Functional Coefficient Panel Modeling with Communal Smoothing Covariates*

Peter C. B. Phillips[a,b,c,d], Ying Wang[e]

[a]Yale University, [b]The University of Auckland,
[c]University of Southampton, [d]Singapore Management University,
[e]Renmin University of China

February 24, 2021

Abstract

Behavior at the individual level in panels is often influenced by aspects of the system in aggregate. In particular, the interaction between individual-specific explanatory variables and an individual dependent variable may be affected by 'global' variables that are relevant in decision making and shared communally by all individuals in the sample. To capture such behavioral features, we employ a functional coefficient panel model in which certain communal covariates may jointly influence panel interactions by means of their impact on the model coefficients. Two classes of estimation procedures are proposed, one based on cross section averaged data, the other on the full panel. The asymptotic properties of these methods are obtained and compared, allowing for sequential and joint expansion of the cross section and time series sample sizes. Limit theory for the associated fixed effects estimators are derived and inferential procedures are developed to test hypotheses concerning the functional coefficients. The finite sample performance of the proposed estimators and tests are examined by simulation. An empirical illustration is provided in which the regional sensitivity of housing rental prices to available job numbers is studied with national labor force participation rate as the communal smoothing covariate. Strong evidence is found supporting the functional coefficient specification with this country-wide smoothing variable.

JEL classification: C14, C23

Keywords: Communal covariates, Fixed effects, Functional coefficients, Housing rental prices, Panel data, Tests of constancy.

---

*The authors thank the Managing Editor, Serena Ng, an Associate Editor, and two referees for many helpful comments and suggestions on earlier versions of the paper. Research support is acknowledged from Marsden Grant 16-UOA-239 at the University of Auckland. Phillips acknowledges support from the Kelly Fund at the University of Auckland, an LKC Fellowship at Singapore Management University, and the NSF under Grant No. SES 18-50860. Peter C. B. Phillips email: peter.phillips@yale.edu; Ying Wang email: wangyingstat@gmail.com.

1 Introduction

Decisions taken at the individual consumer or firm level are frequently affected by prevailing macroeconomic influences such as interest rates, inflation, and aggregate indices of consumer or business sentiment. Likewise, changes in the labor market at the national level can play a role in influencing the nature and intensity of the impact of local workforce changes on the demand for local housing. Similarly, in spatial modeling it is often appropriate to model behavior at individual locations partly in terms of aggregate influences. In modeling climatic change, for instance, average temperature in any given spatial location needs to account for prevailing aggregates such as greenhouse gas concentrations in the atmosphere because of the way such gases are well-mixed in Earth's atmosphere as a whole.

One mechanism by which such individual or local dependencies on aggregates may be modeled in practical work is to use a panel framework in which the coefficients are functionally determined by the relevant aggregate or 'communal' variables. This type of model is closely related to a fixed effects functional coefficient panel data model of the following form

$$y_{it} = \alpha_i + \beta(z_{it})' x_{it} + u_{it}, \quad i = 1, \cdots, N; t = 1, \cdots, T; \tag{1.1}$$

where x_{it} is a p-vector of regressors, z_{it} is a q-vector of covariates that determine the (random) coefficients $\beta(z_{it}) = (\beta_1(z_{it}), \cdots, \beta_p(z_{it}))'$, the α_i are individual fixed effects, and the error u_{it} has zero mean and finite variance σ_u^2. In what follows, we will focus on the case where both x_{it} and z_{it} are exogenous.[1]

Methods of econometric estimation and inference in model (1.1) are reviewed in Su and Ullah (2011). Regarding estimation, the usual differencing method to eliminate fixed effects can be extended to this functional coefficient model. But as indicated in Sun et al. (2009), this approach leads to additive nonparametric components and therefore suffers from the problems of estimating nonparametric additive models as well as the additional complexity of the presence of common functional coefficients in the resulting additive nonparametric regression. Instead, Sun et al. (2009) proposed a profile least squares approach in which the nonparametric component $\beta(\cdot)$ is profiled out first. In later work Su and Ullah (2011) proposed an alternative profile least squares method in which the fixed effects α_i rather than $\beta(\cdot)$ are profiled out first. Both approaches may be employed in the communal panel framework (1.2) that we consider in the present paper. Rodriguez-Poo and Soberón (2015) presented an estimation procedure that employs a within un-smoothed mean deviation transformation of (1.1). Recently, Feng et al. (2017) considered varying-coefficient categorical panel data models where the z_{it} are discrete

[1] When the exogeneity condition $\mathbb{E}(u_{it}|x_{it}) = 0$ fails and there are endogenous regressors, model (1.1) has been examined by Cai and Li (2008) but without fixed effects α_i. These authors proposed a nonparametric GMM estimation method. To our knowledge, models with endogenous covariates z_{it} for which $\mathbb{E}(u_{it}|z_{it}) \neq 0$ have so far only been analyzed by Li and Sun (2019).

covariates.

Our interest in this paper lies in a communal version of the model (1.1). Instead of using individual specific variates z_{it} as the smoothing covariate for the regression coefficients, our framework employs smoothing covariates z_t that are common to all individuals in the panel. This formulation allows for global influences in determining the impact of the individual regressors and intercept. More specifically, we consider the following model

$$y_{it} = \alpha_i + \beta_0(z_t) + x_{it}'\beta(z_t) + u_{it} = \alpha_i + x_{*,it}'\beta_*(z_t) + u_{it}, \tag{1.2}$$

where $x_{*,it} = (1, x_{it}')'$ and $\beta_*(z) = (\beta_0(z), \beta(z)')'$ is a $(p+1)$-vector of coefficients. We allow the individual specific effects α_i to be correlated with z_t and/or x_{it} with an unknown correlation structure, so that (1.2) is treated as a fixed effects model. For identification, we assume that $\sum_{i=1}^N \alpha_i = 0$[2]. We include the intercept $\beta_0(z_t)$ explicitly. Then (1.2) includes the model studied by Lee and Robinson (2015). Thus, when there is no explanatory variable x_{it}, (1.2) reduces to the nonparametric panel data model with fixed effects of Lee and Robinson (2015). For the case without the intercept $\beta_0(z_t)$, analysis can be carried out in a similar fashion and the results are collected together in the Online Supplement to this paper (Phillips and Wang, 2020a).

As indicated at the outset, a primary motivation underlying the specification of (1.2) lies in the fact that z_t may represent global variables that are shared as common influences by all individuals in the panel or all locations in the spatial model. For example, z_t could be some world-wide variables in a panel cross-country study or nation-wide variables in panel cross-state or regional analyses. Similarly, z_t may include certain global variables that are relevant in determining station-level outcomes in a spatial model, as in a model of Earth's climate.[3] The time-varying coefficient panel data model studied by Li et al. (2011) reflects similar considerations in which the parameters may evolve over time. But instead of using covariates such as z_t to drive this evolution, the coefficients are assumed to be directly time-varying. In a similar fashion the model employed in Robinson (2012) considers a nonparametric trending regression where only an intercept function of time is included.

Our first contribution is to provide an analytic study of econometric estimation and inference in the model (1.2). We consider two estimation approaches, explore their respective asymptotic properties, and develop tests to assess constancy of the functional coefficients. One approach uses a cross-section averaged version of (1.2) and the other works directly with the full panel structure. Differences in the asymptotic behavior of the functional coefficient estimators, in-

[2] This is a standard assumption in the literature. A referee pointed out an alternative identification condition $\mathbb{E}\beta_0(z_t) = 0$ for all t. With this condition, estimation can be carried out in similar ways except that a demeaning transformation is needed to ensure the estimator of $\beta_0(z)$ satisfies the zero expectation condition. The assumption $\sum_{i=1}^N \alpha_i = 0$ has no meaningful impact on the test of intercept homogeneity, namely $\alpha_i = \alpha_0$ for all i where α_0 is either given or an unknown constant that needs to be estimated.

[3] A spatial climate econometric model of this type was used as an empirical illustration in the original version of this paper (Phillips and Wang, 2019) with global atmospheric CO_2 equivalent as the communal variate.

cluding convergence rates, of these approaches and their relation to the properties of an oracle estimator are examined. The analysis contributes to multi-index limit theory by developing sequential and joint limits as the sample sizes $(N, T) \to \infty$ in cases where are different convergence rates and potential degeneracies in the asymptotics. The latter present difficulties that require new methods in developing a rigorous joint limit theory. A further contribution of the paper is to provide limit theory for the associated fixed effects estimators, which is new to the literature.

The limit theory is used to construct tests of parametric linear specification against the semiparametric functional coefficient model. Numerical work is conducted to examine the finite sample properties of the estimation and test procedures. A real data analysis is provided to study the sensitivity of housing rental prices to labor market supply in the United Kingdom with regional data. The national labor force participation rate is used as the smoothing communal variate. Strong empirical evidence is found in this application to support a mediating impact of the labor force participation rate on the relationship between regional housing rental prices and regional job numbers. This functional coefficient specification with a communal smoothing variable is also shown to provide improved within sample and one-period ahead forecasting performance.

The remainder of the paper is organized as follows. Section 2 presents the two estimation approaches and derives their asymptotic properties. Testing constancy of the functional coefficients is considered in Section 3. Simulations are conducted in Section 4 to examine the finite sample performance of the two approaches and the test statistics. Section 5 is devoted to the empirical analysis of housing rental prices sensitivity to the number of jobs available in the region and the national level labor force participation rate. Section 6 concludes the paper. The Online Supplement gives further results for the panel model (1.2) with fixed effects and functional coefficients but without an intercept coefficient function.

2 Estimation and asymptotic theory

This section is devoted to the estimation of model (1.2) and the development of asymptotic theory for the proposed estimation procedures. Before presenting these procedures, we discuss the effects of the within and differencing transformations commonly employed to deal with the fixed effects α_i.

Taking a time series average of (1.2) gives

$$y_{iA} = \alpha_i + \frac{1}{T}\sum_{t=1}^{T} x'_{*,it}\beta_*(z_t) + u_{iA}, \quad i = 1, \cdots, N, \tag{2.1}$$

where $y_{iA} = T^{-1}\sum_{t=1}^{T} y_{it}$, and u_{iA} is defined analogously. The within transformation of (1.2) then yields the system

$$\begin{aligned} y_{it} - y_{iA} &= x'_{*,it}\beta_*(z_t) - \frac{1}{T}\sum_{s=1}^{T} x'_{*,is}\beta_*(z_s) + u_{it} - u_{iA} \\ &= \sum_{s=1}^{T} \delta_{ts} x'_{*,is}\beta_*(z_s) + u_{it} - u_{iA}, \quad i = 1, \cdots, N, t = 1, \cdots, T, \end{aligned} \tag{2.2}$$

where $\delta_{ts} = 1 - 1/T$ if $s = t$ and $-1/T$ otherwise. The right-hand side of (2.2) involves a linear combination of $x'_{*,is}\beta_*(z_s)$ of quantities measured at all time periods including time t. Marginal integration methods can be used to estimate $\beta_*(\cdot)$, as in the estimation of nonparametric additive models.

An alternative way to remove fixed effects is to use first differences of (1.2) or to long-difference by deducting the equation at time period 1. This approach again leads to a model that contains a linear combination of $x'_{*,it}\beta_*(z_t)$ at different times t.

Both transformation methods therefore suffer from difficulties similar to those that arise in the estimation of nonparametric additive models. A further difficulty is that if x_{it} contains a time-invariant term whose coefficient has an additive constant, then first order differencing wipes out the additive constant. In consequence, the coefficient cannot be consistently estimated; see Sun et al. (2009) for more discussion.

In view of these difficulties, we adopt a profile method to remove the unknown fixed effects. Profile least squares estimation procedures for model (1.1) were proposed by Sun et al. (2009) and Su and Ullah (2011). These methods can be applied in our model (1.2). We adopt the approach of Su and Ullah (2011), which profiles out the fixed effects first. In what follows, we consider two types of local constant nonparametric estimates. A cross-section averaged profile local constant (APLC) estimation method is considered in Section 2.1. Section 2.2 discusses the profile local constant (PLC) approach.

2.1 APLC estimation

Averaging over i in (1.2) gives, using the setting $\bar{\alpha} = 0$ for identification,

$$y_{At} = \beta_0(z_t) + x'_{At}\beta(z_t) + u_{At} = x'_{*,At}\beta_*(z_t) + u_{At}, t = 1, \cdots, T. \tag{2.3}$$

Let $Y_A = (y_{A1}, ..., y_{AT})'$ be the $T \times 1$ vector, and X_A^* be the $T \times (p+1)$ matrix obtained by stacking the $1 \times (p+1)$ vector $x'_{*,At}$. Using the local level approach to estimate $\beta_*(z)$ yields

$$\hat{\beta}_{*,APLC}(z) = [(X_A^*)'K_T(z)X_A^*]^{-1}(X_A^*)'K_T(z)Y_A, \tag{2.4}$$

where $K_T(z)$ is a $T \times T$ diagonal matrix with t-th central element $K_{tH} = K(H^{-1}(z_t - z))$ and $K(\cdot)$ is a multivariate kernel function. For the diagonal bandwidth matrix $H = diag(h_1, \cdots, h_q)$, define $||H|| = \sqrt{\sum_{j=1}^q h_j^2}$ and $|H| = h_1 \cdots h_q$. We adopt the product kernel $K(v) = \prod_{j=1}^q k(v_j)$, with $v = (v_1, \cdots, v_q)'$.

To establish the asymptotic properties of $\hat{\beta}_{*,APLC}(z)$, we employ the following conditions.

Assumption 1. *The kernel function $k(\cdot)$ is a symmetric bounded probability function with support $[-1, 1]$, $\int k(w)dw = 1$, and $\int wk(w)dw = 0$. Denote $\int w^2k(w)dw = \mu_2$, $\int k^2(w)dw = \nu_0$ and $\int w^2k^2(w)dw = \nu_2$.*

Assumption 2. *(a) $\{(x_i, u_i), i \geq 1\}$ is a sequence of independent and identically distributed (i.i.d.) variates over i, where $x_i = (x_{it}, t \geq 1)$ and $u_i = (u_{it}, t \geq 1)$. Further, for each $i \geq 1$, $\{(x_{it}, z_t, u_{it}), t \geq 1\}$ is stationary and α-mixing with mixing coefficients α_k satisfying $\alpha_k = O(k^{-\tau})$, where $\tau > \frac{\lambda+2}{\lambda}$ for some $\lambda > 0$, as in (b) and (c) below. Furthermore, u_{it} is independent of x_{it} and z_t for all i and t;*

(b) Let $\mathbb{E}(x_{it}|z_t = z) = \eta(z) = \eta$, $\mathbb{E}(x_{it}x_{it}'|z_t = z) = V_{xx}(z) = V_{xx}$, and $\mathbb{V}ar(x_{it}|z_t = z) = \Sigma_{xx}(z) = \Sigma_{xx} = V_{xx} - \eta\eta'$ be positive definite, so that the first and second order conditional moments of x_{it} given $z_t = z$ are independent of z. Furthermore, $\mathbb{E}(||x_{it}||^{2(2+\lambda)}) < C < \infty$, where $||\cdot||$ is Euclidean distance;

(c) The error process $\{u_{it}\}$ satisfies $\mathbb{E}u_{it} = 0$, $\mathbb{E}u_{it}^2 = \sigma_u^2 < \infty$, and $\mathbb{E}|u_{it}|^{2+\lambda} < \infty$ for some $\lambda > 0$;

(d) z_t has bounded probability density $f_z(z)$ and $f_z(z) > 0$. The functions $f_z(\cdot)$ and $\beta_*(\cdot)$ have bounded continuous derivatives to the second order.

Assumption 3. *As $T \to \infty$, $||H|| \to 0$, $T|H| \to \infty$, and $NT|H| \to \infty$.*

Remark 2.1. *(i) The conditions on the kernel function $k(\cdot)$ in Assumption 1 are commonly used for convenience in proofs and can be relaxed. For example, the compact support condition can be replaced with some restrictions on the tail behavior of the kernel function, which can be useful when small bandwidth choices are employed. The product kernel $K(v) = \prod_{j=1}^q k(v_j)$ is standard in multivariate smoothing and satisfies $\int vv'K(v)dv = \mu_2 I_q$, $\int K^2(v)dv = \nu_0^q$, and $\int vv'K^2(v)dv = \nu_2 I_q$.*
(ii) Assumption 2 (a) assumes $\{(x_{it}, u_{it}), i \geq 1\}$ is iid across section but allows for temporal dependence under mixing conditions to facilitate the asymptotic theory. Similar assumptions are used by Li et al. (2011). Endogeneity is ruled out in (a) for simplicity in the present study, as in Sun et al. (2009), but may be accommodated using instrumental variable methods. The first part of the present paper works under these assumptions. However, the case of nonstationary regressors x_{it} is relevant in many applications, including climate change studies of the type that

were considered in the empirical work of the original version of the present paper (Phillips and Wang, 2019). Such extensions obviously require some considerable modification to the conditions and proofs. They are an important line of further research given the prevalence of nonstationarity in many economic, financial and climate time series.

(iii) Assumption 2 (b) assumes that the first two conditional moments of x_{it} given $z_t = z$ are equivalent to unconditional moments, which obviously holds under independence of x_{it} and z_t. This requirement is not crucial to our findings and can be relaxed at the cost of more complex calculations and notation[4]*. In contrast to Li et al. (2011) we allow x_{it} to have a non-zero mean η in Assumption 2 (b). This extension has important consequences and is relevant in practical work. In the asymptotic distributions presented below, we will see the role that a non-zero (conditional) mean η plays. Denoting $x_{it} = x_{it}^0 + \eta$, where $\mathbb{E}(x_{it}^0|z_t = z) = 0$, model (1.2) can be rewritten as $y_{it} = \alpha_i + \beta_0(z_t) + \eta'\beta(z_t) + (x_{it}^0)'\beta(z_t) + u_{it}$, where $\beta_0(z_t) + \eta'\beta(z_t) \equiv \beta_0^*(z_t)$ is a composite intercept. As shown later in Remark 2.7, the estimator of the composite intercept has a faster convergence rate than that of other linear combinations of $\beta_*(z)$. From this perspective, Theorem 2.1 of Li et al. (2011) is nested as a special case of Theorem 2.2 below.*

(iv) Assumption 2 (c) provides standard moment conditions on the equation error u_{it}.

(v) Assumptions 2 (d) and 3 are standard regularity conditions on smoothness of the functional coefficients, the density of z_t, and the bandwidth and effective sample size $T|H| \to \infty$ requirements that are used in kernel estimation.

The following result provides asymptotic theory for the estimator $\hat{\beta}_{*,APLC}(z)$ in various settings depending on whether N is fixed or $N \to \infty$ and whether $\eta = 0$ or $\eta \neq 0$. In all cases, it is presumed that $T \to \infty$[5]. Results with $N \to \infty$ include both sequential limit $(N, T)_{\text{seq}} \to \infty$ theory, where $T \to \infty$ followed by $N \to \infty$, and joint limit $(N, T) \to \infty$ theory, where T, N pass to infinity together. Joint limit theory is obtained by following the double indexed limit theory in Phillips and Moon (1999, theorem 2) but our results provide an important extension that covers cases of multiple convergence rates and possibly degenerate limit distributions. In addition, a re-labeling technique is used when dealing with triple index sequences so that the double index joint limit theory can be employed.

Theorem 2.1. *Under Assumptions 1-3, as $T \to \infty$, we have:*

[4] If the first and second moments of the regressor variables x_{it} given $z_t = z$ depend on z, we need to replace the occurrence of η with $\eta(z)$, V_{xx} with $V_{xx}(z)$ and Σ_{xx} with $\Sigma_{xx}(z)$. Those adjustments will largely change the notations in this paper. The transformation C_η introduced later in Theorem 2.2 will depend on the estimation point z. Upon appropriate requirements on $\eta(z)$, $V_{xx}(z)$ and $\Sigma_{xx}(z)$ and associated conditions that assure validity of the limit theory arguments, our findings continue to hold. However, in the Supplement of this paper, relaxing this assumption will cause further complication in the derivation of the asymptotics for fixed effects estimators. Please refer to the comments on assumptions in the Supplement for more discussion.

[5] This condition precludes sequential limits in which $(T, N)_{\text{seq}} \to \infty$, i.e., $N \to \infty$ first, followed by $T \to \infty$. In such cases, there is partial failure in identification. In particular, when $N \to \infty$ and thus $u_{At} \to_p 0$, the model (2.3) reduces to $y_t = \beta_0(z_t) + \eta'\beta(z_t)$, where $y_{At} \to_p y_t$, and only the composite parameter $\beta_0(z) + \eta'\beta(z)$ is identified when $\eta \neq 0$.

(a) if N is fixed (η may be zero or nonzero), then

$$\sqrt{NT|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) V_{xx,\eta,N}^{-1}), \tag{2.5}$$

where

$$V_{xx,\eta,N} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix}, \tag{2.6}$$

and

$$\mathcal{B}(z) = f_z^{-1}(z)\mu_2 \sum_{s=1}^{q} h_s^2 \left[\frac{\partial f_z(z)}{\partial z_s} \frac{\partial \beta_*(z)}{\partial z_s} + \frac{1}{2} \frac{\partial^2 \beta_*(z)}{\partial^2 z_s} f_z(z) \right]; \tag{2.7}$$

(b) if $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$ jointly with $N \sum_{j=1}^{q} h_j^2 \to 0$, then

(b1) if $\eta = 0$

$$\sqrt{T|H|} D_N (\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) (V_{xx}^*)^{-1}), \tag{2.8}$$

where

$$D_N = \begin{pmatrix} \sqrt{N} & 0 \\ 0 & I_p \end{pmatrix}, \quad V_{xx}^* = \begin{pmatrix} 1 & 0 \\ 0 & V_{xx} \end{pmatrix};$$

(b2) if $\eta \neq 0$

$$\sqrt{T|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) \begin{pmatrix} \eta'\Sigma_{xx}^{-1}\eta & -\eta'\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\eta & \Sigma_{xx}^{-1} \end{pmatrix}\right), \tag{2.9}$$

which is a degenerate normal distribution;

Remark 2.2. *When N is fixed, it is apparent from the definition of $V_{xx,\eta,N}$ in (2.6) and the decomposition*

$$V_{xx,\eta,N} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{N}\Sigma_{xx} \end{pmatrix} + \begin{pmatrix} 1 \\ \eta \end{pmatrix} \begin{pmatrix} 1 & \eta' \end{pmatrix} \tag{2.10}$$

that when the standardized variance matrix $\frac{1}{N}\Sigma_{xx}$ is small or when the mean η is large relative to $\frac{1}{N}\Sigma_{xx}$, the matrix $V_{xx,\eta,N}$ is close to singular. In such cases, the average profile estimator can be very inefficient, as reported in the simulation results of Section 4.

Remark 2.3. *In the case where $N \to \infty$, the bandwidth condition $N \sum_{j=1}^{q} h_j^2 \to 0$ is imposed to control randomness arising from the bias component so that standard limit theory is obtained. In stationary FC regression, randomness that is contained in the conventional bias term is normally of smaller order than that contained in the typical 'variance' term involving the regression error, as demonstrated in Phillips and Wang (2020b). However, the cross-section averaged time series FC regression (2.3) differs from stationary time series FC models by virtue of the effects of averaging the regressors and errors while keeping the intercept unchanged. This difference effectively magnifies the randomness arising from the approximation error of the intercept coefficient $\beta_0(z_t)$. In consequence, the divergence rate of the cross section sample size N needs to be controlled in order to contol the random bias contributed by estimation of $\beta_0(z_t)$. More details are available in the proof of Lemma E.4.*

Remark 2.4. *When $\eta = 0$ and $N \to \infty$ either sequentially or jointly with T, it is clear from the definition of D_N and (2.8) that the convergence rate of $\hat{\beta}_{0,APLC}(z)$ is $\sqrt{NT|H|}$, whereas the convergence rate of $\hat{\beta}_{APLC}(z)$ is $\sqrt{T|H|}$. When $h_s = O(h)$ for $s = 1, ..., q$, the optimal bandwidth order is then given by $h = O((NT)^{-1/(q+4)})$ for $\hat{\beta}_{0,APLC}(z)$ and $h = O(T^{-1/(q+4)})$ for $\hat{\beta}_{APLC}(z)$. Accordingly, a two step procedure may be considered as the mean squared error (MSE) of $\beta_0(z)$ and $\beta(z)$ cannot be minimized simultaneously. The idea is similar to that discussed in Cai et al. (2009). But we do not pursue this direction further in the present paper. For the purpose of simulation we use the order $h = O(T^{-1/(q+4)})$ when implementing the APLC estimators in later comparisons with our alternate approach.*

Remark 2.5. *The singularity of the limit distribution of $\sqrt{T|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z))$ in Theorem 2.1 (b2) arises from the singularity of $V_{xx,\eta,N}$ as $N \to \infty$ and $\eta \neq 0$. Note that $V_{xx,\eta,N}$ is the probability limit of the sample moment matrix $\frac{1}{T|H|}\sum_t x_{*,At} K_{tH} x_{*,At}'$. As $N \to \infty$, $x_{*,At} \xrightarrow{p} (1, \eta')'$ for all t. Thus, there is insufficient signal in the averaged data $x_{*,At}$ asymptotically as $N \to \infty$ to ensure a positive definite limiting sample moment matrix.*[6] *However, when $\eta = 0$, the standardized quantity $\sqrt{N} x_{At}$ converges to a normally distributed random variable, which provides sufficient variation for the sample moment matrix $\frac{N}{T|H|}\sum_t x_{At} K_{tH} x_{At}'$ to be non-singular and the standardized sample moment matrix $\frac{1}{T|H|} P_N \sum_t x_{*,At} K_{tH} x_{*,At}' P_N$, where $P_N = diag(1, \sqrt{N} I_p)$, converges to a non-singular matrix. Consequently, the convergence rate of the coefficient of x_{At} is necessarily slower than that of the intercept coefficient by order $\sqrt{N}$. Correspondingly, the variance of the limit distribution in the direction $(1, \eta')$ in Theorem 2.1 (b2) is zero, because the convergence rate is faster in this direction. Specifically, transforming (2.9) in both the direction $(1, \eta')$ and the orthogonal direction $(-\eta, I_p)$ we have as $(N, T) \to \infty$*

[6] Related consequences arise when $N \to \infty$ before $T \to \infty$, as discussed in footnote 5

$$\sqrt{T|H|}\begin{pmatrix}1 & \eta' \\ -\eta & I_p\end{pmatrix}(\hat{\beta}_{*,APLC}(z)-\beta_*(z)-\mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q\sigma_u^2 f_z^{-1}(z)\begin{pmatrix}0 & 0 \\ 0 & (I_p+\eta\eta')\Sigma_{xx}^{-1}(I_p+\eta\eta')\end{pmatrix}\right), \tag{2.11}$$

since

$$\begin{pmatrix}1 & \eta' \\ -\eta & I_p\end{pmatrix}\begin{pmatrix}\eta'\Sigma_{xx}^{-1}\eta & -\eta'\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\eta & \Sigma_{xx}^{-1}\end{pmatrix}\begin{pmatrix}1 & -\eta' \\ \eta & I_p\end{pmatrix} = \begin{pmatrix}0 & 0 \\ 0 & (I_p+\eta\eta')\Sigma_{xx}^{-1}(I_p+\eta\eta')\end{pmatrix}. \tag{2.12}$$

Remark 2.6. *Following Lee and Robinson (2015), we can consider the construction of an improved APLC estimator. The current estimator assumes the identifying condition* $\sum_{i=1}^N \alpha_i = 0$*, which is arbitrary in terms of weighting the fixed effects. In general, we could require that* $\omega'\alpha = 0$*, for some* $N \times 1$ *weight vector* ω*, and consider choosing an optimal* ω*. In our current design, we assume* $\{u_{it}\}$ *is iid across* i*, stationary across* t *and independent with* $\{z_t\}$*. It is not hard to verify that the optimal* ω *in this setting is simply* $\frac{1}{N}1_N$*, which leads to the same identifying condition* $\sum_{i=1}^N \alpha_i = 0$ *used here and, hence, the same estimator. But in the heteroskedastic situation, for example where* $\mathbb{E}(u_{it}u_{jt}|z_t = z)$ *is a function of* z*, other choices may be optimal. This research direction is not pursued in the present work because we consider in the following section a preferred approach that is based on profiling with the full panel.*

To develop a complete asymptotic theory in the degenerate case (Theorem 2.1 (b2)), we first transform coordinates in the regression (2.3) as follows

$$y_{At} = (x'_{*,At}C_\eta)(C_\eta^{-1}\beta_*(z_t)) + u_{At} \equiv (x'_{*,At}C_\eta)\theta_*(z_t) + u_{At}, \tag{2.13}$$

where $C_\eta = \begin{pmatrix}1 & -\eta' \\ \eta & I_p\end{pmatrix}$ and $\theta_*(z_t) = C_\eta^{-1}\beta_*(z_t)$ is the transformed coefficient vector. Assuming η is known, we can estimate $\theta_*(z)$ using standard local level nonparametric estimation in (2.13) and denote the resulting estimator $\hat{\theta}_{*,APLC}(z)$. This is entirely analogous to the estimation of $\beta_*(z)$ based on (2.3), and the asymptotics follow easily. Theorem 2.2 below gives the asymptotic distribution of $\hat{\theta}_{*,APLC}(z)$, whose proof follows as in Theorem 2.1 and is omitted. In general, of course, η is unknown, in which case it may be estimated using the sample mean $\widehat{\eta} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T x_{it}$. Since $\widehat{\eta} = \eta + O_p(1/\sqrt{NT})$ the results of the following theorem continue to hold in the case of estimated η.

Theorem 2.2. *Under Assumptions 1-3 and as* $(N,T)_{seq} \to \infty$ *or* $(N,T) \to \infty$ *jointly*

$$\sqrt{T|H|}D_N(\hat{\theta}_{*,APLC}(z) - \theta_*(z) - \mathcal{B}_{\theta_*}(z)) \xrightarrow{d} \mathcal{N}(0, \nu_0^q\sigma_u^2 f_z^{-1}(z)(\Sigma_{xx,\eta})^{-1}), \tag{2.14}$$

where

$$\mathcal{B}_{\theta_*}(z) = f_z^{-1}(z)\mu_2 \sum_{s=1}^{q} h_s^2 \left[\frac{\partial f_z(z)}{\partial z_s} \frac{\partial \theta_*(z)}{\partial z_s} + \frac{1}{2} \frac{\partial^2 \theta_*(z)}{\partial^2 z_s} f_z(z) \right], \tag{2.15}$$

and

$$\Sigma_{xx,\eta} = \begin{pmatrix} (1+\eta'\eta)^2 & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}. \tag{2.16}$$

Equivalently,

$$\sqrt{T|H|} D_N C_\eta^{-1} (\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) (\Sigma_{xx,\eta})^{-1}). \tag{2.17}$$

Remark 2.7. *Given the definition of D_N, (2.17) implies that one linear combination of $\hat{\beta}_{*,APLC}(z)$ is $\sqrt{NT|H|}$ consistent while others are $\sqrt{T|H|}$ consistent. More specifically, since*

$$C_\eta^{-1} = \begin{pmatrix} \frac{1}{1+\eta'\eta} & \frac{\eta'}{1+\eta'\eta} \\ -\frac{\eta}{1+\eta'\eta} & (I_p + \eta\eta')^{-1} \end{pmatrix} =: \begin{pmatrix} c_\eta' \\ C_\eta^* \end{pmatrix} \tag{2.18}$$

the following component limits hold when $\eta \neq 0$ and $(N,T) \to \infty$

$$\sqrt{NT|H|} \begin{pmatrix} 1 & \eta' \end{pmatrix} (\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)\right), \tag{2.19}$$

$$\sqrt{T|H|} [I_p + \eta\eta']^{-1} \begin{pmatrix} -\eta & I_p \end{pmatrix} (\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) \Sigma_{xx}^{-1}\right), \tag{2.20}$$

and the limit distributions are independent. It follows that the linear combination $c_\eta' \hat{\beta}_{*,APLC}(z)$ is $\sqrt{NT|H|}$ consistent, whereas the other linear combinations $C_\eta^* \hat{\beta}_{*,APLC}(z)$ are $\sqrt{T|H|}$ consistent. Note that $(1+\eta'\eta) c_\eta' \hat{\beta}_{*,APLC}(z)$ is precisely the estimator of the composite intercept $\beta_0(z) + \eta'\beta(z)$ defined in Remark 2.1 (iii). This finding explains Theorem 2.1 of Li et al. (2011). Under the assumption that $\eta = 0$, the composite intercept is just the usual intercept $\beta_0(z)$. This explains why the estimator $\hat{\beta}_{0,APLC}(z)$ of $\beta_0(z)$ has a faster convergence rate, as shown in Theorem 2.1 (b1). See also Remark 2.4.

2.2 Profile local constant estimation

As shown in Theorem 2.1 the estimator of the slope function $\beta(z)$ converges at rate $\sqrt{T|H|}$. We now propose another estimation procedure that returns a $\sqrt{NT|H|}$ consistent estimator of $\beta(z)$.

In model (1.2), if the $\alpha_i, i = 1, \cdots, N$ were known, we could apply standard local level estimation of $\beta_*(z)$ giving the oracle estimator

$$\hat{\beta}_{*,PLC}^{oracle}(z) = [X_*' K_n(z) X_*]^{-1} X_*' K_n(z)(Y - D^*\alpha^*), \tag{2.21}$$

where $Y = (y_{11}, \cdots, y_{1T}, \cdots, y_{N1}, \cdots, y_{NT})'$, $D^* = (-1_{N-1}, I_{N-1})' \otimes 1_T$ is $NT \times (N-1)$, $\alpha^* = (\alpha_2, \cdots, \alpha_N)'$, X_* is $NT \times (p+1)$ by stacking the $1 \times (p+1)$ vector $x_{*,it}'$, and $K_n(z) = I_N \otimes K_T(z)$ is an $n \times n$ diagonal matrix with $n \equiv NT$.

Let $w_*(z) = [X_*' K_n(z) X_*]^{-1} X_*' K_n(z)$ be the $(p+1) \times NT$ coefficient matrix in the oracle estimator (2.21), which may then be written as $\hat{\beta}_{*,PLC}^{oracle}(z) = w_*(z)(Y - D^*\alpha^*)$. Using this estimator in the model (1.2), gives the adjusted equation

$$y_{it} = \alpha_i + x_{*,it}' w_*(z_t)(Y - D^*\alpha^*) + v_{*,it}, \tag{2.22}$$

where $v_{*,it} = y_{it} - \alpha_i - x_{*,it}' w_*(z_t)(Y - D^*\alpha^*)$. To solve for α from (2.22), we first rewrite the equation as

$$\begin{aligned} y_{it} &= \alpha_i + x_{*,it}' w_*(z_t)(Y - D^*\alpha^*) + v_{*,it} \\ &= \alpha_i + \sum_{d=1}^{p+1} x_{it,d-1} e_{*,d}' w_*(z_t)(Y - D^*\alpha^*) + v_{*,it} \\ &= \alpha_i + \sum_{d=1}^{p+1} x_{it,d-1} w_{*,d}(z_t)(Y - D^*\alpha^*) + v_{*,it}, \end{aligned} \tag{2.23}$$

where $x_{it,0} \equiv 1$, $e_{*,d}$ is a $(p+1) \times 1$ vector with 1 in the d-th entry and 0 elsewhere, and $w_{*,d}(z_t) = e_{*,d}' w_*(z_t)$ is $1 \times NT$. Arranging (2.23) in matrix observation form gives the system

$$\begin{aligned} Y - D^*\alpha^* &= \sum_{d=1}^{p+1} \mathbf{x}_{d-1} \odot [1_N \otimes w_{*,d}(Z)(Y - D^*\alpha^*)] + V^* \\ &= \left[\sum_{d=1}^{p+1} (\mathbf{x}_{d-1} \otimes 1_n') \odot (1_N \otimes w_{*,d}(Z)) \right] (Y - D^*\alpha^*) + V^* \\ &\equiv Q_1^*(Y - D^*\alpha^*) + V^*, \end{aligned}$$

where $\mathbf{x}_d = (x_{11,d}, \cdots, x_{1T,d}, \cdots, x_{N1,d}, \cdots, x_{NT,d})'$, $w_{*,d}(Z) = (w_{*,d}(z_1)', \cdots, w_{*,d}(z_T)')'$ is $T \times NT$, $\odot$ denotes the Hadamard product, Q_1^* is defined by the matrix in square parentheses in the final equation, and $V^* = (v_{*,11}, ..., v_{*,1T}, ..., v_{*,N1}, ..., v_{*,NT})'$. Consequently, we have

$$(I_{NT} - Q_1^*)Y = (I_{NT} - Q_1^*)D^*\alpha^* + V^*. \tag{2.24}$$

Let $M_1^* = I_{NT} - Q_1^*$ and $M_2^* = M_1^* D^*$. Then least squares regression on (2.24) gives

$$\hat{\alpha}_{PLC}^* = [(M_2^*)' M_2^*]^{-1} (M_2^*)' M_1^* Y, \tag{2.25}$$

with the affix "PLC" standing for Profile Local Constant. Plugging the estimator $\hat{\alpha}_{PLC}^*$ into the expression for $\hat{\beta}_{*,PLC}^{oracle}(z)$, we get the PLC estimator of the coefficient function $\beta_*(z)$

$$\hat{\beta}_{*,PLC}(z) = [X_*' K_n(z) X_*]^{-1} X_*' K_n(z)(Y - D^* \hat{\alpha}_{PLC}^*). \tag{2.26}$$

In view of the identifying restriction which requires $\sum_{j=1}^N \alpha_j = 0$, α_1 can be estimated by $\hat{\alpha}_{1,PLC} = -1_{N-1}' \hat{\alpha}_{PLC}^* = -\sum_{j=2}^N \hat{\alpha}_{j,PLC}$. Denote $\hat{\alpha}_{PLC} = (\hat{\alpha}_{1,PLC}, (\hat{\alpha}_{PLC}^*)')'$.

The following assumption imposes a rate requirement on the bandwidth H in relation to the time series sample size T. It is a useful technical condition in establishing limit theory for the estimated fixed effects $\hat{\alpha}_{PLC}$ vector, as discussed in Remark 2.8(ii) below.

Assumption 4. *As $T \to \infty$, (i) $T|H|/\log_2(T|H|) \to \infty$, and (ii) $\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2 \to 0$, where $\log_2(\cdot) := \log\log(\cdot)$.*

The following result gives the asymptotic distributions of the oracle and PLC estimators.

Theorem 2.3. *Under Assumptions 1-3, as $T \to \infty$, we have:*

(a) for the oracle estimator $\hat{\beta}_{*,PLC}^{oracle}(z)$ (with N either fixed or passing to infinity via $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$),

$$\sqrt{NT|H|}(\hat{\beta}_{*,PLC}^{oracle}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}(0, f_z^{-1}(z)\sigma_u^2 \nu_0^q V_{xx,\eta}^{-1}), \tag{2.27}$$

where $\mathcal{B}(z)$ is defined in (2.7), and

$$V_{xx,\eta} = \begin{pmatrix} 1 & \eta' \\ \eta & V_{xx} \end{pmatrix};$$

(b) when Assumption 4 holds, we have for finite N,

$$\sqrt{T}(\hat{\alpha}_{PLC} - \alpha) \xrightarrow{d} \mathcal{N}\left(0, \gamma_u^2 \left[I_N - \frac{1}{N} 1_N 1_N'\right]\right), \tag{2.28}$$

where γ_u^2 is the long run variance of $\{u_{it}\}$ defined by $\gamma_u^2 = \sum_{j=-\infty}^{\infty} \gamma_u(j)$ with $\gamma_u(j) = \mathbb{E}(u_{1t} u_{1,t+j})$; when $N \to \infty$, we have

$$\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j) \xrightarrow{d} \mathcal{N}(0, \gamma_u^2), \quad \text{for } j = 1, 2, ... \tag{2.29}$$

with the property that $\sqrt{T}(\hat{\alpha}_{j,PLC}-\alpha_j)$ and $\sqrt{T}(\hat{\alpha}_{\ell,PLC}-\alpha_\ell)$ are asymptotically independent for $j \neq \ell$;

(c) *with Assumption 4(i) holding and N either fixed or passing to infinity, the feasible estimator $\hat{\beta}_{*,PLC}(z)$ is asymptotically equivalent to the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z)$.*

Remark 2.8. ***(Limit theory of $\hat{\alpha}_{PLC}$)*** *(i) Theorem 2.3 (b) provides limit theory for the estimator of the fixed effects for both fixed N and $N \to \infty$ cases. Since our proposed estimation method profiles out fixed effects first and the feasible estimator $\hat{\beta}_{*,PLC}(z)$ is a plug-in estimator, we need to study the limit property of $\hat{\alpha}^*_{PLC}$ before examining the limit property of the feasible estimator $\hat{\beta}_{*,PLC}(z)$. To the best of our knowledge, this is the first limit result regarding estimators of fixed effects in this literature. In a more general setting with smoothing variable z_{it}, Su and Ullah (2011) profile out the fixed effects but do not provide limit theory for their estimators. The other estimation approach is to profile out the functional coefficient estimator first. That method has been adopted by Sun et al. (2009) in the functional coefficient panel data model with smoothing variable z_{it} and Li et al. (2011) in the time varying coefficient panel data model. These papers provide limit theory for the functional coefficient estimators in those models without considering estimators of the fixed effects. Asymptotic results for fixed effects estimation are important in practice because of the distinguishing role that fixed effects play in determining the dependent variable and because of their importance in forecasting, where limit theory facilitates the construction of forecast intervals.*
(ii) Assumption 4(ii) is imposed to control the bias influence on $\hat{\alpha}^*_{PLC}$ imported by the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z)$ in fixed effects estimation. From the proof of the theorem (in particular, (C.27) in the Appendix) we obtain the following representation of the estimation error

$$\hat{\alpha}^*_{PLC} - \alpha^* = O_p\left(\sqrt{\frac{\log_2(T|H|)}{T|H|}}||H||^2\right) + O_p(1/\sqrt{T}), \tag{2.30}$$

where the first order of magnitude term, $O_p\left(\sqrt{\frac{\log_2(T|H|)}{T|H|}}||H||^2\right)$, originates in the bias term of the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z)$, and the second order of magnitude term comes from the usual variance term involving the equation errors $\{u_{it}\}$. When the condition $\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2 \to 0$ holds, the bias influence from $\hat{\beta}^{oracle}_{*,PLC}(z)$ is $o_p(1/\sqrt{T})$ and can be neglected, leading to $\sqrt{T}$ consistency and limit theory for the fixed effects estimator $\hat{\alpha}^*_{PLC}$ that corresponds to oracle efficient intercept estimation of α as if the functional coefficient $\beta(\cdot)$ were known. We remark that the rate condition $\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2 = \sqrt{\frac{\log_2(Th_1\cdots h_q)}{h_1\cdots h_q}}(\sum_{s=1}^q h_s^2)^2 \to 0$ is achievable when $q \leq 3$. For example, by setting $h_s = O(h)$ for $s = 1, ..., q$, we have $|H| = O(h^q)$, $||H||^2 = O(h^2)$ and $\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2 = O(\sqrt{h^{4-q}\log_2(Th^q)})$. Since $T|H| = O(Th^q) \to \infty$ and $h \to 0$, we have $h^{4-q}\log_2(Th^q) \to 0$ only when $q \leq 3$. However, even without Assumption 4(ii) we still *have* $\hat{\alpha}^*_{PLC} - \alpha^* = o_p(1)$*, so that* $\hat{\alpha}^*_{PLC}$ *remains consistent for* α^**. Moreover, the asymptotic equivalence between the feasible and oracle estimators of* $\beta(z)$ *stated in Theorem 2.3 (c) continues to hold. This is because the bias of* $\hat{\alpha}^*_{PLC}$ *arising from the oracle estimator (i.e., the first order term in (2.30)) becomes smaller than the bias of the oracle estimator itself. Therefore the bias of the feasible estimator is dominated by that of the oracle estimator and there is no need to further control for the bias in* $\hat{\alpha}^*_{PLC}$*.*

(iii) When N *is fixed, we have joint asymptotic normality for the components of* $\hat{\alpha}_{PLC}$*. The distribution on the right side of (2.28) is unsurprisingly degenerate since the covariance matrix* $I_N - \frac{1}{N}1_N1_N'$ *is singular with* $1_N'(I_N - \frac{1}{N}1_N1_N')1_N = 0$ *and* $\hat{\alpha}_{PLC}$ *is perfectly linearly self-related by construction. The subvector estimator* $\hat{\alpha}^*_{PLC}$ *is also asymptotically normal and, as shown in the proof, for fixed* N

$$\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) \xrightarrow{d} \mathcal{N}(0, \gamma_u^2[I_{N-1} - N^{-1}1_{N-1}1_{N-1}']), \tag{2.31}$$

but $I_{N-1} - \frac{1}{N}1_{N-1}1_{N-1}'$ *is positive definite, with* $1_{N-1}'[I_{N-1} - N^{-1}1_{N-1}1_{N-1}']1_{N-1} = 1 - \frac{1}{N}$*. In fact, being the associated linear combination of* $\hat{\alpha}^*_{PLC}$*,* $\hat{\alpha}_{1,PLC}$ *satisfies*

$$\begin{aligned}\sqrt{T}(\hat{\alpha}_{1,PLC} - \alpha_1) = -1_{N-1}'\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) &\xrightarrow{d} \mathcal{N}(0, \gamma_u^2 1_{N-1}'[I_{N-1} - N^{-1}1_{N-1}1_{N-1}']1_{N-1}) \\ &= \mathcal{N}(0, \gamma_u^2(1 - 1/N)).\end{aligned} \tag{2.32}$$

The proof of Theorem 2.3 further shows that when N *is fixed,* $\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j) \xrightarrow{d} \mathcal{N}(0, \gamma_u^2(1 - 1/N))$ *holds uniformly over* $j = 1, ..., N$ *and that* $\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)$ *and* $\sqrt{T}(\hat{\alpha}_{\ell,PLC} - \alpha_\ell)$ *have asymptotic covariance* $-\gamma_u^2/N$ *for all* $j \neq \ell$*. These results together lead to (2.29) and asymptotic independence of the components* $\{\hat{\alpha}_{j,PLC}\}$ *when* $N \to \infty$*. Importantly, as* $N \to \infty$ *the covariance matrix* $I_N - \frac{1}{N}1_N1_N' \to I_\infty$*, which is the nonsingular identity matrix of infinite order* [7]*. Thus, the distinction between the properties of* $\hat{\alpha}_{PLC}$ *and* $\hat{\alpha}^*_{PLC}$ *disappears in the infinite dimensional case. These asymptotics for the fixed effects estimator clearly rely on cross section independence over* i *given in Assumption 2 (a). Allowing dependence over* i *produces additional technical complications and is left for future study.*

Remark 2.9. ***(Optimal bandwidth order)*** *Suppose* $h_s = O(h)$ *for* $s = 1, ..., q$*. Based on the limit theory given in (2.27), the optimal bandwidth order is* $h = O((NT)^{-1/(4+q)})$*. This rate may seem counterintuitive given that smoothing in the nonparametric estimate relates only to* $\{z_t\}$ *over time series observations whereas the optimal order suggests a smaller bandwidth should be used when* N *increases while* T *is fixed. The result is explained by the fact that we use the series* $\{z_t\}$ *repeatedly in estimation with cross section data. For each* i*, we treat* $z_{it} \equiv z_t$*. From*

[7] A nonsingular matrix of infinite order is a matrix of infinite order with both a left and right inverse and whose inverse is unique (see, e.g., MacDuffee (2012)). I_∞ is the identity matrix in an associative matric algebra of infinite order with itself the unique inverse.

this point of view, the effective sample size is $\sqrt{NT|H|}$. *But this conclusion is misleading and, mistakenly, it may seem reasonable to let the bandwidth decrease as* N *increases. Care is needed when* N *is large and* T *is moderate as the 'optimal' rate* $O((NT)^{-1/(q+4)})$ *may lead to a very small bandwidth. In such cases, increasing the cross section sample size* N *does not increase the density of the fixed* T *sample observations* $\{z_t\}$ *in the sample space. In consequence, nonparametric estimation of* $\beta_*(z)$ *is more vulnerable to a weak signal and denominator singularity, viz., that there is no observation point* z_t *within the given bandwidth region for very small* h. *In practice, we can impose some restrictions on the bandwidth to ensure that at least one or two points are available in the selected bandwidth range. Or we can simply use a kernel with unbounded support such as the Gaussian kernel to avoid this problem.*

Remark 2.10. ***(N/T ratio)*** *We emphasize that the limit theory obtained here requires* $T \to \infty$, $T|H| \to \infty$ *and* $T|H|/\log_2(T|H|) \to \infty$. *These asymptotics do not apply when* T *is fixed and* N *goes to infinity*[8]. *The failure is evident from the form of the (cross section averaged) kernel density estimator* $\frac{1}{NT|H|}\sum_i \sum_t K(H^{-1}(z_t - z))$. *When* T *is fixed, there are insufficient time series observations to estimate the density of* z_t *at an arbitrary point* z *in the support. For the kernel density estimate to converge to the true density, we need the time series sample size* T *and its effective sample size* $T|H|$ *to pass to infinity (i.e.,* $T \to \infty$ *and* $T|H| \to \infty$*) in which case the estimate converges to* $f_z(z)$. *To fix ideas, assume that* $h_s = O(h)$ *for* $s = 1, ..., q$ *and the optimal bandwidth order* $h = O((NT)^{-1/(q+4)})$ *is used. To ensure* $T|H|/\log_2(T|H|) = O(Th^q/\log_2(Th^q)) \to \infty$ *in this case, we need* $T^4/N^q \to \infty$. *When* z_t *is univariate, this reduces to* $N/T^4 \to 0$, *which is unlikely to be demanding in practical work unless* T *is very small or* N *is extremely large.*

Remark 2.11. ***(Density estimation)*** *Suppose* z_t *is univariate. To estimate the density* $f_z(z)$, *we recommend using the sample average* $\frac{1}{Th}\sum_t K(h^{-1}(z_t - z))$ *with bandwidth order* $h = O(T^{-1/5})$. *This estimate is common across sections and no further information is available in the cross section observations to raise efficiency.*

Remark 2.12. ***(Asymptotically equivalent estimation of the composite intercept)*** *For the intercept functional coefficient* $\beta_0(z)$, *we have two estimators:* $\hat{\beta}_{0,APLC}(z)$ *with asymptotics given in (2.5) and* $\hat{\beta}_{0,PLC}(z)$ *with asymptotics given in (2.27). Note that the (1,1)th entry of* $V_{xx,\eta,N}^{-1}$ *is* $1 + N\eta'\Sigma_{xx}^{-1}\eta$, *and the (1,1)th entry of* $V_{xx,\eta}^{-1}$ *is* $1 + \eta'\Sigma_{xx}^{-1}\eta$. *Thus these two estimators are asymptotically equivalent when* $N = 1$ *or* $\eta = 0$. *This equivalence is confirmed in simulations. We find that when* $\eta = 0$, *the APLC estimator has averaged MSE (AMSE) very close to the PLC estimator. When* N *is relatively large, the discrepancy grows larger because of the bandwidth problem. See the first comment in Section 4.1 for more discussion.*

More generally, the PLC estimator and the APLC estimator of the composite intercept $\beta_0(z) + \eta'\beta(z)$ *are always asymptotically equivalent. This equivalence is demonstrated in the*

[8]Different limit theory applies in such cases and the relevant asymptotics will be provided in a later study.

following way. From (2.19), the APLC estimator of the composite intercept has asymptotic variance $\frac{1}{NT|H|}\nu_0^q\sigma_u^2 f_z^{-1}(z)$. *From (2.27), the (1,1)th entry of* $(1\ \eta')V_{xx,\eta}^{-1}(1\ \eta')'$ *is 1. So the asymptotic variance of the PLC estimator of the composite intercept is also* $\frac{1}{NT|H|}\nu_0^q\sigma_u^2 f_z^{-1}(z)$. *So these two estimates of the composite intercept* $\beta_0(z)+\eta'\beta(z)$ *are asymptotically equivalent.*

We note that Lee and Robinson (2015) provided one type of estimator for the nonparametric regression function in their model and this estimator corresponds to our APLC estimator. A PLC-type estimator is also possible. But the two estimators are asymptotically equivalent. The equivalence can be seen from the fact that the nonparametric conditional mean function of Lee and Robinson (2015) corresponds to our intercept function $\beta_0(z)$ *as there are no explanatory variables* x_{it} *in their model and we have demonstrated the asymptotic equivalence of APLC and PLC estimation in this scenario.*

3 Testing constancy of the functional coefficients

In practical work it is often useful to test specific parametric forms of functional coefficients. Particularly important in this respect is inference regarding constancy of the regression coefficients. In the context of model (1.2) the relevant hypothesis concerning the functional coefficient $\beta_*(z)$ is whether this vector of coefficient functions can be treated as a constant vector. Tests of such hypotheses can be constructed by examining the discrepancy between the nonparametric estimate of $\beta_*(z)$ and parametric estimate of $\beta_*(z)$. In the present case it is advantageous to use the PLC estimator because of its faster convergence rate. In what follows, we therefore use $\hat{\beta}_{*,PLC}(z)$ as given in (2.26) to construct the test statistic.

3.1 Test statistic and limit distribution under the null

Under the null that $\mathcal{H}_0 : \beta_*(z) = \beta_* = const.\ a.s.$, the model (1.2) can be estimated by least squares, giving the estimate $\hat{\beta}_{*,OLS}$. A constancy test may then be constructed based on the difference between the nonparametric estimate $\hat{\beta}_{*,PLC}(z)$ and the null-restricted estimate $\hat{\beta}_{*,OLS}$ at a fixed number m of distinct points $\{z_s^*\}_{s=1}^m$[9], namely

$$I_m = \sum_{s=1}^{m}[\hat{\beta}_{*,PLC}(z_s^*) - \hat{\beta}_{*,OLS}]'[\hat{\beta}_{*,PLC}(z_s^*) - \hat{\beta}_{*,OLS}]. \tag{3.1}$$

By standard results in linear panel models with fixed effects and stationary data, we know that $\hat{\beta}_{*,OLS}$ is $\sqrt{TN}$ consistent, which is faster than the $O_p(\sqrt{TN|H|})$ convergence rate of $\hat{\beta}_{*,PLC}(z)$.

[9] In practice, the percentiles of $\{z_t\}_{t=1}^T$ may be used to select these distinct points. For example, in the illustrative simulation in Section 4, we consider $m \in \{3, 9, 20\}$ and use the $\{i/(m+1)\}_{i=1}^m$ percentiles of $\{z_t\}$ as the m distinct points.

Then under the null hypothesis and based on Theorem 2.3 (a) and (c) we have

$$\sqrt{TN|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) \xrightarrow{d} \mathcal{N}(0, \Omega(z)), \tag{3.2}$$

where $\Omega(z) \equiv f_z^{-1}(z)\sigma_u^2 \nu_0^q V_{xx,\eta}^{-1}$.

Further, for any m distinct points $\{z_s^*\}_{s=1}^m$ and with undersmoothing[10] in the construction of the nonparametric estimates $\hat{\beta}_{*,PLC}(z_i^*)$, we have

$$\sqrt{NT|H|}\begin{pmatrix} \hat{\beta}_{*,PLC}(z_1^*) - \beta_*(z_1^*) \\ \vdots \\ \hat{\beta}_{*,PLC}(z_m^*) - \beta_*(z_m^*) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \Omega(z_1^*) & \cdots & 0 \\ \vdots & & 0 \\ 0 & \cdots & \Omega(z_m^*) \end{pmatrix}\right). \tag{3.3}$$

To show this result, it suffices to verify that the limit variance matrix is block diagonal, which can be done in the standard manner by showing that $|H|^{-1}\mathbb{E}K(H^{-1}(z_t - z_1^*))K(H^{-1}(z_t - z_2^*)) = o(1)$. In view of (3.2), one can verify that (3.3) continues to hold under the null by replacing $\beta_*(z_s^*)$ $(s = 1, ..., m)$ with $\hat{\beta}_{*,OLS}$. To obtain a suitable pivotal limit theory for the test statistic we normalize the quantity I_m as follows

$$I_m^* = NT|H|\sum_{s=1}^m [\hat{\beta}_{*,PLC}(z_s^*) - \hat{\beta}_{*,OLS}]'\hat{\Omega}^{-1}(z_s^*)[\hat{\beta}_{*,PLC}(z_s^*) - \hat{\beta}_{*,OLS}], \tag{3.4}$$

where $\hat{\Omega}(z)$ is a consistent estimate of $\Omega(z)$. It is easy to see that $I_m^* \xrightarrow{d} \chi^2_{(p+1)m}$ under the null, where χ_k^2 denotes the χ^2 distribution with k degrees of freedom.

Theorem 3.1. *Under Assumptions 1-3, 4(i) and the null $\mathcal{H}_0$, we have $I_m^* \xrightarrow{d} \chi^2_{(p+1)m}$ as $T \to \infty$.*

In addition, when m is large we can construct a test statistic that is asymptotically standard normal under the null. Denote $\delta(z) = \sqrt{NT|H|}\hat{\Omega}^{-1/2}(z)(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS})$ and then $\delta(z) \xrightarrow{d} \mathcal{N}(0, I_{p+1})$ as $T \to \infty$, so that $\delta(z)'\delta(z) \xrightarrow{d} \chi^2_{p+1}$. Consider the quantity $I(m) = m^{-1}I_m^* = m^{-1}\sum_{s=1}^m \delta(z_s^*)'\delta(z_s^*)$, where the $\{z_s^*\}_{s=1}^m$ are m distinct points in the support of z_t. Since $\delta(z_s^*)'\delta(z_s^*)$ and $\delta(z_t^*)'\delta(z_t^*)$ are asymptotically independent for $s \neq t$, we can apply the Lindeberg CLT to $I(m)$, giving as $m \to \infty$,

$$\sqrt{m}(I(m) - \mathbb{E}(\delta(z_t)'\delta(z_t))) \xrightarrow{d} \mathcal{N}(0, \mathbb{V}ar(\delta(z_t)'\delta(z_t))). \tag{3.5}$$

Since $\delta(z)'\delta(z) \xrightarrow{d} \chi^2_{p+1}$, we have $\mathbb{E}(\delta(z)'\delta(z)) \to p+1$ and $\mathbb{V}ar(\delta(z)'\delta(z)) \to 2(p+1)$ as $NT \to \infty$. Then

$$J = \sqrt{\frac{m}{2(p+1)}}[m^{-1}I_m^* - (p+1)] \xrightarrow{d} \mathcal{N}(0,1), \tag{3.6}$$

[10]Undersmoothing requires $\sqrt{NT|H|}||H||^2 \to 0$ as $NT \to \infty$.

as $NT \to \infty$ and $m \to \infty$, using sequential asymptotics where $NT \to \infty$ and $|H| \to 0$ followed by $m \to \infty$. It seems likely that the result holds under joint asymptotics, perhaps with an additional condition controlling the expansion rate of m relative to T, but this is not investigated here.

3.2 Local asymptotic power

We consider the local alternative $\mathcal{H}_1^L : \beta_*(z) = \beta_* + \rho_n g_*(z)$, where $n \equiv NT$, ρ_n is a sequence of constants that goes to 0 as $n \to \infty$, and $g_*(z) = (g_0(z), g(z)')'$ is a $(p+1) \times 1$ uniformly bounded real vector function that satisfies the continuity condition in Assumption 2 (d). To distinguish the alternative from the null we further require that $P(g_0(z_t) = constant) < 1$ and $P(g(z_t) = constant) < 1$, so that $g_*(z)$ is not a constant function. This kind of local alternative is commonly used in the study of nonparametric and semiparametric inference involving stationary and nonstationary data; see, for example, Gao et al. (2009a), Gao et al. (2009b), Wang and Phillips (2012), Chen et al. (2015).

The limit properties of the test statistic I_m^* under $\mathcal{H}_1^L$ are given in the following result.

Theorem 3.2. *Under Assumptions 1-3, 4(i) and the local alternative $\mathcal{H}_1^L$, we have the following limit behavior for fixed m as $T \to \infty$,*

(a) if $\sqrt{n|H|}\rho_n \to 0$, $I_m^* \xrightarrow{d} \chi^2_{(p+1)m}$;

(b) if $\sqrt{n|H|}\rho_n = O(1)$, $I_m^* = O_p(1)$ but no longer has an asymptotic $\chi^2_{(p+1)m}$ distribution;

(c) if $\sqrt{n|H|}\rho_n \to \infty$, $I_m^* = O_p(n|H|\rho_n^2) \to \infty$.

Remark 3.1. *(i) Test performance evidently depends on the local distance from the null, which in turn depends on the rate $\rho_n \to 0$. For $\rho_n = O\left(1/\sqrt{n|H|}\right)$ tests based on I_m^* have non-trivial local asymptotic power and when $\rho_n \to 0$ at a slower rate than $O\left(1/\sqrt{n|H|}\right)$ the test is consistent.*

(ii) Similar results hold for the test statistic J. In particular, if $\sqrt{n|H|}\rho_n \to 0$, we have $J \xrightarrow{d} \mathcal{N}(0,1)$ as $(m,n)_{seq} \to \infty$. If $\sqrt{n|H|}\rho_n = O(1)$, we have $J = O_p(\sqrt{m})$. And if $\sqrt{n|H|}\rho_n \to \infty$, we have $J = O_p(n|H|\rho_n^2/\sqrt{m})$.

The analysis leading to Theorem 3.2 follows standard lines. Under $\mathcal{H}_1^L$, the properties of $\hat{\beta}_{*,OLS}$ depend on ρ_n[11]. There are two components, specifically

$$\hat{\beta}_{*,OLS} - \beta_* = O_p(1/\sqrt{n} + \rho_n). \tag{3.7}$$

[11] More generally, and especially in nonlinear, nonstationary settings, the convergence rate of $\hat{\beta}_{*,OLS}$ can depend on both ρ_n and the properties of the function $g_*(z)$, as discussed in Wang and Phillips (2012). Here we only discuss dependence on ρ_n.

From the decomposition $\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} = \hat{\beta}_{*,PLC}(z) - \beta_*(z) + \beta_* - \hat{\beta}_{*,OLS} + \rho_n g_*(z)$ and the fact that the nonparametric estimate $\hat{\beta}_{*,PLC}(z)$ is $\sqrt{NT|H|}$ consistent under the alternative with $g_*(z)$ bounded, we have

$$\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} = O_p(\rho_n ||H||^2) + O_p(1/\sqrt{n|H|}) + O_p(1/\sqrt{n} + \rho_n) + O_p(\rho_n) = O_p(\rho_n + 1/\sqrt{n|H|}), \tag{3.8}$$

giving

$$\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(\sqrt{n|H|}\rho_n) + O_p(1), \tag{3.9}$$

where the $O_p(1)$ term arises from the asymptotic normal distribution of $\hat{\beta}_{*,PLC}(z)$. The representation (3.9) then reveals the asymptotic local power properties of the test.

(a) If $\sqrt{n|H|}\rho_n \to 0$, (3.9) is dominated by the $O_p(1)$ term. Then the limit theory (3.2) continues to hold. Consequently, I_m^* remains asymptotically $\chi^2_{(p+1)m}$ under the local alternative $\mathcal{H}_1^L$ and the test has asymptotic power equal to size for such alternatives.

(b) If $\sqrt{n|H|}\rho_n = O(1)$, then we have $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(1)$, but there is a departure from normality that is of order $O_p(1)$. Then I_m^* is still $O_p(1)$ but is no longer $\chi^2_{(p+1)m}$ distributed. In this case, the test has non-trivial local asymptotic power.

(c) If $\sqrt{n|H|}\rho_n \to \infty$, then $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) \to \infty$ and $I_m^* \to \infty$. The I_m^* test is consistent in this case.

4 Simulations

This Section reports simulation results on the finite sample performance of the estimation procedures and tests considered in the previous two sections. Section 4.1 examines finite sample performance of the two estimators proposed in Section 2 and Section 4.2 investigates inferential performance of the tests in Section 3.

4.1 Estimation Accuracy

We use the following data generating mechanism

$$\begin{aligned}
y_{it} &= \alpha_i + \beta_0(z_t) + x_{it}\beta_1(z_t) + u_{it}, \\
x_{it} &= \eta + x_{it}^0, \quad x_{it}^0 = \rho_1 x_{it-1}^0 + \xi_{it}, \\
\alpha_i &= c_0 x_{iA}^0 + v_i, \quad i = 1, \cdots, N-1, \quad \alpha_N = -\sum_{i=1}^{N-1} \alpha_i,
\end{aligned}$$

where (u_{it}, ξ_{it}, v_i) is i.i.d. $\mathcal{N}(\mathbf{0}, diag(1, 1+\eta^2, 1))$, z_t is i.i.d. $\mathcal{U}(-1, 1)$. The functional coefficients are $\beta_0(z) = 1 + z$, $\beta_1(z) = 1 + z^2$. The parameter c_0 controls the correlation between α_i and $x_{iA}^0 = T^{-1} \sum_{t=1}^T x_{it}^0$. We use $c_0 = 1$ and $\rho_1 = 0.5$. The bandwidth is determined by $h = \hat{\sigma}_z (NT)^{-1/5}$ for the PLC estimators and $h = \hat{\sigma}_z T^{-1/5}$ for the APLC estimators, where $\hat{\sigma}_z$ is the sample standard deviation of $\{z_t\}_{t=1}^T$. We use a Gaussian kernel to avoid the singularity problem discussed in Remark 2.9.

The process $\{x_{it}^0\}$ has zero mean and thus $\mathbb{E} x_{it} = \eta$. To avoid the near singularity of the matrix $V_{xx,\eta,N}$ defined in (2.6) which arises for large η (c.f., Remark 2.2), we let the variance of ξ_{it} be $1+\eta^2$. We use the values $\eta \in \{0, 1, 5\}$ under different combinations of (N, T). To evaluate estimation accuracy, we report the Averaged MSE, defined as

$$AMSE(\hat{\beta}(z)) = \frac{1}{B} \sum_{l=1}^{B} \left[\frac{1}{T} \sum_{t=1}^{T} (\hat{\beta}^{(l)}(z_t) - \beta(z_t))^2 \right], \tag{4.1}$$

where $\hat{\beta}^{(l)}(z)$ denotes the estimate in the ℓ-th replication. We use $B = 400$ and report the criteria in Table 1 for $\beta_0(z)$ and in Table 2 for $\beta_1(z)$.

Our main findings are as follows:

(1) From Table 1, we see that for $N = 5$ and $\eta = 0$ the APLC and PLC estimators have very close AMSEs. The correspondence is due to the fact that under this scenario the estimates are asymptotically equivalent, as noted in Remark 2.12. When $N = 50$ and $\eta = 0$, the discrepancies are mainly due to the large difference in their respective bandwidths (viz., $h = O(T^{-1/5})$ for the APLC estimator and $h = O((NT)^{-1/5})$ for the PLC estimator). As noted in Remark 2.4, we are not using the optimal bandwidth order for the APLC estimator $\hat{\beta}_{0,APLC}(z)$. But when $\eta \neq 0$, the PLC estimate is more efficient that the APLC estimate, as expected because the APLC estimator converges at the slower rate $\sqrt{T|H|}$ indicated in Theorem 2.1 (a), whereas the PLC estimator still converges at the rate $\sqrt{NT|H|}$.

(2) From Table 2, it is evident that the PLC estimates always outperform the APLC estimates irrespective of whether $\eta = 0$ or $\eta \neq 0$. This outcome is well expected as the APLC estimator of $\beta_1(z)$ is $\sqrt{T|H|}$ consistent and the PLC estimator is $\sqrt{NT|H|}$ consistent, consonant with the findings presented in Theorem 2.1 (a) and Theorem 2.3 (a).

(3) The feasible PLC estimator performs almost as well as the oracle estimator in both Tables 1 and 2, especially when T is large. These results corroborate the asymptotic equivalence of the estimators given in Theorem 2.3 (c).

(4) The $AMSEs$ of the $\beta_0(z)$ estimates reported in Table 1 increase as η increases, whereas the $AMSEs$ of the $\beta_1(z)$ estimates in Table 2 decrease as η increases. This is also explained by the asymptotic theory. For the APLC estimates, from Theorem 2.1 (a) it is easy to verify

that the (1,1)th entry of $V^{-1}_{xx,\eta,N}$ is an increasing function of η and the (2,2)th entry is a decreasing function of η (note that $\Sigma_{xx} = 4(1+\eta^2)/3$ according to our simulation design). For the PLC estimates, from Theorem 2.3 (a) it is apparent that the (1,1)th entry of $V^{-1}_{xx,\eta}$ is an increasing function of η and the (2,2)th entry is a decreasing function of η.

(5) Finally, as T increases, all the AMSEs decrease as expected. For the estimates that have $\sqrt{NT|H|}$ convergence rates (viz., the PLC estimates and the APLC estimate of $\beta_0(z)$ when $\eta = 0$), the AMSEs also decrease as N increases. But for the $\sqrt{T|H|}$ consistent estimates (the APLC estimate of $\beta_0(z)$ with $\eta \neq 0$ and the APLC estimate of $\beta_1(z)$), the AMSEs increase as N increases[12]. The heuristic explanation is that there is information loss from cross section averaging in the first step of APLC estimation, see (2.3). When N increases, the average $x_{At} = N^{-1}\sum_{i=1}^{N} x_{it}$ converges to its mean η at which limit there is insufficient signal variation (information) to jointly identify $\beta_0(z)$ and $\beta(z)$. Correspondingly, as N rises for any given T the estimation accuracy of the APLC procedure naturally deteriorates. An exception occurs for the APLC estimator of $\beta_0(z)$ which has a $\sqrt{NT|H|}$ convergence rate when $\eta = 0$ because there is no intercept contamination from $x'_{At}\beta(z_t)$ as there is when $\eta \neq 0$. The phenomenon is related to the failure of the asymptotic theory when $N \to \infty$ with T fixed, as discussed earlier in Remark 2.10.

4.2 Test performance

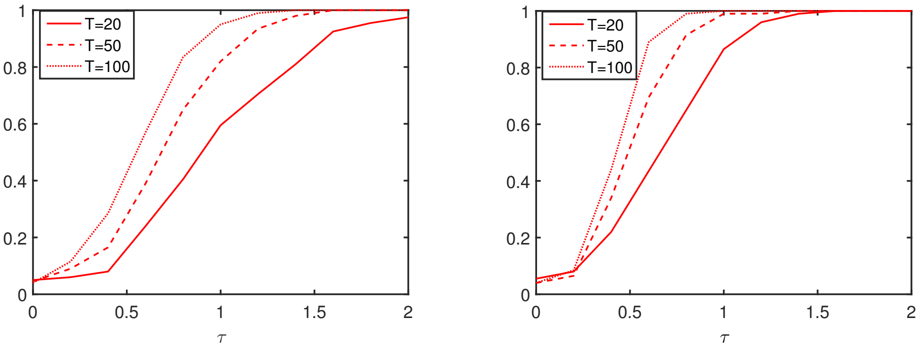
We next consider the finite sample performance of the test statistics defined in (3.4) and (3.6). The DGP is the same as in Section 4.1 except that the functional coefficients are given as

$$\beta_0(z) = 1 + \rho_n z, \beta_1(z) = 1 + \rho_n z^2, \tag{4.2}$$

where ρ_n satisfies $\sqrt{nh}\rho_n \to \infty$ with $n \equiv NT$ such that I^*_m and J are both consistent tests (see Theorem 3.2 and the following remarks). More specifically, we set $\rho_n = O((nh)^{-1/4}) = \tau n^{-3/16}$ since we adopt $h = O(n^{-1/4})$ to achieve undersmoothing of the PLC estimator. Without loss of generality, we set $\eta = 1$. We let $m \in \{3, 9, 20\}$ to examine the impact of the number of distinct points used in the test statistics and the $\{j/(m+1)\}_{j=1}^{m}$ percentiles of $\{z_t\}$ are used for the m points. We let $N \in \{5, 20\}$ and $T \in \{20, 50, 100\}$.

We first examine size performance of the tests I^*_m and J. We set $\tau = 0$. The rejection rates with 5% nominal size are collected in Table 3 with 200 replications. The bandwidth is determined by $h = \hat{\sigma}_z \cdot n^{-1/4}$. Evidently, the two tests share similar performance. When $N = 5$, they are undersized if m is small. As m increases, size also increases, especially for small T. When m=20, the tests are slightly oversized when T is small. When $N = 20$, they are undersized

[12] Li et al. (2011) found a similar phenomenon for their $\sqrt{Th}$ consistent estimator, although their reported AMSEs appear not to be monotonically increasing as N increases.



(a) $N = 5$ (b) $N = 20$

Figure 1: Rejection rates of the bootstrap procedure with 5% nominal size against the local alternative $\mathcal{H}_1^L : \beta_0(z) = 1 + \rho_n z, \beta_1(z) = 1 + \rho_n z^2$ where $\rho_n = \tau n^{-3/16}$

for small m. For m=20, size is close to the nominal size. Overall, test size appears sensitive to m and more sensitive in the case of small sample size.

To achieve better size control we adopt a bootstrap procedure. Details of the procedure are provided in Appendix F[13]. Within each replication, 200 bootstrap samples are used to generate critical values. Since J is a monotone transformation of I_m^*, the bootstrap rejection rates are the same for these two tests. The two tests are therefore not distinguished in the bootstrap setting. To examine the sensitivity to bandwidth, we consider the settings $h = c_h \cdot \hat{\sigma}_z \cdot n^{-1/4}$ with $c_h \in \{0.5, 1.0, 1.5\}$. The results are collected in Table 4. Evidently, the bootstrap test size is close to nominal for most parameter constellations and there is no clear dependence on m, and size performance is very stable with respect to bandwidth. We therefore recommend using the bootstrap procedure to help ensure size control in these tests in practical work.

To assess test power we let the localizing coefficient τ (recall that $\rho_n = \tau n^{-3/16}$) in departures from the null in (B.1) vary from 0.2 to 2 with a step length of 0.2. The rejection rates of the test implemented using the bootstrap procedure with 5% nominal size are plotted in Figure 1 for (a) $N = 5$ and (b) $N = 20$. We show the results with $m = 9$ here as the test based on the bootstrap procedure shows little sensitivity to m.

In Figure 1 the rejection rates show test size when $\tau = 0$ and the results are close to the nominal size (5%), as seen in Table 4. When the localizing coefficient τ increases, the rejection rates are flat for small departures from the null but subsequently rise rapidly with τ. The power curves also rise uniformly as T increases. Comparison of the results for $N = 5$ and $N = 20$ reveals that increasing N also improves power. These findings corroborate the results in Section 3.2 that the test is consistent for departures from the null of the form (B.1) with $\sqrt{nh}\rho_n \to \infty$.

[13] A parametric bootstrap is used for simplicity.

5 Empirical application

As an application of our methods we examine the relationship between housing rental prices and the labor market in the United Kingdom. We use three observational data sets: the index of the private housing rental prices (P_{it}, monthly), total workforce job numbers (W_{it}, quarterly measure of the number of jobs) and the UK labor force participation rate (R_t, annually as percentages of total population ages 15+). The rental prices and workforce jobs data record time series in nine English regions (see the column names in Table 5), thereby leading to a conventional panel dataset with time series observations over individual regions, whereas the national labor force participation rate data varies only over time and is a communal variable. The rental price data is available from the Office for National Statistics, the workforce jobs data is collected from Nomis and the labor force participation rate is obtained from the World Bank.

To meet stationarity requirements in the model framework, the data are transformed. We compute the 12-month percentage growth rate for housing rental prices in each of the English regions. The data is transformed to quarterly by taking averages within each quarter and this quarterly measure of housing rental price growth is denoted p_{it}. For the workforce jobs data, we compute the quarter-on-quarter percentage growth rate, which is denoted w_{it}. The labor force participation rate is found to be a unit root process with drift by standard ADF testing and first differences are used to achieve stationarity.[14] Quarterly readings are obtained by interpolation and this time series is denoted r_t. To ensure data availability for all series we use data over the period 2006Q1 to 2018Q4 for the 9 English regions, giving a panel with $N = 9$ and $T = 52$. The aggregated time series based on these transformed series are plotted in Figure 2.

Figure 2 shows that these three time series move in a closely related manner with w_{At} and r_t evidently acting as possible leading indicators for aggregate movements in p_{At}. To test this role at the regional level we use the functional coefficient model specification

$$p_{it} = \alpha_i + w_{i,t-k_1}\beta(r_{t-k_2}) + u_{it}, \tag{5.1}$$

which allows for potential lagged influences of the growth of the regional workforce $w_{i,t-k_1}$ on current period regional house rental price growth p_{it}. The functional coefficient $\beta(r_{t-k_2})$ in (5.1) allows for the magnitude of the regional workforce influence on prices to be functionally dependent on changes in the national labor force participation rate, thereby reflecting a potential country-wide impact on the regional housing market. The specification allows both these effects to operate with lags that are empirically determined. No functional coefficient intercept is included in the specification (5.1) to achieve greater parsimony and because the individual fixed

[14] It is common for bounded series such as participation rates to display random wandering behavior and to return a positive unit root test outcome. This regularly happens with economic share and exchange rate target zone data as well as political approval ratings data (see, Phillips (2001) for some examples and discussion).

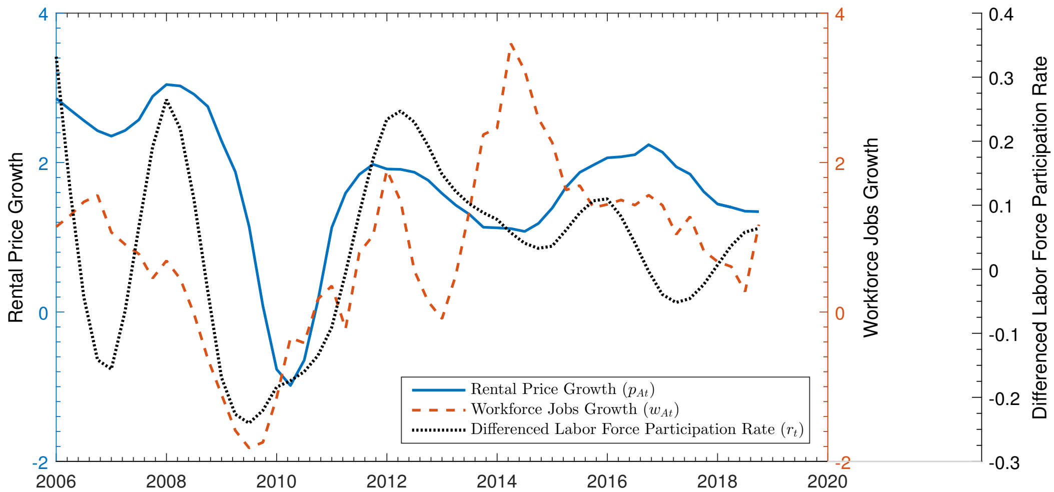


Figure 2: Time series plots of region-averaged housing rental price growth rates ($p_{At} = \frac{1}{N}\sum_{i=1}^{N} p_{it}$, percentage; blue solid line), workforce jobs growth rate ($w_{At} = \frac{1}{N}\sum_{i=1}^{N} w_{it}$, percentage; red dashed line) and first order differenced labor force participation rate r_t (percentage; black dotted kube) ranging from 2006Q1 to 2018Q4.

effects capture regional differences well.[15] No improvements to within sample fit or forecasting performance were achieved with the inclusion of a functional coefficient communal intercept.

To determine the lag specifications (k_1, k_2) and bandwidth h in (5.1) cross validation is employed. More specifically, for given lag choices $1 \leq k_1, k_2 \leq 4$, the cross validation criteria $CV(h; k_1, k_2)$ is minimized with respect to h. The objective function is defined as $CV(h; k_1, k_2) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{N}\sum_{i=1}^{N}(p_{it} - \hat{p}_{it}^{(-t)}(h; k_1, k_2))^2$, where $\hat{p}_{it}^{(-t)}(h; k_1, k_2)$ denotes the fitted value of p_{it} based on model (5.1) using bandwidth h and lags (k_1, k_2) when the data point at time t is removed. In this implementation the rule-of-thumb formula $h = c_h \hat{\sigma}_r n^{-1/5}$ is employed, where $\hat{\sigma}_r$ is the sample standard deviation of $\{r_t\}$. Optimization is then conducted with respect to c_h. We use the range $0.2 \leq c_h \leq 2.5$ in a grid search (with step 0.1) to locate the optimal c_h. Upon obtaining the optimal $c_h = c_h(k_1, k_2)$, conditional on the lags (k_1, k_2), the lags (k_1, k_2) are then chosen to minimize the objective function $CV(h; k_1, k_2)$. This procedure returns the values $k_1 = 3, k_2 = 2$ and $c_h = 0.7$. With these lag and smoothing parameter choices, model (5.1) is finalized and estimated.

The estimated functional coefficient $\hat{\beta}(\cdot)$ obtained by PLC is shown in Figure 3 with pointwise confidence bands computed using the asymptotic formula. We first note that $\hat{\beta}(\cdot)$ is positive throughout the support of r_t and significantly so at the 95% level except for a small region

[15] Asymptotic theory for the case of functional coefficient regression without a functional intercept coefficient is provided in the Online Supplement to this paper (Phillips and Wang, 2020a).

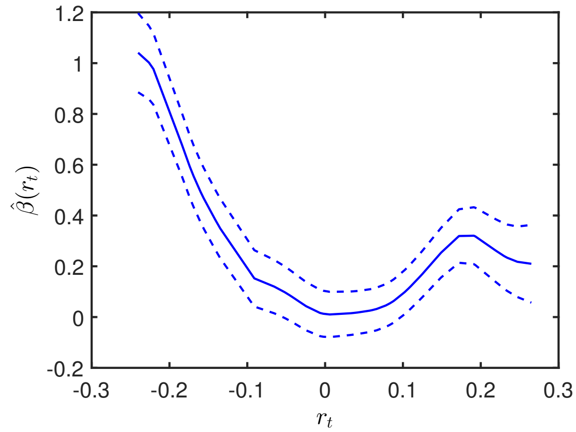


Figure 3: Estimated functional coefficient curve (solid blue line) $\beta(r)$ in the model (5.1) with lag parameter settings $k_1 = 3, k_2 = 2$, bandwidth $h = 0.7\hat{\sigma}_r n^{-1/5}$ and 95% asymptotic confidence bands (dashed blue lines).

around $r_t = 0$. This means that regional workforce growth exerts a positive force on housing rental price growth, conforming to a significant demand side influence on the housing market. The curve $\hat{\beta}(\cdot)$ in the figure shows strong nonlinearity with a U-shaped form that suggests country-wide labor force participation affects the sensitivity of housing rental price growth rate to the workforce jobs growth rate. More specifically, when the labor force participation rate remains stable (with r_t around 0), the housing rental price growth seems less sensitive (via $\hat{\beta}(\cdot)$) to workforce jobs growth than when the participation rate is rising or falling. Changes in the labor force participation rate reflect ongoing macro influences on the stability of the labor market. The empirical results plotted in Figure 3 indicate that the rental market for housing shows greater responsiveness to demand, with the regional workforce serving as a proxy for house rental demand, when the macro influences on the labor market are stronger.

To directly test whether this communal influence of the national labor market participation rate on housing rental responsiveness to demand is significant, we test whether the functional coefficient $\beta(\cdot)$ can be treated as constant. The test proposed in Section 3 is implemented with the bootstrap procedure used for inference and with 20 evenly spaced points in the support of r_t. The results give a p-value of zero and constancy of $\beta(\cdot)$ is rejected.

To assess the within sample fitting performance of model (5.1), we plot the aggregated fitted rental price growth, $\hat{p}_{At} = \frac{1}{N}\sum_{i=1}^{N}\hat{p}_{it}$, in Figure 4. For comparison, we include the analogue forecast, $\tilde{p}_{At}$, which is obtained from the following fixed coefficient linear panel model

$$p_{it} = \alpha_i + w_{i,t-k_1}\beta + v_{it}. \tag{5.2}$$

Estimation of this model (5.2) gives $\hat{\beta} = 0.262$ with the 95% confidence interval $[0.204, 0.319]$.

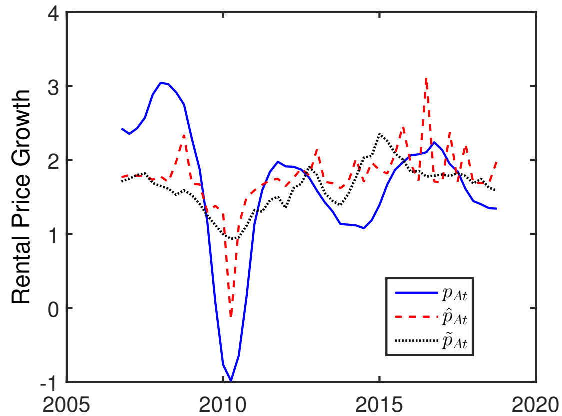


Figure 4: Time series plots of the aggregate housing rental price growth rate and model fitted values: the true value p_{At} (solid blue line), the fitted value $\hat{p}_{At}$ (dashed red line) based on the functional coefficient model (5.1), and the fitted value $\tilde{p}_{At}$ (dotted black line) based on the fixed coefficient linear panel model (5.2).

Noticeably, this confidence interval only covers a small part of the coefficient variation of $\beta(\cdot)$ observed in Figure 3. From the fits shown in Figure 4, it is evident that the functional coefficient model (5.1) assists in capturing some of the more substantial variations in the observed time series p_{At}, especially the large drop in growth around 2010. The MSE for $\hat{p}_{At}$ is 0.534, which is 7.3% lower than that for $\tilde{p}_{At}$ (0.576).

Next we compare the out-of-sample forecasting performance of the two models (5.1) and (5.2). We consider one-quarter-ahead forecasts for the latest 5 years (2014-2018). For simplicity and in keeping with the overall model estimates, we retain the fitted lag order $(k_1, k_2) = (3, 2)$ and use $c_h = 0.7$.[16] Table 5 summarizes the MSE improvement (in percentages) of model (5.1) over model (5.2) in each year and for each region. Evidently, for most regions and most years, model (5.1) has superior forecasting performance over (5.2). In terms of 5-year performance, (5.1) outperforms (5.2) for 8 out of the 9 regions. In terms of the cross-region performance, (5.1) gives improved forecasting performance over (5.2) in 4 out of 5 years. The overall improvement for all the 9 regions and for the full 5 years is 27.19%.

6 Conclusions

Panel modeling of individual behavior and spatial modeling of physical phenomena often need to account for the possible impact of macroeconomic and global influences on individual slope and intercept coefficients. These effects can be captured in panel models by means of functional

[16]Other bandwidth choices involving constants $c_h \in \{0.5, 1, 1.3, 1.5\}$ led to similar forecast performance.

coefficients in which the coefficients depend on observable communal covariates via smooth functions.

The analysis of such functional coefficient panel models in the present paper reveals important differences in convergence rates and asymptotic properties between two general classes of estimators, one (APLC) relying on convenient cross section averaged data and the other (PLC) based on the full panel sample. The differences especially highlight the effects of information loss from cross section averaging on the performance of the APLC estimator that relies on averaged data. When the cross section sample size $N \to \infty$ and the mean of the explanatory regressors is non-zero, it is apparent that there is insufficient signal variation in the regressors to jointly identify the intercept and slope functions. When N and T both pass to infinity this indeterminacy is mitigated and consistency holds but some linear combinations converge faster than others, depending on the direction determined by the mean function. Use of the full panel data in estimation avoids these difficulties and the corresponding PLC estimator has $\sqrt{NT|H|}$ convergence rate and oracle efficient properties, thereby making this method of estimation the recommended procedure for practical work. The PLC estimators may also be used to construct tests for constancy of the functional coefficients.

In developing the limit theory in this paper both sequential limit $(N,T)_{seq} \to \infty$ and joint limit $(N,T) \to \infty$ approaches are used. The latter presents additional difficulty in cases where singularities occur and new methods are employed in deriving joint limit results in those cases. The asymptotic findings on the estimators are corroborated in finite sample performance in the simulations and, with the use of a bootstrap procedure, the constancy tests are shown to have good finite sample size and power performance. The empirical application of these methods to the regional UK housing rental market shows that country-wide labor participation rates play an active role in influencing the nature and intensity of the impact of regional workforce changes on regional housing market prices, particularly during periods of rapid labor market changes.

The present work may be extended and applied in several directions. Nonparametric methods of the type given here may be developed to generalize and test parametric applications, as in Auerbach and Gorodnichenko (2012)[17] analyzing the effect of recessions and expansions on fiscal multipliers using a business cycle index, or in studying the impact of regulatory measures and government policy on variables such as reproduction rates of infections during pandemics. As already indicated, performance in functional coefficient estimation can be affected by the use of very large cross section samples in relation to time series observations. The effects can be exacerbated in the communal covariate case as consistent function estimation is reliant on $T \to \infty$ so that inconsistencies arise in fixed T cases, where the inconsistencies depend on the time series trajectory of the realized sample and the nature of the kernel function. Functional coefficient models with nonstationary, trending variables are also of great importance in empirical work. The nonstationary covariate case is particularly relevant in some applications and involves

[17]We thank the Managing Editor, Serena Ng, for this reference.

additional limit theory challenges (Wang and Phillips, 2009a,b). The authors plan to address these and other issues such as the presence of endogenous covariates in subsequent work.

Appendix

We start with some notation. C is a positive constant that may take different values in different places. For a column vector ξ, we use $|\xi| = (\xi'\xi)^{1/2}$ for the L_2 norm. For a diagonal matrix $A = diag(a_1, \cdots, a_p)$, we use $|A| = a_1 \cdots a_p$. The operator $VecDiag(B)$ takes the diagonal elements of the square matrix B and stacks them as a column vector. According to the context, we use := and =: to signify definitional equality. For any random variables ξ_n and η_n, $\xi_n \sim_a \eta_n$means ξ_n and η_n are asymptotically equivalent, namely $\xi_n = \eta_n\{1 + o_p(1)\}$. All asymptotic results in this paper assume that $T \to \infty$ with N fixed or $N, T \to \infty$ jointly. We use $(N, T)_{\text{seq}} \to \infty$ to denote the sequential limit where $T \to \infty$ followed by $N \to \infty$ and $(N, T) \to \infty$ denotes joint limits.

A Proof of Theorem 2.1

The functional coefficient model (2.3) involves time series of stationary (cross section averaged) data. Some aspects of the proof therefore follow standard lines and the asymptotics presume that $T \to \infty$. But important distinctions from existing theory occur in the case where $N \to \infty$. Both sequential limit $(N, T)_{\text{seq}} \to \infty$ and joint limit $(N, T) \to \infty$ are developed, relying on arguments developed in Lemma E.2. For the joint limit theory, the CLT for double indexed processes in Phillips and Moon (1999, theorem 2) is employed. Special arguments are needed in the case of differing convergence rates and degenerate limit distributions. A re-ordering technique is needed in some cases to ensure rigorous joint limit theory arguments.

First, we re-write model (2.3) in matrix form: $Y_A = VecDiag(X_A^* \beta_*^M) + U_A$, where β_*^M is a $(p+1) \times T$ matrix with the t-th column being the $(p+1) \times 1$ vector $\beta_*(z_t)$. Following (2.4) we have

$$
\begin{aligned}
&\hat{\beta}_{*,APLC}(z) - \beta_*(z) \\
=&[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' K_T(z) Y_A - \beta_*(z) \\
=&[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' K_T(z) [VecDiag(X_A^* \beta_*^M) - X_A^* \beta_*(z)] + [(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' K_T(z) U_A.
\end{aligned} \tag{A.1}
$$

Note that

$$
(X_A^*)' K_T(z) [VecDiag(X_A^* \beta_*^M) - X_A^* \beta_*(z)] = \sum_t x_{*,At} x_{*,At}' [\beta_*(z_t) - \beta_*(z)] K_{tH}
$$

$$= \sum_t x_{*,At}x'_{*,At}\mathbb{E}\xi^*_{\beta t} + \sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t}$$
$$= (X^*_A)'X^*_A\mathbb{E}\xi^*_{\beta t} + \sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t},$$

where $\xi^*_{\beta t} = [\beta_*(z_t) - \beta_*(z)]K_{tH}$ and $\eta^*_{\beta t} = \xi^*_{\beta t} - \mathbb{E}\xi^*_{\beta t}$. As a result,

$$\begin{aligned}&\hat{\beta}_{*,APLC}(z) - \beta_*(z)\\ &= [(X^*_A)'K_T(z)X^*_A]^{-1}(X^*_A)'X^*_A\mathbb{E}\xi^*_{\beta t} + [(X^*_A)'K_T(z)X^*_A]^{-1}\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t} + [(X^*_A)'K_T(z)X^*_A]^{-1}(X^*_A)'K_T(z)U_A.\end{aligned} \tag{A.2}$$

To study the asymptotic distribution of $\hat{\beta}_{*,APLC}(z)$, we examine the limit behavior of the three terms on the right side of (A.2).

The first term is analyzed in Lemma E.3, where it is shown that $[(X^*_A)'K_T(z)X^*_A]^{-1}(X^*_A)'X^*_A\mathbb{E}\xi^*_{\beta t} \sim_a \mathcal{B}(z)$ in eaach of the cases considered, namely N fixed or $N \to \infty$, with $\eta = 0$ or $\eta \neq 0$. Both sequential limits $(N,T)_{seq} \to \infty$ and joint limits $(N,T) \to \infty$ are permitted. Lemma E.4 implies that the second component on the right side of (A.2) is of smaller order than the third term. Therefore, the second term can be ignored asymptotically and the limit distribution is determined by the third term $[(X^*_A)'K_T(z)X^*_A]^{-1}(X^*_A)'K_T(z)U_A$. Consequently, we can write

$$\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z) \sim_a [(X^*_A)'K_T(z)X^*_A]^{-1}(X^*_A)'K_T(z)U_A. \tag{A.3}$$

It remains to analyze the right side of (A.3). The component $(X^*_A)'K_T(z)X^*_A$ is studied in Lemma E.1 for each case and weak convergence of appropriately restandardized forms of the component $(X^*_A)'K_T(z)U_A$ are analyzed in Lemma E.2. Those results cover the three cases of fixed N, sequential asymptotics with $(N,T)_{seq} \to \infty$ and joint asymptotics with $(N,T) \to \infty$. The following summary arguments bring the findings together here to establish the required limit theory.

Part (a) N fixed:
Combining results (i) in Lemmas E.1 and E.2, we have when N is fixed,

$$\begin{aligned}\sqrt{NT|H|}[(X^*_A)'K_T(z)X^*_A]^{-1}(X^*_A)'K_T(z)U_A &= [\frac{1}{T|H|}(X^*_A)'K_T(z)X^*_A]^{-1}\frac{\sqrt{N}}{\sqrt{T|H|}}(X^*_A)'K_T(z)U_A\\ &\xrightarrow{d} \mathcal{N}(0, \nu_0^q\sigma_u^2 f_z^{-1}(z)V^{-1}_{xx,\eta,N}).\end{aligned} \tag{A.4}$$

In consequence, we have

$$\sqrt{NT|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}(0, \nu_0^q\sigma_u^2 f_z^{-1}(z)V^{-1}_{xx,\eta,N}),$$

when N is fixed and $T \to \infty$. This gives the asymptotic distribution in part (a), namely (2.5).

Part (b1) $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$ and $\eta = 0$:
In this case as $N \to \infty$ with $\eta = 0$ we have

$$\begin{aligned}
&\sqrt{T|H|}D_N[(X_A^*)'K_T(z)X_A^*]^{-1}(X_A^*)'K_T(z)U_A \\
&= \left[\frac{1}{T|H|}P_N(X_A^*)'K_T(z)X_A^*P_N\right]^{-1}\sqrt{\frac{N}{T|H|}}P_N(X_A^*)'K_T(z)U_A \\
&\xrightarrow{d} \mathcal{N}(0, \nu_0^q\sigma_u^2 f_z^{-1}(z)(V_{xx}^*)^{-1}),
\end{aligned} \tag{A.5}$$

following from the results given in Lemmas E.1(ii) and E.2(ii). Combining (A.5) with (A.3) gives the limit theory in part (b1) of Theorem 2.1 for both the sequential limit $(N,T)_{seq} \to \infty$ and joint limit $(N,T) \to \infty$.

Part (b2) $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$ and $\eta \neq 0$:
For $(N,T)_{seq} \to \infty$ and $\eta \neq 0$, we can proceed from the result for fixed N that was established in (A.4) and apply sequential limits as $(N,T)_{seq} \to \infty$. First we have $V_{xx,\eta,N}^{-1} = N\begin{pmatrix} 1/N + \eta'\Sigma_{xx}^{-1}\eta & -\eta'\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\eta & \Sigma_{xx}^{-1}\end{pmatrix}$ and $\lim_{N\to\infty} N^{-1}V_{xx,\eta,N}^{-1} = \begin{pmatrix} \eta'\Sigma_{xx}^{-1}\eta & -\eta'\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\eta & \Sigma_{xx}^{-1}\end{pmatrix}$. Therefore, rescaling by $1/\sqrt{N}$ and using (A.4) as $N \to \infty$ when $\eta \neq 0$ we have

$$\sqrt{T|H|}[(X_A^*)'K_T(z)X_A^*]^{-1}(X_A^*)'K_T(z)U_A \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q\sigma_u^2 f_z^{-1}(z)\begin{pmatrix} \eta'\Sigma_{xx}^{-1}\eta & -\eta'\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\eta & \Sigma_{xx}^{-1}\end{pmatrix}\right), \tag{A.6}$$

which is a degenerate limiting normal distribution. In view of (A.3), result (2.9) follows from (A.6), establishing sequential convergence as $(N,T)_{seq} \to \infty$.

Joint convergence as $(N,T) \to \infty$ requires a more subtle argument involving transformation of the system to address the singularity of the limiting covariance matrix in (A.6). The reason is that the signal matrix $(X_A^*)'K_T(z)X_A^*$ in the regression is asymptotically singular upon standardization when $(N,T) \to \infty$. Indeed, as shown in Lemma E.1(iii), we have

$$\frac{1}{T|H|}(X_A^*)'K_T(z)X_A^* \xrightarrow{p} f_z(z)V_\eta = f_z(z)\begin{pmatrix} 1 & \eta' \\ \eta & \eta\eta' \end{pmatrix}, \text{ as}(N,T) \to \infty, \tag{A.7}$$

so that standard arguments using Slutsky's theorem in conjunction with a triangular array CLT, like those leading to (A.4), fail when $(N,T) \to \infty$ jointly, whereas the sequential argument leading to a singular normal limit distribution as $(N,T)_{seq} \to \infty$ is valid.

To establish joint asymptotics for $\hat{\beta}_{*,APLC}(z)$ as $(N,T) \to \infty$ when $\eta \neq 0$ we take account of the singularity that arises in the limiting signal matrix (A.7) in the passage to joint asymptotics. We proceed by using the limit theory established in the proof of Theorem 2.2 that applies to the transformed version (2.13) of the system involving $\theta_*(z) = C_\eta^{-1}\beta_*(z)$, where

$$C_\eta = \begin{bmatrix} 1 & -\eta' \\ \eta & I_p \end{bmatrix} \text{ and } C_\eta^{-1} = \begin{pmatrix} \frac{1}{1+\eta'\eta} & \frac{\eta'}{1+\eta'\eta} \\ -\frac{\eta}{1+\eta'\eta} & (I_p + \eta\eta')^{-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{1+\eta'\eta}[1 & \eta'] \\ (I_p + \eta\eta')^{-1}[-\eta & I_p] \end{pmatrix}. \tag{A.8}$$

In particular, from the results for the transformed system given in Theorem 2.2 and Remark 2.7, the following component limits hold as $(N,T) \to \infty$ when $\eta \neq 0$

$$\frac{\sqrt{NT|H|}}{1+\eta'\eta}\begin{pmatrix} 1 & \eta' \end{pmatrix}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\nu_0^q \sigma_u^2 f_z^{-1}(z)}{(1+\eta'\eta)^2}\right), \tag{A.9}$$

$$\sqrt{T|H|}[I + \eta\eta']^{-1}\begin{pmatrix} -\eta & I \end{pmatrix}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)\Sigma_{xx}^{-1}\right), \tag{A.10}$$

where the limit distributions are independent. These results imply that, as $(N,T) \to \infty$,

$$\sqrt{T|H|}\begin{pmatrix} \frac{1}{1+\eta'\eta} & \frac{\eta'}{1+\eta'\eta} \\ -(I_p + \eta\eta')^{-1}\eta & (I_p + \eta\eta')^{-1} \end{pmatrix}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)\begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{xx}^{-1} \end{pmatrix}\right), \tag{A.11}$$

from which it follows that

$$\begin{aligned} \sqrt{T|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) &\xrightarrow{d} \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) C_\eta \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{xx}^{-1} \end{pmatrix} C_\eta'\right) \\ &= \mathcal{N}\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)\begin{pmatrix} \eta'\Sigma_{xx}^{-1}\eta & -\eta'\Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1}\eta & \Sigma_{xx}^{-1} \end{pmatrix}\right), \end{aligned} \tag{A.12}$$

confirming that (A.6) holds jointly as $(N,T) \to \infty$. ■

B Proof of Theorem 2.2

With $D_N = diag\{\sqrt{N}, I_p\}$ and $D_N^{-1} = \frac{1}{\sqrt{N}}P_N$, where $P_N = diag\{1, \sqrt{N}I_p\}$, we have by transformation of (A.3) and with $\theta_*(z) = C_\eta^{-1}\beta_*(z)$ and $\mathcal{B}_{\theta_*}(z) = C_\eta^{-1}\mathcal{B}(z)$

$$\begin{aligned} &\sqrt{T|H|}D_N(\hat{\theta}_{*,APLC}(z) - \theta_*(z) - \mathcal{B}_{\theta_*}(z)) = \sqrt{T|H|}D_N C_\eta^{-1}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \\ &\sim_a \sqrt{T|H|}D_N C_\eta^{-1}\left[(X_A^*)'K_T(z)X_A^*\right]^{-1}\left[(X_A^*)'K_T(z)U_A\right] \\ &= \sqrt{T|H|}D_N\left[C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta\right]^{-1}\left[C_\eta'(X_A^*)'K_T(z)U_A\right] \\ &= \left[\frac{1}{T\,|H|}D_N^{-1}C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta D_N^{-1}\right]^{-1}\left[\frac{1}{\sqrt{T|H|}}D_N^{-1}C_\eta'(X_A^*)'K_T(z)U_A\right] \end{aligned}$$

$$= \left[\frac{1}{NT\left|H\right|}P_N C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta P_N\right]^{-1}\left[\frac{1}{\sqrt{NT|H|}}P_N C_\eta'(X_A^*)'K_T(z)U_A\right]$$

$$= \left[\frac{1}{T\left|H\right|}P_N C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta P_N\right]^{-1}\left[\frac{\sqrt{N}}{\sqrt{T|H|}}P_N C_\eta'(X_A^*)'K_T(z)U_A\right]. \tag{B.1}$$

We start with the signal matrix $\frac{1}{T|H|}P_N C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta P_N$. We have

$$\frac{1}{T|H|}P_N C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta P_N = \frac{1}{T|H|}P_N\begin{bmatrix} 1 & \eta' \\ -\eta & I_p \end{bmatrix}(X_A^*)'K_T(z)X_A^*\begin{bmatrix} 1 & -\eta' \\ \eta & I_p \end{bmatrix}P_N.$$

Observe that $x_{*,At}'C_\eta = (1, x_{At}')\begin{bmatrix} 1 & -\eta' \\ \eta & I_p \end{bmatrix} = (1 + x_{At}'\eta, x_{At}' - \eta') =: \left(1_{\eta t}, x_{At}^{0\prime}\right)$, where $x_{At}^0 = x_{At} - \eta = \frac{1}{N}\sum_{i=1}^N(x_{it} - \eta) = \frac{1}{N}\sum_{i=1}^N x_{it}^0$. Then

$$\frac{1}{T|H|}C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta = \frac{1}{T|H|}\begin{bmatrix} \sum_t K_{tH}1_{\eta t}^2 & \sum_t K_{tH}1_{\eta t}x_{At}^{0\prime} \\ \sum_t K_{tH}1_{\eta t}x_{At}^0 & \sum_t K_{tH}x_{At}^0 x_{At}^{0\prime} \end{bmatrix},$$

where $1_{\eta t} = 1 + x_{At}'\eta = 1 + \eta'\eta + (x_{At}^0)'\eta$ with $x_{At}^0 = x_{At} - \eta$. Then $1_{\eta t} \to_p 1 + \eta'\eta$ and $\sqrt{N}x_{At}^0 = \frac{1}{\sqrt{N}}\sum_{i=1}^N x_{it}^0 = \xi_{Nt} \to_d \xi_t = \mathcal{N}\left(0, \Sigma_{xx}\right)$, where $\Sigma_{xx} = E\left(x_{it}^0 x_{it}^{0\prime}\right)$. Just as in the proof of Lemma E.1(ii) in the case $\eta = 0$, we now obtain the following limit result as $N \to \infty$

$$\begin{aligned}
&\frac{1}{T|H|}P_N C_\eta'(X_A^*)'K_T(z)X_A^*C_\eta P_N = \frac{1}{T|H|}P_N\begin{bmatrix} \sum_t K_{tH}1_{\eta t}^2 & \sum_t K_{tH}1_{\eta t}x_{At}^{0\prime} \\ \sum_t K_{tH}1_{\eta t}x_{At}^0 & \sum_t K_{tH}x_{At}^0 x_{At}^{0\prime} \end{bmatrix}P_N \\
&= \begin{bmatrix} \frac{1}{T|H|}\sum_t K_{tH}1_{\eta t}^2 & \frac{1}{T|H|}\sum_t K_{tH}1_{\eta t}\left(\sqrt{N}x_{At}^0\right)' \\ \frac{1}{T|H|}\sum_t K_{tH}1_{\eta t}\left(\sqrt{N}x_{At}^0\right) & \frac{1}{T|H|}\sum_t K_{tH}\left(\sqrt{N}x_{At}^0\right)\left(\sqrt{N}x_{At}^0\right)' \end{bmatrix} \\
&\to_p f_z\left(z\right)\begin{bmatrix} \left(1 + \eta'\eta\right)^2 & \frac{1}{T|H|}\sum_t K_{tH}1_{\eta t}\left(\sqrt{N}x_{At}^0\right)' \\ \frac{1}{T|H|}\sum_t K_{tH}1_{\eta t}\left(\sqrt{N}x_{At}^0\right) & \frac{1}{T|H|}\sum_t K_{tH}\left(\sqrt{N}x_{At}^0\right)\left(\sqrt{N}x_{At}^0\right)' \end{bmatrix} \\
&\to_p f_z\left(z\right)\begin{bmatrix} \left(1 + \eta'\eta\right)^2 & 0 \\ 0 & \Sigma_{xx} \end{bmatrix} = f_z\left(z\right)\Sigma_{xx,\eta}.
\end{aligned} \tag{B.2}$$

Convergence of the $(1, 1)$ element of (B.2) is straightforward, the off diagonal elements satisfy

$$\frac{1}{T|H|}\sum_t K_{tH}1_{\eta t}\left(\sqrt{N}x_{At}^0\right) \to_p (1 + \eta'\eta)\mathbb{E}\left(\xi_t\right) = 0,$$

and convergence of the $(2, 2)$ block holds when $N \to \infty$ just as the (2,2) block of $\frac{1}{T|H|}P_N(X_A^*)'K_T(z)X_A^*P_N$ in Lemma E.1 (ii) with $\eta = 0$. Note that (B.2) holds under both sequential limit $(N, T)_{seq} \to \infty$

and joint limit $(N,T)\to\infty$ just as in Lemma E.1 (ii) with $\eta=0$.

Next, consider the second member of (B.1)

$$\frac{\sqrt{N}}{\sqrt{T|H|}}P_N C_\eta'(X_A^*)'K_T(z)U_A=\begin{bmatrix}\frac{1}{\sqrt{T|H|}}\sum_t 1_{\eta t}K_{tH}\left(\sqrt{N}u_{At}\right)\\ \frac{1}{\sqrt{T|H|}}\sum_t\left(\sqrt{N}x_{At}^0\right)K_{tH}\left(\sqrt{N}u_{At}\right)\end{bmatrix}$$
$$=\begin{bmatrix}\frac{1}{\sqrt{T|H|}}\sum_t(1+\eta'\eta)K_{tH}\left(\sqrt{N}u_{At}\right)+\frac{1}{\sqrt{T|H|}}\sum_t(x_{At}^0)'\eta K_{tH}\left(\sqrt{N}u_{At}\right)\\ \frac{1}{\sqrt{T|H|}}\sum_t\left(\frac{1}{\sqrt{N}}\sum_i x_{it}^0\right)K_{tH}\left(\frac{1}{\sqrt{N}}\sum_j u_{jt}\right)\end{bmatrix}$$
$$=\begin{bmatrix}\frac{1}{\sqrt{N}}\sum_i\frac{1}{\sqrt{T|H|}}\sum_t(1+\eta'\eta)K_{tH}u_{it}+\frac{1}{N\sqrt{N}}\sum_{i,j}\frac{1}{\sqrt{T|H|}}\sum_t(x_{jt}^0)'\eta K_{tH}u_{it}\\ \frac{1}{N}\sum_i\frac{1}{\sqrt{T|H|}}\sum_t x_{it}^0K_{tH}u_{it}+\frac{1}{N}\sum_{i\neq j}\frac{1}{\sqrt{T|H|}}\sum_t x_{it}^0K_{tH}u_{jt}\end{bmatrix}. \tag{B.3}$$

We first consider the sequential limit $(N,T)_{seq}\to\infty$. As $T\to\infty$, we have $\frac{1}{\sqrt{T|H|}}\sum_t(1+\eta'\eta)K_{tH}u_{it}\xrightarrow{d}\mathcal{N}(0,f_z(z)\nu_0^q(1+\eta'\eta)^2\sigma_u^2)=:\xi_i$ and ξ_i are zero-mean independent random variables due to cross sectional independence. We then have $\frac{1}{\sqrt{N}}\xi_i\xrightarrow{d}\mathcal{N}(0,f_z(z)\nu_0^q(1+\eta'\eta)^2\sigma_u^2)$ as $N\to\infty$. Combining these results we have $\frac{1}{\sqrt{N}}\sum_i\frac{1}{\sqrt{T|H|}}\sum_t(1+\eta'\eta)K_{tH}u_{it}\xrightarrow{d}\mathcal{N}(0,f_z(z)\nu_0^q(1+\eta'\eta)^2\sigma_u^2)$ as $(N,T)_{seq}\to\infty$. Similarly, for the other components in (B.3), we can show that as $(N,T)_{seq}\to\infty$

$$\frac{1}{N\sqrt{N}}\sum_{i,j}\frac{1}{\sqrt{T|H|}}\sum_t(x_{jt}^0)'\eta K_{tH}u_{it}=O_p(1/\sqrt{N})=o_p(1),$$

$$\frac{1}{N}\sum_i\frac{1}{\sqrt{T|H|}}\sum_t x_{it}^0K_{tH}u_{it}=O_p(1/\sqrt{N})=o_p(1),$$

and

$$\frac{1}{N}\sum_{i\neq j}\frac{1}{\sqrt{T|H|}}\sum_t x_{it}^0K_{tH}u_{jt}\xrightarrow{d}\mathcal{N}(0,f_z(z)\nu_0^q\Sigma_{xx}\sigma_u^2).$$

It follows that

$$\frac{\sqrt{N}}{\sqrt{T|H|}}P_NC_\eta'(X_A^*)'K_T(z)U_A\xrightarrow{d}\mathcal{N}(0,f_z(z)\nu_0^q\sigma_u^2\Sigma_{xx,\eta}). \tag{B.4}$$

Combining this result with (B.2) and (B.1) we have the sequential limit theory

$$\sqrt{T|H|}D_N(\hat{\theta}_{*,APLC}(z)-\theta_*(z)-\mathcal{B}_{\theta_*}(z))$$
$$\sim_a\left[\frac{1}{T\,|H|}P_NC_\eta'(X_A^*)'K_T(z)X_A^*C_\eta P_N\right]^{-1}\left[\frac{\sqrt{N}}{\sqrt{T|H|}}PC_\eta'(X_A^*)'K_T(z)U_A\right]$$

$$\rightarrow_d [f_z(z)\Sigma_{xx,\eta}]^{-1}\mathcal{N}\left(0, f_z(z)v_0^q\sigma_u^2\Sigma_{xx,\eta}\right) = \mathcal{N}\left(0, \frac{v_0^q\sigma_u^2}{f_z(z)}\Sigma_{xx,\eta}^{-1}\right), \tag{B.5}$$

as $(N,T)_{seq} \rightarrow \infty$.

To show joint convergence as $(N,T) \rightarrow \infty$, we follow the same argument as in the proof in Part (c) of Lemma E.2 because the components (up to some constant multipliers) in (B.3) have been tackled there. More specifically, $\frac{1}{\sqrt{N}}\sum_i \frac{1}{\sqrt{T|H|}}\sum_t (1+\eta'\eta)K_{tH}u_{it} = (1+\eta'\eta)S_{1NT}$. Other components in (B.3) can be dealt with in the same way as S_{2NT}. The required joint limit theory therefore follows, showing that (B.5) holds as $(N,T) \rightarrow \infty$ jointly as well as sequentially. The equivalent result for $\hat{\beta}_{*,APLC}$ in (2.17) and the component results given in (2.19) and (2.20) follow immediately. ■

C Proof of Theorem 2.3

Part (a) We first consider the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}$. Model (1.2) can be rewritten in the following matrix form

$$Y = D^*\alpha^* + VecDiag(X_*(1'_N \otimes \beta_*^M)) + U, \tag{C.1}$$

where X_* is $NT \times (p+1)$ by stacking the $1 \times (p+1)$ vector $x'_{*,it} = [1, x'_{it}]$, β_*^M is $(p+1) \times T$ with the t-th column being $\beta_*(z_t)$. Following (2.21) we have

$$\begin{aligned}
&\hat{\beta}^{oracle}_{*,PLC}(z) - \beta_*(z) = [X'_*K_n(z)X_*]^{-1}X'_*K_n(z)(Y - D^*\alpha^*) - \beta_*(z) \\
&= [X'_*K_n(z)X_*]^{-1}X'_*K_n(z)\left[VecDiag(X_*(1'_N \otimes \beta_*^M)) - X_*\beta_*(z)\right] + [X'_*K_n(z)X_*]^{-1}X'_*K_n(z)U.
\end{aligned} \tag{C.2}$$

Note that

$$\begin{aligned}
&X'_*K_n(z)\left[VecDiag(X_*(1'_N \otimes \beta_*^M)) - X_*\beta_*(z)\right] = \sum_{i,t} x_{*,it}x'_{*,it}[\beta_*(z_t) - \beta_*(z)]K_{tH} \\
&= \sum_{i,t} x_{*,it}x'_{*,it}\mathbb{E}\xi^*_{\beta t} + \sum_{i,t} x_{*,it}x'_{*,it}\eta^*_{\beta t} = X'_*X_*\mathbb{E}\xi^*_{\beta t} + \sum_{i,t} x_{*,it}x'_{*,it}\eta^*_{\beta t},
\end{aligned}$$

where $\xi^*_{\beta t} = [\beta_*(z_t) - \beta_*(z)]K_{tH}$ and $\eta^*_{\beta t} = \xi^*_{\beta t} - \mathbb{E}\xi^*_{\beta t}$. As a result,

$$\begin{aligned}
&\hat{\beta}^{oracle}_{*,PLC}(z) - \beta_*(z) \\
=&[X'_*K_n(z)X_*]^{-1}X'_*X_*\mathbb{E}\xi^*_{\beta t} + [X'_*K_n(z)X_*]^{-1}\sum_{i,t} x_{*,it}x'_{*,it}\eta^*_{\beta t} + [X'_*K_n(z)X_*]^{-1}X'_*K_n(z)U.
\end{aligned} \tag{C.3}$$

Using Lemma E.7, we have $[X_*'K_n(z)X_*]^{-1}X_*'X_*\mathbb{E}\xi_{\beta t}^* \sim_a \mathcal{B}(z)$. In view of Lemma E.8, the second term in the RHS of (C.3) is dominated by the third term and can be ignored asymptotically. In view of Lemmas E.5 and E.6, we obtain

$$\sqrt{NT|H|}[X_*'K_n(z)X_*]^{-1}X_*'K_n(z)U \xrightarrow{d} \mathcal{N}(0, f_z^{-1}(z)\sigma_u^2\nu_0^q V_{xx,\eta}^{-1}), \tag{C.4}$$

which holds as $T \to \infty$ under fixed N, sequential convergence $(N,T)_{seq} \to \infty$ and joint convergence $(N,T) \to \infty$. The oracle estimator therefore has the following limit theory

$$\sqrt{NT|H|}(\hat{\beta}_{*,PLC}^{oracle} - \beta_*(z) - \mathcal{B}(z)) \xrightarrow{d} \mathcal{N}(0, f_z^{-1}(z)\sigma_u^2\nu_0^q V_{xx,\eta}^{-1}). \tag{C.5}$$

Part (b) We first consider the estimator $\hat{\alpha}_{PLC}^*$ defined in (2.25). Based on (2.24) we have

$$\hat{\alpha}_{PLC}^* - \alpha^* = [(M_2^*)'M_2^*]^{-1}(M_2^*)'V^*,$$

where $V^* = VecDiag(X_*\{1_N' \otimes [\beta_*^M - \hat{\beta}_{*,PLC}^{oracle,M}]\}) + U$ is $NT \times 1$ with typical element $v_{*,it} = x_{*,it}'[\beta_*(z_t) - \hat{\beta}_{*,PLC}^{oracle}(z_t)] + u_{it}$, β_*^M is $(p+1) \times T$ with t-th column $\beta_*(z_t)$ and $\hat{\beta}_{*,PLC}^{oracle,M}$ is defined in the same way as β_*^M. Then we have

$$\begin{aligned}\hat{\alpha}_{PLC}^* - \alpha^* &= [(M_2^*)'M_2^*]^{-1}(M_2^*)'VecDiag(X_*\{1_N' \otimes [\beta_*^M - \hat{\beta}_{*,PLC}^{oracle,M}]\}) + [(M_2^*)'M_2^*]^{-1}(M_2^*)'U \\ &=: [(M_2^*)'M_2^*]^{-1}(M_2^*)'W + [(M_2^*)'M_2^*]^{-1}(M_2^*)'U, \end{aligned} \tag{C.6}$$

where $W = VecDiag(X_*\{1_N' \otimes [\beta_*^M - \hat{\beta}_{*,PLC}^{oracle,M}]\})$ is $NT \times 1$. Below we analyze the two terms in (C.6). We will show the first term has smaller order than the second term and the asymptotic distribution of $\hat{\alpha}_{PLC}^* - \alpha^*$ is therefore determined by the second term.

We start with an analysis of $M_2^* = (I_{NT} - Q_1^*)D^*$. First consider the matrix Q_1^*. We have

$$\begin{aligned} Q_1^* &= \sum_{d=1}^{p+1}(\mathbf{x}_{d-1} \otimes 1_{NT}') \odot (1_N \otimes w_{*,d}(Z)) \\ &= \begin{pmatrix} \sum_d x_{11,d-1}w_d(z_1) \\ \vdots \\ \sum_d x_{1T,d-1}w_d(z_T) \\ \vdots \\ \sum_d x_{N1,d-1}w_d(z_1) \\ \vdots \\ \sum_d x_{NT,d-1}w_d(z_T) \end{pmatrix} = \begin{pmatrix} x_{*,11}'w_*(z_1) \\ \vdots \\ x_{*,1T}'w_*(z_T) \\ \vdots \\ x_{*,N1}'w_*(z_1) \\ \vdots \\ x_{*,NT}'w_*(z_T) \end{pmatrix}, \end{aligned}$$

where $w_*(z_t) = [X_*'K_n(z_t)X_*]^{-1}X_*'K_n(z_t)$ is $(p+1) \times NT$. Then each row $x_{*,it}'w_*(z_t)$ is $1 \times NT$.

For the matrix $Q_1^* D^*$, the typical row at the $(j-1)$-th column is

$$q_{it,j-1} := x'_{*,it}[X'_* K_n(z_t) X_*]^{-1} \sum_{s=1}^{T} (x_{*,js} - x_{*,1s}) K(H^{-1}(z_s - z_t)) =: x'_{*,it} q_{t,j}, \tag{C.7}$$

for $j = 2, \cdots, N$ and $q_{t,j}$ is $(p+1) \times 1$. We now provide a uniform result regarding the order of $q_{t,j}$. We have shown $\frac{1}{NT|H|} X'_* K_n(z) X_* \xrightarrow{p} f_z(z) V_{xx,\eta}$ in Lemma E.5. Then $X'_* K_n(z) X_* = O_p(NT|H|)$. Further, $X'_* K_n(z_t) X_* = O_p(NT|H|)$ holds uniformly over t since the kernel function is bounded and the uniform convergence result $\sup_{z \in \mathcal{S}_n} \left| \frac{1}{NT|H|} X'_* K_n(z_t) X_* - f_z(z) V_{xx,\eta} \right| = o_p(1)$ holds over $\mathcal{S}_n = \mathcal{S} \cap [-c_n, c_n]$, where $\mathcal{S}$ is the support of z_t and $c_n = O(n^{\phi} \ln n)$ is a sequence of nondecreasing positive numbers for any $\phi > 0$ and $n \equiv NT$. Since $\mathbb{E}(x_{*,js} - x_{*,1s}) = 0$, by appealing to a law of interated logarithm for kernel regression estimates and triangular arrays, we have

$$\lim \sup_{T \to \infty} \pm \sqrt{2 \frac{T|H|}{\log_2(T|H|)}} \frac{1}{T|H|} \sum_{s=1}^{T} (x_{*,js} - x_{*,1s}) K(H^{-1}(z_s - z_t)) \leq C \ a.s.. \tag{C.8}$$

where $\log_2(\cdot) = \log\log(\cdot)$, and C is a constant that is independent of (j, t) by virtue of stationarity over j, t and bounded density $f_z(\cdot)$. Stute (1982), Hardle (1984) and Hall (1991) proved related LIL results for kernel estimators based on *iid* data. More recent research by Huang et al. (2014) established an LIL result of the form (C.8) for recursive kernel regression estimates with weakly dependent sequences. We believe similar results can be expected to hold for standard kernel regression estimates for weakly dependent sequences but have not been able to find a reference to such a result in the literature.[18] Then we have, uniformly for $t = 1, ..., T$ and $j = 2, ..., N$,

$$q_{t,j} = O_p\left(\frac{\sqrt{T|H| \log_2(T|H|)}}{NT|H|} \right) = O_p\left(\frac{1}{N} \sqrt{\frac{\log_2(T|H|)}{T|H|}} \right) = o_p(1) \tag{C.9}$$

because $\frac{\log_2(T|H|)}{T|H|} \to 0$ as $T|H| \to \infty$ from Assumption 4. Note that (C.9) holds for either fixed N or $N \to \infty$.

[18] Hall (1991) and Hardle (1984) used a strong approximation for empirical processes of *iid* sequences to prove an LIL for kernel density and kernel regression estimates. Recent work by Berkes et al. (2009) and Dedecker et al. (2013) provides extensions of such strong approximations to stationary sequences under mixing conditions. It seems that an LIL for kernel regression estimates with dependent data should be obtainable along similar lines under suitable dependence and bandwidth conditions, although an explicit result does not appear to be available presently in the literature.

Denote $q_{i\cdot,j-1} = (q_{i1,j-1}, q_{i2,j-1}, \cdots, q_{iT,j-1})'$, which is a $T \times 1$ vector. We write

$$M_2^* = \begin{pmatrix} -1_T - q_{1\cdot,1} & \cdots & -1_T - q_{1\cdot,N-1} \\ 1_T - q_{2\cdot,1} & \cdots & -q_{2\cdot,N-1} \\ \vdots & \vdots & \vdots \\ -q_{N\cdot,1} & \cdots & 1_T - q_{N\cdot,N-1} \end{pmatrix}, \tag{C.10}$$

and consider the $(N-1) \times (N-1)$ matrix $M_2^{*\prime} M_2^*$. The diagonal $(j-1, j-1)$th $(j = 2, .., N)$ element is

$$\begin{aligned} &\sum_{t=1}^{T}(-1 - q_{1t,j-1})^2 + \sum_{t=1}^{T}(1 - q_{jt,j-1})^2 + \sum_{i\neq 1,j}\sum_{t=1}^{T}(-q_{it,j-1})^2 \\ =&2T + 2\sum_{t=1}^{T} q_{1t,j-1} - 2\sum_{t=1}^{T} q_{jt,j-1} + \sum_{i=1}^{N}\sum_{t=1}^{T} q_{it,j-1}^2 && \text{(C.11)} \\ =&2T\{1 + o_p(1)\}. && \text{(C.12)} \end{aligned}$$

Below we show that (C.12) holds uniformly for $j = 2, ..., N$ with either fixed N or $N \to \infty$, which will hold if the final three terms of (C.11) are of order $o_p(T)$ uniformly for $2 \le j \le N$. This is achieved by noting that (C.9) holds uniformly over $t = 1, ..., T$ and $j = 2, ..., N$. For example, we have $\sum_{i=1}^{N}\sum_{t=1}^{T} q_{it,j-1}^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{*,it}' q_{t,j})^2 = O_p(\frac{\log_2(T|H|)}{N^2 T|H|})\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{*,it}' 1_{p+1})^2 = O_p(\frac{\log_2(T|H|)}{N^2 T|H|} \times NT) = O_p(T) O_p(\frac{\log_2(T|H|)}{NT|H|}) = o_p(T)$ since $\frac{\log_2(T|H|)}{T|H|} \to 0$. For the non-diagonal $(j-1, \ell-1)$th $(j \neq \ell, j, \ell = 2, ..., N)$ element of $M_2^{*\prime} M_2^*$, we have

$$\begin{aligned} &\sum_{t=1}^{T}(-1 - q_{1t,j-1})(-1 - q_{1t,\ell-1}) + \sum_{t=1}^{T}(1 - q_{jt,j-1})(-q_{jt,\ell-1}) + \sum_{t=1}^{T}(-q_{\ell t,j-1})(1 - q_{\ell t,\ell-1}) + \sum_{i\neq 1,j,\ell}\sum_{t=1}^{T}(-q_{it,j-1})(-q_{it,\ell-1}) \\ &= T + \sum_{t=1}^{T} q_{1t,j-1} + \sum_{t=1}^{T} q_{1t,\ell-1} - \sum_{t=1}^{T} q_{jt,\ell-1} - \sum_{t=1}^{T} q_{\ell t,j-1} + \sum_{i=1}^{N}\sum_{t=1}^{T} q_{it,j-1} q_{it,\ell-1} && \text{(C.13)} \\ &= T\{1 + o_p(1)\}. && \text{(C.14)} \end{aligned}$$

Equation (C.14) holds because the final five terms in (C.13) are of order $o_p(T)$ and so (C.14) holds uniformly for $j, \ell = 2, ..., N, j \neq \ell$ with either fixed N or $N \to \infty$. This can be verified in the same way as the earlier proof for (C.11) and the details are omitted. Combining (C.12) and (C.14) we have, if N is fixed,

$$\frac{1}{T} M_2^{*\prime} M_2^* \xrightarrow{p} I_{N-1} + 1_{N-1} 1_{N-1}'. \tag{C.15}$$

When $N \to \infty$, (C.15) continues to hold element wise, namely the diagonal elements of $\frac{1}{T} M_2^{*\prime} M_2^*$ converge to 2 uniformly over the index $j = 2, ..., N$ (see (C.12)) and the off-diagonal elements of $\frac{1}{T} M_2^{*\prime} M_2^*$ converge to 1 uniformly over the indices $j, \ell = 2, ..., N, j \neq \ell$ (see (C.14)).

Next consider $M_2^{*\prime}U$, which is an $(N-1)\times 1$ vector, the $(j-1)$-th element of which is

$$\begin{aligned}
&\sum_{t=1}^{T}(-1-q_{1t,j-1})u_{1t}+\sum_{t=1}^{T}(1-q_{jt,j-1})u_{jt}+\sum_{i\neq 1,j}\sum_{t=1}^{T}(-q_{it,j-1})u_{it}\\
&=-\sum_{t=1}^{T}u_{1t}+\sum_{t=1}^{T}u_{jt}-\sum_{i=1}^{N}\sum_{t=1}^{T}q_{it,j-1}u_{it} && \text{(C.16)}\\
&=\left(\sum_{t=1}^{T}u_{jt}-\sum_{t=1}^{T}u_{1t}\right)\{1+o_p(1)\}. && \text{(C.17)}
\end{aligned}$$

Equation (C.17) holds because, as we now show, the third term of (C.16) is of order $o_p(\sqrt{T})$ and is therefore negligible compared to the first two terms. In particular, we have

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{t=1}^{T}q_{it,j-1}u_{it}&=O_p\left(\frac{1}{N}\sqrt{\frac{\log_2(T|H|)}{T|H|}}\right)\sum_{i=1}^{N}\sum_{t=1}^{T}x_{*,it}'1_{p+1}u_{it}=O_p\left(\frac{1}{N}\sqrt{\frac{\log_2(T|H|)}{T|H|}}\times\sqrt{NT}\right) && \text{(C.18)}\\
&=O_p(\sqrt{T})\times O_p\left(\sqrt{\frac{\log_2(T|H|)}{NT|H|}}\right)=o_p(\sqrt{T}), && \text{(C.19)}
\end{aligned}$$

since $\frac{\log_2(T|H|)}{T|H|}\to 0$. The second equality in (C.18) holds because $\mathbb{E}(x_{*,it}'1_{p+1}u_{it})=0$ and the sum $\sum_{i=1}^{N}\sum_{t=1}^{T}x_{*,it}'1_{p+1}u_{it}=O_p(\sqrt{NT})$ by virtue of standard central limit theory for independent and stationary sequences, thereby contributing the final $\sqrt{NT}$ factor in (C.18).

Returning to (C.17) and using the fact that u_{it} is iid over i, we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_{jt}-u_{1t})\xrightarrow{d}\mathcal{N}(0,2\gamma_u^2)$ and the asymptotic covariance between $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_{jt}-u_{1t})$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_{\ell t}-u_{1t})$ for $j\neq\ell$ is γ_u^2. It follows that any linear combination of $\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_{jt}-u_{1t})\}_{j=2}^{N}$ is asymptotically normal, which yields joint asymptotic normality of $\frac{1}{\sqrt{T}}M_2^{*\prime}U$ in view of (C.17). For fixed N, we therefore have

$$\frac{1}{\sqrt{T}}M_2^{*\prime}U\xrightarrow{d}\mathcal{N}(0,\gamma_u^2(I_{N-1}+1_{N-1}1_{N-1}')). \tag{C.20}$$

When $N\to\infty$, (C.20) continues to hold element wise because (C.17) holds uniformly over $j=2,...,N$.

Combining (C.15) with (C.20), we have for fixed N

$$\sqrt{T}[M_2^{*\prime}M_2^*]^{-1}M_2^{*\prime}U\xrightarrow{d}\mathcal{N}(0,\gamma_u^2(I_{N-1}+1_{N-1}1_{N-1}')^{-1})=\mathcal{N}\left(0,\gamma_u^2\left(I_{N-1}-\frac{1_{N-1}1_{N-1}'}{N}\right)\right). \tag{C.21}$$

The limit theory (C.21) holds element wise uniformly over index $j=2,...,N$ because (C.15) and (C.20) both hold uniformly over all their indices. Moreover, as $N\to\infty$ distinct elements of (C.21) are asymptotically normal and independent each with variance γ_u^2.

Finally we look at the first term in (C.6). We only need to consider $(M_2^*)'W$. Note that the typical row of W is $x_{*,it}'[\beta_*(z_t)-\hat{\beta}_{*,PLC}^{oracle}(z_t)]$, then the $(j-1)$-th row of $(M_2^*)'W$ is

$$\sum_{t=1}^{T}(-1-q_{1t,j-1})x_{*,1t}'[\beta_*(z_t)-\hat{\beta}_{*,PLC}^{oracle}(z_t)]+\sum_{t=1}^{T}(1-q_{jt,j-1})x_{*,jt}'[\beta_*(z_t)-\hat{\beta}_{*,PLC}^{oracle}(z_t)]$$

$$+\sum_{i\neq 1,j}\sum_{t=1}^{T}(-q_{it,j-1})x'_{*,it}[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]$$
$$=\sum_{t=1}^{T}(x'_{*,jt}-x'_{*,1t})[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]-\sum_{i=1}^{N}\sum_{t=1}^{T}q_{it,j-1}x'_{*,it}[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]. \quad (C.22)$$

Since $\mathbb{E}(x'_{*,jt}-x'_{*,1t})=0$, we have $T^{-1/2}\sum_{t=1}^{T}(x'_{*,jt}-x'_{*,1t})=O_p(1)$ uniformly for $j=2,...,N$ in view of stationarity over j. By a uniform convergence extension[19] of the limit theory of the oracle estimator $\sup_{t\leq T}|\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)|=o_p(1)$, the first term in (C.22) is of order $o_p(\sqrt{T})$ uniformly for $j=2,...,N$. Now consider the second term in (C.22), which we show is also $o_p(\sqrt{T})$. In view of (C.9), we have, uniformly over $j=2,...,N$

$$\sum_{i=1}^{N}\sum_{t=1}^{T}q_{it,j-1}x'_{*,it}[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]=\sum_{i=1}^{N}\sum_{t=1}^{T}q'_{t,j}x_{*,it}x'_{*,it}[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]$$
$$=O_p\left(\frac{1}{N}\sqrt{\frac{\log_2(T|H|)}{T|H|}}\right)\sum_{i=1}^{N}\sum_{t=1}^{T}1'_{p+1}x_{*,it}x'_{*,it}[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]. \quad (C.23)$$

In view of (C.2), (C.4) and (C.5), we have $\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)=O_p(||H||^2+\frac{1}{\sqrt{NT|H|}})$ where the $O_p(||H||^2)$ component comes from the bias term and the $O_p(\frac{1}{\sqrt{NT|H|}})$ component comes from the variance term of $[X'_*K_n(z_t)X_*]^{-1}X'_*K_n(z_t)U$. Then

$$\sum_{i=1}^{N}\sum_{t=1}^{T}1'_{p+1}x_{*,it}x'_{*,it}[\beta_*(z_t)-\hat{\beta}^{oracle}_{*,PLC}(z_t)]=O_p(NT||H||^2)+O_p\left(\sqrt{\frac{N}{|H|}}\right) \quad (C.24)$$

where the order $O_p(NT||H||^2)$ element reflects the fact that the bias has non-zero mean, giving rise to the additional factor NT, and the order $O_p\left(\sqrt{\frac{N}{|H|}}\right)$ component reflects the order of the variation term $\sum_{i=1}^{N}\sum_{t=1}^{T}1'_{p+1}x_{*,it}x'_{*,it}[X'_*K_n(z_t)X_*]^{-1}X'_*K_n(z_t)U$. To see this, we first note that this term has zero mean due to the independence between $\{u_{it}\}$ and $\{(x_{it},z_t)\}$. The variance of the component

$$\sum_{t=1}^{T}1'_{p+1}x_{*,it}x'_{*,it}f_z^{-1}(z_t)X'_*K_n(z_t)U=\sum_{t=1}^{T}1'_{p+1}x_{*,it}x'_{*,it}f_z^{-1}(z_t)\sum_{j=1}^{N}\sum_{s=1}^{T}x_{*,js}u_{js}K(H^{-1}(z_s-z_t))$$

is of order $O(NT^2|H|)$, which can be verified in the same way as the standard result that the variance of $X'_*K_n(z)U=\sum_{i=1}^{N}\sum_{t=1}^{T}x_{*,it}u_{it}K(H^{-1}(z_t-z)$ is of order $O(NT|H|)$. It then follows that

$$\sum_{t=1}^{T}1'_{p+1}x_{*,it}x'_{*,it}f_z^{-1}(z_t)X'_*K_n(z_t)U=O_p(T\sqrt{N|H|}),$$

[19] Under suitable conditions on smoothness, bounded densities, and mixing decay rate, uniform convergence may be established using the results of Hansen (2008) for kernel regression with weakly dependent data.

and therefore

$$\sum_{i=1}^{N}\sum_{t=1}^{T} 1'_{p+1}x_{*,it}x'_{*,it}[X'_*K_n(z_t)X_*]^{-1}X'_*K_n(z_t)U = O_p\left(N \times \frac{T\sqrt{N|H|}}{NT|H|}\right) = O_p\left(\sqrt{\frac{N}{|H|}}\right). \tag{C.25}$$

Combining (C.23) and (C.24) we have

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{t=1}^{T} q_{it,j-1}x'_{*,it}[\beta_*(z_t) - \hat{\beta}^{oracle}_{*,PLC}(z_t)] \\
&= O_p\left(\frac{1}{N}\sqrt{\frac{\log_2(T|H|)}{T|H|}}\right)\left[O_p(NT||H||^2) + O_p(\sqrt{\frac{N}{|H|}})\right] \\
&= O_p(\sqrt{T})\left[O_p(\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2) + O_p(\sqrt{\frac{\log_2(T|H|)}{TN|H|}\frac{1}{T|H|}})\right] \\
&= O_p\left(\sqrt{T}\cdot\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2\right) + o_p(\sqrt{T})
\end{aligned} \tag{C.26}$$

where the first term in (C.26) comes from the bias of the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z_t)$. Moreover, (C.26) is true uniformly over $j = 2, ..., N$ because (C.23) holds uniformly over j. Combining (C.15), (C.20), (C.22) and (C.26) in (C.6), we have

$$\hat{\alpha}^*_{PLC} - \alpha^* = O_p\left(\sqrt{\frac{\log_2(T|H|)}{T|H|}}||H||^2\right) + O_p(1/\sqrt{T}) \tag{C.27}$$

where the first order term $O_p(\sqrt{\frac{\log_2(T|H|)}{T|H|}}||H||^2)$ represents bias and the second order term $O_p(1/\sqrt{T})$ represents variation. When $q \le 3$ we have $\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2 \to 0$ and then (C.26) has order $o_p(\sqrt{T})$. It follows that the second term in (C.22), and hence the first term in (C.6), is $o_p(\sqrt{T})$.

Thus, the first term in (C.6) is negligible compared to the second term. In view of (C.21), we then have for fixed N

$$\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) \xrightarrow{d} \mathcal{N}\left(0, \gamma_u^2\left(I_{N-1} - \frac{1}{N}1_{N-1}1'_{N-1}\right)\right), \tag{C.28}$$

and (C.28) holds element wise uniformly over $j = 2, ..., N$. It follows from (C.28) that for fixed N

$$1'_{N-1}\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) \xrightarrow{d} \mathcal{N}(0, \gamma_u^2 1'_{N-1}[I_{N-1} - \frac{1}{N}1_{N-1}1'_{N-1}]1_{N-1}) = \mathcal{N}(0, \gamma_u^2(1 - 1/N)). \tag{C.29}$$

However, $1'_{N-1}\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) = \sqrt{T}\sum_{j=2}^{N}(\hat{\alpha}_{j,PLC} - \alpha_j) = -\sqrt{T}(\hat{\alpha}_{1,PLC} - \alpha_1)$ and so when N is fixed

$$\sqrt{T}(\hat{\alpha}_{1,PLC} - \alpha_1) \xrightarrow{d} \mathcal{N}(0, \gamma_u^2(1 - 1/N)). \tag{C.30}$$

The asymptotic covariance between $\sqrt{T}(\hat{\alpha}_{1,PLC} - \alpha_1)$ and $\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)$ $(j = 2, ..., N)$ is

$$
\begin{aligned}
&AsymCov\left(\sqrt{T}(\hat{\alpha}_{1,PLC} - \alpha_1), \sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)\right) \\
=& - AsymCov\left(\sqrt{T}\sum_{\ell=2}^{N}(\hat{\alpha}_{\ell,PLC} - \alpha_\ell), \sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)\right) \\
=& - AsymVar\left(\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)\right) - AsymCov\left(\sqrt{T}\sum_{\ell\neq j}(\hat{\alpha}_{\ell,PLC} - \alpha_\ell), \sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)\right) \\
=& - \gamma_u^2(1 - 1/N) - \gamma_u^2(-1/N)\cdot(N-2) \\
=& - \gamma_u^2/N. \qquad \text{(C.31)}
\end{aligned}
$$

Combining (C.28) with the asymptotics of $\hat{\alpha}_{1,PLC}$, we deduce that for fixed N

$$\sqrt{T}(\hat{\alpha}_{PLC} - \alpha) \xrightarrow{d} \mathcal{N}\left(0, \gamma_u^2\left(I_N - \frac{1}{N}1_N 1_N'\right)\right). \qquad \text{(C.32)}$$

Since (C.28) holds element wise uniformly over $j = 2, ..., N$ and since (C.30) and (C.31) hold for $\hat{\alpha}_{1,PLC}$, we deduce that (C.32) holds element wise uniformly over $j = 1, ..., N$. It follows that for each $j = 1, ..., N$

$$\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j) \xrightarrow{d} \mathcal{N}(0, \gamma_u^2) \qquad \text{(C.33)}$$

when $N \to \infty$. And for $j \neq \ell$, $\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j)$ is asymptotically independent of $\sqrt{T}(\hat{\alpha}_{\ell,PLC} - \alpha_\ell)$ because the covariance $-\gamma_u^2/N$ goes to zero as $N \to \infty$.

Finally, we mention that the above proof applies to both sequential limit $(N, T)_{seq} \to \infty$ and joint limit $(N, T) \to \infty$. This is because the preliminary results, including the convergences established in Lemmas E.5 and E.6, and hence the established uniform results over $j = 1, ..., N$, remain valid for both sequential and joint joints. For example, (C.9) holds with either $(N, T)_{seq} \to \infty$ or $(N, T) \to \infty$ because Lemma E.5 holds with either $(N, T)_{seq} \to \infty$ or $(N, T) \to \infty$ and (C.8) is a uniform result with respect to j. Then (C.12) and (C.14) continue to hold for both $(N, T)_{seq} \to \infty$ or $(N, T) \to \infty$ and so does the element-wise version of (C.15). The joint convergence of the term $\sum_{i=1}^{N}\sum_{t=1}^{T} x_{*,it}' 1_{p+1} u_{it}$ in (C.18) can be verified in the same way as that in Lemma E.6. Therefore, the element-wise version of (C.20) holds for joint limits and this is also true for (C.21). Similarly, it can be verified that the first term in (C.6) is asymptotically negligible under the joint limit $(N, T) \to \infty$. Therefore, result (C.33) is true with either $(N, T)_{seq} \to \infty$ or $(N, T) \to \infty$.

Part (c) Now consider the feasible estimator $\hat{\beta}_{*,PLC}(z)$. We have

$$
\begin{aligned}
\hat{\beta}_{*,PLC}(z) - \beta_*(z) &= [X_*' K_n(z) X_*]^{-1} X_*' K_n(z)(Y - D^*\hat{\alpha}_{PLC}^*) - \beta_*(z) \\
&= [X_*' K_n(z) X_*]^{-1} X_*' K_n(z)(Y - D^*\alpha^*) - \beta_*(z) + [X_*' K_n(z) X_*]^{-1} X_*' K_n(z) D^*(\alpha^* - \hat{\alpha}_{PLC}^*) \\
&= \hat{\beta}_{*,PLC}^{oracle}(z) - \beta_*(z) + [X_*' K_n(z) X_*]^{-1} X_*' K_n(z) D^*(\alpha^* - \hat{\alpha}_{PLC}^*). \qquad \text{(C.34)}
\end{aligned}
$$

We look at the second term in (C.34). Lemma E.5 have shown that $\frac{1}{NT|H|} X_*' K_n(z) X_* \xrightarrow{p} f_z(z) V_{xx,\eta}$ as

$T \to \infty$ with N either fixed or passing to infinity. Consider the term $X_*' K_n(z) D^*(\alpha^* - \hat{\alpha}^*_{PLC})$. We have

$$\frac{1}{NT|H|} X_*' K_n(z) D^*(\alpha^* - \hat{\alpha}^*_{PLC}) = \begin{pmatrix} 0 \\ \frac{1}{N}\sum_{j=2}^N [\frac{1}{T|H|}\sum_t (x_{jt} - x_{1t}) K_{tH}](\alpha_j - \hat{\alpha}_{j,PLC}) \end{pmatrix}. \tag{C.35}$$

From standard kernel asymptotics we know that $\frac{1}{T|H|}\sum_t (x_{jt} - x_{1t}) K_{tH} \xrightarrow{p} f_z(z)\mathbb{E}(x_{jt} - x_{1t}) = 0$, and $\frac{1}{\sqrt{T|H|}}\sum_t (x_{jt} - x_{1t}) K_{tH} \xrightarrow{d} \mathcal{N}(0, \nu_0^q f_z(z)\mathbb{E}(x_{jt} - x_{1t})(x_{jt} - x_{1t})') = \mathcal{N}(0, 2\nu_0^q f_z(z)\Sigma_{xx})$. Thus

$$\frac{1}{T|H|}\sum_t (x_{jt} - x_{1t}) K_{tH} = O_p(1/\sqrt{T|H|}) \; for \; all \; j = 2, \cdots, N. \tag{C.36}$$

Further, the term $\frac{1}{T|H|}\sum_t (x_{jt} - x_{1t}) K_{tH}$ may be taken out of the summation over j in (C.35). Indeed, by following the same argument as in (C.8), we have

$$\limsup_{T\to\infty} \pm \sqrt{2\frac{T|H|}{\log_2(T|H|)}} \frac{1}{T|H|}\sum_t (x_{jt} - x_{1t}) K_{tH} \leq C \; a.s., \tag{C.37}$$

where C is a constant that is independent of j. We then have

$$\begin{aligned} &\frac{1}{N}\sum_{j=2}^N \left[\frac{1}{T|H|}\sum_t (x_{jt} - x_{1t}) K_{tH}\right](\alpha_j - \hat{\alpha}_{j,PLC}) \\ &= O_p\left(\sqrt{\frac{\log_2(T|H|)}{T|H|}}\right) \times \frac{1}{N}\sum_{j=2}^N (\alpha_j - \hat{\alpha}_{j,PLC}) \\ &= O_p\left(\sqrt{\frac{\log_2(T|H|)}{T|H|}}\right) \times O_p\left(\sqrt{\frac{\log_2(T|H|)}{T|H|}}||H||^2 + \frac{1}{N}\frac{\sqrt{N}}{\sqrt{T}}\right) \\ &= O_p\left(\frac{\log_2(T|H|)}{T|H|}||H||^2\right) + O_p\left(\sqrt{\frac{\log_2(T|H|)}{NT^2|H|}}\right) \qquad \text{(C.38)} \\ &= o_p(||H||^2) + o_p\left(\frac{1}{\sqrt{NT|H|}}\right), \qquad \text{(C.39)} \end{aligned}$$

where the second equality follows because (C.27) holds uniformly over j. Also note that the first term in (C.38) comes from bias and the second term comes from variation. Consequently, the second term of (C.34) is of the same order as (C.38), which is smaller than that of the first term in (C.34) ($O_p(||H||^2 + \frac{1}{\sqrt{NT|H|}})$) as demonstrated in (C.39) because $\frac{\log_2(T|H|)}{T|H|} \to 0$ and $\log_2(T|H|)/T \to 0$ as $T \to \infty$. Therefore, the feasible estimator $\hat{\beta}_{*,PLC}(z)$ is asymptotically equivalent to the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z)$ as long as $T \to \infty$. And this conclusion holds without requiring $q \leq 3$ and $\sqrt{\frac{\log_2(T|H|)}{|H|}}||H||^2 \to 0$. The parameter N may be fixed or may pass to infinity as $T \to \infty$. Further, (C.39) remains valid with either $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$ jointly. This is easy to see because Lemma E.5 holds under both sequential and joint limit, and both (C.27) and (C.37) are uniform results with respect to j. Therefore, the second term in (C.34) is negligible compared to the first term and the asymptotic equivalence remains valid under both sequential $(N,T)_{seq} \to \infty$ and joint $(N,T) \to \infty$ limits. ■

D Proof of Theorem 3.2

For convenience, we first outline the estimation procedure under the null $\mathcal{H}_0$. Let $\beta_* = (\beta_0, \beta')'$. Under $\mathcal{H}_0$, model (1.2) becomes

$$y_{it} = \alpha_i + \beta_0 + x_{it}'\beta + u_{it}, i = 1, \cdots, N, t = 1, \cdots, T. \tag{D.1}$$

Taking averages over t, we get

$$y_{iA} = \alpha_i + \beta_0 + x_{iA}'\beta + u_{iA}, i = 1, \cdots, N. \tag{D.2}$$

Subtracting (D.2) from (D.1) gives

$$y_{it} - y_{iA} = (x_{it}' - x_{iA}')\beta + u_{it} - u_{iA}, i = 1, \cdots, N, t = 1, \cdots, T, \tag{D.3}$$

and then

$$\hat{\beta}_{OLS} = \left[\sum_{i,t}(x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1}\left[\sum_{i,t}(x_{it} - x_{iA})(y_{it} - y_{iA})\right]. \tag{D.4}$$

Plugging $\hat{\beta}_{OLS}$ into (D.1) and noting that $\sum_i \alpha_i = 0$ from the identification condition, we have

$$\hat{\beta}_{0,OLS} = \frac{1}{NT}\sum_{i,t}(y_{it} - x_{it}'\hat{\beta}_{OLS}). \tag{D.5}$$

Next we study the properties of $\hat{\beta}_{*,OLS} = (\hat{\beta}_{0,OLS}, \hat{\beta}_{OLS}')'$ under the alternative $\mathcal{H}_1^L$. Under $\mathcal{H}_1^L$, we have

$$y_{it} = \alpha_i + \beta_0 + \rho_n g_0(z_t) + x_{it}'[\beta + \rho_n g(z_t)] + u_{it}. \tag{D.6}$$

Averaging (D.6) over t gives

$$y_{iA} = \alpha_i + \beta_0 + \rho_n T^{-1}\sum_t g_0(z_t) + x_{iA}'\beta + \rho_n T^{-1}\sum_t x_{it}'g(z_t) + u_{iA}. \tag{D.7}$$

Subtracting (D.7) from (D.6), we have

$$y_{it} - y_{iA} = \rho_n[g_0(z_t) - T^{-1}\sum_t g_0(z_t)] + (x_{it} - x_{iA})'\beta + \rho_n\left[x_{it}'g(z_t) - T^{-1}\sum_t x_{it}'g(z_t)\right] + u_{it} - u_{iA}. \tag{D.8}$$

Then, under the alternative $\mathcal{H}_1^L$ the OLS estimator in (D.4) is

$$\begin{aligned}\hat{\beta}_{OLS} - \beta &= \left[\sum_{i,t}(x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1}\left[\sum_{i,t}(x_{it} - x_{iA})(y_{it} - y_{iA})\right] - \beta \\ &= \rho_n\left[\sum_{i,t}(x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1}\left[\sum_{i,t}(x_{it} - x_{iA})\left(g_0(z_t) - T^{-1}\sum_t g_0(z_t)\right)\right]\end{aligned}$$

$$+\rho_n \left[\sum_{i,t}(x_{it}-x_{iA})(x_{it}-x_{iA})'\right]^{-1}\left[\sum_{i,t}(x_{it}-x_{iA})\left(x_{it}'g(z_t)-T^{-1}\sum_t x_{it}'g(z_t)\right)\right]$$

$$+\left[\sum_{i,t}(x_{it}-x_{iA})(x_{it}-x_{iA})'\right]^{-1}\left[\sum_{i,t}(x_{it}-x_{iA})(u_{it}-u_{iA})\right]. \tag{D.9}$$

Next we analyze the orders of the three terms on the RHS of (D.9). First, the signal moment matrix $\sum_{i,t}(x_{it}-x_{iA})(x_{it}-x_{iA})'=O_p(n)$ and is positive definite uniformly for large enough N and T because $\Sigma_{xx}>0$ and the $\{x_{it}x_{it}'\}_{i,t}$ are uniformly integrable because $\mathbb{E}(||x_{it}||^{2(2+\lambda)})<C<\infty$ (with $\lambda>0$) by Assumption 2 (b). Let $\xi_{it}=(x_{it}-x_{iA})(g_0(z_t)-T^{-1}\sum_t g_0(z_t))$. We will show $\mathbb{E}\xi_{it}\xi_{it}'<\infty$. For notational ease suppose x_{it} is univariate, in which case

$$\mathbb{E}\xi_{it}^2=\mathbb{E}(x_{it}-x_{iA})^2\left(g_0(z_t)-T^{-1}\sum_t g_0(z_t)\right)^2\leq\sqrt{\mathbb{E}[(x_{it}-x_{iA})^4]\mathbb{E}\left[\left(g_0(z_t)-T^{-1}\sum_t g_0(z_t)\right)^4\right]}<\infty$$

uniformly in $\{i,t\}$ because x_{it} has finite fourth moment from Assumption 2 (b) and $g_0(z)$ is uniformly bounded. Therefore, $\xi_{it}=O_p(1)$ uniformly in $\{i,t\}$ and so $\sum_{i,t}\xi_{it}=O_p(n)$. The first term on the right side of (D.9) is then $O_p(\rho_n)$. Analysis of the second term is similar. Let $\eta_{it}=(x_{it}-x_{iA})(x_{it}'g(z_t)-T^{-1}\sum_t x_{it}'g(z_t))$. We can similarly show that $\mathbb{E}\eta_{it}\eta_{it}'<\infty$ since $\mathbb{E}[x_{it}g(z_t)-T^{-1}\sum_t x_{it}g(z_t)]^4\leq C\times\mathbb{E}[x_{it}-T^{-1}\sum_t x_{it}]^4<\infty$ for univariate x_{it}. Therefore $\eta_{it}=O_p(1)$ uniformly in $\{i,t\}$ and the second term on the right side of (D.9) is $O_p(\rho_n)$. Analysis of the third term is standard. By virtue of independence, $\mathbb{E}(x_{it}-x_{iA})(u_{it}-u_{iA})=0$ and $\sum_{it}(x_{it}-x_{iA})(u_{it}-u_{iA})=O_p(\sqrt{n})$. Consequently the third term is $O_p(1/\sqrt{n})$. Combining these results gives

$$\hat\beta_{OLS}-\beta=O_p(1/\sqrt{n}+\rho_n) \tag{D.10}$$

under the local alternative $\mathcal{H}_1^L$.

Next consider the OLS estimator $\hat\beta_{0,OLS}$ given in (D.5). We have

$$\begin{aligned}\hat\beta_{0,OLS}-\beta_0=&\frac{1}{NT}\sum_{i,t}(y_{it}-x_{it}'\hat\beta_{OLS})-\beta_0\\ =&\rho_n\frac{1}{T}\sum_t g_0(z_t)+\frac{1}{n}\sum_{i,t}x_{it}'[\beta-\hat\beta_{OLS}]+\rho_n\frac{1}{n}\sum_{i,t}x_{it}'g(z_t)+\frac{1}{n}\sum_{it}u_{it}.\end{aligned} \tag{D.11}$$

The first term on the RHS of (D.11) is $O_p(\rho_n)$ since $g_0(z)$ is uniformly bounded. The second term has the same order as $\hat\beta_{OLS}-\beta$, which is $O_p(\rho_n+1/\sqrt{n})$. The third term is $O_p(\rho_n)$ because $g(z)$ is uniformly bounded. The last term is $O_p(1/\sqrt{n})$. It follows that

$$\hat\beta_{0,OLS}-\beta_0=O_p(1/\sqrt{n}+\rho_n) \tag{D.12}$$

under the local alternative $\mathcal{H}_1^L$. Combining (D.10) and (D.12), we have

$$\hat\beta_{*,OLS}-\beta_*=O_p(1/\sqrt{n}+\rho_n), \tag{D.13}$$

under the local alternative $\mathcal{H}_1^L$.

Since

$$\begin{aligned}\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} &= \hat{\beta}_{*,PLC}(z) - \beta_*(z) + \beta_*(z) - \hat{\beta}_{*,OLS} \\ &= \hat{\beta}_{*,PLC}(z) - \beta_*(z) + \beta_* - \hat{\beta}_{*,OLS} + \rho_n g_*(z),\end{aligned} \tag{D.14}$$

and since the nonparametric estimator $\hat{\beta}_{*,PLC}(z)$ is $\sqrt{NT|H|}$ consistent under the alternative and $g_*(z)$ is bounded, we have

$$\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} = O_p(\rho_n||H||^2) + O_p\left(\frac{1}{\sqrt{n|H|}}\right) + O_p\left(\frac{1}{\sqrt{n}} + \rho_n\right) + O_p(\rho_n) = O_p\left(\frac{1}{\sqrt{n|H|}} + \rho_n\right). \tag{D.15}$$

Consequently,

$$\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(1) + O_p(\sqrt{n|H|}\rho_n), \tag{D.16}$$

where the $O_p(1)$ term is an asymptotically normal random variable.

If $\sqrt{n|H|}\rho_n \to 0$, then in this case (3.2) continues to hold. For such local alternatives we therefore have $I_m^* \xrightarrow{d} \chi^2_{(p+1)m}$ and $J \xrightarrow{d} \mathcal{N}(0,1)$ under $\mathcal{H}_1^L$, and the tests have asymptotically trivial power equal to size .

If $\sqrt{n|H|}\rho_n = O(1)$, then $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(1)$ but is no longer asymptotically normal. The test statistic $I_m^* = O_p(1)$ but no longer follows a $\chi^2_{(p+1)m}$ distribution and has non-trivial local asymptotic power. Also, we no longer have $\mathbb{E}(\delta(z_t)'\delta(z_t)) \to p+1$ as $n \to \infty$. It then follows that $m^{-1}I_m^* - (p+1) = O_p(1)$ and $J = O_p(\sqrt{m})$ in this case.

If $\sqrt{n|H|}\rho_n \to \infty$, then $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(\sqrt{n|H|}\rho_n)$ diverges. It follows that $I_m^* = O_p(n|H|\rho_n^2) \to \infty$ and $J = O_p(n|H|\rho_n^2/\sqrt{m}) \to \infty$ if $\sqrt{n|H|}\rho_n/\sqrt{(m)} \to \infty$. So the tests are asymptotically powerful and consistent in this case, nesting the fixed alternative where $\rho_n = \rho$ is a constant. ■

E Useful lemmas

Lemma E.1. *Let $P_N = diag(1, \sqrt{N}I_p)$. Under Assumptions 1-3, as $T \to \infty$, we have the following asymptotic forms for the signal matrix*

$$(X_A^*)'K_T(z)X_A^* = \begin{pmatrix} \sum_t K_{tH} & \sum_t K_{tH}x'_{At} \\ \sum_t K_{tH}x_{At} & \sum_t K_{tH}x_{At}x'_{At}\end{pmatrix}.$$

(i) if N is fixed,

$$\frac{1}{T|H|}(X_A^*)'K_T(z)X_A^* \xrightarrow{p} f_z(z)V_{xx,\eta,N},$$

where

$$V_{xx,\eta,N} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix};$$

(ii) if $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$ jointly with $\eta = 0$,

$$\frac{1}{T|H|} P_N (X_A^*)' K_T(z) X_A^* P_N \xrightarrow{p} f_z(z) V_{xx}^*$$

where

$$V_{xx}^* = \begin{pmatrix} 1 & 0 \\ 0 & V_{xx} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}, \text{ when } \eta = 0;$$

(iii) if $(N,T)_{seq} \to \infty$ or $(N,T) \to \infty$ jointly with $\eta \neq 0$,

$$\frac{1}{T|H|} (X_A^*)' K_T(z) X_A^* \xrightarrow{p} f_z(z) V_\eta$$

where

$$V_\eta = \begin{pmatrix} 1 & \eta' \\ \eta & \eta\eta' \end{pmatrix}.$$

Proof

(a) Part (i) N fixed:

Denote $\tilde{x}_t = x_{At}$. Due to cross section independence and stationarity and α-mixing over t, $\tilde{x}_t$ is stationary and α-mixing with the same mixing coefficient α_k, so that results for α-mixing processes can be employed. We have

$$\frac{1}{T|H|}(X_A^*)' K_T(z) X_A^* = \begin{pmatrix} \frac{1}{T|H|}\sum_t K_{tH} & \frac{1}{T|H|}\sum_t K_{tH}\tilde{x}_t' \\ \frac{1}{T|H|}\sum_t K_{tH}\tilde{x}_t & \frac{1}{T|H|}\sum_t K_{tH}\tilde{x}_t\tilde{x}_t' \end{pmatrix} \xrightarrow{p} f_z(z) \begin{pmatrix} 1 & \mathbb{E}(\tilde{x}_t') \\ \mathbb{E}(\tilde{x}_t) & \mathbb{E}(\tilde{x}_t\tilde{x}_t') \end{pmatrix}. \tag{E.1}$$

Further, note that $\mathbb{E}(\tilde{x}_t) = \eta$, $\mathbb{V}ar(x_{it}) = \Sigma_{xx} = V_{xx} - \eta\eta'$, and

$$\mathbb{E}(\tilde{x}_t\tilde{x}_t') = \frac{1}{N^2}\sum_i\sum_j \mathbb{E}x_{it}x_{jt}' = \frac{1}{N^2}\left(\sum_i \mathbb{E}x_{it}x_{it}' + \sum_{i\neq j}\mathbb{E}x_{it}x_{jt}'\right) = \frac{1}{N}V_{xx} + \left(1 - \frac{1}{N}\right)\eta\eta'.$$

Then we have

$$\begin{pmatrix} 1 & \mathbb{E}(\tilde{x}_t') \\ \mathbb{E}(\tilde{x}_t) & \mathbb{E}(\tilde{x}_t\tilde{x}_t') \end{pmatrix} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}V_{xx} + (1-\frac{1}{N})\eta\eta' \end{pmatrix} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix} =: V_{xx,\eta,N},$$

which is close to singular if η is large and the variance of x_{it} is small. Note that (E.1) holds for all fixed N and hence also for large N, in which case $V_{xx,\eta,N}$ is close to singular.

(b) Parts (ii) and (iii) $(N,T)_{\text{seq}} \to \infty$ with $\eta \neq 0$ or $\eta = 0$:
Results for the sequential limit $(N,T)_{\text{seq}} \to \infty$ can be deduced from the results with fixed N by letting N pass to infinity. For the case where $\eta \neq 0$, we have $\lim_{N\to\infty} V_{xx,\eta,N} = V_\eta$, which yields the sequential limit of part (iii). When $\eta = 0$, $\lim_{N\to\infty} V_{xx,\eta,N} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, which is singular. In this case we restandardize the matrix $(X_A^*)'K_T(z)X_A^*$ as $\frac{1}{T|H|}P_N(X_A^*)'K_T(z)X_A^*P_N$. It then follows that

$$\frac{1}{T|H|}P_N(X_A^*)'K_T(z)X_A^*P_N \xrightarrow{p} f_z(z)\begin{pmatrix} 1 & 0 \\ 0 & V_{xx} \end{pmatrix} =: f_z(z)V_{xx}^*, \tag{E.2}$$

which verifies the sequential limit result of part (ii).

(c) Parts (ii) and (iii) $(N,T) \to \infty$ with $\eta \neq 0$ or $\eta = 0$:
Next consider joint limits as $(N,T) \to \infty$ together. We take the (2,2) subblock $\sum_t K_{tH}x_{At}x_{At}'$ to illustrate, treatment of the other elements being simpler. For the case $\eta \neq 0$ we have

$$\begin{aligned}
\frac{1}{T|H|}\sum_t K_{tH}x_{At}x_{At}' &= \frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N}\sum_i x_{it}\frac{1}{N}\sum_j x_{jt}' \\
&= \frac{1}{T|H|}\sum_t K_{tH}\left[\eta + \frac{1}{N}\sum_i (x_{it}-\eta)\right]\left[\eta' + \frac{1}{N}\sum_j (x_{jt}-\eta)'\right] \\
&= \eta\eta'\frac{1}{T|H|}\sum_t K_{tH} + \eta\frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N}\sum_j (x_{jt}-\eta)' + \frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N}\sum_i (x_{it}-\eta)\eta' \\
&+ \frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N}\sum_i (x_{it}-\eta)\frac{1}{N}\sum_j (x_{jt}-\eta)' \\
&= \eta\eta'\frac{1}{T|H|}\sum_t K_{tH} + \eta\frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N}\sum_j (x_{jt}-\eta)' + \frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N}\sum_i (x_{it}-\eta)\eta' \\
&+ \frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N^2}\sum_i (x_{it}-\eta)(x_{it}-\eta)' + \frac{1}{T|H|}\sum_t K_{tH}\frac{1}{N^2}\sum_{i\neq j}\sum_j (x_{it}-\eta)(x_{jt}-\eta)' \\
&=: \eta\eta' M_{1T} + \eta M_{2NT}' + M_{2NT}\eta' + M_{3NT} + M_{4NT},
\end{aligned} \tag{E.3}$$

with definitions of $\{M_{1T}, M_{2NT}, M_{3NT}, M_{4NT}\}$ implied in the final line above. For M_{1T} we have the standard result $M_{1T} \xrightarrow{p} f_z(z)$. In view of the joint weak convergence of $\sqrt{NT|H|}M_{2NT} = \frac{1}{\sqrt{T|H|}}\sum_t K_{tH}\frac{1}{\sqrt{N}}\sum_i (x_{it} - \eta)$ to a normal distribution (established in the same way as S_{1NT} in Lemma E.2), we obtain $M_{2NT} = O_p(1/\sqrt{NT|H|}) = o_p(1)$. For M_{3NT}, we have $NM_{3NT} = \frac{1}{NT|H|}\sum_t\sum_i K_{tH}(x_{it}-\eta)(x_{it}-\eta)' \xrightarrow{p} \mathbb{E}K_{tH}(x_{it}-\eta)(x_{it}-\eta)' = f_z(z)\Sigma_{xx}$ and hence $M_{3NT} = O_p(1/N) = o_p(1)$ as $N \to \infty$. After restandardizing the component M_{4NT}, we have $\sqrt{N^2T|H|}M_{4NT} = O_p(1)$, similar to the analysis of the joint weak convergence of $S_{2NT,2}$ in Lemma E.2, which leads to $M_{4NT} = O_p(1/\sqrt{N^2T|H|}) = o_p(M_{3NT})$. Combining these results, it is evident that when $\eta \neq 0$, $\frac{1}{T|H|}\sum_t K_{tH}x_{At}x_{At}'$ is dominated by the component $\eta\eta' M_{1T}$, thereby giving $\frac{1}{T|H|}\sum_t K_{tH}x_{At}x_{At}' \xrightarrow{p} \eta\eta' f_z(z)$ and establishing joint convergence of the (2,2) subblock in part (iii) as $(N,T) \to \infty$ jointly.

Finally, when $\eta = 0$, we find that M_{3NT} is the leading term of $\frac{1}{T|H|}\sum_t K_{tH}x_{At}x_{At}'$ and in this case we have $\frac{N}{T|H|}\sum_t K_{tH}x_{At}x'_{At} \xrightarrow{p} f_z(z)\Sigma_{xx} = f_z(z)V_{xx}$ as $(N,T)\to\infty$. This proves joint convergence of the (2,2) subblock when $\eta = 0$. Treatment of the remaining elements is straightforward and joint convergence as $(N,T)\to\infty$ is established for case (ii). ■

Lemma E.2. *Under Assumptions 1-3, as $T\to\infty$, we have the following asymptotic forms for*

$$(X_A^*)'K_T(z)U_A = \begin{pmatrix} \sum_t K_{tH}u_{At} \\ \sum_t K_{tH}x_{At}u_{At} \end{pmatrix}, \tag{E.4}$$

(i) if N is fixed,

$$\frac{\sqrt{N}}{\sqrt{T|H|}}(X_A^*)'K_T(z)U_A \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q V_{xx,\eta,N});$$

(ii) if $(N,T)_{seq}\to\infty$ or $(N,T)\to\infty$ jointly with $\eta = 0$,

$$\frac{\sqrt{N}}{\sqrt{T|H|}}P_N(X_A^*)'K_T(z)U_A \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q V_{xx}^*);$$

(iii) if $(N,T)_{seq}\to\infty$ or $(N,T)\to\infty$ jointly with $\eta \neq 0$,

$$\frac{\sqrt{N}}{\sqrt{T|H|}}(X_A^*)'K_T(z)U_A \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q V_\eta).$$

Proof

(a) Part (i) N fixed:

Note that $(X_A^*)'K_T(z)U_A$ has zero mean. Denote $\tilde{u}_t = u_{At}$ and note that $\sigma_{\tilde{u}}^2 = \mathbb{V}ar(\tilde{u}_t) = \sigma_u^2/N$. Then, for the first element in (E.4), we have

$$\mathbb{V}ar\left(\frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}u_{At}\right) = \frac{N}{T|H|}\sum_t \mathbb{E}[K_{tH}^2u_{At}^2] + \frac{N}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}K_{sH}u_{As}] \to f_z(z)\nu_0^q\sigma_u^2, \tag{E.5}$$

because

$$\begin{aligned} \frac{N}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}K_{sH}u_{As}] &= \frac{2N}{T|H|}\sum_{\ell=1}^{T-1}(T-\ell)\mathbb{E}[K_{1H}\tilde{u}_1K_{1+\ell,H}\tilde{u}_{1+\ell}] \\ &\sim_a \frac{CN}{|H|}\sum_{\ell=1}^{T-1}(1-\ell/T)\gamma_{\tilde{u}}(\ell)|H|^2 \\ &\leq C|H|\sum_{\ell=1}^{T-1}|\gamma_u(\ell)| = o(1), \end{aligned} \tag{E.6}$$

where the last argument in (E.6) follows from the bound $|\gamma_u(\ell)| \leq C|\alpha_\ell|^{\lambda/(2+\lambda)} = O(\ell^{-\tau\lambda/(2+\lambda)})$, using Davydov's inequality, and the fact that $\sum_{\ell=1}^{T-1}\ell^{-\tau\lambda/(2+\lambda)} < \infty$ since $\tau > (2+\lambda)/\lambda$. For the second element in (E.4), we have

$$\mathbb{V}ar\left(\frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At}\right) = \frac{N}{T|H|}\sum_t \mathbb{E}[K_{tH}^2 x_{At}x_{At}' u_{At}^2] + \frac{N}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}x_{At}x_{As}'K_{sH}u_{As}]$$
$$\to f_z(z)\nu_0^q\sigma_u^2\mathbb{E}[\tilde{x}_t\tilde{x}_t'] = f_z(z)\nu_0^q\sigma_u^2(N^{-1}\Sigma_{xx}+\eta\eta'), \tag{E.7}$$

because

$$\frac{N}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}x_{At}x_{As}'K_{sH}u_{As}] = \frac{2N}{T|H|}\sum_{\ell=1}^{T-1}(T-\ell)\mathbb{E}[K_{1H}\tilde{u}_1\tilde{x}_1K_{1+\ell,H}\tilde{u}_{1+\ell}\tilde{x}_{1+\ell}']$$
$$\leq C|H|N\sum_{\ell=1}^{T-1}|\gamma_{\tilde{u}}(\ell)||\gamma_{\tilde{x}}(\ell)| = O(|H|/N) = o(1). \tag{E.8}$$

Similarly, we can show that the covariance of the two elements in (E.4) satisfies

$$\mathbb{C}ov\left(\frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}u_{At}, \frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At}\right) \to f_z(z)\nu_0^q\sigma_u^2\mathbb{E}[\tilde{x}_t] = f_z(z)\nu_0^q\sigma_u^2\eta. \tag{E.9}$$

Since $\tilde{x}_t$ and $\tilde{u}_t$ remain stationary and α-mixing with the same mixing coefficient α_k, we are able to employ existing CLT results for α-mixing processes. In particular, combining (E.5), (E.7), (E.9) and using a triangular array CLT (Theorem 2.2 of Peligrad and Utev (1997)) for weighted partial sums of the α-mixing process $\{(\tilde{x}_1,\tilde{u}_1),\cdots,(\tilde{x}_T,\tilde{u}_T)\}$ result (i) holds.

(b) Parts (ii) and (iii) $(N,T)_{\text{seq}}\to\infty$ with $\eta\neq 0$ or $\eta = 0$:
We start with the case $\eta\neq 0$ and take the second component $\sum_t K_{tH}x_{At}u_{At}$ to illustrate. Treatment of the first element $\sum_t K_{tH}u_{At}$ is included as a special case. We have

$$\begin{aligned}\frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At} &= \frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}\frac{1}{N}\sum_i u_{it}\frac{1}{N}\sum_j x_{jt}\\ &= \frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}\frac{1}{N}\sum_i u_{it}\left(\eta+\frac{1}{N}\sum_j(x_{jt}-\eta)\right)\\ &= \eta\frac{1}{\sqrt{N}}\sum_i\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}u_{it} + \frac{1}{N}\frac{1}{\sqrt{N}}\sum_i\sum_j\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}u_{it}(x_{jt}-\eta)\\ &= \eta S_{1NT}+S_{2NT}.\end{aligned} \tag{E.10}$$

As $T\to\infty$, we have

$$\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}u_{it} \xrightarrow{d} U_i = \mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q), \tag{E.11}$$

by standard triangular array CLT arguments for kernel weighted mixing processes. Note that the limit normal variates U_i are iid over i because of independence across section. Then $\frac{1}{\sqrt{N}}\sum_i U_i \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q)$ as $N\to\infty$ and so $S_{1NT}\xrightarrow{d}\mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q)$ as $(N,T)_{\text{seq}}\to\infty$. The analysis for

S_{2NT} is similar and takes account of the centering $x_{jt} - \eta$ in this expression. In particular, following arguments as in (E.7) we find that $\mathbb{V}ar(S_{2NT}) = O(1/N)$ by letting $\eta = 0$ on the right side of (E.7). So when $\eta \neq 0$ in (E.10), it transpires that S_{1NT} is the leading term in (E.10). In this case, we have $\frac{\sqrt{N}}{\sqrt{T|H|}} \sum_t K_{tH} x_{At} u_{At} \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2 \nu_0^q \eta\eta')$. Note that S_{1NT} is also the first element $\frac{\sqrt{N}}{\sqrt{T|H|}} \sum_t K_{tH} u_{At}$ on the right side of (E.4), for which (E.5) and (E.9) continue to hold for large N. Combining these results together yields the stated result (iii) under sequential asymptotics with $(N,T)_{\text{seq}} \to \infty$.

Otherwise, when $\eta = 0$, S_{2NT} is the only non-zero term and a different rate of convergence applies. In this case we write

$$\begin{aligned}\sqrt{N} S_{2NT} &= \frac{1}{N} \sum_i \frac{1}{\sqrt{T|H|}} \sum_t K_{tH} u_{it}(x_{it} - \eta) + \frac{1}{N} \sum_{i \neq j} \sum_j \frac{1}{\sqrt{T|H|}} \sum_t K_{tH} u_{it}(x_{jt} - \eta) \\ &=: S_{2NT,1} + S_{2NT,2}. \end{aligned} \tag{E.12}$$

By standard triangular array CLT arguments we have $\frac{1}{\sqrt{T|H|}} \sum_t K_{tH} u_{it}(x_{it} - \eta) \xrightarrow{d} \mathcal{N}(0, f_z(z)\nu_0^q \sigma_u^2 \Sigma_{xx}) =: U_i^*$ and $\frac{1}{\sqrt{T|H|}} \sum_t K_{tH} u_{it}(x_{jt} - \eta) \xrightarrow{d} \mathcal{N}(0, f_z(z)\nu_0^q \sigma_u^2 \Sigma_{xx}) =: U_{ij}$. Note that the U_i^* are iid over i and the U_{ij} are iid over (i,j) due to cross section independence. Then as $N \to \infty$ we have $\sqrt{N} S_{2NT,1} \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2 \nu_0^q \Sigma_{xx})$ and $S_{2NT,2} \xrightarrow{d} \mathcal{N}(0, f_z(z)\nu_0^q \sigma_u^2 \Sigma_{xx})$. So $S_{2NT,2}$ is the leading term of $\sqrt{N} S_{2NT}$, and thereby also the leading term of $\frac{N}{\sqrt{T|H|}} \sum_t K_{tH} x_{At} u_{At}$ when $\eta = 0$. Therefore in the case where $\eta = 0$, we have $\frac{N}{\sqrt{T|H|}} \sum_t K_{tH} x_{At} u_{At} \xrightarrow{d} \mathcal{N}(0, f_z(z)\nu_0^q \sigma_u^2 \Sigma_{xx}) = \mathcal{N}(0, f_z(z)\nu_0^q \sigma_u^2 V_{xx})$ as $(N,T)_{\text{seq}} \to \infty$. At the same time, (E.9) becomes $\mathbb{C}ov\left(\frac{\sqrt{N}}{\sqrt{T|H|}} \sum_t K_{tH} u_{At}, \frac{N}{\sqrt{T|H|}} \sum_t K_{tH} x_{At} u_{At}\right) \to 0$. Combining this result with $\frac{\sqrt{N}}{\sqrt{T|H|}} \sum_t K_{tH} u_{At} \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2 \nu_0^q)$ yields the stated result (ii) under sequential asymptotics with $(N,T)_{\text{seq}} \to \infty$.

(c) Parts (ii) and (iii) $(N,T) \to \infty$ with $\eta \neq 0$ or $\eta = 0$:

For the joint asymptotics case as $(N,T) \to \infty$, we proceed by application of the joint limit CLT for a double indexed process based on Theorem 2 of Phillips and Moon (1999).

We start with S_{1NT}. Let $U_{i,T} = \frac{1}{\sqrt{T|H|}} \sum_t K_{tH} u_{it}$. Then the $U_{i,T}$ are independent random variables over index i with $\mathbb{E}U_{i,T} = 0$ and $\mathbb{E}U_{i,T}^2 = \Omega_T$, given below in (E.13). From previous analysis and (E.11), we have $\Omega_T \to \Omega \equiv f_z(z)\sigma_u^2 \nu_0^q$ and $U_{i,T} \xrightarrow{d} U_i$, as $T \to \infty$. Further, using (E.5), (E.6) and the boundedness of $|H|$ we have

$$\Omega_T = \mathbb{E}U_{i,T}^2 = \frac{1}{T|H|} \sum_t \mathbb{E}[K_{tH}^2 u_{it}^2] + \frac{1}{T|H|} \sum_{t \neq s} \mathbb{E}[K_{tH} u_{it} K_{sH} u_{is}] \tag{E.13}$$

$$\leq C\sigma_u^2 + C|H| \sum_{\ell=1}^{T-1} |\gamma_u(\ell)| \leq C \sum_{\ell=0}^{\infty} |\gamma_u(\ell)|, \tag{E.14}$$

from which it follows that $\sup_i \sup_T \mathbb{E}U_{i,T}^2 = \sup_T \Omega_T \leq C < \infty$ and the $U_{i,T}$ are uniformly integrable over both i and T. Let $s_{N,T}^2 = N\Omega_T$. Following Theorem 2 of Phillips and Moon (1999), the joint limit theory $\sum_{i=1}^{N} U_{i,T}/s_{N,T} \xrightarrow{d} \mathcal{N}(0,1)$ follows once the Lindeberg condition is established. We have

$$\begin{aligned}
&\lim_{N,T\to\infty} \sum_{i=1}^{N} \mathbb{E}[U_{i,T}^2/s_{N,T}^2 1\{|U_{i,T}/s_{N,T}| > \varepsilon\}] \\
&= \lim_{N,T\to\infty} \Omega_T^{-1}\mathbb{E}[U_{i,T}^2 1\{|U_{i,T}/s_{N,T}| > \varepsilon\}] \\
&= \lim_{N,T\to\infty} \Omega_T^{-1}\mathbb{E}[U_{i,T}^2 1\{|U_{i,T}| > \varepsilon\sqrt{N\Omega_T}\}] = 0. \qquad \text{(E.15)}
\end{aligned}$$

The final line (E.15) follows by uniform integrability of the sequence $\{U_{i,T}\}$, the existence of the limit $\Omega_T \to \Omega = f_z(z)\sigma_u^2\nu_0^q > 0$ and the fact that $\lim_{N,T\to\infty} \varepsilon\sqrt{N\Omega_T} \to \infty$. These results combine to establish the CLT $\frac{1}{s_{N,T}}\sum_{i=1}^{N} U_{i,T} \xrightarrow{d} \mathcal{N}(0,1)$ under joint asymptotics as $(N,T)\to\infty$. Next note that $S_{1NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_{i,T} = \frac{s_{N,T}}{\sqrt{N}}\sum_{i=1}^{N} U_{i,T}/s_{N,T}$ and $\lim_{N,T\to\infty}\frac{s_{N,T}}{\sqrt{N}} \to \sqrt{\Omega}$. The required joint limit theory as $(N,T)\to\infty$,

$$S_{1NT} \xrightarrow{d} \mathcal{N}(0,\Omega), \qquad \text{(E.16)}$$

follows immediately, thereby confirming that the joint limit exists and is the same as the sequential limit of $(N,T)_{\text{seq}} \to \infty$.

The treatment of S_{2NT} requires some extra effort. The component $S_{2NT,1}$ can be dealt with in the same way as S_{1NT} and is omitted. The component $S_{2NT,2}$ has a triple index. To apply the double index CLT of Phillips and Moon (1999), we first re-index the sequence. Specifically, let $U_{ij,T} = \frac{1}{\sqrt{T|H|}}\sum_t K_{tH}u_{it}(x_{jt}-\eta)$ and set

$$\{\xi_{k,T}, k = 1,2,...,N^2-N\} = \{U_{12,T},...,U_{1N,T},U_{21,T},...,U_{2N,T},...,U_{N1,T},...,U_{N(N-1),T}\}.$$

In view of cross section independence the $\xi_{k,T}$ are independent random variables across k. We may then rewrite $S_{2NT,2}$ as a double indexed process $S_{2NT,2} = \frac{1}{N}\sum_{k=1}^{N^2-N}\xi_{k,T} = \frac{\sqrt{N^2-N}}{N}\frac{1}{\sqrt{N^2-N}}\sum_{k=1}^{N^2-N}\xi_{k,T}$. The remaining treatment is analogous to that of S_{1NT} and the joint limit result $S_{2NT,2} \xrightarrow{d} \mathcal{N}(0, f_z(z)\nu_0^q\sigma_u^2\Sigma_{xx})$ follows as $(N,T)\to\infty$.

When $\eta = 0$, the leading term in (E.10) is S_{2NT} and in this case we have $\frac{N}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At} = S_{2NT,2} + o_p(1) \xrightarrow{d} \mathcal{N}(0, f_z(z)\nu_0^q\sigma_u^2 V_{xx})$ as $(N,T)\to\infty$. Combined with the zero limiting covariance result in (E.9) when $\eta = 0$ the joint limit theory as $(N,T)\to\infty$ holds for case (ii).

When $\eta \neq 0$, the leading term in (E.10) is S_{1NT} as earlier and in this case we have

$$\begin{aligned}
\frac{\sqrt{N}}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At} &= \eta\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}(\sqrt{N}u_{At}) + o_p(1) \\
&= \eta\frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_{i,T} + o_p(1) \xrightarrow{d} \mathcal{N}(0, f_z(z)\sigma_u^2\nu_0^q\eta\eta'),
\end{aligned}$$

as $(N,T)\to\infty$ jointly, just as in (E.16). Combining this result with (E.16) for the standardized first element of (E.4) and with the limiting covariance (E.9), the required joint limit theory for case (iii) follows. ■

Lemma E.3. *Under Assumptions 1-3, as $T \to \infty$, for the first term in the RHS of (A.3), the following asymptotic form*

$$[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' X_A^* \mathbb{E}\xi_{\beta t}^* \sim_a \mu_2 f_z^{-1}(z) \sum_{j=1}^{q} h_j^2 [\frac{\partial \beta_*(z)}{\partial z_j} \frac{\partial f_z(z)}{\partial z_j} + f_z(z) \frac{1}{2} \frac{\partial^2 \beta_*(z)}{\partial^2 z_j}] =: \mathcal{B}(z) \quad \text{(E.17)}$$

remains valid with N either fixed or passing to infinity via sequential limit $(N,T)_{seq} \to \infty$ or joint limit $(N,T) \to \infty$ with $\eta = 0$ or $\eta \neq 0$.

Proof Component $(X_A^*)' K_T(z) X_A^*$ has been analyzed in Lemma E.1. The analysis of $(X_A^*)' X_A^*$ is entirely similar to that of $(X_A^*)' K_T(z) X_A^*$. Similar to the results in Lemma E.1 we can get the following results: (i) if N is fixed, $T^{-1}(X_A^*)' X_A^* \xrightarrow{p} V_{xx,\eta,N}$; (ii) if $N \to \infty$ and $\eta = 0$, $T^{-1} P_N (X_A^*)' X_A^* P_N \xrightarrow{p} V_{xx}^*$; (iii) if $N \to \infty$ and $\eta \neq 0$, $T^{-1}(X_A^*)' X_A^* \xrightarrow{p} V_\eta$. Standard calculations give $\mathbb{E}\xi_{\beta t}^* = |H| \mu_2 \sum_{j=1}^{q} h_j^2 [\frac{\partial \beta_*(z)}{\partial z_j} \frac{\partial f_z(z)}{\partial z_j} + f_z(z) \frac{1}{2} \frac{\partial^2 \beta_*(z)}{\partial^2 z_j}] =: |H| \mu_2 \sum_{j=1}^{q} h_j^2 C_j(z) =: |H| \mu_2 C(z)$. Combining all these results, for the first component in the RHS of (A.2), we have:
(i) if N is fixed: Following Lemma E.1 (i) we have

$$\begin{aligned}
&[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' X_A^* \mathbb{E}\xi_{\beta t}^* \\
&= \left[\frac{1}{T|H|} (X_A^*)' K_T(z) X_A^* \right]^{-1} T^{-1} (X_A^*)' X_A^* |H|^{-1} \mathbb{E}\xi_{\beta t}^* \\
&\sim_a [f_z(z) V_{xx,\eta,N}]^{-1} V_{xx,\eta,N} \mu_2 C(z) = \mu_2 f_z^{-1}(z) C(z) =: \mathcal{B}(z). \qquad \text{(E.18)}
\end{aligned}$$

(ii) if $(N,T)_{\text{seq}} \to \infty$ or $(N,T) \to \infty$ jointly with $\eta = 0$: Following Lemma E.1 (ii) we have

$$\begin{aligned}
&[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' X_A^* \mathbb{E}\xi_{\beta t}^* \\
&= P_N \left[\frac{1}{T|H|} P_N (X_A^*)' K_T(z) X_A^* P_N \right]^{-1} T^{-1} P_N (X_A^*)' X_A^* P_N P_N^{-1} |H|^{-1} \mathbb{E}\xi_{\beta t}^* \\
&\sim_a P_N [f_z(z) V_{xx}^*]^{-1} V_{xx}^* P_N^{-1} \mu_2 C(z) = \mu_2 f_z^{-1}(z) C(z) =: \mathcal{B}(z).
\end{aligned}$$

(iii) if $(N,T)_{\text{seq}} \to \infty$ or $(N,T) \to \infty$ jointly with $\eta \neq 0$: For the sequential limit case where $(N,T)_{seq} \to \infty$, we still have $[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' X_A^* \mathbb{E}\xi_{\beta t}^* \sim_a \mathcal{B}(z)$ by letting $N \to \infty$ in (E.18). In the joint limit case where $(N,T) \to \infty$ jointly, we have the singularity problem arising from V_η. In this case, we consider transformed version as shown in (B.2) in the proof of Theorem 2.2. We have

$$\begin{aligned}
&[(X_A^*)' K_T(z) X_A^*]^{-1} (X_A^*)' X_A^* \mathbb{E}\xi_{\beta t}^* \\
&= C_\eta P_N \left(\frac{1}{T|H|} P_N C_\eta' (X_A^*)' K_T(z) X_A^* C_\eta P_N \right)^{-1} T^{-1} P_N C_\eta' (X_A^*)' X_A^* C_\eta P_N P_N^{-1} C_\eta^{-1} |H|^{-1} \mathbb{E}\xi_{\beta t}^* \\
&\sim_a C_\eta P_N (f_z(z) \Sigma_{xx,\eta})^{-1} \Sigma_{xx,\eta} P_N^{-1} C_\eta^{-1} \mu_2 C(z) = \mu_2 f_z^{-1}(z) C(z) =: \mathcal{B}(z).
\end{aligned}$$

Therefore, we have shown that in all the cases considered (E.17) remains to hold. ■

Lemma E.4. *Under Assumptions 1-3 and assume $N \sum_{j=1}^{q} h_j^2 \to 0$, as $T \to \infty$ with N either fixed or* *passing to infinity, we have*

$$\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t} = o_p((X^*_A)'K_T(z)U_A) \tag{E.19}$$

where $\eta^*_{\beta t} = \xi^*_{\beta t} - \mathbb{E}\xi^*_{\beta t}$, $\xi^*_{\beta t} = [\beta_*(z_t) - \beta_*(z)]K_{tH}$.

Proof It is standard to show that $\mathbb{E}\xi^*_{\beta t}\xi^{*\prime}_{\beta t} = O(|H|\sum_{j=1}^q h_j^2)$ and hence $\eta^*_{\beta t} = O_p(\sqrt{|H|\sum_{j=1}^q h_j^2})$.
(i)When N is fixed, we have $\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t} = O_p(\sqrt{T|H|\sum_{j=1}^q h_j^2})$ following invariance principle for stationary mixing time series. In view of Lemma E.2 we have $(X^*_A)'K_T(z)U_A = O_p(\sqrt{T|H|}/\sqrt{N})$ when N is fixed. Then (E.19) follows in a straightforward way.
(ii) When $N \to \infty$ and $\eta = 0$, we verify (E.19) element-wisely. Following Lemma E.2(ii) we have

$$(X^*_A)'K_T(z)U_A = \begin{pmatrix} \sum_t K_{tH}u_{At} \\ \sum_t K_{tH}x_{At}u_{At} \end{pmatrix} = \begin{pmatrix} O_p(\sqrt{\frac{T|H|}{N}}) \\ O_p(\sqrt{\frac{T|H|}{N^2}}) \end{pmatrix}.$$

As in the proof of Lemma E.2, we can justify that invariance principle continues to hold for $\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t}$ upon appropriate standardization, Then it is easy to get

$$\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t} = \sum_t \begin{pmatrix} \eta_{0\beta t} + x'_{At}\eta_{1\beta t} \\ x_{At}\eta_{0\beta t} + x_{At}x'_{At}\eta_{1\beta t} \end{pmatrix} = \begin{pmatrix} O_p(\sqrt{T|H|\sum_{j=1}^q h_j^2}) + O_p(\sqrt{\frac{T|H|\sum_{j=1}^q h_j^2}{N}}) \\ O_p(\sqrt{\frac{T|H|\sum_{j=1}^q h_j^2}{N}}) + O_p(\sqrt{\frac{T|H|\sum_{j=1}^q h_j^2}{N^2}}) \end{pmatrix}.$$

Given the bandwidth condition that $N\sum_{j=1}^q h_j^2 \to 0$, we can see the components of $\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t}$ is of smaller order than that of $(X^*_A)'K_T(z)U_A$. As a result, (E.19) holds.
(iii) If $N \to \infty$ and $\eta \neq 0$, the analysis is the same with that in the case of $\eta = 0$. Following Lemma E.2(iii) we have

$$(X^*_A)'K_T(z)U_A = \begin{pmatrix} \sum_t K_{tH}u_{At} \\ \sum_t K_{tH}x_{At}u_{At} \end{pmatrix} = \begin{pmatrix} O_p(\sqrt{\frac{T|H|}{N}}) \\ O_p(\sqrt{\frac{T|H|}{N}}) \end{pmatrix}.$$

Furthermore,

$$\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t} = \sum_t \begin{pmatrix} \eta_{0\beta t} + x'_{At}\eta_{1\beta t} \\ x_{At}\eta_{0\beta t} + x_{At}x'_{At}\eta_{1\beta t} \end{pmatrix} = \begin{pmatrix} O_p(\sqrt{T|H|\sum_{j=1}^q h_j^2}) \\ O_p(\sqrt{T|H|\sum_{j=1}^q h_j^2}) \end{pmatrix}.$$

With $N\sum_{j=1}^q h_j^2 \to 0$ the elements of $\sum_t x_{*,At}x'_{*,At}\eta^*_{\beta t}$ are of smaller order than that of $(X^*_A)'K_T(z)U_A$. This completes the proof of (E.19). ■

Lemma E.5. *Under Assumptions 1-3, as* $T \to \infty$*, we have*

$$\frac{1}{NT|H|}X'_*K_n(z)X_* \xrightarrow{p} f_z(z)V_{xx,\eta}, \tag{E.20}$$

where

$$V_{xx,\eta} = \begin{pmatrix} 1 & \eta' \\ \eta & V_{xx} \end{pmatrix}.$$

Proof Note that

$$\frac{1}{NT|H|} X_*' K_n(z) X_* = \frac{1}{NT|H|} \sum_i \sum_t \begin{pmatrix} K_{tH} & x_{it}' K_{tH} \\ x_{it} K_{tH} & x_{it} K_{tH} x_{it}' \end{pmatrix}.$$

We only need to show element-wise convergence, which is entirely analogous to that of Lemma E.1, which covers limits as $T \to \infty$ under fixed N, sequential convergence $(N,T)_{seq} \to \infty$ and joint convergence $(N,T) \to \infty$. Details are omitted. ■

Lemma E.6. *Under Assumptions 1-3, as $T \to \infty$, we have*

$$\frac{1}{\sqrt{NT|H|}} X_*' K_n(z) U \xrightarrow{d} \mathcal{N}\left(0, f_z(z) \sigma_u^2 \nu_0^q V_{xx,\eta}\right). \tag{E.21}$$

Proof Note that

$$\frac{1}{\sqrt{NT|H|}} X_*' K_n(z) U = \frac{1}{\sqrt{NT|H|}} \sum_i \sum_t \begin{pmatrix} K_{tH} u_{it} \\ x_{it} K_{tH} u_{it} \end{pmatrix}. \tag{E.22}$$

Analysis of the two components on the RHS of (E.22) is included in the proof of Lemma E.2, which again covers limits under fixed N as $T \to \infty$, sequential convergence $(N,T)_{seq} \to \infty$ and joint convergence $(N,T) \to \infty$. The first component $\frac{1}{\sqrt{NT|H|}} \sum_i \sum_t K_{tH} u_{it}$ is S_{1NT} defined in (E.10). The second component $\frac{1}{\sqrt{NT|H|}} \sum_i \sum_t x_{it} K_{tH} u_{it}$ can be treated in the same way as $S_{2NT,1}$ defined in (E.12). Details are omitted. ■

Lemma E.7. *Under Assumptions 1-3, as $T \to \infty$, we have*

$$[X_*' K_n(z) X_*]^{-1} X_*' X_* \mathbb{E} \xi_{\beta t}^* \sim_a \mathcal{B}(z). \tag{E.23}$$

Proof This is a straightforward application of Lemma E.5 and the derivations in Lemma E.3. Details are omitted. ■

Lemma E.8. *Under Assumptions 1-3, as $T \to \infty$, we have*

$$\sum_{i,t} x_{*,it} x_{*,it}' \eta_{\beta t}^* = o_p(X_*' K_n(z) U). \tag{E.24}$$

Proof Following Lemma E.6 we have $X_*' K_n(z) U = O_p(\sqrt{NT|H|})$. Similar to the derivations in Lemma E.4, we can show $\sum_{i,t} x_{*,it} x_{*,it}' \eta_{\beta t}^* = O_p(\sqrt{NT|H| \sum_{j=1}^q h_j^2})$. Then (E.24) follows in a straightforward way. ■

F Bootstrap Procedure

We describe the bootstrap procedure for implementing the I_m^* test.

Step (i). For the observed sample $\{y_{it}, x_{it}, z_t\}$, $i = 1, ..., N, t = 1, ..., T$, obtain the PLC estimates $\hat{\alpha}_{PLC}$, $\hat{\beta}_{*,PLC}(z)$ at the observations $\{z_t\}_{t=1}^T$ and the selected grid points $\{z_s^*\}_{s=1}^m$, and the OLS estimator $\hat{\beta}_{*,OLS}$. Compute the residual $\hat{u}_{it} = y_{it} - \hat{\alpha}_{i,PLC} - x'_{*,it}\hat{\beta}_{*,PLC}(z_t)$, $\hat{\sigma}_u^2 = n^{-1}\sum_{i,t}\hat{u}_{it}^2$, and the estimator $\hat{\Omega}(z)$ at the grid points $\{z_s^*\}_{s=1}^m$. With these estimates, compute the statistic I_m^* as in (3.4).

Step (ii). Generate $\tilde{u}_{it}$ independently from $N(0, \hat{\sigma}_u^2)$. Compute $\tilde{y}_{it} = \hat{\alpha}_{i,PLC} + x'_{*,it}\hat{\beta}_{*,OLS} + \tilde{u}_{it}$. Compute the bootstrap statistic $\tilde{I}_m^*$ as in *Step (i)*, using the bootstrap sample $\{\tilde{y}_{it}, x_{it}, z_t\}$, $i = 1, ..., N, t = 1, ..., T$.

Step (iii). Repeat Step (ii) a large number of times, say $B = 200$, and use the upper δ-percentile of the bootstrap statistic empirical distribution $\{\tilde{I}_m^{*(b)}\}_{b=1}^B$, c_δ, to approximate the upper δ-percentile critical value of the null distribution of I_m^*.

Step (iv). Reject the null hypothesis if $I_m^* > c_\delta$, or if the p-value$=\frac{1}{B}\sum_{b=1}^B 1(\tilde{I}_m^{*(b)} > I_m^*) < \delta$. Otherwise, the null is not rejected.

Supplementary Material A technical supplement, Phillips and Wang (2020a), to this paper is available online. This document deals with the case where a functional intercept term is not included in the model and provides a full analysis of this case in parallel to the main paper. To view this supplementary material please visit (DOI details and URL to be provided).

References

Auerbach, A. J. and Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Berkes, I., Hörmann, S., and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic processes and their applications*, 119(4):1298–1324.

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

Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.

Chen, J., Gao, J., Li, D., and Lin, Z. (2015). Specification testing in nonstationary time series models. *The Econometrics Journal*, 18(1):117–136.

Dedecker, J., Merlevède, F., Rio, E., et al. (2013). Strong approximation results for the empirical process of stationary sequences. *The Annals of Probability*, 41(5):3658–3696.

Feng, G., Gao, J., Peng, B., and Zhang, X. (2017). A varying-coefficient panel data model with fixed effects: Theory and an application to US commercial banks. *Journal of Econometrics*, 196(1):68–82.

Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory*, 25(6):1869–1892.

Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics*, 37(6B):3893–3928.

Hall, P. (1991). On iterated logarithm laws for linear arrays and nonparametric regression estimators. *The Annals of Probability*, 19(2):740–757.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Hardle, W. (1984). A law of the iterated logarithm for nonparametric regression function estimators. *The Annals of Statistics*, pages 624–635.

Huang, Y., Chen, X., and Wu, W. B. (2014). Recursive nonparametric estimation for time series. *IEEE Transactions on Information Theory*, 60(2):1301–1312.

Lee, J. and Robinson, P. M. (2015). Panel nonparametric regression with fixed effects. *Journal of Econometrics*, 188(2):346–362.

Li, D., Chen, J., and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, 14(3):387–408.

Li, D. and Sun, Y. (2019). Investment regressions with mismeasured Q, endogenous regressors, and nonlinearities. *working paper*.

MacDuffee, C. C. (2012). *The theory of matrices*, volume 5. Springer Science & Business Media.

Peligrad, M. and Utev, S. (1997). Central limit theorem for linear processes. *The Annals of Probability*, 25(1):443–456.

Phillips, P. C. B. (2001). Descriptive econometrics for non-stationary time series with empirical illustrations. *Journal of Applied Econometrics*, 16(3):389–413.

Phillips, P. C. B. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Phillips, P. C. B. and Wang, Y. (2019). Functional coefficient panel modeling with communal smoothing covariates. *Cowles Foundation Discussion Paper, No. 2193, Yale University.*

Phillips, P. C. B. and Wang, Y. (2020a). Functional coefficient panel modeling with communal smoothing covariates: the case without intercept. *Working paper, Yale University.*

Phillips, P. C. B. and Wang, Y. (2020b). When bias contributes to variance: True limit theory in functional coefficient cointegrating regression. *Cowles Foundation Discussion Paper, No. 2250, Yale University.*

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics*, 169(1):4–14.

Rodriguez-Poo, J. M. and Soberón, A. (2015). Nonparametric estimation of fixed effects panel data varying coefficient models. *Journal of Multivariate Analysis*, 133:95–122.

Stute, W. (1982). A law of the logarithm for kernel density estimators. *The Annals of Probability*, 10:414–422.

Su, L. and Ullah, A. (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. *Handbook of Empirical Economics and Finance*, pages 455–497.

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

Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(3):710–738.

Wang, Q. and Phillips, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, 77(6):1901–1948.

Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, 40(2):727–758.

Table 1: $AMSE$ of the estimators of the intercept coefficient $\beta_0(z)$

		η	$\hat{\beta}_{0,APLC}(z)$	$\hat{\beta}^{oracle}_{0,PLC}(z)$	$\hat{\beta}_{0,PLC}(z)$
$N=5$	$T=20$	0	0.0431	0.0371	0.0373
		1	0.1593	0.0524	0.0531
		5	1.1456	0.0954	0.0969
	$T=50$	0	0.0181	0.0165	0.0165
		1	0.0519	0.0237	0.0238
		5	0.2881	0.0402	0.0405
	$T=100$	0	0.0106	0.0098	0.0098
		1	0.0276	0.0132	0.0132
		5	0.1264	0.0222	0.0224
$N=50$	$T=20$	0	0.0163	0.0064	0.0064
		1	0.4648	0.0077	0.0078
		5	9.1880	0.0112	0.0114
	$T=50$	0	0.0080	0.0031	0.0031
		1	0.1077	0.0040	0.0041
		5	2.0186	0.0055	0.0056
	$T=100$	0	0.0049	0.0017	0.0017
		1	0.0450	0.0023	0.0023
		5	0.8907	0.0031	0.0031

Table 2: $AMSE$ of the estimators of the slope coefficient $\beta_1(z)$

		η	$\hat{\beta}_{1,APLC}(z)$	$\hat{\beta}^{oracle}_{1,PLC}(z)$	$\hat{\beta}_{1,PLC}(z)$
$N=5$	$T=20$	0	0.2038	0.0355	0.0375
		1	0.1397	0.0252	0.0263
		5	0.0699	0.0121	0.0123
	$T=50$	0	0.0715	0.0169	0.0171
		1	0.0503	0.0125	0.0127
		5	0.0284	0.0073	0.0074
	$T=100$	0	0.0384	0.0103	0.0104
		1	0.0279	0.0076	0.0076
		5	0.0164	0.0051	0.0051
$N=50$	$T=20$	0	0.4886	0.0066	0.0068
		1	0.4571	0.0048	0.0049
		5	0.3846	0.0032	0.0032
	$T=50$	0	0.1347	0.0035	0.0036
		1	0.1140	0.0029	0.0029
		5	0.0961	0.0021	0.0021
	$T=100$	0	0.0613	0.0022	0.0022
		1	0.0499	0.0018	0.0018
		5	0.0465	0.0014	0.0014

Table 3: Size (in percentage) of the two tests I_m^* and J with bandwidth formula $h = \hat{\sigma}_z n^{-1/4}$ and nominal size 5%

		$N = 5$		$N = 20$	
		I_m^*	J	I_m^*	J
$m = 3$	$T = 20$	5.00	5.50	2.00	3.50
	50	2.50	3.00	2.00	2.50
	100	2.00	2.50	3.50	5.50
$m = 9$	$T = 20$	6.50	8.00	2.50	3.00
	50	2.00	3.50	2.50	2.50
	100	1.50	2.50	1.50	2.00
$m = 20$	$T = 20$	12.50	12.50	5.50	6.50
	50	10.50	11.00	6.50	6.50
	100	3.50	5.00	4.00	4.50

Table 4: Size (in percentage) of the bootstrapped procedure with bandwidth formula $h = c_h \hat{\sigma}_z n^{-1/4}$ and nominal size 5%

		$N = 5$			$N = 20$		
	c_h	0.5	1.0	1.5	0.5	1.0	1.5
$m = 3$	$T = 20$	2.0	6.0	5.5	5.0	7.5	1.5
	50	4.5	6.5	4.5	4.5	3.5	4.0
	100	6.0	3.0	2.0	5.0	6.0	4.0
$m = 9$	$T = 20$	4.0	5.0	6.5	3.0	5.5	6.0
	50	5.0	4.5	7.0	7.0	4.0	6.0
	100	2.5	4.0	3.0	2.5	4.0	3.0
$m = 20$	$T = 20$	4.5	6.0	4.0	5.5	5.5	6.0
	50	7.0	7.5	6.5	3.5	5.5	5.5
	100	4.5	3.0	6.0	6.5	4.0	7.5

Table 5: Out-of-sample forecasting MSE improvement (in percentage) of functional-coefficient model (5.1) over fixed-coefficient linear model (5.2)

	North East	North West	Yorkshire and The Humber	East Midlands	West Midlands	East	London	South East	South West	All
2014	30.52	46.29	23.92	74.56	-34.45	70.98	51.98	83.60	75.89	48.21
2015	47.02	89.13	77.77	-35.02	82.58	-138.91	9.11	-253.80	96.02	50.26
2016	69.44	66.72	88.69	-26.96	15.77	-6.90	21.02	-80.88	45.00	5.03
2017	-29.36	-0.78	58.13	-7.51	29.07	-0.64	-6.22	-134.76	-69.46	-11.84
2018	50.22	-8.00	79.10	31.64	43.84	65.59	0.15	-244.22	57.47	16.40
All	40.69	66.54	59.84	6.82	44.21	2.55	8.79	-37.27	64.89	27.19