FUNCTIONAL COEFFICIENT PANEL MODELING WITH COMMUNAL SMOOTHING COVARIATES

By

Peter C.B. Phillips and Ying Wang

May 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2193

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Functional Coefficient Panel Modeling with Communal Smoothing Covariates*

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

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

May 27, 2019

Abstract

Behavior at the individual level in panels or at the station level in spatial models is often influenced by aspects of the system in aggregate. In particular, the nature of 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 station averaged data the other on the full panel, and their asymptotic properties are derived. Inference regarding the functional coefficient is also considered. The finite sample performance of the proposed estimators and tests are examined by simulation. An empirical spatial model illustration is provided in which the climate sensitivity of temperature to atmospheric CO_2 concentration is studied at both station and global levels.

JEL classification: C14, C23

Keywords: Climate modeling, Communal covariates, Fixed effects, Functional coefficients, Panel data, Spatial modeling.

*Research support from the Kelly Fund and Marsden Grant 16-UOA-239 at the University of Auckland and an LKC Fellowship at Singapore Management University is gratefully acknowledged. Peter C. B. Phillips email: peter.phillips@yale.edu; Ying Wang email: ying.w@auckland.ac.nz

1 Introduction

Decisions taken at the individual consumer or firm level are frequently affected by prevailing macroeconomic influences such as interest rates, inflation, and indices of consumer or business sentiment. Likewise 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 gases 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 model where the z_{it} are discrete covariates.

[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 individual effects α_i. These authors proposed a nonparametric GMM estimation method. Models with endogenous covariates z_{it} for which $\mathbb{E}(u_{it}|z_{it}) \neq 0$ have not yet been analyzed to our knowledge.

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 $\sum_{i=1}^N \alpha_i = 0$. 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, 2019).

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. 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, these authors assume that the coefficients are 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 station averaged version of (1.2) and the other approach works directly with the full panel structure. Differences in the asymptotic behaviors, including convergence rates, of these approaches and their relation to the properties of an oracle estimator are examined. The limit theory is used to construct tests of parametric linear specification against the semiparametric functional coefficient model. In addition and in contrast to Li et al. (2011), we allow x_{it} to have non-zero mean, which leads to new complications of singularities in the asymptotic theory that are resolved in the paper. 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 nature of Earth's climate sensitivity to CO_2 concentrations and the possibility of functional-coefficient global-dependencies in that sensitivity. Strong evidence is found showing the impact of downwelling radiation effects on temperature in addition to greenhouse gas effects. The conclusions of our semiparametric analysis support the linear specifications used in recent climate econometric research by Magnus et al. (2011), Storelvmo et al. (2016), and Phillips et al. (2019).

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 temperature sensitivity to global CO_2 concentrations. Section 6 concludes the paper.

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 individual 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 model.

An alternative way to remove individual 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 contains an additive constant term, 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 station 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 Averaged profile local constant estimation

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

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

or in vector form

$$Y_A = X_A^*\beta_*(z_t) + U_A, \tag{2.4}$$

where X_A^* is $T \times (p+1)$ obtained by stacking the $1 \times (p+1)$ vector $x'_{*,At}$. Using the local constant approach to estimate $\beta_*(z)$ yields

$$\hat{\beta}_{*,APLC}(z) = [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z Y_A, \tag{2.5}$$

where K_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.

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

Assumption 1. *(a) 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 $\int w^2k^2(w)dw = \nu_2$;*

(b) The product kernel $K(v) = \prod_{j=1}^q k(v_j)$, with $v = (v_1, \cdots, v_q)'$, $\int vv'K(v)dv = \mu_2 I_q$, $\int K^2(v)dv = \nu_0^q$, and $\int vv'K^2(v)dv = \nu_2 I_q$.

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

(c) The error process $\{u_{it}\}$ satisfies $\mathbb{E}u_{it} = 0$, $\mathbb{E}u_{it}^2 = \sigma_u^2 < \infty$, $\mathbb{E}|u_{it}|^{2+\lambda} < \infty$, $\sum_{j=0}^{\infty} |\gamma_u(j)| < \infty$ where $\gamma_u(j) = \mathbb{E}(u_{1t}u_{1,t+j})$ and $\gamma_u^2 = \sum_{j=-\infty}^{\infty} \gamma_u(j)$ is the common long-run variance of u_{it};

(d) z_t has probability density $f_z(z)$. $f_z(\cdot)$ and $\beta_*(\cdot)$ have continuous derivatives up to the second order.

Assumption 3. *Define $H = diag(h_1, \cdots, h_q)$, $||H|| = \sqrt{\sum_{j=1}^{q} h_j^2}$ and $|H| = h_1 \cdots h_q$. As $T \to \infty$, $||H|| \to 0$ and $T|H| \to \infty$.*

Remark 2.1. *(i) The product kernel in Assumption 1 (b) is standard in multivariate smoothing. The conditions on the kernel function $k(\cdot)$ in Assumption 1 (a) are commonly used for convenience in proofs and can be relaxed. For example, the compact support condition can be replaced by some restrictions on the tail behaviour of the kernel function.*
(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). The present paper does not extend the asymptotic analysis to the case of nonstationary $\{(x_{it}), i \geq 1\}$ and $\{(z_t)\}$. Such an extension is relevant in many applications, including the climate change application considered later in this paper, but will require different conditions and proofs than those considered here. This is an important line of research and will be considered by the authors in future work. Endogeneity has been ruled out for simplicity as in Sun et al. (2009).
(iii) 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 mean η plays. Denoting $x_{it} = x_{it}^0 + \eta$, where $\mathbb{E}x_{it}^0 = 0$, 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 2.6, 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. Assumption 2 (b) also requires 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 notation.

(iv) Assumption 2 (c) provides standard moment and weak dependence 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 requirements 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 $N \to \infty$ as $T \to \infty$ and whether $\eta = 0$ or $\eta \neq 0$.

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

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

$$\sqrt{NT|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \Rightarrow N(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) \bar{V}_{xx}^{-1}), \tag{2.6}$$

where

$$\bar{V}_{xx} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix}; \tag{2.7}$$

(b) if $N \to \infty$ simultaneously with T and $\eta = 0$,

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

where

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

(c) if $N \to \infty$ simultaneously with T and $\eta \neq 0$,

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

which is a degenerate normal and where

$$\mathcal{B}(z) = f_z^{-1}(z)\mu_2 \sum_{s=1}^{q} h_s^2 \left[\frac{\partial f_z(z)}{\partial z_s}\frac{\partial \beta_*(z)}{\partial z_s} + \frac{1}{2}\frac{\partial^2 \beta_*(z)}{\partial^2 z_s} f_z(z)\right]. \tag{2.10}$$

Remark 2.2. *When N is fixed, it is apparent from the definition of $\bar{V}_{xx}$ in (2.7) that the mean η cannot be too large compared to the variance Σ_{xx}. Otherwise the matrix $\bar{V}_{xx}$ is close to singular. In such cases, the average profile estimator can be very inefficient.*

Remark 2.3. *From the definition of D_N and when $\eta = 0$ and N goes to infinity, 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 $h_1 = h_2 = \cdots = h_q = h$, 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 purposes of simulation we use the bandwidth decay rate $h = O(T^{-1/(q+4)})$ when implementing the APLC estimators in later comparisons with our alternate approach.*

Remark 2.4. *The singularity of the limiting distribution of $\sqrt{T|H|}(\hat{\beta}_{*,APLC}(z) - \beta_*(z))$ in Theorem 2.1 (c) arises from the singularity of $\bar{V}_{xx}$ as $N \to \infty$ and $\eta \neq 0$. Note that $\bar{V}_{xx}$ 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 $x_{*,At}$ asymptotically as $N \to \infty$ to ensure a positive definite limiting sample moment matrix. 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 $\sqrt{N}$. These results match the limit theory given in Theorem 2.1 (b).*

Remark 2.5. *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 $\frac{1}{N}1_{N\times 1}$, 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 another approach based on profiling with the full panel in the following section.*

To develop a complete asymptotic theory in the degenerate case (Theorem 2.1 (c)), we first transform coordinates in the regression (2.4) as follows

$$Y_A = (X_A^* C_\eta)(C_\eta^{-1}\beta_*(z_t)) + U_A \equiv (X_A^* C_\eta)\theta_*(z_t) + U_A, \tag{2.11}$$

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 (2.11) and denote the resulting estimator $\hat{\theta}_{*,APLC}(z)$. This is entirely analogous to the estimation of $\beta_*(z)$ based on (2.4), and the asymptotics follow easily. In Theorem 2.2 below we present the asymptotic distribution of $\hat{\theta}_{*,APLC}(z)$. 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 $\widehat{\eta} = \eta + O_p(1/(NT)^{1/2})$ the results of the following theorem continue to hold.

Theorem 2.2. *Under Assumptions 1-3, as $T \to \infty$ and $N \to \infty$ simultaneously, we have*

$$D_N(\hat{\theta}_{*,APLC}(z) - \theta_*(z) - \mathcal{B}^*(z)) \Rightarrow N(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)(\bar{V}_{xx,\eta})^{-1}), \tag{2.12}$$

where

$$\mathcal{B}^*(z) = f_z^{-1}(z)\mu_2 \sum_{s=1}^{q} h_s^2 \left[\frac{\partial f_z(z)}{\partial z_s}\frac{\partial \theta_*(z)}{\partial z_s} + \frac{1}{2}\frac{\partial^2 \theta_*(z)}{\partial^2 z_s} f_z(z)\right], \tag{2.13}$$

$$\bar{V}_{xx,\eta} = \begin{pmatrix} (1+\eta'\eta)^2 & 0 \\ 0 & \Sigma_{xx} \end{pmatrix}. \tag{2.14}$$

Equivalently,

$$D_N C_\eta^{-1}(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \Rightarrow N(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)(\bar{V}_{xx,\eta})^{-1}). \tag{2.15}$$

Remark 2.6. *Given the definition of D_N, (2.15) 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 $C_\eta^{-1} = \frac{1}{1+\eta'\eta}\begin{pmatrix} 1 & \eta' \\ \eta & (1+\eta'\eta)I_p - \eta\eta' \end{pmatrix} \equiv \frac{1}{1+\eta'\eta}\begin{pmatrix} c_\eta' \\ C_\eta^* \end{pmatrix}$, where $c_\eta' = (1\ \eta')$, 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 $c_\eta'\hat{\beta}_{*,APLC}(z)$ 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 found in Theorem 2.1 (b). See also Remark 2.3.*

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

$$\hat{\beta}^{oracle}_{*,PLC}(z) = [X_*'K(z)X_*]^{-1}X_*'K(z)(Y - D^*\alpha^*), \tag{2.16}$$

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 $K(z) = I_N \otimes K_z$ is an $NT \times NT$ diagonal matrix.

Let $w_*(z) = [X_*'K(z)X_*]^{-1}X_*'K(z)$ be the $(p+1) \times NT$ coefficient matrix in the oracle estimator (2.16), which may then be written as $\hat{\beta}^{oracle}_{*,PLC}(z) = w_*(z)(Y - D^*\alpha^*)$. Using this estimator in the model (1.2), gives the adjusted equation

$$y_{it} = \alpha_i + x_{*,it}'w_*(z_t)(Y - D^*\alpha^*) + v_{*,it}, \tag{2.17}$$

where $v_{*,it} = y_{it} - \alpha_i - x_{*,it}'w_*(z_t)(Y - D^*\alpha^*)$. To solve for α from (2.17), we first rewrite the equation as

$$\begin{aligned} y_{it} &= \alpha_i + x_{*,it}'w_*(z_t)(Y - D^*\alpha^*) + v_{*,it} \\ &= \alpha_i + \sum_{d=1}^{p+1} x_{it,d-1}e_{*,d}'w_*(z_t)(Y - D^*\alpha^*) + v_{*,it} \\ &= \alpha_i + \sum_{d=1}^{p+1} x_{it,d-1}w_{*,d}(z_t)(Y - D^*\alpha^*) + v_{*,it}, \end{aligned} \tag{2.18}$$

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 (2.18) in vector 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, and Q_1^* is defined by the matrix in square parentheses in the final equation. Consequently, we have

$$(I_{NT} - Q_1^*)Y = (I_{NT} - Q_1^*)D^*\alpha^* + V^*. \tag{2.19}$$

Let $M_1^* = I_{NT} - Q_1^*$ and $M_2^* = M_1^*D^*$. Then least squares regression on (2.19) gives

$$\hat{\alpha}^*_{PLC} = [(M_2^*)'M_2^*]^{-1}(M_2^*)'M_1^*Y, \tag{2.20}$$

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

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

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 $\hat{\beta}^{oracle}_{*,PLC}(z)$ (with N either fixed or passing to infinity),

$$\sqrt{NT|H|}(\hat{\beta}^{oracle}_{*,PLC}(z) - \beta_*(z) - \mathcal{B}(z)) \Rightarrow N(0, f_z^{-1}(z)\sigma_u^2 \nu_0^q \tilde{V}_{xx}^{-1}), \tag{2.22}$$

where $\mathcal{B}(z)$ is defined in (2.10), and

$$\tilde{V}_{xx} = \begin{pmatrix} 1 & \eta' \\ \eta & V_{xx} \end{pmatrix};$$

(b) for finite N, the estimator $\hat{\alpha}^*_{PLC}$ has limit theory

$$\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) \Rightarrow N(0, \gamma_u^2[I_{N-1} - \frac{1}{N} 1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}]),$$

where γ_u^2 is the long-run variance of $\{u_{it}\}$ defined in Assumption 2 (c);

(c) with 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 2.7. ***(Asymptotic equivalence)*** *Theorem 2.3 (c) shows that the feasible estimator $\hat{\beta}_{*,PLC}(z)$ is asymptotically equivalent to the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z)$ irrespective of the T/N ratio, provided $T \to \infty$ and $T|H| \to \infty$. Thus, the result holds for N fixed as well as $N \to \infty$ at a faster or slower rate than T. The key that leads to this equivalence lies in the special structure of the asymptotic variance of $\hat{\alpha}^*_{PLC}$ given in Theorem 2.3 (b). First, for each $j = 2, \cdots, N$, we have $\sqrt{T}(\hat{\alpha}_{j,PLC} - \alpha_j) \Rightarrow N(0, \gamma_u^2(1 - 1/N))$. Then $\hat{\alpha}_{j,PLC} - \alpha_j = O_p(1/\sqrt{T})$ for each j. Note that $1_{1\times(N-1)}\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*)$ is asymptotically normal with covariance matrix $\gamma_u^2 1_{1\times(N-1)}[I_{N-1} - \frac{1}{N}1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}]1_{(N-1)\times 1} \equiv \gamma_u^2 \frac{N-1}{N}$. It follows that $\sum_{j=2}^N (\hat{\alpha}_{j,PLC} - \alpha_j)$ is also of order $O_p(1/\sqrt{T})$, even as N goes to infinity. So the total estimation error introduced by the individual effects remains well controlled and the feasible estimator is asymptotically equivalent to the oracle estimator.*

Remark 2.8. ***(Optimal bandwidth order)*** *Suppose $h_1 = h_2 = \cdots = h_q = h$. Based on the limit theory given in (2.22), the optimal bandwidth order is $h = O((NT)^{-1/(4+q)})$. This rate may seem counter-intuitive given that the smoothing in the nonparametric estimate relates* *only to $\{z_t\}$ whereas the optimal order suggests a smaller bandwidth should be used when N increases while T is fixed. This result is explained by the fact that we are actually using the series $\{z_t\}$ repeatedly in estimation. For each i, we treat $z_{it} \equiv z_t$. From this point of view, the effective sample size is $\sqrt{NT|H|}$. Then it may seem reasonable to let the bandwidth decrease as N increases. However, we need to be more careful when N is large and T is moderate. 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 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 2.9. ***(N/T ratio)*** *We emphasize that the limit theory obtained here requires $T \to \infty$ and $T|H| \to \infty$. These asymptotics do not apply when T is fixed and N goes to infinity*[2]*. The failure is evident from the form of the 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 $h_1 = \cdots = h_q = h$ and the optimal bandwidth order $h = O((NT)^{-1/(q+4)})$ is used. To ensure $T|H| = Th^q \to \infty$, 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 2.10. ***(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 $h = O(T^{-1/5})$. There is no need to repeat the series $\{z_t\}$ N times and use $\frac{1}{N}\sum_i \frac{1}{Th}\sum_t K(h^{-1}(z_t - z))$ with $h = O((NT)^{-1/5})$. Repetition brings no more information in this case.*

Remark 2.11. ***(Asymptotically equivalent estimation of the composite intercept)*** *For the intercept coefficient $\beta_0(z)$, we have two estimators: $\hat{\beta}_{0,APLC}(z)$ with asymptotics given in (2.6) and $\hat{\beta}_{0,PLC}(z)$ with asymptotics given in (2.22). Note that the (1,1) entry of $\bar{V}_{xx}^{-1}$ is $1 + N\eta'\Sigma_{xx}^{-1}\eta$, and the (1,1) entry of $\tilde{V}_{xx}^{-1}$ is $1+\eta'\Sigma_{xx}^{-1}\eta$. Thus these two estimators are asymptotically equivalent when $N = 1$ or $\eta = 0$. This equivalence is verified in simulations. We find that when $\eta = 0$, the APLC estimator has averaged MSE (AMSE) very close to the PLC estimator. When*

[2]This case is currently under investigation by the authors.

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 following way. From (2.15), 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 (2.22), the (1,1) entry of* $(1\ \eta')\tilde{V}_{xx}^{-1}(1\ \eta')'$ *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 although there are two estimators for the composite intercept* $\beta_0(z) + \eta'\beta(z)$*, these estimators are asymptotically equivalent.*

We also note that Lee and Robinson (2015) only 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} *and we have demonstrated that the APLC and PLC estimators are asymptotically equivalent 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 (2.21) to construct the test statistic.

3.1 Test statistic and limit distribution under the null

Under the null that $H_0 : \beta_*(z) = \beta_* = const.\ a.s.$, 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 $\{z_s^*\}_{s=1}^m$[3], 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 $\hat{\beta}_{*,PLC}(z)$. Then under the null hypothesis and based on Theorem 2.3 (a) and (c) we have

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

where $\Omega(z) \equiv f_z^{-1}(z)\sigma_u^2 \nu_0^q \tilde{V}_{xx}^{-1}$.

Further, for any m distinct points $\{z_s^*\}_{s=1}^m$ and with undersmoothing[4] in the construction of the nonparametric estimates $\hat{\beta}_{*,PLC}(z_i^*)$, we have

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

To show this result, we need only verify that the limit covariance matrix is block diagonal, which can be done by showing that $|H|^{-1}\mathbb{E}K(H^{-1}(z_t - z_1^*))K(H^{-1}(z_t - z_2^*)) = o(1)$. For ease of exposition and with no loss of generality assume z_t is univariate. Then

$$\begin{aligned} & h^{-1}\mathbb{E}K(\frac{z_t - z_1^*}{h})K(\frac{z_t - z_2^*}{h}) = h^{-1}\int K(\frac{z_t - z_1^*}{h})K(\frac{z_t - z_2^*}{h})f_z(z_t)dz_t \\ &= \int K(u_1)K(u_1 + \frac{z_1^* - z_2^*}{h})f_z(z_1^* + hu_1)du_1 \\ &= \int K(u_1)K(u_1 + \frac{z_1^* - z_2^*}{h})du_1 f_z(z_1^*) + smaller\ order \\ &\to 0 \ as\ h \to 0, \end{aligned}$$

since $K(u) \to 0$ as $u \to \infty$. This justifies (3.3).

To obtain a suitable pivotal limit theory for the test statistic we normalize the quantity I_m as follows

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

[3] 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.

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

where $\hat{\Omega}(z)$ is a consistent estimate of $\Omega(z)$. It is easy to see that $I_m^* \Rightarrow \chi^2_{(p+1)m}$ under the null, where χ^2_k denotes the χ^2 distribution with k degree of freedom. The multivariate case follows directly.

Theorem 3.1. *Under Assumptions 1-3 and under the null H_0 without undersmoothing as $T \to \infty$, we have $I_m^* \Rightarrow \chi^2_{(p+1)m}$.*

Alternatively, 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 $\delta(z) \Rightarrow N(0, I_{p+1})$, so that $\delta(z)'\delta(z) \Rightarrow \chi^2_{p+1}$. 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 a CLT to $I(m)$, giving as $m \to \infty$,

$$\sqrt{m}(I(m) - \mathbb{E}(\delta(z_t)'\delta(z_t))) \Rightarrow N(0, \Gamma(\delta(z_t)'\delta(z_t))), \tag{3.5}$$

where $\Gamma(v_t)$ denotes the long run variance of the series $\{v_t\}$. Since $\delta(z)'\delta(z) \Rightarrow \chi^2_{p+1}$, we have $\mathbb{E}(\delta(z)'\delta(z)) \to p+1$ as $NT \to \infty$. Due to the asymptotic independence of the coordinates, we have $\Gamma(\delta(z)'\delta(z)) \to Var(\delta(z)'\delta(z)) = 2(p+1)$ as $NT \to \infty$. Then

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

as $NT \to \infty$ and $m \to \infty$. The above asymptotic theory uses sequential asymptotics where $NT \to \infty$ and $|H| \to 0$ followed by $m \to \infty$. It is likely that the result also holds under joint asymptotics with some conditions, possibly 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 $H_1^L : \beta_*(z) = \beta_* + \rho_n g_*(z)$, 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 $(p+1) \times 1$ bounded real vector function that satisfies the continuity condition in Assumption 2 (d). 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 theorem below presents the limit properties of the test statistic I_m^* under the local alternative H_1^L.

Theorem 3.2. *Under Assumptions 1-3 and undersmoothing with $\sqrt{NT|H|}||H||^2 \to 0$, and the local alternative H_1^L, we have the following limit behavior for fixed m as $T \to \infty$,*

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

2. *if* $\sqrt{n|H|}\rho_n = O(1)$, I_m^* *has non-central* $\chi^2_{(p+1)m}$ *distribution;*

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

Remark 3.1. *Theorem 3.2 shows that the performance of the test* I_m^* *depends on the rate at which the sequence* ρ_n *in the localizing function* $\rho_n g_*(z)$ *in the alternative hypothesis converges to zero. For sequences* $\rho_n = O(\frac{1}{\sqrt{n|H|}})$*, test based on* I_m^* *has non-trivial local asymptotic power and when* ρ_n *diminishes at a rate slower than* $O(\frac{1}{\sqrt{n|H|}})$*, the test is consistent, nesting the fixed alternative case where* $\rho_n = \rho$*, some fixed constant.*

Similar results hold for the test statistic J*. In particular, if* $\sqrt{n|H|}\rho_n \to 0$*, we have* $J \Rightarrow N(0,1)$ *as* $(n,m) \to \infty$*. 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* $J = O_p(n|H|\rho_n^2/\sqrt{m})$*.*

We outline below the essentials of the argument in the proof of Theorem 3.2. First we show under the local alternative H_1^L, the property of the OLS estimator $\hat{\beta}_{*,OLS}$ depends on ρ_n[5]. More specifically, we have

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

under the local alternative H_1^L, and the bias $\mathcal{B}_* = O_p(\rho_n)$.

Write $\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} = \hat{\beta}_{*,PLC}(z) - \beta_*(z) + \beta_* - \hat{\beta}_{*,OLS} + \rho_n g_*(z)$, and note that the nonparametric estimate $\hat{\beta}_{*,PLC}(z)$ is always $\sqrt{NT|H|}$-consistent under the alternative with $g_*(z)$ bounded. Then, using undersmoothing to remove the bias of $\hat{\beta}_{*,PLC}(z)$, we have

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

where the bias on the left side has order $\mathcal{B}^* = O_p(\rho_n)$. Therefore,

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

where the first term $O_p(\sqrt{n|H|}\rho_n)$ in this expression arises from bias. The representation (3.9) reveals the asymptotic local power properties of the test.

1. 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 H_1^L and the test has asymptotic power equal to size for such alternatives.

[5] 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 $g_*(z)$, as discussed in Wang and Phillips (2012). Here we only discuss dependence on ρ_n.

2. 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 bias term that is of order $O_p(1)$. Then $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS})$ has a non-central limiting normal distribution and I_m^* is asymptotically non-central $\chi^2_{(p+1)m}$. In this case, the test has non-trivial local asymptotic power.

3. 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 Simulation

This Section reports simulation results on the finite sample performance of the estimation procedures and tests considered in the previous two sections. First, in Section 4.1, we examine the finite sample performance of the two estimators proposed in Section 2. The tests proposed in Section 3 are investigated in Section 4.2.

4.1 Estimation Accuracy

We use the following data generating mechanism

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

where (u_{it}, ξ_{it}, v_i) is i.i.d. $N(\mathbf{0}, diag(1, 1+\eta^2, 1))$, z_t is i.i.d. $U(-1, 1)$. 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 2.8.

The process $\{x_{it}^0\}$ has zero mean and thus $\mathbb{E}x_{it} = \eta$. To avoid the near singularity of the matrix $\bar{V}_{xx}$ defined in (2.7) 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

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

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 2.11. 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 2.3, 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, consistent 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 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) entry of $\bar{V}_{xx}^{-1}$ is an increasing function of η and the (2,2) 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) entry of $\tilde{V}_{xx}^{-1}$ is an increasing function of η and the (2,2) 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[6]. 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,

[6] 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.

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 also related to the failure of the asymptotic theory when $N \to \infty$ with T fixed, as discussed earlier in Remark 2.9.

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

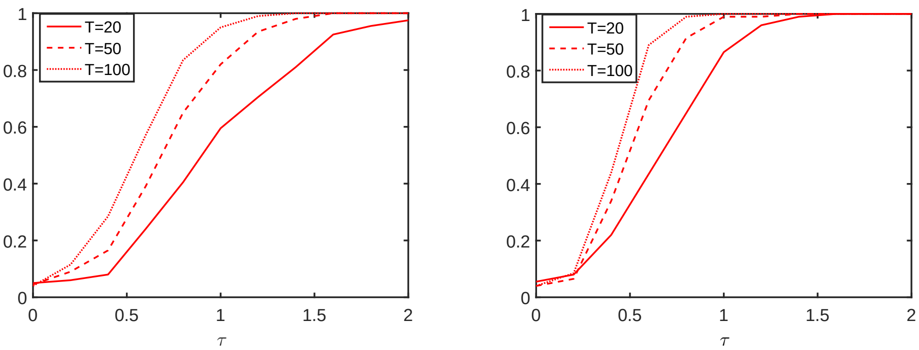
$$\beta_0(z) = 1 + \rho_n z, \beta_1(z) = 1 + \rho_n z^2, \tag{4.2}$$

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. Evidently, the two tests share very 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 slighted oversized when T is small. When $N = 20$, they are undersized for small m. For m=20, size is close to the nominal size. Overall, test size appears sensitive to m for small sample sizes and more sensitive in the case of I_m^* than J.

To achieve better size control we adopt a bootstrap procedure. 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. The results are collected in Table 4. Evidently, size of the bootstrap tests are close to nominal size for most parameter constellations and there is no clear dependence on m. 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 (4.2) 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



(a) $N = 5$

(b) $N = 20$

Figure 1: Rejection rates of the bootstrap procedure with 5% nominal size

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 (4.2) with $\sqrt{nh}\rho_n \to \infty$.

5 Empirical application to climate sensitivity analysis

As an application of our methods we consider the problem of estimating Earth's climate sensitivity to a given increase in atmospheric CO_2 concentration. As described in Storelvmo et al. (2018), this is an issue on which there is much ongoing research, primarily with the use of large scale global climate models. In these global climate modeling exercises analysis relies on computer simulation data generated from immensely detailed models of Earth's climate using an ensemble of initializations of the variables that help to assess model sensitivity. An alternative approach that instead relies on observational data is to use econometric methods to fit much simpler dynamic panel models from which parameter estimates may be obtained to assess the impact on climate of rising atmospheric CO_2 concentrations. The methodology has been developed in recent work by Magnus et al. (2011), Storelvmo et al. (2016) and Phillips et al. (2019). An advantage of the approach is that observationally based confidence intervals may be constructed for the key parameters that are involved in measuring climate change dynamics and the long term impact of rising CO_2 concentrations.

Our application uses three observational data sets: temperature (T_{it}), surface level solar radiation (R_{it}) and CO_2 equivalent greenhouse gas concentrations ($CO_{2,t}$). The temperature and surface radiation data record time series at multiple surface stations and thereby conform to usual panel data with individual station and time series observations, whereas the CO_2 data varies only over the temporal dimension. Since a primary goal of empirical research with this data is to measure the recent historical impact of aggregate CO_2 on Earth's climatic temperature, the panel model framework necessarily involves the use of time and station level series in conjunction with possible communal variables that affect climate in aggregate. Such communal variables in the present case are aggregate CO_2 levels and aggregate solar radiation levels measured at the Earth's surface.

It is convenient in this application to use the same data as in Phillips et al. (2019), which is recorded over the 42-year period from 1964 to 2005 over 1484 land-based observation stations. Using this data enables comparisons of various nonlinear functional coefficient specifications with the linear specifications employed in Phillips et al. (2019). For example, we are able to assess whether the impact of CO_2 as a communal variable works nonlinearly through the model coefficients or simply linearly as a common time effect regressor. For more information about the data, see Phillips et al. (2019) and Storelvmo et al. (2016) and the references therein.[7]

Figure 2 plots the aggregated time series data[8]. An obvious feature of these aggregate data series are their nonstationarity. Trend characteristics of varying types are evident in the global temperature, radiation, and CO_2 series, with greater year-to-year volatility in temperature and radiation than in CO_2. These series were modeled in Phillips et al. (2019) by allowing for stochastic and deterministic linear trends as well as cointegrating linkages among the three aggregate series. As might be expected, much greater variation occurs in the disaggregated station level data. Linear cointegrating regression analysis was used in Phillips et al. (2019) in studying the aggregate series and an asymptotic theory for the estimated coefficients was obtained, including asymptotics for a composite parameter that measures climate sensitivity to CO_2.

In Magnus et al. (2011) and Storelvmo et al. (2016) disaggregated station-level data were employed, standard dynamic panel within group and GMM methods were used in estimation, and no attention was paid to possible nonstationarity in the data. The present application proceeds along similar lines, utilizing the nonparametric panel regression modeling methodology developed here, allowing for the presence of functional regression coefficients that depend on communal aggregate variables. The goal of the application is to assess the impact at the station

[7]For a general discussion of Earth's climate sensitivity to greenhouse gases and for references to recent work in this field of climate science, readers are also referred to Storelvmo et al. (2018).

[8]Minor differences between the CO_2 graphic in Figure 2 and that shown in Phillips et al. (2019) arise because more decimal places are used here than in Phillips et al. (2019)

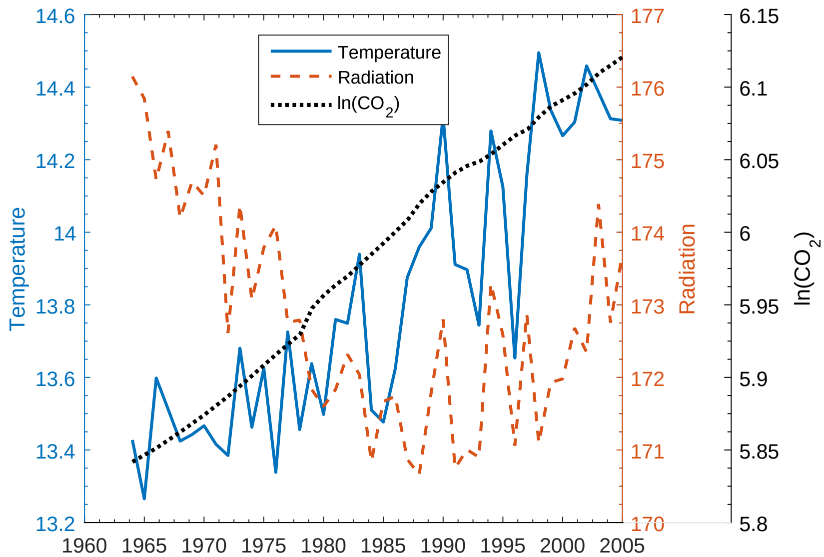
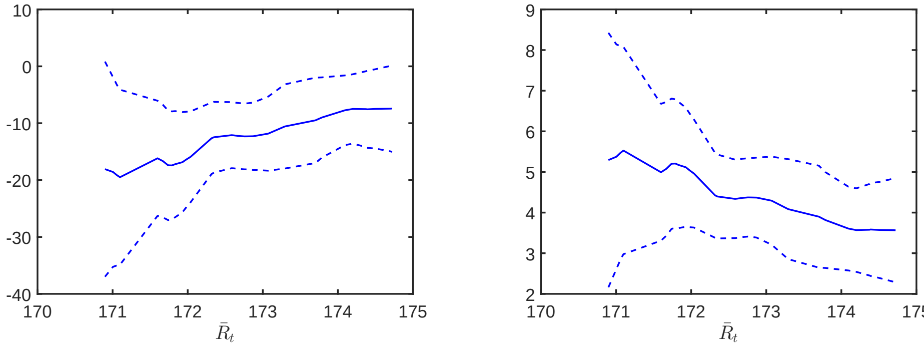


Figure 2: Time series plot of Station-averaged temperature ($\overline{T}_t$, °Celsius; blue solid), downward surface radiation ($\bar{R}_t$, Watts per m^2; red dashed) and logarithms of CO_2 (Pg: metric gigatons; black dotted) ranging from 1964 to 2005

level of the global influences on Earth's temperature of CO_2 and downwelling solar radiation and to determine whether linear specifications of the type used in earlier research is justified. The implementation of this paper's methodology that follows therefore does not specifically address nonstationarity in the aggregate data, so that the tests and confidence intervals obtained in our empirics are not strictly supported by the theory of the present paper. Nonetheless, we expect that the present findings will be indicative[9] and development of supporting limit theory for nonstationary data extensions of the present paper's methodology is planned by the authors for future work.

We first investigate the relation among these three variables using the aggregated time series data plotted in Figure 2. To explore the sensitivity of temperature to CO_2, earlier work in the literature has used linear regression of temperature on CO_2 via linear modeling or linear cointegration modeling, as in Phillips et al. (2019). To allow for a possibly nonlinear coefficient impact of CO_2 on temperature, where aggregate levels of downwelling radiation may potentially

[9]There is good reason to expect that the stationary process kernel regression asymptotics used here retain some validity in nonstationary cases. Indeed, Wang and Phillips (2009a,b) showed that stationary kernel asymptotics, including confidence intervals, remain valid in certain nonstationary nonparametric cases such as regression functions of integrated or near integrated time series.



(a) $\beta_0(\bar{R}_t)$ (b) $\beta_1(\bar{R}_t)$

Figure 3: Estimated functional coefficients of model (5.1)

affect the impact of CO_2, we use the functional coefficient regression formulation

$$\overline{T}_t = \beta_0(\bar{R}_t) + \beta_1(\bar{R}_t)\ln(CO_{2,t}) + e_t, \tag{5.1}$$

where $\overline{T}_t$ and $\bar{R}_t$ denote the cross-station average of T_{it} and R_{it}, respectively. The estimated functional coefficients and the corresponding 95% confidence bands[10] are plotted in Figure 3. The fitted functional coefficient estimates show that the intercept function $\beta_0(\bar{R}_t)$ exhibits an upward trend with radiation, suggesting that the level impact on temperature rises with solar radiation, as expected. On the other hand, the slope coefficient function $\beta_1(\bar{R}_t)$ exhibits a downward trend, suggesting lower correlation between temperature and CO_2 when radiation is higher. This outcome may be partly explained by the fact that downwelling solar radiation rises when atmospheric conditions are clearer with less pollutants (like sulfur dioxide) and in such cases the greenhouse gas effects of rising CO_2 on temperature may be attenuated because of greater infrared radiation into space and aerosol/cloud interactions (Wild, 2012).

To see whether the linear relation assumption between temperature and CO_2 is supported by the data, we consider the following partial linear model

$$\overline{T}_t = \beta_0(\bar{R}_t) + \beta_1\ln(CO_{2,t}) + v_t. \tag{5.2}$$

The estimated curve of $\beta_0(\bar{R}_t)$ is plotted in Figure 4 with 95% confidence bands, and the estimate

[10] The confidence bands plotted in Figure 3 may be inaccurate. In model (5.1), $\bar{R}_t$ is I(1) and $\ln(CO_{2,t})$ has a linear drift as well as a stochastic trend. To the best of our knowledge, asymptotic theory for nonparametric regression within a functional coefficient environment with nonstationary communal covariate, as in the present model, is unavailable in the literature. The current confidence bands are computed based on the asymptotics obtained by Xiao (2009), which studies functional-coefficient cointegration in which the coefficients are functions of a stationary covariate.

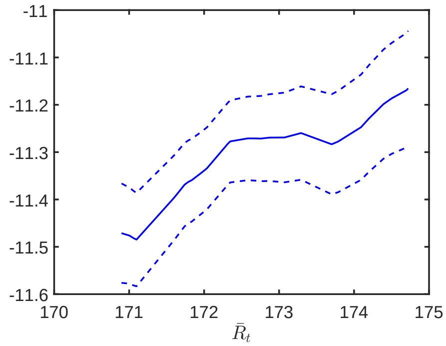


Figure 4: Estimated functional coefficient $\beta_0(\bar{R}_t)$ of model (5.2)

of β_1 is 4.20 with the 95% confidence interval $[3.47, 4.92]$[11]. The estimated curve $\beta_0(\bar{R}_t)$ in Figure 4 shows a clear upward trend, revealing strong positive effects from radiation to temperature. To test the constancy of $\beta_1(\bar{R}_t)$ in model (5.1), or equivalently, the linear relation assumption between temperature and CO_2, we employ a likelihood ratio test to test the null model of (5.2) against the alternative model of (5.1). The test gives a p-value of 0.52, suggesting acceptance of a linear relation between temperature and CO_2 that embodies the positive impact of radiation on temperature through the functional dependence on radiation of the intercept.

To further investigate the nature of the impact of radiation on temperature, we consider testing the constancy of $\beta_0(\bar{R}_t)$ in model (5.2). A likelihood ratio test gives a p-value of 0.019, suggesting rejection of the commonly used simple linear model $\overline{T}_t = \beta_0 + \beta_1 \ln(CO_{2,t}) + \epsilon_t$. This result strongly demonstrates the significant role of radiation on temperature. Moreover, from the estimated curve shown in Figure 4 for the intercept function, it seems that radiation has a near linear impact on temperature. We therefore proceed to test whether $\beta_0(\bar{R}_t)$ in model (5.2) can be accepted as a linear function of $\bar{R}_t$. That is, we test a linear model of the following form

$$\overline{T}_t = \alpha_0 + \alpha_1 \bar{R}_t + \beta_1 \ln(CO_{2,t}) + \varepsilon_t, \tag{5.3}$$

against the partial linear model in (5.2). Using a likelihood ratio test gives a p-value of 0.75, suggesting that the linear model specification in (5.3) is an adequate one to describe the dependence of temperature on radiation and CO_2. The linear model in (5.3) is exactly the global linear (cointegrating) relation among the variables that was used in Phillips et al. (2019).

[11]The confidence bands in Figure 4 and the confidence interval for β_1 may be inaccurate because the asymptotic theory for the model (5.2) with nonstationary data characteristics is presently unavailable in literature. According to the tests in Phillips et al. (2019), $\bar{R}_t$ has a stochastic trend and $\ln(CO_{2,t})$ has a stochastic trend with drift. The stated calculations are currently based on the asymptotic theory of a model of the same form as (5.2), assuming that $\bar{R}_t$ is stationary. However, as discussed in fn. 9, there is good reason to believe that this asymptotic theory may have validity in a wider context that allows for integrated or near integrated communal variates.

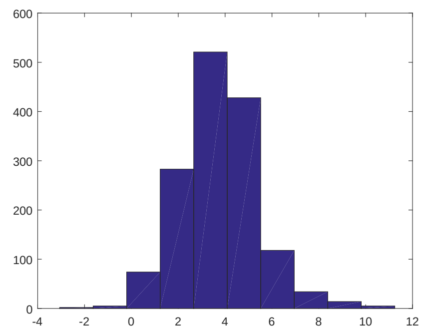


Figure 5: Histogram of the estimates of $\beta_{1,i}$ based on model (5.2) for each station i

We also estimate the partial linear model in (5.2) using the individual time series data for each station to explore possible heterogeneity in the coefficient β_1 across stations. In particular, we estimate the partially linear regression equation $T_{it} = \beta_{0,i}(R_{it}) + \beta_{1,i} \ln(CO_{2,t}) + v_{it}$ for each station i. Figure 5 shows the histogram of the estimates of $\beta_{1,i}$ obtained for all 1484 stations. It is evident that the estimates center on a slope coefficient around 4 but with substantial variation over stations, indicating that the impact of CO_2 on temperature varies considerably over land station locations. These findings on idiosyncratic variation in the magnitudes of β_1 suggest that stations might usefully be divided into subgroups of homogeneous coefficients, thereby capturing between-group heterogeneity in the global impact of CO_2 on temperature. Such grouping might be accomplished using modern econometric methods such as classification Lasso techniques (Su et al., 2016).

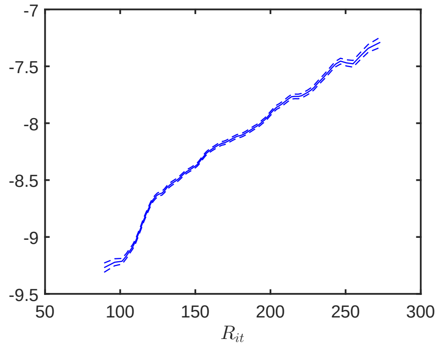
Next we consider full panel specifications, starting with the model

$$T_{it} = \alpha_i + \beta_0(R_{it}) + \beta_1 \ln(CO_{2,t}) + e_{it}. \tag{5.4}$$

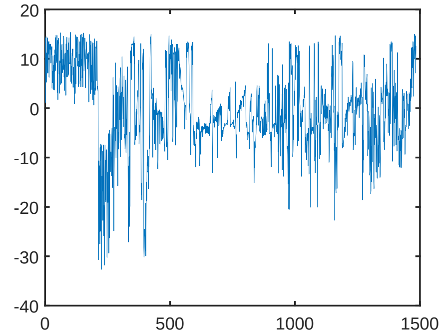
The $\ln(CO_{2,t})$ coefficient estimate is 3.68 for β_1, with a 95% confidence interval $[3.63, 3.72]$, which accords closely with the histogram results of Figure 5. The estimated intercept function $\beta_0(R_{it})$ and estimated fixed effects α_i are plotted in Figure 6. The estimated function $\beta_0(R_{it})$ shows a strong rising influence of radiation on temperature, as would be expected. Moreover, the fitted curve is indicative of a linear relation with a strong linear correlation with temperature. This finding supports the linear panel model specification used in earlier work (Storelvmo et al., 2016; Phillips et al., 2019). The fixed effects displayed in the right panel of Figure 6 also indicate that stations may be usefully subdivided into smaller more homogeneous groups.

In view of the strong evidence of linearity shown in Figure 6 (a), we consider the following linear panel data model as a specialization of (5.4)

$$T_{it} = \alpha_i + \beta_0 R_{it} + \beta_1 \ln(CO_{2,t}) + u_{it}. \tag{5.5}$$



(a) $\beta_0(R_{it})$



(b) Fixed Effects α_i

Figure 6: Estimated functional coefficient $\beta_0(R_{it})$ and the fixed effects α_i of model (5.4)

Within Group estimation is used to estimate (5.5) giving $\hat{\beta}_0 = 0.0103$ with 95% confidence interval $[0.0098, 0.0108]$, and $\hat{\beta}_1 = 3.7042$ with 95% confidence interval $[3.6585, 3.7499]$. The fixed effects are similar to those given in Figure 6 (b) and thus are omitted. We use the likelihood ratio test to test the linear model (5.5) against the alternative partial linear model in (5.4). The p-value is 0 in this test and so the simple linear panel model (5.5) is rejected. The main reason for the statistical rejection is the extremely sharp confidence bounds that are shown in Figure 6 (a), which exclude the strict linear relationship (5.5).[12] These sharp bounds arise from the very large number of total observations, $NT = 62,328$, that are used in estimating the nonparametric function $\beta_0(R_{it})$ in (5.4).

For compliance with the specifications used in this paper, we also consider the following functional coefficient model

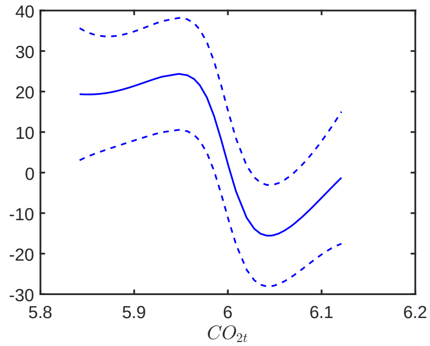
$$T_{it} = \alpha_i + \beta_0(\ln(CO_{2,t})) + R_{it}\beta_1(\ln(CO_{2,t})) + v_{it}. \tag{5.6}$$

This model is consistent with the form of (1.2), except that the data here do not meet the requirements of stationarity[13] in Assumption 2. Despite these inconsistencies with the assumed regularity conditions, we can still estimate (5.6) with the proposed methods. The estimated curves obtained by using the APLC method[14] are shown in Figure 7. From the shape of the $\beta_1(\ln(CO_{2,t}))$ curve shown in panel (b), it is apparent that when atmospheric concentrations of CO_2 are high, radiation R_{it} has a significant positive effect on temperature, whereas the

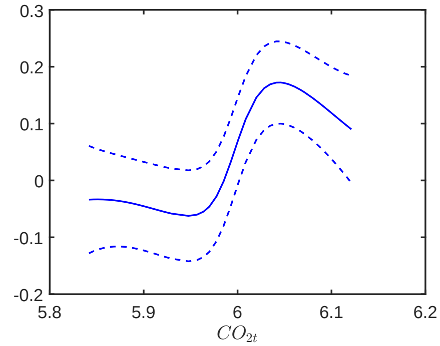
[12] Use of parsimonious parametric forms such as inclusion of a quadratic term in R_{it} to (5.5) still leads to a rejection in favor of the partial linear model with a likelihood ratio p-value again of 0.

[13] As indicated earlier, the $\ln(CO_{2,t})$ data have a stochastic trend with drift, whereas we assume z_t to be stationary; and radiation R_{it} has a stochastic trend, whereas we assume x_{it} to be stationary.

[14] The PLC estimation procedure is extremely computer intensive and time-consuming in this case with a very large NT value, so we only report the APLC results here.



(a) APLC estimate of $\beta_0(\ln(CO_{2,t}))$



(b) APLC estimate of $\beta_1(\ln(CO_{2,t}))$

Figure 7: Estimated coefficient curves of model (5.6)

impact is less significant when CO_2 levels are lower. The patterns shown by the fitted intercept function $\beta_0(\ln(CO_{2,t}))$ in panel (a) appear complementary in form to those of $\beta_1(\ln(CO_{2,t}))$. These results accord with the aggregate time series behavior shown in Figure 2. In particular, high levels of CO_2 occur later in the sample time period because of the strong trend in $\ln(CO_{2,t})$ and during that period radiation appears to comove upwards with temperature, whereas lower levels of CO_2 occur early in the sample time period when radiation levels show little correlation with temperature. These aggregate movements are mirrored in the station level data, leading to the estimated coefficient curves shown in Figure 7. The fitted individual effects are similar to those shown in Figure 6 (b) and are therefore omitted.

These empirical findings show that functional coefficient regression models can be helpful in guiding and justifying parametric specifications where communal covariates are relevant. Both at the aggregate and panel regression levels the functional coefficient estimates reveal that the linear panel regression specifications used in past applied climate econometric research (Magnus et al., 2011; Storelvmo et al., 2016; Phillips et al., 2019) in studying the impact of CO_2 and downwelling solar radiation on temperature receive support from testing these linear specifications against much more complex nonparametric mechanisms where nonlinear effects of CO_2 and radiation on temperature are permitted.

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-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 $(N, T) \to \infty$ 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. 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.

There are many directions in which the current work may be extended. 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. Function-coefficient models with nonstationary, trending variables are also of great importance in empirical work, as the empirical application to climate data illustrates. The nonstationary covariate case is particularly relevant in some applications (e.g., the trending CO_2 series was used in the empirical illustration) and leads to the challenges presented by nonlinear nonstationary kernel regression (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. We use 'smaller order' to represent terms that are of smaller asymptotic order than the terms stated explicitly. C is a constant that may take different values at 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 notation $VecDiag(B)$ means take the diagonal elements of the square matrix B and stack them as a column vector. According to the context, we use := and =: to signify definitional equality.

A Proof of Theorem 2.1

Since (2.3) is a conventional time series functional-coefficient model of stationary data, most of the proof follows standard lines and we only outline the essentials. Important distinctions occur when $\mathbb{E}x_{it} = \eta \neq 0$ in part (c).

First, the d-th element of $\beta_*(z_t)$, $\beta_{d-1}(z_t), d = 1, \cdots, p+1$, admits the following second order Taylor expansion at a given point $z = (z_1, \cdots, z_q)'$

$$\begin{aligned}
\beta_{d-1}(z_t) &= \beta_{d-1}(z) + (z_t - z)'\beta_{d-1}^{(1)}(z) + \frac{1}{2}(z_t - z)'\beta_{d-1}^{(2)}(z)(z_t - z) + o_p(|z_t - z|^2) \\
&= \beta_{d-1}(z) + [H^{-1}(z_t - z)]'H\beta_{d-1}^{(1)}(z) + \frac{1}{2}[H^{-1}(z_t - z)]'H\beta_{d-1}^{(2)}(z)H[H^{-1}(z_t - z)] + o_p(|z_t - z|^2) \\
&=: \beta_{d-1}(z) + \dot{\beta}_{d-1}(z, z_t) + \frac{1}{2}\ddot{\beta}_{d-1}(z, z_t) + o_p(|z_t - z|^2),
\end{aligned}$$

where $H = diag(h_1, \cdots, h_q)$ and $|z_t - z|^2 = \sum_{j=1}^{q}(z_{t,j} - z_j)^2$. The definitions of $\dot{\beta}_{d-1}(z, z_t)$ and $\ddot{\beta}_{d-1}(z, z_t)$ follow from the context and they are scalars. Then the $(p + 1) \times 1$ vector $\beta_*(z_t)$ can be written as follows

$$\beta_*(z_t) = \begin{pmatrix} \beta_0(z_t) \\ \beta_1(z_t) \\ \vdots \\ \beta_p(z_t) \end{pmatrix} = \beta_*(z) + \dot{\beta}_*(z, z_t) + \frac{1}{2}\ddot{\beta}_*(z, z_t) + o_p(|z_t - z|^2)1_{(p+1)\times 1}, \tag{A.1}$$

where $\dot{\beta}_*(z, z_t)$ is a $(p + 1) \times 1$ vector with d-th element $\dot{\beta}_{d-1}(z, z_t)$, $\ddot{\beta}_*(z, z_t)$ is $(p + 1) \times 1$ with d-th element $\ddot{\beta}_{d-1}(z, z_t)$. $1_{p\times 1}$ denotes a $p \times 1$ vector of ones. Using the Taylor expansion of $\beta_*(z_t)$ given in (A.1) in (2.3) we have

$$y_{At} = x'_{*,At}\beta_*(z) + x'_{*,At}\dot{\beta}_*(z, z_t) + \frac{1}{2}x'_{*,At}\ddot{\beta}_*(z, z_t) + u_{At} + smaller\ order, \tag{A.2}$$

or in vector form,

$$Y_A = X_A^*\beta_*(z) + U_A + VecDiag(X_A^*\dot{\beta}_*(z)) + \frac{1}{2}VecDiag(X_A^*\ddot{\beta}_*(z)) + smaller\ order, \tag{A.3}$$

where X_A^* is $T \times (p+1)$ by stacking $x'_{*,At}$, $\dot{\beta}_*(z)$ is $(p+1) \times T$ with the t-th column being the $(p+1) \times 1$ vector $\dot{\beta}_*(z, z_t)$, and $\ddot{\beta}_*(z)$ is analogously defined. Then we have

$$\begin{aligned}
\hat{\beta}_{*,APLC}(z) - \beta_*(z) =& [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z Y_A - \beta_*(z) \\
=& [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z \left(X_A^*\beta_*(z) + U_A \right. \\
& \left. + VecDiag(X_A^*\dot{\beta}_*(z)) + VecDiag(X_A^*\ddot{\beta}_*(z))/2 + smaller\ order \right) - \beta_*(z) \\
=& [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z U_A + [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z VecDiag(X_A^*\dot{\beta}_*(z)) \\
& + 1/2[(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z VecDiag(X_A^*\ddot{\beta}_*(z)) + smaller\ order.
\end{aligned} \tag{A.4}$$

From expression (A.4), it follows that the bias term is $\bar{B}_1^* + \bar{B}_2^*$, where

$$\begin{aligned}
\bar{B}_1^* &= [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z VecDiag(X_A^*\dot{\beta}_*(z)), \\
\bar{B}_2^* &= \frac{1}{2}[(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z VecDiag(X_A^*\ddot{\beta}_*(z)).
\end{aligned}$$

Combining Lemma D.1 and D.3, we can see under all three situations (N is fixed, N goes to infinity and $\mathbb{E}x_{it} = 0$ or N goes to infinity and $\mathbb{E}x_{it} \neq 0$), we have

$$\bar{B}_1^* \xrightarrow{p} \mu_2 f_z^{-1}(z) \sum_{j=1}^{q} h_j^2 \frac{\partial \beta_*(z)}{\partial z_j} \frac{\partial f_z(z)}{\partial z_j}, \tag{A.5}$$

$$\bar{B}_2^* \xrightarrow{p} \frac{1}{2}\mu_2 \sum_{j=1}^{q} h_j^2 \frac{\partial^2 \beta_*(z)}{\partial^2 z_j}. \tag{A.6}$$

Therefore, the asymptotic bias is

$$\mathcal{B}(z) = \mu_2 f_z^{-1}(z) \sum_{j=1}^{q} h_j^2 \left[\frac{\partial \beta_*(z)}{\partial z_j} \frac{\partial f_z(z)}{\partial z_j} + \frac{1}{2} \frac{\partial^2 \beta_*(z)}{\partial^2 z_j} f_z(z) \right].$$

From (A.4), the term that determines the limit distribution is

$$\bar{V}^* = [(X_A^*)'K_z X_A^*]^{-1}(X_A^*)'K_z U_A.$$

Combining Lemma D.1 and D.2, we have when N is fixed,

$$\sqrt{T|H|}\bar{V}^* \Rightarrow N(0, \nu_0^q N^{-1} \sigma_u^2 f_z^{-1}(z) \bar{V}_{xx}^{-1}). \tag{A.7}$$

When $N \to \infty$ and $\mathbb{E}x_{it} = 0$,

$$\begin{pmatrix} \sqrt{NT|H|} & 0 \\ 0 & \sqrt{T|H|}I_p \end{pmatrix} \bar{V}^* = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{N}I_p \end{pmatrix}^{-1} \sqrt{NT|H|}\bar{V}^* \Rightarrow N(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)(V_{xx}^*)^{-1}). \tag{A.8}$$

When $N \to \infty$ and $\mathbb{E}x_{it} \neq 0$, we have

$$\sqrt{T|H|}\bar{V}^* \Rightarrow N\left(0, \nu_0^q \sigma_u^2 f_z^{-1}(z) \begin{pmatrix} \eta' \Sigma_{xx}^{-1} \eta & -\eta' \Sigma_{xx}^{-1} \\ -\Sigma_{xx}^{-1} \eta & \Sigma_{xx}^{-1} \end{pmatrix}\right),$$

which is a degenerate distribution.

Combining (A.5)-(A.8), we see that when N is fixed,

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

and when $N \to \infty$ and $\mathbb{E}x_{it} = 0$,

$$D_N(\hat{\beta}_{*,APLC}(z) - \beta_*(z) - \mathcal{B}(z)) \Rightarrow N(0, \nu_0^q \sigma_u^2 f_z^{-1}(z)(V_{xx}^*)^{-1}),$$

where

$$D_N = \begin{pmatrix} \sqrt{NT|H|} & 0 \\ 0 & \sqrt{T|H|} I_p \end{pmatrix}.$$

When $N \to \infty$ as $T \to \infty$ and $\mathbb{E}x_{it} = \eta \neq 0$, we have

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

which is degenerate normal. ■

B Proof of Theorem 2.3

(a) We first consider the oracle estimator $\hat{\beta}_{*,PLC}^{oracle}$. The Taylor expansion given in (A.1) continues to hold. Using this expansion in model (1.2) we have

$$y_{it} = \alpha_i + x'_{*,it}\left[\beta_*(z) + \dot{\beta}_*(z, z_t) + \frac{1}{2}\ddot{\beta}_*(z, z_t)\right] + u_{it} + smaller\ order, \tag{B.1}$$

or in vector form

$$Y = D^*\alpha^* + X_*\beta_*(z) + VecDiag(X_*[1'_N \otimes \dot{\beta}_*(z)]) + \frac{1}{2}VecDiag(X_*[1'_N \otimes \ddot{\beta}_*(z)]) + U + smaller\ order,$$

where X_* is $NT \times (p+1)$ by stacking the $1 \times (p+1)$ vector $x'_{*,it} = [1, x'_{it}]$. $\ddot{\beta}_*(z)$ is $(p+1) \times T$ with the t-th column being $\ddot{\beta}_*(z, z_t)$. $\dot{\beta}_*(z)$ is analogously defined. Then we have

$$\begin{aligned}
\hat{\beta}_{*,PLC}^{oracle}(z) =& [X'_*K(z)X_*]^{-1}X'_*K(z)(Y - D^*\alpha^*) \\
=& [X'_*K(z)X_*]^{-1}X'_*K(z)\Big[X_*\beta_*(z) + VecDiag(X_*[1'_N \otimes \dot{\beta}_*(z)]) \\
& + \frac{1}{2}VecDiag(X_*[1'_N \otimes \ddot{\beta}_*(z)]) + U + smaller\ order\Big] \\
=& \beta_*(z) + [X'_*K(z)X_*]^{-1}X'_*K(z)U + [X'_*K(z)X_*]^{-1}X'_*K(z)VecDiag(X_*[1'_N \otimes \dot{\beta}_*(z)]) \\
& + [X'_*K(z)X_*]^{-1}X'_*K(z)\frac{1}{2}diag[X_*(1_{1\times N} \otimes \ddot{\beta}_*(z))] + smaller\ order.
\end{aligned}$$

In view of Lemma D.4 and D.5, we have

$$\frac{1}{\sqrt{NT|H|}}[X_*'K(z)X_*]^{-1}X_*'K(z)U \Rightarrow N(0, f_z^{-1}(z)\sigma_u^2\nu_0^q\tilde{V}_{xx}^{-1}).$$

In view of Lemma D.4 and D.6, we have

$$[X_*'K(z)X_*]^{-1}X_*'K(z)VecDiag(X_*[1_N'\otimes\dot{\beta}_*(z)]) \xrightarrow{p} \mu_2 f_z^{-1}(z)\sum_{s=1}^{q}h_s^2\frac{\partial\beta_*(z)}{\partial z_s}\frac{\partial f_z(z)}{\partial z_s},$$

$$[X_*'K(z)X_*]^{-1}X_*'K(z)VecDiag(X_*[1_N'\otimes\ddot{\beta}_*(z)]) \xrightarrow{p} \mu_2\sum_{s=1}^{q}h_s^2\frac{\partial^2\beta_*(z)}{\partial^2 z_s}.$$

Thus the asymptotic bias is

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

Finally, we have

$$\sqrt{NT|H|}(\hat{\beta}_{*,PLC}^{oracle} - \beta_*(z) - \mathcal{B}(z)) \Rightarrow N(0, f_z^{-1}(z)\sigma_u^2\nu_0^q\tilde{V}_{xx}^{-1}).$$

(b) Next consider the estimator $\hat{\alpha}_{PLC}^*$ defined in (2.20). Based on (2.19) we have

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

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

$$\begin{aligned}\hat{\alpha}_{PLC}^* - \alpha^* &= [(M_2^*)'M_2^*]^{-1}(M_2^*)'VecDiag(X_*\{1_N'\otimes[\beta_*(Z)-\hat{\beta}_{*,PLC}^{oracle}(Z)]\}) + [(M_2^*)'M_2^*]^{-1}(M_2^*)'U \\ &=: [(M_2^*)'M_2^*]^{-1}(M_2^*)'W + [(M_2^*)'M_2^*]^{-1}(M_2^*)'U, \qquad \text{(B.2)}\end{aligned}$$

where $W = VecDiag(X_*\{1_N'\otimes[\beta_*(Z)-\hat{\beta}_{*,PLC}^{oracle}(Z)]\})$. Below we analyze the two terms in (B.2). 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.

Note that $M_2^* = (I_{NT} - Q_1^*)D^*$. First consider Q_1^*. We have

$$Q_1^* = \sum_{d=1}^{p+1} (\mathbf{x}_{d-1} \otimes 1_n') \odot (1_N \otimes w_{*,d}(Z))$$

$$= \begin{pmatrix} \sum_d x_{11,d-1} w_d(z_1) \\ \vdots \\ \sum_d x_{1T,d-1} w_d(z_T) \\ \vdots \\ \sum_d x_{N1,d-1} w_d(z_1) \\ \vdots \\ \sum_d x_{NT,d-1} w_d(z_T) \end{pmatrix} = \begin{pmatrix} x_{*,11}' w_*(z_1) \\ \vdots \\ x_{*,1T}' w_*(z_T) \\ \vdots \\ x_{*,N1}' w_*(z_1) \\ \vdots \\ x_{*,NT}' w_*(z_T) \end{pmatrix},$$

where $w_*(z_t) = [X_*' K(z_t) X_*]^{-1} X_*' K(z_t)$ is $(p+1) \times n$, with $n \equiv NT$. Then each row $x_{*,it}' w_*(z_t)$ is $1 \times n$. For the matrix $Q_1^* D^*$, the typical row at the $(j-1)$-th column is

$$x_{*,it}'[X_*' K(z_t) X_*]^{-1} \sum_{s=1}^{T} [-x_{*,1s} K(H^{-1}(z_s - z_t)) + x_{*,js} K(H^{-1}(z_s - z_t))],$$

for $j = 2, \cdots, N$. We already have $\frac{1}{NT|H|} X_*' K(z) X_* \xrightarrow{p} f_z(z) \tilde{V}_{xx}$. Note that

$$\frac{1}{T|H|} \sum_{s=1}^{T} \left[-x_{*,1s} K(H^{-1}(z_s - z_t)) + x_{*,js} K(H^{-1}(z_s - z_t))\right] \xrightarrow{p} f_z(z_t) \left[- \begin{pmatrix} 1 \\ \eta_1 \end{pmatrix} + \begin{pmatrix} 1 \\ \eta_j \end{pmatrix} \right].$$

Since $\eta_1 = \cdots = \eta_N = \eta$, we have $Q_1^* D^* \xrightarrow{p} 0$. then $M_2^* \xrightarrow{p} D^*$. Further, by direct calculation $T^{-1}(D^*)' D^* = I_{N-1} + 1_{(N-1)\times 1} \otimes 1_{1 \times (N-1)}$. Then we have $T^{-1}(M_2^*)' M_2^* \xrightarrow{p} I_{N-1} + 1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}$, and $(M_2^*)' U = (D^* + o_p(1))' U$. Note that the $(j-1)$-th row of $(D^*)'U$ is $-\sum_t u_{1t} + \sum_t u_{jt}$, for $j = 2, \cdots, N$. Since we assume u_{it} is *iid* across i, the asymptotic variance of $\frac{1}{\sqrt{T}}(D^*)'U$ is $\gamma_u^2 [I_{N-1} + 1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}]$. Therefore, for the second term in (B.2), we have

$$\sqrt{T}[(M_2^*)' M_2^*]^{-1} (M_2^*)' U \Rightarrow N(0, \gamma_u^2 [I_{N-1} + 1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}]^{-1}). \tag{B.3}$$

Next consider the first term in (B.2). Note that the typical row of W is $x_{*,it}'[\beta_*(z_t) - \hat{\beta}_{*,PLC}^{oracle}(z_t)]$, so that the $(j-1)$-th row of $(M_2^*)'W$ is $\sum_t [x_{*,jt}' - x_{*,1t}'][\beta_*(z_t) - \hat{\beta}_{*,PLC}^{oracle}(z_t)]$, for $j = 2, \cdots, N$. Now $[x_{*,jt}' - x_{*,1t}']$ has zero mean, $T^{-1/2} \sum_t [x_{*,jt}' - x_{*,1t}'] = O_p(1)$ by central limit theory, and by a uniform convergence extension[15] of the limit theory of the oracle estimator $\sup_{t \le T} |\beta_*(z_t) - \hat{\beta}_{*,PLC}^{oracle}(z_t)| = o_p(1)$. It follows that $T^{-1/2}(M_2^*)'W = o_p(1)$. Thus the first term in (B.2) is of order $o_p(1/\sqrt{T})$, which is smaller than that of the second term.

[15] 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.

Finally, in view of (B.3), we have

$$\sqrt{T}(\hat{\alpha}^*_{PLC} - \alpha^*) \Rightarrow N(0, \gamma_u^2[I_{N-1} + 1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}]^{-1})$$
$$= N(0, \gamma_u^2[I_{N-1} - \frac{1}{N}1_{(N-1)\times 1} \otimes 1_{1\times(N-1)}]). \tag{B.4}$$

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

$$\begin{aligned}\hat{\beta}_{*,PLC}(z) - \beta_*(z) &= [X_*'K(z)X_*]^{-1}X_*'K(z)(Y - D^*\hat{\alpha}^*_{PLC}) - \beta_*(z) \\ &= [X_*'K(z)X_*]^{-1}X_*'K(z)(Y - D^*\alpha^*) - \beta_*(z) + [X_*'K(z)X_*]^{-1}X_*'K(z)D^*(\alpha^* - \hat{\alpha}^*_{PLC}) \\ &= \hat{\beta}^{oracle}_{*,PLC}(z) - \beta_*(z) + [X_*'K(z)X_*]^{-1}X_*'K(z)D^*(\alpha^* - \hat{\alpha}^*_{PLC}). \end{aligned} \tag{B.5}$$

For the second term in (B.5), we already have $\frac{1}{NT|H|}X_*'K(z)X_* \xrightarrow{p} f_z(z)\tilde{V}_{xx}$. Consider the term $X_*'K(z)D^*(\alpha^* - \hat{\alpha}^*_{PLC})$. We have

$$\frac{1}{NT|H|}X_*'K(z)D^*(\alpha^* - \hat{\alpha}^*_{PLC}) = \begin{pmatrix} 0 \\ \frac{1}{N}\sum_{j=2}^N[\frac{1}{T|H|}\sum_t(x_{jt} - x_{1t})K_t](\alpha_j - \hat{\alpha}_{j,PLC}) \end{pmatrix}. \tag{B.6}$$

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

$$\frac{1}{T|H|}\sum_t(x_{jt} - x_{1t})K_t = O_p(1/\sqrt{T|H|})\ for\ all\ j = 2, \cdots, N. \tag{B.7}$$

Further, the term $\frac{1}{T|H|}\sum_t(x_{jt} - x_{1t})K_t$ may be taken out of the summation over j in (B.6) after rescaling by appealing to a law of interated logarithm for kernel regression estimates and triangular arrays, so that

$$\limsup_{T\to\infty} \pm\sqrt{2\frac{T|H|}{\log_2(T|H|)}}\frac{1}{T|H|}\sum_t(x_{jt} - x_{1t})K_t \leq C\ a.s., \tag{B.8}$$

where $\log_2(\cdot) = \log\log(\cdot)$, and C is a constant that is independent of j by virtue of stationarity over j and bounded density $f_z(z)$. 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 (B.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.[16]

[16] 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.

From (B.4), we have

$$1_{1\times(N-1)}\sqrt{T}(\hat{\alpha}^*_{PLC}-\alpha^*) \Rightarrow N(0,\gamma_u^2 1_{1\times(N-1)}[I_{N-1}-\frac{1}{N}1_{(N-1)\times 1}\otimes 1_{1\times(N-1)}]1_{(N-1)\times 1})$$
$$= N(0,\gamma_u^2(N-1)/N).$$

Therefore, $\sqrt{T}\sum_{j=2}^N(\hat{\alpha}_{j,PLC}-\alpha_j) \Rightarrow N(0,\gamma_u^2\frac{N-1}{N})$. Then

$$\frac{1}{N}\sum_{j=2}^N(\alpha_j-\hat{\alpha}_{j,PLC}) = O_p(1/N\sqrt{T}). \tag{B.9}$$

Combining (B.7) and (B.9), we know the second entry of (B.6) is of order $O_p(1/NT\sqrt{|H|})$. Thus the second term of (B.5) is of order $O_p(1/NT\sqrt{|H|})$, which is smaller than that of the first term as $NT\to\infty$. Therefore, the feasible estimator $\hat{\beta}_{*,PLC}(z)$ is asymptotically equivalent to the oracle estimator $\hat{\beta}^{oracle}_{*,PLC}(z)$ as $NT\to\infty$. ■

C Proof of Theorem 3.2

For convenience, we first outline the estimation procedure under the null H_0. Let $\beta_* = (\beta_0,\beta')'$. Under H_0, model (1.2) becomes

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

Taking averages over t, we get

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

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

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

Then we have

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

Plugging $\hat{\beta}_{OLS}$ into (C.1), we get (noting that $\sum_i \alpha_i = 0$ due to the identification condition)

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

Next consider the properties of $\hat{\beta}_{*,OLS} = (\hat{\beta}_{0,OLS},\hat{\beta}_{OLS}')'$ under the alternative H_1^L. Under H_1^L, we have

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

Averaging (C.6) over t gives

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

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

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

Then, under the alternative H_1^L the OLS estimator in (C.4) is

$$\begin{aligned}
\hat{\beta}_{OLS} - \beta &= \left[\sum_{i,t} (x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1} \left[\sum_{i,t} (x_{it} - x_{iA})(y_{it} - y_{iA})\right] - \beta \\
&= \rho_n \left[\sum_{i,t} (x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1} \left[\sum_{i,t} (x_{it} - x_{iA})(g_0(z_t) - T^{-1} \sum_t g_0(z_t))\right] \\
&\quad + \rho_n \left[\sum_{i,t} (x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1} \left[\sum_{i,t} (x_{it} - x_{iA})(x_{it}' g(z_t) - T^{-1} \sum_t x_{it}' g(z_t))\right] \\
&\quad + \left[\sum_{i,t} (x_{it} - x_{iA})(x_{it} - x_{iA})'\right]^{-1} \left[\sum_{i,t} (x_{it} - x_{iA})(u_{it} - u_{iA})\right].
\end{aligned} \tag{C.9}$$

Since $\mathbb{E}[(x_{it} - x_{iA})|z_s] = 0$ due to Assumption 2 (b), $\mathbb{E}(x_{it} - x_{iA})(g_0(z_t) - T^{-1} \sum_t g_0(z_t)) = 0$. Then by central limit theory as $n \to \infty$, $n^{-1/2} \sum_{i,t} (x_{it} - x_{iA})(g_0(z_t) - T^{-1} \sum_t g_0(z_t)) = O_p(1)$. Thus the first term in the RHS of (C.9) is of order $O_p(\rho_n/\sqrt{n})$. For the second term in the RHS of (C.9), in view of Assumption 2 (b), we have

$$\begin{aligned}
&\mathbb{E}[(x_{it} - x_{iA})(x_{it}' g(z_t) - T^{-1} \sum_s x_{is}' g(z_s))] \\
=&\mathbb{E}[(x_{it} - x_{iA}) x_{it}'] \mathbb{E} g(z_t) - \mathbb{E}\left[(x_{it} - x_{iA}) T^{-1} \sum_s x_{is}'\right] \mathbb{E} g(z_s) \\
=&\mathbb{E}\left[(x_{it} - x_{iA})(x_{it} - T^{-1} \sum_s x_{is})'\right] \mathbb{E} g(z_t) \neq 0
\end{aligned} \tag{C.10}$$

since $\mathbb{E} g(z_t) \neq 0$. Therefore, this term is of order $O_p(\rho_n)$. Analysis for the third term is standard and it has zero mean and is of order $O_p(1/\sqrt{NT})$. Combining these results, we have

$$\hat{\beta}_{OLS} - \beta - \mathcal{B} = O_p(1/\sqrt{n}) \tag{C.11}$$

under the local alternative H_1^L, where the bias term $\mathcal{B} = O_p(\rho_n)$.

Now consider the OLS estimator $\hat{\beta}_{0,OLS}$ given in (C.5). We have

$$\begin{aligned}\hat{\beta}_{0,OLS} - \beta_0 =& \frac{1}{NT}\sum_{i,t}(y_{it} - x_{it}'\hat{\beta}_{OLS}) - \beta_0 \\ =& \rho_n \frac{1}{T}\sum_t g_0(z_t) + \frac{1}{n}\sum_{i,t} x_{it}'[\beta - \hat{\beta}_{OLS}] + \rho_n \frac{1}{n}\sum_{i,t} x_{it}' g(z_t) + \frac{1}{n}\sum_{it} u_{it}. \qquad \text{(C.12)}\end{aligned}$$

The first term on the RHS of (C.12) is of order $O_p(\rho_n)$ because $\mathbb{E}g_0(z_t) \neq 0$. The second term is of the same order as $\hat{\beta}_{OLS} - \beta$, which is $O_p(\rho_n + 1/\sqrt{n})$. The third term is of order $O_p(\rho_n)$ because $\mathbb{E}[x_{it}'g(z_t)] \neq 0$. The last term is of order $O_p(1/\sqrt{n})$. So we have

$$\hat{\beta}_{0,OLS} - \beta_0 - \mathcal{B}_0 = O_p(1/\sqrt{n}) \qquad \text{(C.13)}$$

under the local alternative H_1^L, and the bias $\mathcal{B}_0 = O_p(\rho_n)$. Combining (C.11) and (C.13), we have

$$\hat{\beta}_{*,OLS} - \beta_* - \mathcal{B}_* = O_p(1/\sqrt{n}), \qquad \text{(C.14)}$$

under the local alternative H_1^L, and the bias $\mathcal{B}_* = O_p(\rho_n)$.

Since

$$\begin{aligned}\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} &= \hat{\beta}_{*,PLC}(z) - \beta_*(z) + \beta_*(z) - \hat{\beta}_{*,OLS} \\ &= \hat{\beta}_{*,PLC}(z) - \beta_*(z) + \beta_* - \hat{\beta}_{*,OLS} + \rho_n g_*(z), \qquad \text{(C.15)}\end{aligned}$$

and since the nonparametric estimator $\hat{\beta}_{*,PLC}(z)$ is always $\sqrt{NT|H|}$-consistent under the alternative and $g_*(z)$ is bounded, we have (with undersmoothing to remove the bias of $\hat{\beta}_{*,PLC}(z)$)

$$\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS} - \mathcal{B}^* = O_p(1/\sqrt{n|H|}), \qquad \text{(C.16)}$$

where the bias $\mathcal{B}^* = O_p(\rho_n)$. Therefore,

$$\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(\sqrt{n|H|}\rho_n) + O_p(1), \qquad \text{(C.17)}$$

where the first term $O_p(\sqrt{n|H|}\rho_n)$ comes from the bias.

If $\sqrt{n|H|}\rho_n \to 0$, then in this case (3.2) continues to hold. Thus we still have $I_m^* \Rightarrow \chi^2_{(p+1)m}$ and $J \Rightarrow N(0,1)$ under the local alternative H_1^L. This means the tests have asymptotic power equals to size for such alternatives.

If $\sqrt{n|H|}\rho_n = O(1)$, then $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS}) = O_p(1)$, but with $O_p(1)$ bias. Thus $\sqrt{n|H|}(\hat{\beta}_{*,PLC}(z) - \hat{\beta}_{*,OLS})$ is asymptotically non-central normal. Hence the test statistic I_m^* has a non-central limiting $\chi^2_{(p+1)m}$ distribution. 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. Thus, the test I_m^* has non-trivial local asymptotic power.

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

D Useful lemmas

Lemma D.1. *Let $P_N = diag(1, \sqrt{N} I_p)$. Under Assumptions 1-3, as $T \to \infty$, we have the following:*

1. *if N is fixed,*

$$\frac{1}{T|H|}(X_A^*)' K_z X_A^* \xrightarrow{p} f_z(z)\bar{V}_{xx},$$

where

$$\bar{V}_{xx} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix};$$

2. *if $N \to \infty$ and $\mathbb{E}x_{it} = 0$,*

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

where

$$V_{xx}^* = \begin{pmatrix} 1 & 0 \\ 0 & V_{xx} \end{pmatrix};$$

3. *if $N \to \infty$ and $\mathbb{E}x_{it} \neq 0$,*

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

where

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

Proof Note that

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

Denote $\tilde{x}_t = x_{At}$. When N is small but T goes to infinity, we have

$$\frac{1}{T|H|}(X_A^*)'K_z X_A^* \xrightarrow{p} f_z(z)\begin{pmatrix} 1 & \mathbb{E}(\tilde{x}_t') \\ \mathbb{E}(\tilde{x}_t) & \mathbb{E}(\tilde{x}_t\tilde{x}_t') \end{pmatrix} \equiv f_z(z)\bar{V}_{xx}.$$

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

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

Then we have

$$\bar{V}_{xx} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}V_{xx} + (1-\frac{1}{N})\eta\eta' \end{pmatrix} = \begin{pmatrix} 1 & \eta' \\ \eta & \frac{1}{N}\Sigma_{xx} + \eta\eta' \end{pmatrix},$$

which is close to singular if η is large and the variance of x_{it} is small.

However, if $\mathbb{E}x_{it} = 0$, then $x_{At} = O_p(1/\sqrt{N}) = o_p(1)$ as $N \to \infty$. If we still divide $(X_A^*)'K_z X_A^*$ by $T|H|$, we have

$$\frac{1}{T|H|}(X_A^*)'K_z X_A^* \xrightarrow{p} f_z(z)\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

which is again non-invertible. However, since $\sqrt{N}x_{At} \Rightarrow N(0, V_{xx})$, we have

$$\frac{1}{T|H|}P_N(X_A^*)'K_z X_A^* P_N \xrightarrow{p} f_z(z)\begin{pmatrix} 1 & 0 \\ 0 & V_{xx} \end{pmatrix} \equiv f_z(z)V_{xx}^*,$$

where $P_N = diag(1, \sqrt{N}I_p)$.

If $\mathbb{E}x_{it} = \eta \neq 0$, then $x_{At} \xrightarrow{p} \eta$, as $N \to \infty$. This gives

$$\frac{1}{T|H|}(X_A^*)'K_z X_A^* \xrightarrow{p} f_z(z)\begin{pmatrix} 1 & \eta' \\ \eta & \eta\eta' \end{pmatrix} \equiv f_z(z)V_\eta,$$

which is non-invertible. ■

Lemma D.2. *Under Assumptions 1-3, as $T \to \infty$, we have*

1. *if N is fixed,*

$$\frac{1}{\sqrt{T|H|}}(X_A^*)'K_z U_A \Rightarrow N(0, f_z(z)\sigma_u^2 N^{-1}\nu_0^q \bar{V}_{xx});$$

2. *if $N \to \infty$ as $T \to \infty$ and $\mathbb{E}x_{it} = 0$,*

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

3. if $N \to \infty$ as $T \to \infty$ and $\mathbb{E}x_{it} \neq 0$,

$$\frac{1}{\sqrt{T|H|}}(X_A^*)'K_zU_A \Rightarrow N(0, f_z(z)\sigma_u^2\nu_0^q V_\eta).$$

Proof Note first that

$$(X_A^*)'K_zU_A = \begin{pmatrix} \sum_t K_{tH}u_{At} \\ \sum_t K_{tH}x_{At}u_{At} \end{pmatrix}. \tag{D.1}$$

has zero mean. When N is small, denote $\tilde{u}_t = u_{At}$ and $\sigma_{\tilde{u}}^2 = Var(\tilde{u}_t) = \sigma_u^2/N$. Then for the first element in (D.1), we have

$$Var(\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}u_{At}) = \frac{1}{T|H|}\sum_t \mathbb{E}[K_{tH}^2 u_{At}^2] + \frac{1}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}K_{sH}u_{As}] \to f_z(z)\nu_0^q\sigma_{\tilde{u}}^2,$$

because

$$\begin{aligned}\frac{1}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}K_{sH}u_{As}] =& \frac{2}{T|H|}\sum_{\ell=1}^{T-1}(T-\ell)\mathbb{E}[K_{1H}\tilde{u}_1 K_{1+\ell,H}\tilde{u}_{1+\ell}] \\ =& \frac{C}{|H|}\sum_{\ell=1}^{T-1}(1-\ell/T)\gamma_{\tilde{u}}(\ell)|H|^2 + \textit{smaller order} \\ \leq& C|H|N^{-1}\sum_{\ell=1}^{T-1}|\gamma_u(\ell)| = o(1). \end{aligned} \tag{D.2}$$

For the second element in (D.1), we have

$$\begin{aligned}Var(\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At}) =& \frac{1}{T|H|}\sum_t \mathbb{E}[K_{tH}^2 x_{At}x_{At}' u_{At}^2] + \frac{1}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}x_{At}x_{As}'K_{sH}u_{As}] \\ &\to f_z(z)\nu_0^q\sigma_{\tilde{u}}^2\mathbb{E}[\tilde{x}_t\tilde{x}_t'],\end{aligned}$$

because

$$\begin{aligned}\frac{1}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}x_{At}x_{As}'K_{sH}u_{As}] =& \frac{2}{T|H|}\sum_{\ell=1}^{T-1}(T-\ell)\mathbb{E}[K_{1H}\tilde{u}_1\tilde{x}_1 K_{1+\ell,H}\tilde{u}_{1+\ell}\tilde{x}_{1+\ell}'] \\ \leq& C|H|\sum_{\ell=1}^{T-1}\tilde{\alpha}_\ell^{1-2/\delta} = O(|H|) = o(1),\end{aligned} \tag{D.3}$$

where the $\tilde{\alpha}_\ell$ are the α-mixing coefficients of the process $\tilde{u}_t$ and $\tilde{x}_t$, which satisfy Assumption 2 (a).

Similarly, we can show that the covariance of the two elements in (D.1) satisfies

$$Cov(\frac{1}{\sqrt{T|H|}}\sum_t K_{tH}u_{At}, \frac{1}{\sqrt{T|H|}}\sum_t K_{tH}x_{At}u_{At}) \to f_z(z)\nu_0^q\sigma_{\tilde{u}}^2\mathbb{E}[\tilde{x}_t].$$

Therefore, by the CLT for the α-mixing process $\{(\tilde{x}_1,\tilde{u}_1),\cdots,(\tilde{x}_T,\tilde{u}_T)\}$, we have

$$\frac{1}{\sqrt{T|H|}}(X_A^*)'K_zU_A \Rightarrow N(0, f_z(z)\nu_0^q\sigma_{\tilde{u}}^2\bar{V}_{xx}).$$

Results for N goes to infinity and $\eta \neq 0$ can be obtained by letting $N \to \infty$. The analysis is the same and is omitted.

When N goes to infinity and $\mathbb{E}x_{it} = 0$, for the first element in (D.1), we have

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

by similar arguments to those of (D.2). For the second element in (D.1), we have

$$\begin{aligned}Var(\sqrt{\frac{N^2}{T|H|}}\sum_t K_{tH}x_{At}u_{At}) &= \frac{N^2}{T|H|}\sum_t \mathbb{E}[K_{tH}^2x_{At}x_{At}'u_{At}^2] + \frac{N^2}{T|H|}\sum_{t\neq s}\mathbb{E}[K_{tH}u_{At}x_{At}x_{As}'K_{sH}u_{As}] \\ &\to f_z(z)\nu_0^q\sigma_u^2V_{xx},\end{aligned}$$

by an argument similar to (D.3). We can further show that the covariance of the two elements in (D.1) satisfies

$$Cov(\sqrt{\frac{N}{T|H|}}\sum_t K_{tH}u_{At}, \sqrt{\frac{N^2}{T|H|}}\sum_t K_{tH}x_{At}u_{At}) \to 0.$$

Therefore, when N goes to infinity and $\mathbb{E}x_{it} = 0$, we have

$$\sqrt{\frac{N}{T|H|}}P_N(X_A^*)'K_zU_A \Rightarrow N(0, f_z(z)\nu_0^q\sigma_u^2V_{xx}^*).$$

■

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

1. *if N is fixed,*

$$\frac{1}{T|H|}(X_A^*)'K_z diag(X_A^*\ddot{\beta}_*(z)) \xrightarrow{p} f_z(z)\mu_2\bar{V}_{xx}\sum_{j=1}^q h_j^2\frac{\partial^2\beta_*(z)}{\partial^2 z_j};$$

2. if $N \to \infty$ and $\mathbb{E}x_{it} = 0$,

$$\frac{1}{T|H|}P_N(X_A^*)'K_z diag(X_A^* P_N \ddot{\beta}_*(z)) \xrightarrow{p} f_z(z)\mu_2 V_{xx}^* \sum_{j=1}^{q} h_j^2 \frac{\partial^2 \beta_*(z)}{\partial^2 z_j};$$

3. if $N \to \infty$ and $\mathbb{E}x_{it} \neq 0$,

$$\frac{1}{T|H|}(X_A^*)'K_z diag(X_A^* \ddot{\beta}_*(z)) \xrightarrow{p} f_z(z)\mu_2 V_\eta \sum_{j=1}^{q} h_j^2 \frac{\partial^2 \beta_*(z)}{\partial^2 z_j}.$$

Proof The proofs are routine and are omitted. ■

Lemma D.4. *Under Assumptions 1-3, as $T \to \infty$, we have*

$$\frac{1}{NT|H|}X_*'K(z)X_* \xrightarrow{p} f_z(z)\tilde{V}_{xx}, \tag{D.4}$$

where

$$\tilde{V}_{xx} = \begin{pmatrix} 1 & \eta' \\ \eta & V_{xx} \end{pmatrix}.$$

Proof Note that

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

The stated convergence follows from standard kernel limit theory for stationary data. ■

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

$$\frac{1}{\sqrt{NT|H|}}X_*'K(z)U \Rightarrow N\left(0, f_z(z)\sigma_u^2 \nu_0^q \tilde{V}_{xx}\right). \tag{D.5}$$

Proof Note that

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

The required weak convergence follows by standard methods for stationary data. ■.

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

$$\frac{1}{NT|H|}X_*'K(z)diag[X_*(1_{1\times N} \otimes \dot{\beta}_*(z))] \xrightarrow{p} \mu_2 \tilde{V}_{xx} \sum_{s=1}^{q} h_s^2 \frac{\partial \beta_*(z)}{\partial z_s}\frac{\partial f_z(z)}{\partial z_s},$$

$$\frac{1}{NT|H|}X_*'K(z)diag[X_*(1_{1\times N} \otimes \ddot{\beta}_*(z))] \xrightarrow{p} f_z(z)\mu_2 \tilde{V}_{xx} \sum_{s=1}^{q} h_s^2 \frac{\partial^2 \beta_*(z)}{\partial^2 z_s}.$$

Proof The proofs are straightforward and the computations are omitted. ■.

References

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.

Magnus, J. R., Melenberg, B., and Muris, C. (2011). Global warming and local dimming: the statistical evidence. *Journal of the American Statistical Association*, 106(494):452–464.

Phillips, P. C. B., Leirvik, T., and Storelvmo, T. (2019). Econometric estimates of Earth's transient climate sensitivity. *Journal of Econometrics*.

Phillips, P. C. B. and Wang, Y. (2019). Functional coefficient panel modeling with communal smoothing covariates: the case without intercept. *Working paper*.

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.

Storelvmo, T., Heede, U. K., Leirvik, T., Phillips, P. C. B., Arndt, P., and Wild, M. (2018). Lethargic response to aerosol emissions in current climate models. *Geophysical Research Letters*, 45(18):9814–9823.

Storelvmo, T., Leirvik, T., Lohmann, U., Phillips, P. C. B., and Wild, M. (2016). Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity. *Nature Geoscience*, 9(4):286.

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

Su, L., Shi, Z., and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.

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. In *Nonparametric econometric methods*, pages 101–129. Emerald Group Publishing Limited.

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.

Wild, M. (2012). Enlightening global dimming and brightening. *Bulletin of the American Meteorological Society*, 93(1):27–37.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.

Table 1: $AMSE(\beta_0(z))$

			$\hat{\beta}_{APLC,0}(z)$	$\hat{\beta}_0^{oracle}(z)$	$\hat{\beta}_{PLC,0}(z)$
$N=5$	$T=20$	$\eta=0$	0.0431	0.0371	0.0373
		1	0.1593	0.0524	0.0531
		5	1.1456	0.0954	0.0969
	$T=50$	$\eta=0$	0.0181	0.0165	0.0165
		1	0.0519	0.0237	0.0238
		5	0.2881	0.0402	0.0405
	$T=100$	$\eta=0$	0.0106	0.0098	0.0098
		1	0.0276	0.0132	0.0132
		5	0.1264	0.0222	0.0224
$N=50$	$T=20$	$\eta=0$	0.0163	0.0064	0.0064
		1	0.4648	0.0077	0.0078
		5	9.1880	0.0112	0.0114
	$T=50$	$\eta=0$	0.0080	0.0031	0.0031
		1	0.1077	0.0040	0.0041
		5	2.0186	0.0055	0.0056
	$T=100$	$\eta=0$	0.0049	0.0017	0.0017
		1	0.0450	0.0023	0.0023
		5	0.8907	0.0031	0.0031

Table 2: $AMSE(\beta_1(z))$

			$\hat{\beta}_{APLC,1}(z)$	$\hat{\beta}_1^{oracle}(z)$	$\hat{\beta}_{PLC,1}(z)$
$N=5$	$T=20$	$\eta=0$	0.2038	0.0355	0.0375
		1	0.1397	0.0252	0.0263
		5	0.0699	0.0121	0.0123
	$T=50$	$\eta=0$	0.0715	0.0169	0.0171
		1	0.0503	0.0125	0.0127
		5	0.0284	0.0073	0.0074
	$T=100$	$\eta=0$	0.0384	0.0103	0.0104
		1	0.0279	0.0076	0.0076
		5	0.0164	0.0051	0.0051
$N=50$	$T=20$	$\eta=0$	0.4886	0.0066	0.0068
		1	0.4571	0.0048	0.0049
		5	0.3846	0.0032	0.0032
	$T=50$	$\eta=0$	0.1347	0.0035	0.0036
		1	0.1140	0.0029	0.0029
		5	0.0961	0.0021	0.0021
	$T=100$	$\eta=0$	0.0613	0.0022	0.0022
		1	0.0499	0.0018	0.0018
		5	0.0465	0.0014	0.0014

Table 3: Size of the two tests (in percentage, with 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
	$T = 50$	2.50	3.00	2.00	2.50
	$T = 100$	2.00	2.50	3.50	5.50
$m = 9$	$T = 20$	6.50	8.00	2.50	3.00
	$T = 50$	2.00	3.50	2.50	2.50
	$T = 100$	1.50	2.50	1.50	2.00
$m = 20$	$T = 20$	12.50	12.50	5.50	6.50
	$T = 50$	10.50	11.00	6.50	6.50
	$T = 100$	3.50	5.00	4.00	4.50

Table 4: Size of the bootstrapped test (in percentage, nominal size=5%)

		$N = 5$	$N = 20$
$m = 3$	$T = 20$	6.00	7.50
	$T = 50$	6.50	3.50
	$T = 100$	3.00	6.0
$m = 9$	$T = 20$	5.00	5.50
	$T = 50$	4.50	4.00
	$T = 100$	4.00	4.00
$m = 20$	$T = 20$	6.00	5.50
	$T = 50$	7.50	5.50
	$T = 100$	3.00	4.00