Nonparametric Testing for Smooth Structural Changes in Panel Data Models

Bin Chen[a,*] University of Rochester

Liquan Huang[b] University of Rochester

Abstract: Detecting and modelling structural changes in time series models have attracted great attention while relatively little effort has been paid to the test of structural changes in panel data models despite their increasing importance in economics and finance. In this paper, we propose a new approach to testing structural changes in panel data models with cross-sectional dependence. The idea is to compare the fitted values of a time-varying parameter panel data model and a constant parameter panel data model, where the time-varying parameters are estimated by a local linear dummy variable regression and the constant parameters are estimated by a least squared dummy variable estimation. The test does not require any prior information about the alternatives of structural changes. It has an asymptotic N(0,1) distribution under the null hypothesis of parameter constancy and is consistent against a vast class of smooth structural changes as well as abrupt structural breaks with possibly unknown break points. To further gauge possible sources of structural changes, a diagnostic test is supplemented to check potential time-varying interaction while allowing for a common trend. Simulation studies show that the tests provide reliable inference in finite samples.

JEL Classifications: *C12, C14, C23*

Key words: Local smoothing, Panel data, Parameter constancy, Smooth structural change

---

We are grateful to Zongwu Cai, A. Ronald Gallant, Cheng Hsiao, Guido Kuersteiner, Ingmar Prucha and seminar participants at the University of Maryland, the 2013 International Symposium on Analysis of Panel Data, the 23rd Annual Meeting of the Midwest Econometrics Group and the 9th Greater New York Metropolitan Area Econometrics Colloquium for their useful comments and discussions. Any remaining errors are solely ours.

a: Department of Economics, University of Rochester, Rochester, NY 14627, USA. E-mail: bchen8@mail.rochester.edu.

b: Department of Economics, University of Rochester, Rochester, NY 14627, USA. E-mail: l.huang@rochester.edu.

*: Corresponding author.

1. INTRODUCTION

The last fifty years have seen the development of a large and still growing literature on the modelling and testing of structural changes. In time series analysis, a classical econometric test for structural changes is Chow's (1960) test, which checks whether there exists a structural break at a known time point. A great deal of effort has been made in this literature, allowing multiple breaks and/or unknown time points (e.g., Andrews 1993; Andrews and Ploberger 1994; Bai 1996, 1999; Bai and Perron 1998, Elliott and Müller 2006, Perron 2006, Chen and Hong 2012, as well as references therein).

Relatively little effort has been paid to the testing of structural changes in panel data models despite the fact that panel data models have become increasingly popular among both theoretical and applied researchers (see, e.g., Baltagi 2001, Arellano 2003, Hsiao 2003, 2007). Modelling and testing structural changes are particularly relevant for panel data over a long time horizon since the underlying economic mechanism is likely to be disturbed by various factors such as preference changes, institutional changes and technological progress. The world economy is an increasingly globalized economy and thus policy changes and technological development are no longer within a country's borders. Global environmental issues have also gained enormous attention in recent years. On the other hand, model stability is crucial for statistical inference, forecasts, and sensible policy implications drawn from the model. Detecting structural changes in panel data models provides a way to better understand controversial issues such as uneven cross-country growth and global climate changes.

Some tests have been proposed to detect structural changes in panel data models in the literature. For example, De Wachter and Tzavalis (2012) propose a likelihood ratio test for a single structural break at an unknown break point in linear dynamic panel data models. Horváth and Hušková (2012) propose a CUSUM-based test for the means of panel data models. They focus on a single break although their test could be extended for multiple breaks. Feng, Kao and Lazarová (2009) study the estimation of a single change point in panel models via a Wald-type statistic and Baltagi, Kao and Liu (2012) extend it to allow for nonstationary regressors and innovations.

Almost all existing change-point tests for panel data models are constructed for abrupt changes. From a practical point of view, slowly-changing breaks may be more realistic. Various economic events, such as liberalization of emerging markets, integration of world equity markets, changes in exchange rate or interest rate regimes, may lead to structural changes in panel data models. The changes induced by policy switch, preference changes and technological progress usually exhibit evolutionary changes in the long term. Despite the importance of smooth structural changes in panel data models, to our knowledge, there is only one test designed explicitly for smooth structural changes in the literature. González, Teräsvirta and van Dijk (2005) develop a Lagrange Multiplier (LM) test against a time-varying panel smooth transition regression model. While the test might have best power against the assumed alternative, usually no prior information is available on the form of structural changes for practitioners. Therefore, it is desirable to develop consistent tests that have good power against all-round alternatives of structural changes.

Recently, a time-varying parameter panel data model has appeared as a novel tool to identify the trend

function and capture the evolutionary behavior of economic relationship. Robinson (2012) introduces a nonparametric trending regression for panel data with cross-sectional dependence and considers a simple nonparametric trend estimate. Chen, Gao and Li (2012) extend Robinson's (2012) work to the semiparametric partially linear panel data model where all individuals share a common trend. Atak, Linton and Xiao (2011) develop a semiparametric panel model to explain the trend in UK regional temperatures and other weather outcomes over the last century. The trend is allowed to evolve in a continuous manner and a nonparametric profile likelihood estimation is developed. Li, Chen and Gao (2011) generalize Cai's (2007) time-varying coefficient model to the panel data framework. One advantage of the evolutionary time-varying parameter panel data model is that little restriction is imposed on the functional forms of coefficients, except for the regularity condition that they evolve over time smoothly. Motivated by its flexibility, we will use this model as the alternative to test smooth structural changes for a panel data model with fixed effects.

We develop a Wald-type test for smooth structural changes as well as sudden structural breaks. Such a test will complement the existing tests for sudden changes in the literature and avoid the difficulty associated with whether there are multiple breaks and/or whether the time points of changes are unknown. We estimate the slowly-changing parameters by local linear dummy variable (LLDV) regression, and compare them to least squares dummy variable (LSDV) estimators. As shown in Li *et al.* (2011), the LLDV approach removes fixed effects by deducting a smoothed version of cross-time average from each individual and hence is more efficient than the averaged LL estimation, which eliminates fixed effects by taking cross-sectional averages. Moreover, it ideally suits the present problem at hand. The proposed Wald-type test can be viewed as a generalization of Hausman's (1978) test from the parametric framework to the nonparametric framework. Compared with the existing tests for structural breaks in panel data models in the literature, the proposed approach has a number of appealing features.

First, the proposed test is consistent against a large class of smooth time-varying parameter alternatives. It is also consistent against multiple sudden structural breaks in panel data models with known or unknown break points. Second, no prior information on a structural change alternative is needed. In particular, we do not need to know whether the structural changes are smooth or abrupt, and in the cases of abrupt structural breaks, we do not need to know the dates or the number of breaks. Third, unlike most tests for structural breaks in the literature, which often have nonstandard asymptotic distributions, the proposed test has a null asymptotic N(0,1) distribution. The only inputs required are LLDV and LSDV estimators. Any standard econometric software can carry out computational implementation easily. Fourth, a diagnostic test is supplemented to check possible sources of structural changes. Specifically, the diagnostic test can detect time-varying interactive effect between the dependent variable and regressors while allowing for a common trend. Fifth, the LLDV estimator can capture the local behavior of time-varying parameters. Because only local information is employed in estimating parameters at each time point, the proposed test has symmetric power against structural breaks that occur either in the first or second half of the sample period. This is different from the CUSUM-based tests that may have asymmetric power against structural breaks that have same sizes but occur at different time points. No trimming procedure is needed for the proposed test and hence the proposed test is expected to have

nontrivial powers for structural changes near the boundary regions of time, provided that the sample size is large enough. Moreover, the LLDV estimator can provide some insight into the economic relationship.

In Section 2, we introduce the time-varying panel framework and hypotheses of interest. Section 3 develops a Wald-type test, derives its asymptotic null distribution and investigates its asymptotic power property. In Section 4, a diagnostic test is proposed to check time-varying interactive effect while allowing for a common trend. Section 5 conducts a simulation study to examine the finite sample performance of the tests. Section 6 provides concluding remarks. All mathematical proofs are collected in the appendix. Throughout the paper, C denotes a generic bounded constant.

2. TIME-VARYING PANEL DATA MODEL AND HYPOTHESES OF INTEREST

Consider a nonparametric time-varying coefficient panel data model:

$$Y_{it} = X_{it}^{\intercal}\beta_t + \alpha_i + \lambda_t + \varepsilon_{it}, \qquad i = 1, ..., N,\ t = 1, ..., T, \tag{1}$$

where Y_{it} is a scalar, X_{it} is a $d \times 1$ vector of explanatory variables, β_t and λ_t are $d \times 1$ and 1×1 possibly time-varying parameter vector and scalar respectively, α_i represents an unobserved individual-specific effect and ε_{it} is weakly serially dependent and cross-sectionally dependent. We allow α_i to be correlated with X_{it} through some unknown structure and hence both fixed effects and random effects are covered. For the purpose of identification, we assume that

$$\sum_{i=1}^{N} \alpha_i = 0.$$

A keen interest in econometrics is whether the parameters of (1) β_t and λ_t are changing over time. The null hypothesis is that

$$\mathbb{H}_0 : \beta_t = \beta \text{ and } \lambda_t = \lambda \text{ for all } t.$$

The alternative hypothesis $\mathbb{H}_A$ is that $\mathbb{H}_0$ is false. Under the null, we have

$$Y_{it} = X_{it}^{\intercal}\beta + \alpha_i + \lambda + \varepsilon_{it}, \qquad i = 1, ..., N,\ t = 1, ..., T,$$

where λ can be viewed as a "mean intercept" (Hsiao, 2003). We estimate β and λ via the LSDV estimation:

$$\begin{aligned} \hat{\beta} &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i)^{\intercal}\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)(Y_{it} - \bar{Y}_i), \\ \hat{\lambda} &= \bar{Y} - \bar{X}^{\intercal}\hat{\beta}, \end{aligned}$$

where $\bar{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$, $\bar{Y}_i = T^{-1}\sum_{t=1}^{T} Y_{it}$, $\bar{X} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}$ and $\bar{Y} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}$. The consistency and asymptotic normality of $\hat{\beta}$ and $\hat{\lambda}$ are established in Section 3.

Under the alternative $\mathbb{H}_A$, β_t and λ_t are time-varying. Examples include single break and multiple break models considered in Feng *et al.* (2009) and Baltagi *et al.* (2012), González *et al.*'s (2005) time-varying panel smooth transition regression model. Tests for parametric structural change alternatives such as González *et al.*'s (2005) LM test have best power against the assumed alternative. Unfortunately, usually no prior information about the structural change alternative is available in practice. To cover a wide range of alternatives, we consider the following smooth time-varying parameter panel data model:

$$Y_{it} = X_{it}^{\intercal}\beta(t/T) + \alpha_i + \lambda(t/T) + \varepsilon_{it}, \qquad i = 1, ..., N,\ t = 1, ..., T, \tag{2}$$

where $\beta : [0,1] \to \mathbb{R}^d$ and $\lambda : [0,1] \to \mathbb{R}$ are unknown smooth functions except for a finite number of points on $[0,1]$. Discontinuities of $\beta(\cdot)$ and $\lambda(\cdot)$ at a finite number of points on $[0,1]$ allow sudden breaks. This model is studied in Li *et al.* (2011) and Chen *et al.* (2012). It covers Robinson's (2012) panel trending regression as a special case. The specification that parameters $\beta(\cdot)$ and $\lambda(\cdot)$ are some functions of ratio t/T rather than time t only is a common scaling scheme in the literature (see, e.g., Phillips and Hansen 1990, Robinson 1991 and Cai 2007). The reason for this specification is that nonparametric estimators for β_t and λ_t will not be consistent unless the amount of data on which they depend increases, and merely increasing the sample size will not necessarily improve estimation of β_t and λ_t at some fixed point t, even if some smoothness conditions are imposed.

We will assume that $\beta(\cdot)$ and $\lambda(\cdot)$ are continuous except for a finite number of points on $[0,1]$. Therefore, single structural break or multiple breaks with known or unknown break points are special cases of model (2). For example, suppose $\beta(u) = \beta_0$ and $\lambda(u) = \lambda_0$ if $u \leq u_0$, and $\beta(u) = \beta_1$ and $\lambda(u) = \lambda_1$ otherwise. Then we obtain a single break panel data model.

3. NONPARAMETRIC TESTING FOR STRUCTURAL CHANGES

We shall propose a consistent test for smooth structural changes in panel data models. Under the alternative, we have a time-varying parameter panel data model and we follow Li *et al.* (2011) to estimate β_t and λ_t via an LLDV regression. The idea of the LLDV regression is summarized below.

(i) Let $\theta(\tau) = [\lambda(\tau)\ \beta^{\intercal}(\tau)]^{\intercal}$. For each given $\tau \in (0,1)$, we minimize

$$\left\{Y - M(\tau)\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]^{\intercal} - D\alpha\right\}^{\intercal} \mathcal{K}(\tau) \left\{Y - M(\tau)\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]^{\intercal} - D\alpha\right\} \tag{3}$$

with respect to $\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]$ and α, where $\theta'(\tau) = d\theta(\tau)/d\tau$, $Y = (Y_1^{\intercal}, ..., Y_N^{\intercal})^{\intercal}$, $Y_i = (Y_{i1}, ..., Y_{iT})^{\intercal}$, $\alpha = (\alpha_2, ..., \alpha_N)^{\intercal}$, $M^{\intercal}(\tau) = [M_1^{\intercal}(\tau), ..., M_N^{\intercal}(\tau)]$ with

$$M_i(\tau) = \begin{bmatrix} 1 & X_{i1}^{\intercal} & \frac{1-\tau T}{Th} & \frac{1-\tau T}{Th} X_{i1}^{\intercal} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{iT}^{\intercal} & \frac{T-\tau T}{Th} & \frac{T-\tau T}{Th} X_{iT}^{\intercal} \end{bmatrix},$$

$\mathcal{K}(\tau) = I_N \otimes K(\tau)$, $D = (-1_{N-1}\ I_{N-1})^{\intercal} \otimes 1_T$, $K(\tau) = diag\left[k\left(\frac{1-\tau T}{Th}\right), ..., k\left(\frac{T-\tau T}{Th}\right)\right]$, I_d is a $d \times d$ identity matrix and 1_d is a $d \times 1$ vector of ones. The kernel $k(\cdot) : [-1,1] \to \mathbb{R}^+$ is a prespecified symmetric

probability density and $h \equiv h(NT)$ is a bandwidth. Examples of $k(\cdot)$ include the uniform, Epanechnikov and quartic kernels.

The first order condition of (3) with respect to α yields:

$$\hat{\alpha} = [D^{\intercal}\mathcal{K}(\tau)D]^{-1}D^{\intercal}\mathcal{K}(\tau)\left\{Y - M(\tau)\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]^{\intercal}\right\}. \tag{4}$$

(ii) Plugging (4) into (3), we get the concentrated weighted least squares:

$$\left\{Y - M(\tau)\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]^{\intercal}\right\}^{\intercal}\mathcal{W}(\tau)\left\{Y - M(\tau)\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]^{\intercal}\right\}, \tag{5}$$

where $\mathcal{W}(\tau) = W^{\intercal}(\tau)\mathcal{K}(\tau)W(\tau)$ and $W(\tau) = I_{NT} - D[D^{\intercal}\mathcal{K}(\tau)D]^{-1}D^{\intercal}\mathcal{K}(\tau)$. Minimizing (5) with respect to $\left[\theta^{\intercal}(\tau), h\left(\theta'(\tau)\right)^{\intercal}\right]$, we obtain the LLDV estimator of $\theta(\tau)$:

$$\hat{\theta}(\tau) = [I_{d+1}\ 0_{d+1}]\left[M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)\right]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)Y,$$

where 0_{d+1} is a $(d+1)\times(d+1)$ null matrix. Li *et al.* (2011) assume that (i) (X_t, ε_t) is a sequence of independent and identically distributed (*i.i.d.*) variables, where $X_t = (X_{1t}, ..., X_{Nt})^{\intercal}$ and $\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})^{\intercal}$; (ii) the error process $\{\varepsilon_{it}\}$ is independent of $\{X_{it}\}$; (iii) neither $\{\varepsilon_{it}\}$ nor $\{X_{it}\}$ is allowed to have cross-sectional dependence. We relax their assumptions and derive the asymptotic property of $\hat{\theta}(\tau)$ under the following set of assumptions.

Assumption A.1: *(i)* (X_t, ε_t) *is an* $N\times(d+1)$ *β-mixing random matrix with mixing coefficients* $\beta(j) = \sup_s E[\sup_{A\in\mathcal{G}_{s+j}^{\infty}}|P(A|\mathcal{G}_{-\infty}^{s}) - P(A)|]$, *where* $\mathcal{G}_s^t$ *is the σ-field generated by* $\{(X_k, \varepsilon_k) : k = s, \ldots, t\}$ *and* $\{\beta(j)\}$ *satisfies* $\beta(j) \le C_\beta\rho^j$ *for* $0 < C_\beta < \infty$ *and* $0 < \rho < 1$; *(ii) For any t, we have* $E(X_{it}) = \mu_X(\frac{t}{T})$, *where* $\mu_X(\tau)$ *is continuously differentiable up to the second order on* $[0, 1]$.

Assumption A.2: *(i)* $\{\varepsilon_t\}$ *is a martingale difference sequence (m.d.s.) such that* $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, *where* $\mathcal{F}_{t-1}$ *is the σ-field generated by* $\{X_t, X_{t-1}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$; *(ii)* $\sup_t E|\varepsilon_{it}|^{4(1+\eta)} < \infty$ *for some* $\eta > 0$; *(iii) As* $N \to \infty$,

$$\sup_t E\left|\sum_{i=1}^{N}\varepsilon_{it}\right|^{\delta} = O(N^{\delta/2}),\ \textit{for some}\ \delta \in (2, 4].$$

Assumption A.3: *(i) The* $d\times d$ *matrix* $\Sigma_X(\frac{t}{T}) = E\{[X_{it} - \mu_X(\frac{t}{T})][X_{it} - \mu_X(\frac{t}{T})]^{\intercal}\}$ *is positive definite. In addition,* $\Sigma_X(\tau)$ *is continuously differentiable with respect to* $\tau \in [0, 1]$*; (ii)* $\sup_t E(\|X_{it}\|^{4(1+\eta)}) < \infty$; *(iii) As* $N \to \infty$, $\sup_t E\|\sum_{i=1}^{N}[X_{it} - \mu_X(\frac{t}{T})]\|^{\delta} = O(N^{\delta/2})$; *(iv) As* $N \to \infty$, $\sup_t E\|\sum_{i=1}^{N}\{[X_{it} - \mu_X(\frac{t}{T})][X_{it} - \mu_X(\frac{t}{T})]^{\intercal} - \Sigma_X(\frac{t}{T})\}\|^{\delta} = O(N^{\delta/2})$.

Assumption A.4: *There exist a* $d\times d$ *positive definite matrix* $\Sigma_{X\varepsilon}(\frac{t}{T})$ *and* $0 < \sigma_\varepsilon^2(\frac{t}{T}) < \infty$ *such that*

as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left\{\left[X_{it}-\mu_X\left(\frac{t}{T}\right)\right]\left[X_{jt}-\mu_X\left(\frac{t}{T}\right)\right]^{\intercal}\right\}\sigma_{\varepsilon,t}(i,j) \to \Sigma_{X\varepsilon}\left(\frac{t}{T}\right),$$

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{\varepsilon,t}(i,j) \to \sigma_{\varepsilon}^2\left(\frac{t}{T}\right),$$

$$\sup_t \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} E[(\varepsilon_{it}\varepsilon_{jt})^{2(1+\eta)}|\mathcal{F}_{t-1}]^{\frac{1}{2(1+\eta)}} = O(1),$$

$$\sup_t \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} (E\|X_{it}X_{jt}^{\intercal}\|^{2(1+\eta)})^{\frac{1}{2(1+\eta)}} = O(1),$$

where $\sigma_{\varepsilon,t}(i,j) = E(\varepsilon_{it}\varepsilon_{jt}|\mathcal{F}_{t-1})$, $\sigma_{\varepsilon}^2(\tau)$ *and* $\Sigma_{X\varepsilon}(\tau)$ *are continuously differentiable on* $[0,1]$.

Assumption A.5: $\beta(\tau)$ *and* $\lambda(\tau)$ *have continuous derivatives up to the second order.*

Assumption A.6: $k:[-1,1]\to\mathbb{R}^+$ *is a symmetric and Lipschitz continuous probability density function.*

Assumption A.7: *As* $N\to\infty$ *and* $T\to\infty$, *(i) the bandwidth* h *satisfies that* $h\to 0, Th\to\infty$, $\frac{\sqrt{NT}h}{\log(NT)}\to\infty$ *and* $\frac{T^{1-\frac{2}{\delta}}h}{\log(NT)}\to\infty$; *(ii)* $\log(NT)NT^{1/2-2/\delta}\to\infty$ *and* $Th^2\to\infty$; *(iii)* $N^2/T^{\delta}\to 0$ *and* $NTh^6\to 0$.

The β-mixing condition in Assumption A.1 imposes a restriction on the temporal dependence in (X_t,ε_t). A similar condition has been used in Chen and Hong (2012), Juhl and Xiao (2013) and Kristensen (2012) in a time series context. Unlike Li *et al.* (2011), we allow for dependence between X_{it} and ε_{it} and we relax their assumptions on cross-sectional independence and stationarity. Therefore, we include the important class of dynamic panel data models in our framework. Assumption A.2 allows panel data models with potential conditional heteroscedasticity of unknown form. Assumption A.2 requires the linear panel data model to be correctly specified under $\mathbb{H}_0$ and the violation of correct model specification may lead to spurious rejection of model stability. Assumptions A.3 and A.4 impose moment conditions on $\{X_{it}\}$ and $\{\varepsilon_{it}\}$. Similar to Chen, *et al.* (2012), we allow for cross-sectional dependence in $\{\varepsilon_{it}\}$ and the degree of cross-sectional dependence is controlled by the moment conditions in Assumption A.4. Assumption A.5 is to guarantee that the asymptotic bias and variance of the LLDV are well-defined. Assumption A.6 implies $\int_{-1}^{1}k(u)du=1$, $\int_{-1}^{1}uk(u)du=0$ and $\int_{-1}^{1}u^2k(u)du<\infty$. All examples noted in Section 2 satisfy this assumption. Assumption A.7 imposes mild conditions on the bandwidth h and sample sizes N and T. Note that we allow $N/T\to C$ for $0\le C\le\infty$ as $(N,T)\to\infty$ and it covers the optimal rate $h\propto(NT)^{-\frac{1}{5}}$ of the nonparametric estimation for $\theta(\tau)$.[1]

We now state the asymptotic property of $\hat{\theta}(\tau)$.

Proposition 1. *Suppose Assumptions A.1, A.2(i)-(ii), A.3-A.7(i) hold. Then for any* $\tau\in(0,1)$, *as* $N\to\infty$ *and* $T\to\infty$,

$$\sqrt{NTh}[\hat{\theta}(\tau)-\theta(\tau)-B(\tau)]\xrightarrow{d}N(\bar{0}_{d+1},\Sigma(\tau)),$$

[1] Our derivation can still go through with a fixed N. But as a tradeoff, we need to impose a stronger assumption on T.

where $\bar{0}_{d+1}$ *is a* $(d+1)$*-dimensional null vector,*

$$B(\tau) = \frac{h^2\theta''(\tau)\int_{-1}^{1} u^2 k(u)\,du}{2} + o_P(h^2),$$

$$\Sigma(\tau) = \nu_0 \begin{bmatrix} \sigma_\varepsilon^2(\tau) + \mu_X^\intercal(\tau)\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau)\Sigma_X^{-1}(\tau)\mu_X(\tau) & -\mu_X^\intercal(\tau)\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau)\Sigma_X^{-1}(\tau) \\ -\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau)\Sigma_X^{-1}(\tau)\mu_X(\tau) & \Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau)\Sigma_X^{-1}(\tau) \end{bmatrix}, \ \theta(\tau) = [\lambda(\tau), \beta_1(\tau), ..., \beta_d(\tau)]^\intercal, \ \theta''(\tau) = d^2\theta(\tau)/d\tau^2 \text{ and } \nu_0 = \int_{-1}^{1} k^2(u)du.$$

Proposition 1 extends Theorem 2.2 of Li *et al.* (2011) to allow for cross-sectional dependence and nonstationarity. We only impose a mild condition on the relative rates of growth between T and N, thus the proposed estimator and test below are applicable to panel data with various size combinations of T and N. Both $\hat{\beta}(\tau)$ and $\hat{\lambda}(\tau)$ achieve the same convergence rate.

Under the null hypothesis, $\beta(\cdot)$ and $\lambda(\cdot)$ are constant and we estimate them via the LSDV estimation, which is discussed in Section 2. Let $\theta = [\lambda \ \beta^\intercal]^\intercal$ and $\hat{\theta} = [\hat{\lambda} \ \hat{\beta}^\intercal]^\intercal$. The asymptotic property of $\hat{\theta}$ is established below.

Proposition 2. *Suppose Assumptions A.1, A.2(i)-(ii), A.3(i)-(ii), A.4 hold. Then under* $\mathbb{H}_0$*, as* $T \to \infty$ *and* $N \to \infty$*,*

$$\sqrt{NT}(\hat{\theta} - \theta) \xrightarrow{d} N(\bar{0}_{d+1}, \Sigma_\theta),$$

where $\Sigma_\theta = \Xi V \Xi^\intercal$*,* $A = \int_0^1 \Sigma_X(\tau)d\tau + \int_0^1 \mu_X(\tau)\mu_X^\intercal(\tau)\,d\tau - \int_0^1 \mu_X(\tau)\,d\tau \int_0^1 \mu_X^\intercal(\tau)\,d\tau$*,*

$$\Xi = \begin{bmatrix} 1 + \int_0^1 \mu_X^\intercal(\tau)d\tau A^{-1}\int_0^1 \mu_X(\tau)d\tau & -\int_0^1 \mu_X^\intercal(\tau)d\tau A^{-1} & -\int_0^1 \mu_X^\intercal(\tau)d\tau A^{-1} \\ -A^{-1}\int_0^1 \mu_X(\tau)d\tau & A^{-1} & A^{-1} \end{bmatrix}$$

$$and V = \begin{bmatrix} \int_0^1 \sigma_\varepsilon^2(\tau)d\tau & \int_0^1 \sigma_\varepsilon^2(\tau)\mu_X^\intercal(\tau)d\tau & \bar{0}_d^\intercal \\ \int_0^1 \sigma_\varepsilon^2(\tau)\mu_X(\tau)d\tau & \int_0^1 \sigma_\varepsilon^2(\tau)\mu_X(\tau)\mu_X^\intercal(\tau)d\tau & 0_d \\ \bar{0}_d & 0_d & \int_0^1 \Sigma_{X\varepsilon}(\tau)d\tau \end{bmatrix}.$$

The LSDV estimator has been commonly used in practice, but its asymptotic property under non-stationarity, both cross-sectional and serial dependence has not yet been developed to our knowledge. Proposition 2 thus fills the gap in the literature. Under the null hypothesis $\mathbb{H}_0$, the LSDV estimator $\hat{\theta}$ and the LLDV estimator $\hat{\theta}_t$ converge to the same probability limit; under the alternative hypothesis $\mathbb{H}_A$, they depart from each other. Therefore, to check parameter constancy, we consider a Wald-type test that compares these two estimators via a weighted quadratic form:

$$\hat{Q} = N\sqrt{h}\sum_{t=1}^{T}\left(\hat{\theta}_t - \hat{\theta}\right)^\intercal \hat{\Omega}_t \left(\hat{\theta}_t - \hat{\theta}\right),$$

where the weighting matrix $\hat{\Omega}_t = N^{-1}\sum_{i=1}^{N}[1 \ X_{it}^\intercal]^\intercal[1 \ X_{it}^\intercal]$. The statistic $(NT\sqrt{h})^{-1}\hat{Q}$ converges to 0 under $\mathbb{H}_0$, but to a strictly positive constant under $\mathbb{H}_A$, giving our one-sided test asymptotic unit power. Any significant departure from 0 is evidence of structural changes. Formally, our generalized Hausman

test is a standardized version of $\hat{Q}$:

$$\hat{H} = (\hat{Q} - \hat{C})/\sqrt{\hat{S}},$$

where

$$\begin{aligned}
\hat{C} &= h^{-1/2}\nu_0 \int_0^1 \left[\hat{\sigma}_\varepsilon^2(\tau) + trace(\hat{\Sigma}_X^{-1}(\tau)\hat{\Sigma}_{X\varepsilon}(\tau))\right] d\tau, \\
\hat{S} &= 4\int_0^1 \left[\hat{\sigma}_\varepsilon^4(\tau) + trace(\hat{\Sigma}_X^{-1}(\tau)\hat{\Sigma}_{X\varepsilon}(\tau)\hat{\Sigma}_X^{-1}(\tau)\hat{\Sigma}_{X\varepsilon}(\tau))\right] d\tau \int_0^1 [\int_{-1}^1 k(u)k(u+v)du]^2 dv
\end{aligned}$$

are centering and scaling factors, $\hat{\sigma}_\varepsilon^2(\tau) = N^{-1}\sum_{i=1}^N \sum_{j=1}^N \hat{\sigma}_{\varepsilon,\tau}(i,j)$, $\hat{\sigma}_{\varepsilon,\tau}(i,j) = (Th)^{-1}\sum_{t=1}^T \hat{\varepsilon}_{it}\hat{\varepsilon}_{jt}k(\frac{t-\tau T}{Th})$, $\hat{\Sigma}_X(\tau) = N^{-1}\sum_{i=1}^N \hat{M}_\tau^{ii}$, $\hat{\Sigma}_{X\varepsilon}(\tau) = N^{-1}\sum_{i=1}^N \sum_{j=1}^N \hat{M}_\tau^{ij}\hat{\sigma}_{\varepsilon,\tau}(i,j)$, $\hat{M}_\tau^{ij} = (Th)^{-1}\Sigma_{t=1}^T[X_{it} - \hat{\mu}_X^i(\frac{t}{T})][X_{jt} - \hat{\mu}_X^j(\frac{t}{T})]^\intercal k(\frac{t-\tau T}{Th})$, $\hat{\mu}_X^i(\tau) = (Th)^{-1}\sum_{t=1}^T k(\frac{t-\tau T}{Th})X_{it}$ and ν_0 is defined in Proposition 1.

We now state the asymptotic distribution of $\hat{H}$ under $\mathbb{H}_0$.

Assumption A.8: *(i) As* $N \to \infty$, $\sup_t \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \{E[(\varepsilon_{it}\varepsilon_{jt})^{4(1+\eta)}|\mathcal{F}_{t-1}]\}^{\frac{1}{4(1+\eta)}} = O(1)$; *(ii) As* $N \to \infty$, $\sup_t \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \{E[(\varepsilon_{it}\varepsilon_{jt})^{8(1+\eta)}\|X_{it}X_{jt}\|^{8(1+\eta)}]\}^{\frac{1}{8(1+\eta)}} = O(1)$; *(iii)* $\sup_t E\|X_{it}\|^{8(1+\eta)} < \infty$ and $\sup_t E|\varepsilon_{it}|^{8(1+\eta)} < \infty$.

Assumption A.9: *As* $N \to \infty$, *(i)* $E\|\sum_{i=1}^N \sum_{j=1}^N [\varepsilon_{it}\varepsilon_{jt} - E(\varepsilon_{it}\varepsilon_{jt})]\|^\delta = O(N^{\delta/2})$; *(ii)* $E\|\sum_{i=1}^N \sum_{j=1}^N \{\varepsilon_{it}\varepsilon_{jt}[X_{it} - \mu_X(\frac{t}{T})][X_{jt} - \mu_X(\frac{t}{T})]^\intercal - E\{\varepsilon_{it}\varepsilon_{jt}[X_{it} - \mu_X(\frac{t}{T})][X_{jt} - \mu_X(\frac{t}{T})]^\intercal\}\}\|^\delta = O(N^{\delta/2})$.

Theorem 1: *Suppose Assumptions A.1, A.2(i)(ii)-A.4, A.6, A.7(i)(ii)-A.9 and* $\mathbb{H}_0$ *hold. Then*

$$\hat{H} \xrightarrow{d} N(0,1)$$

as $T \to \infty$ *and* $N \to \infty$.

The $\hat{H}$ test has a convenient null asymptotic N(0,1) distribution. This is quite appealing in light of the facts that most existing tests for structural changes in panel data models have nonstandard distributions which may depend on the DGPs. The proposed test does not require formulation of an alternative and is applicable when one has no prior information of the alternative. Moreover, no trimming is needed. As an important feature of the $\hat{H}$ test, the use of the LSDV estimator $\hat{\theta}$ in place of the true parameter θ under $\mathbb{H}_0$ has no impact on the limit distribution of $\hat{H}$. Intuitively, the parametric estimator $\hat{\theta}$ converges to θ at a $\sqrt{NT}$-rate, which is faster than the nonparametric estimator $\hat{\theta}_t$. Consequently, the asymptotic distribution of $\hat{H}$ is solely determined by the nonparametric estimator $\hat{\theta}_t$ and is nuisance parameter free.

Next, we investigate the asymptotic power property of $\hat{H}$ under $\mathbb{H}_A$.

Assumption A.10: *Except for a finite number of points on* $[0,1]$, *(i)* $\beta(\tau)$ *has continuous derivatives up to the second order; (ii)* $\lambda(\tau)$ *has continuous derivatives up to the second order.*

Assumption A.11: *(i)* $\sup_{v\in(0,1)} \|\lim_{\tau\to v^+} \beta(\tau) - \lim_{\tau\to v^-} \beta(\tau)\| \le C$; *(ii)* $\sup_{v\in(0,1)} \|\lim_{\tau\to v^+} \lambda(\tau) - \lim_{\tau\to v^-} \lambda(\tau)\| \le C$.

Theorem 2: *Suppose Assumptions A.1, A.2(i)(ii)-A.4, A.6-A.7(i)(ii), A.8-A.11 hold. Then for any sequence of nonstochastic constants* $\{C_T = o(NT\sqrt{h})\}$, $P(\hat{H} > C_T) \to 1$ *under* $\mathbb{H}_A$ *as* $T \to \infty$ *and* $N \to \infty$.

Assumption A.10 allows for both smooth structural changes and abrupt structural breaks with known or unknown break points. We permit $\theta(\cdot)$ to have a fixed number of discontinuities. Hence, single structural break and multiple breaks with known or unknown break points, which are often considered in this literature, are included as special cases. For abrupt structural breaks, the break size is bounded by Assumption A.11. Theorem 2 suggests that the $\hat{H}$ test is consistent against all alternatives to $\mathbb{H}_0$, subject to a set of regularity conditions. Thus, the proposed test will be able to detect any structural changes in panel data models as long as T and N are sufficiently large. This is appealing in light of the fact that no prior information about the alternative of structural changes is available in practice. It avoids the blindness of searching for possible alternatives of structural changes in practice.

4. DIAGNOSTIC TESTING FOR TIME-VARYING INTERACTION

When structural changes are detected by the $\hat{H}$ test, it would be interesting to explore possible sources of the rejection. That is, whether the rejection is from a time-varying intercept or slope, which corresponds to a time trend or a time-varying interactive effect in economic applications. Such information, if any, will be valuable in reconstructing the model and studying the relationship between economic variables. For example, in a simple wage equation, it might be interesting to check whether structural changes exist in the return to schooling while allowing for time variation in the intercept. On the other hand, researchers may have some prior information (from e.g., economic theories or empirical evidences) on the existence of a time trend. Therefore, the structural break tests could mainly focus on the slope coefficients. For example, in the model of regional economic growth, it is important and challenging to study the stability of the impact of various economic factors (e.g., capital stock, labor input, technology, etc.) on growth rate, while a time trend is usually assumed.

In the past few years, panel data models with a common trend, which specifies the time-specific effect with some unknown functions rather than dummy variables, have become popular. For example, Atak, Linton and Xiao (2012) develop a semiparametric panel model to explain the trend in UK temperatures over the last century using data observed at the twenty six Meteorological Office stations. Chen, *et al.* (2012) study a semiparametric fixed effect model to capture the nonlinear trending phenomenon in panel data analysis and develop a pooled semiparametric profile dummy variable estimation (PSPDV). Zhang, Su and Phillips (2012) propose a nonparametric test for common trend in semiparametric panel data models with fixed effects. Hence, we build on this rich literature and focus on testing whether the interactive effect (β_t) is time-varying while allowing for a time trend in this section. Namely, the DGP is

$$Y_{it} = X_{it}^{\intercal}\beta_t + \alpha_i + \lambda_t + \varepsilon_{it}, \qquad i = 1, ..., N,\ t = 1, ..., T,$$

the null hypothesis is

$$\mathbb{H}_0^* : \beta_t = \beta \text{ for all } t$$

and the alternative hypothesis $\mathbb{H}_A^*$ is the negation of $\mathbb{H}_0^*$. We shall follow Chen *et al.* (2012) to estimate β via a PSPDV estimation under the null hypothesis $\mathbb{H}_0^*$. The idea of the PSPDV estimation is briefly summarized below.

(i) For given α and β, we estimate $\lambda(\tau)$ by

$$\begin{pmatrix} \hat{\lambda}_\beta(\tau) \\ \hat{\lambda}'_\beta(\tau) \end{pmatrix} = \arg\min_\gamma [Y - X\beta - D\alpha - Z(\tau)\gamma]^\intercal \mathcal{K}(\tau) [Y - X\beta - D\alpha - Z(\tau)\gamma], \tag{6}$$

where $X = (X_1^\intercal, ..., X_N^\intercal)^\intercal$, $X_i = (X_{i1}, ..., X_{iT})^\intercal$, $Z(\tau) = 1_N \otimes z(\tau)$, $z(\tau) = \begin{pmatrix} 1 & \cdots & 1 \\ \frac{1-\tau T}{Th} & \cdots & \frac{T-\tau T}{Th} \end{pmatrix}^\intercal$. The first order condition of (6) yields

$$\hat{\lambda}_\beta(\tau) = (1,0) S(\tau) (Y - X\beta - D\alpha),$$

where $S(\tau) = [Z^\intercal(\tau) \mathcal{K}(\tau) Z(\tau)]^{-1} Z^\intercal(\tau) \mathcal{K}(\tau)$.

(ii) Let $\hat{\Lambda}_\beta = 1_N \otimes [\hat{\lambda}_\beta(1/T), ..., \hat{\lambda}_\beta(T/T)]^\intercal$. We estimate $(\alpha^\intercal, \beta^\intercal)^\intercal$ by

$$\min_{(\alpha^\intercal, \beta^\intercal)^\intercal} \left(Y - X\beta - D\alpha - \hat{\Lambda}_\beta\right)^\intercal \left(Y - X\beta - D\alpha - \hat{\Lambda}_\beta\right). \tag{7}$$

By solving the optimization problem in (7), we obtain the solution:

$$\hat{\beta}_P = \left(\tilde{X}^\intercal \tilde{D} \tilde{X}\right)^{-1} \tilde{X}^\intercal \tilde{D} \tilde{Y}, \tag{8}$$

where $\tilde{X} = (I_{NT} - \tilde{S})X$, $\tilde{Y} = (I_{NT} - \tilde{S})Y$, $\tilde{D} = I_{NT} - D^*(D^{*\intercal} D^*)^{-1} D^{*\intercal}$, $D^* = (I_{NT} - \tilde{S})D$ and $\tilde{S} = 1_N \otimes \{[(1,0) S(1/T)]^\intercal, ..., [(1,0) S(T/T)]^\intercal\}^\intercal$.

We now state the asymptotic property of $\hat{\beta}_P$ in (8).

Proposition 3. *Suppose Assumptions A.1, A.2, A.3(i)(ii)(iii), A.4, A.6, A.7(i)(iii), A.10(ii) and A.11(ii) hold. Then under $\mathbb{H}_0^*$, as $T \to \infty$ and $N \to \infty$,*

$$\sqrt{NT}(\hat{\beta}_P - \beta) \xrightarrow{d} N(\bar{0}_d, \Sigma_\beta),$$

where $\Sigma_\beta = \left[\int_0^1 \Sigma_X(\tau) d\tau\right]^{-1} \int_0^1 \Sigma_{X\varepsilon}(\tau) d\tau \left[\int_0^1 \Sigma_X(\tau) d\tau\right]^{-1}$.

Proposition 3 is an extension of Theorem 3.1 in Chen *et al.* (2012). We relax their stationarity assumption so that lagged dependent variables can be included as regressors. The smoothness condition of the trend function $\lambda(\tau)$ is also relaxed to allow for a finite number of jumps. When the null hypothesis $\mathbb{H}_0^*$ holds, the PSPDV estimator $\hat{\beta}_P$ and the LLDV estimator $\hat{\beta}_t$ will be close to each other while under $\mathbb{H}_A^*$, $\hat{\beta}_P$ and $\hat{\beta}_t$ will converge to different probability limits. Therefore, the test statistic for the time-varying interaction can be constructed as

$$\hat{H}_1 = \left(\hat{Q}_1 - \hat{C}_1\right) / \sqrt{\hat{S}_1},$$

where

$$\begin{aligned}\hat{Q}_1 &= N\sqrt{h}\sum_{t=1}^{T}\left(\hat{\beta}_t-\hat{\beta}_P\right)^{\intercal}\hat{M}_t\left(\hat{\beta}_t-\hat{\beta}_P\right),\\ \hat{M}_t &= N^{-1}\sum_{i=1}^{N}X_{it}X_{it}^{\intercal},\end{aligned}$$

and

$$\begin{aligned}\hat{C}_1 &= h^{-1/2}\nu_0\int_0^1 trace(\hat{\Psi}_X(\tau)\hat{\Sigma}_{X\varepsilon}(\tau))d\tau,\\ \hat{S}_1 &= 4\int_0^1 trace(\hat{\Psi}_X(\tau)\hat{\Sigma}_{\varepsilon X}(\tau)\hat{\Psi}_X(\tau)\hat{\Sigma}_{\varepsilon X}(\tau))d\tau\int_0^1[\int_{-1}^1 k(u)k(u+v)du]^2dv\end{aligned}$$

are centering and scaling factors respectively, where $\hat{\Psi}_X(\tau) = \hat{\Sigma}_X^{-1}(\tau)+\hat{\Sigma}_X^{-1}(\tau)\hat{\mu}_X(\tau)\hat{\mu}_X^{\intercal}(\tau)\hat{\Sigma}_X^{-1}(\tau)$, $\hat{\mu}_X(\tau)=(NTh)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}k(\frac{t-\tau T}{Th})X_{it}$ and ν_0 is defined in Proposition 1.

The asymptotic null distribution and the asymptotic power property of $\hat{H}_1$ are stated below.

Theorem 3: *Suppose Assumptions A.1-A.4, A.6-A.8, A.9(ii), A.10(ii), A.11(ii) and* $\mathbb{H}_0^*$ *hold. Then*

$$\hat{H}_1 \xrightarrow{d} N(0,1)$$

as $T\to\infty$ *and* $N\to\infty$.

Theorem 4: *Suppose Assumptions A.1-A.4, A.6-A.8, A.9(ii), A.10 and A.11 hold. Then for any sequence of nonstochastic constants* $\{C_T = o(NT\sqrt{h})\}$, $P(\hat{H}_1 > C_T)\to 1$ *under* $\mathbb{H}_A^*$ *as* $T\to\infty$ *and* $N\to\infty$.

Similar to the $\hat{H}$ test, the use of the PSPDV estimator $\hat{\beta}_P$ in place of the true parameter β under $\mathbb{H}_0^*$ has no impact on the limit distribution of $\hat{H}_1$. That is due to the fact that $\hat{\beta}_P$ could achieve a $\sqrt{NT}$ rate, which is faster than the convergence rate of the nonparametric estimator $\hat{\beta}_t$. Theorem 4 allows $\beta(\cdot)$ to have a fixed number of discontinuities and thus includes single structural break and multiple breaks with known or unknown break points as special cases.

5. FINITE SAMPLE PERFORMANCE

Next, we study the finite sample performance of our tests. Theorems 1 and 3 provide the null asymptotic $N(0,1)$ distribution of $\hat{H}$ and $\hat{H}_1$. Thus, one can implement our tests for $\mathbb{H}_0$ and $\mathbb{H}_0^*$ by comparing $\hat{H}$ and $\hat{H}_1$ with a $N(0,1)$ critical value. However, like many other nonparametric tests in the literature, the size of $\hat{H}$ and $\hat{H}_1$ in finite samples may differ significantly from the prespecified asymptotic significance level. Our analysis suggests that the asymptotic theory may not work well even for relatively large sample sizes, because the asymptotically negligible higher order terms in $\hat{H}$ and $\hat{H}_1$ are close in order of magnitude to the dominant U-statistic that determines the limit distribution of $\hat{H}$ and $\hat{H}_1$. To overcome this problem, we consider a residual-based bootstrap:

Step (i): Use the sample $\{(Y_{it}, X_{it}^{\intercal}), i = 1, ..., N, t = 1, ..., T\}$ to estimate the model via LSDV (PSPDV for the $\hat{H}_1$ test) and LLDV regression respectively and compute the $\hat{H}$ and $\hat{H}_1$ statistics and the nonparametric residual $\hat{\varepsilon}_{it} = Y_{it} - \hat{\lambda}_t - \hat{\alpha}_i - X_{it}^{\intercal}\hat{\beta}_t$; *Step* (ii): Obtain a bootstrap residual $\hat{\varepsilon}_{it}^*$ by random sampling with replacement from the centered nonparametric residual $\bar{\varepsilon}_{it} = \hat{\varepsilon}_{it} - N^{-1}T^{-1}\Sigma_{i=1}^{N}\Sigma_{t=1}^{T}\hat{\varepsilon}_{it}$ and construct a bootstrap sample $\{(Y_{it}^*, X_{it}^{\intercal}), i = 1, ..., N, t = 1, ..., T\}$, where $Y_{it}^* = X_{it}^{\intercal}\hat{\beta} + \hat{\alpha}_i + \hat{\lambda} + \hat{\varepsilon}_{it}^*$ ($Y_{it}^* = X_{it}^{\intercal}\hat{\beta} + \hat{\alpha}_i + \hat{\lambda}_t + \hat{\varepsilon}_{it}^*$ for the $\hat{H}_1$ test); *Step* (iii): Compute the bootstrap statistics $\hat{H}^*$ and $\hat{H}_1^*$, in the same way as $\hat{H}$ and $\hat{H}_1$ respectively, with $\{(Y_{it}^*, X_{it}^{\intercal}), i = 1, ..., N, t = 1, ..., T\}$ replacing the original sample $\{(Y_{it}, X_{it}^{\intercal}), i = 1, ..., N, t = 1, ..., T\}$; *Step* (iv): Repeat steps (ii) and (iii) B times to obtain B bootstrap test statistics $\{\hat{H}_l^*\}_{l=1}^{B}$ and $\{\hat{H}_{1l}^*\}_{l=1}^{B}$, where B is sufficiently large; *Step* (v): Compute the bootstrap p-values $p^* \equiv B^{-1}\Sigma_{l=1}^{B}\mathbf{1}(\hat{H}_l^* > \hat{H})$ and $p_1^* \equiv B^{-1}\Sigma_{l=1}^{B}\mathbf{1}(\hat{H}_{1l}^* > \hat{H}_1)$, where $\mathbf{1}(\cdot)$ is the indicator function.

To examine the size of our test $\hat{H}$ under $\mathbb{H}_0$, we consider the following DGP:
DGP S.1 [*No Structural Change*]:

$$Y_{it} = \bar{\lambda} + \bar{\beta}X_{it} + \alpha_i + \varepsilon_{it},$$

where $\bar{\lambda}$ and $\bar{\beta}$ are the average of $\lambda(t/T)$ and $\beta(t/T)$ respectively with $t = 1, ..., T$:

$$\begin{aligned} \lambda(\tau) &= \tau^2 + \tau + 1 \\ \beta(\tau) &= sin(\pi\tau), \end{aligned} \tag{9}$$

$$X_{it} = \frac{1}{2}X_{i,t-1} + v_{it},$$

and

$$\begin{aligned} \alpha_i &= \theta_0\bar{X}_i + u_i, \qquad i = 1, \cdots, N-1,\ \theta_0 = 1,\ u_i \sim i.i.d.N(0,1), \\ \alpha_N &= -\sum_{i=1}^{N-1}\alpha_i. \end{aligned}$$

To check the robustness of our test $\hat{H}$, we generate v_{it} and ε_{it} in two ways: one is from an $i.i.d.N(0,1)$ distribution and the other has cross-sectional dependence following Chen *et al.* (2012). Let $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})^{\intercal}$, which is generated as a sequence of an N-dimensional vector of independent Gaussian random variables with zero mean and covariance matrix (c_{ij}), where

$$c_{ij} = 0.8^{|j-i|}.$$

We generate v_{it} similarly, but with covariance matrix (d_{ij}) where $d_{ij} = 0.5^{|j-i|}$. We generate 500 data sets of a random sample $\{(Y_{it}, X_{it}^{\intercal}), i = 1, ..., N, t = 1, ..., T\}$ for $N, T = 15, 20, 25, 30$. To investigate the power of our test $\hat{H}$ in detecting structural changes in panel data models, we consider three alternatives:

DGP P.1 [*Single Structural Break*]:

$$Y_{it} = \begin{cases} \bar{\lambda} + \bar{\beta} X_{it} + \alpha_i + \varepsilon_{it}, & \text{if } t \leq 0.5T, \\ (\bar{\lambda} + 0.5) + (\bar{\beta} + 0.5) X_{it} + \alpha_i + \varepsilon_{it}, & \text{otherwise.} \end{cases}$$

DGP P.2 [*Multiple Structural Breaks*]:

$$Y_{it} = \begin{cases} (\bar{\lambda} + 0.5) + (\bar{\beta} + 0.5) X_{it} + \alpha_i + \varepsilon_{it}, & \text{if } 0.3T \leq t \leq 0.7T, \\ \bar{\lambda} + \bar{\beta} X_{it} + \alpha_i + \varepsilon_{it}, & \text{otherwise.} \end{cases}$$

DGP P.3 [*Smooth Structural Changes*]:

$$Y_{it} = \bar{\lambda} + \phi \left[\lambda \left(t/T \right) - \bar{\lambda} \right] + \left\{ \bar{\beta} + \phi \left[\beta \left(t/T \right) - \bar{\beta} \right] \right\} X_{it} + \alpha_i + \varepsilon_{it},$$

where $\phi = 0.4$.

The single break has been a structural change with classical importance. Under DGP P.1, an abrupt break occurs in the panel data model at some unknown time t. DGP P.2 admits nonmonotonic multiple breaks. DGP P.3 is the time-varying coefficient panel model considered in Li *et al.* (2011).

For the proposed $\hat{H}$ test, we use the uniform kernel. In fact, our simulation experience (not reported here) suggests that the choice of $k(\cdot)$ has little impact on the performance of the test. For simplicity, we choose the bandwidth $h = C(NT)^{-\frac{1}{5}}$ with $C = 0.5, 1$ and 1.5. We use the bootstrap procedure described above with the number of bootstrap iterations $B = 99$.

Table 1 reports the rejection rates of $\hat{H}$ under DGP S.1 at the 10% and 5% significance levels, using bootstrap critical values (BCVs). When ε_{it} is $i.i.d$, the $\hat{H}$ test has good size with the rejection rates close to the nominal levels. When ε_{it} is cross-sectionally dependent, the size performance of $\hat{H}$ enjoys similar pattern, which suggests that our $\hat{H}$ test is robust to potential cross-sectional dependence. We also note that the size of our $\hat{H}$ test is robust to different choices of bandwidth.

Table 2 reports the rejection rates of $\hat{H}$ with BCVs under DGPs P.1-P.3 at the 10% and 5% levels. The $\hat{H}$ test has reasonable all-around power against smooth and abrupt structural changes. The power increases as either N or T increases. The rejection rate is about 52% at the 5% level even when (N, T) is as small as $(15, 15)$, and approaches unity when $(N, T) = (30, 30)$. The choice of bandwidth has some effect on the power of our test when the sample size is small. However, with the increase of N and T, the power becomes rather robust to choices of bandwidth.

Next, we turn to the finite sample performance of our diagnostic test $\hat{H}_1$. We use a similar DGP as DGP S.1 for size while replacing $\bar{\lambda}$ with the time varying function (9):

DGP S1.1 [*No Structural Break in the Interactive Effect*]:

$$Y_{it} = \lambda_t + \bar{\beta} X_{it} + \alpha_i + \varepsilon_{it},$$

where $\lambda_t = \lambda(t/T) = (t/T)^2 + (t/T) + 1$ and $\bar{\beta}$, X_{it}, α_i and ε_{it} are generated as before.

To check the power of $\hat{H}_1$, we consider three types of structural changes in the interactive effect while

allowing for the common time trend λ_t.

DGP P1.1 [*Single Structural Break in the Interactive Effect*]:

$$Y_{it} = \begin{cases} \lambda_t + \bar{\beta} X_{it} + \alpha_i + \varepsilon_{it}, & \text{if } t \leq 0.5T, \\ \lambda_t + (\bar{\beta} + 0.5) X_{it} + \alpha_i + \varepsilon_{it}, & \text{otherwise.} \end{cases}$$

DGP P1.2 [*Multiple Structural Breaks in the Interactive Effect*]:

$$Y_{it} = \begin{cases} \lambda_t + (\bar{\beta} + 0.5) X_{it} + \alpha_i + \varepsilon_{it}, & \text{if } 0.3T \leq t \leq 0.7T, \\ \lambda_t + \bar{\beta} X_{it} + \alpha_i + \varepsilon_{it}, & \text{otherwise.} \end{cases}$$

DGP P1.3 [*Smooth Structural Changes in the Interactive Effect*]:

$$Y_{it} = \lambda_t + \left\{ \bar{\beta} + \phi \left[\beta \left(t/T \right) - \bar{\beta} \right] \right\} X_{it} + \alpha_i + \varepsilon_{it},$$

where $\phi = 0.4$.

Tables 3 and 4 report the empirical size and power of $\hat{H}_1$ under DGPs S1.1 and P1.1-1.3. Similar to the $\hat{H}$ test, $\hat{H}_1$ has reasonable size performance and is robust to potential cross-sectional dependence and the bandwidth selection. As expected, the $\hat{H}_1$ test has a bit lower rejection rate than the $\hat{H}$ test. But the rejection rate increases with both N and T and approaches to unity when $(N, T) = (30, 30)$.

To confirm that our tests have the right power when the distance between the null and alternative hypotheses is increased, we plot in Figures 1 and 2 the empirical power of $\hat{H}$ and $\hat{H}_1$ as functions of ϕ for DGPs P.3 and P1.3. When $\phi = 0$, we are back to our null models DGPs S.1 and S1.1. Figures 1 and 2 show that the power functions increase monotonically with ϕ. When the magnitude of ϕ is increased to a larger extent, the power of our tests is reaching unity.

To sum up, we observe that both $\hat{H}$ and $\hat{H}_1$ tests have good sizes in finite samples when the residual-based bootstrap is applied. They also have reasonable powers against both sudden structural breaks and smooth structural changes.

6. CONCLUSION

The modelling of structural changes in panel data models has attracted increasing attention in econometrics. We have complemented the literature by proposing a Wald-type test for smooth structural changes as well as abrupt structural breaks in panel data models, which has not been attempted in the previous literature. Our generalized Hausman's (1978) test is intuitively appealing and straightforward to compute. It has a convenient null asymptotic N(0,1) distribution, does not require trimming data, does not require prior information on the possible alternative, and is consistent against all smooth structural changes as well as multiple abrupt structural breaks in panel data models. Moreover, only a mild condition is imposed on the relative rates of growth between T and N, thus our approach can be applied to panel data with various size combinations of T and N. Our omnibus test is supplemented by a diagnostic test, which allows for a common trend and focuses on potential structural changes in the interactive effect. Such information is useful for practitioners in reconstructing a misspecified model

and studying the relationship between economic variables. To overcome the adverse impact of the first stage nonparametric estimation of the time-varying parameters, we use residual-based bootstrap, which provides reasonable size and power for the proposed tests in finite samples.

REFERENCES

Andrews, D.W.K. (1993) Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* **61**, 821-856.

Andrews, D.W.K. & Ploberger, W. (1994) Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative. *Econometrica* **62**, 1383-1414.

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

Atak, A., Linton, O. & Xiao, Z. (2011) A Semiparametric Panel Model for Unbalanced Data with Application to Climate Change in the United Kingdom. *Journal of Econometrics* **164**, 92-115.

Bai, J. (1996) Testing for Parameter Constancy in Linear Regressions: an Empirical Distribution Function Approach. *Econometrica* **64**, 597-622.

Bai, J. (1999) Likelihood Ratio Tests for Multiple Structural Changes. *Journal of Econometrics* **91**, 299-323.

Bai, J. & Perron, P. (1998) Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* **66**, 47-78.

Baltagi, B. (2001) Econometric Analysis of Panel Data. Wiley.

Baltagi, B., Kao, C. & Liu, L. (2012) Estimation and Identification of Change Points in Panel Models with Nonstationary or Stationary Regressors and Error Term. Working paper.

Bosq, D., (1998) Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, second edition. In Lecture Notes in Statistics, vol. 110. Springer-Verlag.

Cai, Z. (2007) Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors. *Journal of Econometrics* **136**, 163-188.

Chen, J., Gao, J. & Li, D. (2012) Semiparametric Trending Panel Data Models with Cross-sectional Dependence. *Journal of Econometrics* **171**, 71-85.

Chen, B. & Hong, Y. (2012) Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression. *Econometrica* **80**, 1157-1183.

Chow, G.C. (1960) Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* **28**, 591-605.

De Wachter, S. & Tzavalis, E. (2012) Detection of Structural Breaks in Linear Dynamic Panel Data Models. *Computational Statistics and Data Analysis* **56**, 3020-3234.

Elliott, G. & Müller, U (2006) Efficient Tests for General Persistent Time Variation in Regression Coefficients. *Review of Economic Studies* **73**, 907-940.

Feng, Q., Kao, C. & Lazarová, Š. (2009) Estimation and Identification of Change Points in Panel Models. Working paper.

González, A., Teräsvirta, T. & van Dijk, D. (2005) Panel Smooth Transition Regression Models. Working paper.

Hall, P. & Heyde, C.C. (1980) Martingale Limit Theory and its Applications. Academic Press.

Hausman, J. (1978) Specification Tests in Econometrics. *Econometrica* **46**, 1251-1271.

Horváth, L. & Hušková, M. (2012) Change-point Detection in Panel Data. *Journal of Time Series Analysis* **33**, 631-648.

Hsiao, C. (2003) Analysis of Panel Data, Cambridge. Cambridge University Press.

Hsiao, C. (2007) Panel Data Analysis Advantages and Challenges. *Test* **16**, 1-22.

Juhl, T. & Xiao, Z. (2013) Nonparametric Tests of Moment Condition Stability *Econometric Theory* **29**, 90-114.

Kristensen, D. (2012) Non-parametric Detection and Estimation of Structural Change. *Econometrics Journal* **15**, 420-461.

Li, D., Chen, J. & Gao, J. (2011) Non-parametric Time-varying Coefficient Panel Data Models with Fixed Effects. *Econometrics Journal* **14**, 387–408.

Lin, Z. & Lu, C. (1996) Limit Theorems of Mixing Dependent Random Variables. Science Press.

Peligrad, M. & Utev, S. (1997) Central Limit Theorem for Linear Processes. *Annals of Probability* **25**, 443-456.

Perron, P. (2006) Dealing with Structural Breaks. In Patterson, K. & Mills, T.C. (eds.), *Palgrave Handbook of Econometrics*, Vol. 1: Econometric Theory. Palgrave Macmillan.

Phillips, P.C.B. & Hansen, B.E. (1990) Statistical Inference in Instrumental Variables Regression with I(1) Processes. *Review of Economic Studies* **57**, 99-125.

Robinson, P.M. (1991) Time-Varying Nonlinear Regression. In Hackl, P. (ed.), *Economic Structural Change Analysis and Forecasting.* Springer.

Robinson, P.M. (2012) Nonparametric Trending Regression with Cross-sectional Dependence. *Journal of Econometrics* **169**, 4-14.

Yoshihara, K. (1976) Limiting Behavior of $U-$Statistics for Stationary, Absolutely Regular Processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **35**, 237-252.

Zhang, Y., Su, L. & Phillips, P.C.B. (2012) Testing for Common Trends in Semi-parametric Panel Data Models with Fixed Effects. *Econometrics Journal* **15**, 56-100.

TABLE I

Empirical Size of the $\hat{H}$ Test

N/T	15		20		25		30	
	10%	5%	10%	5%	10%	5%	10%	5%
with i.i.d innovations								
15	0.100	0.050	0.104	0.046	0.096	0.052	0.116	0.044
20	0.106	0.040	0.106	0.044	0.086	0.048	0.082	0.034
25	0.110	0.054	0.102	0.042	0.110	0.056	0.066	0.036
30	0.076	0.030	0.092	0.052	0.090	0.048	0.112	0.058
with cross-sectionally dependent innovations								
15	0.106	0.052	0.102	0.036	0.088	0.042	0.084	0.034
20	0.106	0.064	0.110	0.054	0.110	0.052	0.112	0.050
25	0.084	0.038	0.092	0.046	0.080	0.040	0.114	0.050
30	0.104	0.056	0.100	0.042	0.090	0.050	0.078	0.038

Notes: (1) $\hat{H}$ denotes the test for the overall model stability; (2) 500 iterations; (3) Rejection rates are based on bootstrap critical values with B=99.

TABLE II

Empirical Power of $\hat{H}$ under DGPs P.1-P.3

N/T	15		20		25		30	
	10%	5%	10%	5%	10%	5%	10%	5%
DGP P.1-Single Structural Break								
15	0.832	0.708	0.928	0.876	0.980	0.964	0.998	0.992
20	0.924	0.850	0.984	0.968	1.000	0.992	1.000	1.000
25	0.964	0.916	0.998	0.986	1.000	1.000	1.000	1.000
30	0.986	0.954	1.000	0.998	1.000	1.000	1.000	1.000
DGP P.2- Multiple Structural Breaks								
15	0.692	0.522	0.736	0.564	0.932	0.866	0.928	0.836
20	0.812	0.628	0.878	0.742	0.974	0.926	0.990	0.948
25	0.880	0.758	0.906	0.762	0.982	0.952	0.990	0.958
30	0.914	0.808	0.912	0.810	0.994	0.984	0.996	0.986
DGP P.3-Smooth Structural Changes								
15	0.882	0.776	0.938	0.898	0.988	0.960	1.000	0.996
20	0.940	0.862	0.994	0.972	0.998	0.990	1.000	0.998
25	0.974	0.954	0.990	0.984	1.000	0.998	1.000	1.000
30	0.994	0.974	1.000	0.996	1.000	1.000	1.000	1.000

Notes: (1) $\hat{H}$ denotes the test for the overall model stability; (2) 500 iterations; (3) Rejection rates are based on bootstrap critical values with B=99.

TABLE III

Empirical Size of the $\hat{H}_1$ Test

N/T	15		20		25		30	
	10%	5%	10%	5%	10%	5%	10%	5%
with i.i.d innovations								
15	0.110	0.064	0.130	0.078	0.116	0.056	0.098	0.046
20	0.096	0.046	0.124	0.072	0.094	0.054	0.092	0.044
25	0.104	0.068	0.124	0.070	0.118	0.060	0.086	0.034
30	0.092	0.044	0.100	0.050	0.096	0.056	0.094	0.052
with cross-sectionally dependent innovations								
15	0.110	0.068	0.096	0.056	0.098	0.050	0.084	0.038
20	0.108	0.046	0.100	0.046	0.096	0.050	0.098	0.054
25	0.084	0.032	0.090	0.046	0.084	0.036	0.110	0.058
30	0.106	0.046	0.076	0.022	0.086	0.040	0.090	0.046

Notes: (1) $\hat{H}_1$ denotes the diagnostic test for the potential time-varying interaction; (2) 500 iterations; (3) Rejection rates are based on bootstrap critical values with B=99.

TABLE IV

Empirical Power of $\hat{H}_1$ under DGPs P1.1-P1.3

N/T	15		20		25		30	
	10%	5%	10%	5%	10%	5%	10%	5%
DGP P1.1-Single Structural Break								
15	0.636	0.504	0.820	0.656	0.894	0.812	0.956	0.892
20	0.714	0.580	0.866	0.790	0.956	0.890	0.984	0.964
25	0.800	0.684	0.954	0.892	0.986	0.964	0.996	0.992
30	0.868	0.752	0.968	0.914	0.996	0.988	1.000	1.000
DGP P1.2- Multiple Structural Breaks								
15	0.468	0.290	0.530	0.360	0.776	0.604	0.718	0.572
20	0.520	0.326	0.666	0.444	0.838	0.688	0.832	0.690
25	0.630	0.450	0.626	0.422	0.874	0.740	0.882	0.756
30	0.670	0.524	0.740	0.520	0.930	0.814	0.942	0.860
DGP P1.3-Smooth Structural Changes								
15	0.292	0.198	0.420	0.288	0.532	0.390	0.612	0.472
20	0.358	0.254	0.542	0.394	0.620	0.500	0.752	0.584
25	0.466	0.324	0.622	0.476	0.746	0.620	0.798	0.696
30	0.550	0.372	0.748	0.592	0.808	0.658	0.896	0.798

Notes: (1) $\hat{H}_1$ denotes the diagnostic test for the potential time-varying interaction; (2) 500 iterations; (3) Rejection rates are based on bootstrap critical values with B=99.

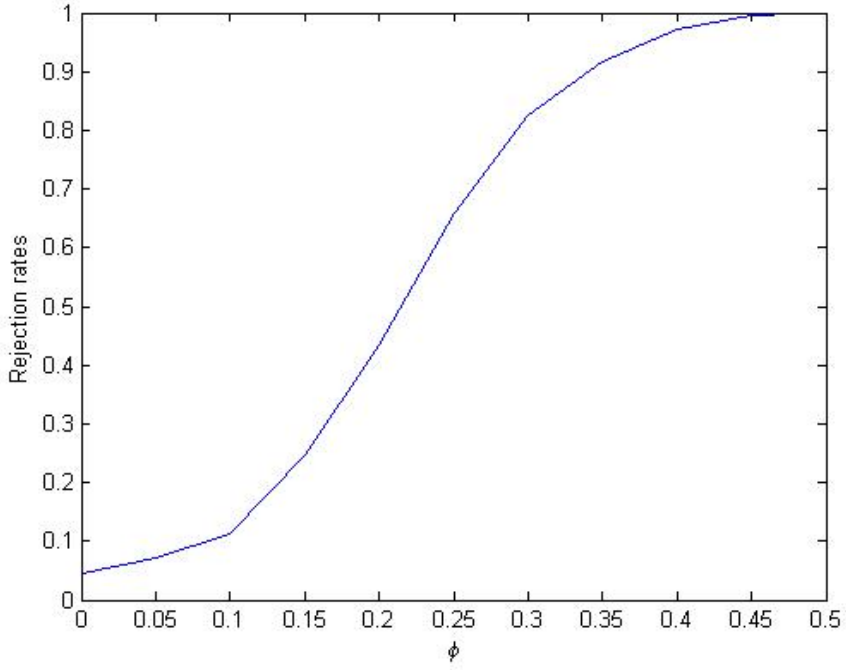

Figure 1: Power curve of the $\hat{H}$ test

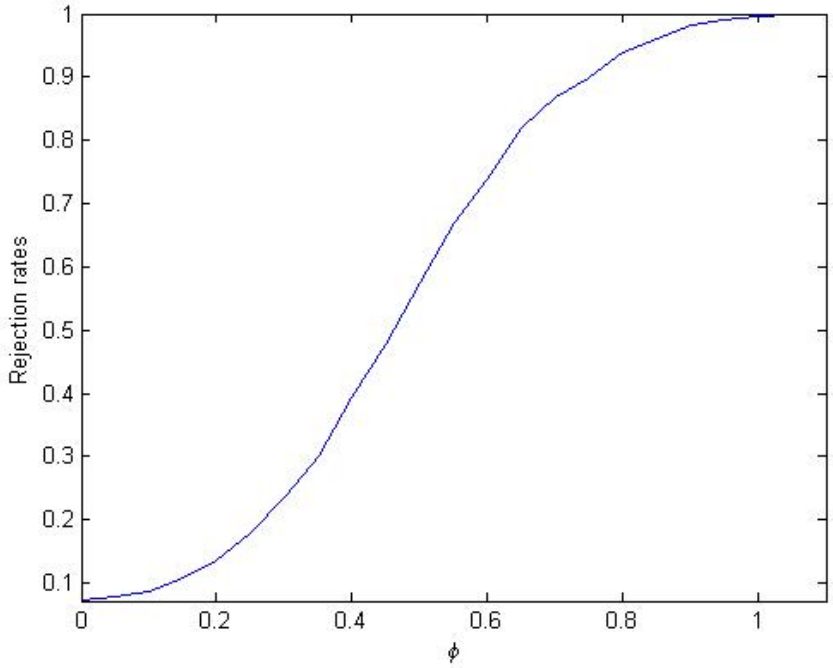

Figure 2: Power curve of the $\hat{H}_1$ test

MATHEMATICAL APPENDIX

Throughout the appendix, we assume $\mu_j = \int_{-1}^{1} u^j k(u)\,du$, $\nu_j = \int_{-1}^{1} u^j k^2(u)\,du$ and $C \in (1,\infty)$ a generic bounded constant.

Proof of Proposition 1: By definition, we have

$$\begin{aligned}\hat{\theta}(\tau) - \theta(\tau) &= [I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)(\Lambda+\psi) - \theta(\tau) \\ &\quad + [I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)D\alpha \\ &\quad + [I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)\varepsilon \\ &= I_\tau(1) + I_\tau(2) + I_\tau(3)\,,\end{aligned}$$

where $\Lambda = 1_N \otimes [\lambda_1, ..., \lambda_T]^{\intercal}$ and $\psi = (X_{11}^{\intercal}\beta_1, ..., X_{1T}^{\intercal}\beta_T, X_{21}^{\intercal}\beta_1, ..., X_{NT}^{\intercal}\beta_T)^{\intercal}$.

By definition, $I_\tau(2) = \bar{0}_{d+1}$, where $\bar{0}_{d+1}$ is a $(d+1)\times 1$ vector of zeros. We shall prove the rest via the following propositions.

Proposition A.1: Under the assumptions of Proposition 1,

$$\frac{1}{NTh}M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau) = \Phi_\mu \otimes \Phi_X(\tau) + o_P(1)$$

uniformly for $\tau \in [0,1]$, where $\Phi_\mu = \mathrm{diag}(\mu_0, \mu_2)$ and $\Phi_X(\tau) = \begin{bmatrix} 1 & \mu_X^{\intercal}(\tau) \\ \mu_X(\tau) & \Sigma_X(\tau) + \mu_X(\tau)\mu_X^{\intercal}(\tau) \end{bmatrix}$.

Proposition A.2: Under the assumptions of Proposition 1,

$$I_\tau(1) = \frac{1}{2}\mu_2\theta''(\tau)h^2 + o_P(h^2)\,.$$

Proposition A.3: Under the assumptions of Proposition 1,

$$\sqrt{NTh}I_\tau(3) \xrightarrow{d} N(\bar{0}_{d+1}, \nu_0\Phi_X^{-1}(\tau)\Phi_{X\varepsilon}(\tau)\Phi_X^{-1}(\tau))\,,$$

where $\Phi_{X\varepsilon}(\tau) = \begin{bmatrix} \sigma_\varepsilon^2(\tau) & \sigma_\varepsilon^2(\tau)\mu_X^{\intercal}(\tau) \\ \sigma_\varepsilon^2(\tau)\mu_X(\tau) & \Sigma_{X\varepsilon}(\tau) + \sigma_\varepsilon^2(\tau)\mu_X(\tau)\mu_X^{\intercal}(\tau) \end{bmatrix}$.

Proof of Proposition A.1: By definition, we have

$$M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau) = M^{\intercal}(\tau)\mathcal{K}(\tau)M(\tau) - M^{\intercal}(\tau)\mathcal{K}(\tau)D[D^{\intercal}\mathcal{K}(\tau)D]^{-1}D^{\intercal}\mathcal{K}(\tau)M(\tau) \tag{A1}$$

and the first term in (A1) is

$$\begin{aligned}M^{\intercal}(\tau)\mathcal{K}(\tau)M(\tau) &= \sum_{i=1}^{N} M_i^{\intercal}(\tau)K(\tau)M_i(\tau) \\ &= \begin{bmatrix} \sum_{i=1}^{N}\sum_{t=1}^{T} k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{\intercal}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{t-\tau T}{Th}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{t-\tau T}{Th}X_{it}^{\intercal}k(\frac{t-\tau T}{Th}) \\ \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{\intercal}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}\frac{t-\tau T}{Th}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{\intercal}\frac{t-\tau T}{Th}k(\frac{t-\tau T}{Th}) \\ \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{t-\tau T}{Th}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{\intercal}\frac{t-\tau T}{Th}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} (\frac{t-\tau T}{Th})^2k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} (\frac{t-\tau T}{Th})^2X_{it}^{\intercal}k(\frac{t-\tau T}{Th}) \\ \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{t-\tau T}{Th}X_{it}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{\intercal}\frac{t-\tau T}{Th}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} (\frac{t-\tau T}{Th})^2X_{it}k(\frac{t-\tau T}{Th}) & \sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{\intercal}(\frac{t-\tau T}{Th})^2k(\frac{t-\tau T}{Th}) \end{bmatrix}.\end{aligned}$$

Let $v_{it} = X_{it} - \mu_X(\frac{t}{T})$ and $v_t = (v_{1t}, \ldots, v_{Nt})^{\intercal}$. By Assumptions A.1 and A.3, (v_t, ε_t) is a non-stationary β-mixing $N\times(d+1)$ random matrix with mean zero and $E(v_{it}v_{it}^{\intercal}) = \Sigma_X(\frac{t}{T})$. The fact $\frac{1}{NTh}M^{\intercal}(\tau)\mathcal{K}(\tau)M(\tau) = \Phi_\mu \otimes \Phi_X(\tau) + o_P(1)$ follows the arguments below:

1. By the definition of Riemann integral and Assumption A.1, it is clear to see that, for $j = 0, 1, 2$, we have $\frac{1}{Th}\sum_{t=1}^{T}(\frac{t-\tau T}{Th})^j k(\frac{t-\tau T}{Th}) = \mu_j + O(\frac{1}{Th})$, $\frac{1}{Th}\sum_{t=1}^{T}\mu_X(\frac{t}{T})(\frac{t-\tau T}{Th})^j k(\frac{t-\tau T}{Th}) = \mu_j\mu_X(\tau) + O(\frac{1}{Th})$ and $\frac{1}{Th}\sum_{t=1}^{T}\mu_X(\frac{t}{T})\mu_X^\intercal(\frac{t}{T})(\frac{t-\tau T}{Th})^j k(\frac{t-\tau T}{Th}) = \mu_j\mu_X(\tau)\mu_X^\intercal(\tau) + O(\frac{1}{Th})$ uniformly for $\tau \in [0, 1]$. Also, by Assumption A.6, we have $\mu_1 = 0$.

2. We have $\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}(\frac{t-\tau T}{Th})^j k(\frac{t-\tau T}{Th})v_{it} = o_P(1)$, given $\frac{\sqrt{NTh}}{\log(NT)} \to \infty$, where we have used
$$\sup_{0\le\tau\le 1}\left\|\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{t-\tau T}{Th}\right)^j k\left(\frac{t-\tau T}{Th}\right)v_{it}\right\| = O_P\left(\sqrt{\frac{\log(NT)}{NTh}}\right), \tag{A2}$$
for $j = 0, 1, 2$. A stationary version of (A2) has been shown in Chen *et al.* (2012). It's straightforward to show that the result still holds under Assumption A.1.

3. Finally, we have for $j = 0, 1, 2$,
$$\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}v_{it}^\intercal\left(\frac{t-\tau T}{Th}\right)^j k\left(\frac{t-\tau T}{Th}\right) = \mu_j\Sigma_X(\tau) + o_P(1), \tag{A3}$$
uniformly for $\tau \in [0, 1]$. (A3) can be shown in a similar way as (A2) under Assumption A.3(iv).

Next, we shall prove that the second term of (A1) is $o_P(NTh)$. Let $Z_\tau = \sum_{t=1}^{T} k(\frac{t-\tau T}{Th})$. The inverse part of the second term in (A1) becomes
$$[D^\intercal\mathcal{K}(\tau)D]^{-1} = \begin{bmatrix} \frac{1}{Z_\tau} - \frac{1}{NZ_\tau} & -\frac{1}{NZ_\tau} & \cdots & -\frac{1}{NZ_\tau} \\ -\frac{1}{NZ_\tau} & \frac{1}{Z_\tau} - \frac{1}{NZ_\tau} & \cdots & -\frac{1}{NZ_\tau} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{NZ_\tau} & -\frac{1}{NZ_\tau} & \cdots & \frac{1}{Z_\tau} - \frac{1}{NZ_\tau} \end{bmatrix}.$$
Denote $\bar{Z}_\tau(i) = \sum_{t=1}^{T} k(\frac{t-\tau T}{Th})v_{it}$ and $\tilde{Z}_\tau(i) = \sum_{t=1}^{T}(\frac{t-\tau T}{Th})k(\frac{t-\tau T}{Th})v_{it}$. The second term becomes
$$M^\intercal(\tau)\mathcal{K}(\tau)D[D^\intercal\mathcal{K}(\tau)D]^{-1}D^\intercal\mathcal{K}(\tau)M(\tau) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & A_\tau & 0 & C_\tau \\ 0 & 0 & 0 & 0 \\ 0 & D_\tau & 0 & B_\tau \end{bmatrix}$$
where
$$\begin{aligned} A_\tau &= \sum_{k=2}^{N} A_\tau(k)[\bar{Z}_\tau(k) - \bar{Z}_\tau(1)]^\intercal, \quad B_\tau = \sum_{k=2}^{N} B_\tau(k)[\bar{Z}_\tau(k) - \bar{Z}_\tau(1)]^\intercal, \\ C_\tau &= \sum_{k=2}^{N} A_\tau(k)[\tilde{Z}_\tau(k) - \tilde{Z}_\tau(1)]^\intercal, \quad D_\tau = \sum_{k=2}^{N} B_\tau(k)[\tilde{Z}_\tau(k) - \tilde{Z}_\tau(1)]^\intercal, \end{aligned}$$
and
$$A_\tau(k) = \frac{\bar{Z}_\tau(k)}{Z_\tau} - \frac{1}{NZ_\tau}\sum_{i=1}^{N}\bar{Z}_\tau(i), \quad B_\tau(k) = \frac{\tilde{Z}_\tau(k)}{Z_\tau} - \frac{1}{NZ_\tau}\sum_{i=1}^{N}\tilde{Z}_\tau(i).$$
Note that $\sum_{i=2}^{N} A_\tau(i) = -A_\tau(1)$ and $\sum_{i=2}^{N} B_\tau(i) = -B_\tau(1)$. As we have shown $\frac{1}{Th}Z_\tau = \mu_0 + o(1)$, Proposition A.1 follows from the lemma below.

Lemma A.1: For each $i \ge 1$, we have $\frac{1}{Th}\bar{Z}_\tau(i) = o_P(1)$ and $\frac{1}{Th}\tilde{Z}_\tau(i) = o_P(1)$ uniformly for $\tau \in [0, 1]$.

Proof of Lemma A.1: We shall prove that $\sup_{\tau\in[0,1]}\|\frac{1}{Th}\sum_{t=1}^{T}(\frac{t-\tau T}{Th})^j k(\frac{t-\tau T}{Th})v_{it}\| = O_P((\log T/Th)^{1/2})$ for $j=0,1$. The proof is similar to that of (B.10) in Chen *et al.* (2012), with $\bar{Q}_{t,N}(v)=\frac{1}{N}\sum_{i=1}^{N}v_{it}$ replaced by $\bar{Q}_t(v)=v_{it}$ and the stationarity condition relaxed. For truncation, we choose $\tilde{Q}_t(v)=\bar{Q}_t(v)I\{\|\bar{Q}_t(v)\|\le T^{1/\delta}l(T)\}$. By choosing the following parameters

$$l(T)\to\infty,\ \frac{T^{1-2/\delta}h}{l^4(T)\log T}\to\infty,\ b=CT^{-1+1/\delta}h^{-1}l(T),\ q=T/(2p)\,,$$
$$p=\frac{1}{\epsilon_* l^2(T)}\sqrt{\frac{T^{1-2/\delta}h}{\log T}},\ \epsilon=\epsilon_* T^{-1}l(T)\sqrt{\log T/Th}\text{ and }\frac{2\sigma^2(q)}{p^2}+\frac{b\epsilon}{2}\le\frac{C}{T^2hp}\,,$$

the desired result follows. ■■

Proof of Proposition A.2: It follows by a second-order Taylor expansion and Proposition A.1. ■

Proof of Proposition A.3: We have

$$M^\intercal(\tau)\mathcal{W}(\tau)\varepsilon = M^\intercal(\tau)\mathcal{K}(\tau)\varepsilon - M^\intercal(\tau)\mathcal{K}(\tau)D[D^\intercal\mathcal{K}(\tau)D]^{-1}D^\intercal\mathcal{K}(\tau)\varepsilon. \tag{A4}$$

Then the first term in (A4) becomes

$$\frac{1}{\sqrt{NTh}}M^\intercal(\tau)\mathcal{K}(\tau)\varepsilon=\begin{bmatrix}\frac{1}{\sqrt{NTh}}\sum_{i=1}^{N}\sum_{t=1}^{T}k(\frac{t-\tau T}{Th})\varepsilon_{it}\\ \frac{1}{\sqrt{NTh}}\sum_{i=1}^{N}\sum_{t=1}^{T}k(\frac{t-\tau T}{Th})X_{it}\varepsilon_{it}\\ \frac{1}{\sqrt{NTh}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\frac{t-\tau T}{Th})k(\frac{t-\tau T}{Th})\varepsilon_{it}\\ \frac{1}{\sqrt{NTh}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\frac{t-\tau T}{Th})k(\frac{t-\tau T}{Th})X_{it}\varepsilon_{it}\end{bmatrix}.$$

Let $L_j(\frac{t-\tau T}{Th})=(\frac{t-\tau T}{Th})^j k(\frac{t-\tau T}{Th})$. It is clear to see for $j=0,1$, we have

$$\frac{1}{\sqrt{NTh}}\sum_{i=1}^{N}\sum_{t=1}^{T}L_j\left(\frac{t-\tau T}{Th}\right)\begin{bmatrix}1\\X_{it}\end{bmatrix}\varepsilon_{it}\xrightarrow{d}N(\bar{0}_{d+1},\nu_{2j}\Phi_{X\varepsilon}(\tau))\,,$$

by applying Theorem 2.2 in Peligrad and Utev (1997). Next, we show that the second term of (A4) is of small order, which is $M^\intercal(\tau)\mathcal{K}(\tau)D[D^\intercal\mathcal{K}(\tau)D]^{-1}D^\intercal\mathcal{K}(\tau)\varepsilon=o_P(\sqrt{NTh})$.

Let $\varepsilon_\tau(i)=\sum_{t=1}^{T}k(\frac{t-\tau T}{Th})\varepsilon_{it}$. We have $D^\intercal\mathcal{K}(\tau)\varepsilon=[\varepsilon_\tau(2)-\varepsilon_\tau(1),\cdots,\varepsilon_\tau(N)-\varepsilon_\tau(1)]^\intercal$ and thus

$$M^\intercal(\tau)\mathcal{K}(\tau)D[D^\intercal\mathcal{K}(\tau)D]^{-1}D^\intercal\mathcal{K}(\tau)\varepsilon=(0,U_\tau^\intercal,0,V_\tau^\intercal)^\intercal\,,$$

where

$$U_\tau=\sum_{k=2}^{N}A_\tau(k)[\varepsilon_\tau(k)-\varepsilon_\tau(1)]=\sum_{k=1}^{N}A_\tau(k)\varepsilon_\tau(k)\,,$$

and

$$V_\tau=\sum_{k=2}^{N}B_\tau(k)[\varepsilon_\tau(k)-\varepsilon_\tau(1)]=\sum_{k=1}^{N}B_\tau(k)\varepsilon_\tau(k)\,.$$

To show U_τ and V_τ are $o_P(\sqrt{NTh})$, we use Chebyshev's inequality and shall verify that $E(U_\tau)=o(\sqrt{NTh})$, $E(V_\tau)=o(\sqrt{NTh})$, $E(U_\tau U_\tau^\intercal)=o(NTh)$ and $E(V_\tau V_\tau^\intercal)=o(NTh)$. It's easy to see the

condition for mean holds. To save space, we only derive for the second moment here. We have

$$\frac{1}{NTh}E(U_\tau U_\tau^\intercal) = \frac{1}{NTh}E\Bigg\{\Bigg[\sum_{k=1}^{N} A_\tau(k)\varepsilon_\tau(k)\Bigg]\Bigg[\sum_{k=1}^{N} A_\tau(k)\varepsilon_\tau(k)\Bigg]^\intercal\Bigg\}$$

$$= \frac{1}{NTh}\frac{1}{Z_T^2}\Bigg\{E\Bigg[\sum_{k=1}^{N}\bar{Z}_\tau(k)\varepsilon_\tau(k)\sum_{k'=1}^{N}\bar{Z}_\tau^\intercal(k')\varepsilon_\tau(k')\Bigg] - E\Bigg[\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_\tau(i)\sum_{k=1}^{N}\varepsilon_\tau(k)\sum_{k'=1}^{N}\bar{Z}_\tau^\intercal(k')\varepsilon_\tau(k')\Bigg]$$

$$- E\Bigg[\sum_{k=1}^{N}\varepsilon_\tau(k)\bar{Z}_\tau(k)\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_\tau^\intercal(i)\sum_{k'=1}^{N}\varepsilon_\tau(k')\Bigg] + E\Bigg\{\Bigg[\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_\tau(i)\Bigg]\Bigg[\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_\tau(i)\Bigg]^\intercal\Bigg[\sum_{k=1}^{N}\varepsilon_\tau(k)\Bigg]^2\Bigg\}\Bigg\}. \quad \text{(A5)}$$

We only show the first term of (A5) is $o(1)$ and the rest can be derived in a similar way.

For simplicity, let $k_{t,\tau} = k(\frac{t-\tau T}{Th})$. Given $\frac{1}{Th}Z_\tau = O_P(1)$, we shall show

$$\frac{1}{NT^3h^3}E\Bigg[\sum_{k=1}^{N}\bar{Z}_\tau(k)\varepsilon_\tau(k)\sum_{k'=1}^{N}\bar{Z}_\tau^\intercal(k')\varepsilon_\tau(k')\Bigg]$$

$$= \frac{1}{T^3h^3}\sum_{t=1}^{T}\sum_{t'=1}^{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}k_{t,\tau}k_{t',\tau}k_{s,\tau}k_{s',\tau}E\Bigg[\frac{1}{N}\Bigg(\sum_{k=1}^{N}v_{kt}\varepsilon_{ks}\Bigg)\Bigg(\sum_{k'=1}^{N}v_{k't'}^\intercal\varepsilon_{k's'}\Bigg)\Bigg] = o(1)$$

To save space, we only prove for the case $t > t' > s > s'$ here. Suppose $t - t'$ is the largest among $\{t-t', t'-s, s-s'\}$, then

$$\Bigg\|\frac{1}{T^3h^3}\sum_{t=4}^{T}\sum_{t'=3}^{t-1}\sum_{s=2}^{t'-1}\sum_{s'=1}^{s-1}k_{t,\tau}k_{t',\tau}k_{s,\tau}k_{s',\tau}E\Bigg[\frac{1}{N}\Bigg(\sum_{k=1}^{N}v_{kt}\varepsilon_{ks}\Bigg)\Bigg(\sum_{k'=1}^{N}v_{k',t'}^\intercal\varepsilon_{k',s'}\Bigg)\Bigg]\Bigg\|$$

$$\leq \frac{C}{T^3h^3}\sum_{t=4}^{T}\sum_{t'=3}^{t-1}\sum_{s=2}^{t'-1}\sum_{s'=1}^{s-1}k_{t,\tau}k_{t',\tau}k_{s,\tau}k_{s',\tau}C_M^{\frac{1}{1+\eta}}\beta^{\frac{\eta}{1+\eta}}(t-t')$$

$$\leq \frac{C}{T^3h^3}\sum_{t=1}^{T}\sum_{s=1}^{T}k_{t,\tau}k_{s,\tau}\sum_{l=1}^{T-1}l\beta^{\frac{\eta}{1+\eta}}(l) = O\Bigg(\frac{1}{Th}\Bigg),$$

where we have used Lemma 1 in Yoshihara (1976). We have $C_M = \max(C_M^1, C_M^2) < \infty$ by Assumptions A.1, A.2(ii), A.3(ii) and A.4, where $C_M^1 = \sup_{t,s,t',s'} E\|\frac{1}{N}(\sum_{k=1}^{N}v_{kt}\varepsilon_{ks})(\sum_{k'=1}^{N}v_{k't'}^\intercal\varepsilon_{k's'})\|^{1+\eta}$ and $C_M^2 = \sup_{t,s,t',s'}\int\int\|\frac{1}{N}(\sum_{k=1}^{N}v_{kt}\varepsilon_{ks})(\sum_{k'=1}^{N}v_{k't'}^\intercal\varepsilon_{k's'})\|^{1+\eta}dF(v_t)dF(v_{t'},\varepsilon_s,\varepsilon_{s'})$. The other cases are similar. ■

The desired result of Proposition 1 thus follows Propositions A.1-A.3 and the block matrix inversion formula. ■

Proof of Proposition 2: By definition, we have

$$\hat{\beta} - \beta = \Bigg[\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}-\bar{X}_i)^\intercal\Bigg]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(\varepsilon_{it}-\bar{\varepsilon}_i).$$

We can show that the inverse part converges to a constant at the rate NT and $\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(\varepsilon_{it}-\bar{\varepsilon}_i)$ is $O_P(\sqrt{NT})$. Replacing X_{it} with $v_{it} + \mu_X(\frac{t}{T})$, where v_{it} is defined in the proof of Proposition 1, inside

the inverse, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}-\bar{X}_i)^{\intercal} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}v_{it}^{\intercal} - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}v_{is}^{\intercal}$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\mu_X^{\intercal}\left(\frac{t}{T}\right) - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}v_{it}\mu_X^{\intercal}\left(\frac{s}{T}\right)$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X\left(\frac{t}{T}\right)v_{it}^{\intercal} - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mu_X\left(\frac{t}{T}\right)v_{is}^{\intercal}$$
$$+\frac{1}{T}\sum_{t=1}^{T}\mu_X\left(\frac{t}{T}\right)\mu_X^{\intercal}\left(\frac{t}{T}\right) - \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mu_X\left(\frac{t}{T}\right)\mu_X^{\intercal}\left(\frac{s}{T}\right). \quad \text{(A6)}$$

For the first term of (A6), we have $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}v_{it}^{\intercal} = \int_0^1 \Sigma_X(\tau)\,d\tau + o_P(1)$ by the LLN for the β-mixing process (Corollary 8.5.1, Lin and Lu, 1996). The second term of (A6) is of order $O_p(T^{-1})$ by Chebyshev's inequality and Lemma 1 in Yoshihara (1976). For the third to the sixth term, they are of order $O_p((NT)^{-1/2})$ by applying the LLN for β-mixing process. By the definition of Riemann integral, the last two terms become

$$\frac{1}{T}\sum_{t=1}^{T}\mu_X\left(\frac{t}{T}\right)\mu_X^{\intercal}\left(\frac{t}{T}\right) - \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mu_X\left(\frac{t}{T}\right)\mu_X^{\intercal}\left(\frac{s}{T}\right) = \int_0^1 \mu_X(\tau)\mu_X^{\intercal}(\tau)\,d\tau - \int_0^1 \mu_X(\tau)\,d\tau\int_0^1 \mu_X^{\intercal}(\tau)\,d\tau + o(1).$$

Next, it is clear to see that

$$\sqrt{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(\varepsilon_{it}-\bar{\varepsilon}_i) = \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}[v_{it}\varepsilon_{it} + \mu_X\left(\frac{t}{T}\right)\varepsilon_{it} - \frac{1}{T}\sum_{s=1}^{T}\mu_X\left(\frac{s}{T}\right)\varepsilon_{it}]$$
$$-\sqrt{NT}\frac{1}{NT^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}v_{it}\varepsilon_{is}\,,$$

where the first term converges by the CLT for the m.d.s. (Hall and Heyde, 1980), and the second term is of small order $O_p(1/\sqrt{T})$ by Lemma 1 in Yoshihara (1976), Assumptions A.1 and A.4. Therefore, we have

$$\sqrt{NT}(\hat{\beta}-\beta) = A^{-1}(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}[v_{it}\varepsilon_{it} + \mu_X\left(\frac{t}{T}\right)\varepsilon_{it} - \frac{1}{T}\sum_{s=1}^{T}\mu_X(\frac{s}{T})\varepsilon_{it}] + o_P(1)\,,$$

where

$$A = \int_0^1 \Sigma_X(\tau)\,d\tau + \int_0^1 \mu_X(\tau)\,\mu_X^{\intercal}(\tau)\,d\tau - \int_0^1 \mu_X(\tau)\,d\tau\int_0^1 \mu_X^{\intercal}(\tau)\,d\tau. \quad \text{(A7)}$$

For the intercept, we have

$$\sqrt{NT}(\hat{\lambda}-\lambda) = -\bar{X}^{\intercal}\sqrt{NT}(\hat{\beta}-\beta) + \sqrt{NT}\bar{\varepsilon}\,,$$

where $\bar{\varepsilon} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}$. Therefore,

$$\sqrt{NT}\begin{bmatrix}\hat{\lambda}-\lambda\\ \hat{\beta}-\beta\end{bmatrix} = \begin{bmatrix}1 & -\bar{X}^{\intercal}\\ \bar{0}_d & I_d\end{bmatrix}\sqrt{NT}\begin{bmatrix}\bar{\varepsilon}\\ \hat{\beta}-\beta\end{bmatrix}$$

$$= \begin{bmatrix}1 & -\int_0^1 \mu_X^{\intercal}(\tau)\,d\tau\\ \bar{0}_d & I_d\end{bmatrix}\begin{bmatrix}1 & \bar{0}_d^{\intercal} & \bar{0}_d^{\intercal}\\ -A^{-1}\int_0^1 \mu_X(\tau)\,d\tau & A^{-1} & A^{-1}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu_X(\frac{t}{T})\varepsilon_{it}\\ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\varepsilon_{it}\end{bmatrix} + o_P(1)\,.$$

By applying Theorem 2.2 in Peligrad and Utev (1997), we have

$$\sqrt{NT}\begin{bmatrix}\hat{\lambda}-\lambda\\ \hat{\beta}-\beta\end{bmatrix} \xrightarrow{d} N(\bar{0}_{d+1}, \Xi V \Xi^{\intercal})\,,$$

where Ξ and V are defined in Proposition 2.

Proof of Proposition 3: We first assume the trend function is smooth, and show near the end of the proof that the same result holds when $\lambda(\cdot)$ has a finite number of bounded jumps. Define

$$s(\tau) = (1,0)S(\tau) = (1,0)[Z^{\intercal}(\tau)\mathcal{K}(\tau)Z(\tau)]^{-1}Z^{\intercal}(\tau)\mathcal{K}(\tau)\,,$$
$$\tilde{\lambda} = (I_{NT}-\tilde{S})\Lambda \text{ and } \tilde{\varepsilon} = (I_{NT}-\tilde{S})\varepsilon\,.$$

We have

$$\hat{\beta}_P - \beta = (\tilde{X}^{\intercal}\tilde{D}\tilde{X})^{-1}\tilde{X}^{\intercal}\tilde{D}\tilde{\lambda} + (\tilde{X}^{\intercal}\tilde{D}\tilde{X})^{-1}\tilde{X}^{\intercal}\tilde{D}D^{*}\alpha + (\tilde{X}^{\intercal}\tilde{D}\tilde{X})^{-1}\tilde{X}^{\intercal}\tilde{D}\tilde{\varepsilon}$$
$$= \Pi_{NT}(1) + \Pi_{NT}(2) + \Pi_{NT}(3)\,.$$

It's obvious that $\Pi_{NT}(2) = \bar{0}_d$. Similar to Chen *et al.* (2012), we have $(\frac{1}{NT}\tilde{X}^{\intercal}\tilde{D}\tilde{X})^{-1} = (\int_0^1 \Sigma_X(\tau)\,d\tau)^{-1} + o_P(1)$ and $\sqrt{NT}\Pi_{NT}(1) = o_P(1)$. For $\Pi_{NT}(3)$, we know that

$$\tilde{X}^{\intercal}\tilde{D}\tilde{\varepsilon} = \tilde{X}^{\intercal}\tilde{\varepsilon} - X^{\intercal}D(D^{\intercal}D)^{-1}D^{\intercal}\varepsilon = \Pi_{NT}^{*}(1) + \Pi_{NT}^{*}(2)\,.$$

For $\Pi_{NT}^{*}(1)$, it is clear to see that

$$\frac{1}{\sqrt{NT}}\Pi_{NT}^{*}(1) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\varepsilon_{it} + o_P(1) \xrightarrow{d} N(\bar{0}_d, \int_0^1 \Sigma_{X\varepsilon}(\tau)\,d\tau)$$

by the CLT for m.d.s. (Hall and Heyde, 1980).

Let $A_T(k) = \sum_{t=1}^{T}X_{kt} - \sum_{t=1}^{T}X_{1t} = \sum_{t=1}^{T}v_{kt} - \sum_{t=1}^{T}v_{1t}$ and $B_T(k) = \sum_{t=1}^{T}\varepsilon_{kt} - \sum_{t=1}^{T}\varepsilon_{1t}$. Also note that $A_T(1) = \bar{0}_d$ and $B_T(1) = 0$.

By definition, we have

$$\Pi_{NT}^{*}(2) = \frac{1}{T}\sum_{k=1}^{N}A_T(k)B_T(k) - \frac{1}{NT}\sum_{k=1}^{N}A_T(k)\sum_{k=1}^{N}B_T(k)$$
$$= \frac{1}{T}\sum_{k=1}^{N}\sum_{t=1}^{T}v_{kt}\sum_{t=1}^{T}\varepsilon_{kt} - \frac{1}{NT}\sum_{k=1}^{N}\sum_{t=1}^{T}v_{kt}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}$$
$$= \Pi_{NT}^{*}(2,1) + \Pi_{NT}^{*}(2,2)\,.$$

As $\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}v_{kt} = O_P(1)$ and $\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt} = O_P(1)$, we know $\Pi_{NT}^{*}(2,2) = o_P(\sqrt{NT})$.

We shall show $\Pi^*_{NT}(2,1) = o_P(\sqrt{NT})$. Note that

$$E\left(\frac{1}{T}\sum_{k=1}^{N}\sum_{t=1}^{T} v_{kt}\sum_{t=1}^{T}\varepsilon_{kt}\right) = \frac{1}{T}\sum_{k=1}^{N}\sum_{t=2}^{T} E(v_{kt}\varepsilon_{kt}) + \frac{1}{T}\sum_{k=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T} E(v_{kt}\varepsilon_{ks})$$
$$= \frac{\sqrt{N}}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\frac{1}{\sqrt{N}}\sum_{k=1}^{N} E(v_{kt}\varepsilon_{ks})$$
$$\leq \frac{C\sqrt{N}}{T}\sum_{t=1}^{T}\sum_{l=1}^{T} C_M^{\frac{1}{1+\eta}}\beta^{\frac{\eta}{1+\eta}}(l) = O(\sqrt{N}) = o(\sqrt{NT})\,,$$

where we have used Lemma 1 in Yoshihara (1976). We can show $C_M = \max(C_M^1, C_M^2)$ is bounded given Assumptions A.4, where

$$C_M^1 = \sup_{t,s} E\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N} v_{kt}\varepsilon_{ks}\|^{1+\eta} \leq \sup_{t,s}\left[E(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\|v_{it}v_{jt}^{\intercal}\varepsilon_{is}\varepsilon_{js}\|)^{1+\eta}\right]^{\frac{1}{2}}$$
$$\leq \left[\sup_{t,s}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}(E\|v_{it}v_{jt}^{\intercal}\varepsilon_{is}\varepsilon_{js}\|^{1+\eta})^{\frac{1}{1+\eta}}\right]^{\frac{1+\eta}{2}}$$
$$\leq \left\{\sup_{t,s}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[E(\|v_{it}v_{jt}^{\intercal}\|)^{2(1+\eta)}\right]^{\frac{1}{2(1+\eta)}}\left[E(\varepsilon_{is}\varepsilon_{js})^{2(1+\eta)}\right]^{\frac{1}{2(1+\eta)}}\right\}^{\frac{1+\eta}{2}} < \infty\,.$$

$$C_M^2 = \sup_{t,s}\int\int\|\frac{1}{\sqrt{N}}\sum_{k=1}^{N} v_{kt}\varepsilon_{ks}\|^{1+\eta}\,dF(v_t)dF(\varepsilon_s)$$
$$\leq \left[\sup_{t,s}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\int\int\|v_{it}v_{jt}^{\intercal}\|^{1+\eta}|\varepsilon_{is}\varepsilon_{js}|^{1+\eta}\,dF(v_t)dF(\varepsilon_s)^{\frac{1}{1+\eta}}\right]^{\frac{1+\eta}{2}}$$
$$= \left[\sup_{t,s}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}(E\|v_{it}v_{jt}^{\intercal}\|^{1+\eta})^{\frac{1}{1+\eta}}(E|\varepsilon_{is}\varepsilon_{js}|^{1+\eta})^{\frac{1}{1+\eta}}\right]^{\frac{1+\eta}{2}} < \infty\,.$$

To check for the variance, which is $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{t'=1}^{T}\sum_{s'=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E(\varepsilon_{it}\varepsilon_{jt'}v_{is}v_{js'}^{\intercal}) = o(NT)$, there are four cases:

Case a: $t = s = t' = s'$;

Case b: $t = t' = s'$, $s = s' = t$ and etc.;

Case c: $t = t'$, $s = s'$, $t = s$ and etc.;

Case d: $t \neq t' \neq s \neq s'$.

For Case a, we have $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} E(\varepsilon_{it}\varepsilon_{jt}v_{it}v_{jt}) = O(\frac{N}{T})$. By applying Lemma 1 in Yoshihara (1976), we have $O(\frac{N}{T})$ for Case b. Case c and Case d are both of order $O(N)$. Finally, by Chebyshev's inequality, $\Pi^*_{NT}(2,1) = o_P(\sqrt{NT})$. The desired result thus follows.

We shall show the same result holds when λ_t has a finite number of bounded jumps. Without loss of generality, we assume that λ_t has only one jump at $t = t_0$ which is $\lambda(t/T) = \lambda_1(t/T)1_{t<t_0} + \lambda_2(t/T)1_{t\geq t_0}$ for some smooth functions $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$. By the definition of $\hat{\beta}_P$, we only need to show that $\sqrt{NT}(\tilde{X}^{\intercal}\tilde{D}\tilde{X})^{-1}\tilde{X}^{\intercal}\tilde{D}\tilde{\lambda} = o_P(1)$ still holds. For $\tilde{X}^{\intercal}\tilde{D}\tilde{\lambda}$, we have

$$\tilde{X}^{\intercal}\tilde{D}\tilde{\lambda} = \tilde{X}^{\intercal}\tilde{\lambda} - X^{\intercal}D(D^{\intercal}D)^{-1}D^{\intercal}\Lambda = \rho_1 + \rho_2.$$

It's obvious that $\rho_2 = \bar{0}_d$. By definition, we have

$$\begin{aligned}\frac{1}{\sqrt{NT}}\rho_1 &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}[\lambda(\frac{t}{T}) - s(\frac{t}{T})\Lambda] - \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v^{\intercal}s^{\intercal}(\frac{t}{T})[\lambda(\frac{t}{T}) - s(\frac{t}{T})\Lambda] \\ &\quad + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mu_X(\frac{t}{T}) - \mu_X^{\intercal}s^{\intercal}(\frac{t}{T})][\lambda(\frac{t}{T}) - s(\frac{t}{T})\Lambda] \\ &= \rho_{11} + \rho_{12} + \rho_{13},\end{aligned}$$

where $v = (v_{11}, \ldots, v_{1T}, \ldots, v_{NT})^{\intercal}$, $\mu_X = 1_N \otimes (\mu_X(\frac{1}{T}), \ldots, \mu_X(1))^{\intercal}$ and $s(\tau)\Lambda = (1\ 0)S(\tau)\Lambda = \frac{1}{Th}\sum_{t=1}^{T} k(\frac{t-\tau T}{Th})\lambda(\frac{t}{T}) + o(1)$.

For each h, we construct a continuously differentiable function λ_h, such that $\lambda_h(\tau) = \lambda(\tau)$ for $\tau \in [0, t_0/T - h) \cup [t_0/T + h, 1]$ and $\lambda_h(\tau) = \lambda(t_0/T - h) + (\tau - t_0/T + h)^2\frac{\lambda(t_0/T+h)-\lambda(t_0/T-h)}{4h^2}$ for $\tau \in [t_0/T - h, t_0/T + h)$. We rewrite ρ_{11} as

$$\begin{aligned}\rho_{11} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}[\lambda(\frac{t}{T}) - \frac{1}{Th}\sum_{s=1}^{T} k(\frac{s-t}{Th})\lambda(\frac{s}{T})] + o_P(1) \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}[\lambda_h(\frac{t}{T}) - \frac{1}{Th}\sum_{s=1}^{T} k(\frac{s-t}{Th})\lambda_h(\frac{s}{T})] \\ &+ \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}[(\lambda - \lambda_h)(\frac{t}{T}) - \frac{1}{Th}\sum_{s=1}^{T} k(\frac{s-t}{Th})(\lambda - \lambda_h)(\frac{s}{T})] + o_P(1) \\ &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=t_0-2\lfloor Th\rfloor}^{t_0+2\lceil Th\rceil} v_{it}[(\lambda - \lambda_h)(\frac{t}{T}) - \frac{1}{Th}\sum_{s=t-\lfloor Th\rfloor}^{t+\lceil Th\rceil} k(\frac{s-t}{Th})(\lambda - \lambda_h)(\frac{s}{T})] + o_P(1) \\ &\leq C\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=t_0-2\lfloor Th\rfloor}^{t_0+2\lceil Th\rceil} v_{it} + o_P(1) \\ &= O_P(\sqrt{h})\,,\end{aligned}$$

as we know $\sup_{\tau\in[t_0/T-3h, t_0/T+3h]}|\lambda(\tau) - \lambda_h(\tau)| \leq C$. Similarly, we have

$$\rho_{12} \leq C\frac{1}{\sqrt{NT}}N\frac{1}{NTh}\sum_{s=t_0-3\lfloor Th\rfloor}^{t_0+3\lceil Th\rceil}\sum_{i=1}^{N} v_{is}\sum_{t=t_0-2\lfloor Th\rfloor}^{t_0+2\lceil Th\rceil} k(\frac{s-t}{Th}) = O_P(\sqrt{h})\,,$$

and also $\rho_{13} = O_P(\sqrt{NT}h^3) = o_P(1)$. Therefore, we complete the proof of Proposition 3. ■

Proof of Theorem 1: We have that

$$\begin{aligned}\hat{Q} &= N\sqrt{h}\sum_{t=1}^{T}(\hat{\theta}_t - \theta + \theta - \hat{\theta})^{\intercal}\hat{\Omega}_t(\hat{\theta}_t - \theta + \theta - \hat{\theta}) \\ &= N\sqrt{h}\sum_{t=1}^{T}(\hat{\theta}_t - \theta)^{\intercal}\hat{\Omega}_t(\hat{\theta}_t - \theta) - 2N\sqrt{h}\sum_{t=1}^{T}(\hat{\theta}_t - \theta)^{\intercal}\hat{\Omega}_t(\hat{\theta} - \theta) + N\sqrt{h}\sum_{t=1}^{T}(\hat{\theta} - \theta)\hat{\Omega}_t(\hat{\theta} - \theta) \\ &= J_1 - 2J_2 + J_3\,.\end{aligned}$$

We shall decompose the proof to three sub-theorems, which show that the asymptotic mean and variance are determined by nonparametric estimation and the estimation uncertainty of parametric estimation is asymptotically negligible.

Theorem A.1: Under the assumptions of Theorem 1, $(J_1 - \hat{C})/\hat{S} \xrightarrow{d} N(0,1)$.

Theorem A.2: Under the assumptions of Theorem 1, $J_2 \xrightarrow{p} 0$.

Theorem A.3: Under the assumptions of Theorem 1, $J_3 \xrightarrow{p} 0$.

Proof of Theorem A.1: Let $k_{st} = k(\frac{s-t}{Th})$. Expand J_1

$$
\begin{aligned}
J_1 &= N\sqrt{h}\sum_{t=1}^{T}\left\{[I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)\varepsilon\right\}^{\intercal} \\
&\quad \times \begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N} X_{it}^{\intercal} \\ \frac{1}{N}\sum_{i=1}^{N} X_{it} & \frac{1}{N}\sum_{i=1}^{N} X_{it}X_{it}^{\intercal} \end{bmatrix} \left\{[I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)\varepsilon\right\} \\
&= \frac{1}{NT^2h^{3/2}}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Bigg[\varepsilon_{is_1}\varepsilon_{js_2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t} + \varepsilon_{is_1}\varepsilon_{js_2}v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js_2} \\
&\quad + \varepsilon_{is_1}\varepsilon_{js_2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\mu_X^{\intercal}\left(\frac{t}{T}\right)\Sigma_X^{-1}\left(\frac{t}{T}\right)\mu_X\left(\frac{t}{T}\right) - \varepsilon_{is_1}\varepsilon_{js_2}\mu_X^{\intercal}\left(\frac{s_1}{T}\right)\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\Sigma_X^{-1}\left(\frac{t}{T}\right)\mu_X\left(\frac{t}{T}\right) \\
&\quad - \varepsilon_{is_1}\varepsilon_{js_2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\mu_X^{\intercal}\left(\frac{t}{T}\right)\Sigma_X^{-1}\left(\frac{t}{T}\right)\mu_X\left(\frac{s_2}{T}\right) + \varepsilon_{is_1}\varepsilon_{js_2}\mu_X^{\intercal}\left(\frac{s_1}{T}\right)\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\Sigma_X^{-1}\left(\frac{t}{T}\right)\mu_X\left(\frac{s_2}{T}\right) \\
&\quad - \varepsilon_{is_1}\varepsilon_{js_2}v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\Sigma_X^{-1}\left(\frac{t}{T}\right)\mu_X\left(\frac{t}{T}\right) + \varepsilon_{is_1}\varepsilon_{js_2}v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\Sigma_X^{-1}\left(\frac{t}{T}\right)\mu_X\left(\frac{s_2}{T}\right) \\
&\quad - \varepsilon_{is_1}\varepsilon_{js_2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\mu_X^{\intercal}\left(\frac{t}{T}\right)\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js_2} + \varepsilon_{is_1}\varepsilon_{js_2}\mu_X^{\intercal}\left(\frac{s_1}{T}\right)\sum_{t=1}^{T}k_{s_1t}k_{s_2t}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js_2}\Bigg] + o_P(1) \\
&= \frac{1}{NT^2h^{3/2}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}k_{st}^2\varepsilon_{is}\varepsilon_{js} + \frac{1}{NT^2h^{3/2}}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}k_{st}^2\varepsilon_{is}v_{is}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)\varepsilon_{js}v_{js} \\
&\quad + \frac{1}{NT^2h^{3/2}}\sum_{t=1}^{T}\sum_{s_1=1}^{T}\sum_{s_2\neq s_1,s_2=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}k_{s_1t}k_{s_2t}\varepsilon_{is_1}\varepsilon_{js_2}[1 + v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js_2}] + o_P(1) \\
&= J_{11} + J_{12} + J_{13} + o_P(1)\,,
\end{aligned}
$$

where we have used Lemma A.2 below for the second equality and Riemann integral for the third equality.

Lemma A.2: Under the assumptions of Theorem 1, we have

$$
\sup_{s_1,s_2\in[0,1]}\left\|\frac{1}{NTh^2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}H_t\sum_{k=1}^{N}v_{kt}g_t^{\intercal}\right\| = O_P\left(\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\right), \tag{A8}
$$

$$
\sup_{s_1,s_2\in[0,1]}\left\|\frac{1}{NTh^2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}H_t\sum_{k=1}^{N}v_{kt}v_{kt}^{\intercal}G_t - \frac{1}{Th^2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}H_t\Sigma_X\left(\frac{t}{T}\right)G_t\right\| = O_P\left(\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\right),
$$

where $H_t = H(\frac{t}{T})$ is either some $1\times d$ vector or $d\times d$ matrix, $G_t = G(\frac{t}{T})$ is either some $d\times 1$ vector or $d\times d$ matrix and $g_t = g(\frac{t}{T})$ is either some $1\times d$ or 1×1 vector with each element (of the matrix or the vector) a continuously differentiable function with respect to t. For instance, let $H_t = G_t = \Sigma_X^{-1}(\frac{t}{T})$ and $g_t = 1$.

Proof of Lemma A.2: They are similar to (B.10) in Chen *et al.* (2012). For simplicity, we only show (A8) and concentrate on the difference between the current proof and that of (B.10). Let $l(\cdot)$ be any

positive function that satisfies $l(n) \to \infty$ as $n \to \infty$. Then, it suffices to show

$$\sup_{s_1,s_2\in[0,1]}\left\|\frac{1}{Th^2}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}H_t\frac{1}{N}\sum_{k=1}^{N}v_{kt}g_t\right\| = o_P\left(l(NT)\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\right).$$

We can cover the interval $[0,1]^2$ by a finite number of squares $\{B_{l_1,l_2}\}$ with center at (b_{l_1}, b_{l_2}) and with width and length $\delta_{NT} = o(h^3)$. In total, we have U_{NT} squares, where $U_{NT} = O(\delta_{NT}^{-2})$. Define $\bar{k}_t(s_1,s_2) = \frac{1}{Th^2}k_{s_1t}k_{s_2t}H_t$ and $\bar{Q}_t = \frac{1}{N}\sum_{k=1}^{N}v_{kt}$. Obviously, H_t and g_t are bounded on [0,1]. Then we have

$$\begin{aligned}\sup_{s_1,s_2\in[0,1]}\left\|\sum_{t=1}^{T}\bar{k}_t(s_1,s_2)\bar{Q}_tg_t\right\| &\le max_{1\le l_1,l_2\le\sqrt{U_{NT}}}\sup_{(s_1,s_2)\in B_{l_1,l_2}}\left\|\sum_{t=1}^{T}\bar{k}_t(s_1,s_2)\bar{Q}_tg_t-\sum_{t=1}^{T}\bar{k}_t(b_{l_1},b_{l_2})\bar{Q}_tg_t\right\| \\ &+ max_{1\le l_1,l_2\le\sqrt{U_{NT}}}\sup_{(s_1,s_2)\in B_{l_1,l_2}}\left\|\sum_{t=1}^{T}\bar{k}_t(b_{l_1},b_{l_2})\bar{Q}_tg_t\right\| = \Theta_1+\Theta_2\,.\end{aligned}$$

Take $\delta_{NT} = O(l(NT)\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}h^3)$, we have $\Theta_1 = O_P(\delta_{NT}/h^3E\|\bar{Q}_T\|) = o_P(l(NT)\sqrt{\frac{\log(NT)}{\sqrt{NT}h}})$. We use truncation technique by defining

$$\begin{aligned}\tilde{Q}_t &= \bar{Q}_tI\{\|\bar{Q}_t\| \le N^{-1/2}T^{1/\delta}l(NT)\} \\ \tilde{Q}_t^c &= \bar{Q}_t - \tilde{Q}_t\,.\end{aligned}$$

Then

$$\begin{aligned}\Theta_2 &\le \max_{1\le l_1,l_2\le\sqrt{U_{NT}}}\|\sum_{t=1}^{T}\bar{k}_t(b_{l_1},b_{l_2})\tilde{Q}_tg_t\| + \max_{1\le l_1,l_2\le\sqrt{U_{NT}}}\|\sum_{t=1}^{T}\bar{k}_t(b_{l_1},b_{l_2})\tilde{Q}_t^cg_t\| \\ &= \Theta_3+\Theta_4\,.\end{aligned}$$

It's easy to see that $\Theta_4 = o_P(l(NT)\sqrt{\frac{\log(NT)}{\sqrt{NT}h}})$ by Chebyshev's inequality and Assumption A.3(iii). Observe $\|\bar{k}_t(b_{l_1},b_{l_2})\tilde{Q}_tg_t\| \le CN^{-1/2}T^{-1+1/\delta}h^{-2}l(NT)$. We choose $l(\cdot)$ to satisfy

$$\frac{T^{1-\frac{2}{\delta}}}{l^4(NT)\log(NT)h}\to\infty\,,\ \text{and}\ l^2(NT)\log(NT)\sqrt{\frac{N}{T}}Th^2\to\infty\,,$$

and such $l(NT)$ exists by Assumption A.7(iii). So for some $\epsilon_* > 0$, we apply Theorem 1.3(b) in Bosq (1998) by assuming

$$\begin{aligned}&q = T/(2p),\quad p = \frac{1}{\epsilon_*l^2(NT)}\sqrt{\frac{T^{1-2/\delta}}{\log(NT)h}},\quad b = CN^{-1/2}T^{-1+1/\delta}h^{-2}l(NT) \\ &\epsilon = \epsilon_*T^{-1}l(NT)\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\ \text{and}\ \frac{2\sigma^2(q)}{p^2}+\frac{b\epsilon}{2}\le\frac{C}{NT^2h^3p}\,.\end{aligned}$$

Note that $\sigma^2(q) \le C\frac{p}{NT^2h^3}$. We have

$$\begin{aligned}P\left(\Theta_3 > \epsilon_*l(NT)\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\right) &\le C\delta_{NT}^{-2}\exp\left[-\frac{\epsilon_*^2l^2(NT)q\frac{\log(NT)}{\sqrt{NT}T^2h}}{\frac{C}{NT^2h^3p}}\right]+C\delta_{NT}^{-2}\left(1+\frac{4b}{\epsilon}\right)^{\frac{1}{2}}q\rho^{[\frac{T}{2q}]} \\ \le C\delta_{NT}^{-2}\exp\left[-\frac{\epsilon_*^2l^2(NT)\log(NT)Th^2}{2C}\sqrt{\frac{N}{T}}\right]&+C\delta_{NT}^{-2}\left(1+\frac{4b}{\epsilon}\right)^{\frac{1}{2}}q\rho^{[\frac{T}{2q}]}\,.\end{aligned}$$

Then, as $l^2(NT)\log(NT)Th^2\sqrt{\frac{N}{T}}\to\infty$ and the mixing coefficient decays with the exponential rate, the

desired result follows. ■

We will show J_{11} and J_{12} determine asymptotic mean and J_{13} determines asymptotic variance in Proposition A.4 and Proposition A.5 respectively.

Proposition A.4: Under the assumptions of Theorem 1, we have

$$J_{11}+J_{12}=h^{-1/2}\int_{-1}^{1}k^2(u)\,du\int_0^1\left[\hat{\sigma}_\varepsilon^2(\tau)+trace(\hat{\Sigma}_X^{-1}(\tau)\hat{\Sigma}_{X\varepsilon}(\tau))\right]d\tau+o_P(1).$$

Proposition A.5: Under the assumptions of Theorem 1, we have

$$J_{13}/S\xrightarrow{d}N(0,1),$$

where $S=4[\int_0^1\sigma_\varepsilon^4(\tau)\,d\tau+\int_0^1 trace(\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau)\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau))\,d\tau]\int(k*k)^2\,dv$ with $\int(k*k)^2\,dv=\int_0^1[\int_{-1}^1 k(u)k(u+v)\,du]^2dv$.

Proof of Proposition A.4: The result follows by Chebyshev's inequality, $var(J_{11})=O(\frac{1}{Th})$ and $var(J_{12})=O(\frac{1}{Th})$. To save space, the detailed proof is omitted. ■

Proof of Proposition A.5: Let

$$V_{s_1}=\frac{1}{Th^{1/2}}\sum_{s_2=1}^{s_1-1}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is_1}\varepsilon_{js_2}\left[\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}+\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js_2}\right],$$

which is a m.d.s., and $U_t=\sum_{s_1=2}^{t}V_{s_1}$. We have $J_{13}=2U_T$. We will apply the CLT for m.d.s. (Hall and Heyde, 1980) to get $\frac{U_T}{s_T}\xrightarrow{d}N(0,1)$, where $s_T^2=E(U_T^2)=\sum_{s_1=2}^{T}E(V_{s_1}^2)$. We need to show that

$$s_T^{-2}\sum_{s_1=2}^{T}V_{s_1}^2\xrightarrow{p}1 \tag{A9}$$

and

$$s_T^{-4}\sum_{s_1=2}^{T}E(V_{s_1}^4)\to 0\,. \tag{A10}$$

First, we compute the asymptotic variance of U_T. To save space, we define $\bar{k}_{st}=\frac{1}{Th}\sum_{\tau=1}^{T}k_{\tau s}k_{\tau t}$ and

$$\overline{\varepsilon v}_{st}=\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is}\varepsilon_{jt}\left[\frac{1}{Th}\sum_{\tau=1}^{T}k_{s\tau}k_{t\tau}+\frac{1}{Th}\sum_{\tau=1}^{T}k_{s\tau}k_{t\tau}v_{is}^{\intercal}\Sigma_X^{-1}(\frac{\tau}{T})v_{jt}\right].$$

We have

$$\begin{aligned}
s_T^2=&\frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s_2=1}^{s_1-1}E\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is_1}\varepsilon_{js_2}\left[\bar{k}_{s_1s_2}+\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js_2}\right]\right\}^2\\
&+\frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s=1}^{s_1-1}\sum_{s'\neq s}^{s_1-1}E\left\{\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is_1}\varepsilon_{js}\left[\bar{k}_{s_1s}+\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{st}v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js}\right]\right\}\right.\\
&\left.\times\left\{\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is_1}\varepsilon_{js'}\left[\bar{k}_{s_1s'}+\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s't}v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)v_{js'}\right]\right\}\right\}\\
=&B_1+B_2\,.
\end{aligned} \tag{A11}$$

For the first term in (A11), we have

$$B_1 = \frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s_2=1}^{s_1-1}\Bigg\{E\bigg(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\varepsilon_{js_2}\varepsilon_{j's_2}\bar{k}_{s_1s_2}^2\bigg)$$
$$+ E\bigg[\frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\varepsilon_{js_2}\varepsilon_{j's_2}\bar{k}_{s_1s_2}\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}v_{i's_1}^{\intercal}\Sigma_X^{-1}(\frac{t}{T})v_{j's_2}\bigg]$$
$$+ E\bigg[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{i'=1}^{N}\sum_{j'=1}^{N}\varepsilon_{is_1}\varepsilon_{js_2}\varepsilon_{i's_1}\varepsilon_{j's_2}\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}v_{is_1}^{\intercal}\Sigma_X^{-1}(\frac{t}{T})v_{js_2}\frac{1}{Th}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}v_{i's_1}^{\intercal}\Sigma_X^{-1}(\frac{t}{T})v_{j's_2}\bigg]\Bigg\}$$
$$= B_{11} + B_{12} + B_{13}\,. \tag{A12}$$

For the first term in (A12), we have

$$B_{11} = \frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s_2}^{s_1-1}E\bigg(\frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\bigg)E\bigg(\frac{1}{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\varepsilon_{js_2}\varepsilon_{j's_2}\bigg)\bar{k}_{s_1s_2}^2$$
$$+ \frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s_2}^{s_1-1}\bigg[E\bigg(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\sum_{j=1}^{N}\sum_{j'=1}^{N}\varepsilon_{js_2}\varepsilon_{j's_2}\bigg)$$
$$- E\bigg(\frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\bigg)E\bigg(\frac{1}{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\varepsilon_{js_2}\varepsilon_{j's_2}\bigg)\bigg]\bar{k}_{s_1s_2}^2$$
$$= B_{11,a} + B_{11,b}\,. \tag{A13}$$

By Lemma 1 in Yoshihara (1976), for the second term of (A13), we have

$$|B_{11,b}| \le \frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s_2=1}^{s_1-1}4C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(s_1-s_2)\bar{k}_{s_1s_2}^2 = O\bigg(\frac{1}{Th}\bigg),$$

where $C_M = \max(C_M^1, C_M^2)$, $C_M^1 = E|\frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\sum_{j=1}^{N}\sum_{j'=1}^{N}\varepsilon_{js_2}\varepsilon_{j's_2}|^{1+\eta} < \infty$ and $C_M^2 = E|\frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}|^{1+\eta}E|\frac{1}{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\varepsilon_{js_2}\varepsilon_{j's_2}|^{1+\eta} < \infty$ by Assumptions A.4. For the first term of (A13), we have $B_{11,a} = \int_0^1\sigma_\varepsilon^4(\tau)\,d\tau\int(k*k)^2\,dv + o(1)$. B_{12} and B_{13} can be decomposed in a similar way. It's easy to see that B_{12} is of small order and the leading term of B_{13} is

$$B_{13,a} = \frac{1}{T^2h}\sum_{s_1=2}^{T}\sum_{s_2=1}^{s_1-1}\Bigg\{\frac{1}{T^2h^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}k_{s_1t}k_{s_2t}k_{s_1t'}k_{s_2t'}$$
$$\times trace\left(E\left[\frac{1}{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\Sigma_X^{-1}\left(\frac{t}{T}\right)\varepsilon_{js_2}\varepsilon_{j's_2}v_{js_2}v_{j's_2}^{\intercal}\right]E\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{i'=1}^{N}\varepsilon_{is_1}\varepsilon_{i's_1}\Sigma_X^{-1}\left(\frac{t'}{T}\right)v_{i's_1}v_{is_1}^{\intercal}\right]\right)\Bigg\}$$
$$= \int_0^1 trace(\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau)\Sigma_X^{-1}(\tau)\Sigma_{X\varepsilon}(\tau))\,d\tau\int(k*k)^2\,dv + o(1).$$

Next, the second term in (A11) is $B_2 = \frac{2}{T^2h}\sum_{s_1=3}^{T}\sum_{s=2}^{s_1-1}\sum_{s'=1}^{s-1} E(\overline{\varepsilon v}_{s_1 s}\overline{\varepsilon v}_{s_1 s'})$. Suppose $s-s' > s_1 - s$, we have

$$\begin{aligned}
|B_2| &= \left|\frac{2}{T^2h}\sum_{s_1=3}^{T}\sum_{s=2}^{s_1-1}\sum_{s'=1}^{s-1} E(\overline{\varepsilon v}_{s_1 s}\overline{\varepsilon v}_{s_1 s'})\right| \\
&\leq \frac{2}{T^2h}\sum_{s'=1}^{T-2}\sum_{s=s'+1}^{T-1}\sum_{s_1=s+1}^{s+(s-s')} C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(s-s') \\
&\leq \frac{C}{T^2h}\sum_{s'=1}^{T-2}\sum_{l=1}^{T} l C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(l) = O\left(\frac{1}{Th}\right),
\end{aligned}$$

where $C_M = \max(C_M^1, C_M^2)$ and

$$\begin{aligned}
C_M^1 &= \sup_{s,s',s_1}\int |\overline{\varepsilon v}_{s_1 s}\overline{\varepsilon v}_{s_1 s'}|^{1+\eta}\, dF(\varepsilon_{s_1}, \varepsilon_s, \varepsilon_{s'}, v_{s'}, v_{s_1}, v_s) \\
&\leq \sup_{s,s',s_1} E[(\overline{\varepsilon v}_{s_1 s})^{2(1+\eta)}]^{\frac{1}{2}} E[(\overline{\varepsilon v}_{s_1 s'})^{2(1+\eta)}]^{\frac{1}{2}}, \\
C_M^2 &= \sup_{s,s',s_1}\int |\overline{\varepsilon v}_{s_1 s}\overline{\varepsilon v}_{s_1 s'}|^{1+\eta}\, dF(\varepsilon_{s'}, v_{s'}) dF(\varepsilon_s, v_s, \varepsilon_{s_1}, v_{s_1}).
\end{aligned}$$

By the definition of $\overline{\varepsilon v}_{st}$, it's easy to check $C_M^1, C_M^2 = O(1)$ by Minkowski's inequality and Assumptions A.3(i), A.4 and A.8(iii). The case where $s - s' < s_1 - s$ is similar.

Now we study $E(\sum_{s_1=2}^{T} V_{s_1}^2 - s_T^2)^2 = E(\sum_{s_1=2}^{T} V_{s_1}^2)^2 - s_T^4$, where the first term can be written as

$$\begin{aligned}
E(\sum_{s_1=2}^{T} V_{s_1}^2)^2 &= \sum_{s_1=2}^{T} E(V_{s_1}^4) + 2\sum\sum_{2\leq t'<t\leq T} E(V_t^2 V_{t'}^2) \\
&= \frac{1}{T^4h^2}\sum_{s_1=2}^{T}\sum_{s=1}^{s_1-1}\sum_{s'=1}^{s_1-1}\sum_{r=1}^{s_1-1}\sum_{r'=1}^{s_1-1} E(\overline{\varepsilon v}_{s_1 s}\overline{\varepsilon v}_{s_1 s'}\overline{\varepsilon v}_{s_1 r}\overline{\varepsilon v}_{s_1 r'}) \\
&\quad + \frac{2}{T^4h^2}\sum_{t=3}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{r=1}^{t'-1}\sum_{r'=1}^{t'-1} E(\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t'r}\overline{\varepsilon v}_{t'r'}) \\
&= C_1 + C_2 .
\end{aligned}$$

Decompose C_1,

$$\begin{aligned}
C_1 &= \frac{3}{T^4h^2}\sum_{s_1=2}^{T}\sum_{s=1}^{s_1-1}\sum_{s'=1}^{s_1-1} E(\overline{\varepsilon v}_{s_1 s}^2\overline{\varepsilon v}_{s_1 s'}^2) + \frac{6}{T^4h^2}\sum_{s_1=2}^{T}\sum_{s}^{s_1-1}\sum_{r}^{s_1-1}\sum_{r'\neq r\neq s}^{s_1-1} E(\overline{\varepsilon v}_{s_1 s}^2\overline{\varepsilon v}_{s_1 r}\overline{\varepsilon v}_{s_1 r'}) \\
&\quad + \frac{1}{T^4h^2}\sum_{s_1=2}^{T}\sum_{s}^{s_1-1}\sum_{s'}^{s_1-1}\sum_{r}^{s_1-1}\sum_{r'\neq r\neq s'\neq s}^{s_1-1} E(\overline{\varepsilon v}_{s_1 s}\overline{\varepsilon v}_{s_1 s'}\overline{\varepsilon v}_{s_1 r}\overline{\varepsilon v}_{s_1 r'}) \\
&\quad + \frac{1}{T^4h^2}\sum_{s_1=2}^{T}\sum_{s=1}^{s_1-1} E(\overline{\varepsilon v}_{s_1 s}^4) + \frac{4}{T^4h^2}\sum_{s_1=2}^{T}\sum_{s}^{s_1-1}\sum_{s'\neq s}^{s_1-1} E(\overline{\varepsilon v}_{s_1 s}^3\overline{\varepsilon v}_{s_1 s'}) \\
&= C_{11} + C_{12} + C_{13} + C_{14} + C_{15} .
\end{aligned}$$

We further decompose C_2,

$$
\begin{aligned}
C_2 = & \frac{2}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} E(\overline{\varepsilon v}_{ts}^2\overline{\varepsilon v}_{t's'}^2) + \frac{4}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} E(\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t's}\overline{\varepsilon v}_{t's'}) \\
& + \frac{4}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s}^{t-1}\sum_{s'}^{t-1}\sum_{r\neq s\neq s'}^{t'-1} E(\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t'r}^2) \\
& + \frac{8}{T^4h^2}\sum_{t=2}^{T}\sum_{t'}^{t-1}\sum_{s}^{t-1}\sum_{s'}^{t-1}\sum_{r\neq s\neq s'}^{t'-1} E(\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t's}\overline{\varepsilon v}_{t'r}) \\
& + \frac{2}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s}^{t-1}\sum_{s'}^{t-1}\sum_{r}^{t'-1}\sum_{r'\neq r\neq s\neq s'}^{t'-1} E(\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t'r}\overline{\varepsilon v}_{t'r'}) \\
& + \frac{2}{T^4h^2}\sum_{t=3}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t'-1} E(\overline{\varepsilon v}_{ts}^2\overline{\varepsilon v}_{t's}^2) + \frac{4}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s}^{t'-1}\sum_{s'\neq s}^{t'-1} E(\overline{\varepsilon v}_{ts}^2\overline{\varepsilon v}_{t's}\overline{\varepsilon v}_{t's'}) \\
& + \frac{4}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s}^{t-1}\sum_{s'\neq s}^{t'-1} E(\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t's'}^2) \\
= & \, C_{21} + C_{22} + C_{23} + C_{24} + C_{25} + C_{26} + C_{27} + C_{28}\,.
\end{aligned}
$$

We also have

$$
\begin{aligned}
s_T^4 = & \frac{2}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s}^{t-1}\sum_{s'\neq s}^{t'-1} E(\overline{\varepsilon v}_{ts})^2 E(\overline{\varepsilon v}_{t's'})^2 \\
& + \frac{2}{T^4h^2}\sum_{t=3}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t'-1} E(\overline{\varepsilon v}_{ts})^2 E(\overline{\varepsilon v}_{t's})^2 + o(1) \\
= & \, C_{31} + C_{32} + o(1)\,.
\end{aligned}
$$

For the rest of the proof, we will use the fact that for any $s_1 > s$, $\sup_{s_1,s} E(\overline{\varepsilon v}_{s_1s})^4 \leq C$. This is because

$$
\begin{aligned}
& \sup_{s_1,s} E(\overline{\varepsilon v}_{s_1s})^4 \\
= & \sup_{s_1,s} E\Bigg\{ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\Bigg[\varepsilon_{is_1}\varepsilon_{i's_1}\varepsilon_{js}\varepsilon_{j's}\bar{k}_{s_1s}^2 + 2\varepsilon_{is_1}\varepsilon_{i's_1}\varepsilon_{js}\varepsilon_{j's}\bar{k}_{s_1s}\frac{1}{Th}\sum_{t=1}^{T} k_{s_1t}k_{st} v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right) v_{js} \\
& + \varepsilon_{is_1}\varepsilon_{i's_1}\varepsilon_{js}\varepsilon_{j's}\frac{1}{Th}\sum_{t=1}^{T} k_{s_1t}k_{st} v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right) v_{js}\frac{1}{Th}\sum_{t=1}^{T} k_{s_1t}k_{st} v_{i's_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right) v_{j's}\Bigg]\Bigg\}^2 \\
\leq & \sup_{s_1,s}\Bigg\{ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\Bigg\{ \bar{k}_{s_1s}^2[E(\varepsilon_{is_1}\varepsilon_{i's_1})^2]^{1/2}[E(\varepsilon_{js}\varepsilon_{j's})^2]^{1/2} \\
& + 2\bar{k}_{s_1s}E[(\varepsilon_{is_1}\varepsilon_{i's_1})^2]^{1/2}\sup_t[E\|v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)\|^4]^{1/4}(E\|v_{js}\|^4)^{1/4}E[(\varepsilon_{js}\varepsilon_{j's})^4]^{1/4} + \bar{k}_{s_1s}^2[E(\varepsilon_{is_1}\varepsilon_{i's_1})^2]^{1/2} \\
& \times [E(\varepsilon_{js}\varepsilon_{j's})^8]^{1/8}(E\|v_{js}v_{j's}^{\intercal}\|^8)^{\frac{1}{8}}\sup_t(E\|v_{is_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)\|^4)^{\frac{1}{4}}\sup_t(E\|v_{i's_1}^{\intercal}\Sigma_X^{-1}\left(\frac{t}{T}\right)\|^8)^{\frac{1}{8}}\Bigg\}\Bigg\}^2 \\
\leq & \sup_{s_1s}\bar{k}_{s_1s}O(1) = O(1)\,.
\end{aligned}
$$

By Cauchy-Schwarz inequality, we have $\sup_{s_1,s} E(\overline{\varepsilon v}^2_{s_1 s}\overline{\varepsilon v}^2_{s_1 s'}) \leq C$. Then $C_{11} = O(\frac{1}{Th^2})$ follows. From the above, we also have for any $s_1 > s$, $E(\overline{\varepsilon v}_{s_1 s})^4 \leq C\bar{k}_{s_1 s}$. Similarly, we have $C_{12}, C_{22}, C_{27}, C_{28} = O(h)$, $C_{14} = O(\frac{1}{T^2h^2})$ and $C_{15}, C_{26}, C_{32} = O(\frac{1}{Th^2})$.

The derivations for C_{13} and $C_{23} - C_{25}$ are similar, and we only show $C_{25} = O(\frac{1}{Th^2})$ here to save space. Without loss of generality, we assume that $r' < r$ and $s' < s$. Denote $r' < r < s' < s < t' < t$ as case 1, and assume the largest distance between two adjacent points as d_1, then we have the following subcases:

Case 1a: $r - r' = d_1$, Case 1b: $s' - r = d_1$, Case 1c: $s - s' = d_1$,

Case 1d: $t' - s = d_1$, Case 1e: $t - t' = d_1$.

For Case 1a, we have

$$\begin{aligned} |C_{25}| &\leq \frac{1}{T^4h^2}\sum_{r'=1}^{T-5}\sum_{r=r'+1}^{T-4}\sum_{s'=r+1}^{T-3}\sum_{s=s'+1}^{T-2}\sum_{t'=s+1}^{T-1}\sum_{t=t'+1}^{T} C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(r-r') \\ &\leq \frac{C}{T^4h^2}\sum_{r'=1}^{T-5}\sum_{r=r'+1}^{T-4}(r-r')^4\beta^{\eta/(1+\eta)}(r-r') \\ &= O\bigg(\frac{1}{T^3h^2}\bigg), \end{aligned}$$

where $C_M = \max(C_M^1, C_M^2)$ is bounded. This is because

$$\begin{aligned} C_M^1 &= \sup_{t,t',s,s',r,r'} E(|\overline{\varepsilon v}_{ts}\overline{\varepsilon v}_{ts'}\overline{\varepsilon v}_{t'r}\overline{\varepsilon v}_{t'r'}|^{1+\eta}) \\ &\leq \sup_{t,s} E(\overline{\varepsilon v}_{ts}^{4(1+\eta)})^{\frac{1}{4}} \sup_{t,s'} E(\overline{\varepsilon v}_{ts'}^{4(1+\eta)})^{\frac{1}{4}} \sup_{t',r} E(\overline{\varepsilon v}_{t'r}^{4(1+\eta)})^{\frac{1}{4}} \sup_{t',r'} E(\overline{\varepsilon v}_{t'r'}^{4(1+\eta)})^{\frac{1}{4}}. \end{aligned}$$

Therefore, it's sufficient to show $\sup_{t,s} E(\overline{\varepsilon v}_{ts}^{4(1+\eta)})^{\frac{1}{4}} = O(1)$, which is similar to that of $\sup_{t,s} E(\overline{\varepsilon v}_{ts}^4) = O(1)$. For C_M^2, we can show it in a similar manner and apply the sub-multiplicative property of Frobenius norm (L^2 norm). For Case 1b, we have

$$\begin{aligned} |C_{25}| &\leq \frac{1}{T^4h^2}\sum_{r'=1}^{T-5}\sum_{r=r'+1}^{T-4}\sum_{s'=r+1}^{T-3}\sum_{s=s'+1}^{T-2}\sum_{t'=s+1}^{T-1}\sum_{t=t'+1}^{T} C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(s'-r) \\ &\leq \frac{C}{T^4h^2}\sum_{r'=1}^{T-5}\sum_{r=r'+1}^{T-4}\sum_{s'=r+1}^{T-3}(s'-r)^3\beta^{\eta/(1+\eta)}(s'-r) = O(\frac{1}{T^2h^2}). \end{aligned}$$

Following the same steps, for Case 1c, 1d and 1e, we have $|C_{25}| = O(\frac{1}{Th^2})$, $O(\frac{1}{T^2h^2})$ and $O(\frac{1}{T^3h^2})$ respectively. By a similar argument, we have C_{13}, C_{23} and C_{24} are at most $O(\frac{1}{T^2h^2})$. Finally, we have

$$C_{21} - C_{31} = \frac{2}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1}(E\overline{\varepsilon v}_{ts}^2\overline{\varepsilon v}_{t's'}^2 - E\overline{\varepsilon v}_{ts}^2 E\overline{\varepsilon v}_{t's'}^2).$$

There are three cases $s' < t' < s < t$, $s' < s < t' < t$ and $s < s' < t' < t$. When $s' < t' < s < t$, we have by Lemma 1 in Yoshihara (1976),

$$|E\overline{\varepsilon v}_{ts}^2\overline{\varepsilon v}_{t's'}^2 - E\overline{\varepsilon v}_{ts}^2 E\overline{\varepsilon v}_{t's'}^2| \leq C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(s-t').$$

Note that C_M is finite by a similar argument as in C_{25}. Then

$$|C_{21} - C_{31}| \leq \frac{2}{T^4h^2}\sum_{t=2}^{T}\sum_{t'=2}^{t-1}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} C_M^{1/(1+\eta)}\beta^{\eta/(1+\eta)}(s-t') = O\bigg(\frac{1}{Th^2}\bigg).$$

The other two cases are similar, but we need to use Lemma 1 in Yoshihara (1976) multiple times. To sum up, we have $E(\sum_{s_1=2}^{T} V_{s_1}^2 - s_T^2)^2 = o(1)$, so (A9) holds. Since $s_T^2 = O(1)$ and $C_1 = O(T^{-1}h^{-2} + h)$, (A10) holds. The desired result thus follows. ■■

Proof of Theorem A.2: Similar to J_1, we need the following uniform results.

Lemma A.3: Under the assumptions of Theorem 1, we have

$$\sup_{s\in[0,1]}\left\|\frac{1}{NTh}\sum_{t=1}^{T} k_{st}H_t\sum_{i=1}^{N} v_{it}g_t^{\intercal}\right\| = O_P\left(\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\right),$$

$$\sup_{s\in[0,1]}\left\|\frac{1}{NTh}\sum_{t=1}^{T} k_{st}H_t\sum_{i=1}^{N} v_{it}v_{it}^{\intercal}G_t - \frac{1}{Th}\sum_{t=1}^{T} k_{st}H_t\Sigma_X^{-1}\left(\frac{t}{T}\right)G_t\right\| = O_P\left(\sqrt{\frac{\log(NT)}{\sqrt{NT}h}}\right),$$

where $H_t = H(\frac{t}{T})$, $G_t = G(\frac{t}{T})$ and $g_t = g(\frac{t}{T})$ are defined in the same way as Lemma A.2. For instance, we can have $H_t = G_t = \Sigma_X^{-1}(\frac{t}{T})$ and $g_t = 1$.

Proof of Lemma A.3: The proof is similar to that of Lemma A.2, hence it is omitted here. ■

Then we can simplify J_2 as

$$J_2 = \sqrt{h}\sqrt{NT}(\hat{\theta}-\theta)\frac{1}{h\sqrt{NT^3}}\begin{bmatrix}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{is}k(\frac{s-t}{Th}) \\ \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\varepsilon_{is}X_{is}k(\frac{s-t}{Th})\end{bmatrix} + o_P(1) = O_P(\sqrt{h}),$$

where we have used Proposition 2, the CLT for m.d.s (Hall and Heyde, 1980), Assumptions A.1, A.2(i)(ii) and A.6. ■

Proof of Theorem A.3:

$$J_3 = N\sqrt{h}(\hat{\theta}-\theta)^{\intercal}\left(\sum_{t=1}^{T}\hat{\Omega}_t\right)(\hat{\theta}-\theta) = \sqrt{h}\sqrt{NT}(\hat{\theta}-\theta)^{\intercal}\frac{1}{T}\left(\sum_{t=1}^{T}\hat{\Omega}_t\right)\sqrt{NT}(\hat{\theta}-\theta) = o_P(1),$$

where we have used the fact that $\sqrt{NT}(\hat{\theta}-\theta) = O_P(1)$ and $\frac{1}{T}\sum_{t=1}^{T}\hat{\Omega}_t = O_P(1)$. ■■

Proof of Theorem 2: Under the alternative $\mathbb{H}_A$, we have

$$\begin{aligned}(NT)^{-1}h^{-1/2}\hat{Q} &= \frac{1}{T}\sum_{t=1}^{T}(\hat{\theta}_t-\theta_t+\theta_t-\hat{\theta})^{\intercal}\hat{\Omega}_t(\hat{\theta}_t-\theta_t+\theta_t-\hat{\theta}) \\ &= \frac{1}{T}\sum_{t=1}^{T}(\hat{\theta}_t-\theta_t)^{\intercal}\hat{\Omega}_t(\hat{\theta}_t-\theta_t) - \frac{2}{T}\sum_{t=1}^{T}(\hat{\theta}_t-\theta_t)^{\intercal}\hat{\Omega}_t(\hat{\theta}-\theta_t) + \frac{1}{T}\sum_{t=1}^{T}(\hat{\theta}-\theta_t)\hat{\Omega}_t(\hat{\theta}-\theta_t) \\ &= J_4 - 2J_5 + J_6\,.\end{aligned}$$

Under the alternative hypothesis, we have

$$\begin{aligned}\hat{\beta} &= A^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}^{\intercal}\beta_t - \frac{1}{T}\sum_{s=1}^{T}X_{is}^{\intercal}\beta_s) + A^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(\lambda_t - \frac{1}{T}\sum_{s=1}^{T}\lambda_s) \\ &\quad + A^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(\varepsilon_{it}-\bar{\varepsilon}_i) + o_P(1) \\ &= N_1 + N_2 + N_3 + o_P(1)\,,\end{aligned}$$

where we have used $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}-\bar{X}_i)^{\intercal} = A + o_P(1)$ from Proposition A.2, and A is defined in (A7).

For N_1, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(X_{it}^{\intercal}\beta_t-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\intercal}\beta_s)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}^{\intercal}\beta_t-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{X}_iX_{it}^{\intercal}\beta_t$$

$$=\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}X_{it}X_{it}^{\intercal}\beta_t-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}^{\intercal}\beta_t-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}X_{is}X_{it}^{\intercal}\beta_t$$

$$=\int_0^1[\Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)]\beta(\tau)\,d\tau-\int_0^1\mu_X(\tau)\,d\tau\int_0^1\mu_X^{\intercal}(\tau)\beta(\tau)\,d\tau+o_P(1),$$

which is obtained by the LLN for the β-mixing process (Lin and Lu, 1996) and Lemma 1 in Yoshihara (1976). Similarly, for N_2, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\bar{X}_i)(\lambda_t-\frac{1}{T}\sum_{t=1}^{T}\lambda_t)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\lambda_t-\frac{1}{NT}\sum_{i=1}^{N}\bar{X}_i\sum_{t=1}^{T}\lambda_t$$

$$=\int_0^1\mu_X(\tau)\lambda(\tau)\,d\tau-\int_0^1\mu_X(\tau)\,d\tau\int_0^1\lambda(\tau)\,d\tau+o_P(1)\,.$$

We already know that $\sqrt{NT}N_3=O_P(1)$ from Proposition 2. Hence, we have

$$\hat{\beta}-\beta^*=o_P\,(1)\,,$$

where

$$\begin{aligned}\beta^* &= A^{-1}\{\int_0^1[\Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)]\beta(\tau)\,d\tau-\int_0^1\mu_X(\tau)\,d\tau\int_0^1\mu_X^{\intercal}(\tau)\beta(\tau)\,d\tau\\ &\quad+\int_0^1\mu_X(\tau)\lambda(\tau)\,d\tau-\int_0^1\mu_X(\tau)\,d\tau\int_0^1\lambda(\tau)\,d\tau\}.\end{aligned}$$

For the intercept, we have

$$\hat{\lambda}=\bar{Y}-\bar{X}^{\intercal}\hat{\beta}=\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{it}^{\intercal}\beta_t-\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}X_{it}^{\intercal}\hat{\beta}+\frac{1}{T}\sum_{t=1}^{T}\lambda_t+\bar{\varepsilon}\,.$$

Similarly, we have $\hat{\lambda}-\lambda^*=o_p(1)$, where

$$\lambda^*=\int_0^1\mu_X^{\intercal}(\tau)\beta(\tau)\,d\tau-\int_0^1\mu_X^{\intercal}(\tau)\,d\tau\beta^*+\int_0^1\lambda(\tau)\,d\tau+o_P(1)\,.$$

Then J_6 becomes

$$J_6=\frac{1}{T}\sum_{t=1}^{T}\hat{\theta}^{\intercal}\hat{\Omega}_t\hat{\theta}-\frac{2}{T}\sum_{t=1}^{T}\hat{\theta}^{\intercal}\hat{\Omega}_t\theta_t+\frac{1}{T}\sum_{t=1}^{T}\theta_t^{\intercal}\hat{\Omega}_t\theta_t$$

$$=\theta^{*\intercal}\begin{bmatrix}1 & \int_0^1\mu_X^{\intercal}(\tau)\,d\tau\\ \int_0^1\mu_X(\tau)\,d\tau & \int_0^1[\Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)]\,d\tau\end{bmatrix}\theta^*-2\int_0^1\theta^{\intercal}(\tau)\begin{bmatrix}1 & \mu_X^{\intercal}(\tau)\\ \mu_X(\tau) & \Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)\end{bmatrix}d\tau\theta^*$$

$$+\int_0^1\theta^{\intercal}(\tau)\begin{bmatrix}1 & \mu_X^{\intercal}(\tau)\\ \mu_X(\tau) & \Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)\end{bmatrix}\theta(\tau)\,d\tau+o_p(1)\,,$$

where $\theta^* = [\lambda^* \ \beta^{*\intercal}]^\intercal$. We also have,

$$
\begin{aligned}
J_4 &= \frac{1}{T}\sum_{t=1}^{T}\Big[\frac{1}{Th}\sum_{s=1}^{T}k_{st}\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N}X_{is}^\intercal \\ \frac{1}{N}\sum_{i=1}^{N}X_{is} & \frac{1}{N}\sum_{i=1}^{N}X_{is}X_{is}^\intercal \end{bmatrix}(\theta_s-\theta_t)\Big]^\intercal \Phi_X^{-1}(\frac{t}{T})\hat{\Omega}_t\Phi_X^{-1}(\frac{t}{T}) \\
&\quad \times\Big[\frac{1}{Th}\sum_{s=1}^{T}k_{st}\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N}X_{is}^\intercal \\ \frac{1}{N}\sum_{i=1}^{N}X_{is} & \frac{1}{N}\sum_{i=1}^{N}X_{is}X_{is}^\intercal \end{bmatrix}(\theta_s-\theta_t)\Big] \\
&+ \frac{2}{T}\sum_{t=1}^{T}\Big[\frac{1}{Th}\sum_{s=1}^{T}k_{st}\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N}X_{is}^\intercal \\ \frac{1}{N}\sum_{i=1}^{N}X_{is} & \frac{1}{N}\sum_{i=1}^{N}X_{is}X_{is}^\intercal \end{bmatrix}(\theta_s-\theta_t)\Big]^\intercal \Phi_X^{-1}(\frac{t}{T})\hat{\Omega}_t\Phi_X^{-1}(\frac{t}{T})\Big(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{s=1}^{T}k_{st}\begin{bmatrix}1\\X_{is}\end{bmatrix}\varepsilon_{is}\Big) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{s=1}^{T}k_{st}\begin{bmatrix}1\\X_{is}\end{bmatrix}\varepsilon_{is}\Big)^\intercal \Phi_X^{-1}(\frac{t}{T})\hat{\Omega}_t\Phi_X^{-1}(\frac{t}{T})\Big(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{s=1}^{T}k_{st}\begin{bmatrix}1\\X_{is}\end{bmatrix}\varepsilon_{is}\Big)+o_P(1) \\
&= J_{41}+J_{42}+J_{43}+o_P(1)\,.
\end{aligned}
$$

By Assumption A.3(ii), A.4 and Lemma 1 in Yoshihara (1976), we have $\sup_t E[trace(\frac{1}{Th}\sum_{s=1}^{T}k_{st}\hat{\Omega}_s)^2]^{1/2} = O(1)$. Then, we know that

$$
E\|\frac{1}{Th}\sum_{s=1}^{T}k_{st}\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N}X_{is}^\intercal \\ \frac{1}{N}\sum_{i=1}^{N}X_{is} & \frac{1}{N}\sum_{i=1}^{N}X_{is}X_{is}^\intercal \end{bmatrix}(\theta_s-\theta_t)\| \le \sup_{|s-t|\le Th}\|\theta_s-\theta_t\|E[trace(\frac{1}{Th}\sum_{s=1}^{T}k_{st}\hat{\Omega}_s)^2]^{1/2}\,,
$$

which converges to 0 if t/T belongs to continuity points and C if t/T belongs to discontinuity points. As the number of discontinuity points is finite, we have $J_{41} = o_P(1)$. Similarly, we have $J_{42} = o_P(1)$. Following the proof of J_1 in Theorem A.1, we have $J_{43} = o_P(1)$. Using a similar argument, we have the cross term $J_5 = o_P(1)$. Moreover, $(NT)^{-1}h^{-1/2}\hat{C} = o(1)$ and $\hat{S} = O(1)$. Therefore, we have $P(\hat{H} > C_T) \to 1$ for any $C_T = o(NT\sqrt{h})$. ■

Proof of Theorem 3: Similarly to Theorem 1, we have

$$
\begin{aligned}
\hat{Q}_1 &= N\sqrt{h}\sum_{t=1}^{T}(\hat{\beta}_t-\beta+\beta-\hat{\beta}_P)^\intercal\hat{M}_t(\hat{\beta}_t-\beta+\beta-\hat{\beta}_P) \\
&= N\sqrt{h}\sum_{t=1}^{T}(\hat{\beta}_t-\beta)^\intercal\hat{M}_t(\hat{\beta}_t-\beta) - 2N\sqrt{h}\sum_{t=1}^{T}(\hat{\beta}_t-\beta)^\intercal\hat{M}_t(\hat{\beta}_P-\beta) + N\sqrt{h}(\hat{\beta}_P-\beta)^\intercal\hat{M}_t(\hat{\beta}_P-\beta)] \\
&= L_1 - 2L_2 + L_3\,.
\end{aligned}
$$

We can also show that $L_1 = O_P(1)$ and $L_2\,, L_3 = O_P(\sqrt{h})$. We expand L_1 and apply Lemma A.2,

$$
\begin{aligned}
L_1 &= N\sqrt{h}\sum_{t=1}^{T}\left\{[\bar{0}_d\ I_d][I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)\varepsilon\right\}^{\intercal} \\
&\quad \times\hat{M}_t\left\{[\bar{0}_d\ I_d][I_{d+1}\ 0_{d+1}][M^{\intercal}(\tau)\mathcal{W}(\tau)M(\tau)]^{-1}M^{\intercal}(\tau)\mathcal{W}(\tau)\varepsilon\right\} \\
&= \frac{1}{NT^2h^{3/2}}\sum_{t=1}^{T}\left[\begin{array}{cc} \sum_{s=1}^{T}\sum_{i=1}^{N}k_{st}\varepsilon_{is} & \sum_{s=1}^{T}\sum_{i=1}^{N}k_{st}\varepsilon_{is}X_{is}^{\intercal}\end{array}\right] \\
&\quad \times\begin{bmatrix}\mu_X^{\intercal}(\frac{t}{T})\Sigma_X^{-1}(\frac{t}{T})\hat{M}_t\Sigma_X^{-1}(\frac{t}{T})\mu_X(\frac{t}{T}) & -\mu_X^{\intercal}(\frac{t}{T})\Sigma_X^{-1}(\frac{t}{T})\hat{M}_t\Sigma_X^{-1}(\frac{t}{T}) \\ -\Sigma_X^{-1}(\frac{t}{T})\hat{M}_t\Sigma_X^{-1}(\frac{t}{T})\mu_X(\frac{t}{T}) & \Sigma_X^{-1}(\frac{t}{T})\hat{M}_t\Sigma_X^{-1}(\frac{t}{T})\end{bmatrix}\left[\begin{array}{c}\sum_{s=1}^{T}\sum_{i=1}^{N}k_{st}\varepsilon_{is} \\ \sum_{s=1}^{T}\sum_{i=1}^{N}k_{st}\varepsilon_{is}X_{is}\end{array}\right] + o_P(1) \\
&= \frac{1}{NT^2h^{3/2}}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is_1}\varepsilon_{js_2}\Bigg\{v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\Sigma_X^{-1}+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right)v_{js_2} \\
&+ \sum_{t=1}^{T}k_{s_1t}k_{s_2t}[\mu_X^{\intercal}\Sigma_X^{-1}\mu_X+(\mu_X^{\intercal}\Sigma_X^{-1}\mu_X)^2]\left(\frac{t}{T}\right)-\mu_X^{\intercal}\left(\frac{s_1}{T}\right)\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\Sigma_X^{-1}\mu_X+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1}\mu_X)\left(\frac{t}{T}\right) \\
&- \sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\mu_X^{\intercal}\Sigma_X^{-1}+\mu_X^{\intercal}\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right)\mu_X\left(\frac{s_2}{T}\right)+\mu_X^{\intercal}\left(\frac{s_1}{T}\right)\sum_{t=1}^{T}k_{s_1t}k_{s_2t} \\
&\quad \times(\Sigma_X^{-1}+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right)\mu_X\left(\frac{s_2}{T}\right)-v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\Sigma_X^{-1}\mu_X+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1}\mu_X)\left(\frac{t}{T}\right) \\
&+ v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\Sigma_X^{-1}+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right)\mu_X\left(\frac{s_2}{T}\right)-\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\mu_X^{\intercal}\Sigma_X^{-1}+\mu_X^{\intercal}\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right) \\
&\quad \times v_{js_2}+\mu_X^{\intercal}\left(\frac{s_1}{T}\right)\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\Sigma_X^{-1}+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right)v_{js_2}\Bigg\}+o_P(1) \\
&= \frac{1}{NT^2h^{3/2}}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{is_1}\varepsilon_{js_2}v_{is_1}^{\intercal}\sum_{t=1}^{T}k_{s_1t}k_{s_2t}(\Sigma_X^{-1}+\Sigma_X^{-1}\mu_X\mu_X^{\intercal}\Sigma_X^{-1})\left(\frac{t}{T}\right)v_{js_2}+o_P(1)\,,
\end{aligned}
$$

which is very similar to J_1 in the proof of Theorem 1. As we have $\sqrt{NT}(\hat{\beta}_P-\beta)=O_P(1)$ by Proposition 3, the rest of the proof is similar to that of Theorem 1 and hence is omitted here. ■

Proof of Theorem 4: Under the alternative $\mathbb{H}_A^*$, we have

$$
\begin{aligned}
(TN)^{-1}h^{-1/2}\hat{Q}_1 &= \frac{1}{T}[\sum_{t=1}^{T}(\hat{\beta}_t-\beta_t)^{\intercal}\hat{M}_t(\hat{\beta}_t-\beta_t)-2\sum_{t=1}^{T}(\hat{\beta}_t-\beta_t)^{\intercal}\hat{M}_t(\hat{\beta}_P-\beta_t)+\sum_{t=1}^{T}(\hat{\beta}_P-\beta_t)^{\intercal}\hat{M}_t(\hat{\beta}_P-\beta_t)] \\
&= L_4-2L_5+L_6\,.
\end{aligned}
$$

Similar to the proof of Proposition 3, we have $\hat{\beta}_P-\beta_P^*=o_p(1)$, where

$$
\beta_P^*=(\int_0^1\Sigma_X(\tau)\,d\tau)^{-1}\int_0^1\Sigma_X(\tau)\beta(\tau)\,d\tau\,.
$$

Then, similar to the proof of Theorem 2, it's clear to see that

$$
\begin{aligned}
L_6 &= \int_0^1\beta^{\intercal}(\tau)[\Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)]\beta(\tau)\,d\tau+\beta_P^{*\intercal}\int_0^1[\Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)]\,d\tau\beta_P^* \\
&\quad -2\int_0^1\beta^{\intercal}(\tau)[\Sigma_X(\tau)+\mu_X(\tau)\mu_X^{\intercal}(\tau)]\,d\tau\beta_P^*+o_P(1)\,.
\end{aligned}
$$

Since we have $\hat{\beta}_t - \beta_t = [\bar{0}_d \; I_d](\hat{\theta}_t - \theta_t)$, we rewrite L_4 as

$$
\begin{aligned}
L_4 = \frac{1}{T}\sum_{t=1}^{T}\Bigg[\frac{1}{Th}\sum_{s=1}^{T} k_{st}
\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N} X_{is}^{\intercal} \\ \frac{1}{N}\sum_{i=1}^{N} X_{is} & \frac{1}{N}\sum_{i=1}^{N} X_{is}X_{is}^{\intercal} \end{bmatrix} (\theta_s - \theta_t)\Bigg]^{\intercal} \Phi_X^{-1}(\frac{t}{T}) \begin{bmatrix} 0 & \bar{0}_d^{\intercal} \\ \bar{0}_d & I_d \end{bmatrix} \hat{\Omega}_t \\
\times \begin{bmatrix} 0 & \bar{0}_d^{\intercal} \\ \bar{0}_d & I_d \end{bmatrix} \Phi_X^{-1}(\frac{t}{T})\Bigg[\frac{1}{Th}\sum_{s=1}^{T} k_{st}
\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N} X_{is}^{\intercal} \\ \frac{1}{N}\sum_{i=1}^{N} X_{is} & \frac{1}{N}\sum_{i=1}^{N} X_{is}X_{is}^{\intercal} \end{bmatrix} (\theta_s - \theta_t)\Bigg] \\
+ \frac{2}{T}\sum_{t=1}^{T}\Bigg[\frac{1}{Th}\sum_{s=1}^{T} k_{st}
\begin{bmatrix} 1 & \frac{1}{N}\sum_{i=1}^{N} X_{is}^{\intercal} \\ \frac{1}{N}\sum_{i=1}^{N} X_{is} & \frac{1}{N}\sum_{i=1}^{N} X_{is}X_{is}^{\intercal} \end{bmatrix} (\theta_s - \theta_t)\Bigg]^{\intercal} \Phi_X^{-1}(\frac{t}{T}) \begin{bmatrix} 0 & \bar{0}_d^{\intercal} \\ \bar{0}_d & I_d \end{bmatrix} \hat{\Omega}_t \\
\times \begin{bmatrix} 0 & \bar{0}_d^{\intercal} \\ \bar{0}_d & I_d \end{bmatrix} \Phi_X^{-1}(\frac{t}{T})\Bigg(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{s=1}^{T} k_{st}
\begin{bmatrix} 1 \\ X_{is} \end{bmatrix} \varepsilon_{is}\Bigg) \\
+ \frac{1}{T}\sum_{t=1}^{T}\Bigg(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{s=1}^{T} k_{st}
\begin{bmatrix} 1 \\ X_{is} \end{bmatrix} \varepsilon_{is}\Bigg)^{\intercal} \Phi_X^{-1}(\frac{t}{T}) \begin{bmatrix} 0 & \bar{0}_d^{\intercal} \\ \bar{0}_d & I_d \end{bmatrix} \hat{\Omega}_t \begin{bmatrix} 0 & \bar{0}_d^{\intercal} \\ \bar{0}_d & I_d \end{bmatrix} \Phi_X^{-1}(\frac{t}{T}) \\
\times \Bigg(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{s=1}^{T} k_{st}
\begin{bmatrix} 1 \\ X_{is} \end{bmatrix} \varepsilon_{is}\Bigg) + o_P(1) \\
= L_{41} + L_{42} + L_{43} + o_P(1)\,.
\end{aligned}
$$

We can expand L_5 in a similar way. Then the rest of the proof follows closely to that of Theorem 2 and hence is omitted. ■