CROSS SECTION CURVE DATA AUTOREGRESSION

By

Peter C. B. Phillips and Liang Jiang

June 2026

COWLES FOUNDATION DISCUSSION PAPER NO. 2439R1

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Cross Section Curve Data Autoregression*

Peter C. B. Phillips[†] and Liang Jiang[††]

[†] Yale University, University of Auckland & Singapore Management University,

[††] International School of Finance, Fudan University

June 5, 2026

Abstract

This paper develops and applies new asymptotic theory for estimation and inference in parametric autoregression and vector autoregression with function valued cross section curve time series. The study provides a new approach to dynamic panel regression with high dimensional dependent cross section data. We deal with the stationary case and provide a full set of results extending those of standard Euclidean space autoregression, showing how function space curve cross section data raises efficiency and reduces bias in estimation and shortens confidence intervals. Wild-bootstrap methods are also developed to improve testing and inference. The findings reveal that function space models with wide-domain and narrow-domain cross section dependence provide insights on the effects of various forms of cross section dependence in discrete dynamic panel models with fixed and interactive fixed effects. The methodology is applicable to panels of high dimensional wide datasets that are now available in many longitudinal studies. An empirical illustration is provided that sheds light on household Engel curves among ageing seniors in Singapore using the Singapore life panel longitudinal dataset.

Keywords: Curved Cross Section Data, Dynamic panel, Functional AR model, Hilbert space, Specification test.

JEL classification: C21, C23

1 Introduction

Hilbert space formulations of regression models have been studied for several decades. Much of the methodology and inferential machinery has been developed in mathematics and is exposited in texts by Bosq (2000), Bosq and Blanke (2008), and Horváth and Kokoszka (2012).

*This paper belongs to a wider study on curve time series regression, deals with stationary autoregression, and has a sequel on unit root curve autoregressions (Phillips and Jiang, 2025). An early working paper that contained much of the limit theory findings in both papers was entitled 'Parametric Autoregression in Function Space' (Phillips and Jiang, 2016). That paper was not circulated as the authors awaited longitudinal data to emerge from Singapore life panel study. Parts of the present paper were presented at the TSF Symposium at the University of Sydney in November, 2023, and ESAM 2024 at Monash University. Phillips acknowledges partial research support from the Kelly Fund at the University of Auckland and an LKC Fellowship at Singapore Management University

In recognition of the availability of function space data in many scientific disciplines, descriptive methods and modeling techniques for function valued variables have also been developing rapidly (e.g. James (2010), Morris, 2014). This progress in methodology is matched by concurrent rapid expansion in the use of high-dimensional data and associated data science methods for practical implementation.

In economic theory, models often involve function valued variables that take the form of time varying densities or curves over a population of individuals, a spatial framework of trade, or human migration in economic geography. In these models, key notions such as income inequality are defined as functionals of underlying distributions whose natural support is a function space. Modern rational expectations models and dynamic stochastic general equilibrium models may also allow for function valued state variables and may be functionally linearized to produce Hilbert space models amenable to function space regression techniques (Childers, 2018). Much recent work in econometrics (Chang et al., 2016; Park and Qian, 2012; Hu et al., 2016; Beare, 2017; Beare et al., 2017; Beare and Seo, 2020; Li et al., 2025), among others, has utilized the statistical machinery of function space regression to perform estimation and inference in models involving variates such as state densities using functional principal component analysis and to study co-movement amongst function valued time series. Recent research on Wasserstein regression in statistics (Zhang et al., 2022) is also relevant to the present work and studies stationary time series methods for modeling probability distributions in optimal transport problems using the Wasserstein metric. This metric provides a measure of distance between distributions and enables transformation of the densities of these distributions to a linear Hilbert space by means of a log quantile density transformation (Chen et al., 2023). As we show later, the results of the present paper include this approach to studying distributions by using a weighted Hilbert space setting.

Function space regression and autoregression have many similarities in common with standard regression. For instance, sample moments of the data may be used to estimate population moments and central limit theory may be used to characterize Gaussian process limit theory for these moments in a function space setting under stationarity conditions. But there are important differences between function space and Euclidean space regressions. One complicating difference is that the coefficients or matrices of coefficients that arise in finite dimensional models are typically replaced in function space models by infinite dimensional operators whose estimation presents special challenges in view of covariance operator inversion difficulties. These challenges relate to those that arise in ill-posed inversion problems in nonparametric instrumental variable estimation of structural equations which require functional regularization techniques for resolution and typically lead to bias and reductions in convergence rates – see Hall and Horowitz (2005), Darolles et al. (2011), and Horowitz (2014) for a recent review. Bosq (2000) and Horváth and Kokoszka, 2012 describe these inversion challenges and possible solutions in function space regression in detail.

While these general challenges are important and have attracted attention both in the mathematical statistics and econometrics literatures, they are not the concern of the present paper which has more specific and modest goals in function space regression that are well suited to empirical applications. A primary motivation for our work comes from the many longitudinal studies that involve extremely wide panels. These high dimensional cross section panels offer new possibilities for inference that allow for nonparametric forms of cross section dependence, while having time series of observations on the functional variates. This data availability opens up the option of parametric Hilbert space methods. The present paper and it sequel explore this option. In particular, we consider parametric autoregressions with function valued time series variates, such as the first order autoregression

$$X_t(r) = \alpha(r) + \theta X_{t-1}(r) + u_t(r), \quad t = 1, 2, \ldots, n;\ r \in [a, b], \tag{1}$$

involving function valued time series $\{X_t, u_t\}_{t=1}^n$ where the observed variate X_t, the unobserved error u_t, and the constant fixed effect curve $\alpha(r)$ all lie in some Hilbert space such as $L_2[a, b] = \left\{ f : \int_a^b f^2 < \infty \right\}$ with inner product $\langle f, g \rangle = \int_a^b fg$ taken over some finite observational interval $[a, b]$. It is particularly convenient in applications to use càdlàg functions in $L_2[a, b]$ because this allows for well-defined realized values with potential discontinuities but also allows for specifications in which the curve fixed effect $\alpha(r)$ may have regions of constancy, thereby accommodating cross section cluster effects, on which much panel data empirical work in econometrics has focussed (MacKinnon et al., 2023). The model (1) is a special case of a first order autoregression in Hilbert space (an ARH(1)) in which the operator θ here is a simple scalar multiplier, just as it is in the usual finite dimensional Euclidean space case.

An important advantage of the model (1) is its close relationship to popular dynamic panel models such as the simple panel autoregression

$$X_{it} = \alpha_i + \theta X_{it-1} + u_{it},\ i = 1, ..., N; t = 1, ..., n \tag{2}$$

in which the AR coefficient θ is homogeneous across individuals i in the panel and the α_i are individual fixed effects. In (1) the α_i of (2) are replaced by functional fixed effects $\alpha(r)$ which give a nonparametric formulation of the intercept that enables sensitivity of the functional data $X_t(r)$ to individual position $r \in [a, b]$ in the domain of the function space. Similarly, two way fixed effects models in which (2) is augmented by discrete temporal intercepts β_t have functional analogues extending (1) by the inclusion of either a discrete or curve trajectory intercept (c.f., Li et al. (2020)). But whereas it is often conventional to assume that the error components u_{it} in the panel (2) are independent over i or at least independent conditional on common shocks or factors, in the ARH model (1) the error component u_t is assumed to be a stochastic process with a general covariance kernel $k_u(r, s) = \mathbb{E}(u_t(r) u_t(s)) \in L_2([a, b]^2)$ that allows great flexibility in terms of the permissable cross section dependence structure in the model. Notably, $k_u(r, s)$ may be non-zero for all pairs (r, s) so that there may be cross section dependence as well as heterogeneity throughout the interval $[a, b]$. This framework seems well suited to practical work where general forms of cross section dependence may be desirable. There is another connection between the two models at the other extreme where the u_{it} in

(2) are assumed to be iid. In model (1), this case corresponds to a covariance kernel where $k_u(r,s) = k(|r-s|) = 0, \forall r \neq s$, as in a Dirac delta function $k(\cdot)$. As explained in Remarks 1(d) and 1(e) in Section 2.2 below, this case may be treated in the present framework by way of a limiting process that links the curve autoregression and dynamic panel models and their associated rates of convergence by employing a suitably designed covariance kernel $k_N(\cdot)$ and letting $N \to \infty$. Second, model (1) allows for a continuum of fixed effects, including possible cluster effects over subintervals, in terms of the function ordinate $r \in [a,b]$ rather than the countable infinite set of incidental parameters $\{\alpha_i\}$ in (2). A third advantage of (1) is that the model readily accommodates unit root (UR) persistence in which $\theta = 1$, as well as roots in the vicinity of unity, such as local unit roots (LURs) with $\theta_n = 1 + c/n$ for some constant c, functional local unit roots (FLURs) with $\theta_{tn} = 1 + c(t/n)/n$ for some constant function $c(\cdot)$ (Bykhovskaya and Phillips, 2018, 2020) and various extensions of these specifications that allow for mildly integrated and mildly explosive roots (Phillips and Magdalinos, 2007; Phillips, 2023). Models with such persistent data are treated in Phillips and Jiang (2025). A further option available in the parametric setting of (1) is specification testing of a simple null such as $\mathcal{H}_0 : \theta = \theta_0$, involving some specific parameter value θ_0 against a general functional alternative such as $\mathcal{H}_1 : \theta = \theta_0 + \frac{\psi(r)}{\sqrt{n}}$ for $\psi \in L_2[a,b]$, which enables a statistical check on whether the scalar operator form of θ in the model (1) is supported by the data against a more complex operator formulation of the ARH(1).

Thus, in spite of its apparent simplicity, the parametric ARH (1) opens up the study of stationary and stochastically nonstationary function valued processes that provide linkages to dynamic panel regression and a first step towards fully infinite dimensional extensions of autoregressions involving operator forms of the parameter θ. Finally, together with nonparametric smoothing methods, (1) is easy to apply to unbalanced or pseudo panel datasets, or datasets with missing observations. The estimated function values that are obtained from nonparametric smoothing or methods such as random forest can then be utilized in the ARH model for regression. It is known that under certain conditions such nonparametric fitting of the curves from discrete data does not disturb asymptotic properties of the ARH parametric estimates (see, for example, Zhang and Chen (2007); Cho et al. (2022)).

In the context of the ARH model (1) and higher order extensions, this paper develop an asymptotic theory of estimation and inference for stationary cases. Comparisons are provided with conventional scalar autoregression, showing how function space data can raise efficiency in estimation and shorten confidence intervals in inference about θ. Results for the ARH model are explicitly related to those of dynamic panel regression, showing the impact of a continuum of fixed effects on inference in function space regression. Specifications tests are also provided for cross section breaks and against more general functional alternatives.

A final contribution of the paper provides an empirical illustration with data on household Engel curves in Singapore for seniors with ages from 50 to 70 years. This application employs the recently developed Singapore life panel (SLP), which comprises a wide panel of 10,000 individuals whose consumption behavior over multiple commodity groupings has been observed

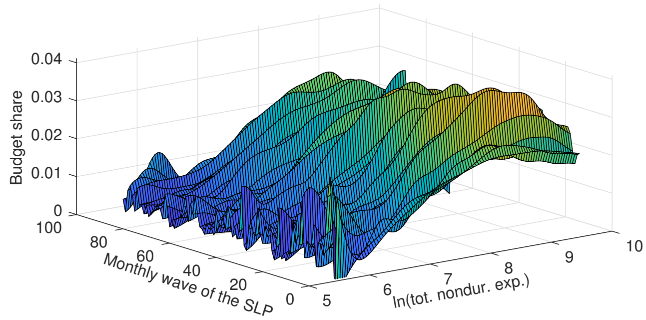


Figure 1: Clothing Engel Curves for Singaporeans aged 50-70

monthly over the time period since 2015. The SLP directly monitors senior Singaporean citizen behavior as demographics evolve with an ageing population, providing interpretable findings based on curve time series data revealing the dynamics of household Engel curves over time for various nondurable expenditures. This application to the SLP falls into the category of high dimensional wide panel data analysis and shows that our methods have potential applicability in many other comparable longitudinal settings.

Consider, for example, the smoothed Engel curve for clothing expenditure in Figure 1. The x-axis plots the log of total non-durable expenditure, the y-axis indicates the date (month or wave of the SLP survey), and the vertical axis shows the budget share for clothing. This figure highlights how expenditure on clothing changes with income levels: starting high for lower-income groups, decreasing in the middle-income bracket, then increasing again for higher incomes as spending on luxury items grows, before decreasing once more at the highest income levels. This general pattern of expenditure is largely sustained over time, which underscores the significance of analyzing function valued cross-section time series.

It is important that our empirical analysis uses curved cross section data as time series observations that live in the convenient and commonly used space $L_2[a,b]$ with the standard inner product. This formulation is convenient and appropriate in the present setting because the space that contains curved cross section positive expenditure data such as that in Figure 1 lies in $C_+[a,b] \subset L_2[a,b]$ and so it is possible to think of the space $C_+[a,b]$ as being embedded in $L_2[a,b]$, which therefore provides a Hilbert geometry that is particularly useful in regression analysis and the development of its limit theory.

The paper is organized as follows. The model, its parametric form, estimation, inference, bias analysis, and asymptotic theory in the stationary case together with cross section break and specification testing are all given in Section 2. Section 3 considers functional fixed effects in the ARH model with comparisons to the dynamic fixed effects panel regression model. Simulations are reported in Section 4. Section 5 reports an empirical investigation of household Engel curves for ageing seniors using data from the Singapore Life Panel (SLP). Section 6 concludes. Proofs, additional technical material, and additional simulations are in Appendices in Sections 7, 8, 9, and 10. Institutional details of the SLP are given in Section 11.

2 Parametric Hilbert Space Autoregression

The present paper studies a simple class of ARH model where the Hilbert space regression is parameterized in a finite dimensional way so that the operator is simply a scalar parameter or group of scalar parameters to be estimated. This class is especially useful in applied econometrics because it has advantages in empirical work with wide short panels with functional curve data in the cross section dimension. We start with a first order autoregression and later provide higher order ARH(p) and VAR extensions.

2.1 The functional ARH(1)

To fix ideas some basic concepts on functional regression are outlined here. We start by considering a pure functional ARH(1) model with no functional intercept (or functional fixed effects) and time series sample size n, viz.,

$$X_t = AX_{t-1} + u_t,\ t = 1, ..., n, \tag{3}$$

where u_t and X_t are random elements in the Hilbert space $\mathcal{H} = L_2[a,b]$ for some real $a < b$ with inner product $\langle x, y\rangle = \int_a^b xy$ for $x, y \in L_2[a,b]$ and norm $\|x\| = \langle x, x\rangle^{1/2}$. In (3) A is an operator on $L_2[a,b]$ such that $Ax \in L_2[a,b]$ for all $x \in L_2[a,b]$. In a commonly used example, A is a kernel operator defined by $A_\ell(x)(r) = \int_a^b \ell(s,r)\, x(s)ds$, with $x \in L_2[a,b]$, $r \in [a,b]$, and kernel function ℓ satisfying $\int_a^b \int_a^b \ell(s,r)^2 dsdr < \infty$. Then, $A_\ell x \in L_2[a,b]$ and A_ℓ is a bounded linear operator on $L_2[a,b]$ with operator norm

$$\|A_\ell\|_{op} = \left(\sup_{\|x\|\le 1} \left\| \int_a^b \ell(s,r)\, x(s) ds \right\|^2\right)^{1/2} \le \left(\int_a^b \int_a^b \ell(s,r)^2\, dsdr\right)^{1/2} < \infty.$$

With this operator the model (3) has the useful coordinate form

$$X_t(r) = \int_a^b \ell(s,r) X_{t-1}(s) ds + u_t(r), \quad r \in [a,b] \tag{4}$$

which may be considered a function space extension of a common vector autoregression (VAR) in $\mathbb{R}^m$. As the dimension m of a VAR increases, the number of coefficients to be estimated increases at the rate $O(m^2)$; and problems of finite sample autoregressive bias correspondingly increase, approximately proportional to the VAR order m (see Abadir et al. (1999); Lawford and Stamatogiannis (2009)). Similar problems arise in high dimensional cointegrated VAR settings where inference can be severaly impacted, as demonstrated in recent work by Bykhovskaya and Gorin (2022). In practical work with large VAR systems, there is less interest in the individual VAR coefficients and more attention is given to their implications in terms of forecast error variance decompositions and impulse reponse paths, although the relevance of large sample asymptotics derived from the fitted model is inevitably affected. In the function space ARH (4) the coefficient operator A_ℓ is infinite dimensional and estimation is substantially more complex because of operator inversion problems in moving from covariance operators to autocorrelation operators which arise in the estimation of the operator A, e.g., Bosq (2000, ch. 8). Especially when the time series sample size is small or moderate, as it often is in short and extremely wide panels, there is therefore interest in using more parsimonious systems than (4).

In the work that follows, we will assume that A is the simple scalar multiplier operator defined by $A_\theta (x) (r) = \theta x (r)$ where $\theta \in \mathbb{R}$. Again, A_θ is a bounded linear operator with operator norm $\|A_\theta\|_{op} = \left(\sup_{\|x\|\leq 1} \|\theta x (r)\|^2\right)^{1/2} = |\theta| < 1$, which assures stationarity in (3). The coordinate form of the model in this case is simply

$$X_t(r) = \theta X_{t-1}(r) + u_t(r), \quad r \in [a, b], \tag{5}$$

which we call a parametric autoregression in Hilbert space. Functional least squares estimation of (5) involves estimation of the scalar parameter θ, just as in scalar autoregression, and nonparametric estimation of the covariance kernel function $k_u (r, s) = \mathbb{E} [u_t (r) u_t (s)]$ of the error process $u_t(r)$. In the stable case ($|\theta| < 1$) it is often assumed that u_t is an L_2- valued *iid* process (pure noise) over time with zero mean, conditional variance $\mathbb{E}u_t(r)^2 = k_u(r, r)$ and overall variance $\mathbb{E} \|u_t\|^2 = \int_a^b k_u (r, r) dr < \infty$ (Bosq, 2000), which we write as $iid (0, k_u)$. This framework allows for cross section interdependence via $k_u(r, s)$ for $r \neq s$ and heterogeneity in the variance $k_u (r, r)$ over $r \in [a, b]$. Several examples are given later in this section. In nonstationary cases the assumption of temporal independence may be relaxed in favor of some general form of weak dependence in u_t, as considered in Phillips and Jiang (2025), enabling extension of the methods of nonstationary scalar autoregression (Phillips, 1987a,b). It is sometimes convenient to write $u_t = \sigma\varepsilon_t$ where ε_t is $\mathcal{H}$- valued with covariance kernel $k (r, s) = \mathbb{E} (\varepsilon_t (r) \varepsilon_t (s))$ and $\sigma \in \mathbb{R}^+$ is a scale parameter, in which case a normalization such as $\int_a^b k(r, r)dr = 1$ can be employed. Extensions to a framework in which there may be unconditional temporal heterogeneity in σ will naturally be of interest in some applications but these are not treated here.

2.2 Stationary and ergodic parametric ARH(1)

The following assumption is used to develop a limit theory for the parametric ARH(1). We work with the stationary case, so that (5) mirrors a standard AR(1) model (where r is a fixed constant), and let u_t be an $L_2[a, b]$- valued martingale difference process in discrete time t. We use the following conditions.

Assumption A1.

(i) $\{u_t\}$ *is a stationary and ergodic* $L_2[a, b]$*-valued martingale difference sequence over* t *accompanied by the natural filtration* $\mathcal{F}_t = \sigma (u_t, u_{t-1}, ...)$, *covariance kernel* $k_u(r, s) \in L_2 [a, b]^2$, *and aggregate variance* $\sigma^2_{u,ab} = \int_a^b \mathbb{E}u_0(r)^2 dr = \int_a^b k_u(r, r)dr$, *which we collectively write as* $u_t \sim mds (0, k_u)$. *Fourth order moments of* u_t *are finite and* u_t *has càdlàg realized sample paths* $u_t(r)$.

(ii) $|\theta| < 1$.

Under A1(i) & (ii) and with initial conditions in the infinite past, the following elementary properties hold and follow directly from standard theory for stationary ARH models. Readers are referred to Bosq (2000) for an introduction to ARH models, and to Métivier (2011) for a discussion of spaces of Hilbert-valued martingales and stochastic integration with respect to such processes.

First, X_t has the linear process representation $X_t = \sum_{j=0}^{\infty} \theta^j u_{t-j}$ and its covariance operator is $C_X = \sum_{j=0}^{\infty} \theta^{2j} C_u = \left(1-\theta^2\right)^{-1} C_u$ where C_u is the covariance operator of u_t defined by $C_u(x) = \mathbb{E}\left[\langle u_t, x\rangle u_t\right] = \mathbb{E}\left\{\int_a^b u_t(s)x(s)ds u_t\right\}$, for all $x \in L_2[a,b]$, so that $C_u(x)(r) = \int_a^b k_u(s,r)x(s)ds$. Then, $C_X(x)(r) = \left(1-\theta^2\right)^{-1}\int_a^b k_u(s,r)x(s)ds = \int_a^b k_X(s,r)x(s)ds$ where $k_X(s,r) = \mathbb{E}\left[X_t(s)X_t(r)\right] = \left(1-\theta^2\right)^{-1} k_u(s,r)$ and the autocovariance operator $C_{X_t,X_{t+h}} = \theta^{|h|}C_X$. The martingale difference condition in A1(i) allows for conditional heterogeneity over time in the functional error process u_t and heterogeneity across section by way of the covariance kernel variation in $k_u(r,r)$ over r, but does not permit unconditional temporal heterogeneity over t. The càdlàg condition in A1(i) ensures that $\{u_t, X_t\}$ have well-defined realized sample paths in $L_2[a,b]$.

Least squares estimation of the parameter θ is achieved by minimization of the squared L_2 distance summed over the sample observations, giving

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \int_a^b \left[X_t(r) - \theta X_{t-1}(r)\right]^2 dr = \frac{\sum_{t=1}^{n}\int_a^b X_t(r)X_{t-1}(r)dr}{\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)dr}. \tag{6}$$

Simple least squares in (6) may be replaced by weighted least squares given by

$$\widehat{\theta}_w = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \int_a^b \left[X_t(r) - \theta X_{t-1}(r)\right]^2 w(r)dr = \frac{\sum_{t=1}^{n}\int_a^b X_t(r)X_{t-1}(r)w(r)dr}{\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)w(r)dr}, \tag{7}$$

for some positive, bounded, non-random weight function $w(r) \in L_2[a,b]$. For instance, in applications such as Wasserstein autoregression (Zhang et al., 2022), it may be useful to employ a density over $L_2[a,b]$ as the weight function or in other cases use weights associated with cross section heterogeneity in $k(r,r)$ over $r \in [a,b]$, as in (24) below. In such cases, regression takes place in the weighted Hilbert space $L_2[a,b;w(\cdot)]$, leading to (7).

Since the pair (X_t, u_t) are stationary and ergodic and the autocovariance operator $C_{X_{t-1},u_t} = 0$ under A1(i) and A1(ii), we have

$$\widehat{\theta} - \theta = \frac{\frac{1}{n}\sum_{t=1}^{n}\int_a^b X_t(r)u_t(r)dr}{\frac{1}{n}\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)dr} \overset{a.s.}{\to} \frac{\int_a^b \mathbb{E}\left[X_{t-1}(r)u_t(r)\right]dr}{\int_a^b \mathbb{E}\left[X_{t-1}^2(r)\right]dr} = 0,$$

so that $\widehat{\theta}$, and similarly $\widehat{\theta}_w$, are consistent for θ. Limit distributions follow by the martingale central limit theorem (MGCLT).

Theorem 1. *Under Assumption A1 and for model (5),* $\widehat{\theta}, \widehat{\theta}_w \overset{a.s.}{\to} \theta$, $\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}\left(0, \omega_\theta^2\right)$ *and* $\sqrt{n}(\widehat{\theta}_w - \theta) \rightsquigarrow \mathcal{N}\left(0, \omega_{\theta_w}^2\right)$, *where* $\omega_\theta^2 = \left(1 - \theta^2\right)\rho_u^2$, $\omega_{\theta,w}^2 = \left(1 - \theta^2\right)\rho_{u,w}^2$ *and*

$$\rho_u^2 = \frac{\int_a^b \int_a^b k_u(r,s)^2 ds dr}{\left(\int_a^b k_u(r,r) dr\right)^2}; \quad \rho_{u,w}^2 = \frac{\int_a^b \int_a^b k_u(r,s)^2 w(r) w(s) ds dr}{\left(\int_a^b k_u(r,r) w(r) dr\right)^2}. \tag{8}$$

Remarks

1(a) In their study of Wasserstein autoregression Zhang et al. (2022, Theorem 3.2) gave the limit theory for the weighted autoregressive estimator $\hat{\theta}_w$ with explicit probability density weights in place of $w(r)$. As discussed in the following Remark 1(b) both estimators $\hat{\theta}$ and $\hat{\theta}_w$ are generally more efficient than simple least squares autoregression with scalar time series, for which the asymptotic variance has the familiar form $1 - \theta^2$. But, as explored further in Remark 1(e), the relationship between the limit theory of $\hat{\theta}$ and $\hat{\theta}_w$ for weighted and unweighted autoregression is more subtle and, in fact, neither is universally asymptotically more efficient.

1(b) By Cauchy-Schwarz, $k_u(r,s)^2 = \left(\mathbb{E}\left[u_t(r)u_t(s)\right]\right)^2 \leq \mathbb{E}\left[u_t(r)^2\right]\mathbb{E}\left[u_t(s)^2\right] = k_u(r,r)k_u(s,s)$, so that

$$\int_a^b \int_a^b k_u(r,s)^2 ds dr \leq \int_a^b \int_a^b k_u(r,r) k_u(s,s) ds dr = \left(\int_a^b k_u(r,r) dr\right)^2, \tag{9}$$

ensuring the squared correlation functional $\rho_u^2 \leq 1$, and in a similar manner $\rho_{u,w}^2 \leq 1$. It follows that the asymptotic variances of both $\sqrt{n}(\widehat{\theta} - \theta)$ and $\sqrt{n}(\widehat{\theta}_w - \theta)$ are bounded above by $1 - \theta^2$, which is the variance of the limit distribution of the least squares or Gaussian maximum likelihood estimator of θ in the simple scalar AR(1) model. Thus, functional least squares estimation of θ in (5) is asymptotically efficient relative to simple least squares estimation in a scalar AR(1) (or equivalently, in (5) with a single fixed ordinate r) whenever strict inequality holds in (9), i.e., whenever $u_t(r) \neq \lambda u_t(s)$ for some (r,s) on a set of positive Lebesgue measure in $\mathbb{R}^2$. In other words, efficiency gains occur in functional least squares regression on (5) provided the functional correlation coefficient $\rho_u(r,s) = k_u(r,s)/\left\{k_u(r,r)k_u(s,s)\right\}^{1/2}$ between $u_t(r)$ and $u_t(s)$ satisfies $\left|\rho_u(r,s)\right| < 1$ on a set of positive measure in $\mathbb{R}^2$. Observe that when $\rho_u(r,s) = 1, a.e.$ in $\mathbb{R}^2$, the model reduces to the scalar AR(1) case.

1(c) The criterion $\left|\rho_u(r,s)\right| < 1$ implies that there are regions of positive Lebesgue measure in $[a,b]^2$ where there is less than perfect dependence between the random elements $u_t(r)$ and $u_t(s)$ at ordinates (r,s). This is a weak condition that can be expected to apply in almost all cases in practice. Thus, we may expect function space data to raise efficiency in parametric least squares regression even when there is considerable cross section dependence in the error covariance kernel $k_u(r,s)$. The Brownian motion and Brownian bridge examples (i) and (ii) below and those in Section 8.2 of Appendix B illustrate some of the efficiency gains that are attainable in specific cases.

1(d) There is an important connection between Theorem 1 and corresponding asymptotic theory for the dynamic panel model

$$X_{it} = \theta X_{it-1} + u_{it}, \ \ i = 1, ..., N; t = 1, ..., n; \ u_{it} \sim_{iid} \left(0, \sigma^2\right) \text{ over } i \text{ and } t, \tag{10}$$

that is, model (2) with no fixed effects. It is well known that in this model the least squares estimate $\widehat{\theta}$ is consistent and has limit theory

$$\sqrt{nN}\left(\widehat{\theta} - \theta\right) \underset{(N,n)\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, 1-\theta^2\right). \tag{11}$$

However, this result holds under independence over the cross section $i = 1, ..., N$, which raises the convergence rate from $\sqrt{n}$ to $\sqrt{nN}$. When there is perfect dependence over i and $u_{it} = u_t$ $a.s.$ for all i, then the panel model is equivalent to a scalar AR(1) and the limit theory is $\sqrt{n}\left(\widehat{\theta} - \theta\right) \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, 1-\theta^2\right)$. Theorem 1 shows that there is an intermediate class of model in which the asymptotic variance is reduced through cross section averaging but where the rate of convergence remains $\sqrt{n}$ because of cross section error dependence in the covariance kernel $k_u(r, s)$. Note that if $k_u(r, s) = 0$ for all $r \neq s$ then $\int_a^b \int_a^b k_u(r, s)^2 ds dr = 0$ and the limit variance degenerates in (8). This degeneracy aligns with the higher rate of convergence that applies in the panel data limit theory (11) when the errors are *iid* across section as well as over time as in (10).

1(e) Theorem 1 can be used to consider cases where cross section dependence in the error random elements $u_t(r)$ goes to zero outside small neighborhoods, i.e., local cross section dependence. If the neighborhoods shrink in size within the interval domain $[a, b]$, then this specification may be regarded as analogous to a form of weak cross section dependence in the discrete space of a dynamic panel model or sparsity in large covariance matrix estimation. To fix ideas, consider a triangular array model defined by

$$X_{t,N}(r) = \theta X_{t-1,N}(r) + u_{t,N}(r), \quad r \in [a, b],$$

where $u_{t,N}(r)$ has covariance kernel

$$k_{u,N}(r, s) = \mathbb{E}\left[u_{t,N}(r)u_{t,N}(s)\right] = \sigma^2 N \ \mathbf{1}\{|r-s| \leq \frac{1}{2N}\}. \tag{12}$$

The formulation (12) involves a (hard) threshold that sets covariances to zero beyond the distance $1/2N$ that tends to zero as $N \to \infty$, thereby behaving in the limit like a sequence of functions of $|r - s|$ that tend to a Dirac delta function. Other formulations might use: (i) a different scheme of downweighting covariances with soft rather than hard thresholding, such as $k_{u,N}(r, s) = \sigma^2 e^{-a|r-s|N}$ for some fixed $a > 0$; or (ii) an extended versions of (12) such as $k_{u,N}(r,s) = \sigma^2 \rho(r,s) N \; \mathbf{1}\{|r-s| \leq \frac{1}{2N}\}$ where $\rho(r,s)$ is the covariance kernel of a specific error process that may be relevant in applications, such as a Brownian motion with $\rho(r,s) = r \wedge s$ or Brownian bridge with $\rho(r,s) = r \wedge s - rs$. Figure 2 plots for $N = 3$ the covariance kernel (12) and the corresponding kernels with thresholded Brownian motion and Brownian bridge. For such covariance kernels, the random function $u_{t,N}(r)$ is weakly dependent across ordinates, having non-zero covariance only within the (shrinking as $N \to \infty$) narrow band $|r-s| \leq \frac{1}{2N}$ in which the error variance is correspondingly rescaled by N. The error variance rescaling ensures that the total cross section dependence $\int_a^b k_{u,N}(r,s)ds = \sigma^2$ is constant for all $r \in \left[a + \frac{1}{2N}, b - \frac{1}{2N}\right]$ in (12). This function valued specification emulates a panel model such as (10) in which the error process u_{it} is nearly *iid* over i, or more precisely u_{it} is correlated only with respect to a finite number of neighbors as the cross section sample size $N \to \infty$. Calculations in Appendix B in Section 8.1 show that

$$\sqrt{n}(\widehat{\theta} - \theta) \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, \left(1-\theta^2\right) \frac{\int_a^b \int_a^b k_{u,N}(r,s)^2 ds dr}{\left(\int_a^b k_{u,N}(r,r)dr\right)^2}\right) = \mathcal{N}\left(0, \frac{1}{N}\frac{1-\theta^2}{b-a}\right),$$

and then

$$\sqrt{nN}(\widehat{\theta} - \theta) \underset{(N,n)_{\text{seq}}\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, \frac{1-\theta^2}{b-a}\right), \tag{13}$$

where $(N,n)_{\text{seq}} \to \infty$ signifies sequential divergence with $n \to \infty$ followed by $N \to \infty$. When $b - a = 1$, the limit variance in (13) is $1 - \theta^2$ and the result matches the dynamic panel case (11) exactly. It is useful to compare the limit result (13) with the case where $k_u(r,s) = 1$ uniformly for all (r,s), in which case it follows directly from (8) that

$$\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}\left(0, \left(1-\theta^2\right) \frac{(b-a)^2}{(b-a)^2}\right) = \mathcal{N}\left(0, 1-\theta^2\right). \tag{14}$$

This limit theory corresponds to the earlier upper bound case discussed in Remark (a) because $k_u(r,s) = 1$ uniformly for all (r,s) implies that the functional correlation coefficient $\rho_u(r,s) = 1$. The differences between (13) and (14) are clear. In the near cross section independence case (13), we get accelerated convergence and a limit variance $\frac{1-\theta^2}{b-a}$ that depends on the size of domain $b-a$. Thus, the near independence of the data across section (in the continuous parameter r) delivers more information about θ and leads to the faster convergence rate $\sqrt{nN}$. On the other hand, when there is cross section dependence over the full domain as in (14), there is no acceleration in the $\sqrt{n}$ convergence rate and the limit variance $1-\theta^2$ is retained, corresponding to the scalar AR(1) to which the model reduces in that case.

1(f) Some further heuristics may be considered as the width $R = b - a$ of the cross section domain $[a,b]$ changes. When $R \to 0$, then (14) still holds because the model simply

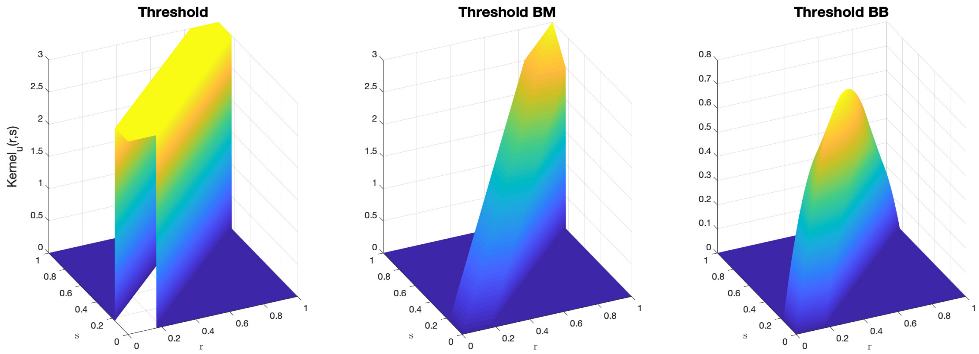


Figure 2: Examples of covariance kernels $k_{u,N}(r,s) = \sigma^2 \rho(r,s) N \ \mathbf{1}\{|r-s| \leq \frac{1}{2N}\}$ with hard thresholding set with $N = 3$ and where $\rho(r,s)$ is the covariance kernel of a constant $\rho(r,s) = 1$, standard Brownian motion $\rho(r,s) = r \wedge s$, and standard Brownian bridge $\rho(r,s) = r \wedge s - rs$.

corresponds to the standard AR(1) case where $a = b = r$ and $\sqrt{n}$ convergence is retained. But in case (13) the limit variance $\left(1-\theta^2\right)/R \to \infty$ as $R \to 0$, and the limit theory fails because there is no extra information in the data as $N \to \infty$ from independence in the errors over $[a,b]$ because the observation interval of the functional data shrinks to a single point. So in this case the scaling $\sqrt{nN}$ is excessive and leads to divergence. On the other hand, if $R \to \infty$, $\sqrt{nN}(\widehat{\theta} - \theta) \to_p 0$ because we have asymptotically independent errors over an infinite region in $L_2(\mathbb{R})$ space. Then, $\sqrt{nNR}(\widehat{\theta}-\theta) \rightsquigarrow_{(R,N,n)_{seq}\to\infty} \mathcal{N}\left(0, 1-\theta^2\right)$, with the faster convergence rate $\sqrt{nNR}$ that accounts for the accumulated information over the widening observation interval as $R \to \infty$. Note that there is no gain in (14) in that case because the errors $u_t(r)$ are still perfectly correlated over the wider cross section domain $\mathbb{R}$ as $R \to \infty$. These findings show that the function space model with wide-domain and narrow-domain cross section dependence sheds new light on the effects of various forms of cross section dependence in discrete dynamic panel models.

Some specific examples of kernel functions help in exploring the potential efficiency gains associated with the use of functional data. They also turn out to be useful in developing shrinkage estimators of the covariance kernel $k_u(r,s)$ that are needed for inference. Remarks 1(a)-1(e) above already show that the nature of the cross section dependence in $u_t(r)$ can play a big role in the limit theory. To illustrate further we consider the following simple cases involving spatial Gaussian stochastic process errors. Additional examples are discussed in Appendix B in Section 8.2. For convenience of exposition let the observation interval $[a,b] = [0,1]$ and consider the asymptotic variance of the functional least squares estimator $\widehat{\theta}$.

(i) **Brownian motion errors** Define the function valued error process

$$u_t(r) = \sigma\left(W_{t+r} - W_t\right), \ r \in [0,1], \ t = 1, 2, \cdots, n, \tag{15}$$

where W is standard Brownian motion, so that $\{u_t\}_{t=1}^{\infty}$ is a sequence of independent

Brownian motions each living in $C[0,1] \subset L_2[0,1]$. In this case, $k(r,s) = r \wedge s$ and the variance of the limit distribution of $\widehat{\theta}$ reduces to

$$\omega_\theta^2 = \left(1-\theta^2\right) \frac{\int_0^1 \int_0^1 k_\varepsilon(r,s)^2 dsdr}{\left(\int_0^1 k(r,r)dr\right)^2} = (1-\theta^2)\frac{2\int_0^1 \int_0^s r^2 drds}{\left(\int_0^1 rdr\right)^2} = \frac{2}{3}(1-\theta^2). \tag{16}$$

Thus, with Brownian motion errors as in (15), the use of functional least squares regression reduces asymptotic variance in scalar autoregression by a third.

(ii) **Brownian bridge errors** Define $u_t(r) = \sigma\left\{(W_{t+r} - W_t) - r\left(W_{t+1} - W_t\right)\right\}$, so that $k(r,s) = r \wedge s - rs$. The asymptotic variance of $\widehat{\theta}$ is now

$$\omega_\theta^2 = \left(1-\theta^2\right) \frac{\int_0^1 \int_0^1 k_\varepsilon(r,s)^2 dsdr}{\left(\int_0^1 k(r,r)dr\right)^2} = (1-\theta^2)\frac{\int_0^1 \int_0^1 \left\{(r \wedge s) - rs\right\}^2 drds}{\left(\int_0^1 r(1-r)dr\right)^2} = \frac{2}{5}(1-\theta^2). \tag{17}$$

It follows that under a similar configuration as Example (i) the limit variance (17) of $\widehat{\theta}$ is smaller under Brownian bridge innovations than under Brownian motion innovations. The reason is that the Brownian bridge is tied down at the ends of the interval $[0,1]$, thereby reducing the variation in the function space equation error, which correspondingly reduces the asymptotic variance of $\widehat{\theta}$ compared with a Brownian motion error process.

(iii) **Linear diffusion errors** Define

$$u_t(r) = \sigma J_{c,t}(r) := \sigma \int_0^r e^{c(r-p)} dW_t(p), \quad r \in [0,1], \tag{18}$$

where $\{W_t(p)\}_{t=1}^n$ is a sequence of independent standard Brownian motions with $p \in [0,1]$, thereby giving a sequence over time of independent linear diffusion processes $\{J_{c,t}(r)\}_{t=1}^n$, with diffusion coefficient $c \in \mathbb{R}$, that model the cross section curved error process $u_t(r)$. On an equispaced grid $\{r_i\}_{i=0}^m$ with $\Delta_r = r_i - r_{i-1}$, this process has the exact discrete representation

$$J_{c,t}(r_i) = e^{c\Delta_r} J_{c,t}(r_{i-1}) + V_{t,r_i}, \quad V_{t,r_i} = \int_{r_{i-1}}^{r_i} e^{c(r_i - p)} dW_t(p). \tag{19}$$

For fitted residual curves $\{\hat{u}_t\}$, the diffusion coefficient can be estimated from this exact discrete representation by

$$\hat{c} = \underset{c \in (-\infty,\infty)}{\operatorname{argmin}} \sum_{t=1}^n \sum_{i=1}^m \left[\hat{u}_t(r_i) - e^{c\Delta_r} \hat{u}_t(r_{i-1})\right]^2. \tag{20}$$

The cross section error covariance kernel of $u_t(r) = J_{c,t}(r)$ is then $k_u(r,s) = \sigma^2 k(r,s)$ where $k(r,s) = \frac{e^{|r-s|c} - e^{(r+s)c}}{-2c} = \frac{e^{(r+s)c}[1-e^{2c(r \wedge s)}]}{2c}$ and $k(r,r) = \frac{1-e^{2rc}}{-2c}$. The asymptotic variance of $\widehat{\theta}$ is then dependent on the value of the diffusion coefficient c and, after some calculations, we find

$$\omega_\theta(c)^2 = (1-\theta^2)\frac{\int_0^1\int_0^1 \left(e^{|r-s|c} - e^{(r+s)c}drds\right)^2}{\left(\int_0^1 \left(1-e^{2rc}\right)dr\right)^2} = (1-\theta^2)\frac{\frac{e^{4c}-1}{4c^2} - \frac{2}{c}e^{2c} + \frac{e^{2c}-1}{c^2} - \frac{1}{c}}{\left(\frac{e^{2c}-1}{2c}-1\right)^2}$$
$$= (1-\theta^2)\frac{e^{4c} - 8ce^{2c} + 4e^{2c} - 4c - 5}{\left(e^{2c}-2c-1\right)^2}. \tag{21}$$

The implications of this asymptotic variance formula differ for negative and positive values of the diffusion coefficient c as we now discuss.

(a) When $c = -1$ we have $\omega_\theta^2(-1) \approx 0.4983 \times \left(1-\theta^2\right)$, reducing the asymptotic variance of $\hat{\theta}$ in the Brownian motion case of $\frac{2}{3}\left(1-\theta^2\right)$ and by around a half compared with the scalar AR(1) model; and for $c = -5$ we have $\omega_\theta^2(-5) \approx 0.1852 \times \left(1-\theta^2\right)$. So the asymptotic variance of $\widehat{\theta}$ reduces rapidly as c becomes increasingly negative, in which case cross section dependence correspondingly reduces. In fact, as $c \to -\infty$, the correlation function $R_\varepsilon(r,s) = \frac{k(r,s)}{\{k(r,r)k(s,s)\}^{1/2}} \sim_a \frac{1}{-c}$ and $\lim_{c\to-\infty}\omega_\theta(c)^2 = 0$. Hence, $k(r,s), R_\varepsilon(r,s) \to 0$ as $c \to -\infty$ for all $r \neq s$ and so cross section dependence goes to zero, effectively leading to independence. The limiting stochastic process is then pure noise across section in continuous r, which is a generalized random process. The intuition for the zero limiting variance of $\omega_\theta(c)^2$ is that, with pure noise innovations $u_t(r)$ across the domain of $r \in [0,1]$, $\hat{\theta}$ converges faster than the $\sqrt{n}$ rate of Theorem 1. In fact, it is easy to see that, upon sequential convergence of $n \to \infty$ followed by $c \to -\infty$ denoted $(-c,n)_{seq}$, we have[1]

$$\sqrt{n(-c)}\left(\widehat{\theta}-\theta\right) \underset{(-c,n)_{seq}\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, 1-\theta^2\right), \tag{22}$$

which is analogous to the limit theory (11) in dynamic panel regression with (assumed) N *iid* cross section errors, where the convergence rate rises to $\sqrt{nN}$. In both cases there is more information in the data arising from either effective or assumed cross section independence that raises the convergence rate.

(b) When $c = 1$ we have $\omega_\theta^2(1) \approx 0.83275 \times \left(1-\theta^2\right)$, which raises the asymptotic variance compared with the Brownian motion cross section error case; and for $c = 5$ we have $\omega_\theta^2(5) \approx 0.99936 \times \left(1-\theta^2\right)$, so the asymptotic variance rises towards the level $\left(1-\theta^2\right)$ in the scalar AR(1). As $c \to \infty$ the asymptotic variance rapidly rises to $\left(1-\theta^2\right)$. The intuition is that as $c \to \infty$ cross section dependence continues to rise until there is no further information in the cross section data than there is in a

[1] Under certain relative rate conditions on c and n of the type employed in Phillips and Moon (1999) the limit given in (22) should remain valid under joint convergence but this is not pursued here.

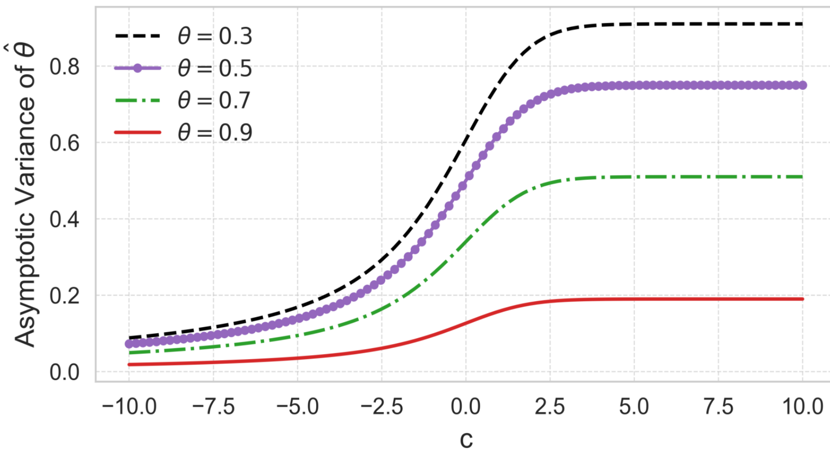


Figure 3: Examples of the asymptotic variance $\omega_\theta(c)^2$ in (21) for various values of the diffusion coefficient c.

scalar AR(1). In fact as evident from (21), we have $\omega_\theta^2(c) \underset{c\to\infty}{\to} \left(1-\theta^2\right)$ and then

$$\sqrt{n}\left(\widehat{\theta}-\theta\right) \underset{(c,n)_{seq}\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, 1-\theta^2\right), \tag{23}$$

just as in scalar AR(1) limit theory with no cross section curve observations.

The ARH model with diffusion process cross section innovations $u_t(r) = \sigma J_{c,t}(r)$ therefore covers a wide class of intermediate cases of dependent observations across section while also allowing for the limiting extreme cases of independence (with $c \to -\infty$) and perfect dependence ($c \to \infty$). As discussed below, this coverage of spatial dependence makes the ARH-diffusion model potentially useful in estimating the covariance kernel $k_u(r, s)$, which is needed for inference. The asymptotic variance $\omega_\theta(c)^2$ in (21) is shown for various values of the diffusion coefficient c in Figure 3. The steady decline in asymptotic variance as c becomes increasingly negative, reflecting greater indepencence, is apparent as well as the stabilization of the asymptotic variance as c becomes increasingly positive, reflecting greater dependence. The progressive reduction in variance as the autoregressive coefficient θ rises towards unity is also clear in the plots.

In view of the cross section dependence in the model (5) weighted least squares alternatives to $\hat{\theta}$ might be considered, taking account of possible knowledge (or estimates) of the covariance kernel $k_u(r, s)$. The simplest of these would be the (infeasible) weighted least squares estimator

$$\widehat{\theta}_w = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \int_a^b \left(\frac{X_t(r) - \theta X_{t-1}(r)}{k_u(r,r)^{1/2}}\right)^2 dr = \frac{\sum_{t=1}^n \int_a^b X_t(r) X_{t-1}(r)/k_u(r,r)dr}{\sum_{t=1}^n \int_a^b X_{t-1}^2(r)/k_u(r,r)dr}, \tag{24}$$

with weight function $w(r) = 1/k(r, r)$, accounting for heterogeneity in the cross section variance

$k_u(r,r)$ over $r \in [a,b]$. The estimator $\widehat{\theta}_w$ has the following asymptotic theory in Theorem 1

$$\sqrt{n}(\widehat{\theta}_w - \theta) \rightsquigarrow \mathcal{N}\left(0, \omega^2_{\widehat{\theta}_w}\right), \tag{25}$$

with $\omega^2_{\widehat{\theta}_w} = \frac{1-\theta^2}{(b-a)^2} \int_a^b \int_a^b \frac{k(r,s)^2}{k(r,r)k(s,s)} dsdr$. Since $k(r,s)^2 \leq k(r,r)k(s,s)$ by Cauchy-Schwarz, we deduce that $\omega^2_{\widehat{\theta}_w} \leq 1-\theta^2$, so that $\widehat{\theta}_w$, just like $\widehat{\theta}$, is asymptotically more efficient than the least squares estimator of θ in the scalar AR(1) model. A feasible version of $\widehat{\theta}_w$ requires estimation of the covariance kernel diagonal elements $k_u(r,r)$, which can be achieved using sample variances $\hat{k}_u(r,r) = n^{-1}\sum_{t=1}^n \hat{u}_t(r)^2$ of the least squares residuals $\hat{u}_t(r) = X_t(r) - \hat{\theta}X_{t-1}(r)$, which are consistent for $u_t(r)$. The explicit relationship between the variances of the two estimators in the examples above is easily calculated. For instance, in Example 1 with $k_u(r,s) = \sigma^2 r \wedge s$ and the interval $[a,b] = [0,1]$, we have $\omega^2_{\widehat{\theta}} = \frac{2}{3}(1-\theta^2)$, from (16), whereas

$$\omega^2_{\widehat{\theta}_w} = (1-\theta^2)\int_0^1\int_0^1 \frac{k(r,s)^2}{k(r,r)k(s,s)} dsdr = 2(1-\theta^2)\int_0^1\int_0^r \frac{s^2}{rs} dsdr = \frac{1}{2}(1-\theta^2),$$

so that weighted OLS is more efficient asymptotically than the OLS regression (reducing asymptotic variance by by $\frac{1}{6}(1-\theta^2)$), as might be expected by taking account of the cross section heterogeneity in $k_u(r,r)$. Interestingly, in the general framework with cross section covariance kernel $k(r,s)$, no universal ordering of the asymptotic variances of OLS and WOLS holds. In terms of the squared functional correlation coefficient $\rho_u^2(r,s) - k(r,s)^2/(k(r,r)k(s,s))$ and the aggregate cross section variance $M = \int_0^1 k(r,r)dr$, the asymptotic variance of OLS exceeds that of WLS (i.e., $\omega^2_{\widehat{\theta}_{ols}} > \omega^2_{\widehat{\theta}_w}$) iff $\int_0^1\int_0^1 \rho_u^2(r,s)[k(r,r)k(s,s) - M^2]drds > 0$. This difference depends on the cross section weighted difference between the variance product $k(r,r)k(s,s)$ and the squared mean variance M^2 over the domain $[0,1]^2$. As shown above, both BM and BB kernels give efficiency gains for WLS with $\omega^2_{\widehat{\theta}_{ols}} > \omega^2_{\widehat{\theta}_w}$. An example where the reverse is true is the algebraic kernel $k_A(r,s) = (1 + 4(r-\frac{1}{2})(s-\frac{1}{2}))$, which is continuous and positive definite over $[0,1]^2$. In this case, computations show that $\rho_u^2 = 0.625 < 0.6629 = \rho^2_{u,\omega}$, so that $\omega^2_{\widehat{\theta}_{ols}} < \omega^2_{\widehat{\theta}_w}$ and weighted regression in the presence of cross section dependence can actually reduce efficiency in some cases.

In addition to weighted least squares leading to $\widehat{\theta}_w$, generalized least squares involving the full covariance kernel $k_u(r,s)$ might be considered as a means of achieving asymptotically optimal estimation. While this is typically straightforward in finite dimensional cases involving the inversion of a finite dimensional variance matrix and use of a suitably consistent estimator, in the infinite dimensional Hilbert space case inversion of the operator corresponding the covariance kernel generally leads to inverse problems because, even in simple cases such as Examples 1 and 2 involving Brownian motion or Brownian bridge error processes $u_t(r)$, the covariance kernel does not satisfy Picard's criterion which provides necessary and sufficient conditions for a solution involving the inverse operator to exist – see Carrasco and Florens (2000, Lemma 5 and Example 1). In such cases, the inverse operator is actually a differential operator, which is not a compact operator, and for practical implementation either projection on a finite dimensional space (Bosq, 2000, chapter 8) or regularization is typically required, as in nonparametric instrumental variable regression (Hall and Horowitz, 2005). This approach to estimation and inference is of interest but is not pursued in the present work.

2.3 Inference

For inference and confidence interval construction about θ we need to estimate the limit variance $\omega_\theta^2 = \left(1-\theta^2\right)\rho_u^2$, which involves the functional squared correlation coefficient

$$\rho_u^2 = \frac{\int_a^b \int_a^b k_u(r,s)^2 dsdr}{\left(\int_a^b k_u(r,r)dr\right)^2} = \frac{\int_a^b \int_a^b k(r,s)^2 dsdr}{\left(\int_a^b k(r,r)dr\right)^2} = \rho_\varepsilon^2, \tag{26}$$

and is independent of σ^2. Using the residual function $\hat{u}_t(r) = X_t(r) - \widehat{\theta}X_{t-1}(r)$, we may construct a t ratio statistic for θ of the form $\widetilde{t}_\theta = \frac{\widehat{\theta}-\theta}{\widetilde{s}_\theta}$ where $\widetilde{s}_\theta^2$ is the continuous sandwich form variance estimate

$$\widetilde{s}_\theta^2 := \left(\sum_{t=1}^n \int_a^b X_{t-1}^2(r)dr\right)^{-2} \left(\sum_{t=1}^n \int_a^b \int_a^b X_{t-1}(r)\hat{u}_t(r)\hat{u}_t(s)X_{t-1}(s)drds\right). \tag{27}$$

It is shown in the proof of Theorem 2 below that upon standardization $n\widetilde{s}_\theta^2 \to_{a.s} \omega_\theta^2$, giving a consistent estimator of the asymptotic variance of $\hat{\theta}$.

Alternatively, we can estimate the covariance kernel of $u_t(r)$ directly using the residuals $\hat{u}_t(r)$ to construct a sample covariance kernel estimate such as $\hat{k}_u(r,s) = n^{-1}\sum_{t=1}^n \hat{u}_t(r)\hat{u}_t(s)$. As shown in the proof of Theorem 2, the estimate $\hat{k}_u(r,s)$ is consistent as $n \to \infty$. But achieving satisfactory performance in finite sample estimation of $k_u(r,s)$ is naturally challenging. It is analogous to that of large dimensional covariance matrix estimation but possibly even more complex because sparsity may not relevant in many applied contexts with curved cross sectional data of the type considered in our empirical work. A new method of improving finite sample performance will be discussed later in this section. Using the unrestricted estimate $\hat{k}_u(r,s)$ directly, the corresponding t ratio statistic $\hat{t}_\theta = \frac{\widehat{\theta}-\theta}{\hat{s}_\theta}$ is formed by using a standard error estimate $\hat{s}_\theta$ obtained from the variance estimate

$$\hat{s}_\theta^2 = \frac{1-\widehat{\theta}^2}{n}\int_a^b \int_a^b \hat{k}_u(r,s)^2 dsdr / \left(\int_a^b \hat{k}_u(r,r)dr\right)^2. \tag{28}$$

In view of the first factor of (28) and the support of $\widehat{\theta}$, the estimate $\hat{s}_\theta^2$ may be non-positive. To avoid this practical difficulty we can use the following simple estimate

$$\widetilde{s}_\theta^{*2} = \left(\sum_{t=1}^n \int_a^b X_{t-1}^2(r)dr\right)^{-1} \frac{\int_a^b \int_a^b \hat{k}_u(r,s)^2 dsdr}{\left(\int_a^b \hat{k}_u(r,r)dr\right)}, \tag{29}$$

giving the corresponding t ratio $\widetilde{t}_\theta^* = \frac{\widehat{\theta}-\theta}{\widetilde{s}_\theta^*}$. These t ratios are asymptotically standard normal and may be used for testing and confidence interval construction with functional data $\{X_t(r)\}_{t=0}^n$.

Theorem 2. *Under Assumption A1, $\hat{t}_\theta$, $\widetilde{t}_\theta$, $\widetilde{t}_\theta^* \rightsquigarrow \mathcal{N}(0,1)$ as $n \to \infty$.*

In parametric ARH regression, just as in dynamic panel models, inference must account for cross section dependence in the equation errors. The asymptotic variance formulae in (27)–(29) do this through residual-based estimates of the covariance kernel $k_u(r,s)$ or through the continuous sandwich form. In finite samples, however, direct covariance-kernel estimation can be noisy because the curve dimension is large. This is the same practical difficulty that arises in high dimensional covariance matrix estimation, where sampling variation in the unrestricted covariance estimate can affect the accuracy of standard errors and tests.

To reduce reliance on the normal critical values in such samples, we use a residual wild bootstrap to obtain critical values and confidence intervals for the three statistics $\hat{t}_\theta$, $\widetilde{t}_\theta$, and $\widetilde{t}_\theta^*$. The bootstrap uses the empirical residual curves and therefore preserves the estimated cross section dependence across r, while avoiding a parametric specification of $k_u(r,s)$.

To test $\mathcal{H}_0 : \theta = \theta_0$, first estimate the ARH coefficient and compute the recentered residual curves $\hat{u}_t(r) = X_t(r) - \widehat{\theta} X_{t-1}(r)$. Let $\xi_t{}_{t=1}^n$ be i.i.d. standard normal multipliers and define the wild-bootstrap innovations by $\hat{u}_t^*(r) = \xi_t \hat{u}_t(r)$. Since a single multiplier is applied to the entire residual curve at each date t, this construction preserves the estimated cross-sectional dependence across r.

Bootstrap samples are generated under the null using $X_t^*(r) = \theta_0 X_{t-1}^*(r) + \hat{u}_t^*(r)$, with the initial curve fixed at its observed value. For each bootstrap sample, the coefficient is re-estimated and the statistic $T_j^*(\theta_0)$, $j \in \hat{t}, \widetilde{t}, \widetilde{t}^*$, is recomputed using the same variance formula as in the original sample. The two-sided level-α critical value is the conditional $(1-\alpha)$ quantile of $|T_j^*(\theta_0)|$ given the data, and $\mathcal{H}_0$ is rejected when $|T_j(\theta_0)|$ exceeds this value. Confidence intervals are obtained from the corresponding recentered bootstrap-t distribution. Section 4 reports simulations on the finite-sample performance of the wild-bootstrap tests based on $\hat{t}_\theta$, $\widetilde{t}_\theta$, and $\widetilde{t}_\theta^*$.

2.4 Bias expansion in the stationary parametric ARH(1)

It is well known that least squares autoregression produces downward biased autoregressive coefficient estimators in finite samples. In the literature, bias approximations and expansions have been derived for various autoregressive models and parameter domains covering both stationary and nonstationary cases (White, 1961; Shenton and Johnson, 1965; Phillips, 1977; Vinod and Shenton, 1996; Phillips, 2012). The following result is obtained from the Edgeworth expansion of the distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ given in (Phillips, 2025, Theorem 1) in the ARH(1) model (5) when $|\theta| < 1$.

Theorem 3. *Under Assumption A1 and an additional zero third moment condition $k_u(r,s,p) :=$*

$\mathbb{E}u_t(r)u_t(s)u_t(p) = 0$ *for all* $\{r, s, p \in [a, b]\}$, $\widehat{\theta}$ *has the following bias expansion*

$$\mathbb{E}\left(\widehat{\theta}\right) - \theta = -\frac{2\theta}{n}\rho_u^2 + o\left(\frac{1}{n}\right). \tag{30}$$

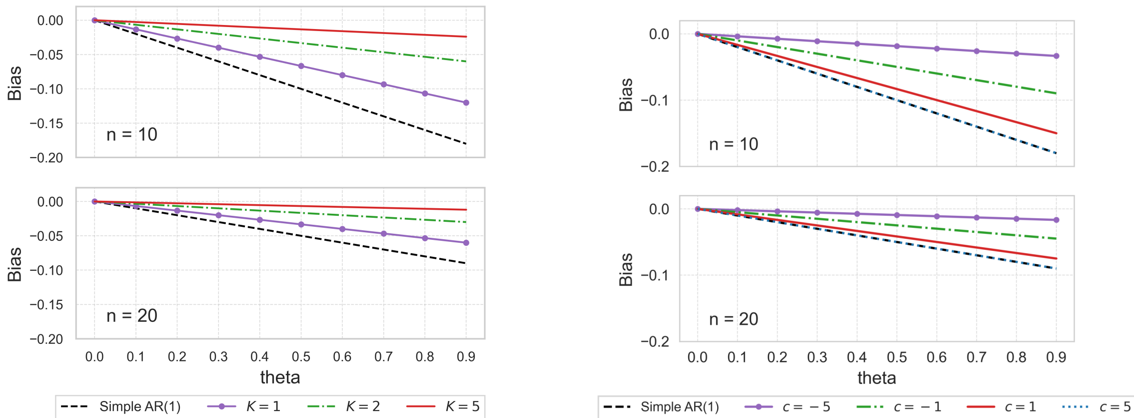
Remarks

2(a) The bias expansion (30) shows that bias is negative and linear in θ to order $O\left(n^{-1}\right)$, but with a slope coefficient that differs from the simple scalar AR(1) case. By Cauchy-Schwarz $\rho_u^2 \leq 1$ from which it follows that the magnitude of the ARH(1) bias is bounded above by that of the simple scalar AR(1) model in the stationary case, i.e., $\left|\frac{2\theta}{n}\rho_u^2\right| \leq \left|\frac{2\theta}{n}\right|$. Hence, the ARH(1) bias is less than the AR(1) bias except when $|\rho_u(r, s)| = 1$, a.e in $[a, b]^2$ where $\rho_u(r, s) = k_u(r, s)/\left\{k_u(r, r)k_u(s, s)\right\}^{1/2}$. Thus, whenever there is less than perfect dependence across ordinates (r, s) on a set of positive measure in $\mathbb{R}^2$, there will be gain for bias reduction in functional least squares regression on (5).

2(b) Perfect dependence over r of $u_t(r)$ implies that the model is effectively one dimensional. Then the bias is given by the familiar formula $-\frac{2\theta}{n}$, as in the simple scalar AR(1) case discussed in White (1961). On the other hand, if $k_u(r, s) = 0$ for $r \neq s$, then $\int_a^b \int_a^b k_u(r, s)^2 dsdr = 0$ and the asymptotics change, as discussed earlier in Remarks **1(c)** and **1(d)**, with the faster rate of convergence in $\widehat{\theta}$ induced by cross section averaging over independent data. The case where there are functional fixed effects in the model is considered later in Section 3 and in Remark **4(e)** when there is narrow band cross section dependence.

2(c) The zero third moment condition is added to the assumptions in Theorem 3 to accord with the conditions used in Phillips (2025), which were in turn used to align with the earlier work on such expansions that were commonly obtained under stricter Gaussian error conditions as for example in Phillips (1977). The zero third moment condition holds for Gaussian processes like Brownian motion and Brownian bridge errors u_t.

Figure 4 shows the asymptotic bias function expansion of $\mathbb{E}(\widehat{\theta}|\theta) - \theta$ of the OLS estimator $\widehat{\theta}$ in the curve AR(1) model for $n = 10, 20$ with: (a) segmented Brownian motion errors as in Section 8.2 and (84) with segment numbers $K = 1, 2, 5$; and (b) diffusion process errors as given in Example (iii) above and the exact discrete model (19). In Figure 4(a) the downward bias is greater when $K = 1$, i.e., when there is more dependence across ordinates. The bias almost disappears when $K = 5$ and $n = 20$. In Figure 4(b) the downward bias is substantially smaller when c is negative than when c is positive. Notably, when $c = 5$ the bias is the same as in the scalar AR(1) case, demonstrating the effect of considerable cross section dependence. So there are substantial gains in bias reduction in functional least squares regression when dependence across ordinates is relatively small. Further, just as in scalar autoregression, bias rises when the autoregressive coefficient approaches unity, and bias declines as n increases.



(a) Bias expansion: segmented BM kernel

(b) Bias expansion: diffusion process kernel

Figure 4: The bias function expansion of $\widehat{\theta}$, calculated by using the analytic formulae for $n = 10, 20$ with cross section covariance kernels: (a) segmented Brownian motion kernel with $K = 1, 2, 5$ segments; and diffusion kernel for various diffusion coefficients c.

2.5 Specification test for a cross section break

In empirical applications it may be useful to employ a test of the parametric specification. The simplest approach is to allow for a break in the AR coefficient θ to distinguish different segments in the dynamics by explicit dependence on the functional cross section argument r. This approach is shown below to provide power against more general functional alternatives. Consider, for example, the following dual regression specification[2] induced by cross section dependence

$$X_t(r) = \begin{cases} \theta_1 X_{t-1}(r) + u_t(r) & \text{if } a \le r < b \\ \theta_2 X_{t-1}(r) + u_t(r) & \text{if } b \le r \le c \end{cases}, \quad t = 1, \cdots, T \tag{31}$$

Here the cross section break point b is assumed known and assumption A1 holds with both $|\theta_1|, |\theta_2| < 1$. To test for such a break point a function space version of a standard Wald test is possible with null hypothesis $\mathcal{H}_0 : \theta_1 = \theta_2$. Least squares estimates $(\hat{\theta}_1, \hat{\theta}_2)$ over the two subintervals are obtained and the Wald statistic $\mathcal{W}_n = n\left(\widehat{\theta}_1 - \widehat{\theta}_2\right) \widehat{V}_{1,2}^{-1} \left(\widehat{\theta}_1 - \widehat{\theta}_2\right)$ is constructed, with estimated asymptotic variance

$$\widehat{V}_{1,2} = (1 - \widehat{\theta}^2) \left\{ \frac{\int_a^b \int_a^b \widehat{k}_u(r,s)^2 ds dr}{\left(\int_a^b \widehat{k}_u(r,r) dr\right)^2} + \frac{\int_b^c \int_b^c \widehat{k}_u(r,s)^2 ds dr}{\left(\int_b^c \widehat{k}_u(r,r) dr\right)^2} - 2 \frac{\int_a^b \int_b^c \widehat{k}_u(r,s)^2 ds dr}{\int_a^b \widehat{k}_u(r,r) dr \int_b^c \widehat{k}_u(r,r) dr} \right\}, \tag{32}$$

of $\sqrt{n}\left(\widehat{\theta}_1 - \widehat{\theta}_2\right)$ under the null. The variance estimate $\hat{V}_{1,2}$ in (32) relies on the global parameter estimate $\widehat{\theta}$ under the null and global estimates of the covariance kernel $k_u(r,s)$ of $u_t(r)$

[2]The specification (31) may be given a formal operator representation as $X_t(r) = \int_a^c \ell(s,r) X_{t-1}(s) ds + u_t(r)$, in which $\ell(s,r) = \theta_1 \delta(s-r) \mathbf{1}\{s \in [a,b)\} + \theta_2 \delta(s-r) \mathbf{1}\{s \in [b,c]\}$, where $\delta(\cdot)$ is the dirac delta function.

over the two domains. In constructing the latter, explicit kernels such as Brownian motion, Brownian bridge or diffusion process kernels may be used, or fully nonparametric consistent covariance kernel estimates based on regression residuals, as discussed earlier. To avoid potential negativity a sandwich form based on the null can be used in place of the factor $(1-\widehat{\theta}^2)$, as in (29).

Theorem 4. *Under model (31) with Assumption A1 and parameters $|\theta_1| < 1$ and $|\theta_2| < 1$, the following asymptotics hold as $n \to \infty$:*
(i) Under the null $\mathcal{H}_0 : \theta_2 = \theta_1$, $\mathcal{W}_n \rightsquigarrow \chi_1^2$, central chi-squared with a single degree of freedom;
(ii) Under the local alternative $\mathcal{H}_1 : \theta_2 = \theta_1 + \frac{\psi}{\sqrt{n}}$ for some constant ψ, $\mathcal{W}_n \rightsquigarrow \chi_1^2(\delta_\psi)$, noncentral chi-squared with noncentrality parameter $\delta_\psi = \psi^2 V_{1,2}^{-1}$, where $V_{1,2} = (1-\theta^2)V_{\theta_1,\theta_2}$ and

$$V_{\theta_1,\theta_2} = \left\{ \frac{\int_a^b \int_a^b k_u(r,s)^2 dsdr}{\left(\int_a^b k_u(r,r)dr\right)^2} + \frac{\int_b^c \int_b^c k_u(r,s)^2 dsdr}{\left(\int_b^c k_u(r,r)dr\right)^2} - 2\frac{\int_a^b \int_b^c k_u(r,s)^2 dsdr}{\int_a^b k_u(r,r)dr \int_b^c k_u(r,r)dr} \right\}; \tag{33}$$

(iii) Under the local functional alternative $\mathcal{H}_2 : \theta_2 = \theta_1 + \frac{\psi(r)}{\sqrt{n}}$ for some continuous function $\psi(r)$ over $r \in [b,c]$, $\mathcal{W}_n \rightsquigarrow \chi_1^2(\delta_\Psi)$, noncentral chi-squared with noncentrality parameter $\delta_\Psi = \Psi^2 V_{1,2}^{-1}$, where $\Psi = \frac{\int_b^c \psi(r) k_u(r,r)dr}{\int_b^c k_u(r,r)dr}$.

The Wald statistic $\mathcal{W}_n$ is useful as a specific break test for parameter cross section constancy against a specific alternative with given break point $b \in (a,c)$. If such parameter instability is suspected and its location is unknown then this is a case where the change point parameter appears only under the alternative. In such a case a search procedure such as that of Andrews (1993) could be conducted but here the test would focus on cross section instability with an unknown change point that is estimated as part of the procedure in contrast to a simple time series change point. In that event, the limit theory would differ from Theorem 4*(i)* and the details are left for future study.

Theorem 4*(iii)* shows that the simple test based on $\mathcal{W}_n$ has local power as a more general omnibus test against a nonparametric alternative with noncentrality parameter δ_Ψ that depends on the cross section functional form $\psi(r)$ of the departure from the null of constancy. This test may also be interpreted as a form of robustness specification check on whether the simple scalar operator form of θ in the model (5) is supported by the data against a more complex operator formulation of the ARH(1). In a similar manner we can also test hypotheses such as the null $\mathcal{H}_0 : \theta = \theta_0$, involving some specific value against a general functional alternative such as $\mathcal{H}_1 : \theta = \theta_0 + \frac{\psi(r)}{\sqrt{n}}$. Specific tests of this type will also be relevant in the unit root case.

2.6 The curved VAR case

This Section briefly extends model (3) to the following functional VARH(1) case

$$X_t = AX_{t-1} + u_t, \; t = 1, ..., n, \tag{34}$$

where A is an $m \times m$ matrix whose eigenvalues all lie in the unit circle and where u_t and X_t are now m-vector random elements in the Hilbert space $\mathcal{H} = \times_1^m L_2[a,b]$ of m-vector functions $\{f = (f_i)\}$ with inner product $\langle f, g \rangle = \int_a^b f'g$ and norm $||f||_2 = (\int_a^b f'f)^{1/2}$ where integration is over the finite observational interval $[a,b]$, as in (3). With initial conditions of (34) in the infinite past, back substitution yields the usual representation $X_t = \sum_{j=0}^{\infty} A^j u_{t-j}$. The following conditions are assumed.

Assumption A2.

(i) The m-vector sequence u_t in (34) is a stationary and ergodic $\times_1^m L_2[a,b]$-valued mds over t accompanied by the natural filtration $\mathcal{F}_t = \sigma(u_t, u_{t-1}, ...)$, positive definite matrix kernel $K_u(r,s) = \mathbb{E}u_t(r)u_t(s)' \in \times_1^{m\times m} L_2[a,b]^2$, and positive definite aggregate variance matrix $\Sigma_{u,ab} = \int_a^b \mathbb{E}u_0u_0' =: \int_a^b K_u(r,r)dr$, which we collectively write as $u_t \sim mds(0, K_u)$. Fourth order moments of u_t are finite and u_t has càdlàg realized sample paths $u_t(r)$.

(ii) The eigenvalues $\lambda_i : i = 1, \ldots, m$ of A all lie within the unit circle, so that $|\lambda_i| < 1, \forall i$.

Least squares estimation of (34) gives

$$\widehat{A} = A + \left(\sum_{t=1}^{n} \int_a^b u_t X_{t-1}'\right)\left(\sum_{t=1}^{n} \int_a^b X_{t-1}X_{t-1}'\right)^{-1}. \tag{35}$$

By stationarity and ergodicity, $\frac{1}{n}\sum_{t=1}^{n}\int_a^b X_{t-1}X_{t-1}' \overset{a.s.}{\to} \int_a^b \mathbb{E}X_{t-1}X_{t-1}' = \int_a^b K_X(r,r)dr = \sum_{j=0}^{\infty} A^j \Sigma_{u,ab} A^{j\prime}$ since $X_t = \sum_{j=0}^{\infty} A^j u_{t-j}$, and analogous to the scalar case, using martingale central limit theory we have $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\int_a^b u_t \otimes X_{t-1} \rightsquigarrow \mathcal{N}(0, \int_a^b (K_u(r,r) \otimes K_X(r,r))\,dr$. Then, as shown in the proof of Theorem 5, the asymptotic variance matrix of $\sqrt{n}\left(\mathrm{vec}(\widehat{A} - A)\right)$ is

$$V_{AA} = \Upsilon_X^{-1}\left(\int_a^b\int_a^b K_u(r,s) \otimes K_X(r,s)\right)\Upsilon_X^{-1}, \text{ with } \Upsilon_X = I_m \otimes \int_a^b K_X(r,r)dr, \tag{36}$$

where $K_X(r,s) = \sum_{j=0}^{\infty} A^j K_u(r,s) A^{j\prime}$, so that $\mathrm{vec}(K_X(r,s)) = \left(\sum_{j=0}^{\infty} A^j \otimes A^j\right)\mathrm{vec}(K_u(r,s)) = (I_m - A \otimes A)^{-1}\mathrm{vec}(K_u(r,s))$. Extending Theorem 1 we have the following limit theory.

Theorem 5. *Under Assumption A2 and for model (34), $\widehat{A} \overset{a.s.}{\to} A$ and $\sqrt{n}(\widehat{A}-A) \rightsquigarrow \mathcal{N}(0, V_{AA})$.*

In addition to providing limit theory for the stationary and ergodic VARH(1) model (34), Theorem 5 also covers asymptotics for higher order stationary and ergodic VARH(p) models upon use of the corresponding companion VARH(1) system. More particularly, the results cover the stationary and ergodic p-th order scalar ARH(p)

$$y_t = \sum_{j=1}^{p} \theta_j y_{t-j} + e_y = \theta' \underset{\sim}{y}_t + e_t, \quad t = 1, \ldots, n \tag{37}$$

where e_t and y_t are random elements in $L_2[a,b]$, with e_t satisfying Assumption A1 with its covariance kernel $k_e(r,s) \in L_2([a,b]^2)$ and the roots of $z^p - \sum_{j=1}^{p} \theta_j z^{p-j} = 0$ all lying within the unit circle. Writing $Y_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'$, the companion form representation of (37) is the VARH (34) with $Y_t = AY_{t-1} + u_t$, $p \times p$ companion matrix A and innovation process u_t given in the usual way for scalar time series by

$$A = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_{p-1} & \theta_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad u_t = \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{38}$$

where u_t has $p \times p$ diagonal matrix covariance kernel $K_u(r,s) = \text{diag}[k_e(r,s), 0, \dots, 0]$. Since $Y_t = \sum_{j=0}^{\infty} A^j u_{t-j}$ is stationary and ergodic as before, the covariance kernel of Y_t has the same series form $K_Y(r,s) = \mathbb{E}[Y_t(r)Y_t(s)'] = \sum_{j=0}^{\infty} A^j K_u(r,s) A^{j\prime}$ and we again have the limiting form $\frac{1}{n}\sum_{t=1}^{n}\int_a^b Y_{t-1}Y_{t-1}' \overset{a.s.}{\to} \int_a^b \mathbb{E}Y_{t-1}Y_{t-1}' = \int_a^b K_Y(r,r)dr = \sum_{j=0}^{\infty} A^j \Sigma_{u,ab} A^{j\prime}$, but the matrix $\Sigma_{u,ab} = \int_a^b K_u(r,r)dr = \text{diag}[\int_a^b k_e(r,r)dr, 0, \dots, 0]$ is now singular. However, the covariance matrix $\Sigma_{Y,ab} := \int_a^b \mathbb{E}Y_{t-1}Y_{t-1}' = \sum_{j=0}^{\infty} A^j \Sigma_{u,ab} A^{j\prime}$ is positive definite, just as it is in the usual (scalar time series) case (Anderson, 1971, Lemma 5.5.5) for the same reason, viz., the sum of the positive semidefinite matrices $A^j \Sigma_{u,ab} A^{j\prime}$ is positive definite.[3]

Least squares estimation of (37) gives $\widehat{\theta} = \left(\sum_{t=1}^{n}\int_a^b Y_{t-1}Y_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n}\int_a^b Y_{t-1}u_t\right)$ and Theorem 5 then specializes to the following result.

Corollary 1. *In the stationary and ergodic ARH(p) model (37), $\widehat{\theta} \overset{a.s.}{\to} \theta$ and*

$$\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}\left(0, \rho_e^2 V_{\theta\cdot}\right), \text{ with } \rho_e^2 = \frac{\int_a^b\int_a^b k_e(r,s)^2 drds}{(\int_a^b k_e(r,r)dr)^2}, \; V_\theta = \left(\sum_{j=0}^{\infty} A^j E_p A^{j\prime}\right)^{-1}, \tag{39}$$

and $E_p = diag\{1, 0, \dots, 0\}$.

The weighted autoregression case follows similarly by replacing the factor ρ_e^2 with $\rho_{e,w}^2$, defined just as in (8). Corollary 1 reduces to the ARH(1) case since $A = \theta$ is scalar and $E_p = 1$ when $p = 1$, so that $V_\theta = \sum_{j=0}^{\infty} \theta^{2j} = (1-\theta^2)^{-1}$.

3 Functional fixed effects and dynamic panel comparisons

This section considers the model

$$X_t = \alpha + \theta X_{t-1} + u_t, \tag{40}$$

[3] Suppose, for any p-vector $w = (w_1, \cdots, w_p)'$, $w'(\int_a^b K_Y(r,r)dr)w = \sum_{j=0}^{\infty} w'A^j\Sigma_{u,ab}A^{j\prime}w = 0$. Then, for $j = 0$ we have $w'\Sigma_{u,ab}w = 0$, which implies the first element $w_1 = 0$ since $\sigma_{e,ab}^2 = \int_a^b k_e(r,r)dr > 0$. Similarly, $w'A^j\Sigma_{u,ab}A^{j\prime}w = 0$ for all $j = 1, 2, \dots$ implies the first element of $w'A^j$ is zero for all $j \geq 1$ and this, in turn, implies $w = 0$. To see this, note that for the first element of $w'A$ to be zero requires $w_1\theta_1 + w_2 = 0$, so $w_2 = 0$ because $w_1 = 0$. Similarly, for successive first elements of $w'A^j$ for $j > 2$. Thus, the first element of $w'A^2$, for instance, is $w_1(\theta_1^2 + \theta_2) + w_2(\theta_1) + w_3$ so for this element to be zero requires $w_3 = 0$ and so on until we have established that $w = 0$. Hence, $\Sigma_{Y,ab} = \sum_{j=0}^{\infty} A^j\Sigma_{u,ab}A^{j\prime}$ is positive definite.

in $L_2[a,b]$ with coordinate form $X_t(r) = \alpha(r) + \theta X_{t-1}(r) + u_t(r)$ which involves an unknown continuum of fixed effects, $\alpha \in L_2[a,b]$ over the cross section space in addition to the autoregressive parameter θ. For this model in the stationary case $|\theta| < 1$, the curves X_t have nonzero mean level $\mathbb{E}X_t = \frac{\alpha}{1-\theta}$ that is curve dependent, which will typically be realistic in applications. For example, the SLP clothing data shown in Figure 1 evidently have nonzero mean clothing expenditure whose share of expenditure depends on the overall expenditure level as it fluctuates over time.

In what follows it is convenient to assume that the intercept $\alpha(r)$ is a càdlàg function in $L_2[a.b]$ with possible jump discontinuities, which allows $\alpha(r)$ to represent cluster fixed effects, which are commonly used in dynamic panel modeling. The model (40) may be considered a function space version of a simple dynamic panel autoregression with individual fixed effects (Hahn and Kuersteiner, 2002) and interactive fixed effects models (Phillips and Sul, 2007; Moon and Weidner, 2017).[4] In the latter case $X_{it} = \theta X_{it-1} + u_{it}$ with $u_{it} = \lambda_i' f_t + e_{it}$ which involves additive errors $e_{it} \sim_{iid} (0, \sigma_e^2)$ combined with unknown factors f_t and factor loadings λ_i that induce cross section dependence with second moments $\mathbb{E}(u_{it}u_{jt}) = \lambda_i'\mathbb{E}(f_t f_t')\lambda_j|\sigma_e^2 =: \omega_{ij}$, which capture cross section dependence in a manner that relates to the covariance kernel $k_u(r,s)$ of the functional fixed effects model (40).[5] Additional complexity may be introduced by allowing for temporal dependence in the factors f_t and introducing curve dependence in factor loadings as in Li et al. (2025). But these extensions are not explored in the present work.

Profiling out the fixed functional effect α, given θ, we have

$$\widetilde{\alpha}(\theta) = \underset{\alpha}{\operatorname{argmin}} \frac{1}{n} \sum_{t=1}^{n} \|X_t - \alpha - \theta X_{t-1}\|^2 = \overline{X} - \theta \overline{X}_{-1},$$

using sample mean notation $\overline{X} = n^{-1} \sum_{t=1}^{n} X_t$. Taking deviations from these functional means gives $\widetilde{X}_t = X_t - \overline{X}$, $\widetilde{X}_{t-1} = X_{t-1} - \overline{X}_{-1}$ and the functional least squares estimators

$$\widetilde{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \int_a^b \left[\widetilde{X}_t(r) - \theta \widetilde{X}_{t-1}(r)\right]^2 dr = \frac{\sum_{t=1}^{n} \int_a^b \widetilde{X}_t(r)\widetilde{X}_{t-1}(r)dr}{\sum_{t=1}^{n} \int_a^b \widetilde{X}_{t-1}^2(r)dr} \tag{42}$$

$$\widetilde{\alpha} = \overline{X} - \widetilde{\theta}\,\overline{X}_{-1}, \tag{43}$$

with the resulting fitted regression $X_t = \widetilde{\alpha} + \widetilde{\theta} X_t + \widetilde{u}_t$. Test statistics using t-ratios for inference concerning the autoregressive parameter θ in (40) are constructed in the same manner as before

[4]The model is also a special case of an incidental polynomial trend curved autoregression

$$X_t(r) = \sum_{j=0}^{J} \beta_j(r) t^j + \theta X_{t-1}(r) + u_t(r), \quad r \in [a,b]. \tag{41}$$

with functional fixed effects $\beta_0(r)$ and incidental trend functions $\sum_{j=1}^{J} \beta_j(r) t^j$, extending the framework of dynamic panels with scalar incidental linear trends (Moon and Perron, 2004; Moon et al., 2007).

[5]Moon and Weidner (2017) consider only the stationary interactive fixed effects case with $|\theta| < 1$ but allow for heterogeneity in e_{it} over both i and t. Phillips and Sul (2007) consider both stationary and unit root interactive fixed effects models and give bias formulae for both cases.

with $\widetilde{\theta}$ and the regression residuals $\widetilde{u}_t(r)$. For example, we have $\widetilde{t}_{\theta,fe} = \frac{\widetilde{\theta}-\theta}{\widetilde{s}_{\theta,fe}}$, where the standard error $\widetilde{s}^2_{\theta,fe}$ is

$$\widetilde{s}^2_{\theta,fe} := \left(\sum_{t=1}^{n}\int_a^b \widetilde{X}^2_{t-1}(r)dr\right)^{-2}\left(\sum_{t=1}^{n}\int_a^b\int_a^b \widetilde{X}_{t-1}(r)\widetilde{u}_t(r)\widetilde{u}_t(s)\widetilde{X}_{t-1}(s)drds\right), \tag{44}$$

as in (27), but using the residuals $\widetilde{u}_t(r)$ and demeaned curve observations $\widetilde{X}_{t-1}(r)$. Similar constructions lead to the test statistics $\hat{t}_{\theta,fe}$, $\widetilde{t}^*_{\theta,fe}$.

The following result extends aspects of Theorems 1 and 3 to the functional fixed effects dynamic panel case.

Theorem 6. *Under Assumption A1 with $|\theta| < 1$, as $n \to \infty$, $\widetilde{\theta} \overset{a.s.}{\to} \theta$, $\widetilde{\alpha} \overset{a.s.}{\to} \alpha$, and*

$$\sqrt{n}\left(\widetilde{\theta}-\theta\right) \rightsquigarrow \mathcal{N}\left(0, V_\theta\right), \text{ with } V_\theta := \left(1-\theta^2\right)\rho^2_\varepsilon, \tag{45}$$

$$\sqrt{n}\left(\widetilde{\alpha}-\alpha\right) \rightsquigarrow \mathcal{G}\left(0, k_\alpha(r,s)\right), \tag{46}$$

$$\hat{t}_{\theta,fe},\ \widetilde{t}_{\theta,fe},\ \widetilde{t}^*_{\theta,fe} \rightsquigarrow \mathcal{N}\left(0,1\right), \tag{47}$$

where $\mathcal{G}$ is a zero mean Gaussian process on $L_2[a,b]$ with covariance kernel

$$k_\alpha(r,s) = k_u(r,s) + \left(\frac{\alpha(r)\alpha(s)}{(1-\theta)^2}\right)V_\theta. \tag{48}$$

Under the additional requirement of zero third moments in u_t we have the bias expansion

$$\mathbb{E}\left(\widetilde{\theta}\right) - \theta = -\frac{2\theta}{n}\rho^2_u - \frac{1+\theta}{n} + o\left(\frac{1}{n}\right), \tag{49}$$

Remarks

3(a) Stationarity and ergodicity deliver consistency and martingale central limit theory gives the asymptotic distribution of $\widetilde{\theta}$, just as in the proof of Theorem 1. The limit theory (45) for $\widetilde{\theta}$ is the same as the simple functional autoregression without fixed effects and has an asymptotic variance that again reduces variation because of the additional information present in the use of curve data. The variance again carries the effects of curve dependence through the error covariance kernel $k(r,s)$. Because the limit distribution embodies these cross section dependence effects there is no acceleration of convergence in contrast to the scalar dynamic panel model with fixed effects, where averaging over N *iid* cross section errors with $N \to \infty$ raises the convergence rate by $\sqrt{N}$ and introduces an $O\left(\frac{1}{n}\right)$ time series bias into the limit theory, depending on the expansion rate of N in relation to the time series sample size n (Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003).[6]

[6] When cross section dependence in the equation error $u_t(r)$ decays to zero outside small shrinking neighborhoods of width $O(1/N)$, then a similar $\sqrt{N}$ acceleration in the convergence rate of $\widetilde{\theta}$ occurs, as discussed in Remark **1(e)** and the limit theory (13). So the limit theory in the scalar dynamic panel model with fixed effects is

3(b) Limit theory for the fixed effects estimator $\widetilde{\alpha}$ relies on a Hilbert space central limit theorem leading to a limiting Gaussian process $\mathcal{G}$ in $L_2[a,b]$. There are two components of the limiting covariance kernel $k_\alpha(r,s)$ shown in (48): the first is the cross section covariance kernel $k_u(r,s)$ of the equation error u_t; and the second depends on the limiting variance V_θ of $\widetilde{\theta}$ as well as the fixed effect α. The fixed effect limit theory therefore differs in these two respects from the corresponding fixed effect limit theory in dynamic panels with *iid* errors, which ensures that in such panels the estimate of θ converges at a faster rate than the fixed effect estimates whose distributions are therefore unaffected in the limit (Hahn and Kuersteiner, 2002).

3(c) Expression (49) gives the asymptotic bias of the functional fixed effects autoregressive estimator $\widetilde{\theta}$ to order $O\left(\frac{1}{n}\right)$, revealing the impact of cross section curve dependence in autoregressive estimation with functional fixed effects and generalizing the earlier bias expression (30) given in Theorem 3 without fixed effects. Importantly, that expression (30) remains present in the new formula as the first term on the right side of (49) and arises for precisely the same reasons. The second term in the bias expression (49) is related to corresponding results in the literature on dynamic panel bias in first order autoregression with fixed effects. In fact, the downward bias term $-\frac{1+\theta}{n}$ in (49) relates to the early findings of Marriott and Pope (1954) on simple serial correlation coefficient bias with a mean correction and later formulae for dynamic panel autoregression with cross section and time series sample size asymptotics obtained by Nickell (1981); Beggs and Nerlove (1988) and more recently Hahn and Kuersteiner (2002); Alvarez and Arellano (2003); Phillips and Sul (2007). Importantly, in the present case the bias expression (49) differs from those earlier findings because of the primary term of (49), which carries the effects of cross section dependence in the innovations. In that respect, it relates to the formulae obtained in Phillips and Sul (2007, equation (29)) and Moon and Weidner (2017, theorem 4.3) for bias in stationary interactive fixed effects autoregressions with cross section dependence induced by the presence of unknown factors in the regression.

3(d) Theorem 6 gives limit theory as the time series sample size $n \to \infty$. This contrasts with the short wide panel regression literature that focuses on limit theory for datasets with small fixed time series sample size n and large $N \to \infty$ cross sections with *iid* errors. In such cases, time series bias removal can be effected using a wide variety of instrumental variable and GMM methods on which there is now an extensive literature following Anderson and Hsiao (1981); Arellano and Bond (1991) with recent corrections and unit root cases (Phillips and Han, 2015; Phillips, 2018). Similar techniques may be

captured within this more general context as a special case through narrow band cross section dependence. In that case, if $N \to \infty$ faster than $n \to \infty$, a bias term such as (49) will appear in the limit theory. In this general context as we have shown, two terms appear in the bias expression (49), whereas in the scalar dynamic panel model with fixed effects only one term, viz. $\frac{1+\theta}{n}$, appears in the bias correction. This difference is explained by the fact that the analysis of the dynamic panel model bias in Alvarez and Arellano (2003) assumes *iid* innovations across section, which implies that $k_u(r,s) = k(r,s) = 0$ for all $r \neq s$, which removes the first bias term in (49).

considered in the present function space setting when cross section dependence in the equation error $u_t(r)$ decays to zero outside small shrinking neighborhoods of the type considered in Remark **1(d)** and fn.6. But these are not pursued in the present work.

3(e) The limit theory (45) for $\widetilde{\theta}$ differs in another important way from the dynamic panel literature. As shown by Alvarez and Arellano (2003); Hahn and Kuersteiner (2002), in asymptotics that involve both large time series ($n \to \infty$) and large *iid* cross section dynamics ($N \to \infty$), the bias term $\frac{1+\theta}{n}$ figures and must be included in the limit theory unless $n \to \infty$ faster than N. In the latter case the time series asymptotics dominate because of the $\sqrt{nN}$ convergence rate that applies in stationary panels with no cross section dependence, which implies that $\sqrt{nN}\frac{1+\theta}{n} \to 0$. In the functional data case the additional term $\frac{1+\theta}{n}$ is always present in the bias expansion (49) with fitted functional fixed effects. In empirical work with a small time series sample size n, curved data and functional fixed effects, there may be an advantage in using a bias corrected estimator of the form $\widetilde{\theta}^* = \widetilde{\theta} + \frac{2}{n}\widetilde{\theta}\widehat{\rho}_u + \frac{1+\widetilde{\theta}}{n}$ in inference. From Theorem 6 it follows directly that $\sqrt{n}\left(\widetilde{\theta}^* - \theta\right) \rightsquigarrow \mathcal{N}(0, V_\theta)$.

3(f) In finite samples, inference in the functional fixed effects model (40) can be implemented by the same residual wild-bootstrap principle used for the simple ARH model. For a null hypothesis $\mathcal{H}_0 : \theta = \theta_0$, estimate (40), form the fitted residual curves $\widetilde{u}_t(r) = X_t(r) - \widetilde{\alpha}(r) - \widetilde{\theta}X_{t-1}(r)$, and recenter them over t. Let $\{\xi_t\}_{t=1}^n$ be iid multipliers with mean zero and variance one, and define $\widetilde{u}_t^*(r) = \xi_t\widetilde{u}_t(r)$. Bootstrap samples are generated under the null by $X_t^*(r) = \widetilde{\alpha}(r) + \theta_0 X_{t-1}^*(r) + \widetilde{u}_t^*(r)$, with the initial curve fixed at its sample value. For each bootstrap sample, $\widetilde{\theta}$ and the relevant statistic $\hat{t}_{\theta,fe}$, $\widetilde{t}_{\theta,fe}$, or $\widetilde{t}^*_{\theta,fe}$ are recomputed using the same fixed-effect demeaning and variance formulae as in the original sample. Because a single multiplier is applied to the entire residual curve at date t, the bootstrap preserves the estimated cross section dependence in $u_t(r)$ while imposing no parametric diffusion law for the covariance kernel. Critical values and bootstrap-t confidence intervals are then obtained from the conditional bootstrap distribution of the recomputed statistics.

4 Simulations

This section assesses the finite sample behavior of the stationary curve autoregression tests developed above. The main concern is whether inference remains reliable when the covariance structure of the curve innovations is estimated rather than known, and whether the additional cross section information in the curves translates into useful power gains relative to a scalar autoregression. We therefore focus on wild-bootstrap critical values, which provide nonparametric inference for the test statistics and do not impose a parametric diffusion law in the bootstrap step.

The baseline data generating process uses the diffusion model $J_{c,t}(r)$ with $c \in \{-5, -3, 0, 2\}$ to generate *iid* error curve innovations $u_t(r)$ on $r \in [0, 1]$. To examine robustness to departures from the diffusion family, we also generate *iid* error curves from the segmented Brownian motion process in (84), shown in Figure 9, with segment numbers $K \in \{1, 2, 3\}$. Across designs the initial value is $X_0(r) = 0$, the grid frequency is $h = 0.01$, the sample sizes are $n \in \{20, 60, 100\}$, and all rejection frequencies are based on $5,000$ replications. Results for models with functional fixed effects are reported separately in Section 10.

Figure 5 reports local power curves for testing the stationary null hypothesis $\mathcal{H}_0 : \theta = 0.5$ under the diffusion design. The three functional tests are the statistics $\widetilde{t}_\theta$, $\hat{t}_\theta$, and $\widetilde{t}^*_\theta$, with standard errors computed from the variance estimates in (27), (28), and (29). In each case $k_u(r, s)$ is estimated from sample regression residuals, either in covariance-kernel form or in the sandwich form in (27). As a benchmark, the figure also reports the usual scalar AR(1) t-test computed from a no-intercept autoregression with $n = 100$ and the two-sided normal critical value.

The diffusion design fingings are as follows. First, at the null value $\theta = 0.5$, Figure 5 shows that the wild-bootstrap tests have rejection probabilities close to the nominal 5% level. The statistic $\widetilde{t}_\theta$ is especially close to nominal size, while $\hat{t}_\theta$ and $\widetilde{t}^*_\theta$ show some oversizing in the smallest samples that declines as n increases. Second, the same figure shows the expected increase in power as θ moves away from the null and as the time series dimension grows. Third, the functional tests are much more powerful than the scalar AR(1) benchmark when cross section dependence is weak or moderate, especially for negative diffusion coefficients. When $c \geq 0$, stronger dependence over r reduces the effective cross section information, so the incremental power gain from using curves is smaller.

Figure 6 examines robustness by replacing diffusion errors with segmented Brownian motion errors. This design changes the covariance mechanism while leaving the coefficient testing problem and wild-bootstrap implementation unchanged. At the null value, Figure 6 shows that the size performance of the three coefficient tests remains close to the diffusion benchmark. The power curves in the same figure are also stable across $K = 1, 2, 3$, indicating that the wild-bootstrap tests do not rely on the diffusion family including the actual error process. This robustness is important because, in applications, the covariance kernel is estimated from residual curves rather than specified by a known parametric model.

The final design focuses on the cross section break specification test. The model follows (31) with diffusion curve errors and break point $b = 0.5$. Under the null, $\mathcal{H}_0 : \theta_1 = \theta_2 = 0.7$; under the alternative, θ_1 remains fixed at 0.7 and θ_2 varies over $[0.5, 0.9]$. The Wald statistic $\mathcal{W}_n$ is constructed as in Section 2.5, and critical values are again obtained by wild bootstrap.

At the null value $\theta_2 = \theta_1 = 0.7$, Figure 7 shows good size control for the Wald test across sample sizes $n \in \{20, 60, 100\}$ and diffusion coefficients $c \in \{-5, -3, 0, 2\}$. The same figure then shows that rejection probabilities rise as θ_2 moves away from θ_1. The evidence therefore supports the break test as a useful diagnostic for coefficient homogeneity across the curve domain, complementing the coefficient tests in the baseline stationary autoregression.

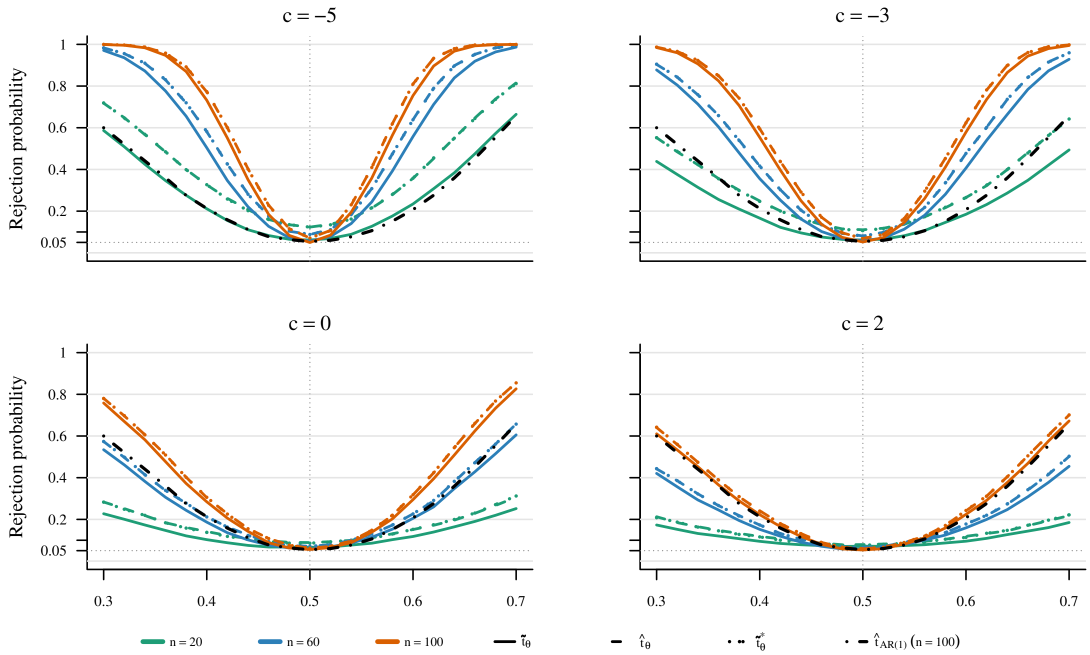


Figure 5: Wild-bootstrap local power curves of the autoregressive coefficient tests $\widetilde{t}_\theta, \hat{t}_\theta$, and $\widetilde{t}^*_\theta$ with stationary curve time series generated by diffusion processes with diffusion coefficient $c \in \{-5, -3, 0, 2\}$, together with the scalar AR(1) benchmark $\hat{t}_{\mathrm{AR}(1)}$ for $n = 100$.

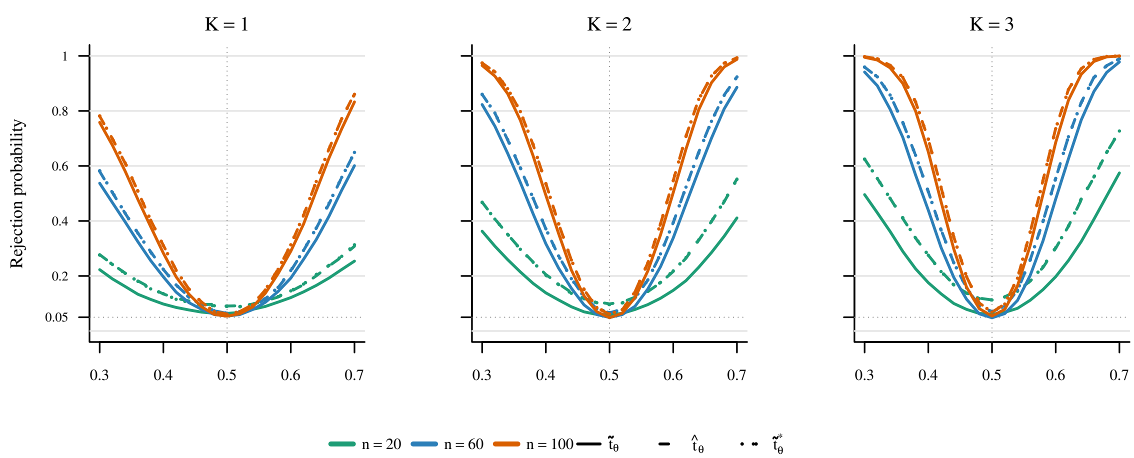


Figure 6: Wild-bootstrap local power curves of the autoregressive coefficient tests $\widetilde{t}_\theta, \hat{t}_\theta$, and $\widetilde{t}^*_\theta$ with stationary curve time series generated by segmented Brownian motion error curves with segment numbers $K \in \{1, 2, 3\}$.

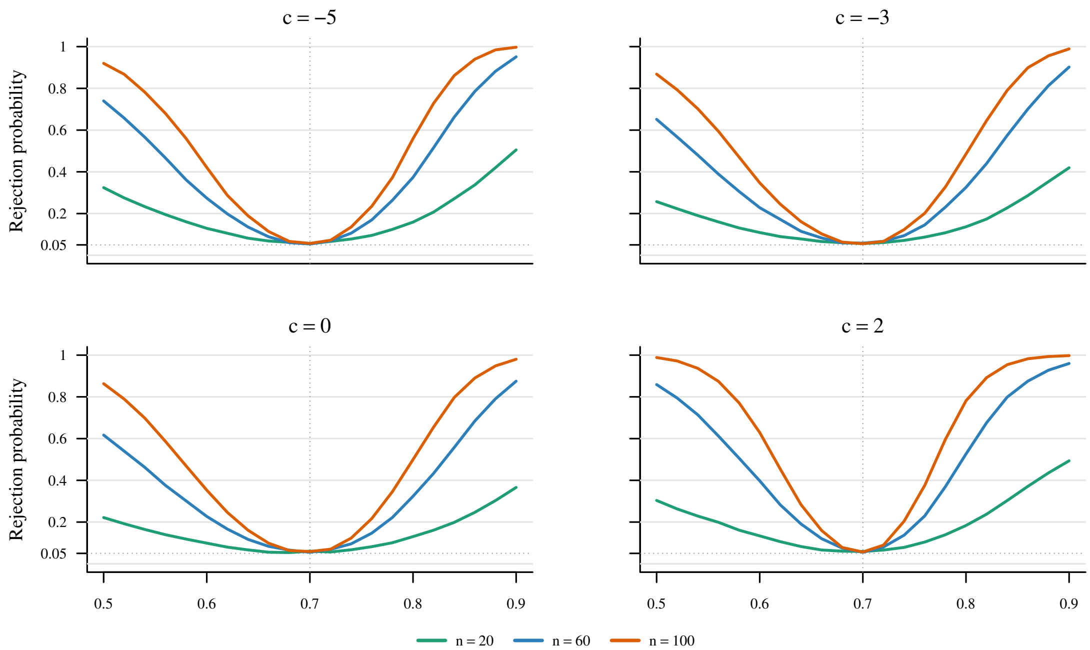


Figure 7: Power curves of the wild-bootstrap specification test $\mathcal{W}_n$ of the null $\mathcal{H}_0 : \theta_1 = \theta_2 = 0.7$ under the alternative of a cross section break in an autoregression generated with diffusion process curves using coefficients $c \in \{-5, -3, 0, 2\}$ with a cross section break point at $b = 0.5$.

5 Empirical Dynamics of Household Engel Curves

Previous studies have documented nonlinear Engel curves for many goods and services.[7] These curves can also shift over time as household income, demographics, prices, and economic conditions change (Lewbel, 2008). The empirical question studied here is whether such shifts are transitory around a stable functional relationship or whether the full Engel curves display near-permanent persistence. The curve time series framework developed in this paper is well suited to this question because it treats each monthly Engel curve as a functional observation and tests temporal dependence in the entire curve, rather than at a small number of expenditure points.

We draw on data from the Singapore Life Panel (SLP), which covers September 2015 to May 2023 and yields 93 monthly observations.[8] Using cross section data for Singaporeans aged 50 to 70, we first follow Blundell et al. (1998, 2003) to estimate monthly Engel curves nonparametrically. These estimated monthly curves are then used as the functional time series observations in the ARH analysis.

[7] See, for example, Banks et al. (1997), Blundell et al. (1998, 2003, 2007), and Reaños and Wölfing (2018).
[8] Section 11 provides details on the SLP dataset.

Specifically, following Blundell et al. (1998), we adopt the generalized Working-Leser model:

$$y_{ij} = g_j(\ln x_i) + u_{ij},$$

where y_{ij} is the budget share of the j-th category for household i, $\ln x_i$ is the log of total nondurable expenditure for household i, and u_{ij} is the error term. We use local linear nonparametric regressions to estimate the monthly Engel curves. Figures 1-8 display the smoothed curves for nine spending categories from September 2015 to May 2023. Several categories, including clothing and transport, show pronounced nonlinearities, whereas food and utilities are closer to monotone or approximately linear patterns, consistent with earlier Engel curve evidence (Lewbel, 2008). The figures also show visible month-to-month movement in the full curves, suggesting persistent changes in how budget shares vary with nondurable expenditure. Small irregularities near the endpoints are most apparent where the data are sparse; these endpoint features reflect the lower precision that is typical in nonparametric curve estimation (Blundell et al., 2003).

We next apply the fitted fixed-effect ARH model (40) to these monthly curve observations. The empirical null hypothesis is $\mathcal{H}_0 : \theta = 0.9$, tested against the one-sided alternative $\mathcal{H}_1 : \theta < 0.9$. This null represents very high persistence in the full Engel curve, while rejection points to a still persistent but more clearly stationary dynamic response. Guided by the simulation evidence, Table 1 reports the fixed-effect estimate $\widetilde{\theta}$, its bias-corrected version $\widetilde{\theta}^*$, and the statistic $\tilde{t}_\theta$, which has the strongest finite-sample performance among the coefficient-based tests, with wild-bootstrap inference. The bias correction uses $\widetilde{\theta}^* = \widetilde{\theta} + \frac{2}{n}\widetilde{\theta}\widehat{\rho}^2_{u,wb} + \frac{1+\widetilde{\theta}}{n}$, where $\widehat{\rho}^2_{u,wb}$ is computed from the residual covariance kernel preserved by the wild-bootstrap procedure. The table also reports wild-bootstrap 95% confidence intervals for θ based on the bootstrap-t distribution of $\tilde{t}_\theta$.

The estimates show substantial persistence in all nine Engel curve series, but the strength of that persistence varies across expenditure categories. The wild-bootstrap test rejects $\mathcal{H}_0$ at the 1% level for clothing, education, health, housing, insurance, utility, and transport. These categories therefore display highly autoregressive but clearly mean-reverting curve dynamics. Food and leisure are the two most persistent categories: food has a bias-corrected estimate above 0.9 but a bootstrap confidence interval below unity, while leisure has a confidence interval that extends slightly above unity and does not reject $\theta = 0.9$. Overall, the evidence favors stationary but highly persistent functional dynamics for most household Engel curves in the SLP data.

Table 1: Wild-bootstrap functional t-test of the null hypothesis $\mathcal{H}_0 : \theta = 0.9$ against the one-sided alternative hypothesis $\mathcal{H}_1 : \theta < 0.9$ for stationary Engel curves with the full SLP dataset. The model estimated is an ARH(1) with fitted fixed-effect curves as in equation (40). The table reports $\widetilde{\theta}$, the bias-corrected estimate $\widetilde{\theta}^*$ computed with the wild-bootstrap residual-covariance estimate $\widehat{\rho}^2_{u,wb}$, the corresponding wild-bootstrap 95% confidence interval, and the statistic $\tilde{t}_\theta$.

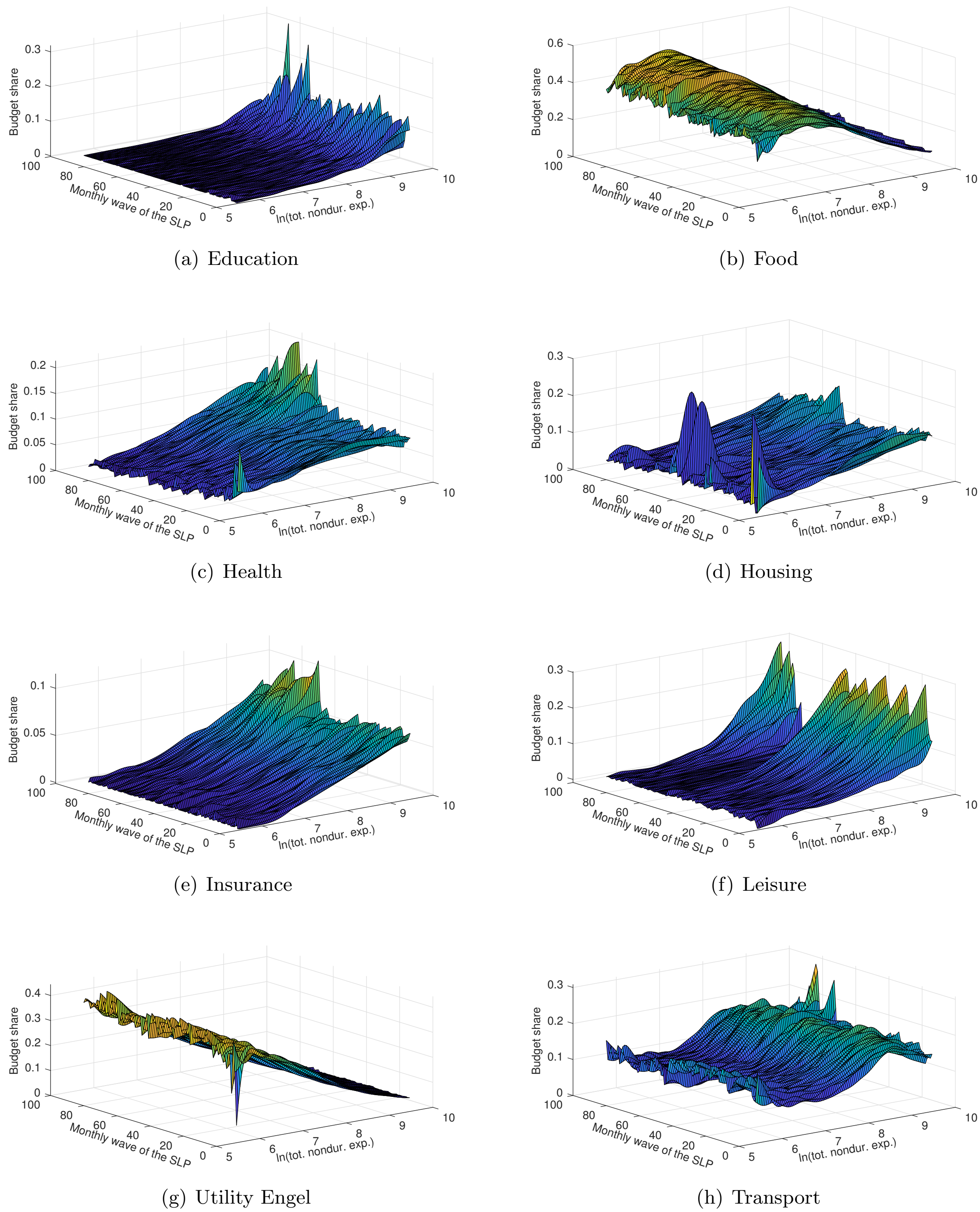
Engel Curves	$\widetilde{\theta}$	$\widetilde{\theta}^*$	95% CI	$\tilde{t}_\theta$
Clothing	0.64	0.66	[0.45, 0.86]	-2.51***
Education	0.48	0.50	[0.34, 0.65]	-5.36***
Food	0.91	0.93	[0.89, 0.96]	1.81
Health	0.76	0.78	[0.69, 0.87]	-2.73***
Housing	0.30	0.32	[0.15, 0.48]	-8.49***
Insurance	0.68	0.70	[0.58, 0.81]	-3.60***
Leisure	0.85	0.88	[0.72, 1.00]	-0.30
Utility	0.31	0.32	[0.21, 0.42]	-11.53***
Transport	0.61	0.63	[0.51, 0.74]	-5.03***

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively, using one-sided wild-bootstrap p-values from 1,000 iid multiplier draws. The confidence intervals are bootstrap-t intervals based on $\tilde{t}_\theta$.

In sum, the empirical application shows how the proposed functional approach converts a sequence of estimated Engel curves into direct evidence on full-curve temporal dependence. The results are not driven by a single point of the Engel curve or by a scalar expenditure summary: instead, they measure persistence in the whole budget-share function. For ageing seniors in Singapore, the estimated Engel curves are highly persistent over time, but most categories exhibit mean reversion that is statistically distinguishable from the high-persistence benchmark $\theta = 0.9$. Food and leisure are the main exceptions and appear closest to near-unit persistence. These findings will be further assessed using functional unit root tests developed in Phillips and Jiang (2025) and reported in other work.

6 Conclusions

This paper has developed asymptotic theory for parametric estimation and inference in first order autoregression with function valued curve time series. The results show that autoregressive models in function space can allow for great flexibility in modeling cross section dependence while retaining many of the advantages of parametric specifications. Cross section curve formulations provide a simple way to embody high dimensional wide panel datasets in empirical work, raising efficiency in estimation and shortening confidence intervals in inference about the AR coefficient in both stationary and unit root cases. Much of the limit theory and bias analyses in scalar autoregressions are shown to have extensions in the function space setting that are easy to interpret in application. The limit theory here covers stationary autoregression including dynamic panel implementations with curve fixed effects and explores the differences and links to standard panel data models. The new methods are illustrated using high dimensional consumption data for ageing seniors obtained from the Singapore Life Panel. The results



(a) Education (b) Food (c) Health (d) Housing (e) Insurance (f) Leisure (g) Utility Engel (h) Transport

Figure 8: Smoothed Engel curves for selected expenditure categories in the SLP dataset

provide an empirical dynamic analysis of household Engel curves for expenditure on various categories of goods and services.

The methods and results presented here can be extended in various directions that utilize high dimensional cross section data. Unit root and local unit root extensions provide mechanisms for testing nonstationarity and are considered in a related paper (Phillips and Jiang, 2025). There is also scope for further specification testing that could provide a more direct interface with models that involve Hilbert space operator formulations of autoregressions. In addition to structural break analysis in the cross section data studied here, it will be of interest in some applications, including those in the SLP data, to assess potential temporal heterogeneity or structural breaks in the cross section curves over time. The latter may be of particular interest to assess changes that may have occurred over the Covid-19 pandemic period.

Curve data is prominent in other empirical areas, such as the study of lifetime income profiles where both mean regression and quantile regression techniques have been employed with functional datasets and parametric models (Chen and Müller, 2012; Cho et al., 2022). There is also scope for tensor extensions in which cross section data in multiple dimensions are incorporated into functional autoregressive frameworks. All these areas of research present their own individual challenges but all share the common goal of methodological advances that enable curve data to be used successfully in empirical econometric research.

7 Appendix A

Proof of Theorem 1 Center and scale $\widehat{\theta}$ so that

$$\sqrt{n}(\widehat{\theta}-\theta)=\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\int_a^b X_{t-1}(r)u_t(r)dr}{\int_a^b \frac{1}{n}\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)dr}. \tag{50}$$

The denominator involves the stationary and ergodic sequence X_{t-1}^2 and by moment existence we have $n^{-1}\sum_{t=1}^{n}X_{t-1}^2 \to_{a.s.} \mathbb{E}\left[X_{t-1}^2\right]$. The real valued measurable function $\int_a^b X_{t-1}^2(r)dr$ is also stationary and ergodic so that

$$\frac{1}{n}\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)dr \to_{a.s.} \int_a^b k_X(r,r)dr=\frac{1}{1-\theta^2}\int_a^b k_u(r,r)dr, \tag{51}$$

where $k_X(r,s)=\mathbb{E}\left[X_t(r)X_t(s)\right]$. The numerator of (50) involves the component $X_{t-1}u_t$, which is a $\mathcal{H}$ - valued stationary martingale difference sequence (mds) with filtration $\mathcal{F}_t$. The real valued sequence $\int_a^b X_{t-1}(r)u_t(r)dr$ is then a simple stationary mds with variance

$$\begin{aligned}\mathbb{E}\left(\int_a^b X_{t-1}(r)u_t(r)dr\right)^2 &= \int_a^b\int_a^b \mathbb{E}\left\{X_{t-1}(r)X_{t-1}(s)\right\}\mathbb{E}\left\{u_t(r)u_t(s)\right\}dr\\ &=\frac{\sigma^2}{1-\theta^2}\int_a^b\int_a^b k_u(r,s)^2dsdr=\frac{\sigma^4}{1-\theta^2}\int_a^b\int_a^b k(r,s)^2dsdr. \end{aligned}\tag{52}$$

The required limit theory follows directly by the martingale central limit theorem (MGCLT). The stability condition holds because the martingale conditional variance is

$$\left\langle \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \int_a^b X_{t-1}(r) u_t(r) dr \right\rangle = \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\left(\int_a^b X_{t-1}(r) u_t(r) dr \right)^2 |\mathcal{F}_{t-1} \right]$$

$$= \frac{1}{n} \sum_{t=1}^{n} \int_a^b \int_a^b X_{t-1}(r) X_{t-1}(s) \mathbb{E}\left[u_t(r) u_t(s) |\mathcal{F}_{t-1} \right] drds$$

$$\to_{a.s.} \mathbb{E}\left[\int_a^b \int_a^b X_{t-1}(r) X_{t-1}(s) k_u(r,s) drds \right] = \frac{\sigma^4}{1-\theta^2} \int_a^b \int_a^b k^2(r,s) drds < \infty. \tag{53}$$

The Lindeberg condition requires that for all $\delta > 0$

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\left(\int_a^b X_{t-1}(r) u_t(r) \right)^2 \mathbf{1}\left(\left| \int_a^b X_{t-1}(r) u_t(r) \right| > \delta\sqrt{n} \right) \mid \mathcal{F}_{t-1} \right] \overset{p}{\to} 0, \tag{54}$$

where $\mathbf{1}(A)$ is the indicator of A. It is sufficient to show L_1 convergence, which follows by stationarity if

$$\mathbb{E}\left[\left(\int_a^b X_{t-1}(r) u_t(r) \right)^2 \mathbf{1}\left(\left| \int_a^b X_{t-1}(r) u_t(r) \right| > \delta\sqrt{n} \right) \right] \to 0,$$

which holds by virtue of Cauchy-Schwarz since

$$\mathbb{E}\left[\left(\int_a^b X_{t-1}(r) u_t(r) dr \right)^2 \right] \le \int_a^b \mathbb{E}\left\{ X_{t-1}^2(r) \right\} dr \int_a^b \mathbb{E}\left\{ u_t^2(r) \right\} dr < \infty.$$

Using the MGCLT, we then have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \int_a^b X_{t-1}(r) u_t(r) dr \rightsquigarrow \mathcal{N}\left(0, \frac{\sigma^4}{1-\theta^2} \int_a^b \int_a^b k_\varepsilon(r,s)^2 dsdr \right). \tag{55}$$

Combining (55) and (51) gives the limit distribution $\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}\left(0, \left(1-\theta^2\right) \rho_u^2\right)$ of θ, as stated. ■

Proof of Theorem 2 Since $\widehat{\theta} \to_{a.s.} \theta$, we have

$$\hat{u}_t = X_t - \widehat{\theta} X_{t-1} = u_t - \left(\widehat{\theta} - \theta \right) X_{t-1} = u_t + o_{a.s.}(1), \tag{56}$$

in $L_2[a,b]$ and then the measurable real valued average kernel estimates are consistent, giving

$$\int_a^b \int_a^b \hat{k}_u(r,s) drds = \frac{1}{n} \sum_{t=1}^{n} \int_a^b \int_a^b \hat{u}_t(r) \hat{u}_t(s) drds \overset{a.s.}{\to} \int_a^b \int_a^b k_u(r,s) drds, \tag{57}$$

$$\int_a^b \hat{k}_u(r,r)dr = \frac{1}{n}\sum_{t=1}^n \int_a^b \hat{u}_t(r)\hat{u}_t(r)dr \overset{a.s.}{\to} \int_a^b k_u(r,r)dr. \tag{58}$$

It follows that the estimate $\hat{\omega}_u^2 := \left(\int_a^b \hat{k}_u(r,r)dr\right)^{-2}\left(\int_a^b\int_a^b \hat{k}_u(r,s)^2 dsdr\right) \to_{a.s.} \omega_u^2$. Then $\hat{\omega}_\theta^2 := \left(1-\widehat{\theta}^2\right)\hat{\omega}_u^2 \to_{a.s.} \omega_\theta^2 = \left(1-\theta^2\right)\omega_u^2$. Setting the variance estimate $\hat{s}_\theta^2 = n^{-1}\left(1-\widehat{\theta}^2\right)\hat{\omega}_u$, we deduce that

$$\hat{t}_\theta = \frac{\widehat{\theta}-\theta}{\hat{s}_\theta} = \frac{\sqrt{n}\left(\widehat{\theta}-\theta\right)}{\hat{\omega}_u\left(1-\hat{\theta}^2\right)^{1/2}} \rightsquigarrow \mathcal{N}(0,1).$$

Next, from (51) we know that

$$\frac{1}{n}\sum_{t=1}^n \int_a^b X_{t-1}^2(r)dr \to_{a.s.} \int_a^b k_X(r,r)dr = \frac{\sigma^2}{1-\theta^2}\int_a^b k_\varepsilon(r,r)dr. \tag{59}$$

From (27) the continuous sandwich formula estimate is

$$\widetilde{s}_\theta^2 = \left(\sum_{t=1}^n \int_a^b X_{t-1}^2(r)dr\right)^{-2}\left(\sum_{t=1}^n \int_a^b\int_a^b X_{t-1}(r)\hat{u}_t(r)\hat{u}_t(s)X_{t-1}(s)drds\right).$$

In view of (56) we have

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n \int_a^b\int_a^b X_{t-1}(r)\hat{u}_t(r)\hat{u}_t(s)X_{t-1}(s)drds \\
&= \frac{1}{n}\sum_{t=1}^n \int_a^b\int_a^b X_{t-1}(r)u_t(r)u_t(s)X_{t-1}(s)drds + o_{a.s.}(1) \\
&= \int_a^b\int_a^b \mathbb{E}\left[X_{t-1}(r)X_{t-1}(s)\right]\mathbb{E}\left[u_t(r)u_t(s)\right]drds + o_{a.s.}(1) \\
&\overset{a.s.}{\to} \int_a^b\int_a^b k_X(r,s)k_u(r,s)drds = \frac{\sigma^4}{1-\theta^2}\int_a^b\int_a^b k(r,s)^2 drds.
\end{aligned} \tag{60}$$

Using (59) and (60) we have $n\widetilde{s}_\theta^2 \overset{a.s.}{\to} \omega_\theta^2$. Then the corresponding t-ratio statistic

$$\widetilde{t}_\theta = \frac{\widehat{\theta}-\theta}{\widetilde{s}_\theta} = \frac{\sqrt{n}(\widehat{\theta}-\theta)}{\left(n\widetilde{s}_\theta^2\right)^{1/2}} \rightsquigarrow \mathcal{N}(0,1), \tag{61}$$

as required. Using these results, it is similarly found that the t ratio $\widetilde{t}_\theta^* = \frac{\widehat{\theta}-\theta}{\widetilde{s}_\theta^*} \rightsquigarrow \mathcal{N}(0,1)$, where $\widetilde{s}_\theta^{*2}$ is given by (29). So these constructions all provide valid asymptotic tests and confidence intervals. ■

Proof of Theorem 3. The proof makes use of the Edgeworth expansion of the distribution of $\sqrt{n}(\widehat{\theta}-\theta)$ and the corresponding moment expansion in Phillips (2025) that gives the bias expression directly. Phillips (2025) assumed third moments of u_t to be zero, in line with the earlier work in Phillips (1977) which directly assumed Gaussian innovations u_t in $\mathbb{R}$. The same zero third moment behavior holds in the present case, for instance, with both Brownian motion and Brownian bridge innovations u_t. Under these conditions, we have

$$\mathbb{E}(\widehat{\theta}) - \theta = -\frac{2\theta}{n}\frac{\int_a^b \int_a^b k_u(r,q)^2 drdq}{\left(\int_a^b k_u(r,r)dr\right)^2} + o\left(\frac{1}{n}\right) = -\frac{2\theta\rho_u^2}{n} + o\left(\frac{1}{n}\right),$$

giving the stated result. ■

Proof of Theorem 4 Part(i) Write $\sqrt{n}\left(\widehat{\theta}_1 - \widehat{\theta}_2\right) = \sqrt{n}\left(\widehat{\theta}_1 - \theta_1\right) - \sqrt{n}\left(\widehat{\theta}_2 - \theta_2\right)$ and the limit theory for the two components is obtained as in the proof of Theorem 2. In particular,

$$\begin{bmatrix} \sqrt{n}\left(\widehat{\theta}_1 - \theta_1\right) \\ \sqrt{n}\left(\widehat{\theta}_2 - \theta_2\right) \end{bmatrix} = \begin{bmatrix} \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_a^b X_{t-1}(r)u_t(r)dr}{\frac{1}{n}\sum_{t=1}^n \int_a^b X_{t-1}^2(r)dr} \\ \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_b^c X_{t-1}(r)u_t(r)dr}{\frac{1}{n}\sum_{t=1}^n \int_b^c X_{t-1}^2(r)dr} \end{bmatrix} \sim_a \frac{1}{\sqrt{n}}\sum_{t=1}^n \begin{bmatrix} \frac{1}{A}\int_a^b X_{t-1}(r)u_t(r)dr \\ \frac{1}{C}\int_b^c X_{t-1}(r)u_t(r)dr \end{bmatrix}, \tag{62}$$

where $\sim_a$ signifies having the same asymptotic distribution, $A := \mathbb{E}\int_a^b X_{t-1}^2(r)dr = \int_a^b k_{X_{\theta_1}}(r,r)dr = \frac{1}{1-\theta_1^2}\int_a^b k_u(r,r)dr$, and $C := \mathbb{E}\int_b^c X_{t-1}^2(r)dr = \int_b^c k_{X_{\theta_2}}(r,r)dr = \frac{1}{1-\theta_2^2}\int_b^c k_u(r,r)dr$. The elements of the vector on the right side of (62) are martingale differences with variance matrix

$$\begin{bmatrix} \frac{1}{A^2}\mathbb{E}\left(\int_a^b X_{t-1}(r)u_t(r)dr\right)^2 & \frac{1}{AC}\int_a^b \int_b^c \mathbb{E}X_{t-1}(r)X_{t-1}(s)u_t(r)u_t(s)drds \\ \cdot & \frac{1}{C^2}\mathbb{E}\left(\int_b^c X_{t-1}(r)u_t(r)dr\right)^2 \end{bmatrix}. \tag{63}$$

Under the null with $\theta_1 = \theta_2$, we have $\mathbb{E}\left(\int_a^b X_{t-1}(r)u_t(r)dr\right)^2 = \frac{\sigma^4}{1-\theta^2}\int_a^b \int_a^b k(r,s)^2 dsdr$, as in (52), and by similar calculations

$$\int_a^b \int_b^c \mathbb{E}X_{t-1}(r)X_{t-1}(s)u_t(r)u_t(s)drds = \frac{\sigma^4}{1-\theta^2}\int_a^b \int_b^c k(r,s)^2 dsdr. \tag{64}$$

The stability and Lindeberg conditions hold as before and we have the following limit theory under the null

$$\begin{aligned} \begin{bmatrix} \sqrt{n}\left(\widehat{\theta}_1 - \theta_1\right) \\ \sqrt{n}\left(\widehat{\theta}_2 - \theta_2\right) \end{bmatrix} &\rightsquigarrow \mathcal{N}\left(0, \frac{\sigma^4}{1-\theta^2}\begin{bmatrix} \frac{1}{A^2}\int_a^b \int_a^b k(r,s)^2 dsdr & \frac{1}{AC}\int_a^b \int_b^c k(r,s)^2 dsdr \\ \frac{1}{AC}\int_a^b \int_b^c k(r,s)^2 dsdr & \frac{1}{C^2}\int_b^c \int_b^c k(r,s)^2 dsdr \end{bmatrix}\right) \\ &= \mathcal{N}\left(0, (1-\theta^2)\begin{bmatrix} \frac{\int_a^b \int_a^b k(r,s)^2 dsdr}{\left(\int_a^b k(r,r)dr\right)^2} & \frac{\int_a^b \int_b^c k(r,s)^2 dsdr}{\left(\int_a^b k(r,r)dr\right)\left(\int_b^c k(r,r)dr\right)} \\ \frac{\int_a^b \int_b^c k(r,s)^2 dsdr}{\left(\int_a^b k(r,r)dr\right)\left(\int_b^c k(r,r)dr\right)} & \frac{\int_b^c \int_b^c k(r,s)^2 dsdr}{\left(\int_b^c k(r,r)dr\right)^2} \end{bmatrix}\right), \end{aligned} \tag{65}$$

giving $\sqrt{n}\left(\widehat{\theta}_1 - \hat{\theta}_2\right) \rightsquigarrow \mathcal{N}\left(0, (1-\theta^2)V_{\theta_1,\theta_2}\right)$ where V_{θ_1,θ_2} is defined in (33). The limit distribution of the Wald statistic $\mathcal{W}_n$ follows directly given consistent estimates of the covariance kernel $k_u(r,s)$ in the formation of the variance estimate $\hat{V}_{\theta_1,\theta_2}$ in $\mathcal{W}_n$.

Part(ii) Under $\mathcal{H}_1$, $\theta_2 = \theta_1 + \frac{\psi}{\sqrt{n}}$ and $\sqrt{n}\left(\widehat{\theta}_1 - \widehat{\theta}_2\right) = \sqrt{n}\left(\widehat{\theta}_1 - \theta_1\right) - \sqrt{n}\left(\widehat{\theta}_2 - \theta_2\right) - \psi$. Equation (62) still holds but we now have $\mathbb{E}\left(\int_a^b X_{t-1}(r)u_t(r)dr\right)^2 = \frac{\sigma^4}{1-\theta_1^2}\int_a^b\int_a^b k(r,s)^2 dsdr$, $\mathbb{E}\left(\int_b^c X_{t-1}(r)u_t(r)dr\right)^2 = \frac{\sigma^4}{1-\theta_2^2}\int_b^c\int_b^c k(r,s)^2 dsdr$, and

$$\begin{aligned}\int_a^b\int_b^c \mathbb{E}X_{t-1}(r)X_{t-1}(s)u_t(r)u_t(s)drds &= \frac{\sigma^4}{(1-\theta_1\theta_2)}\int_a^b\int_b^c k(r,s)^2 dsdr \\ &= \frac{\sigma^4}{(1-\theta_1^2)}\int_a^b\int_b^c k(r,s)^2 dsdr + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{66}$$

Using these results and proceeding in the same way as in Part(i) with martingale central limit theory we find that

$$\begin{bmatrix}\sqrt{n}\left(\widehat{\theta}_1-\theta_1\right)\\ \sqrt{n}\left(\widehat{\theta}_2-\theta_2\right)\end{bmatrix} \rightsquigarrow \mathcal{N}\left(0,(1-\theta_1^2)\begin{bmatrix}\frac{\int_a^b\int_a^b k(r,s)^2dsdr}{\left(\int_a^b k(r,r)dr\right)^2} & \frac{\int_a^b\int_b^c k(r,s)^2dsdr}{\left(\int_a^b k(r,r)dr\right)\left(\int_b^c k(r,r)dr\right)}\\ \frac{\int_a^b\int_b^c k(r,s)^2dsdr}{\left(\int_a^b k(r,r)dr\right)\left(\int_b^c k(r,r)dr\right)} & \frac{\int_b^c\int_b^c k(r,s)^2dsdr}{\left(\int_b^c k(r,r)dr\right)^2}\end{bmatrix}\right). \tag{67}$$

Hence $\left[\sqrt{n}\left(\widehat{\theta}_1-\theta_1\right)-\sqrt{n}\left(\widehat{\theta}_2-\theta_2\right)\right] \rightsquigarrow \mathcal{N}(0,V_{1,2})$ and $\sqrt{n}\left(\widehat{\theta}_1-\hat{\theta}_2\right) \rightsquigarrow \mathcal{N}(-\psi,V_{1,2})$, where $V_{1,2} = (1-\theta_1^2)V_{\theta_1,\theta_2}$. It follows that the Wald statistic $\mathcal{W}_n$ has the following noncentral χ^2 limit theory

$$\mathcal{W}_n = \left[\sqrt{n}\left(\widehat{\theta}_1-\hat{\theta}_2\right)\right]\hat{V}_{1,2}^{-1}\left[\sqrt{n}\left(\widehat{\theta}_1-\hat{\theta}_2\right)\right] \rightsquigarrow \chi_1^2(\delta),$$

where $\delta = \psi^2 V_{1,2}^{-1}$ and $\hat{V}_{1,2}$ is a consistent estimate of $V_{1,2}$..

Part(iii) Next consider the local functional alternative in which $\mathcal{H}_2$, $\theta_2 = \theta_1 + \frac{\psi(r)}{\sqrt{n}} =: \theta_2(r)$ for some continuously differentiable function $\psi(r)$ over $r \in [b,c]$. In this case we have

$$\begin{aligned}\mathbb{E}\left(\int_b^c X_{t-1}(r)u_t(r)dr\right)^2 &= \sigma^4\int_b^c\int_b^c \frac{k(r,s)^2dsdr}{(1-\theta_2(s))(1-\theta_2(r))} \\ &= \frac{\sigma^4}{1-\theta_1^2}\int_b^c\int_b^c k(r,s)^2dsdr + O_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{68}$$

and similarly

$$\begin{aligned}\int_a^b\int_b^c \mathbb{E}X_{t-1}(r)X_{t-1}(s)u_t(r)u_t(s)drds &= \sigma^4\int_a^b\int_b^c \frac{k(r,s)^2dsdr}{1-\theta_1(\theta_1+\psi(s)/\sqrt{n})} \\ &= \frac{\sigma^4}{1-\theta_1^2}\int_a^b\int_b^c k(r,s)^2dsdr + O\left(\frac{1}{\sqrt{n}}\right).\end{aligned} \tag{69}$$

Next note that under $\mathcal{H}_2$, $\theta_2 = \theta_1 + \frac{\psi(r)}{\sqrt{n}}$, we have

$$\widehat{\theta}_2 = \frac{\int_b^c \frac{1}{n}\sum_{t=1}^n X_{t-1}(r)X_t(r)dr}{\int_b^c \frac{1}{n}\sum_{t=1}^n X_{t-1}^2(r)dr} = \frac{\int_b^c \frac{1}{n}\sum_{t=1}^n \left[(\theta_1 + \psi(r)/\sqrt{n})X_{t-1}^2(r)dr + X_{t-1}(r)u_t(r)\right]}{\int_b^c \frac{1}{n}\sum_{t=1}^n X_{t-1}^2(r)dr} \tag{70}$$

and so

$$\begin{aligned}\widehat{\theta}_2 - \theta_1 &= \frac{1}{\sqrt{n}}\frac{\int_b^c \frac{1}{n}\sum_{t=1}^n \left[\psi(r)X_{t-1}^2(r)dr + X_{t-1}(r)u_t(r)\right]}{\int_b^c \frac{1}{n}\sum_{t=1}^n X_{t-1}^2(r)dr} \\ &\sim_a \frac{1}{\sqrt{n}}\frac{\int_b^c \psi(r)K_X(r,r)dr}{\int_b^c K_X(r,r)dr} + \frac{1}{\sqrt{n}}\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_b^c X_{t-1}(r)u_t(r)}{\int_b^c K_X(r,r)dr},\end{aligned} \tag{71}$$

where $K_X(r,r) = \mathbb{E}X_t^2(r) = \frac{k_u(r,r)}{1-(\theta_1+\psi(r)/\sqrt{n})^2} = \frac{k_u(r,r)}{1-\theta_1^2} + O\left(\frac{1}{\sqrt{n}}\right)$. It follows that under $\mathcal{H}_2$

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}_2 - \theta_1\right) &\sim_a \frac{\int_b^c \psi(r)K_X(r,r)dr}{\int_b^c K_X(r,r)dr} + \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_b^c X_{t-1}(r)u_t(r)}{\int_b^c K_X(r,r)dr} \\ &\sim_a \mathcal{N}\left(\frac{\int_b^c \psi(r)k_u(r,r)dr}{\int_b^c k_u(r,r)dr}, (1-\theta_1^2)\frac{\int_b^c\int_b^c k_u(r,s)^2 dsdr}{\left(\int_b^c k_u(r,r)dr\right)^2}\right),\end{aligned} \tag{72}$$

and in place of (67) we have

$$\begin{bmatrix}\sqrt{n}\left(\widehat{\theta}_1 - \theta_1\right) \\ \sqrt{n}\left(\widehat{\theta}_2 - \theta_1\right)\end{bmatrix} \rightsquigarrow \mathcal{N}\left(\begin{bmatrix}0 \\ \frac{\int_b^c \psi(r)k_u(r,r)dr}{\int_b^c k_u(r,r)dr}\end{bmatrix}, (1-\theta_1^2)\begin{bmatrix}\frac{\int_a^b\int_a^b k(r,s)^2dsdr}{\left(\int_a^b k(r,r)dr\right)^2} & \frac{\int_a^b\int_b^c k(r,s)^2dsdr}{\left(\int_a^b k(r,r)dr\right)\left(\int_b^c k(r,r)dr\right)} \\ \cdot & \frac{\int_b^c\int_b^c k(r,s)^2dsdr}{\left(\int_b^c k(r,r)dr\right)^2}\end{bmatrix}\right). \tag{73}$$

Then, $\sqrt{n}\left(\widehat{\theta}_1 - \hat{\theta}_2\right) = \left(\widehat{\theta}_1 - \theta_1\right) - \left(\widehat{\theta}_2 - \theta_1\right)$ and using (73) we have

$$\sqrt{n}\left(\widehat{\theta}_1 - \hat{\theta}_2\right) \rightsquigarrow \mathcal{N}\left(-\Psi, V_{1,2}\right), \text{ with } \Psi = \frac{\int_b^c \psi(r)k_u(r,r)dr}{\int_b^c k_u(r,r)dr}, \tag{74}$$

where again $V_{1,2} = (1-\theta_1^2)V_{\theta_1,\theta_2}$. It follows that under the alternative $\mathcal{H}_2$ the Wald statistic $\mathcal{W}_n$ has the following noncentral χ^2 limit theory

$$\mathcal{W}_n = \left[\sqrt{n}\left(\widehat{\theta}_1 - \hat{\theta}_2\right)\right]\hat{V}_{1,2}^{-1}\left[\sqrt{n}\left(\widehat{\theta}_1 - \hat{\theta}_2\right)\right] \rightsquigarrow \chi_1^2(\delta_\Psi),$$

where $\delta_\Psi = \Psi^2 V_{1,2}^{-1}$ and $\hat{V}_{1,2}$ is, as before, a consistent estimate of $V_{1,2}$..

Proof of Theorem 5 The proof follows as in the proof of Theorem 1, using a multivariate MGCLT to establish $\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_a^b u_t \otimes X_{t-1} \rightsquigarrow \mathcal{N}\left(0, \int_a^b\int_a^b \left(K_u(r,s)\otimes K_X(r,s)\right)drds\right)$ with

$\frac{1}{n}\sum_{t=1}^{n}\int_a^b X_{t-1}X_{t-1}' \to_{a.s} \int_a^b K_X(r,r)dr$ holding by ergodicity for the measurable real valued matrix $\int_a^b X_{t-1}X_{t-1}$. Using row vectorization we then have

$$\sqrt{n}\left(\text{vec}(\widehat{A}-A)\right) = \left(I_p \otimes \frac{1}{n}\sum_{t=1}^{n}\int_a^b X_{t-1}X_{t-1}'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\int_a^b u_t \otimes X_{t-1}\right)$$
$$\rightsquigarrow \left(I_p \otimes \int_a^b K_X(r,r)dr\right)^{-1}\mathcal{N}\left(0, \int_a^b\int_a^b \left(K_u(r,s)\otimes K_X(r,s)\right)drds\right) = \mathcal{N}\left(0, V_{AA}\right),$$

with $V_{AA} = \Upsilon_X^{-1}\left(\int_a^b\int_a^b K_u(r,s)\otimes K_X(r,s)drds\right)\Upsilon_X^{-1}$, and $\Upsilon_X = I_m \otimes \int_a^b K_X(r,r)dr$, as stated in (36). Further, $K_X(r,s) = \sum_{j=0}^{\infty} A^j K_u(r,s)A^{j\prime}$, A has eigenvalues within the unit circle and $\left(\sum_{j=0}^{\infty} A^j \otimes A^j\right) = (I_m - A\otimes A)^{-1}$, so that $\text{vec}(K_X(r,s)) = (I_m - A\otimes A)^{-1}\,\text{vec}(K_u(r,s))$. ■

Proof of Corollary 1 In view of the companion form representation of (37) the proof can be deduced from Theorem 5. We need to obtain the explicit form of the variance of the limiting normal distribution. Since the first row of A in (37) is the coefficient vector θ', the asymptotic variance matrix of $\widehat{\theta}$ has a simpler form without Kronecker products in the ARH(p) case. First, note that the average variance matrix $\Sigma_{u,ab} = \int_a^b \mathbb{E}u_t u_t'$ has the diagonal form $\text{diag}\{\int_a^b k_e(r,r)dr, 0, \ldots, 0\} =: \sigma_{e,ab}^2 E_p$, where $E_p = \text{diag}\{1, 0, \ldots, 0\}$ and $\sigma_{e,ab}^2 = \int_a^b k_e(r,r)dr$. Similarly, the average variance matrix of $Y_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$ is

$$\Sigma_{Y,ab} := \int_a^b K_Y(r,r)dr = \int_a^b \mathbb{E}Y_tY_t' = \sum_{j=0}^{\infty} A^j\Sigma_{u,ab}A^{j\prime} = \sigma_{e,ab}^2\sum_{j=0}^{\infty}A^jE_pA^{j\prime}.$$

Then, specializing the component factors of the asymptotic covariance matrix V_{AA} in Theorem 5 to the present case we have $\Upsilon_Y = \int_a^b K_Y(r,r)dr = \int_a^b k_e(r,r)dr\sum_{j=0}^{\infty}A^jE_pA^{j\prime}$ and $K_Y(r,s) = k_e(r,s)\sum_{j=0}^{\infty}A^jE_pA^{j\prime}$. It follows that the asymptotic variance of $\widehat{\theta}$ is

$$\Upsilon_Y^{-1}\left(\int_a^b\int_a^b k_e(r,s)K_Y(r,s)drds\right)\Upsilon_Y^{-1}$$
$$= \frac{1}{(\int_a^b k_e(r,r)dr)^2}\left(\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)^{-1}\left(\int_a^b\int_a^b k_e(r,s)K_Y(r,s)drds\right)\left(\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)^{-1}$$
$$= \frac{1}{(\int_a^b k_e(r,r)dr)^2}\left(\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)^{-1}\left(\int_a^b\int_a^b k_e(r,s)^2drds\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)\left(\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)^{-1}$$
$$= \rho_e^2\left(\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)^{-1} =: \rho_e^2V_\theta \text{ with } \rho_e^2 = \frac{\int_a^b\int_a^b k_e(r,s)^2}{(\int_a^b k_e(r,r)dr)^2},\quad V_\theta = \left(\sum_{j=0}^{\infty}A^jE_pA^{j\prime}\right)^{-1}. \quad (75)$$

From Theorem 5 we deduce that $\sqrt{n}(\widehat{\theta}-\theta) \rightsquigarrow \mathcal{N}(0, \rho_e^2V_\theta)$, giving the stated result. ■

Proof of Theorem 6 Write $X_t = \mathbb{E}X_t + X_t^0$ with $\mathbb{E}X_t = \frac{\alpha}{1-\theta} \in L_2[a,b]$ and $X_t^0 = \theta X_{t-1}^0 + u_t$. Taking deviations from time series averages gives $\widetilde{X}_t := X_t - \overline{X} =: \widetilde{X}_t^0 = X_t^0 - \overline{X}^0$. By ergodicity $\overline{X}^0 = \frac{1}{n}\sum_{t=1}^n X_t^0 \to_{a.s.} \mathbb{E}X_t^0 = 0$. Then $\overline{X} = \frac{1}{n}\sum_{t=1}^n X_t \to_{a.s} \mathbb{E}X_t = \frac{\alpha}{1-\theta}$ and, noting that $\widetilde{X}_{t-1} = X_{t-1}^0 - \overline{X}_{-1}^0 = X_{t-1}^0 + o_{a.s}(1)$, we have

$$\widetilde{\theta} = \frac{\sum_{t=1}^n \int_a^b \widetilde{X}_t \widetilde{X}_{t-1}}{\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^2} = \theta + \frac{\frac{1}{n}\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^0 u_t}{\frac{1}{n}\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^2} \overset{a.s.}{\to} \theta, \tag{76}$$

$$\widetilde{\alpha} = \overline{X} - \widetilde{\theta}\overline{X}_{-1} = \alpha + \overline{u} - (\widetilde{\theta} - \theta)\overline{X}_{-1} \overset{a.s.}{\to} \alpha, \tag{77}$$

giving consistency. Next

$$\sqrt{n}(\widetilde{\theta} - \theta) = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^0 u_t}{\frac{1}{n}\sum_{t=1}^n \int_a^b (\widetilde{X}_{t-1}^0)^2} = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^0 u_t}{A + \frac{1}{n}\sum_{t=1}^n \left(\int_a^b (\widetilde{X}_{t-1}^0)^2 - A\right)}, \tag{78}$$

where $A_n := \frac{1}{n}\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^0(r)^2 dr \overset{a.s.}{\to} \int_a^b \mathbb{E}X_{t-1}^0(r)^2 dr = \frac{\int_a^b k_u(r,r)dr}{1-\theta^2} =: A$. Setting $\xi_t^0 = \int_a^b X_{t-1}^0(r) u_t(r) dr$, the same martingale limit theory used in Theorem 1 establishes that

$$\sqrt{n}\left(\widetilde{\theta} - \theta\right) = \frac{1}{A\sqrt{n}}\sum_{t=1}^n \int_a^b \widetilde{X}_{t-1}^0(r) u_t(r) dr + O_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{A\sqrt{n}}\sum_{t=1}^n \xi_t^0 + O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$\rightsquigarrow \mathcal{N}\left(0, (1-\theta^2)\frac{\int_a^b\int_a^b k_u(r,q)^2 dr dq}{(\int_a^b k_u(r,r)dr)^2}\right) = \mathcal{N}\left(0, (1-\theta^2)\rho_u^2\right), \tag{79}$$

so that $\sqrt{n}\left(\widetilde{\theta} - \theta\right) \rightsquigarrow \mathcal{N}(0, V_\theta)$, with $V_\theta = (1-\theta^2)\rho_u^2$, giving (45). To establish the limit theory for $\widetilde{\alpha}$ in (46), first note from (77) and $\overline{X}_{-1} \to_{a.s.} \mathbb{E}X_t^0$ that

$$\sqrt{n}\left(\widetilde{\alpha} - \alpha\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^n u_t - \sqrt{n}(\widetilde{\theta} - \theta)\overline{X}_{-1} = \frac{1}{\sqrt{n}}\sum_{t=1}^n u_t - \sqrt{n}(\widetilde{\theta} - \theta)\left(\frac{\alpha}{1-\theta}\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^n u_t - \frac{1}{A\sqrt{n}}\sum_{t=1}^n \xi_t^0 \left(\frac{\alpha}{1-\theta}\right) + o_p(1), \tag{80}$$

wherein the two sequences $\{u_t\}$ and $\{\xi_t^0\}$ are uncorrelated because u_t is a martingale difference and $\mathbb{E}u_s\xi_t^0 = 0$ for all $\{s,t\}$. The next step requires use of a Hilbert space central limit theorem (CLT) for $U_n := \frac{1}{\sqrt{n}}\sum_{t=1}^n u_t$. This result follows directly from Bosq (2000, Theorem 2.7) for the case of iid $\{u_t\}$, giving $U_n \rightsquigarrow \mathcal{G}(0, k_u(r,s))$ and the same result also holds for martingale differences $\{u_t\}$. In particular, the MGCLT delivers weak convergence of the finite dimensional distributions $\langle x, U_n\rangle = \frac{1}{\sqrt{n}}\sum_{t=1}^n \int_a^b x(r) u_t(r) dr \rightsquigarrow \mathcal{N}(0, \int_a^b\int_a^b x(r) k_u(r,s) x(s) dr ds)$ for all $x \in L_2[a,b]$, and tightness follows as in the proof of Bosq (2000, Theorem 2.7). Then, since U_n and $\xi_n := \frac{1}{A\sqrt{n}}\sum_{t=1}^n \xi_t^0$ are uncorrelated and asymptotically independent, the components of (80) satisfy a CLT in $L_2[a,b]$ and we have the limit theory[9]

$$\sqrt{n}\left(\widetilde{\alpha}-\alpha\right) \rightsquigarrow \mathcal{G}\left(0, k_u(r,s)+\frac{\alpha(r)\alpha(s)}{(1-\theta)^2}V_\theta\right), \tag{81}$$

giving the stated result (46).

Next, using the limit theory for $\sqrt{n}\left(\widetilde{\theta}-\theta\right)$ in (45), we find in the same way as (61) that $\widetilde{t}_{\theta,fe} \rightsquigarrow \mathcal{N}(0,1)$. Similar derivations show that $\hat{t}_{\theta,fe}, \widetilde{t}^*_{\theta,fe} \rightsquigarrow \mathcal{N}(0,1)$, which establishes (47).

Finally, using the representation (78) again, together with the Edgeworth expansion in Phillips (2025, Theorem 2) and zero third order moments of u_t, we have the stated bias expansion $\mathbb{E}\left(\widetilde{\theta}-\theta\right) = -\frac{1+\theta}{n} - \frac{2\theta\rho_u^2}{n} + o\left(\frac{1}{n}\right)$ in (49). ■

8 Appendix B

8.1 Limit theory under local cross section dependence

Here we have a triangular array model defined by $X_{t,N}(r) = \theta X_{t-1,N}(r) + u_{t,N}(r), \quad r \in [a,b]$, where $u_{t,N}(r)$ has covariance kernel

$$k_{u,N}(r,s) = \mathbb{E}\left[u_{t,N}(r)u_{t,N}(s)\right] = \begin{cases} \sigma^2 N & |r-s| \le \frac{1}{2N} \\ 0 & |r-s| > \frac{1}{2N} \end{cases},$$

in which the random function $u_{t,N}(r)$ is specified to be weakly dependent across ordinates r with non-zero covariance only within the (shrinking as $N \to \infty$) narrow band $|r-s| \le \frac{1}{2N}$. This specification emulates a panel model such as (10) in which the error process u_{it} is *iid* over both i and t. We have for the denominator

$$\int_a^b \frac{1}{n}\sum_{t=1}^n X_{t-1,N}^2(r)dr \sim \int_a^b \mathbb{E}\left[X_{t-1,N}^2(r)\right]dr = \frac{\sigma^2}{1-\theta^2}\int_a^b k_{\varepsilon,N}(r,r)dr \sim \frac{\sigma^2 N}{1-\theta^2}(b-a).$$

Then $\frac{1}{N}\int_a^b \frac{1}{n}\sum_{t=1}^n X_{t-1}^2(r)dr \sim \frac{\sigma^2}{1-\theta^2}(b-a)$. The martingale CLT also needs modification. In particular, the stability condition is now

$$\left\langle \frac{1}{\sqrt{n}}\sum_{t=1}^n \int_a^b X_{t-1,N}(r)u_t(r)dr \right\rangle = \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left[\left(\int_a^b X_{t-1,N}(r)u_t(r)dr\right)^2 |\mathcal{F}_{t-1}\right]$$

[9] The two random cmponents of (80) are $(\xi_n, U_n) := (\frac{1}{A\sqrt{n}}\sum_{t=1}^n \xi_t^0, \frac{1}{\sqrt{n}}\sum_{t=1}^n u_t)$ in the space $\mathbb{R}\times L_2[a,b]$ and these components are asymptotically independent Gaussian processes, which ensures the joint weak convergence $(\xi_n, U_n) \rightsquigarrow (\xi, U)$ in $\mathbb{R}\times L_2[a,b]$, with $\xi \sim \mathcal{N}(0,\sigma_\xi^2)$ in $\mathbb{R}$, where $\sigma_\xi^2 = (1-\theta^2)\rho_u^2 = V_\theta$, and $U \sim \mathcal{G}(0,k_u)$ in $L_2[a,b]$. Then $Z_n := U_n + \frac{\xi_n}{1-\theta}\alpha \rightsquigarrow Z := U + \frac{\xi}{1-\theta}\alpha$, so that the limit process Z is a Gaussian process in $L_2[a,b]$ space with zero mean and covariance operator $C_Z = C_U + \frac{\sigma_\xi^2}{(1-\theta)^2}\alpha\otimes\alpha$, where C_U is the covariance operator of U and $\alpha\otimes\alpha$ is the rank-one operator on $L_2[a,b]$ for which $(\alpha\otimes\alpha)f = (\int_a^b f(s)\alpha(s)ds)\alpha$ for any $f \in L_2[a,b]$. The pointwise form of the covariance operator C_Z is then $c_z(r,s) = k_u(r,s) + \frac{\sigma_\xi^2}{(1-\theta)^2}\alpha(r)\alpha(s) = k_u(r,s) + \frac{\alpha(r)\alpha(s)}{(1-\theta)^2}V_\theta$, as given in (81) above and (46). This pointwise representation exists because α is a càdlàg function in $L_2[a,b]$ space and therefore has a well-defined pointwise representation in $L_2[a,b]$.

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n}\int_a^b\int_a^b X_{t-1,N}(r)X_{t-1,N}(s)\mathbb{E}\left[u_t(r)u_t(s)|\mathcal{F}_{t-1}\right]dsdr \\
&= \frac{1}{n}\sum_{t=1}^{n}\int_{a+\frac{1}{N}}^{b-\frac{1}{N}} X_{t-1,N}(r)\int_{r-\frac{1}{2N}}^{r+\frac{1}{2N}} X_{t-1,N}(s)\mathbb{E}\left[u_t(r)u_t(s)\right]dsdr \\
&= \frac{\sigma^2}{n}\sum_{t=1}^{n}\int_{a+\frac{1}{N}}^{b-\frac{1}{N}} X_{t-1,N}(r)\int_{r-\frac{1}{2N}}^{r+\frac{1}{2N}} X_{t-1,N}(s)dsdr\times N \\
&\sim \sigma^2\int_{a+\frac{1}{N}}^{b-\frac{1}{N}}\frac{1}{n}\sum_{t=1}^{n}X_{t-1,N}(r)^2dr\times\frac{1}{N}N \\
&\sim \sigma^2\int_a^b \mathbb{E}X_{t-1}(r)^2dr = \frac{\sigma^4}{1-\theta^2}\int_a^b k_{\varepsilon N}\left(r,r\right)dr = \frac{\sigma^4}{1-\theta^2}N\left(b-a\right),
\end{aligned}
$$

based on equivalence of variation, namely:

$$
k_{u,N}(r,s) = \mathbb{E}\left[u_{t,N}(r)u_{t,N}(s)\right] = \begin{cases} \sigma^2 N & |r-s|\le\frac{1}{2N} \\ 0 & |r-s|>\frac{1}{2N} \end{cases},
$$

so that overall variation

$$
\begin{aligned}
\mathbb{E}\left[\int_a^b u_{t,N}(r)dr\right]^2 &= \int_a^b\int_a^b \mathbb{E}\left[u_t(r)u_t(s)\right]dsdr = \int_{a+\frac{1}{N}}^{b-\frac{1}{N}}\int_{r-\frac{1}{2N}}^{r+\frac{1}{2N}}\mathbb{E}\left[u_t(r)u_t(s)\right]dsdr \\
&= \int_{a+\frac{1}{N}}^{b-\frac{1}{N}}\frac{\sigma^2 N}{N}dr \sim \sigma^2\left(b-a\right).
\end{aligned}
$$

This ensures that $\mathbb{E}\left[\int_a^b u_{t,N}(r)dr\right]^2$ has the same asymptotic value as $N\to\infty$ as when $\mathbb{E}\left[u_t(r)u_t(s)\right]=\sigma^2$ uniformly for all (r,s), noting that if $\mathbb{E}\left[u_t(r)u_t(s)\right]=\sigma^2$ uniformly for all (r,s) we would have $\mathbb{E}\left[\int_a^b u_t(r)dr\right]^2=\int_a^b\int_a^b\mathbb{E}\left[u_t(r)u_t(s)\right]dsdr=\sigma^2\left(b-a\right)^2$ and

$$
\widehat{\theta}-\theta = \frac{\frac{1}{nN}\sum_{t=1}^{n}\int_a^b X_{t-1}(r)u_t(r)dr}{\frac{1}{nN}\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)dr} \sim_{a.s.} \frac{\frac{1}{N}\int_a^b\mathbb{E}\left[X_{t-1}(r)u_t(r)\right]dr}{\frac{1}{N}\int_a^b\mathbb{E}[X_{t-1}^2(r)]dr} = 0,
$$

so that $\widehat{\theta}$ is consistent for θ. The limit distribution follows in a straightforward way using the martingale central limit theorem, giving

$$
\sqrt{nN}(\widehat{\theta}-\theta) = \frac{\frac{1}{\sqrt{nN}}\sum_{t=1}^{n}\int_a^b X_{t-1}(r)u_t(r)dr}{\frac{1}{nN}\sum_{t=1}^{n}\int_a^b X_{t-1}^2(r)dr} \rightsquigarrow \frac{\mathcal{N}\left(0,\frac{\sigma^4}{1-\theta^2}\left(b-a\right)\right)}{\frac{\sigma^2}{1-\theta^2}\left(b-a\right)} = \mathcal{N}\left(0,\frac{1-\theta^2}{b-a}\right).
$$

and the result in (13).

Now compare the following cases. First, if $k(r,s)=1$ for all (r,s) we have

$$\sqrt{n}(\widehat{\theta}-\theta)\rightsquigarrow\mathcal{N}\left(0,\left(1-\theta^2\right)\frac{\int_a^b\int_a^b k(r,s)^2dsdr}{\left(\int_a^b k(r,r)dr\right)^2}\right)=\mathcal{N}\left(0,\left(1-\theta^2\right)\frac{(b-a)^2}{(b-a)^2}\right)=\mathcal{N}\left(0,\left(1-\theta^2\right)\right). \tag{82}$$

Next, if $k(r,s)=1\left\{|r-s|\leq\frac{1}{2N}\right\}$. Then $\int_a^b\int_a^b k(r,s)^2dsdr=\int_a^b\int_a^b 1\left\{|r-s|\leq\frac{1}{2N}\right\}dsdr=\frac{b-a}{N}-\frac{1}{4N^2}$, and $\left(\int_a^b k(r,r)dr\right)^2=\left(\int_a^b dr\right)^2=(b-a)^2$. Hence $\sqrt{n}(\widehat{\theta}-\theta)\rightsquigarrow\mathcal{N}\left(0,\frac{(1-\theta^2)}{N(b-a)}\right)$, giving

$$\sqrt{nN}(\widehat{\theta}-\theta)\rightsquigarrow\mathcal{N}\left(0,\frac{1-\theta^2}{b-a}\right) \tag{83}$$

The intuition should be clear. When $b-a\to 0$, then (82) still holds because the model simply corresponds to the standard AR(1) case where $a=b=r$ and we still get $\sqrt{n}$ convergence. However in the case of (83) $\frac{(1-\theta^2)}{(b-a)}\to\infty$ and the limit theory fails because there is no longer extra information in the data as $N\to\infty$ from independence in the errors over $[a,b]$. So in this case the scaling $\sqrt{nN}$ is excessive and leads to divergence. On the other hand, when $b-a\to\infty$, $\sqrt{nN}(\widehat{\theta}-\theta)\to_p 0$ because we have independent errors asymptotically over an infinite region in $L_2(\mathbb{R})$ space. Note that there is no gain in (82) in that case because the errors $u_t(r)$ are now remain correlated over the wider cross section domain $\mathbb{R}$ as $b-a\to\infty$.

8.2 Further error process illustrations

Two further examples are included below to illustrate the impact of different error processes $u_t(r)$ on the limit distribution and variance of the OLS estimator $\widehat{\theta}$. These examples allow for segmentation in the cross section domain allowing for local correlation and heterogeneity. A common domain for $r\in[0,1]$ is assumed within which K individual segments $A_k=[\frac{k-1}{K},\frac{k}{K})$ are included.

Segmented Brownian motion errors Define

$$u_t(r)=\sigma\sum_{k=1}^{K}\left(W_{t+r}^{(k)}-W_{t+\frac{k-1}{K}}^{(k)}\right)\times\mathbf{1}\left(r\in A_k\right), \tag{84}$$

where $\left\{W^{(1)},\ldots,W^{(K)}\right\}$ are K independent standard Brownian motions. In this case $\{u_t(r)\}_{t=1}^{\infty}$ is a time series sequence of independent processes comprised of the spatial superposition (over $r\in[0,1]$) of K independent standard Brownian motion segments each originating at the origin. The covariance kernel is similarly given by the superposition of the covariance kernels of these

Brownian motions within those segments and zero elsewhere, i.e.,

$$k_u(r,s) = \sigma^2 \sum_{k=1}^{K} \left(r - \frac{k-1}{K} \right) \wedge \left(s - \frac{k-1}{K} \right) \times \mathbf{1}\left(r, s \in A_k \right).$$

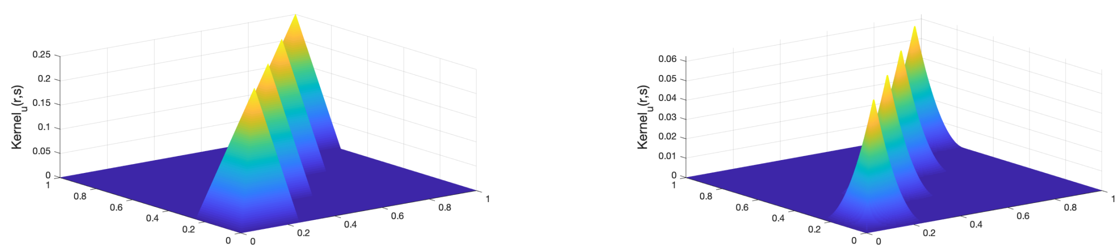
Panel (a) in Figure 9 draws this covariance kernel when $K = 4$. The variance of the limit distribution of $\widehat{\theta}$ is

$$\left(1-\theta^2\right) \frac{\int_0^1 \int_0^1 k_\varepsilon(r,s)^2 ds dr}{\left(\int_0^1 k(r,r) dr \right)^2} = \left(1-\theta^2\right) \frac{K \int_0^{1/K} \int_0^{1/K} (r \wedge s)^2 dr ds}{\left(K \int_0^{1/K} r dr \right)^2} = \frac{2}{3K}\left(1-\theta^2\right), \tag{85}$$

giving $\sqrt{n}\left(\widehat{\theta}-\theta\right) \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, \frac{2}{3K}\left(1-\theta^2\right)\right)$. Then, as the number of independent segments increase we obtain the sequential limit

$$\sqrt{nK}\left(\widehat{\theta}-\theta\right) \underset{(K,n)_{\text{seq}}\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, \frac{2}{3}\left(1-\theta^2\right)\right), \tag{86}$$

which gives the analogue of (13) for a segmented Brownian motion error process. In this case the interval $[0,1]$ is subdivided into the K sectors $\{A_k\}_{k=1}^K$ and the error process $u_t(r)$ is independent and identically distributed across these sectors (rather than uniform over a shrinking interval around $r = s$, as in (12)). As the number of sectors $K \to \infty$, the variance of the limit distribution of $\sqrt{n}\left(\widehat{\theta}-\theta\right)$ declines at the rate $O\left(K^{-1}\right)$. In sequential convergence the convergence rate rises to $\sqrt{nK}$ with limit (86). Although it is not explored rigorously here, we expect the same result to apply in joint convergence as $(K,n) \to \infty$ jointly, using the methods of Phillips and Moon (1999).



(a) Segmented Brownian Motion

(b) Segmented Brownian Bridge

Figure 9: The covariance kernel $k_u(r,s)$ for the segmented Brownian motion/bridge errors when the number of the segments $K = 4$.

Segmented Brownian bridge errors Define

$$u_t(r) = \sigma \sum_{k=1}^{K} \left\{ \left(W_{t+r}^{(k)} - W_{t+\frac{k-1}{K}}^{(k)} \right) - K\left(r - \frac{k-1}{K} \right) \left(W_{t+\frac{k}{K}}^{(k)} - W_{t+\frac{k-1}{K}}^{(k)} \right) \right\} \times \mathbf{1}\left(r \in A_k \right),$$

where $A_k = [\frac{k-1}{K}, \frac{k}{K})$ and $\{W^{(1)}, \ldots, W^{(K)}\}$ are K independent standard Brownian motions. With this specification $\{u_t(r)\}_{t=1}^{\infty}$ is a time series sequence of independent processes comprised of the superposition of K independent standard Brownian bridge segments of length K^{-1}. The covariance kernel is

$$k_u(r,s) = \sigma^2 \left\{ \left(r - \frac{k-1}{K}\right) \wedge \left(s - \frac{k-1}{K}\right) - K\left(r - \frac{k-1}{K}\right)\left(s - \frac{k-1}{K}\right) \right\} \times \mathbf{1}\left(r, s \in A_k\right).$$

Panel (b) in Figure (9) draws this covariance kernel when $K = 4$ and the variance of the limit distribution of $\widehat{\theta}$ is

$$\left(1-\theta^2\right) \frac{\int_0^1 \int_0^1 k_\varepsilon(r,s)^2 ds dr}{\left(\int_0^1 k(r,r) dr\right)^2} = (1-\theta^2) \frac{K \int_0^{1/K} \int_0^{1/K} \left\{(r \wedge s) - Krs\right\}^2 dr ds}{\left(K \int_0^{1/K} r dr\right)^2} = \frac{2}{5K}(1-\theta^2), \quad (87)$$

giving $\sqrt{n}\left(\widehat{\theta} - \theta\right) \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}\left(0, \frac{2}{5K}\left(1-\theta^2\right)\right)$. It follows that under the same configuration as Example 2 the limit variance (87) of $\widehat{\theta}$ is smaller under Brownian bridge innovations than under Brownian motion innovations. The reason is that each Brownian bridge segment is tied down at the ends of the segment, thereby reducing the variation in the function space equation error, which correspondingly reduces the asymptotic variance of $\widehat{\theta}$. Note that in both these examples there is homogeneity in the covariance kernel across segments so that $\int_{\frac{k-1}{K}}^{\frac{k}{K}} k(r,r) dr = \int_0^{\frac{1}{K}} k(r,r) dr$ for all $k = 1, ..., K$, so that using (9) we have the inequality

$$\left(1-\theta^2\right) \frac{\int_0^1 \int_0^1 k_\varepsilon(r,s)^2 ds dr}{\left(\int_0^1 k(r,r) dr\right)^2} \le (1-\theta^2) \frac{K\left(\int_0^{\frac{1}{K}} k_\varepsilon(r,r) dr\right)^2}{K^2 \left(\int_0^{\frac{1}{K}} k_\varepsilon(r,r) dr\right)^2} = \frac{1-\theta^2}{K},$$

which bounds the asymptotic variance when $u_t(r)$ is composed of K independent segments, as it is in these two examples.

9 Appendix C: Wild-bootstrap size results

Tables 2-4 collect the wild-bootstrap size results that support the simulation findings in Section 4 for models without fitted functional fixed effects. Tables 2 and 3 first report empirical sizes for the three autoregressive coefficient tests under diffusion and segmented Brownian motion curve errors, matching the order of the corresponding local power figures. Table 4 then reports empirical sizes for the Wald cross section break specification test. Across these designs, the wild-bootstrap procedures generally deliver size close to the nominal 5% level; the main departures occur for $\hat{t}_\theta$ and $\widetilde{t}_\theta^*$ in the smallest samples, while size improves as the time series sample size increases.

Table 2: Empirical sizes for wild-bootstrap autoregressive coefficient tests of the null hypothesis $\mathcal{H}_0 : \theta = 0.5$ in stationary ARH(1) models with diffusion curve errors

c	20			60			100		
	$\hat{t}_\theta$	$\tilde{t}^*_\theta$	$\tilde{t}_\theta$	$\hat{t}_\theta$	$\tilde{t}^*_\theta$	$\tilde{t}_\theta$	$\hat{t}_\theta$	$\tilde{t}^*_\theta$	$\tilde{t}_\theta$
-5	0.1234	0.1232	0.0614	0.0878	0.0880	0.0560	0.0708	0.0712	0.0494
-3	0.1106	0.1068	0.0614	0.0810	0.0798	0.0576	0.0686	0.0684	0.0520
0	0.0888	0.0856	0.0662	0.0714	0.0710	0.0580	0.0614	0.0616	0.0556
2	0.0792	0.0758	0.0716	0.0660	0.0648	0.0578	0.0590	0.0594	0.0532

Table 3: Empirical sizes for wild-bootstrap autoregressive coefficient tests of the null hypothesis $\mathcal{H}_0 : \theta = 0.5$ in stationary ARH(1) models with segmented Brownian motion curve errors

K	20			60			100		
	$\hat{t}_\theta$	$\tilde{t}^*_\theta$	$\tilde{t}_\theta$	$\hat{t}_\theta$	$\tilde{t}^*_\theta$	$\tilde{t}_\theta$	$\hat{t}_\theta$	$\tilde{t}^*_\theta$	$\tilde{t}_\theta$
1	0.0908	0.0896	0.0640	0.0638	0.0644	0.0542	0.0618	0.0610	0.0538
2	0.0986	0.0962	0.0594	0.0662	0.0664	0.0492	0.0608	0.0620	0.0482
3	0.1128	0.1128	0.0610	0.0698	0.0702	0.0480	0.0682	0.0686	0.0544

Table 4: Empirical sizes for wild-bootstrap functional autoregression Wald specification tests of the null hypothesis $\mathcal{H}_0 : \theta_1 = \theta_2 = 0.7$ of a cross section break in a stationary ARH(1) with diffusion curve errors and diffusion coefficient c

$c \backslash n$	20	40	60	80	100
-5	0.0552	0.0536	0.0546	0.0572	0.0570
-3	0.0564	0.0546	0.0580	0.0584	0.0558
0	0.0606	0.0562	0.0552	0.0556	0.0588
2	0.0590	0.0592	0.0552	0.0540	0.0564

10 Appendix D: Functional fixed effects simulations

Table 5 shows wild-bootstrap simulation results for the tests based on $\widetilde{t}_{\theta,fe}$, $\hat{t}_{\theta,fe}$, and $\widetilde{t}^*_{\theta,fe}$ of the null hypothesis $\mathcal{H}_0 : \theta = 0.5$ in models with functional fixed effects and diffusion curve errors generated with coefficients $c \in \{-5, -3, 0, 2\}$. The tests mirror those in Section 4 but incorporate fitted functional fixed effects. The results show some oversizing in small time series samples, with size improving as n increases.

Figure 10 displays local power curves for the same t-ratio coefficient tests $\widetilde{t}_{\theta,fe}$, $\hat{t}_{\theta,fe}$, and $\widetilde{t}^*_{\theta,fe}$ applied to stationary curve time series generated by diffusion processes with $c \in \{-5, -3, 0, 2\}$. All statistics account for functional fixed effects, and critical values are obtained by wild bootstrap. The oversizing in small time series samples is evident in the graphs,

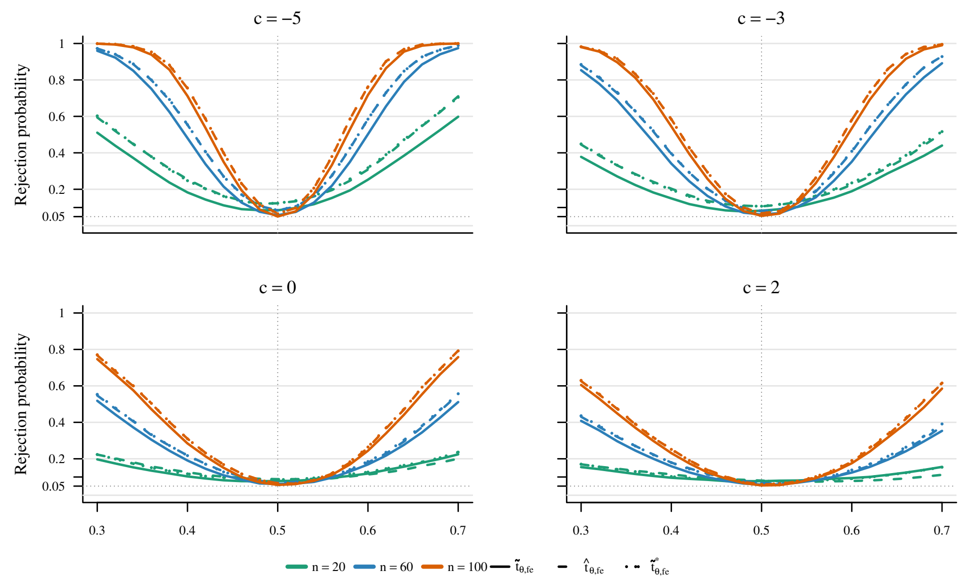


Figure 10: Local power curves of the wild-bootstrap autoregressive coefficient tests $\widetilde{t}_{\theta,fe}, \hat{t}_{\theta,fe}, \widetilde{t}^*_{\theta,fe}$ with stationary curve time series generated by diffusion processes with diffusion coefficient $c \in \{-5, -3, 0, 2\}$. All the test statistics are computed by taking into account functional fixed effects.

as in Table 5. The power curves also show that there is less power in the tests when $c \geq 0$.

Table 5: Empirical sizes for wild-bootstrap autoregressive coefficient tests of the null hypothesis $\mathcal{H}_0 : \theta = 0.5$ in stationary ARH(1) models with fitted functional fixed effects and diffusion curve errors

c	20			60			100		
	$\hat{t}_{\theta,fe}$	$\widetilde{t}^*_{\theta,fe}$	$\widetilde{t}_{\theta,fe}$	$\hat{t}_{\theta,fe}$	$\widetilde{t}^*_{\theta,fe}$	$\widetilde{t}_{\theta,fe}$	$\hat{t}_{\theta,fe}$	$\widetilde{t}^*_{\theta,fe}$	$\widetilde{t}_{\theta,fe}$
-5	0.1232	0.1232	0.0848	0.0844	0.0848	0.0576	0.0670	0.0660	0.0506
-3	0.1062	0.1072	0.0826	0.0812	0.0798	0.0616	0.0682	0.0672	0.0542
0	0.0882	0.0858	0.0778	0.0686	0.0688	0.0618	0.0656	0.0656	0.0566
2	0.0802	0.0766	0.0770	0.0666	0.0650	0.0584	0.0592	0.0592	0.0538

11 Appendix E: The Singapore Life Panel (SLP)

Singapore Management University founded the Centre for Research on the Economics of Ageing (CREA) in 2014 (now the Centre for Research on Successful Ageing, ROSA[10]) to study the economics of ageing. CREA then launched the Singapore Life Panel (SLP), a representative

[10] see https://rosa.smu.edu.sg/

group of citizens and permanent residents aged 50 to 70, using 25,000 addresses from the Department of Statistics. Between May and July 2015, invitation letters were sent, followed by in-person visits and phone calls, yielding 11,511 households (15,212 individuals) at a 52% response rate. A pilot survey with 1,000 participants began in August 2015, and the full survey started in September 2015, collecting monthly data on spending, earnings, health, and employment.

The SLP is a longitudinal study, meaning it retains its members throughout without adding new older participants for its monthly surveys. Consequently, the panel initially focused on individuals aged 50 to 70, and as time progresses, the age range shifts accordingly—for example, moving to 51-71-year-olds in the following year. The SLP maintains relatively stable participation rates, with around 60% of respondents completing surveys each month. For analytical purposes, we exclude any missing data on a monthly basis, applying nonparametric methods to estimate the curve using the available observations.

Focusing on two-adult households—where consumption is reported for the couple, thus excluding children expenditure—we winsorize the extreme 1% on both ends of the monthly nondurable expenditure distribution. This process yields about 5,000 households per month, totaling 418,525 household-month observations. Nondurable spending is divided into nine categories: clothing, education, food, health, housing, insurance, leisure, utilities, and transport. Table 6 provides descriptive statistics for the key variables.

Table 6: Descriptive statistics for the SLP data

	Means	Std. Dev.	10%	50%	90%
Clothing share	0.016	0.038	0.000	0.000	0.053
Education share	0.019	0.078	0.000	0.000	0.026
Food share	0.341	0.203	0.102	0.311	0.629
Health share	0.062	0.111	0.000	0.008	0.186
Housing share	0.065	0.201	0.000	0.005	0.194
Insurance share	0.026	0.069	0.000	0.000	0.091
Leisure share	0.050	0.118	0.000	0.000	0.143
Utility share	0.185	0.141	0.051	0.150	0.365
Transport share	0.141	0.156	0.015	0.088	0.348
Log tot. nondur. exp.	7.591	0.898	6.438	7.581	8.776
Total number of observations	418,525				

References

Abadir, K. M., K. Hadri, and E. Tzavalis (1999): "The influence of VAR dimensions on estimator biases," *Econometrica*, 67, 163–181. 6

Alvarez, J. and M. Arellano (2003): "The time series and cross-section asymptotics of dynamic panel data estimators," *Econometrica*, 71, 1121–1159. 25, 26, 27

Anderson, T. W. (1971): *The statistical analysis of time series*, John Wiley & Sons. 23

Anderson, T. W. and C. Hsiao (1981): "Estimation of dynamic models with error components," *Journal of the American statistical Association*, 76, 598–606. 26

Andrews, D. W. K. (1993): "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 821–856. 21

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 58, 277–297. 26

Banks, J., R. Blundell, and A. Lewbel (1997): "Quadratic Engel curves and consumer demand," *Review of Economics and statistics*, 79, 527–539. 30

Beare, B. K. (2017): "The Chang-Kim-Park model of cointegrated density-valued time series cannot accommodate a stochastic trend," *Econ Journal Watch*, 14, 133. 2

Beare, B. K., J. Seo, and W.-K. Seo (2017): "Cointegrated linear processes in Hilbert space," *Journal of Time Series Analysis*, 38, 1010–1027. 2

Beare, B. K. and W.-K. Seo (2020): "Representation of I (1) and I (2) autoregressive Hilbertian processes," *Econometric Theory*, 36, 773–802. 2

Beggs, J. J. and M. Nerlove (1988): "Biases in dynamic models with fixed effects," *Economics Letters*, 26, 29–31. 26

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV estimation of shape-invariant Engel curves," *Econometrica*, 75, 1613–1669. 30

Blundell, R., A. Duncan, and K. Pendakur (1998): "Semiparametric estimation and consumer demand," *Journal of applied econometrics*, 13, 435–461. 30, 31

Blundell, R. W., M. Browning, and I. A. Crawford (2003): "Nonparametric Engel curves and revealed preference," *Econometrica*, 71, 205–240. 30, 31

Bosq, D. (2000): *Linear processes in function spaces: theory and applications*, vol. 149, Springer Science & Business Media. 1, 2, 7, 8, 17, 41

Bosq, D. and D. Blanke (2008): *Inference and prediction in large dimensions*, John Wiley & Sons. 1

Bykhovskaya, A. and V. Gorin (2022): "Cointegration in large VARs," *The Annals of Statistics*, 50, 1593–1617. 6

Bykhovskaya, A. and P. C. B. Phillips (2018): "Boundary limit theory for functional local to unity regression," *Journal of Time Series Analysis*, 39, 523–562. 4

——— (2020): "Point optimal testing with roots that are functionally local to unity," *Journal of Econometrics*, 219, 231–259. 4

Carrasco, M. and J.-P. Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16, 797–834. 16

Chang, Y., C. S. Kim, and J. Y. Park (2016): "Nonstationarity in time series of state densities," *Journal of Econometrics*, 192, 152–167. 2

Chen, K. and H.-G. Müller (2012): "Conditional quantile analysis when covariates are functions, with application to growth data," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74, 67–89. 34

Chen, Y., Z. Lin, and H.-G. Müller (2023): "Wasserstein regression," *Journal of the American Statistical Association*, 118, 869–882. 2

Childers, D. (2018): "Solution of rational expectations models with function valued states," *Manuscript, Carnegie Mellon University.* 2

Cho, J. S., P. C. B. Phillips, and J. Seo (2022): "Parametric conditional mean inference with functional data applied to lifetime income curves," *International Economic Review*, 63, 391–456. 4, 34

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): "Nonparametric instrumental regression," *Econometrica*, 79, 1541–1565. 2

Hahn, J. and G. Kuersteiner (2002): "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large," *Econometrica*, 70, 1639–1657. 24, 25, 26, 27

Hall, P. and J. L. Horowitz (2005): "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statistics*, 33, 2904–2929. 2, 17

Horowitz, J. L. (2014): "Ill-posed inverse problems in economics," *Annual Review of Economics*, 6, 21–51. 2

Horváth, L. and P. Kokoszka (2012): *Inference for functional data with applications*, vol. 200, Springer Science & Business Media. 1, 2

Hu, B., J. Y. Park, and J. Qian (2016): "Analysis of distributional dynamics for repeated cross-sectional and intra-period observations," Tech. rep., Working paper, presented at SETA/NZESG meeting. 2

James, G. (2010): "The Oxford handbook of functional data analysis," *Chapter Sparseness And Functional Data Analysis*, 298–326. 2

Lawford, S. and M. P. Stamatogiannis (2009): "The finite-sample effects of VAR dimensions on OLS bias, OLS variance, and minimum MSE estimators," *Journal of Econometrics*, 148, 124–130. 6

Lewbel, A. (2008): "Engel curves," *The new Palgrave dictionary of economics*, 2, 1–4. 30, 31

Li, D., Y. Li, and P. C. B. Phillips (2025): "Large-scale time series with common stochastic trends," *Cowles Foundation Discussion Paper, No. 2460, Yale University*, 44 pages. 2, 24

Li, D., P. M. Robinson, and H. L. Shang (2020): "Long-range dependent curve time series," *Journal of the American Statistical Association*, 115, 957–971. 3

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2023): "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*, 232, 272–299. 3

Marriott, F. and J. Pope (1954): "Bias in the estimation of autocorrelations," *Biometrika*, 41, 390–402. 26

Métivier, M. (2011): *Semimartingales: A Course in Stochastic Processes*, de Gruyter. 8

Moon, H. R. and B. Perron (2004): "Testing for a unit root in panels with dynamic factors," *Journal of econometrics*, 122, 81–126. 24

Moon, H. R., B. Perron, and P. C. B. Phillips (2007): "Incidental trends and the power of panel unit root tests," *Journal of Econometrics*, 141, 416–459. 24

Moon, H. R. and M. Weidner (2017): "Dynamic linear panel regression models with interactive fixed effects," *Econometric Theory*, 33, 158–195. 24, 26

Morris, J. S. (2014): "Functional Regression," Tech. rep., arXiv:1406.4068v1. The University of Texas. 2

Nickell, S. (1981): "Biases in dynamic models with fixed effects," *Econometrica*, 1417–1426. 26

Park, J. Y. and J. Qian (2012): "Functional regression of continuous state distributions," *Journal of Econometrics*, 167, 397–412. 2

Phillips, P. C. B. (1977): "Approximations to some finite sample distributions associated with a first-order stochastic difference equation," *Econometrica*, 463–485. 18, 19, 37

——— (1987a): "Time series regression with a unit root," *Econometrica*, 277–301. 7

——— (1987b): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535–547. 7

——— (2012): "Folklore theorems, implicit maps, and indirect inference," *Econometrica*, 80, 425–454. 18

——— (2018): "Dynamic panel Anderson-Hsiao estimation with roots near unity," *Econometric Theory*, 34, 253–276. 26

——— (2023): "Estimation and inference with near unit roots," *Econometric Theory*, 39, 221–263. 4

——— (2025): "Edgeworth Expansions in Cross Section Curve Data Autoregression," *Cowles Foundation Discussion Paper, No. 2464, Yale University*, 29 pages. 18, 19, 36, 37, 42

Phillips, P. C. B. and C. Han (2015): "The true limit distributions of the Anderson–Hsiao IV estimators in panel autoregression," *Economics Letters*, 127, 89–92. 26

Phillips, P. C. B. and L. Jiang (2016): "Parametric Autoregression in Function Space," *Working Paper, Singapore Management University*, 56 pages. 1

——— (2025): "Cross Section Curve Autoregression: the Unit Root Case," *Cowles Foundation Discussion Paper, No. 2454, Yale University*. 1, 4, 7, 32, 34

Phillips, P. C. B. and T. Magdalinos (2007): "Limit theory for moderate deviations from a unit root," *Journal of Econometrics*, 136, 115–130. 4

Phillips, P. C. B. and H. R. Moon (1999): "Linear regression limit theory for nonstationary panel data," *Econometrica*, 67, 1057–1111. 14, 45

Phillips, P. C. B. and D. Sul (2007): "Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence," *Journal of Econometrics*, 137, 162–188. 24, 26

Reaños, M. A. T. and N. M. Wölfing (2018): "Household energy prices and inequality: Evidence from German microdata based on the EASI demand system," *Energy Economics*, 70, 84–97. 30

Shenton, L. and W. L. Johnson (1965): "Moments of a serial correlation coefficient," *Journal of the Royal Statistical Society: Series B (Methodological)*, 27, 308–320. 18

Vinod, H. D. and L. Shenton (1996): "Exact Moments for Autor1egressive and Random walk Models for a Zero or Stationary Initial Value," *Econometric Theory*, 12, 481–499. 18

White, J. S. (1961): "Asymptotic expansions for the mean and variance of the serial correlation coefficient," *Biometrika*, 48, 85–94. 18

Zhang, C., P. Kokoszka, and A. Petersen (2022): "Wasserstein autoregressive models for density time series," *Journal of Time Series Analysis*, 43, 30–52. 2, 8, 9

Zhang, J.-T. and J. Chen (2007): "Statistical inferences for functional data," *Annals of Statistics*, 35, 1052–1079. 4