{\beta}_t \equiv \boldsymbol{\beta}(\frac{t}{T}) = \boldsymbol{\beta}(\delta_0) + O\left(\frac{t}{T} - \delta_0\right), \tag{13}$$

when t/T is in a small neighbourhood of δ_0.

For the given $0 < \delta_0 < 1$, define $\widetilde{M}' = [M_1', \cdots, M_N']$ with $M_i = [X_{i,1}, \cdots, X_{i,T}]'$. Based on the local approximation of $\boldsymbol{\beta}$ in (13), we have $\widetilde{B}(\mathbf{X}, \boldsymbol{\beta}) \approx \widetilde{M}\boldsymbol{\beta}(\delta_0)$.

Let $K(\cdot)$ denote the kernel function and h be a bandwidth, denote $\widetilde{W}(\delta_0) = I_N \otimes W(\delta_0)$ with $W(\delta_0) = \text{diag}\left[K(\frac{1-\delta_0 T}{Th}), \cdots, K(\frac{T-\delta_0 T}{Th})\right]$. The *pooled nonparametric dummy variable estimation* method is given as follows.

For given $0 < \delta_0 < 1$, minimize the loss function

$$L(\boldsymbol{\beta}, \alpha) = \left[\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0) - D\alpha\right]' \widetilde{W}(\delta_0) \left[\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0) - D\alpha\right] \tag{14}$$

with respect to $\boldsymbol{\beta}(\delta_0)$ and α.

Taking the derivative of (14) with respect to α and setting the result to zero, we obtain

$$\widehat{\alpha} := \widehat{\alpha}(\delta_0) = \left[D'\widetilde{W}(\delta_0)D\right]^{-1} D'\widetilde{W}(\delta_0)\left\{\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0)\right\}. \tag{15}$$

Replacing α in (14) by $\widehat{\alpha}$, we obtain the concentrated weighted least squares:

$$\begin{aligned}
&\left[(\mathrm{I}_{NT} - \mathcal{S}_{NT})(\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0))\right]' \widetilde{W}(\delta_0) \left[(\mathrm{I}_{NT} - \mathcal{S}_{NT})(\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0))\right] \\
=&\left[\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0)\right]' \widetilde{W}^*(\delta_0) \left[\widetilde{Y} - \widetilde{M}\boldsymbol{\beta}(\delta_0)\right]
\end{aligned} \tag{16}$$

where $\widetilde{W}^*(\delta_0) = \widetilde{\mathcal{K}}(\delta_0)'\widetilde{W}(\delta_0)\widetilde{\mathcal{K}}(\delta_0)$ and $\widetilde{\mathcal{K}}(\delta_0) \equiv I_{NT} - \mathcal{S}_{NT} = I_{NT} - D[D'\widetilde{W}(\delta_0)D]^{-1}D'\widetilde{W}(\delta_0)$. Observe that for any δ_0, $\widetilde{K}(\delta_0)D\alpha = 0$.

Hence, the fixed effects term $D\alpha$ is eliminated in (14). By the definition of $\widetilde{W}^*(\delta_0)$ and the fact $\widetilde{K}(\delta_0)D\alpha = 0$, we have

$$\begin{aligned}
\widetilde{W}^*(\delta_0) =& \widetilde{\mathcal{K}}(\delta_0)'\widetilde{W}(\delta_0)\widetilde{\mathcal{K}}(\delta_0) = \widetilde{W}(\delta_0)\widetilde{\mathcal{K}}(\delta_0) = \widetilde{W}(\delta_0)(I_{NT} - \mathcal{S}_{NT}) \\
=& \widetilde{W}(\delta_0) - \widetilde{W}(\delta_0)D[D'\widetilde{W}(\delta_0)D]^{-1}D'\widetilde{W}(\delta_0).
\end{aligned} \tag{17}$$

Minimizing (16) with respect to $\boldsymbol{\beta}$, we obtain the estimate of $\boldsymbol{\beta}(\delta_0)$ as

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\delta_0) =& \left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1} \widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{\mathbf{Y}} \\
=&(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}' K_{th}(\delta_0)(x_{i,t} - \bar{x}_i + \bar{x}))^{-1}(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}' K_{th}(\delta_0)(Y_{i,t} - \bar{Y}_i + \bar{Y})),
\end{aligned} \tag{18}$$

where $\bar{x}_i$ and $\bar{x}$ is kernel-weighted average. The $\widehat{\boldsymbol{\beta}}(\delta_0)$ is called NW type local dummy variable estimator of $\boldsymbol{\beta}(\delta_0)$ and its asymptotic distribution is given in the following theorem.

We need to introduce the following regularity conditions to establish the asymptotic results. Here and in the sequel, define $\mu_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$ for $j = 0, 1, 2$.

Assumption 1. *The probability kernel function $K(\cdot)$ is symmetric and Lipschitz continuous with a compact support $[-1, 1]$ with $\mu_0 = 1$.*

Assumption 2. *The coefficient function $\boldsymbol{\beta}(\cdot)$ is continuous with $|\boldsymbol{\beta}(\delta_0 + z) - \boldsymbol{\beta}(\delta_0)| = O(|z|^\gamma)$ as $z \to 0$ for some $1/2 < \gamma \le 1$ and any $\delta_0 \in (0, 1)$.*

Assumption 3. *(i)Let $\varepsilon_{i,t}$ is d-dimensional random vectors across i and over t and assume that $\varepsilon_{i,t}$ are possibly correlated, and heteroscedastic over the cross section with $E(\varepsilon_{i,t}) = 0$, $E(\varepsilon_{i,t}\varepsilon'_{i,t}) = \Lambda_i$, $E(\varepsilon_{i,t}\varepsilon'_{j,t}) = \Lambda_{i,j}$ for $i \neq j$, $E(\varepsilon_{i,t}\varepsilon'_{j,s}) = 0$, for any i, j and $t \neq s$; and $E[||\varepsilon_{i,t}||^{4+\gamma_0}] < \infty$ for $\gamma_0 > 0$; (ii) There exists positive definite matrices Λ_0 and Σ_Λ, such that as $N \to \infty$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \Lambda_i = \Lambda_0 \quad \textit{and} \quad \lim_{N\to\infty} \frac{1}{N} \sum_{i,j=1}^{N} \Lambda_{i,j} = \Sigma_\Lambda.$$

Furthermore, there exists $\Sigma_{\nu,u} > 0$, such that as $N \to \infty$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i,j=1}^{N} E(\Phi(1)'\varepsilon_{i,s}\varepsilon'_{i,t}\psi(1)\psi(1)'\varepsilon_{j,t}\varepsilon'_{j,s}\Phi(1)|F_{t-1,N}) = \Sigma_{\nu,u}, \tag{19}$$

where $F_{t,N} = \sigma\{\varepsilon_{i,s} : 1 \le i \le N, 1 \le s \le t\}$ is a σ-field.

(iii)The linear process (nonrandom) coefficient matrices $C_j \equiv [\Phi_j, \psi_j]'_{(d+1)\times d}$ satisfy $\sum_{j=0}^{\infty} j^3 \, ||C_j|| < \infty$.

Assumption 4. *The bandwidth h satisfies that $Th \to \infty$ and $Nh \to \infty$, and $N/Th \to 0$, as $T, N \to \infty$ simultaneously.* [1]

Let

$$\Delta^t_{\nu\bar{u}}(\delta_0) = \lim_{N\to\infty} \frac{1}{N} \sum_{l\le t} \left\{ (\sum_{s=1}^{T} K_{sh}(\delta_0))^{-1} \sum_{s=1}^{T} K_{sh}(\delta_0) \mathrm{E}[\nu_{i,l} u_{i,s}] \right\}, \tag{20}$$

[1] The condition $Nh \to \infty$ and $N/Th \to 0$ are required to eliminate the bias effect. The condition $(N/Th) \to 0$ indicates that the limit theory is most likely to be useful in practice when N is moderate and T is large. (We can expect such data configurations in multi-country macroeconomic data, for example, when we restrict attention to groups of countries such as OECD nations or developing countries.)

and denote

$$
\begin{aligned}
\mathcal{C}_{K^*}(2) =& \nu_0 \int_{-1}^{1} (\int_{s}^{1} K(t)dt)^2 ds + (\int_{-1}^{1} K(t) \int_{t}^{1} K(u)dudt)^2, \quad \text{and} \\
\mathcal{C}_{K^*}(1,2) =& \int_{-1}^{1} K^2(t)(\int_{-1}^{t}\int_{s}^{1} K(u)duds)dt + \int_{-1}^{1} K(t)(\int_{t}^{1} K(u)du)(\int_{-1}^{t} K(s)ds)dt. \qquad (21)
\end{aligned}
$$

Theorem 1. *Suppose assumptions 1-4 are satisfied and $N^{1/2}Th^{1+\gamma} = o(1)$. Then for any fixed $0 < \delta_0 < 1$,*

(a) as $T, N \to \infty$ simultaneously

$$
\begin{aligned}
& N^{1/2}Th \left\{ \widehat{\boldsymbol{\beta}}(\delta_0) - \boldsymbol{\beta}(\delta_0) - \left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1} \left(N \sum_{t=1}^{T} K_{th}(\delta_0) \left(\Delta_{\nu u} - \Delta_{\nu \bar{u}}^{t}(\delta_0)\right)\right)\right\} \\
& \xrightarrow{\mathcal{L}} N\left(0, \frac{\mathcal{C}_{K^*}}{(1-\mathcal{C}_K)^2}\Omega_\nu^{-1}\Sigma_{\nu,u}\Omega_\nu^{-1}\right), \qquad (22)
\end{aligned}
$$

where

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

$\mathcal{C}_{K^*} \equiv \mathcal{C}_{K^*}(1) + \mathcal{C}_{K^*}(2) - 2\mathcal{C}_{K^*}(1,2)$ with $\mathcal{C}_{K^*}(1) = \nu_0$, $\mathcal{C}_{K^*}(2)$, $\mathcal{C}_{K^*}(1,2)$ are defined in (21), $\Delta_{\nu u}$, $\Delta_{\nu\bar{u}}^{t}(\delta_0)$, Ω_ν and $\Sigma_{\nu,u}$ are defined in (8), (20), (6) and (19) respectively;

(b) Specially, if $\nu_{i,t}$ and $u_{i,t}$ are uncorrelated across i and over t, then as $N, T \to \infty$ with $N/Th \to 0$,

$$
N^{1/2}Th \left\{\widehat{\boldsymbol{\beta}}(\delta_0) - \boldsymbol{\beta}(\delta_0)\right\} \xrightarrow{\mathcal{L}} N\left(0, \frac{\mathcal{C}_{K^*}}{(1-\mathcal{C}_K)^2}\Omega_\nu^{-1}\Sigma_{\nu,u}\Omega_\nu^{-1}\right). \qquad (23)
$$

Remark 1. *If $\{u_{i,t}\}$ and $\{\nu_{i,t}\}$ are correlated, we do not attain $N^{1/2}Th$ consistency with pooled dummy variable estimator $\hat{\boldsymbol{\beta}}(\delta_0)$, because of the persistence of bias effects (which has an order of root N). A fully modified (FM) regression technique based on Phillips et al. (2017) is developed to eliminate the bias effect in this non-stationary panel data case.*

2.2. Pooled FM-nonparametric kernel estimation

Let $\hat{\Delta}_{\nu u}$ and $\hat{\Delta}_{\nu\bar{u}}^{t}(\delta_0)$ denote the consistent estimates of $\Delta_{\nu u}$ and $\Delta_{\nu\bar{u}}^{t}(\delta_0)$ satisfying $\sqrt{N}(\hat{\Delta}_{\nu u} - \hat{\Delta}_{\nu u}) = o_p(1)$ and

$$
\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)||\sqrt{N}(\hat{\Delta}_{\nu\bar{u}}^{t}(\delta_0) - \Delta_{\nu\bar{u}}^{t}(\delta_0))|| = o_p(1),
$$

for any $0 < \delta_0 < 1$.

Recall that $\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{Y}$ can be rewritten as $\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{Y} = \sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}(Y_{i,t} - \bar{Y}_i + \bar{Y})K_{th}(\delta_0)$. We define the "bias-corrected" FM kernel estimator of the functional coefficient as

$$\widehat{\boldsymbol{\beta}}_{PFM}(\delta_0) = \left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\left(\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{\mathbf{Y}} - N\sum_{t=1}^{T} K_{th}(\delta_0)(\hat{\Delta}_{\nu u} - \hat{\Delta}^t_{\nu\bar{u}}(\delta_0))\right)$$
$$\equiv \left(\sum_{i=1}^{N}\sum_{t=1}^{T} x'_{i,t}(x_{i,t} - \bar{x}_i + \bar{x})K_{th}(\delta_0)\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} K_{th}(\delta_0)\left\{x'_{i,t}(Y_{i,t} - \bar{Y}_i + \bar{Y}) - (\hat{\Delta}_{\nu u} - \hat{\Delta}^t_{\nu\bar{u}}(\delta_0))\right\}\right), \tag{24}$$

where $\bar{x}_i = \left(\sum_{t=1}^{T} K_{th}(\delta_0)\right)^{-1}\sum_{t=1}^{T} K_{th}(\delta_0)x_{i,t}$ and $\bar{x} = \left(N\sum_{t=1}^{T} K_{th}(\delta_0)\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{th}(\delta_0)x_{i,t}$.

Since $\sqrt{N}(\hat{\Delta}_{\nu u} - \Delta_{\nu u}) = o_p(1)$ and $\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)||\sqrt{N}(\hat{\Delta}^t_{\nu\bar{u}}(\delta_0) - \Delta^t_{\nu\bar{u}}(\delta_0))|| = o_p(1)$ for any $0 < \delta_0 < 1$, the asymptotic distribution of $\hat{\beta}_{PFM}(\delta_0)$ is obtained directly from Theorem 1.

Theorem 2. *Suppose that the assumptions in Theorem 1 are satisfied. We then have*

$$N^{1/2}Th\left(\widehat{\boldsymbol{\beta}}_{PFM}(\delta_0) - \boldsymbol{\beta}(\delta_0)\right) \overset{\mathcal{L}}{\Longrightarrow} N(0, \frac{\mathcal{C}_{K^*}}{(1-\mathcal{C}_K)^2}\Omega_\nu^{-1}\Sigma_{\nu,u}\Omega_\nu^{-1}), \tag{25}$$

as $N, T \to \infty$, for any fixed $0 < \delta_0 < 1$.

Practical implementation of FM-nonparametric kernel regression requires the estimation of the one-sided long-run covariance matrix $\Delta_{\nu u}$ and $\Delta^t_{\nu\bar{u}}(\delta_0)$. Consistent estimates of Ω_ω are likewise required to construct a consistent estimate of the covariance matrix. Following the approach of Phillips et al. (2017) for time series data, consistent estimates of $\hat{\Delta}_{\nu u}$, $\hat{\Delta}^t_{\nu\bar{u}}(\delta_0)$ and $\hat{\Omega}_\omega$ can be constructed using averages over $i = 1, \cdots, N$. More specifically, let $\hat{\omega}_{i,t} = (\nu'_{i,t}, \hat{u}_{i,t})'$, $\hat{u}_{i,t} = y_{i,t} - \mathbf{X}'_{i,t}\hat{\boldsymbol{\beta}}(t/T) - \hat{\alpha}_i$ be the estimated residuals. Since $\nu_{i,t} = x_{i,t} - x_{i,t-1}$, let $0 < \tau_* < 1/2$, which can be arbitrary small. We may construct the estimated autocovariances

$$\hat{\Gamma}_{i,\omega}(j) = \frac{1}{\tau^*_T - \tau_T}\sum_{t=\tau_T+1}^{\tau^*_T} \hat{\omega}_{i,t-j}\hat{\omega}'_{i,t}, \quad j = 0, 1, \cdots, l_T(\equiv l_i),\ i = 1, 2, \cdots, N,$$

where $\tau_T = \lfloor \tau_* T\rfloor$ and $\tau^*_T = \lfloor (1-\tau_*)T\rfloor$, which are used to define the averaged kernel estimators

$$\hat{\Delta}_\omega = \frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_{i,\omega}, \qquad \hat{\Delta}_{i,\omega} = \sum_{j=0}^{l_T} W(\frac{j}{l_T})\hat{\Gamma}_{i,\omega}(j),$$
$$\hat{\Omega}_\omega = \frac{1}{N}\sum_{i=1}^{N}\hat{\Omega}_{i,\omega}, \qquad \hat{\Omega}_{i,\omega} = \sum_{j=-l_T}^{l_T} W(\frac{j}{l_T})\hat{\Gamma}_{i,\omega}(j), \tag{26}$$

we then have the following estimator

$$\hat{\Delta}_{\nu u}=\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_{i,\nu u},\qquad \hat{\Delta}^{t}_{\nu\bar{u}}(\delta_0)=\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}^{t}_{i,\nu\bar{u}}(\delta_0),\tag{27}$$

with

$$\hat{\Delta}^{t}_{i,\nu\bar{u}}(\delta_0)=(\sum_{s=1}^{T}K_{sh}(\delta_0))^{-1}\sum_{l=1}^{T}\sum_{s=1}^{T}K_{sh}(\delta_0)\mathrm{I}(l\leq t)W(\frac{s-l}{l_T})\hat{\Gamma}_{i,\nu u}(s-l),\quad -l_T\leq s-l\leq l_T,$$

where $W(\cdot)$ is a lag kernel function and $(l_i\equiv)l_T<T$ is the lag truncation number which tends to infinity as $T\to\infty$ To ensure the consistency of $\hat{\Delta}_{\nu u}$ and $\hat{\Delta}^{t}_{\nu\bar{u}}$, the kernel function $W(\cdot)$ is assumed to be bonded $W(0)=1$, and $W(-x)=W(x)$ such that $\int_{-1}^{1}W^2(x)dx<\infty$ and with Parzen's exponent $q\in[0,\infty)$ such that $k_q=\lim_{x\to 0}\frac{1-W(x)}{|x|^q}<\infty$ (Andrews (1991)) As is well known in the nonparametric literature, the choice of the bandwidth l_i is important in the limit behavior of $\hat{\Omega}_{\omega}$ and $\hat{\Delta}_{\omega}$. In the asymptotic theorem, we need the stronger result that satisfies $\sqrt{N}(\hat{\Delta}_{\nu u}-\Delta_{\nu u})=o_p(1)$ and $\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)||\sqrt{N}(\hat{\Delta}^{t}_{\nu\bar{u}}(\delta_0)-\Delta^{t}_{\nu\bar{u}}(\delta_0))||=o_p(1)$.

In the present nonparametric case, kernel methods are used to estimate the varying-coefficient functions, which in turn complicates the form of the estimated residuals and makes the proof of the consistency much more difficult. On the other hand, the asymptotic bias of the kernel estimates also affects the consistency of $\hat{\Delta}_{\omega}$ and $\hat{\Omega}_{\omega}$.

Assumption 5. *The lag kernel $W(\cdot)$ has Parzen exponent $q>1/2$, and the bandwidth parameter l_T tends to infinity with $l_T/T\to 0$, and $l_T^{2q}/T\to\epsilon>0$ when $l_T\to\infty$ as $T\to\infty$.*

Proposition 1. *Suppose that the assumptions in Theorem 1 and assumption 5 are satisfied, and $l_T=o\left(\frac{T}{\log(Th)}\right)$, we have*

$$\sqrt{N}(\hat{\Delta}_{\nu u}-\Delta_{\nu u})=o_p(1),$$

$$\sqrt{N}\left|\left|\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)(\hat{\Delta}^{t}_{\nu\bar{u}}(\delta_0)-\Delta^{t}_{\nu\bar{u}}(\delta_0))\right|\right|=o_p(1),$$

and

$$\sqrt{N}(\hat{\Omega}_{\omega}-\Omega_{\omega})=o_p(1),$$

as $N,T\to\infty$ simultaneously.

The "bias-corrected" FM kernel estimator of the functional coefficient can now be expressed as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{PFM}(\delta_0)&=\left[\widetilde{M}'\widetilde{W}^{*}(\delta_0)\widetilde{M}\right]^{-1}\left(\widetilde{M}'\widetilde{W}^{*}(\delta_0)\widetilde{\mathbf{Y}}-N\sum_{t=1}^{T}K_{th}(\delta_0)(\hat{\Delta}_{\nu u}-\hat{\Delta}^{t}_{\nu\bar{u}}(\delta_0))\right)\\&\equiv\left[\widetilde{M}'\widetilde{W}^{*}(\delta_0)\widetilde{M}\right]^{-1}\left(\widetilde{M}'\widetilde{W}^{*}(\delta_0)\widetilde{\mathbf{Y}}-\text{bias}\right),\end{aligned}\tag{28}$$

and

$$\text{bias} = \left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\left\{N\sum_{t=1}^{T}K_{th}(\delta_0)(\hat{\Delta}_{\nu u} - \hat{\Delta}^t_{\nu\bar{u}}(\delta_0))\right\}$$
$$= \left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\left(NZ_T\hat{\Delta}_{\nu u} - N\sum_{t=1}^{T}K_{th}(\delta_0)\hat{\Delta}^t_{\nu\bar{u}}(\delta_0)\right),$$

where

$$\hat{\Delta}_{\nu,u} = \sum_{j=0}^{l_T} W(\frac{j}{l_T})\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Gamma}_i(j)\right) = \sum_{j=0}^{l_T} W(\frac{j}{l_T})\left(\frac{1}{N(\tau_T^* - \tau_T)}\sum_{i=1}^{N}\sum_{t=\tau_T+1}^{\tau_T^*}\hat{u}_{i,t}\nu'_{i,t+j}\right),$$

and,

$$\sum_{t=1}^{T}K_{th}(\delta_0)\hat{\Delta}^t_{\nu\bar{u}}(\delta_0)$$
$$=(Z_T)^{-1}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\sum_{s=1}^{T}\sum_{l=1}^{T}I(l\le t)K_{sh}(\delta_0)W(\frac{s-l}{l_T})\left(\frac{1}{N}\sum_{i=1}^{N}\Gamma_i(|s-l|)\right)\right)$$
$$=(Z_T)^{-1}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\sum_{s=1}^{T}\sum_{l=1}^{T}I(l\le t)K_{sh}(\delta_0)W(\frac{s-l}{l_T})\left(\frac{1}{N(\tau_T^* - \tau_T)}\sum_{i=1}^{N}\sum_{t=\tau_T+1}^{\tau_T^*}\hat{u}_{i,t}\nu'_{i,t+|s-l|}\right)\right).$$

Meanwhile, we also need the estimator of α to obtain the residual $\hat{u}_{i,t}$, which is given by the average across t

$$\hat{\alpha} = \frac{1}{\tau_T^* - \tau_T}\sum_{t=\tau_T+1}^{\tau_T^*}\hat{\alpha}(\frac{t}{T}).$$

3. Adding Unobservable Factors

The use of panel data models like

$$Y_{i,t} = \alpha_i + \mu_t + \mathbf{X}'_{i,t}\beta + u_{it}, \tag{29}$$

with $\mathbf{X}_{i,t} = (X_{i,t,1}, \ldots, X_{i,t,d})'$ and $\beta = (\beta_1, \ldots, \beta_d)'$, are more flexible to account for country and time heterogeneities such as geographical location or size, business cycles and bias from the omission of country-specific variables. Panel data models, such as (29), assume that heterogeneity is constant over time for each cross-section i. However, this might not be true for large T. Recently, Pesaran (2006), Bai (2009) and Kneip et al. (2012) have developed panel data models with a factor structure in the error term,

$$Y_{i,t} = \mathbf{X}'_{i,t}\beta + F_t\lambda_i + u_{i,t} \tag{30}$$

which allows the unobservable individual effects to vary with time. Term $\lambda_i = (\lambda_{1i}, \ldots, \lambda_{ri})'$ are unobserved individual loadings and $F_t = (F_{1,t}, \ldots, F_{r,t})$ are the unobserved common factors. Thus, factors represent common shocks in time such as business cycles, technological shocks or health crises. Loadings on the other hand, express the heterogeneous impact of those shocks for different countries. Note that when $r = 2$, $F_{1t} = 1$ for all t and $\lambda_{2i} = 1$ for all i, then model (30) reduces to the classical individual and time effects model (29). Model (30) is estimated in three steps: i) a classical panel data model is fitted assuming there are no factors, ii) factors are calculated from the residuals using principal components, and iii) coefficients b are estimated using a panel data model without factors with a new dependent variable, $Y_{i,t} - \hat{F}_t\hat{\lambda}_i$.

Section 2 shows the inference of the Nadaraya-Watson (NW) estimator for model (2) which is an extension of (29) with coefficients varying over time. This model accounts for unobserved fixed effects in coefficient α_i and for fixed time effects if $X_{i,t,1}$ is one for all i and t. The FM estimator corrects for the bias generated by non-stationary variables in the model, but it fails to correct for the bias from omitted variables with heterogeneous effects.

The contribution of this section is the proposition of an extension of model (30) with time-varying coefficients, which will automatically pick up changes in relationships over time, allow for non-stationary regressors and account for heterogenous effects of unobservable factors. The model proposed is,

$$Y_{i,t} = \mathbf{X}'_{i,t}\boldsymbol{\beta}_t + \alpha_i + F_t\lambda_i + u_{i,t}$$

with $\mathbf{X}_{i,t}$ possibly a unit root process. Coefficients $\boldsymbol{\beta}_t$ are estimated with a three steps procedure like for model (30):

i) Using (24) to estimate $\boldsymbol{\beta}_t$ for each time t as if the data process was of type (2).

ii) If other unknown factors explain the dependent variable, then $W_{i,t} = Y_i - \mathbf{X}'_{i,t}\boldsymbol{\beta}_t = F_t\lambda_i + \epsilon_{i,t}$, which also can be written like $W_i = F\lambda_i + \epsilon_i$ with W_i and ϵ_i vectors of length T and F a $T \times r$ matrix. The least square objective function is

$$\mathrm{tr}[(W - F\Lambda')(W - F\Lambda')'],$$

defining $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$. Thus, the common factor F is obtain with principal component analysis from matrix WW'/N to ensure the identificability of F. This differs from

Gao and Xia (2017) whose time-varying factors are estimated using nonparametric techniques, assuming that these factors are the same for each cross-section. The estimation of $F\Lambda'$ can be inconsistent for large values of N and fixed T. However, it is consistent under large N, T as explained in Bai (2009).

iii) Given F and Λ, the new estimate

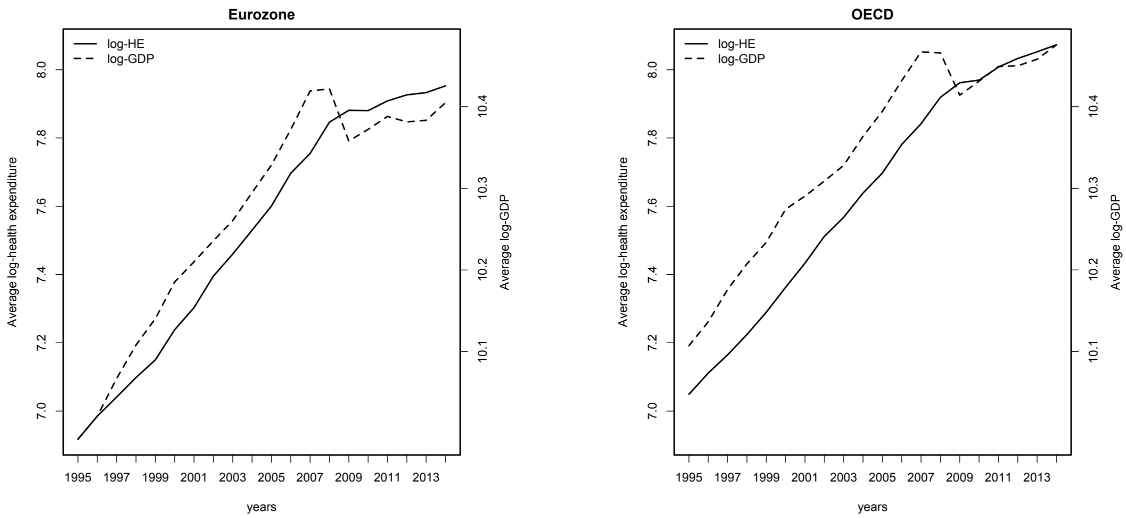
$$\widehat{\boldsymbol{\beta}}_{PFM}(\delta_0) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x'_{i,t}(x_{i,t} - \bar{x}_i + \bar{x}) K_{th}(\delta_0) \right)^{-1} \cdot \left(\sum_{i=1}^{N} \sum_{t=1}^{T} K_{th}(\delta_0) \left\{ x'_{i,t}(Y_{i,t} - \bar{Y}_i + \bar{Y} - F_t \lambda_i) - (\hat{\Delta}_{\nu u} - \hat{\Delta}^t_{\nu \bar{u}}(\delta_0)) \right\} \right) \tag{31}$$

There is a long literature on how to choose the number of unobserved factors, r, in step ii). Thus, Kneip et al. (2012) propose a sequential testing procedure to find the best dimension, Onatski (2010) propose the eigenvalues differences which can work well for stationary and non-stationary factors and Ahn and Horenstein (2013) propose the eigenvalue ration and growth ratio criteria which work well in small samples. A comprehensive survey of these methodologies can be found in Bada and Liebl (2014). Due to the possibility of non-stationary variables in our sample, Onatski (2010) criterion is chosen for this paper's application in Section 4.

4. Application

There is growing concern about the sustainability of health care systems in developed countries. Figure 1 shows a steady increase of average health care expenditure (HCE) during the past two decades in the Eurozone and OECD (continuous lines). The rate of increase reduces only slightly after the GFC, even though the drop in GDP (dashed lines) in 2008 is very important. Thus, Morgan and Astolfi (2015) explain that the share of GDP devoted to health care has steadily increased in the OECD from 2000 to 2009, suffering an important downturn in 2010 and 2011, increasing at a slower pace thereafter. On the other hand, the WHO's report on the effects of the GFC in European health systems (Mladovsky et al. (2012)) concludes that the response has been heterogeneous amongst 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 GFC represents an economic shock which has trigged the implementation of new health policies aim at reducing the sensibility of health systems to these economic shocks and at making them less dependent on public revenues.

Fig. 1. Mean value of log-HCE (continuous lines) and log-GDP (dashed lines) per capita in the Eurozone and the OECD countries during the period 1995-2014.



The consensus in the literature is that the main factor that drives HCE is income. Previous work using parametric cross-sectional data show elasticities of HCE from around 1.20 to 1.50 in the OECD(Kleiman, 1974; Newhouse, 1977; Getzen, 2000), meaning that health care in the OECD is a luxury good. Technically, the cross-sectional models look like,

$$\bar{Y}_i = \alpha + \bar{\mathbf{X}}_i'\beta + u_t, \quad i = 1, \ldots, N \tag{32}$$

where N is the number of countries in the sample. Variables are recorded over a number of years $t = 1, \ldots, T$ and the dependent variable, $\bar{Y}_i$, is calculated as the average value of $Y_{i,t}$ over all years of country i. Similarly, the regressors $\bar{\mathbf{X}}_i = (\bar{\mathbf{X}}_{i,1}, \ldots, \bar{\mathbf{X}}_{i,d})'$ are calculated for each country where $\bar{\mathbf{X}}_{i,k}$ is the average over time of $X_{i,t,k}$. The error term must be uncorrelated with the regressors. The coefficients of model (32) are easily estimated with ordinary least squares (OLS) and often the model is fitted over different time periods to understand the dynamics of coefficients $\beta = (\beta_1, \ldots, \beta_d)^T$.

Model (32) can suffer from sample bias, country effects and endogeneity. Some previous results in panel data models such as (29), see Gerdtham et al. (1992), continue showing elasticities greater than 1 for the OECD. New studies have shown that most variables in this system are non-stationary and when the model reflects this fact, results state that health care is a necessity rather than a luxury (see Baltagi and Moscone, 2010; Samadi and Homaie Rad, 2013). A detailed summary of papers in this field can be found in Table 1 in Lago-Peñas et al. (2012). As mentioned in Section 3, panel data models such as (29) assume that heterogeneity is constant over time for each cross-section i. However, this might not be true for large T. This is corrected by adding unobservable factors as in Pesaran (2006), Bai (2009) and Kneip et al. (2012) amongst others with models such as (30), which is used in Baltagi and Moscone (2010) to estimate the long-run income elasticity of HCE for the OECD, obtaining values of $\hat{\beta}$ statistically smaller than 1. The inclusion of time-varying coefficients and non-stationary regressors panel data models (31) may be the answer to correct all possible biases arising in this problem and to show the evolution of the relationship between income and HCE over time.

4.1. Data

The dependent variable in our study is the log of total HCE per capita PPP (constant 2011 international \$) for country i and year t, denoted by lhe. The proxy for income is the log of the GDP per capita PPP (constant 2011 international \$) in our model variable, $lgdp$. It is reasonable to think that these elasticities change over time in a smooth way, these changes cannot be reflected with constant coefficient models. Authors such as Jewell et al. (2003) show that both series $lhe_{i,t}$ and $lgdp_{i,t}$ in panel data model (29) are stationary but for a few structural breaks and that time specific effects (μ_t) must be included in the model to mitigate the cross-section dependence. On the other hand, Baltagi and Moscone (2010) show that $lhe_{i,t}$ and $lgdp_{i,t}$ are non-stationary, which causes erroneous inference in the results of classical models that must satisfy the assumption of stationarity in all variables.

The literature has also validated some demographic variables such as the population ratio over 65 and under 15 years old as possible determinants of HCE (see Leu, 1986; Hitiris and Posnett, 1992, amongst others), denoted in the model as $Pop65$ and $Pop14$, respectively. These studies also report a positive significant relationship between the public finance share of health care and the total HCE. The $Public$ variable in our study is calculated as the % government expenditure in public health care. In a nutshell, the set of variables in the study are: $Y_{i,t} = lhe_{i,t}$, $\mathbf{X}_{i,t} = (lgdp_{i,t}, Pop65_{i,t}, Pop14_{i,t}, Public_{i,t})'$.

The period under study starts in 1995 and runs until 2014 for 20 countries in the Eurozone (Austria, Belgium, Cyprus, Germany, Estonia, Finland, France, Greece, Ireland, Italy, Lithuania, Luxembourg, Latvia, Macedonia, Malta, Netherlands, Portugal, Slovak Republic, Slovenia and Spain) and for 34 countries in the OECD (Australia, Austria, Belgium, Canada, Chile, Czech Republic, Germany, Denmark, Estonia, Finland, France, Greece, Hungary, Ireland, Iceland, Israel, Italy, Japan, Luxembourg, Latvia, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and United States). The two main questions under study are: 1) the relationship between per capita income and HCE; and 2) the influence of non-income variables on HCE. All variables were downloaded from the Worldbank dataset.

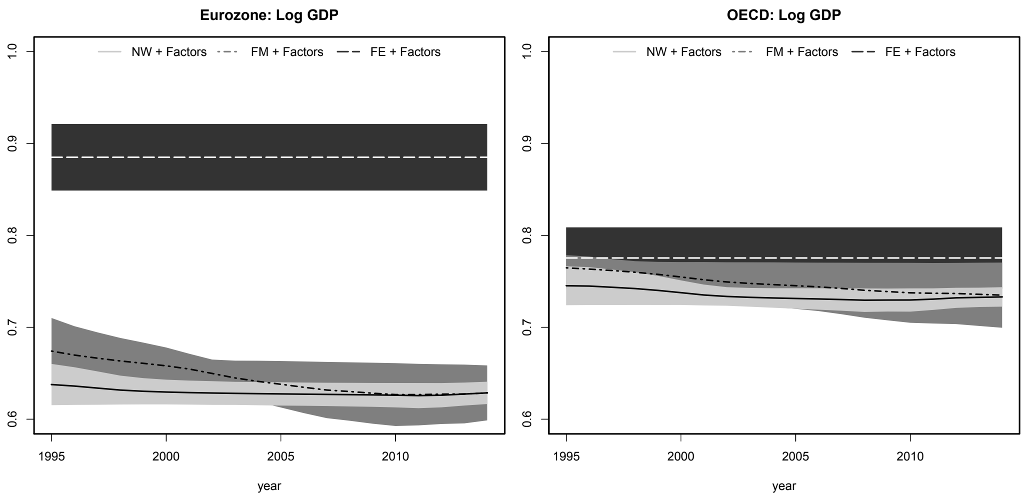
4.2. Is health care a luxury in developed countries?

Results in Table 1 show the estimates of models (32), (29) and (30) and their 95% confidence intervals in brackets. Functions in R package *plm* by Croissant and Millo (2008) are used to obtained coefficients of the last two models and also R package *phtt* is used to obtain the unobserved factors of (30). The *lgdp* coefficients of (32) for both the Eurozone and the OECD are over 1, even when other non-income variables are included in the model. This is in concordance with Kleiman (1974); Newhouse (1977); Leu (1986); Getzen (2000). Whereas, the FE estimates with time and individual fixed effects are significantly below 1 at 5% level. This differs from the results of Hitiris and Posnett (1992) who use a pooled estimator for panel data with two dummy variables to mimic α_i in model (29) and obtains values of $\hat{\beta}$ over 1. The OLS high R^2 adjusted and the non-stationarity variables suggest a case of spurious regression. The coefficients of models (30) are similar to those obtained by Baltagi and Moscone (2010), with long-run income elasticities under 1.

Focussing on panel data models with fixed effects, Figure 2 shows the income elasticity estimates of HCE from FE models with unobserved factors and the corresponding with time-varying coefficients, models (30) and (31) respectively. The estimated number of unobserved factors is calculated using criterion in Onatski (2010) and implemented in the *phtt* R package by Bada and Liebl (2014). The continuous blue line corresponds to the income elasticity estimates of model (30) and the light blue band is its 95% confidence interval. Similarly, the black line and grey confidence interval corresponds to the NW estimates of (31) and the red line with orange confidence intervals corresponds to FM estimates of (31). The latter corrects for the bias arising when there is correlation between the regressor innovations and the error term.

The FE with unobservable factors income elasticities (white line and black bands) are significantly smaller than 1 at 5% level, with values around 0.9 for the Eurozone and under 0.8

Fig. 2. Estimated income elasticity of health care expenditure during the period 1990-2014 for the Eurozone and the OECD countries (left to right). The plots show the estimated values of the coefficients at each year and their 95% confidence interval (bands). The white line with dark bands refer to the FE with unobserved factors models, the continuous line with light grey band refer to the NW estimates of a time-varying coefficient FE model with unobserved factors, and the dash line with medium grey band refers to the FM estimates of a time-varying coefficient FE model with unobserved factors.



for the OECD. This model implies that the expected price of health care is greater in the Eurozone than in the OECD countries. On the other hand, the time-varying coefficient models with unobservable factors (denoted as NW and FM below) show that the price of health care is lower for the Eurozone. The NW income elasticity estimates of HE (continuous line with light grey bands) are stable over the time period 1995-2014 with values a bit over 0.6 in the Eurozone and 0.75 in the OECD. The FM income elasticity estimates (dash lines with medium grey bands) are a bit higher than the NW estimates during the pre-GFC, but the 95% confidence intervals of both estimators overlap during the whole sample period. Most importantly, both estimates display a decreasing trend over time, which is explained by the fact that most countries in the Eurozone, first the Nordic countries and the rest after the GFC, have applied new health care policies to make their systems more efficient (Mladovsky et al. (2012)). This same trend appears in the time-varying coefficient estimates of the OECD, which in this case are not statistically different from the FE estimates.

4.3. Do age structure and public expenditure affect the price of health care?

As reported in Fisher (1990); Alemayehu and Warner (2004), the health bill of the young is the lowest in a health care system while the largest, by a large difference, 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). Thus, it is expected to obtain a positive relationship of HCE and $Pop65$ and a negative relationship of HCE and $Pop14$. In Table 1, as in Leu (1986); Hitiris and Posnett (1992), the OLS estimated coefficients of (32) are non-significant for $Pop65$ and $Pop14$ variables. $Pop65$ variable is significant for the FE and FE model with unobserved factors, but with counterintuitive negative signs. The NW and FM coefficient estimates of $Pop65$ in Figure 3 are positive and slightly increasing with higher values for the Eurozone. Regarding the $Pop14$ variable, the NW and FM coefficient estimates of model (31) are significant and, as expected, negative. This negative relationship is larger for the OECD.

Finally, all estimators report a positive significant relationship between government public investment and HCE (Table 1 and third row of Figure 3). Interestingly, the NW and FM estimates show a decreasing trend accentuated after the GFC. As Liaropoulos and Goranitis (2015) report, the source of financing health care is the core of all developed countries health policies. However, a universal health system based on employment contributions alone might not be feasible in a society that is becoming older, with people living longer and that has the same retirement age as before. The findings indicate that the latest reforms aiming at dissociate health care from public funding are working in the right direction.

In conclusion, the FM estimates of a time-varying coefficients FE model with unobserved factors report very fitting results to the initial expectations and in concordance with the latest health care policies. These estimates converge to the NW values at the end of the sample period. Although, the pointwise 95% confidence intervals do always overlap. The expected price of health care is lower in the Eurozone than in the OECD, although the Eurozone countries, in average, pay more to keep their elderly and young healthy. There is a similar relationship between the government financing of health care and the HCE in both regions.

5. Conclusions

A time-varying coefficient panel data model using fixed effects is estimated with nonparametric kernel smoothing techniques in this paper. The two main theoretical contributions are: i) the asymptotic theory of this estimator which shows a faster rate of convergence than other nonparametric estimators of non-linear models with time-varying coefficients and local stationary variables; and ii) the derivation of a second bias-corrected estimator to tackle the bias arising from the correlation between the regressor innovations and the equation error. In addition, the inclusion of a term of heterogenous unobserved factors has been proposed.

The application of these new methodologies to shine a light in the evolution of the price of health care in developed countries is the empirical contribution of this paper. The estimation of time-varying income elasticities of HCE show that health care is more expensive in the OECD than in the Eurozone, but in any case it is far away from becoming a luxury good. In fact, the income elasticities are decreasing. Age demographics and government funding rate are also significantly related to the HCE. In particular, the relationship between government funding and HCE is decreasing after the GFC.

There are some limitations in this paper. This paper assumes that the regressor innovations and equation error are individually independent but serially correlated. A future topic is to accommodate such dependence on nonparametric estimates of these panel data models.

Acknowledgements

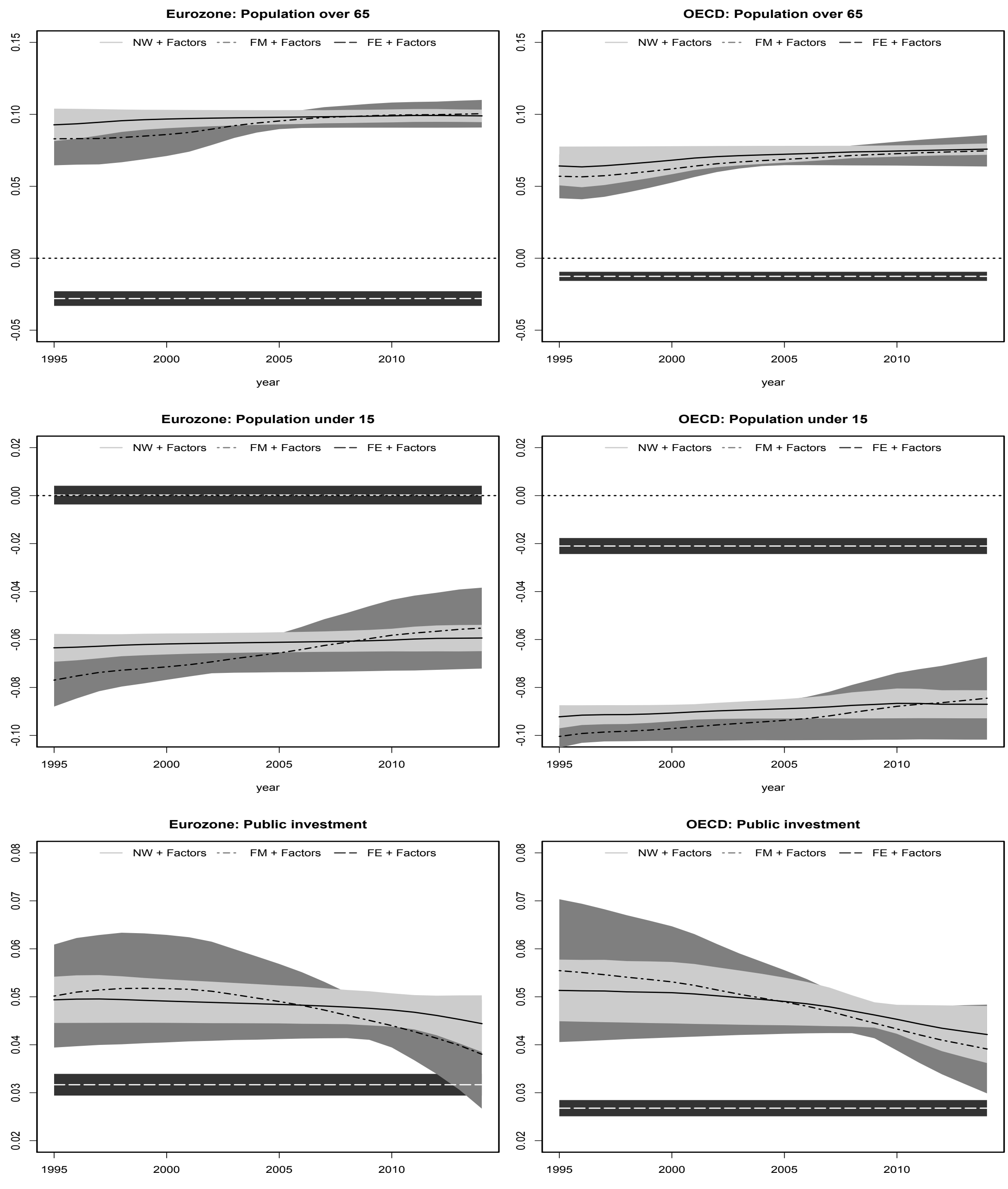
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n. 657182).

Table 1

OLS, FE and FE with unobservable factors coefficient estimates of log GDP as a predictor of HCE. The 95% confidence intervals are displayed in brackets. P-value codes: *** < 0.001, ** < 0.01, * < 0.05, . < 0.1.

	Eurozone			OECD		
	$\hat{b}_{\text{OLS}}$	$\hat{b}_{\text{FE}}$	$\hat{b}_{\text{FE+Factors}}$	$\hat{b}_{\text{OLS}}$	$\hat{b}_{\text{FE}}$	$\hat{b}_{\text{FE+Factors}}$
(Intercept)	-4.771**			-5.738***		
	(-7.951, -1.592)			(-8.075, -3.400)		
lgdp	1.092***	0.921***	0.885***	1.169***	0.810**	0.775***
	(0.860, 1.324)	(0.836, 1.005)	(0.848, 0.922)	(0.916, 1.423)	(0.719, 0.902)	(0.742, 0.809)
Pop65	0.017	-0.030***	-0.028***	0.027	-0.012***	-0.013***
	(-0.066, 0.099)	(-0.041, -0.017)	(-0.033, -0.023)	(-0.025, 0.079)	(-0.021, -0.004)	(-0.016, -0.009)
Pop14	0.004	0.001	0.000	0.011	-0.020***	-0.021 ***
	(-0.097, 0.104)	(-0.008, 0.010)	(-0.004, 0.004)	(-0.028, 0.051)	(-0.029, -0.011)	(-0.024, -0.018)
Public	0.058*	0.034 ***	0.032***	0.048*	0.028***	0.027***
	(0.013, 0.103)	(0.028, 0.039)	(0.029, 0.034)	(0.010, 0.085)	(0.024, 0.033)	(0.025, 0.028)
R^2 adj.	0.913	0.641	0.904	0.915	0.481	0.875

Fig. 3. Coefficient estimates of *Pop*65, *Pop*14 and *Public* variables (top to bottom) during the period 1990-2014 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 white line with dark bands refer to the FE with unobserved factors models, the continuous line with light grey band refer to the NW estimates of a time-varying coefficient FE model with unobserved factors, and the dash line with medium grey band refers to the FM estimates of a time-varying coefficient FE model with unobserved factors.



Proofs of main results

Proof of Theorem 1 (a). Observe that

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\delta_0)-\boldsymbol{\beta}(\delta_0)=&\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{\mathbf{Y}}-\boldsymbol{\beta}_0(\delta_0)\\
=&\left\{\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{B}(\mathbf{X},\beta)-\boldsymbol{\beta}(\delta_0)\right\}\\
&+\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(\delta_0)D\alpha+\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(\delta_0)\tilde{u}\\
=&\Xi_{NT}(1)+\Xi_{NT}(2)+\Xi_{NT}(3). \qquad (33)
\end{aligned}$$

By the definition of $\widetilde{W}^*(\delta_0)$, we have

$$\begin{aligned}
\Xi_{NT}(2)=&\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(\delta_0)D\alpha\\
=&\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{\mathcal{K}}(\delta_0)'\widetilde{W}(\delta_0)\left[\widetilde{\mathcal{K}}(\delta_0)D\alpha\right]. \qquad (34)
\end{aligned}$$

Observe that for any δ_0, $\widetilde{K}(\delta_0)D\alpha=0$. We have $\Xi_{NT}(2)=0$. For $\Xi_{NT}(1)$, by Assumption 2 and Taylor approximation (13), we find that

$$\begin{aligned}
\Xi_{NT}(1)=&\left[\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{B}(\mathbf{X},\boldsymbol{\beta})-\boldsymbol{\beta}(\delta_0)\\
=&O_P(h^\gamma). \qquad (35)
\end{aligned}$$

Then, using Lemmas 1-2, (35) in conjunction with the condition $N^{1/2}Th^{1+\gamma}=o(1)$, we can prove (22) in Theorem 1.

Lemma 1. *Suppose that Assumptions 1, 3 and 4 are satisfied. Then as $T,N\to\infty$ simultaneously, for any $0<\delta_0<1$*

$$\frac{1}{NT^2h^2}\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}\xrightarrow{p}(1-\mathcal{C}_K)\Omega_\nu, \qquad (36)$$

where $\Omega_\nu\equiv\Phi(1)'\Lambda_0\Phi(1)$ defined in (6), $\mathcal{C}_K=\int_{-1}^1\int_{-1}^1\min(s+1,r+1)K(s)K(r)dsdr$.

Proof. Take a neighborhood $\mathcal{N}_{T,\delta_0}=[\lfloor(\delta_0-h)T\rfloor,\lfloor(\delta_0+h)T\rfloor]$ of $\lfloor\delta_0T\rfloor$ and let $\delta(T)=\lfloor(\delta_0-h)T\rfloor$. From the BN decomposition (Phillips and Solo (1992)), we have for $t\geq\delta(T)$

$$\begin{aligned}
x_{i,t}=&\sum_{s=1}^t\nu_{i,s}+x_{i,0}=\sum_{s=1}^t\bar{\bar{\nu}}_{i,s}+\tilde{\nu}_{i,0}-\tilde{\nu}_{i,t}+x_{i,0}\\
=&[\sum_{s=1}^{\delta(T)}\bar{\bar{\nu}}_{i,s}+\tilde{\nu}_{i,0}-\tilde{\nu}_{i,\delta(T)}+x_{i,0}]+[\sum_{s=\delta(T)+1}^{t}\bar{\bar{\nu}}_{i,s}]+[\tilde{\nu}_{i,\delta(T)}-\tilde{\nu}_{i,t}]\\
\equiv&x_{i,\delta(T)}+\eta_{i,t}+\xi_{i,t}, \qquad (37)
\end{aligned}$$

where $\bar{\bar{\nu}}_{i,t} = (\sum_{j=0}^{\infty} \Phi'_j)\varepsilon_{i,t} \equiv \Phi(1)'\varepsilon_{i,t}$, and $\tilde{\nu}_{i,t} = \sum_{j=0}^{\infty} \tilde{\Phi}'_j \varepsilon_{i,t-j}$ with $\tilde{\Phi}_j = \sum_{k=j+1}^{\infty} \Phi_k$. Note that the summability condition $\sum_{j=0}^{\infty} j||\Phi_j|| < \infty$ in Assumption 3 ensures $\sum_{j=0}^{\infty} ||\tilde{\Phi}_j|| < \infty$ (Phillips and Solo (1992)), so that $\xi_{i,t} = O_p(1)$.

We first prove the following asymptotic representation

$$\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M} = \sum_{i=1}^{N}\sum_{t=1}^{T} K_{th}(\delta_0)\left(\eta_{i,t}\eta'_{i,t} - \bar{\eta}_i\bar{\eta}'_i\right) + O_p(N(Th)^{3/2} + T^2h), \tag{38}$$

where $\bar{\eta}_i = \frac{1}{Z_T}\sum_{t=1}^{T} K_{th}(\delta_0)\eta_{i,t}$ and $Z_T = \sum_{t=1}^{T} K_{th}(\delta_0)$.

By the definition $\widetilde{W}^*(\delta_0)$ and (17), we have

$$\begin{aligned}\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M} =& \widetilde{M}'\widetilde{W}(\delta_0)(\mathrm{I}_{NT} - \mathcal{S}_{NT})\widetilde{M}\\ =& \widetilde{M}'\widetilde{W}(\delta_0)\widetilde{M} - \widetilde{M}'\widetilde{W}(\delta_0)\mathcal{S}_{NT}\widetilde{M},\end{aligned}$$

where $\mathcal{S}_{NT} = D[D'\widetilde{W}(\delta_0)D]^{-1}D'\widetilde{W}(\delta_0)$. We will prove that

$$\begin{aligned}\widetilde{M}'\widetilde{W}(\delta_0)\mathcal{S}_{NT}\widetilde{M} =& \frac{1}{Z_T}\sum_{i=1}^{N}\left[\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]\left[\sum_{t=1}^{T} x_{it}K_{th}(\delta_0)\right]'\\ &- \frac{1}{NZ_T}\left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]\left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]'.\end{aligned} \tag{39}$$

To do so, we first consider the term $[D'\widetilde{W}(\delta_0)D]^{-1}$. We have

$$\left[D'\widetilde{W}(\delta_0)D\right]^{-1} = \begin{pmatrix} \frac{1}{Z_T} - \frac{1}{NZ_T} & -\frac{1}{NZ_T} & \cdots & -\frac{1}{NZ_T} \\ -\frac{1}{NZ_T} & \frac{1}{Z_T} - \frac{1}{NZ_T} & \cdots & -\frac{1}{NZ_T} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{NZ_T} & -\frac{1}{NZ_T} & \cdots & \frac{1}{Z_T} - \frac{1}{NZ_T} \end{pmatrix}.$$

By standard arguments, we have

$$\begin{aligned}&\widetilde{M}'\widetilde{W}(\delta_0)\mathcal{S}_{NT}\widetilde{M}(\delta_0) = \widetilde{M}'\widetilde{W}(\delta_0)D[D'\widetilde{W}(\delta_0)D]^{-1}D'\widetilde{W}(\delta_0)\widetilde{M}\\ =& \frac{1}{Z_T}\sum_{i=1}^{N}\sum_{s,t=1}^{T} x_{i,t}x'_{is}K_{th}(\delta_0)K_{sh}(\delta_0) - \frac{1}{NZ_T}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T} x_{i,t}x'_{js}K_{sh}(\delta_0)K_{th}(\delta_0)\\ =& \frac{1}{Z_T}\sum_{i=1}^{N}\left[\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]\left[\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]' - \frac{1}{NZ_T}\left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]\left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0)\right]'.\end{aligned}$$

Combining (37) and (39), we have

$$
\begin{aligned}
&\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M} = \widetilde{M}'\widetilde{W}(\delta_0)(\mathrm{I}_{NT} - \mathcal{S}_{NT})\widetilde{M} \\
=&\sum_{i=1}^N \sum_{t=1}^T x_{i,t}x_{i,t}' K_{th}(\delta_0) - \frac{1}{Z_T}\sum_{i=1}^N \left[\sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]\left[\sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]' \\
&+ \frac{1}{NZ_T}\left[\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]\left[\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]' \\
=&\sum_{i=1}^N \sum_{t=1}^T (x_{i,t} - x_{i,\delta(T)})(x_{i,t} - x_{i,\delta(T)})' K_{th}(\delta_0) + \frac{1}{NZ_T}\left[\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]\left[\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]' \\
&- \frac{1}{Z_T}\sum_{i=1}^N \left\{\sum_{t=1}^T (x_{i,t} - x_{i,\delta(T)})K_{th}(\delta_0)\right\}\left\{\sum_{t=1}^T (x_{i,t} - x_{i,\delta(T)})' K_{th}(\delta_0)\right\} \\
=&\sum_{i=1}^N \sum_{t=1}^T \left\{\eta_{i,t}\eta_{i,t}' - \bar{\eta}_i\bar{\eta}_i'\right\} K_{th}(\delta_0) + 2\sum_{i=1}^N \sum_{t=1}^T \eta_{i,t}\xi_{i,t}' K_{th}(\delta_0) + \sum_{i=1}^N \sum_{t=1}^T \xi_{i,t}\xi_{i,t}' K_{th}(\delta_0) \\
&- \frac{2}{Z_T}\sum_{i=1}^N \left\{\sum_{t=1}^T K_{th}(\delta_0)\eta_{i,t}\right\}\left\{\sum_{t=1}^T K_{th}(\delta_0)\xi_{i,t}'\right\} - \frac{1}{Z_T}\sum_{i=1}^N \left\{\sum_{t=1}^T K_{th}(\delta_0)\xi_{i,t}\right\}\left\{\sum_{t=1}^T K_{th}(\delta_0)\xi_{i,t}'\right\} \\
&+ \frac{1}{NZ_T}\left[\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]\left[\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}\right]' \\
=&\sum_{i=1}^N \sum_{t=1}^T \left(\eta_{i,t}\eta_{i,t}' - \bar{\eta}_i\bar{\eta}_i'\right) K_{th}(\delta_0) + 2\sum_{i=1}^N \sum_{t=1}^T \eta_{i,t}\xi_{i,t}' K_{th}(\delta_0) + \sum_{i=1}^N \sum_{t=1}^T \xi_{i,t}\xi_{i,t}' K_{th}(\delta_0) \\
&- 2Z_T\sum_{i=1}^N \bar{\eta}_i\bar{\xi}_i' - Z_T\sum_{i=1}^N \bar{\xi}_i\bar{\xi}_i' + NZ_T\bar{x}\bar{x}' \\
\equiv&\sum_{i=1}^N \sum_{t=1}^T \left(\eta_{i,t}\eta_{i,t}' - \bar{\eta}_i\bar{\eta}_i'\right) K_{th}(\delta_0) + 2R_{1,NT} + R_{2,NT} - 2R_{3,NT} - R_{4,NT} + R_{5,NT},
\end{aligned}
\tag{40}
$$

where $\bar{\xi}_i = \frac{1}{Z_T}\sum_{t=1}^T K_{th}(\delta_0)\xi_{i,t}$ and $\bar{x} = \frac{1}{NZ_T}\sum_{i=1}^N \sum_{t=1}^T K_{th}(\delta_0)x_{i,t}$.

Next, note that

$$
\begin{aligned}
E||R_{1,NT}|| \leq & \sum_{i=1}^{N}\sum_{t=1}^{T} K_{th}(\delta_0)\mathrm{E}||\eta_{it}||||\xi_{it}|| \\
\leq & \sqrt{2Th}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{th}(\delta_0)\mathrm{E}||\frac{\eta_{it}}{\sqrt{2Th}}||||\xi_{it}|| \\
\leq & \sqrt{2Th}\sum_{i=1}^{N}\sum_{t=1}^{T} K_{th}(\delta_0)\sqrt{\mathrm{E}||\frac{\eta_{it}}{\sqrt{2Th}}||^2\mathrm{E}||\xi_{it}||^2} \\
= & \sqrt{2Th}\sum_{t=1}^{T} K_{th}(\delta_0)O(\sum_{i=1}^{N}||\Lambda_i||) = O(N(Th)^{3/2}),
\end{aligned} \tag{41}
$$

and

$$
\begin{aligned}
& E||R_{3,NT}|| \leq \sum_{i=1}^{N} Z_T\mathrm{E}||\bar{\eta}_i||||\bar{\xi}_i|| \\
\leq & \sqrt{2}(Th)^{5/2}(Z_T)^{-1}\sum_{i=1}^{N}\left\{\mathrm{E}||\frac{1}{Th}\sum_{t=1}^{T}\frac{\eta_{it}}{\sqrt{2Th}}K_{th}(\delta_0)||^2\mathrm{E}||\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\xi_{it}||^2\right\}^{1/2} \\
= & \sqrt{2}(Th)^{5/2}(Z_T)^{-1}O(\sum_{i=1}^{N}||\Lambda_i||) = O(N(Th)^{3/2}).
\end{aligned} \tag{42}
$$

Similar argument above, we can show that

$$
E||R_{2,NT}|| = O(NTh), \qquad \text{and} \qquad E||R_{4,NT}|| = O(NTh). \tag{43}
$$

Noting that $\{\sum_{i=1}^{N}\varepsilon_{i,t}, F_{t,N}\}$ is a martingale difference array with mean 0, where $F_{t,N} = \sigma\{\varepsilon_{i,s} : 1 \leq i \leq N, 1 \leq s \leq t\}$. We next use the central limit theorem for a martingale difference array (Hall and Heyde (1980)), we can prove that

$$
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} x_{i,\delta(T)} \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{\delta(T)}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Phi(1)'\varepsilon_{i,t}\right) \xrightarrow{\mathcal{L}} N(0, \delta_0\Phi(1)'\Sigma_\Lambda\Phi(1)),
$$

as both T and N tend to infinity. So we have

$$
\begin{aligned}
& \sum_{i=1}^{N}\sum_{t=1}^{T} x_{i,t}K_{th}(\delta_0) = \sum_{i=1}^{N} x_{i,\delta(T)}\sum_{t=1}^{T} K_{th}(\delta_0) + \sum_{i=1}^{N}\sum_{t=1}^{T}(\eta_{it}+\xi_{it})K_{th}(\delta_0) \\
= & O_p(Th\sqrt{NT} + N^{1/2}(Th)^{3/2}),
\end{aligned}
$$

and the we have

$$
\begin{aligned}
R_{5,NT} = & \frac{1}{NZ_T}\left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}K_{th}(\delta_0)\right]\left[\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}K_{th}(\delta_0)\right]^\top \\
= & O(\frac{1}{NTh}) \times O_P((Th\sqrt{NT})^2 + (N^{1/2}(Th)^{3/2})^2) = O_P(T^2h).
\end{aligned} \tag{44}
$$

Combing (40)–(44), (38) is proved.

By assumption 4, we have $O_P(N(Th)^{3/2}+T^2h)=o_p(NT^2h^2)$ in (38). Next we only need to prove

$$\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\xrightarrow{p}\Omega_\nu(1-\mathcal{C}_K),\tag{45}$$

as $N,T\to\infty$ simultaneously. Then we only need to prove that

$$E\left[\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\right]\longrightarrow\Omega_\nu(1-\mathcal{C}_K),\tag{46}$$

and

$$E\left(\left[\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\right]-E\left[\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\right]\right)^2\longrightarrow 0,\tag{47}$$

where

$$\begin{aligned}
&E\left[\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\right]\\
=&\frac{\Phi'(1)}{NT^2h^2}\sum_{t=1}^{T}K_{th}(\delta_0)\sum_{i=1}^{N}E\left[\sum_{s=\delta(T)+1}^{t}\varepsilon_{i,s}\varepsilon_{i,s}'-\frac{1}{Z_T^2}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)\sum_{l=\delta(T)+1}^{s\wedge t}\varepsilon_{i,l}\varepsilon_{i,l}'\right]\Phi(1)\\
=&\frac{\Phi'(1)}{T^2h^2}\sum_{t=1}^{T}K_{th}(\delta_0)\left[(t-\delta(T)-1)-\frac{1}{Z_T^2}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)(s\wedge t-\delta(T)-1)\right](\frac{1}{N}\sum_{i=1}^{N}\Lambda_i)\Phi(1)\\
\longrightarrow&(1-\mathcal{C}_K)\Phi'(1)\Lambda_0\Phi(1)=(1-\mathcal{C}_K)\Omega_\nu,
\end{aligned}$$

and

$$\begin{aligned}
&E\left[\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\right]^2\\
=&\frac{1}{(NT^2h^2)^2}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)E\left[\left(\eta_{i,t}\eta_{i,t}'-\bar{\eta}_i\bar{\eta}_i'\right)\left(\eta_{j,s}\eta_{j,s}'-\bar{\eta}_j\bar{\eta}_j'\right)\right]\\
=&\frac{1}{(NT^2h^2)^2}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)E(\eta_{i,t}\eta_{i,t}'\eta_{j,s}\eta_{j,s}')+\frac{1}{(NT^2h^2)^2}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)E(\bar{\eta}_i\bar{\eta}_i'\bar{\eta}_j\bar{\eta}_j')\\
&-\frac{2}{(N^2T^2h^2)^2}Z_T\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)E(\eta_{i,t}\eta_{i,t}'\bar{\eta}_j\bar{\eta}_j')\doteq I(1)+I(2)+I(3),
\end{aligned}$$

with

$$
\begin{aligned}
I(1) =& \frac{1}{(NT^2h^2)^2}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)E(\eta_{i,t}\eta_{i,t}'\eta_{j,s}\eta_{j,s}') \\
=& \frac{1}{(NT^2h^2)^2}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}K_{th}(\delta_0)K_{sh}(\delta_0)E\left(\Phi'(1)\sum_{t_1,t_2=\delta(T)+1}^{t}\varepsilon_{i,t_1}\varepsilon_{i,t_2}\Phi(1)\Phi'(1)\sum_{s_1,s_2=\delta(T)+1}^{s}\varepsilon_{j,s_1}\varepsilon_{j,s_2}\Phi(1)\right) \\
\leq& \frac{1}{N^2T^2h^2}\sum_{s,t=1}^{T}\frac{K_{th}(\delta_0)}{Th}\frac{K_{sh}(\delta_0)}{Th}\sum_{i,j=1}^{N}\sum_{t_1,t_2,s_1,s_2=\delta(T)+1}^{t\vee s}E\left(\Phi'(1)\varepsilon_{i,t_1}\varepsilon_{i,t_1}'\Phi(1)\Phi'(1)\varepsilon_{j,s_1}\varepsilon_{j,s_1}'\Phi(1)\right) \\
=& \frac{1}{N}\sum_{s,t=1}^{T}\frac{K_{th}(\delta_0)}{Th}\frac{K_{sh}(\delta_0)}{Th}\left(\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{k,l=\delta(T)+1}^{t\vee s}E(\Phi'(1)\varepsilon_{i,k}\varepsilon_{i,k}'\Phi(1)\Phi'(1)\varepsilon_{i,l}\varepsilon_{i,l}'\Phi(1))\right. \\
&\left. + \frac{1}{NT^2h^2}\sum_{i,j=1}^{N}\sum_{k,l=\delta(T)+1}^{t\vee s}E\left(\Phi'(1)\varepsilon_{i,k}\varepsilon_{i,l}'\Phi(1)\Phi'(1)\varepsilon_{j,k}\varepsilon_{j,l}'\Phi(1)\right)\right) \longrightarrow 0. \qquad (48)
\end{aligned}
$$

Similar argument with (48), we can also prove that $I(2) \to 0$, and $I(3) \to 0$. Then (47) is proved. Thus, (36) is completely proved. $\square$

Lemma 2. *Suppose that Assumptions 1, 3-4 are satisfied. Then, we have, for any* $0 < \delta_0 < 1$,

$$
\frac{1}{\sqrt{N}Th}\left\{\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{u} - N\sum_{t=1}^{T}K_{th}(\delta_0)(\Delta_{\nu u} - \Delta_{\nu\bar{u}}^{t}(\delta_0))\right\} \xRightarrow{\mathcal{L}} N(0, \mathcal{C}_{K^*}\Sigma_{\nu,u}), \qquad (49)
$$

as $T, N \to \infty$ *simultaneously, where* $\mathcal{C}_{K^*} \equiv \mathcal{C}_{K^*}(1) + \mathcal{C}_{K^*}(2) - 2\mathcal{C}_{K^*}(1,2)$ *with* $\mathcal{C}_{K^*}(1) = \nu_0$,

$$
\mathcal{C}_{K^*}(2) = \nu_0\int_{-1}^{1}(\int_{s}^{1}K(t)dt)^2ds + \left(\int_{-1}^{1}K(t)(\int_{t}^{1}K(u)du)dt\right)^2,
$$

and

$$
\mathcal{C}_{K^*}(1,2) = \int_{-1}^{1}K^2(t)(\int_{-1}^{t}\int_{s}^{1}K(u)duds)dt + \int_{-1}^{1}K(t)(\int_{t}^{1}K(u)du)(\int_{-1}^{t}K(s)ds)dt,
$$

$\Delta_{\nu u}$ *and* $\Delta_{\nu\bar{u}}^{t}$ *are defined in* (8) *and* (20)*, respectively.*

Proof. Similar argument with $\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{M}$ in Lemma 1 and denote $\bar{u}_i = \frac{1}{Z_T}\sum_{t=1}^{T}K_{th}(\delta_0)u_{i,t}$, we have

$$
\begin{aligned}
&\widetilde{M}'\widetilde{W}^*(\delta_0)\widetilde{u} = \widetilde{M}'\widetilde{W}(\delta_0)\widetilde{u} - \widetilde{M}'\widetilde{W}(\delta_0)\mathcal{S}_{NT}\widetilde{u} \\
=& \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - x_{i,\delta(T)})u_{it}K_{th}(\delta_0) - \frac{1}{Z_T}\sum_{i=1}^{N}\left\{\sum_{t=1}^{T}(x_{it} - x_{i,\delta(T)})K_{th}(\delta_0)\right\}\left\{\sum_{t=1}^{T}u_{it}K_{th}(\delta_0)\right\} \\
&+ \frac{1}{NZ_T}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}K_{th}(\delta_0)\right]\left[\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}K_{th}(\delta_0)\right] \\
=& \sum_{i=1}^{N}\sum_{t=1}^{T}(\eta_{it} + \xi_{it})u_{it}K_{th}(\delta_0) - \frac{1}{Z_T}\sum_{i=1}^{N}\left\{\sum_{t=1}^{T}(\eta_{it} + \xi_{it})K_{th}(\delta_0)\right\}\left\{\sum_{t=1}^{T}u_{it}K_{th}(\delta_0)\right\} \\
&+ \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{K_{th}(\delta_0)}{Z_T}x_{it}\right]\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}K_{th}(\delta_0)\right] \\
=& \sum_{i=1}^{N}\sum_{t=1}^{T}(\eta_{i,t} + \xi_{i,t})(u_{it} - \bar{u}_i)K_{th}(\delta_0) + O_p(Th^{1/2}) \\
\equiv& \sum_{i=1}^{N}\sum_{t=1}^{T}S_{i,t}(u_{it} - \bar{u}_i)K_{th}(\delta_0) + O_p(Th^{1/2}).
\end{aligned} \tag{50}
$$

Next, we only need to prove that

$$
\frac{1}{\sqrt{N}Th}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}S_{i,t}(u_{it} - \bar{u}_i)K_{th}(\delta_0) - N\sum_{t=1}^{T}K_{th}(\delta_0)(\Delta_{\nu u} - \Delta_{\nu\bar{u}}^{t})\right\} \stackrel{\mathcal{L}}{\Longrightarrow} N(0, \mathcal{C}_{K^*}\Sigma_{\nu,u}). \tag{51}
$$

From the BN decomposition, we have for $t \geq \delta(T)$, $u_{i,t} = \bar{\bar{u}}_{i,t} + (\tilde{u}_{i,t-1} - \tilde{u}_{i,t})$, where $\bar{\bar{u}}_{i,t} = (\sum_{j=0}^{\infty}\psi_j')\varepsilon_t \equiv \psi(1)'\varepsilon_{i,t}$, and $\tilde{u}_{i,t} = \sum_{j=0}^{\infty}\tilde{\psi}_j'\varepsilon_{i,t-j}$ with $\tilde{\psi}_j = \sum_{k=j+1}^{\infty}\psi_k$. Note that

$$
\begin{aligned}
&\sum_{t=1}^{T}K_{th}(\delta_0)S_{i,t}\Delta\tilde{u}_{i,t} = \sum_{t=1}^{T}K_{th}(\delta_0)S_{i,t}\tilde{u}_{i,t} - \sum_{t=1}^{T}K_{th}(\delta_0)S_{i,t}\tilde{u}_{i,t-1} \\
=& \sum_{t=1}^{T}K_{th}(\delta_0)S_{i,t}\tilde{u}_{i,t} - \sum_{t=1}^{T}K_{th}(\delta_0)\nu_{i,t}\tilde{u}_{i,t-1} - \left\{\sum_{t=1}^{T}K_{(t-1)h}(\delta_0)S_{i,t-1}\tilde{u}_{i,t-1} + \sum_{t=1}^{T}\Delta K_{th}(\delta_0)S_{i,t-1}\tilde{u}_{i,t-1}\right\} \\
=& K_{Th}(\delta_0)S_{i,T}\tilde{u}_{i,T} - \sum_{t=1}^{T}K_{th}(\delta_0)\nu_{i,t}\tilde{u}_{i,t-1} - \sum_{t=1}^{T}\Delta K_{th}(\delta_0)S_{i,t-1}\tilde{u}_{i,t-1},
\end{aligned}
$$

and

$$\begin{aligned}&\sum_{t=1}^{T} K_{th}(\delta_0)\Delta\tilde{u}_{i,t}\\ =&\sum_{t=1}^{T} K_{th}(\delta_0)\tilde{u}_{i,t} - \sum_{t=1}^{T}\left\{K_{(t-1)h}(\delta_0) + (K_{th}(\delta_0) - K_{(t-1)h}(\delta_0))\right\}\tilde{u}_{i,t-1}\\ =&K_{Th}(\delta_0)\tilde{u}_{i,T} - \sum_{t=1}^{T}\Delta K_{th}(\delta_0)\tilde{u}_{i,t-1}.\end{aligned}$$

By virtue of Assumption 1, $K_{Th}(\delta_0) = 0$ with probability 1, which indicates that

$$\sum_{t=1}^{T} K_{th}(\delta_0)S_{i,t}(-\Delta\tilde{u}_{i,t}) = \sum_{t=1}^{T} K_{th}(\delta_0)\nu_{i,t}\tilde{u}_{i,t-1} + \sum_{t=1}^{T}\Delta K_{th}(\delta_0)S_{i,t-1}\tilde{u}_{i,t-1} \tag{52}$$

and

$$\sum_{t=1}^{T} K_{th}(\delta_0)(-\Delta\tilde{u}_{i,t}) = \sum_{t=1}^{T}\Delta K_{th}(\delta_0)\tilde{u}_{i,t-1}. \tag{53}$$

Let $V_{i,t} = \sum_{j=\delta(T)+1}^{t}\varepsilon_{i,j}$. Using BN decomposition again, $S_{i,t} = x_{i,t} - x_{i,\delta(T)} = \Phi(1)'V_{i,t} + \tilde{\nu}_{i,\delta(T)} - \tilde{\nu}_{i,t}$, and (52), we have

$$\begin{aligned}&\sum_{t=1}^{T} K_{th}(\delta_0)S_{i,t}u_{i,t} = \sum_{t=1}^{T} K_{th}(\delta_0)\left(S_{i,t}\varepsilon_{i,t}'\psi(1) - S_{i,t}\Delta\tilde{u}_{i,t}\right)\\ =&\sum_{t=1}^{T} K_{th}(\delta_0)(\Phi(1)'V_{i,t} + \tilde{\nu}_{i,\delta(T)} - \tilde{\nu}_{i,t})\varepsilon_{i,t}'\psi(1) + \left(\sum_{t=1}^{T} K_{th}(\delta_0)\nu_{i,t}\tilde{u}_{i,t-1} + \sum_{t=1}^{T}\Delta K_{th}(\delta_0)S_{i,t-1}\tilde{u}_{i,t-1}\right),\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\left(S_{i,t}u_{i,t} - \Delta_{\nu u}\right)\right)\\ =&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\Phi(1)'V_{i,t-1}\varepsilon_{i,t}'\psi(1)\right) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\Phi(1)'(\varepsilon_{i,t}\varepsilon_{i,t}' - \Lambda_0)\psi(1)\right)\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T-1} K_{th}(\delta_0)(\tilde{u}_{i,t}\nu_{i,t+1} - \sum_{j=0}^{\infty}\tilde{\psi}_j'\Phi_{j+1})\right) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\psi(1)'(\varepsilon_{i,t}\tilde{\nu}_{i,t} - \tilde{\Phi}_0)\right)\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}\Delta K_{th}(\delta_0)S_{i,t-1}\tilde{u}_{t-1}\right) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\tilde{\nu}_{i,\delta(T)}\varepsilon_{i,t}'\psi(1)\right)\\ \equiv&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\Phi(1)'V_{i,t-1}\varepsilon_{i,t}'\psi(1) + R_{1,i,T}(1) + R_{1,i,T}(2) + R_{1,i,T}(3) + O_p(\sqrt{\frac{N}{Th}})\right).\end{aligned} \tag{54}$$

We show that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{1,i,T}(k) \to_p 0, k = 1,2,3$ as $N, T \to \infty$ with $N/Th \to 0$. Note that

$$\begin{aligned}
&\mathrm{E}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} R_{1,i,T}(1)\right|\right|^2 \leq \frac{1}{N}\sum_{i,j=1}^{N} E||R_{1,i,T}R'_{1,j,T}|| \\
\leq&\frac{1}{N}\sum_{i,j=1}^{N}\frac{1}{T^2h^2}\sum_{t=1}^{T} K_{th}^2(\delta_0)E||(\varepsilon_{i,t}\varepsilon'_{i,t}-\Lambda_0)(\varepsilon_{j,t}\varepsilon'_{j,t}-\Lambda_0)'||||\Phi(1)||^2||\psi(1)||^2 \\
=&O(\frac{1}{Th}).
\end{aligned} \tag{55}$$

Thus, $(1/\sqrt{N})\sum_{i=1}^{N} R_{1,i,T}(1) = o_p(1)$. Next, we show that $(1/\sqrt{N})\sum_{i=1}^{N} R_{1,i,T}(2) \to_p 0$ by proving $\mathrm{E}||(1/\sqrt{N})\sum_{i=1}^{N} R_{1,i,T}(2)||^2 \to 0$ as $N, T \to \infty$. Note that $\mathrm{E}||(1/\sqrt{N})\sum_{i=1}^{N} R_{1,i,T}(2)||^2 \leq NE||R_{1,i,T}(2)||^2$ and following lemma 16 on pp.1105 in Phillips and Moon (1999),

$$\begin{aligned}
&\mathrm{E}\left||R_{1,i,T}(2)|\right|^2 = \mathrm{tr}\left(\mathrm{E}(vec(R_{1,i,T}(2))vec(R_{1,i,T}(2))')\right) \quad \text{since } \mathrm{E}(R_{1,i,T}(2)) = 0 \\
=&\frac{1}{T^2h^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1} K_{th}(\delta_0)K_{sh}(\delta_0)\mathrm{E}\left\{\begin{array}{c} (\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\tilde{\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}\tilde{\psi}_q) \\ -(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})' \end{array}\right\} \\
=&\frac{2}{T^2h^2}\sum_{t=\delta(T)+1}^{\lfloor 2Th\rfloor} K_{th}(\delta_0)\sum_{l=0}^{\lfloor 2Th\rfloor-1} K_{(t+l)h}(\delta_0)\left(\begin{array}{c} \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \\ \mathrm{E}(\tilde{\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}\tilde{\psi}_q) \\ -(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})' \end{array}\right) \\
=&\frac{2}{T^2h^2}\sum_{t=\delta(T)+1}^{\lfloor 2Th\rfloor} K_{th}^2(\delta_0)\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} \\ \mathrm{E}(\tilde{\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}\tilde{\psi}_q) \\ -(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})' \end{array}\right) \\
&+\frac{2}{T^2h^2}\sum_{t=\delta(T)+1}^{\lfloor 2Th\rfloor} K_{th}^2(\delta_0)\sum_{l=0}^{\lfloor 2Th\rfloor-1}\Delta K_{th}^l(\delta_0)\left(\begin{array}{c} \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \\ \mathrm{E}(\tilde{\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}\tilde{\psi}_q) \\ -(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})' \end{array}\right)
\end{aligned}$$

where $\Delta K_{th}^l(\delta_0) = K_{(t+l)h}(\delta_0) - K_{th}(\delta_0) < \frac{l}{Th}$. If we show

$$\sum_{l=0}^{\infty}\left(\begin{array}{c} \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \\ \mathrm{E}(\tilde{\psi}'_j\varepsilon_{i,t-j}\varepsilon'_{i,t+1-k}\Phi_k)(\Phi'_p\varepsilon_{i,t+l+1-p}\varepsilon'_{i,t+l-q}\tilde{\psi}_q) \\ -(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})(\sum_{j=0}^{\infty}\tilde{\psi}'_j\Phi_{j+1})' \end{array}\right) < \infty,$$

then it follows that $\mathrm{E}||(1/\sqrt{N})\sum_{i=1}^{N} R_{1,i,T}(2)||^2 = O(\frac{N}{Th}) \to 0$, observe that

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

$$\begin{aligned}
=&\sum_{l=0}^{\infty}\left(\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\mathrm{tr}(\Phi_k'\Phi_{k+l}\otimes\tilde{\psi}_j'\tilde{\psi}_{j+l})\right)+\sum_{l=0}^{\infty}\left(\sum_{j=0}^{\infty}\sum_{k=0\vee(1-l)}^{\infty}\mathrm{tr}\left\{(\Phi_k'\tilde{\psi}_{k+l-1}\otimes\tilde{\psi}_j'\Phi_{j+l+1})K_d\right\}\right) \\
&+(\nu^4-3)\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\mathrm{tr}\left((\Phi_{j+1}'\otimes\tilde{\psi}_j')(\sum_{l=1}^{d}e_{l,l}\otimes e_{l,l})(\Phi_{j+l+1}\otimes\tilde{\psi}_{j+l})\right) \\
=&\sum_{l=0}^{\infty}\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\mathrm{tr}(\Phi_k'\Phi_{k+l})\tilde{\psi}_j'\tilde{\psi}_{j+l}\right)+\sum_{l=0}^{\infty}\left(\sum_{j=0}^{\infty}\sum_{k=0\vee(1-l)}^{\infty}\tilde{\psi}_{k+l-1}'\Phi_k\Phi_{j+l+1}'\tilde{\psi}_j\right) \\
&+(\nu^4-3)\sum_{l=0}^{\infty}\sum_{j=0}^{\infty}\mathrm{tr}\left((\Phi_{j+1}'\otimes\tilde{\psi}_j')(\sum_{l=1}^{d}e_{l,l}\otimes e_{l,l})(\Phi_{j+l+1}\otimes\tilde{\psi}_{j+l})\right) \\
\equiv&\mathrm{I}+\mathrm{II}+\mathrm{III},
\end{aligned}$$

where $e_{l,l}$ is the $(d\times d)$ matrix where the $(l,l)^{th}$ element is one and other elements are zeros. Since $\mathrm{tr}(A\otimes B)=\mathrm{tr}(A)\mathrm{tr}(B)$ and $\mathrm{tr}(A)\le(\mathrm{rows}(A))^{1/2}||A||$ (see lemma 9 in Phillips and Moon (1999)), we have

$$\begin{aligned}
\mathrm{I}&=\sum_{l=0}^{\infty}\mathrm{tr}(\sum_{k=0}^{\infty}\Phi_k'\Phi_{k+l})(\sum_{j=0}^{\infty}\tilde{\psi}_j'\tilde{\psi}_{j+l}) \\
&\le\left[\sum_{l=0}^{\infty}\left|\mathrm{tr}(\sum_{k=0}^{\infty}\Phi_k'\Phi_{k+l})\right|\right]\left[\sum_{l=0}^{\infty}\left|\sum_{j=0}^{\infty}\mathrm{tr}(\tilde{\psi}_j\tilde{\psi}_{j+l}')\right|\right] \\
&\le d\left(\sum_{k=0}^{\infty}||\Phi_k||\right)^2\left(\sum_{k=0}^{\infty}||\tilde{\psi}_k||\right)^2<\infty, \qquad \text{by Assumption 3}
\end{aligned}$$

and

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

Similarly, we can show that for some $M > 0$

$$\mathrm{III} \leq M \left(\sum_{j=0}^{\infty} ||\Phi_j|| \right)^2 \left(\sum_{j=0}^{\infty} ||\psi_j|| \right)^2 < \infty.$$

Thus, we prove that

$$\mathrm{E}||(1/\sqrt{N}) \sum_{i=1}^{N} R_{1,i,T}(2)||^2 = O(\frac{1}{Th}) \to 0. \tag{56}$$

Also, we can show by modifying the arguments used above that

$$\mathrm{E} \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R_{1,i,T}(3) \right\|^2 = O(\sqrt{\frac{1}{Th}}),$$

which combining (54)-(56), we have the first term in (50)

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{1}{Th} \sum_{t=1}^{T} K_{th}(\delta_0) \left(S_{i,t} u_{it} - \Delta_{\nu u} \right) \right) \\
=&\frac{1}{Th} \sum_{t=1}^{T} K_{th}(\delta_0) \Phi(1)' \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} V_{i,t-1} \varepsilon_{i,t}' \right) \psi(1) + O_p(\sqrt{\frac{N}{Th}}) \\
=&\frac{1}{Th} \sum_{t=1}^{T} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{s<t} K_{th}(\delta_0) \Phi(1)' \varepsilon_{i,s} \varepsilon_{i,t}' \psi(1) \right) \\
\overset{a.s.}{=}&\frac{1}{Th} \sum_{t=1}^{T} W_{1,t,N}.
\end{aligned} \tag{57}$$

Similar argument with (54), we next prove the sencod term in (50)

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{1}{Th} \sum_{t=1}^{T} K_{th}(\delta_0) \left(S_{i,t} \bar{u}_i - \Delta_{\nu \bar{u}}^{t}(\delta_0) \right) \right) \\
=&\frac{1}{Th} \sum_{t=1}^{T} K_{th}(\delta_0) \Phi(1)' \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{1}{Z_T} \sum_{l=1}^{T} \sum_{l \neq s \leq t} K_{lh}(\delta_0) \varepsilon_{i,s} \varepsilon_{i,l}' \right) \right) \psi(1) + O_p(\sqrt{\frac{N}{Th}}) \\
\overset{a.s.}{\equiv}&\frac{1}{Th} \sum_{t=1}^{T} W_{2,t,T}.
\end{aligned} \tag{58}$$

Recall that $\sum_{t=1}^{T} K_{th}(\delta_0)(-\Delta \tilde{u}_{i,t}) = \sum_{t=1}^{T} \Delta K_{th}(\delta_0) \tilde{u}_{i,t-1}$ and denote $\bar{\varepsilon}_i = \frac{1}{Z_T} \sum_{t=1}^{T} K_{th}(\delta_0) \varepsilon_{i,t}$, Similar argument with (54) and using BN decomposition, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\left(S_{i,t}\bar{u}_i-\Delta_{\nu,\bar{u}}^{t}(\delta_0)\right)\right)\\
\equiv&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\left(S_{i,t}\bar{u}_i-\sum_{s=1}^{T}\frac{K_{sh}(\delta_0)}{Z_T}\sum_{j=\delta(T)+1}^{t}\mathrm{E}(\nu_{i,j}u_{i,s})\right)\right)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\Phi(1)'\left(\frac{1}{Z_T}\sum_{l=1}^{T}\sum_{l\neq s\leq t}K_{lh}(\delta_0)\varepsilon_{i,s}\varepsilon_{i,l}'\right)\psi(1)\right)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\Phi(1)'\left(\frac{1}{Z_T}\sum_{s\leq t}K_{sh}(\delta_0)\varepsilon_{i,s}\varepsilon_{i,s}'-\frac{1}{Z_T}\sum_{s\leq t}K_{sh}(\delta_0)\right)\psi(1)\right)\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\left(\tilde{\nu}_{i,t}\bar{\varepsilon}_i-\sum_{s\leq t}\frac{K_{sh}(\delta_0)}{Z_T}\tilde{\Phi}_{t-s}\right)'\psi(1)\right)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\left\{\frac{1}{Z_T}\sum_{s=1}^{T}\Delta K_{sh}(\delta_0)\left(S_{i,t}\tilde{u}_{i,s-1}-\mathrm{E}(S_{i,t}\tilde{u}_{i,s-1})\right)\right\}\right)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\tilde{\nu}_{i,\delta(T)}\bar{\varepsilon}_i'\psi(1)\right)\\
\equiv&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\Phi(1)'\left(\frac{1}{Z_T}\sum_{l=1}^{T}\sum_{l\neq s\leq t}K_{lh}(\delta_0)\varepsilon_{i,s}\varepsilon_{i,l}'\right)\psi(1)\\
+&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(R_{2,i,T}(1)+R_{2,i,T}(2)+O_p(\sqrt{\frac{N}{Th}})\qquad a.s., say.
\end{aligned}\tag{59}
$$

Let $\mathcal{W}_{t,s}=\frac{1}{Z_T}\left(K_{th}(\delta_0)\sum_{l\geq s}K_{lh}(\delta_0)+K_{sh}(\delta_0)\sum_{l\geq t}K_{lh}(\delta_0)\right)$, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{Th}\sum_{t=1}^{T}K_{th}(\delta_0)\Phi(1)'\left(\frac{1}{Z_T}\sum_{l=1}^{T}\sum_{l\neq s\leq t}K_{lh}(\delta_0)\varepsilon_{i,s}\varepsilon_{i,l}'\right)\psi(1)\\
=&\frac{1}{\sqrt{N}Th}\frac{1}{Z_T}\Phi(1)'\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\sum_{s\neq t}K_{th}(\delta_0)\sum_{l\geq s}K_{lh}(\delta_0)\varepsilon_{i,s}\varepsilon_{i,t}'\right)\psi(1)\\
=&\frac{1}{2\sqrt{N}Th}\frac{1}{Z_T}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}\left\{K_{th}(\delta_0)\sum_{l\geq s}K_{lh}(\delta_0)+K_{sh}(\delta_0)\sum_{l\geq t}K_{lh}(\delta_0)\right\}\Phi(1)'\varepsilon_{i,s}\varepsilon_{i,t}'\psi(1)\\
=&\frac{1}{Th}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{s<t}\left\{\mathcal{W}_{t,s}(\delta_0)\Phi(1)'\varepsilon_{i,s}\varepsilon_{i,t}'\psi(1)\right\}\right)\\
\equiv&\frac{1}{Th}\sum_{t=1}^{T}W_{2,t,N}
\end{aligned}\tag{60}
$$

Next we only need to show that $1/\sqrt{N}\sum_{i=1}^{N} R_{2,i,T}(k) \to_p 0, k=1,2.$ which follows by modifying the arguments used above in (55) and (56). Thus, (58) is proved.

So far, combining (57) and (58), we have

$$\frac{1}{\sqrt{N}Th}\left\{\sum_{i=1}^{N}\sum_{t=1}^{T} S_{i,t}(u_{it}-\bar{u}_i)K_{th}(\delta_0) - N\sum_{t=1}^{T} K_{th}(\delta_0)(\Delta_{\nu u}-\Delta^t_{\nu,\bar{u}}(\delta_0))\right\}$$
$$\overset{a.s.}{\equiv}\frac{1}{Th}\sum_{t=1}^{T}(W_{1,t,T}-W_{2,t,T}) \equiv \frac{1}{Th}\sum_{t=1}^{T}\mathcal{U}_{t,T}.$$

To prove (51), we only need to prove

$$\frac{1}{Th}\sum_{t=1}^{T}\mathcal{U}_{t,N} \overset{\mathcal{L}}{\Longrightarrow} N(0, \mathcal{C}_{K^*}\Omega_\nu\Omega_u). \tag{61}$$

Denote $\mathcal{W}^*_{t,s}(\delta_0) = K_{th}(\delta_0) - \mathcal{W}_{t,s}(\delta_0)$, we have

$$\mathcal{U}_{t,N} \equiv W_{1,t,N} - W_{2,t,N}$$
$$= \sum_{i=1}^{N}\left(\frac{1}{\sqrt{N}}\sum_{s<t}\mathcal{W}^*_{t,s}(\delta_0)\Phi(1)'\varepsilon_{i,s}\varepsilon'_{i,t}\psi(1)\right)$$

$\{\mathcal{U}_{t,N}, F_{t,N}\}$ is a martingale difference array with mean 0, where $F_{t,N} = \sigma\{\varepsilon_{i,s} : 1 \le i \le N, 1 \le s \le t\}$ is a σ-filed. Next we use the central limit theory for martingale difference array. For any $\varepsilon > 0$, it is easy to check that

$$\frac{1}{T^2h^2}E(\mathcal{U}_{t,N}\mathcal{U}'_{t,N}I(||\mathcal{U}_{t,N}||^2 \ge \varepsilon Th)|F_{t-1,N}) \to 0,$$

as $N, T \to \infty$ simultaneously. Furthermore, we can also prove that

$$\frac{1}{T^2h^2}\sum_{t=1}^{T}E(\mathcal{U}_{t,N}\mathcal{U}'_{t,N}|F_{t-1,N})$$
$$= \frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}\mathcal{W}^{*2}_{t,s}(\delta_0)\left(\frac{1}{N}\sum_{i,j=1}^{N}\mathrm{E}\left(\Phi(1)'\varepsilon_{i,s}\varepsilon'_{i,t}\psi(1)\psi(1)'\varepsilon_{j,t}\varepsilon'_{j,s}\Phi(1)|F_{t-1,N}\right)\right)$$
$$= \mathcal{C}_{K^*}\Sigma_{\nu,u} + o_p(1)$$

because of

$$\frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}\mathcal{W}^{*2}_{t,s}(\delta_0) = \frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}(K_{th}(\delta_0) - \mathcal{W}_{t,s}(\delta_0))^2$$
$$= \frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}\left(K^2_{th}(\delta_0) + \mathcal{W}^2_{t,s}(\delta_0) - 2\ K_{th}(\delta_0)\mathcal{W}_{t,s}(\delta_0)\right)$$
$$\to \mathcal{C}_{K^*} \equiv \mathcal{C}_{K^*}(1) + \mathcal{C}_{K^*}(2) - 2\mathcal{C}_{K^*}(1,2),$$

where

$$\frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}K_{th}^2(\delta_0)=\frac{1}{T^2h^2}\sum_{t=1}^{T}K_{th}^2(\delta_0)(t-\delta(T))\to\int_{-1}^{1}K^2(u)du\equiv\nu_0\equiv\mathcal{C}_{K^*}(1),$$

and

$$\begin{aligned}
&\frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}\mathcal{W}_{t,s}^2(\delta_0)=\frac{1}{T^2h^2}\frac{1}{2}\sum_{t=1}^{T}\sum_{s\neq t}\mathcal{W}_{t,s}^2(\delta_0)\\
=&\frac{1}{T^2h^2}\sum_{t=1}^{T}K_{th}^2(\delta_0)\sum_{s\neq t}^{T}(\sum_{l\geq s}\frac{K_{lh}(\delta_0)}{Z_T})^2+\frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s\neq t}K_{th}(\delta_0)K_{sh}(\delta_0)(\sum_{l\geq s}\frac{K_{lh}(\delta_0)}{Z_T})(\sum_{l\geq t}\frac{K_{lh}(\delta_0)}{Z_T})\\
=&\nu_0\int_{-1}^{1}(\int_{s}^{1}K(t)dt)^2ds+\left(\int_{-1}^{1}K(t)(\int_{t}^{1}K(u)du)dt\right)^2\equiv\mathcal{C}_{K^*}(2)
\end{aligned}$$

by noting $\mathcal{W}_{t,s}=\mathcal{W}_{s,t}$, and

$$\begin{aligned}
&\frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{s<t}K_{th}(\delta_0)\mathcal{W}_{t,s}(\delta_0)\\
=&\frac{1}{T^2h^2}\sum_{t=1}^{T}K_{th}(\delta_0)\sum_{s<t}\left\{K_{th}(\delta_0)\sum_{l\geq s}\frac{K_{lh}(\delta_0)}{Z_T}+K_{sh}(\delta_0)\sum_{l\geq t}\frac{K_{lh}(\delta_0)}{Z_T}\right\}\\
=&\frac{1}{T^2h^2}\sum_{t=1}^{T}K_{th}^2(\delta_0)\sum_{s<t}\sum_{l\geq s}\frac{K_{lh}(\delta_0)}{Z_T}+\frac{1}{T^2h^2}\sum_{t=1}^{T}K_{th}(\delta_0)\sum_{l\geq t}\frac{K_{lh}(\delta_0)}{Z_T}\sum_{s<t}K_{sh}(\delta_0)\\
\to&\int_{-1}^{1}K^2(t)(\int_{-1}^{t}\int_{s}^{1}K(u)duds)dt+\int_{-1}^{1}K(t)(\int_{t}^{1}K(u)du)(\int_{-1}^{t}K(s)ds)dt\equiv\mathcal{C}_{K^*}(1,2).
\end{aligned}$$

Thus, we complete the proof of lemma 2. □

Proof of Proposition 1.

Proof. Let $\hat{\boldsymbol{\beta}}_t=\hat{\boldsymbol{\beta}}(\frac{t}{T})$ and recall that $\boldsymbol{\beta}_t=\boldsymbol{\beta}(\frac{t}{T})$. Observe that

$$\begin{aligned}
\hat{\widetilde{u}}=&\widetilde{Y}-\widetilde{M}\hat{\boldsymbol{\beta}}(\delta_0)-D\hat{\alpha}\\
=&\widetilde{u}-\widetilde{B}(\mathbf{X},\boldsymbol{\beta})-\widetilde{M}\hat{\boldsymbol{\beta}}(\delta_0)-D(\hat{\alpha}-\alpha)\\
=&\left(\mathrm{I}_{NT}-D[D'\widetilde{W}(\delta_0)D]^{-1}D'\widetilde{W}(\delta_0)\right)\left(\widetilde{u}-\widetilde{M}(\hat{\boldsymbol{\beta}}(\delta_0)-\boldsymbol{\beta}(\delta_0))\right)-(\widetilde{B}(\mathbf{X},\boldsymbol{\beta})-\widetilde{M}\boldsymbol{\beta}(\delta_0))\\
\equiv&(\mathrm{I}_{NT}-\mathcal{S}_{NT})\left(\widetilde{u}-\widetilde{M}(\hat{\boldsymbol{\beta}}(\delta_0)-\boldsymbol{\beta}(\delta_0))(1+o(1))\right),
\end{aligned}$$

where the term $\mathcal{S}_{NT}=o(1)$. Without loss of generality, we let

$$\hat{u}_{i,t}=u_{i,t}-\mathbf{X}_{i,t}'(\hat{\boldsymbol{\beta}}_t(\delta_0)-\boldsymbol{\beta}_t(\delta_0)),$$

which implies that

$$
\begin{aligned}
\hat{\Gamma}_{i,\nu u}(j) &= \frac{1}{\tau_T^* - \tau_T} \sum_{t=\tau_T}^{\tau_T^*+1} \nu_{i,t-j} \hat{u}_{i,t} \\
=&\frac{1}{\tau_T^* - \tau_T} \sum_{t=\tau_T+1}^{\tau_T^*} \nu_{i,t-j} u_{i,t} - \frac{1}{\tau_T^* - \tau_T} \sum_{t=\tau_T+1}^{\tau_T^*} \nu_{i,t-j} \mathbf{X}'_{i,t} (\hat{\boldsymbol{\beta}}_t(\delta_0) - \boldsymbol{\beta}_t(\delta_0)) \\
\equiv& \bar{\Gamma}_{i,\nu u}(j) - \tilde{\Gamma}_{i,\nu u}(j) \qquad (62)
\end{aligned}
$$

for $j = 0, 1, \cdots, l_i$, where $\tau_T = \lfloor \tau_* T \rfloor$ and $\tau_T^* = \lfloor (1 - \tau_* T) \rfloor$ for $0 < \tau_* < 1/2$. Using (62), we have

$$
\begin{aligned}
\hat{\Delta}_{i,\nu u} =& \sum_{j=0}^{l_T} W(\frac{j}{l_i}) \hat{\Gamma}_{i,\nu u}(j) \\
=& \sum_{j=0}^{l_T} W(\frac{j}{l_T}) \bar{\Gamma}_{i,\nu u}(j) - \sum_{j=0}^{l_T} W(\frac{j}{l_T}) \tilde{\Gamma}_{i,\nu u}(j) \\
\equiv& \bar{\Delta}_{i,\nu u} + \tilde{\Delta}_{i,\nu u}. \qquad (63)
\end{aligned}
$$

We show that

$$
\sqrt{N}(\hat{\Delta}_{\nu u} - \Delta_{\nu u}) = \sqrt{N}(\bar{\Delta}_{\nu u} - \Delta_{\nu u}) + \sqrt{N} \bar{\tilde{\Delta}}_{\nu u} = o_p(1), \qquad (64)
$$

where $\bar{\Delta}_{\nu u} = \frac{1}{N} \sum_{i=1}^N \bar{\Delta}_{i,\nu u}$ and $\bar{\tilde{\Delta}}_{\nu u} = \frac{1}{N} \sum_{i=1}^N \bar{\tilde{\Delta}}_{i,\nu u}$. We first prove that $\sqrt{N} \bar{\tilde{\Delta}}_{\nu u}$ is asymptotically negligible. Note that

$$
\begin{aligned}
\sqrt{N} \bar{\tilde{\Delta}}_{\nu u} &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \tilde{\Delta}_{i,\nu u} \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{j=0}^{l_T} W(\frac{j}{l_{T_i}}) \tilde{\Gamma}_{i,\nu u}(j) \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \sum_{j=0}^{l_T} W(\frac{j}{l_T}) \left(\frac{1}{\tau_T^* - \tau_T} \sum_{t=\tau_T+1}^{\tau_T^*} \nu_{i,t-j} \mathbf{X}'_{i,t} (\hat{\boldsymbol{\beta}}_t(\delta_0) - \boldsymbol{\beta}_t(\delta_0)) \right) \right\},
\end{aligned}
$$

where

$$
\begin{aligned}
&\mathrm{E} \left\| \sum_{j=0}^{l_T} W(\frac{j}{l_T}) (\frac{1}{\tau_T^* - \tau_T} \sum_{t=\tau_T+1}^{\tau_T^*} \nu_{i,t-j} \mathbf{X}'_{i,t}) \right\| \\
=& \mathrm{E} \left\| \sum_{j=0}^{l_T} W(\frac{j}{l_T}) (\frac{1}{\tau_T^* - \tau_T} \sum_{t=\tau_T+1}^{\tau_T^*} \nu_{i,t-j} (\mathbf{X}_{i,0} + \sum_{s=1}^{t} \nu_{i,s})') \right\| \\
=& O_p(Th\sqrt{l_T/T}) = O_p(h\sqrt{l_T T}).
\end{aligned}
$$

We will prove that

$$||\widehat{\boldsymbol{\beta}}_t(\delta_0) - \boldsymbol{\beta}_t(\delta_0)|| = O_p(h^\gamma + \sqrt{\frac{\log(Th)}{NT^2h^2}}) = O_p(h^\gamma + \sqrt{\frac{\log(Th)}{NT^2h^2}}) \tag{65}$$

uniformly for $t = \tau_T + 1, \cdots, \tau_T^*$. Then we have

$$\sqrt{N}\bar{\bar{\Delta}}_{\nu u} = O_p\left(\sqrt{l_T}\sqrt{\frac{\log(Th)}{T}}\right) = o_p(1), \tag{66}$$

as $l_T = o(\frac{T}{\log(Th)})$. To prove (65), let $\delta_0 = \frac{t}{T}$, we have from (33)

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t =& \left\{\left[\widetilde{M}'\widetilde{W}^*(t/T)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(t/T)\widetilde{B}(\mathbf{X},\beta) - \boldsymbol{\beta}_t\right\} + \left[\widetilde{M}'\widetilde{W}^*(t/T)\widetilde{M}\right]^{-1}\widetilde{M}'\widetilde{W}^*(t/T)\tilde{u}\\ \equiv& \Xi_{NT}^t(1) + \Xi_{NT}^t(3), \quad t = \tau_T + 1, \cdots, \tau_T^*. \end{aligned} \tag{67}$$

From (50) and Lemma B.3 (formula (B.5)) and B.4 (formula (B.24))in Phillips et al. (2017), we have

$$\widetilde{M}'\widetilde{W}^*(t/T)\tilde{u} = \sum_{i=1}^N\sum_{s=1}^T(x_{i,s} - x_{i,\lfloor t-Th\rfloor})(u_{i,s} - \bar{u}_i)K_{sh}(\delta_0) + o_p(\sqrt{N}Th).$$

We show that

$$\sup_t\left|\left|\frac{1}{NT^2h^2}\widetilde{M}'\widetilde{W}^*(t/T)\tilde{u}\right|\right| = \max_{\tau_T\le t\le\tau_T^*}\left|\left|\frac{1}{NT^2h^2}\widetilde{M}'\widetilde{W}^*(t/T)\tilde{u}\right|\right| = \sqrt{\frac{\log(Th)}{NT^2h^2}}.$$

Let $\bar{Q}_{i,T}^t = \sum_{s=1}^T K_{sh}(t/T)(x_{i,s} - x_{i,\lfloor t-Th\rfloor})(u_{i,s} - \bar{u}_i)$

$$\begin{aligned}&\max_{\tau_T+1\le t\le\tau_T^*}\left|\left|\sum_{i=1}^N\sum_{s=1}^T K_{sh}(\delta_0)(x_{i,s} - x_{i,\lfloor t-Th\rfloor})(u_{i,s} - \bar{u}_i)\right|\right|\\ \le&\max_{\tau_T+1\le t\le\tau_T^*}\left|\left|\sum_{i=1}^N(\bar{Q}_{i,T}^t - \mathrm{E}\bar{Q}_{i,T}^t)\right|\right| + \max_{\tau_T+1\le t\le\tau_T^*}||N\mathrm{E}\bar{Q}_{i,T}^t||\\ =&o_p(\sqrt{N}Th(\log(Th))^{1/2}) + O(NTh) = O_p(\sqrt{N}Th(\log(Th))^{1/2}) \quad (\text{if } N/\log(Th)\to 0)\end{aligned}$$

by letting c_0 be some positive constant and using the Bernstein inequality, for sufficient large $C > c_0 > 0$, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{\tau_T+1\le t\le\tau_T^*}\left|\left|\sum_{i=1}^N(\bar{Q}_{i,T}^t - \mathrm{E}\bar{Q}_{i,T}^t)\right|\right| > C\sqrt{NT^2h^2\log(NT)}\right)\\ \le&\sum_{t=\tau_T+1}^{\tau_T^*}\mathbb{P}\left(\left|\left|\sum_{i=1}^N(\bar{Q}_{i,T}^t - \mathrm{E}\bar{Q}_{i,T}^t)\right|\right| > C\sqrt{NT^2h^2\log(NT)}\right)\\ \le&T\exp\left\{-\frac{CNT^2h^2\log(NT)}{c_0NT^2h^2}\right\} = O(T\cdot(NT)^{-C/c_0}) \to 0.\end{aligned} \tag{68}$$

Next, we show that

$$\sup_{\tau_T+1\leq t\leq\tau_T^*}\left|\left|\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{s=1}^{T}K_{sh}(t/T)(\eta_{is}\eta_{is}'-\bar{\eta}_i\bar{\eta}_i')-(1-\mathcal{C}_K)\Omega_\nu\right|\right|=o_p(1). \tag{69}$$

Let $\varepsilon_{i,t}^*$ be an independent copy of $\varepsilon_{i,t}$ and satisfy Assumption 3, where $\eta_{i,s}^*=\sum_{t=1}^{s}\Phi(1)'\varepsilon_{i,t}^*$ and $G_{sh}(t/T)=hK_{sh}(t/T)$. Using Lemma B.4 (formula (B.23) and (B.25)) in Phillips et al. (2017), we have

$$\begin{aligned}\frac{1}{Th}\sum_{s=1}^{2Th}K(\frac{s-Th}{Th})=&\frac{1}{Th}\sum_{s=1}^{T}K_{sh}(t/T)\\ \frac{1}{2T^2h^2}\sum_{s=1}^{T}\eta_{is}^*(\eta_{is}^*)'K(\frac{s-Th}{Th})\overset{\mathcal{L}}{=}&\frac{1}{2T^2h^2}\sum_{s=1}^{T}\eta_{is}(\eta_{is})'G_{sh}(t/T)\end{aligned}$$

do not rely on t. From (40), we have

$$\begin{aligned}&\frac{1}{NT^2h^2}\widetilde{M}'\widetilde{W}^*(\lfloor T\delta_0\rfloor/T)\widetilde{M}\\ =&\frac{1}{NT^2h^2}\sum_{i=1}^{N}\sum_{s=1}^{T}K_{sh}(\lfloor T\delta_0\rfloor/T)(\eta_{is}\eta_{is}'-\bar{\eta}_i\bar{\eta}_i')+o_P(1)\\ \overset{\mathcal{L}}{=}&\frac{1}{NT^2h^2}\left(\sum_{i=1}^{N}\sum_{s=1}^{T}K(\frac{s-Th}{Th})(\eta_{i,s}^*\eta_{i,s}^{*'}-\bar{\eta}_i\bar{\eta}_i')\right),\end{aligned}\tag{70}$$

uniformly for $0<\delta_0<1$. Thus, (69) is proved.

Now, we turn to consider $\sqrt{N}(\bar{\Delta}_{\nu u}-\Delta_{\nu u})=o_p(1)$ in (64). By Proposition 1 of Andrews (1991) and using Assumption 5, we have

$$\begin{aligned}\mathrm{E}||\sqrt{N}(\bar{\Delta}_{\nu u}-\Delta_{\nu u})||^2=&\mathrm{E}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\bar{\Delta}_{i,\nu u}-\mathrm{E}(\bar{\Delta}_{i,\nu u})+\mathrm{E}(\bar{\Delta}_{i,\nu u})-\Delta_{\nu u})\right|\right|^2\\ =&\mathrm{E}\left|\left|\bar{\Delta}_{i,\nu u}-\mathrm{E}(\bar{\Delta}_{i,\nu u})\right|\right|^2+N||\mathrm{E}(\bar{\Delta}_{i,\nu u})-\Delta_{\nu u}||^2\\ =&\left(\frac{l_T}{\tau_T^*-\tau_T}+\frac{N}{l_T^{2q}}\right)O(1)\end{aligned}\tag{71}$$

since the bandwidth parameter $l_T\to\infty$ and $\tau_T^*-\tau_T=(1-2\tau_*)T\to\infty$ for $\tau_*\in(0,1/2)$ with $l_T/T\to 0$, and $l_T^{2q}/T\to\epsilon>0$ for some $q>1/2$ by Assumption 5, it follows that $\mathrm{E}||\sqrt{N}(\bar{\Delta}_{\nu u}-\Delta_{\nu u})||^2\to 0$ with $N/T\to 0$. Using (66) and (71), we complete the proof of (64). The similar argument can be applied to $\hat{\Omega}_\omega$. In consequence, we have $\sqrt{N}(\hat{\Omega}_\omega-\Omega_\omega)=o_p(1)$.

Next we need to prove that $\sqrt{N}||\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)(\hat{\Delta}^t_{\nu\bar{u}}(\delta_0) - \Delta^t_{\nu\bar{u}}(\delta_0))|| = o_p(1)$. Note that

$$\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)||\sqrt{N}(\hat{\Delta}^t_{\nu\bar{u}}(\delta_0) - \Delta^t_{\nu\bar{u}}(\delta_0))|| \\
=&\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\sum_{l\le t}\left\{(\sum_{s=1}^{T} K_{sh}(\delta_0))^{-1}\sum_{s=1}^{T} K_{sh}(\delta_0)||\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(W(\frac{s-l}{l_T})\hat{\Gamma}_i(l-s) - \mathrm{E}[\nu_{i,l}u_{i,s}])||\right\} \\
\le&\frac{1}{Th}\sum_{t=1}^{T} K_{th}(\delta_0)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\sum_{s,l}||W(\frac{s-l}{l_T})\hat{\Gamma}_i(l-s) - \mathrm{E}[\nu_{i,l}u_{i,s}]||) = o_p(1).
\end{aligned}$$

□

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*, 627–642.

Andrews, D. W. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica: Journal of the Econometric Society*, 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, 1–34.

Bai, J. (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Baltagi, B. and Moscone, F. (2010), "Health Care Expenditure and Income in the OECD Reconsidered: Evidence from Panel Data," Center for Policy Research Working Papers 120, Center for Policy Research, Maxwell School, Syracuse University.

Baltagi, B. H. (1995), *Econometric analysis of panel data*, vol. 2, Wiley New York.

Cai, Z. (2007a), "Trending time-varying coefficient time series models with serially correlated errors," *Journal of Econometrics*, 136, 163–188.

— (2007b), "Trending time-varying coefficient time series models with serially correlated errors," *Journal of Econometrics*, 136, 163–188.

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

Casas, I., Ferreira, E., and Orbe, S. (2017), "Time-varying coefficient estimation in SURE models. Application to portfolio management," Working paper, CREATES, Aarhus University.

Chang, Y. and Martinez-Chombo, E. (2003), "Electricity Demand Analysis Using Cointegration and Error-Correction Models with Time Varying Parameters: The Mexican Case," Working papers, Rice University, Department of Economics.

Chen, J., Gao, J., and Li, D. (2012), "Semiparametric trending panel data models with cross-sectional dependence," *Journal of Econometrics*, 171, 71–85.

Chen, X. B., Gao, J., Li, D., and Silvapulle, P. (2017), "Nonparametric Estimation and Forecasting for Time-Varying Coefficient Realized Volatility Models," *journal of Business & Economic Statistics*, online, 1–13.

Croissant, Y. and Millo, G. (2008), "Panel Data Econometrics in R: The plm Package," *Journal of Statistical Software*, 27, 1–43.

Fan, J. and Li, R. (2004), "New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis," *Journal of the American Statistical Association*, 99, 710–723.

Fisher, C. R. (1990), "Differences by Age Groups in Health Care Spending," *Health Care Financing Review*, 65–90.

Gao, J. and Hawthorne, K. (2006), "Semiparametric estimation and testing of the trend of temperature series," *The Econometrics Journal*, 9, 332–355.

Gao, J. and Xia, K. (2017), "Heterogeneous panel data models with cross-sectional dependence," Monash Econometrics and Business Statistics Working Papers 16/17, Monash University, Department of Econometrics and Business Statistics.

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.

Getzen, T. (2000), "Forecasting Health Expenditures: Short, Medium and Long (Long) Term," *Journal of Health Care Finance*, 26, 56–72.

Hall, P. and Heyde, C. (1980), *Martingale Limit Theory and Its Applications*, Academic Press, New York.

Hitiris, T. and Posnett, J. (1992), "The determinants and effects of health expenditure in developed countries," *Journal of Health Economics*, 11, 173–181.

Hjellvik, V., Chen, R., and Tjøstheim, D. (2004), "Nonparametric estimation and testing in panels of intercorrelated time series," *Journal of Time Series Analysis*, 25, 831–872.

Hsiao, C. (2003), “Analysis of panel data, 2nd,” *Cambridge: Cambridge University Press. Kose, MA, ES Prasad and ME Terrones (2003), Financial Integration and Macroeconomic Volatility, IMF Staff Papers*, 50, 119–142.

Jewell, 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, 313–323.

Kleiman, E. (1974), “Thee Determinants of National Outlay on Health,” *Proceedings of a conference held by the International Economic Association at Tokyo*, 66–88.

Kneip, A., Sickles, R. C., and Song, W. (2012), “A new panel data treatment for heterogeneity in time trends,” *Econometric Theory*, 28, 590–628.

Kristensen, D. (2012), “Nonparametric detection and estimation of structural change,” *Econometrics journal*, 15, 420–461.

Lago-Peñas, S., Cantarero, D., and Blzquez, C. (2012), “On the Relationship between GDP and Health Care Expenditure: A New Perspective,” Instituto de Estudios Fiscales Working Papers 4, Instituto de Estudios Fiscale (Spain).

Leu, R. E. (1986), “The public-private mix and international health care costs,” in *Public and Private health services*, eds. Culyer, T. and Jonsson, B., Oxford: Basil Blackwell.

Liaropoulos, L. and Goranitis, I. (2015), “Health care financing and the sustainability of health systems,” *International Journal for Equity in Health*, 14, 80.

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.

Morgan, D. and Astolfi, R. (2015), “Financial impact of the GFC: health care spending across the OECD,” *Health Economics, Policy and Law*, 10, 7–19.

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

Onatski, A. (2010), "Determining the Number of Factors from Empirical Distribution of Eigenvalues," *The Review of Economics and Statistics*, 92, 1004–1016.

Orbe, S., Ferreira, E., and Rodriguez-Poo, J. (2005), "Nonparametric estimation of time varying parameters under shape restrictions," *journal of Econometrics*, 126, 53–77.

Pesaran, M. H. (2006), "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74, 967–1012.

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

Phillips, P. C. and Hansen, B. E. (1990), "Statistical inference in instrumental variables regression with I (1) processes," *The Review of Economic Studies*, 57, 99–125.

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

Phillips, P. C. and Moon, H. R. (1999), "Linear regression limit theory for nonstationary panel data," *Econometrica*, 67, 1057–1111.

Phillips, P. C. and Solo, V. (1992), "Asymptotics for linear processes," *The Annals of Statistics*, 971–1001.

Reikard, G. (2009), "Forecasting ocean wave energy: Tests of time-series models," *Ocean Engineering*, 36, 348–356.

Robinson, P. M. (1989), "Nonparametric estimation of time-varying parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, Springer, pp. 253–264.

— (2012), "Nonparametric trending regression with cross-sectional dependence," *Journal of Econometrics*, 169, 4–14.

Samadi, A. and Homaie Rad, E. (2013), "Determinants of Healthcare Expenditure in Economic Cooperation Organization (ECO) Countries: Evidence from Panel Cointegration Tests," *International Journal of Health Policy and Management*, 1, 63–68.

Zhang, W., Fan, J., and Sun, Y. (2009), "A semiparametric model for cluster data," *Annals of statistics*, 37, 2377.

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

Research Papers 2018

Center for Research in Econometric Analysis of Time Series

2018-12: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Economic Policy Uncertainty and Long-Run Stock Market Volatility and Correlation

2018-13: Emilio Zanetti Chini: Forecasting dynamically asymmetric fluctuations of the U.S. business cycle

2018-14: Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Models with Multiplicative Decomposition of Conditional Variances and Correlations

2018-15: Changli He, Jian Kang, Timo Teräsvirta and Shuhua Zhang: The Shifting Seasonal Mean Autoregressive Model and Seasonality in the Central England Monthly Temperature Series, 1772-2016

2018-16: Ulrich Hounyo and Rasmus T. Varneskov: Inference for Local Distributions at High Sampling Frequencies: A Bootstrap Approach

2018-17: Søren Johansen and Morten Ørregaard Nielsen: Nonstationary cointegration in the fractionally cointegrated VAR model

2018-18: Giorgio Mirone: Cross-sectional noise reduction and more efficient estimation of Integrated Variance

2018-19: Kim Christensen, Martin Thyrsgaard and Bezirgen Veliyev: The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing

2018-20: Ruijun Bu, Kaddour Hadri and Dennis Kristensen: Diffusion Copulas: Identification and Estimation

2018-21: Kim Christensen, Roel Oomen and Roberto Renò: The drift burst hypothesis

2018-22: Russell Davidson and Niels S. Grønborg: Time-varying parameters: New test tailored to applications in finance and macroeconomics

2018-23: Emilio Zanetti Chini: Forecasters' utility and forecast coherence

2018-24: Tom Engsted and Thomas Q. Pedersen: Disappearing money illusion

2018-25: Erik Christian Montes Schütte: In Search of a Job: Forecasting Employment Growth in the US using Google Trends

2018-26: Maxime Morariu-Patrichi and Mikko Pakkanen: State-dependent Hawkes processes and their application to limit order book modelling

2018-27: Tue Gørgens and Allan H. Würtz: Threshold regression with endogeneity for short panels

2018-28: Mark Podolskij, Bezirgen Veliyev and Nakahiro Yoshida: Edgeworth expansion for Euler approximation of continuous diffusion processes

2018-29: Isabel Casas, Jiti Gao and Shangyu Xie: Modelling Time-Varying Income Elasticities of Health Care Expenditure for the OECD