Testing a Constant Mean Function Using Functional Data

JIN SEO CHO

School of Humanities and Social Sciences, Beijing Institute of Technology, Haidan, Beijing 100081, China
School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, Korea
jinseocho@yonsei.ac.kr

MENG HUANG

Quantitative Analytics & Model Development Group, PNC, Washington D.C., US
nkhuangmeng@gmail.com

HALBERT WHITE

Department of Economics, University of California, San Diego, 9500 Gilman Dr. 0508, La Jolla, CA 92093-0508, US

This version: September 2020

Abstract

In this paper, we study functional ordinary least squares estimator and its properties in testing the hypothesis of a constant zero mean function or an unknown constant non-zero mean function. We exploit the recent work by Cho, Phillips and Seo (2020) and show that the associated Wald test statistics have standard chi-square limiting null distributions, standard non-central chi-square distributions for local alternatives converging to zero at a $\sqrt{n}$ rate, and are consistent against global alternatives. These properties permit computationally convenient tests of hypotheses involving nuisance parameters. In particular, we develop new alternatives to tests for regression misspecification, which involves nuisance parameters identified only under the alternative. In Monte Carlo studies, we find that our tests have well behaved levels. We also find that functional ordinary least squares tests can have power better than existing methods that do not exploit this covariance structure, like the specification testing procedures of Bierens (1982, 1990) or Stinchcombe and White (1998).

Key Words: Davies Test; Functional Data; Hypothesis Testing; Integrated Conditional Moment Test; Misspecification; Mixture Distributions; Nuisance Parameters; Wald Test.

Subject Classification: C11, C12, C80.

Acknowledgements: The third author (Halbert White) passed away while the earlier version was circulating as a discussion paper. He initiated the project and outlined the paper. We have also benefited from discussions with Peter C.B. Phillips, Juwon Seo, and seminar participants at Yonsei University and EJC09.

1 Introduction

A considerable variety of useful testing procedures involve "nuisance" parameters. Examples are those considered in the work of Davies (1977, 1987), Bierens (1982, 1990), Bierens and Ploberger (1997), Andrews and Ploberger (1994), and Stinchcombe and White (1998). In these examples, as well as in this context generally, test statistics are constructed by "integrating out" the nuisance parameters, yielding nuisance parameter-free tests. A general consequence of this approach is that the limiting null distributions of the resulting test statistics are highly context specific, requiring special purpose computations to obtain suitable critical values.

In this paper, we consider a different approach, useful in this context, that yields statistics having standard chi-square limiting null distributions. In some cases, our procedures can have better power than previous procedures. For example, this is illustrated by the specification tests of Bierens (1982, 1990) and Stinchcombe and White (1998). Note that the tests of Bierens (1982, 1990) and Stinchcombe and White (1998) do not take account of correlations among the elements of the Gaussian process underlying the test statistic as for the test in Davies (1987). Our procedures also do not take account of the correlations, but this affords computational convenience, analogous to the way that tests based on heteroskedasticity-consistent covariance matrices yield convenient tests of proper size by neglecting efficiency improvements that could be gained by modeling the heteroskedasticity.

The approach taken here is that of hypothesis testing in *functional regression* by applying the functional least squares estimator in Cho, Phillips and Seo (2020), who examine estimating a parametric model for the conditional mean of a continuous functional observation by the functional least squares estimation. We apply their approach by supposing possibly discontinuous functional data and a model constructed by non-random functions attached to linear coefficients to infer the mean function, and from this we provide an estimator and its asymptotic properties along with properly tailored regularity conditions for the estimator. Specifically, the dependent variable is a random function (of $\gamma \in \Gamma$, say) rather than a random variable, and the regressors are user-specified non-random functions of γ chosen to give a good approximation to the mean function of the dependent variable. Under the null hypotheses of interest here, this mean function is either the zero function or an unknown non-zero constant function. We analyze testing procedures designed to have power against the alternatives to either of these nulls by specifying a linear model constructed by deterministic functions with unknown linear coefficients. An appealing consequence of using functional regression is that the resulting test statistics have standard chi-square limiting distributions under the null. Wald, Lagrange multiplier, and quasi-likelihood ratio versions of these statistics are available. For concreteness and conciseness, our focus here is on the use of Wald statistics.

Although functional regression is of theoretical interest in its own right, our focus here is also in illustrating its usefulness in specific application areas. In one sense, functional regression is familiar, in that standard panel data structures can be viewed as examples of functional data. We illustrate this with a running example focused on tests of random effects structure in panel data. On the other hand, the functions of interest arising in the analysis of models involving nuisance parameters identified only under the alternative can also be viewed as instances of functional data. We exploit this here to provide appealing ways of testing hypotheses concerning unidentified nuisance parameters. We pay specific attention to specification testing, as in Bierens (1982, 1990) and Stinchcombe and White (1998).

Functional data analysis is getting popular. We mention a few recent developments. Crambes, Gannoun, and Henchiri (2013) estimate a quantile regression function with a functional covariate by the support vector machine when the dependent variable is a real random variable. Zhang and Chen (2007) first apply the local polynomial kernel estimation to discrete data and next estimate the functional coefficients of the covariates consisted of random variables according to the so-called "smoothing first, then estimation" principle. The authors show that the influence of the smoothing process is diluted as the sample size tends to infinity under some mild regularity conditions. Li, Robinson, and Shang (2020) examine time series of function space curves with long range dependence to yield the limit theory for the sample average of functional observations. They further estimate the covariance kernel function of the functional data by using the functional principal component analysis. Chang, Hu, and Park (2019) estimate a functional autoregressive model with serially correlated functional data, establishing limit theory of their estimator. Phillips and Jiang (2019) study parametric autoregression with function valued time series in stationary and nonstationary cases and also establish limit theory of their estimation. Finally, when the conditioning variable is a random variable, Cho, Phillips, and Seo (2020) examine estimating the conditional mean of functional data, which is nonlinear with respect to unknown parameters. These papers assume the use of functional data but differ from the current paper as the current study focuses on the inference of a parametric population mean function.

The plan of this paper is as follows. In Section 2, we motivate and formally describe the data generating process underlying functional regression, illustrating with examples involving random effect structure in the context of panel data and specification testing. In Section 3, we introduce the functional ordinary least squares (FOLS) and two-stage FOLS (TSFOLS) estimators that are obtained by imposing the linear structure to the functional regression in Cho, Phillips, and Seo (2020). We provide conditions under which these estimators are consistent and asymptotically normal, and we provide consistent estimators of their asymptotic covariance matrices in parallel to Cho, Phillips, and Seo (2020). In Section 4, we specify the null hypotheses of interest and introduce Wald statistics useful for testing these. As we show, these statistics have standard chi-square distributions under the null. We analyze their global and local power properties. Globally, our procedures are consistent; locally we obtain standard non-central chi-square distributions for alternatives converging at the parametric $\sqrt{n}$ rate, where n is the sample size. Section 5 applies the theory developed in the preceding sections to obtain test statistics for our panel data and specification testing examples. Section 6 provides a Monte Carlo analysis, where we study the finite and large sample properties of tests based on the statistics developed in Section 5. Section 7 contains a summary and concluding remarks.

Before proceeding, we introduce some mathematical notation used throughout. First, integrals of functions will be often used in this paper, and we let $\int g \, d\mathbb{P}$ and $\int h \, d\mathbb{P} d\mathbb{Q}$ respectively denote $\int g(x) d\mathbb{P}(x)$ and $\int\int h(x,y) d\mathbb{P}(x) d\mathbb{Q}(y)$

for brevity, unless confusion otherwise arises. When there is no possible ambiguity, we may further abbreviate these to $\int g$ and $\int\int h$. Unless explicitly noted otherwise, limits are taken as $n \to \infty$.

2 The Data Generating Process and Functional Regression

In this section, we motivate and formally describe the data generating process underlying functional regression.

2.1 The Data Generating Process

We consider data generated as follows:

Assumption 1 (DGP-A). *(i) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let (Γ, ρ) be a compact metric space;*

(ii) For $i = 1, 2, ...$, let $G_i : \Omega \times \Gamma \mapsto \mathbb{R}$ be such that for each $\gamma \in \Gamma$, $G_i(\,\cdot\,, \gamma)$ is measurable and independently and identically distributed (IID). $\square$

Often in econometrics, such a function G_i is used to define a *model*, that is a collection of functions $\mathbb{G}_i := \{G_i(\,\cdot\,, \gamma) : \gamma \in \Gamma\}$ that, when "correctly specified," includes some functional of a data generating process for random variables of interest. (See, for example, White, 1994, ch. 2.2.) For example, in that context, $G_i(\omega, \,\cdot\,)$ might represent the log-likelihood function for observation i, determined by the realization $\omega \in \Omega$. Correct specification occurs when there is $\gamma_0 \in \Gamma$ such that $G_i(\,\cdot\,, \gamma_0)$ represents the log density of the data generating process (DGP) for observation i.

Here, we view G_i rather differently. Specifically, we view the observed data not as realizations of random variables, as is common, but as realizations of random functions $\gamma \mapsto G_i(\,\cdot\,, \gamma)$. That is, we observe $G_i(\omega, \,\cdot\,) : \Gamma \mapsto \mathbb{R}$, $i = 1, 2, ...$ for some $\omega \in \Omega$. The IID condition is not essential, but we impose it to keep the main ideas clear. Because our interest is primarily on G_i as a random function of γ, we may abbreviate $G_i(\,\cdot\,, \gamma)$ as $G_i(\gamma)$ for notational simplicity.

The data structure in Assumption 1 is more general than that assumed by Cho, Phillips, and Seo (2020). Note that we do not impose any functional restriction to the functional data, whereas they restrict their interest to continuous functional data. As our model detailed below is linear with respect to unknown coefficients, we can relax the data restriction and examine the asymptotic properties of the estimated coefficients.

To illustrate, we discuss two examples. First, we show how the familiar case of panel data falls into the present framework. As we show later, this supports tests for features of interest in panel data, such as random effects structure. We operate within the panel data setting nicely exposited by Wooldridge (2002, ch.10.4).

Example 1 (Panel Random Effects): Let $\gamma \in \Gamma := \{1, 2, ..., T\}$, and suppose data are generated as $Y_i(\gamma) = X_i(\gamma)'\beta_0 + V_i(\gamma)$, $i = 1, 2, ...,$ where $\beta_0 \in \mathbb{R}^d$ and $V_i(\gamma) := C_i + U_i(\gamma)$. We assume that $(Y_i, X_i')' : \Omega \times \Gamma \mapsto \mathbb{R}^{1+d}$ is IID. $U_i : \Omega \times \Gamma \mapsto \mathbb{R}$ and $C_i : \Omega \mapsto \mathbb{R}$ are unobserved. Let $\mathbf{X}_i := (X_i(1), X_i(2), \cdots, X_i(T))'$, $\mathbf{V}_i := (V_i(1), V_i(2), \cdots, V_i(T))'$, and assume that $\mathbf{\Sigma} := E[\mathbf{V}_i\mathbf{V}_i']$ is finite and positive definite, with $\text{rank}(E[\mathbf{X}_i'\mathbf{\Sigma}^{-1}\mathbf{X}_i]) = d$. The data exhibit *random effects* structure when, for $i = 1, 2, ...,$

1. $U_i(\gamma)$ is IID with respect to γ, and $E[U_i(\gamma)|X_i(\gamma), C_i] = 0$ for each $\gamma \in \Gamma$; and
2. $E[C_i|X_i(\gamma)] = E[C_i] = 0$ for each $\gamma \in \Gamma$.

Under these assumptions, we may write $\sigma_u^2 := E[U_i(\gamma)^2]$ for all $\gamma \in \Gamma$ and $\sigma_c^2 := E[C_i^2]$. The covariance matrix $\mathbf{\Sigma}$ has the form

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_u^2 + \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_u^2 + \sigma_c^2 & \cdots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_u^2 + \sigma_c^2 \end{pmatrix}.$$

When $\sigma_c^2 = 0$, the unobserved effect C_i is absent, and V_i is identical to U_i.

Now consider $G_i(\gamma) = V_i(1)V_i(\gamma)$. Under random effects with $E[G_i(\gamma)] = 0$ for all $\gamma \in \Gamma \setminus \{1\}$, the conventional pooled OLS estimator for β_0 is efficient, and we can use pooled OLS to conduct efficient statistical inference. On the other hand, when $E[G_i(\gamma)] = \sigma_c^2 > 0$ for $\gamma \in \Gamma \setminus \{1\}$, the feasible generalized least squares (FGLS) estimator that exploits the structure of $\mathbf{\Sigma}$ is more efficient than pooled OLS. Moreover, the presence of the unobserved effect C_i may necessitate the use of methods appropriate for handling unobserved fixed effects. $\square$

Another leading case of interest here is associated with what is known in the literature as "nuisance parameters identified only under the alternative." See, for example, Davies (1977, 1987), Andrews (2001), Cho and White (2007, 2010), Baek et al. (2015), Cho and Phillips (2018) and the references therein. An important example involving nuisance parameters present only the alternative is the specification testing framework of Bierens (1990) and its extensions (e.g., Stinchcombe and White, 1998 (SW)).

Example 2 (Specification Testing): Let $\{(Y_i, X_i')' \in \mathbb{R}^{1+d}\}$ be IID, and suppose $E[Y_i|X_i]$ is modeled by a set of functions, say $\mathbb{M} := \{f(X, \theta) : \theta \in \Theta \subset \mathbb{R}^m\}$, where d and m are finite integers. Further, for $\gamma \in \Gamma$, let $G_i(\gamma) = [Y_i - f(X_i, \theta^*)]\psi\{\gamma' X_i\}$, where θ^* is the probability limit of an estimator $\widehat{\theta}_n$, e.g., the nonlinear least squares (NLS) estimator

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} \; \frac{1}{n}\sum_{i=1}^{n}[Y_i - f(X_i, \theta)]^2;$$

and $\psi : \mathbb{R} \to \mathbb{R}$ is a given function. Bierens (1990) specifies $\psi = \exp$; SW consider large families of choices for ψ, notably the comprehensively revealing (CR) and the generically CR (GCR) families[1].

This choice for G_i is easily seen to satisfy Assumption 1 under mild conditions on f and ψ. Further, G_i has remarkable and useful properties. Specifically, as Bierens (1990) and SW show, when $\mathbb{M}$ is correctly specified (so that there exists $\theta_0 \in \Theta$ such that $E[Y_i|X_i] = f(X_i, \theta_0)$), provided that $\theta^* = \theta_0$ (as holds for the NLS estimator as well as for linear exponential family-based quasi-maximum likelihood estimators generally), then $E[G_i(\gamma)] = 0 \quad$ for all $\gamma \in \Gamma$; whereas

[1]To ensure boundedness, Bierens (1990) replaces $\mathbf{X}_i$ with $\Phi(\mathbf{X}_i)$, a $d \times 1$ vector of measurable bounded one-to-one mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$, such as $\Phi(\mathbf{X}_i) := [\tan^{-1}(X_{1i}),\ \tan^{-1}(X_{2i}),\ \cdots,\ \tan^{-1}(X_{di})]'$. We leave this implicit here.

when $\mathbb{M}$ is *not* correctly specified and ψ is GCR (e.g., $\psi = \exp$ is GCR), SW show $E[G_i(\gamma)] \neq 0$ for almost all $\gamma \in \Gamma$. Bierens (1990) and SW exploit this property to construct tests for model misspecification. Their test statistics are based on

$$Z_n(\gamma) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i(\gamma). \qquad \square$$

As these examples suggest, our main interest here concerns the population mean functional μ of G_i (when it exists) defined by

$$\mu(\gamma) := E_{\mathbb{P}}[G_i(\gamma)] := \int G_i(\gamma) d\mathbb{P}, \quad \gamma \in \Gamma.$$

We exploit the identical distribution assumption to drop the i subscript for μ.

We pay particular attention to certain functionals of μ. To specify these, we introduce the notion of an *adjunct* probability measure $\mathbb{Q}$ on Γ. This measure should be viewed as one selected by the researcher; it corresponds to the familiar notion of a regression design. We specify its properties formally as follows:

Assumption 2 (Adjunct Probability Measure). *(i)* $(\Gamma, \mathcal{G}, \mathbb{Q})$ *and* $(\Omega \times \Gamma, \mathcal{F} \otimes \mathcal{G}, \mathbb{P} \cdot \mathbb{Q})$ *are complete probability spaces; (ii) For* $i = 1, 2, ...$, G_i *is measurable* $-\mathcal{F} \otimes \mathcal{G}$. $\square$

The sample space is now the Cartesian product, $\Omega \times \Gamma$; the sigma field $\mathcal{F} \otimes \mathcal{G}$ is the product sigma field generated by $\mathcal{F}$ and $\mathcal{G}$. Because (Γ, ρ) is a metric space, there exists a topology generated by ρ. We may take $\mathcal{G}$ to be the Borel sigma field generated by this topology. The product probability measure $\mathbb{P} \cdot \mathbb{Q}$ governs events jointly involving ω and γ. Because of its product structure, we have independence, in the usual sense that $\mathbb{P} \cdot \mathbb{Q}[F \times G] = \mathbb{P}[F] \cdot \mathbb{Q}[G]$ for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$.

The assumed joint measurability for G_i follows, for example, by Stinchcombe and White (1992, lemma 2.15), if $G_i(\cdot, \gamma)$ is measurable for each $\gamma \in \Gamma$ and $G_i(\omega, \cdot)$ is continuous on Γ for all $\omega \in F$, $\mathbb{P}[F] = 1$.

Under suitable integrability conditions, our assumptions ensure that integrals of the form $\int \int H_i(\omega, \gamma) d\mathbb{Q}(\gamma) d\mathbb{P}(\omega)$ are well defined. Of immediate interest is the integral arising when $H_i(\omega, \gamma) = \{G_i(\omega, \gamma) - m(\gamma)\}^2$, yielding

$$\int \int \{G_i - m\}^2 d\mathbb{Q} d\mathbb{P} = \int \int \{G_i(\omega, \gamma) - m(\gamma)\}^2 d\mathbb{Q}(\gamma) d\mathbb{P}(\omega).$$

This is the $\mathbb{Q}-$*functional mean squared error* ($\mathbb{Q}-$FMSE) for m as a predictor of G_i. As we show next, for every $\mathbb{Q}$, the function m^* minimizing the $\mathbb{Q}-$FMSE is essentially the functional mean, μ. To establish this, we introduce some notation and add some suitable regularity. First, we write $L_2(\mathbb{P}) := \{f : \int |f(\omega)|^2 d\mathbb{P}(\omega) < \infty\}$ and similarly $L_2(\mathbb{Q}) := \{f : \int |f(\gamma)|^2 d\mathbb{Q}(\gamma) < \infty\}$, where f is measurable-$\mathcal{F}$ in the first instance and measurable-$\mathcal{G}$ in the second.

Assumption 3 (Domination). *For random variables* $M_i \in L_2(\mathbb{P})$, $\sup_{\gamma \in \Gamma} |G_i(\gamma)| \leq M_i$, $i = 1, 2, ...$ $\square$

From this, it follows that μ as defined above exists and is measurable $-\mathcal{G}$, and that $\mu \in L_2(\mathbb{Q})$.

Clearly, the optimized $\mathbb{Q}-$FMSE depends on $\mathbb{Q}$. In particular, if for some $\gamma_0 \in \Gamma$, $\mathbb{Q}$ is selected so that $\mathbb{Q}(G) = 1$ if $\gamma_0 \in G \in \mathcal{G}$ and $\mathbb{Q}(G) = 0$ otherwise, then $m^* = \mu$ a.s.$-\mathbb{Q}$ holds for the constant function $m^* = \mu(\gamma_0)$, and the minimized $\mathbb{Q}-$FMSE is $\int \text{var}_{\mathbb{P}}[G_i(\gamma)]d\mathbb{Q}(\gamma) = \text{var}_{\mathbb{P}}[G_i(\gamma_0)]$. This replicates the familiar result for random variables that the expectation $\mu(\gamma_0)$ is the best mean-squared error predictor for the random variable $G_i(\gamma_0)$. Analogously, the function defined by $\mu(\gamma)$ provides a $\mathbb{Q}-$FMSE optimal prediction for the random function defined by $G_i(\,\cdot\,, \gamma)$.

2.2 Functional Regression

Our primary interest attaches to testing hypotheses about μ. For example, given a known function $m^* \in L_2(\mathbb{Q})$, suppose we are interested in testing

$$\mathbb{H}_o : \mu = m^* \quad a.s. - \mathbb{Q} \qquad \text{vs.} \qquad \mathbb{H}_A : \mathbb{H}_o \text{ is false.}$$

Because m^* is known, this is equivalent to testing

$$\mathbb{H}_o : \mu^* = 0 \quad a.s. - \mathbb{Q} \qquad \text{vs.} \qquad \mathbb{H}_A : \mathbb{H}_o \text{ is false,}$$

where $\mu^* := \mu - m^* = E_{\mathbb{P}}[G_i^*]$, with $G_i^*(\gamma) := G_i(\gamma) - m^*(\gamma)$.

We may be also interested in testing

$$\mathbb{H}_o : \mu^* = c \quad a.s. - \mathbb{Q} \qquad \text{vs.} \qquad \mathbb{H}_A : \mathbb{H}_o \text{ is false,}$$

where c is an unknown real constant. For example, in our panel data example, this case is relevant in testing the null of no serial correlation in U_i with respect to γ versus serial correlation in U_i in the possible presence of the unobserved effect C_i.

In what follows, we drop the superscript *, letting any recentering by known m^* be implicit, and just consider testing

$$\mathbb{H}_{1o} : \mu = 0 \quad a.s. - \mathbb{Q} \qquad \text{vs.} \qquad \mathbb{H}_{1A} : \mathbb{H}_{1o} \text{ is false;} \quad \textit{and} \quad \mathbb{H}_{2o} : \mu = c \quad a.s. - \mathbb{Q} \qquad \text{vs.} \qquad \mathbb{H}_{2A} : \mathbb{H}_{2o} \text{ is false.}$$

Power against particular alternatives may be enhanced by making use of non-constant basis functions $g_j : \Gamma \mapsto \mathbb{R}$, $j = 1, 2, ..., k$; we write $\mathbf{g} := (g_1, g_2, ..., g_k)'$. The next assumption specifies their properties. We let $\lambda_{\min}(\,\cdot\,)$ and $\lambda_{\max}(\,\cdot\,)$ denote the minimum and maximum eigenvalues respectively of a given matrix.

Assumption 4 (Basis Functions)**.** *(i) For each $j = 1, 2, ..., k$, $g_j : \Gamma \mapsto \mathbb{R}$ is measurable$-\mathcal{G}$;*

(ii) For each $j = 1, 2, ..., k$, $g_j \in L_2(\mathbb{Q})$; and

(iii) $\lambda_{\min}(\mathbf{A}) > 0$, where

$$\mathbf{A} := \begin{bmatrix} 1 & \int \mathbf{g}(\gamma)' d\mathbb{Q}(\gamma) \\ \int \mathbf{g}(\gamma) d\mathbb{Q}(\gamma) & \int \mathbf{g}(\gamma)\mathbf{g}(\gamma)' d\mathbb{Q}(\gamma) \end{bmatrix}. \qquad \square$$

Assumption 4(*ii*) ensures that $\lambda_{\max}(\mathbf{A}) < \infty$. Assumption 4(*iii*) ensures that the elements of $\mathbf{g}$ are non-constant and non-redundant. As both $\mathbf{g}$ and $\mathbb{Q}$ are under the researcher's control, verifying Assumption 4 is in principle straightforward.

We use $\mathbf{g}$ to approximate μ. Specifically, we consider linear approximations to μ of the form $m(\,\cdot\,, \boldsymbol{\delta}) = \delta_0 + \mathbf{g}(\,\cdot\,)'\delta$. Thus, m belongs to the linear model

$$\mathcal{A}(\mathbf{g}) := \{\delta_0 + \mathbf{g}(\,\cdot\,)'\delta : (\delta_0, \delta) \in \mathbb{R}^{1+k}\}.$$

A "trivial" but important special case for $\mathbf{g}$ is that in which $\mathbf{g}$ has no elements. This gives the simplest test of $\mathbb{H}_{1o}$, although this choice is not relevant for testing $\mathbb{H}_{2o}$. The most convenient non-trivial choice for $\mathbf{g}$ is $\mathbf{g}(\gamma) = \gamma$, which yields a linear functional regression.

The model structure of $\mathcal{A}(\mathbf{g})$ is simpler than that considered by Cho, Phillips, and Seo (2020) in the sense that their model consists of random functions characterized by random conditioning variables and is nonlinearly parameterized. Due to this simple structure, we can below provide better tailored regularity conditions for our estimator without imposing many regularity conditions.

More elaborate choices of $\mathbf{g}$ are often relevant. In some cases, the alternative may provide specific knowledge about relevant choices for $\mathbf{g}$. Alternatively, one can use series functions, such as suitably chosen polynomials in γ, just as when one approximates a standard conditional expectation. The key idea is that power may be gained by selecting $\mathbf{g}$ to capture salient features of μ under important or plausible alternatives.

When $\mathbb{H}_{1o}$ holds, we have the regression representation

$$G_i(\,\cdot\,) = \delta_0^\dagger + \mathbf{g}(\,\cdot\,)'\delta^\dagger + \varepsilon_i(\,\cdot\,), \tag{1}$$

where $\delta_0^\dagger = 0, \delta^\dagger = \mathbf{0}$, $E_{\mathbb{P}}[\varepsilon_i(\,\cdot\,)] = 0$, and $E_{\mathbb{P}}[\mathbf{g}(\,\cdot\,)\varepsilon_i(\,\cdot\,)] = \mathbf{0}$. When $\mathbb{H}_{2o}$ holds we have the same representation, but now with $\delta_0^\dagger = c, \delta^\dagger = \mathbf{0}$. We call a representation of the form given by eq.(1) a *functional regression*.

We let δ_0^* and δ^* index the $\mathbb{Q}-$FMSE optimizer. That is, $m(\,\cdot\,, \boldsymbol{\delta}^*)$ solves

$$\inf_{m \in \mathcal{A}(\mathbf{g})} \int\int \{G_i - m\}^2 d\mathbb{Q} d\mathbb{P} = \int \mathrm{var}_{\mathbb{P}}[G_i] d\mathbb{Q} + \inf_{\delta_0, \delta} \int \{\mu - \delta_0 - \mathbf{g}'\delta\}^2 d\mathbb{Q}.$$

The first-order conditions for the optimum are

$$\int \mu(\gamma) d\mathbb{Q}(\gamma) = \delta_0^* + \int \mathbf{g}(\gamma)'\delta^* d\mathbb{Q}(\gamma); \quad \text{and} \quad \int \mu(\gamma)\mathbf{g}(\gamma) d\mathbb{Q}(\gamma) = \int (\delta_0^* + \mathbf{g}(\gamma)'\delta^*)\mathbf{g}(\gamma) d\mathbb{Q}(\gamma).$$

These yield convenient expressions for δ_0^* and δ^*, analogous to the standard regression approximation case (see, e.g., White, 1980):

$$\boldsymbol{\delta}^* := \begin{bmatrix} \delta_0^* \\ \delta^* \end{bmatrix} := \begin{bmatrix} E_{\mathbb{Q}}[\mu] \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} -E_{\mathbb{Q}}[\,\mathbf{g}\,]'\text{cov}_{\mathbb{Q}}[\mathbf{g},\mathbf{g}]^{-1}\text{cov}_{\mathbb{Q}}[\mathbf{g},\mu] \\ \text{cov}_{\mathbb{Q}}[\mathbf{g},\mathbf{g}]^{-1}\text{cov}_{\mathbb{Q}}[\mathbf{g},\mu] \end{bmatrix},$$

where $E_{\mathbb{Q}}[\mu] := \int \mu \, d\mathbb{Q}$, $E_{\mathbb{Q}}[\,\mathbf{g}\,] := \int \mathbf{g} \, d\mathbb{Q}$;

$$\text{cov}_{\mathbb{Q}}[\mathbf{g},\mathbf{g}] := \int \mathbf{g}(\gamma)\mathbf{g}(\gamma)' \, d\mathbb{Q}(\gamma) - \left(\int \mathbf{g}(\gamma) \, d\mathbb{Q}(\gamma)\right)\left(\int \mathbf{g}(\gamma)' \, d\mathbb{Q}(\gamma)\right); \quad \text{and}$$

$$\text{cov}_{\mathbb{Q}}[\mathbf{g},\mu] := \int \mathbf{g}(\gamma)\mu(\gamma) \, d\mathbb{Q}(\gamma) - \left(\int \mathbf{g}(\gamma) \, d\mathbb{Q}(\gamma)\right)\left(\int \mu(\gamma) \, d\mathbb{Q}(\gamma)\right).$$

It is readily verified that if $\mu = 0 \; a.s. - \mathbb{Q}$ ($\mathbb{H}_{1o}$ holds) then $\boldsymbol{\delta}^* = [0, \mathbf{0}']'$. If instead, for unknown constant c, $\mu = c \; a.s. - \mathbb{Q}$ ($\mathbb{H}_{2o}$ holds) then $\boldsymbol{\delta}^* = [c, \mathbf{0}']'$. Thus, δ_0^* and δ^* coincide with the coefficients of the functional regression representation for $G_i(\,\cdot\,)$ under $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$.

On the other hand, if $\mathbb{H}_{1o}$ does not hold, then δ_0^* or δ^* need not equal zero, as $\text{cov}_{\mathbb{Q}}[\mathbf{g},\mu]$ is not necessarily $\mathbf{0}$ under $\mathbb{H}_{1A}$. Similarly, if $\mathbb{H}_{2o}$ does not hold, then δ^* need not equal zero. This behavior gives our tests their power. We emphasize that in these cases, the optimizer $m(\,\cdot\,, \boldsymbol{\delta}^*)$ generally does not coincide with μ, as $m(\,\cdot\,, \boldsymbol{\delta}^*)$ is essentially a misspecified approximation to μ under the specified alternatives.

3 Functional Ordinary Least Squares (FOLS) Estimation

We construct hypothesis testing procedures based on estimators for δ_0^* and δ^*. For this, we minimize with respect to δ_0 and δ the sample analog of the $\mathbb{Q}-$FMSE,

$$\frac{1}{n}\sum_{i=1}^{n}\int\{G_i(\gamma) - \delta_0 - \mathbf{g}(\gamma)'\delta\}^2 d\mathbb{Q}(\gamma).$$

The resulting estimator is the *functional ordinary least squares* (FOLS) estimator, denoted $(\widehat{\delta}_{0n}, \widehat{\delta}_n')'$. This has the convenient representation

$$\widehat{\boldsymbol{\delta}}_n := \begin{bmatrix} \widehat{\delta}_{0n} \\ \widehat{\delta}_n \end{bmatrix} := \begin{bmatrix} 1 & \int \mathbf{g} \\ \int \mathbf{g} & \int \mathbf{g}\,\mathbf{g}' \end{bmatrix}^{-1} \begin{bmatrix} n^{-1}\sum \int G_i \\ n^{-1}\sum \int \mathbf{g}\, G_i \end{bmatrix},$$

where the integration is always with respect to $d\mathbb{Q}$. Note that this estimator corresponds to the functional least squares estimator in Cho, Phillips, and Seo (2020) obtained by restricting nonlinear structures to the model.

3.1 Consistency of FOLS

The asymptotic properties of the FOLS estimator depend on the properties of G_i. We first require that $n^{-1}\sum_{i=1}^{n} G_i$ obeys the strong uniform law of large numbers (SULLN).

Assumption 5 (SULLN). $\sup_{\gamma\in\Gamma}\left|n^{-1}\sum_{i=1}^{n} G_i(\gamma) - \mu(\gamma)\right| \to 0$ *a.s.*$-\mathbb{P}$. □

Given the domination condition of Assumption 3, this holds under mild additional conditions on $\{G_i\}$. Specifically, if $G_i(\omega,\,\cdot\,)$ is continuous on Γ, then the SULLN of Le Cam (1953) (see also Jennrich, 1969) applies. Additional relevant references are Andrews (1987), Pötscher and Prucha (1989), and Newey (1991).

The dominated convergence theorem (DCT) permits us to first let n tend to infinity before integrating the relevant random functions with respect to $\mathbb{Q}$ involved in $\widehat{\delta}_{0n}$ and $\widehat{\delta}_n$. The key assumptions permitting this are Assumptions 3 and 4(*ii*). With this, we obtain the consistency of the FOLS estimator.

Theorem 1. *Given Assumptions 1, 2, 3, 4, and 5,* $\widehat{\boldsymbol{\delta}}_n \to \boldsymbol{\delta}^*$ *a.s.* $-\mathbb{P}$. □

3.2 Asymptotic Normality of FOLS

The FOLS estimator has the joint normal distribution asymptotically. For this, we impose a functional central limit theorem (FCLT).

Assumption 6 (FCLT). *(i)* $n^{-1/2}\sum_{i=1}^{n}(G_i - \mu) \Rightarrow \mathcal{Z}$, *where* $\mathcal{Z} : \Omega \times \Gamma \mapsto \mathbb{R}$ *is a mean zero Gaussian process such that for* $\gamma, \widetilde{\gamma} \in \Gamma$, $E_{\mathbb{P}}[\mathcal{Z}(\gamma)\mathcal{Z}(\widetilde{\gamma})] = \kappa(\gamma,\widetilde{\gamma}) < \infty$, *where* $\kappa : \Gamma\times\Gamma \mapsto \mathbb{R}$ *is such that for each* $j, \widetilde{j} \in \{1, 2, ..., k\}$,

$$\int\int \kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) < \infty, \quad \int\int g_j(\gamma)\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) < \infty, \quad \textit{and}$$

$$\int\int g_j(\gamma)\kappa(\gamma,\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) < \infty; \quad \textit{and}$$

(ii) Let

$$\mathbf{B} := \begin{bmatrix} \int\int \kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) & \int\int \kappa(\gamma,\widetilde{\gamma})\mathbf{g}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \\ \int\int \mathbf{g}(\gamma)\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) & \int\int \mathbf{g}(\gamma)\kappa(\gamma,\widetilde{\gamma})\mathbf{g}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \end{bmatrix},$$

and suppose that $\lambda_{\min}(\mathbf{B}) > 0$. □

There is an extensive literature providing primitive conditions for the FCLT. Billingsley (1968, 1999) provides primitive conditions when Γ is a compact subset of the real line and G_i belongs to a set of right-continuous functions with left-limits. These results are extended by Bickel and Wichura (1971) to the case where Γ is a compact subset of a finite dimensional Euclidean space. When, as is assumed here, (Γ, ρ) is a compact metric space, Jain and Marcus (1975) provide sufficient conditions for the FCLT[2]. For additional literature developing these conditions under various contexts, see, for example, Shorack and Wellner (1986) and van den Vaart and Wellner (1996).

By construction, $\kappa(\gamma, \widetilde{\gamma})$ defines a measurable symmetric function. Many useful choices for $\mathbf{g}$ are bounded; in such cases, only the first of the integrability conditions in Assumption 6(*i*) is needed. Further, Assumption 6(*i*) ensures that $\lambda_{\max}(\mathbf{B}) < \infty$. Assumption 6(*ii*) ensures that the asymptotic distribution of the FOLS estimator is not degenerate. For example, Assumption 6(*ii*) fails if κ is constant over $\Gamma \times \Gamma$. Constant κ occurs when G_i is a random constant function.

We can now give the asymptotic distribution of the FOLS estimator.

Theorem 2. *Given Assumptions 1, 2, 3, 4, 5 and 6,* $\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{A}{\sim} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right)$. □

The asymptotic normality ensured by this result makes it easy to construct tests of our hypotheses of interest.

Observe that the asymptotic covariance matrix has the sandwich form common to estimators of misspecified models (see, e.g., Huber, 1967; White, 1982, 1994). Nevertheless, this matrix does not simplify further even under $\mathbb{H}_{1o}$ or $\mathbb{H}_{2o}$ (where functional form misspecification is absent) because the functional data contain a stochastic dependence structure captured by κ; this is the analog of neglected heteroskedasticity. We accept this in order to avoid undertaking the intensive effort that would otherwise be required to model and accommodate κ.

3.3 Two-Stage FOLS

In applications, we often encounter situations in which an estimator $\widehat{G}_i(\,\cdot\,, \gamma)$ appears in place of $G_i(\,\cdot\,, \gamma)$. Our Examples 1 and 3 are relevant instances. To handle these cases in a general way, it suffices to assume that

$$G_i(\,\cdot\,, \gamma) := \widetilde{G}_i(\,\cdot\,, \gamma, \theta^*)$$

for some suitably regular function $\widetilde{G}_i$, where θ^* is an unknown $m \times 1$ vector (m finite) in Θ, say. We then form

$$\widehat{G}_i(\,\cdot\,, \gamma) := \widetilde{G}_i(\,\cdot\,, \gamma, \widehat{\theta}_n),$$

where $\widehat{\theta}_n$ is a suitable estimator of θ^*, computed in a first stage. From this, we can construct the two-stage FOLS (TSFOLS) estimator

$$\widetilde{\boldsymbol{\delta}}_n := \begin{bmatrix} \widetilde{\delta}_{0n} \\ \widetilde{\delta}_n \end{bmatrix} := \begin{bmatrix} 1 & \int \mathbf{g} \\ \int \mathbf{g} & \int \mathbf{g}\,\mathbf{g}' \end{bmatrix}^{-1} \begin{bmatrix} n^{-1}\sum \int \widehat{G}_i \\ n^{-1}\sum \int \widehat{G}_i\, \mathbf{g} \end{bmatrix},$$

corresponding to the two-stage functional least squares estimator in Cho, Phillips, and Seo (2020).

[2]Jain and Marcus (1975) provide sufficient conditions for FCLT for random functions G_i with various properties. For example, their theorem 1 states that given our DGP conditions, if G_i is Lipschitz continuous on Γ a.s. $-\mathbb{P}$, so that a.s. $-\mathbb{P}$, for all $\gamma, \widetilde{\gamma} \in \Gamma$, $|G_i(\gamma) - G_i(\widetilde{\gamma})| \le K_i \rho(\gamma, \widetilde{\gamma})$ for some K_i such that $E[K_i^2] < \infty$; and if for any $\epsilon \in (0, 1)$, $\int_0^\epsilon H_\rho^{1/2}(\Gamma, u)du < \infty$, then the FCLT holds, where $H_\rho(\Gamma, u) := \log[N_\rho(\Gamma, u)]$, and $N_\rho(\Gamma, u)$ is the minimal number of $\rho-$balls of radius less than or equal to u covering Γ.

When $\widehat{\theta}_n$ is consistent for θ^* and $\widetilde{G}_i$ is mildly regular, the consistency of TSFOLS follows straightforwardly. In addition, the asymptotic distribution of the TSFOLS estimator is obtained by accommodating the asymptotic distribution of the nuisance parameter estimator. To sketch the main ideas driving the asymptotic distribution result for TSFOLS, we consider

$$\frac{1}{\sqrt{n}}\sum\left[\begin{array}{c}\int \widehat{G}_i-\mu \\ \int \mathbf{g}\,(\widehat{G}_i-\mu)\end{array}\right]=\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}\,(\widehat{G}_i-\mu),$$

where $\widetilde{\mathbf{g}}:=(1,\mathbf{g}')'$. This is the analog of the term whose asymptotic distribution drives the result of Theorem 2 for FOLS.

Writing the integral on the left more explicitly and taking a mean value expansion at θ^* (interior to Θ) gives

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}(\gamma)[\widetilde{G}_i(\,\cdot\,,\gamma,\widehat{\theta}_n)-\mu(\gamma)]d\mathbb{Q}(\gamma)\\ &\quad=\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}(\gamma)[\widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)-\mu(\gamma)]d\mathbb{Q}(\gamma)+\frac{1}{n}\sum\int \widetilde{\mathbf{g}}(\gamma)[\nabla_\theta'\widetilde{G}_i(\,\cdot\,,\gamma,\bar{\theta}_{n,\gamma})]d\mathbb{Q}(\gamma)\ \sqrt{n}(\widehat{\theta}_n-\theta^*),\end{aligned}\tag{2}$$

where the mean value $\bar{\theta}_{n,\gamma}$ lies between $\widehat{\theta}_n$ and θ^* and, as indicated, depends on γ. With $G_i(\,\cdot\,,\gamma):=\widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)$, we recognize the first term as that arising for the simple FOLS estimator. The second term is new and may alter the asymptotic distribution of TSFOLS from that of FOLS.

Under mild domination conditions, the first part of the second term converges:

$$\frac{1}{n}\sum\int \widetilde{\mathbf{g}}(\gamma)[\nabla_\theta'\widetilde{G}_i(\,\cdot\,,\gamma,\bar{\theta}_{n,\gamma})]\,d\mathbb{Q}(\gamma)\to\mathbf{D}^*:=\int \widetilde{\mathbf{g}}(\gamma)\,E_{\mathbb{P}}[\nabla_\theta'\widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)]\,d\mathbb{Q}(\gamma)\ \ a.s.-\mathbb{P}.\tag{3}$$

The second part, $\sqrt{n}(\widehat{\theta}_n-\theta^*)$, generally converges in distribution.

When $E_{\mathbb{P}}[\nabla_\theta\widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)]=0$ for all $\gamma\in\Gamma$, as can happen in important special cases, then $\mathbf{D}^*=\mathbf{0}$. It is then enough that $\sqrt{n}(\widehat{\theta}_n-\theta^*)=O_{\mathbb{P}}(1)$ to ensure that

$$\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}(\gamma)[\widetilde{G}_i(\,\cdot\,,\gamma,\widehat{\theta}_n)-\mu(\gamma)]\,d\mathbb{Q}(\gamma)=\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}(\gamma)[G_i(\,\cdot\,,\gamma)-\mu(\gamma)]\,d\mathbb{Q}(\gamma)+o_{\mathbb{P}}(1),$$

in which case TSFOLS and FOLS are asymptotically equivalent and thus have the same asymptotic covariance matrix.

When $\mathbf{D}^*\neq\mathbf{0}$, then some further mild assumptions deliver a straightforward result. Specifically, suppose that $\widehat{\theta}_n$ is asymptotically linear in the sense that $\sqrt{n}[\widehat{\theta}_n-\theta^*]=-H^{*-1}\sqrt{n}s_n^*+o_{\mathbb{P}}(1)$, where H^* is a nonstochastic finite nonsingular $m\times m$ matrix and s_n^* is an $m\times 1$ random vector such that for some nonstochastic finite symmetric positive semi-definite $m\times m$ matrix I^*, $\sqrt{n}s_n^*\overset{\mathrm{A}}{\sim}N(\mathbf{0},I^*)$. Many estimators used in practice are asymptotically linear. Examples include quasi-maximum likelihood estimators, GMM estimators, and estimators based on U-statistics. In this case,

$$\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}(\gamma)[\widetilde{G}_i(\,\cdot\,,\gamma,\widehat{\theta}_n)-\mu(\gamma)]\,d\mathbb{Q}(\gamma)=\frac{1}{\sqrt{n}}\sum\int \widetilde{\mathbf{g}}(\gamma)[G_i(\,\cdot\,,\gamma)-\mu(\gamma)]\,d\mathbb{Q}(\gamma)-\mathbf{D}^*H^{*-1}\sqrt{n}s_n^*+o_P(1),$$

and an asymptotic normality result follows straightforwardly under mild conditions.

We collect together additional conditions ensuring the validity of the above heuristic arguments as follows:

Assumption 7 (DGP-B). *(i) Let Assumptions 1(i) and 2(i) hold, and let $\Theta \subset \mathbb{R}^m, m \in \mathbb{N}$, be compact;*

(ii) For $i = 1, 2, ...$, let $\widetilde{G}_i : \Omega \times \Gamma \times \Theta \mapsto \mathbb{R}$ be such that for each $\theta \in \Theta$, $\widetilde{G}_i(\,\cdot\,, \,\cdot\,, \theta)$ is measurable$-\mathcal{F} \otimes \mathcal{G}$ and IID;

(iii) Θ is convex, and for each $(\omega, \gamma) \in \Omega \times \Gamma$, $\widetilde{G}_i(\omega, \gamma, \,\cdot\,)$ is continuously differentiable on Θ, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\widetilde{G}_i(\,\cdot\,, \gamma, \theta)| \leq M_i$, and $\sup_{j=1,\dots,m} \sup_{(\gamma,\theta)\in\Gamma\times\Theta} |(\partial/\partial\theta_j)\widetilde{G}_i(\,\cdot\,, \gamma, \theta)| \leq M_i$, $i = 1, 2, ...$ □

Assumptions 7(*i* and *ii*) ensure that Assumptions 1 and 2 hold for $G_i(\,\cdot\,, \gamma) := \widetilde{G}_i(\,\cdot\,, \gamma, \theta^*)$, where θ^* is formally specified next. We use Assumption 7(*iii*) in proving consistency for the FOLS estimator, as well as in obtaining the asymptotic distribution of statistics involving $\widehat{G}_i$.

Assumption 8 (Parameter Estimator-A). *There exist $\theta^* \in \Theta$ and a sequence of measurable functions $\{\widehat{\theta}_n : \Omega \mapsto \Theta\}$ such that*

(i) $\widehat{\theta}_n \to \theta^*$ $a.s. - \mathbb{P}$;

(ii) $\theta^* \in int(\Theta)$ and

(a) $\mathbf{D}^* = \mathbf{0}$ and $\sqrt{n}(\widehat{\theta}_n - \theta^*) = O_{\mathbb{P}}(1)$; or

(b) $\mathbf{D}^* \neq \mathbf{0}$ and there exist a nonstochastic finite nonsingular $m \times m$ matrix H^* and a sequence of measurable random vectors $\{s_n^* : \Omega \mapsto \mathbb{R}^m\}$ such that $\sqrt{n}[\widehat{\theta}_n - \theta^*] = -H^{*-1}\sqrt{n}s_n^* + o_{\mathbb{P}}(1)$. □

Assumption 8(*i*) helps ensure the consistency of estimators involving $\widehat{G}_i$. Assumption 8(*ii*) plays a key role in obtaining the asymptotic distribution of statistics involving $\widehat{G}_i$.

When Assumption 8(*ii.b*) applies, we require one further condition, ensuring the joint convergence of $\sqrt{n}s_n^*$ and $n^{-1/2}\sum_{i=1}^n (G_i - \mu)$. This condition implies Assumption 6.

Assumption 9 (Joint Convergence-A). *(i) For $G_i(\,\cdot\,, \gamma) := \widetilde{G}_i(\,\cdot\,, \gamma, \theta^*)$,*

$$\begin{bmatrix} \sqrt{n}s_n^* \\ n^{-1/2}\sum_{i=1}^n (G_i - \mu) \end{bmatrix} \Rightarrow \boldsymbol{Z} := \begin{bmatrix} \mathcal{Z}_0 \\ \mathcal{Z} \end{bmatrix},$$

where $\boldsymbol{Z} : \Omega \times \Gamma \mapsto \mathbb{R}^{m+1}$ is a mean zero Gaussian process such that for $\gamma, \widetilde{\gamma} \in \Gamma$,

$$E_{\mathbb{P}}[\boldsymbol{Z}(\gamma)\boldsymbol{Z}(\widetilde{\gamma})'] = \begin{bmatrix} I^* & \kappa_0(\widetilde{\gamma}) \\ \kappa_0(\gamma)' & \kappa(\gamma, \widetilde{\gamma}) \end{bmatrix},$$

where I^* is a nonstochastic finite symmetric positive semi-definite $m \times m$ matrix; $\kappa_0 : \Gamma \mapsto \mathbb{R}^m$ belongs to $L_2(\mathbb{Q})$; and κ is as in Assumption 6; and

(ii) $\lambda_{\min}(\mathbf{B}^*) > 0$, where we let $\mathbf{B}^* := \mathbf{B} - \mathbf{D}^* H^{*-1}\mathbf{K}^* - \mathbf{K}^{*\prime} H^{*-1\prime}\mathbf{D}^{*\prime} + \mathbf{D}^* H^{*-1} I^* H^{*-1\prime}\mathbf{D}^{*\prime}$ and $\mathbf{K}^* := \int \kappa_0(\gamma)\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma)$. □

Observe that when $\mathbf{D}^* = \mathbf{0}$, we have $\mathbf{B}^* = \mathbf{B}$.

The consistency result for the TSFOLS estimator is

Theorem 3. *Given Assumptions 3, 4, 5, 7, and 8(i) for $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)$, $\widetilde{\boldsymbol{\delta}}_n \to \boldsymbol{\delta}^*$ a.s. $-\mathbb{P}$.* □

The asymptotic normality result for the TSFOLS estimator is

Theorem 4. *Given Assumptions 3, 4, 5, 6, 7, and 8(i) for $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)$,*

(i) if Assumption 8(ii.a) also holds, then $\sqrt{n}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{A}{\sim} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right)$.

(ii) If Assumptions 8(ii.b) and 9 also hold, then $\sqrt{n}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{A}{\sim} N\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}^*\mathbf{A}^{-1}\right)$. □

Note that Theorems 1 to 4 are parallel to those obtained by the (two-stage) functional least squares estimator in Cho, Phillips, and Seo (2020). Although the asymptotic consistency and distributions are obtained analogously to their estimators, the FOLS and TSFOLS estimators are obtained without supposing continuous functional observations with probability 1 as they suppose.

3.4 Consistent Asymptotic Covariance Matrix Estimation

A consistent estimator of the FOLS asymptotic covariance matrix is $\mathbf{A}^{-1}\widehat{\mathbf{B}}_n\mathbf{A}^{-1}$, where $\widehat{\mathbf{B}}_n$ is a consistent estimator for $\mathbf{B}$. Unlike the situation for standard regression estimation, we do not need to estimate $\mathbf{A}$, as it is known.

Let the functional regression estimated residuals $\widehat{\varepsilon}_{in} : \Omega \times \Gamma \mapsto \mathbb{R}$ be defined by

$$\widehat{\varepsilon}_{in}(\,\cdot\,,\gamma) := G_i(\,\cdot\,,\gamma) - \widehat{\delta}_{0n} - \mathbf{g}(\gamma)'\widehat{\delta}_n.$$

For convenience, we write $\widehat{\varepsilon}_{in}(\gamma)$ as a shorthand for $\widehat{\varepsilon}_{in}(\,\cdot\,,\gamma)$. We consider estimators of the form

$$\widehat{\mathbf{B}}_n := n^{-1}\sum_{i=1}^{n}\begin{bmatrix} \int\int \widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) & \int\int \widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})\mathbf{g}(\widetilde{\gamma})'d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \\ \int\int \mathbf{g}(\gamma)\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) & \int\int \mathbf{g}(\gamma)\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})\mathbf{g}(\widetilde{\gamma})'d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \end{bmatrix}.$$

To ensure the consistency of this estimator, we add the following assumption:

Assumption 10 (FOLS Covariance Matrix Estimation)**.** $\sup_{(\gamma,\widetilde{\gamma})\in\Gamma\times\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n} G_i(\gamma)G_i(\widetilde{\gamma}) - E_{\mathbb{P}}[G_i(\gamma)G_i(\widetilde{\gamma})]\right| \to 0$ *a.s.* $-\mathbb{P}$. □

If we take all of these conditions together, Assumptions 1, 2, 3, 4, 5, 6, and 10 are the functional regression analogs of conditions for heteroskedasticity-consistent covariance estimation (cf. White, 2001, ch.6). Formally, we have

Theorem 5. *Given Assumptions 1, 2, 3, 4, 5, 6, and 10, $\widehat{\mathbf{B}}_n \to \mathbf{B}$ $a.s. - \mathbb{P}$.* □

For the TSFOLS estimator, we use the second-stage residuals $\widetilde{\varepsilon}_{in} : \Omega \times \Gamma \mapsto \mathbb{R}$ defined by

$$\widetilde{\varepsilon}_{in}(\,\cdot\,,\gamma) := \widehat{G}_i(\,\cdot\,,\gamma) - \widetilde{\delta}_{0n} - \mathbf{g}(\gamma)'\widetilde{\delta}_n.$$

When TSFOLS and FOLS are asymptotically equivalent, we simply replace $\widehat{\varepsilon}_{in}$ with $\widetilde{\varepsilon}_{in}$ in the formula for $\widehat{\mathbf{B}}_n$ above, and denote this $\widetilde{\mathbf{B}}_n$.

Otherwise, we construct the estimator

$$\widetilde{\mathbf{B}}_n^* := \widetilde{\mathbf{B}}_n - \widetilde{\mathbf{D}}_n\widehat{H}_n^{-1}\widetilde{\mathbf{K}}_n - \widetilde{\mathbf{K}}_n{}'\widehat{H}_n'^{-1}\widetilde{\mathbf{D}}_n' + \widetilde{\mathbf{D}}_n\widehat{H}_n^{-1}\widehat{I}_n\widehat{H}_n'^{-1}\widetilde{\mathbf{D}}_n',$$

where we let

$$\widetilde{\mathbf{D}}_n := \frac{1}{n}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}(\gamma)\,\nabla_\theta'\widetilde{G}_i(\,\cdot\,,\gamma,\widehat{\theta}_n)\,d\mathbb{Q}(\gamma), \quad \widetilde{\mathbf{K}}_n := \frac{1}{n}\sum_{i=1}^{n}\int s_i(\,\cdot\,,\widehat{\theta}_n)\widetilde{\varepsilon}_{in}(\,\cdot\,,\gamma)\widetilde{\mathbf{g}}(\gamma)'\,d\mathbb{Q}(\gamma), \quad \text{and}$$

$$\widehat{I}_n := \frac{1}{n}\sum_{i=1}^{n} s_i(\,\cdot\,,\widehat{\theta}_n)s_i(\,\cdot\,,\widehat{\theta}_n)'$$

such that $s_i : \Omega \times \Theta \mapsto \mathbb{R}^m$, $\sqrt{n}s_n^* = n^{-1/2}\sum_{i=1}^n s_i(\,\cdot\,,\theta^*) + o_{\mathbb{P}}(1)$, and $\widehat{H}_n$ is a consistent estimator of H^*. For example, $\widehat{H}_n = n^{-1}\sum_{i=1}^n \nabla s_i(\,\cdot\,,\widehat{\theta}_n)$.

We provide further conditions ensuring the consistency of $\widetilde{\mathbf{B}}_n$ and $\widetilde{\mathbf{B}}_n^*$ as follows:

Assumption 11 (Joint Convergence-B)**.** *(i) For $i = 1, 2, ...$, there exists $s_i : \Omega \times \Theta \mapsto \mathbb{R}^m$ such that $s_i(\,\cdot\,,\theta)$ is measurable$-\mathcal{F}$ for each $\theta \in \Theta$ and $s_i(\omega,\,\cdot\,)$ is continuous on Θ for all $\omega \in F \in \mathcal{F}$, $\mathbb{P}(F) = 1$; $\sqrt{n}s_n^* = n^{-1/2}\sum_{i=1}^n s_i(\,\cdot\,,\theta^*) + o_{\mathbb{P}}(1)$; and $\widehat{I}_n \to I^*$ a.s. $-\,\mathbb{P}$; and*

(ii) For $n = 1, 2, ...$, there exists $\widehat{H}_n : \Omega \mapsto \mathbb{R}^{m\times m}$ such that $\widehat{H}_n$ is measurable$-\mathcal{F}$ and $\widehat{H}_n \to H^*$ a.s. $-\,\mathbb{P}$. □

Assumption 12 (TSFOLS Covariance Matrix Estimation)**.**

(i) $\sup_{(\gamma,\widetilde{\gamma},\theta)\in\Gamma\times\Gamma\times\Theta}\left|n^{-1}\sum\widetilde{G}_i(\gamma,\theta)\widetilde{G}_i(\widetilde{\gamma},\theta) - E_{\mathbb{P}}[\widetilde{G}_i(\gamma,\theta)\widetilde{G}_i(\widetilde{\gamma},\theta)]\right| \to 0$ *a.s.*$-\mathbb{P}$*;*

(ii) For each $\gamma \in \Gamma$, $\sup_{\theta\in\Theta}\left|n^{-1}\sum s_i(\theta)\widetilde{G}_i(\gamma,\theta) - E_{\mathbb{P}}[s_i(\theta)\widetilde{G}_i(\gamma,\theta)]\right| \to 0$ *a.s.*$-\mathbb{P}$. □

Note that Assumption 12(*i*) implies Assumption 10, because $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)$. Assumption 12(*ii*) helps ensure the consistency of $\widetilde{\mathbf{K}}_n$.

We can now state the desired consistency results:

Theorem 6. *(i) Given Assumptions 3, 4, 5 for $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)$, 7, 8(i), and 12(i), $\widetilde{\mathbf{B}}_n \to \mathbf{B}$ a.s. $-\mathbb{P}$;*

(ii) Given Assumptions 3, 4, 5 for $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)$, 7, 8, 9, 11, and 12, $\widetilde{\mathbf{B}}_n^* \to \mathbf{B}^*$ a.s. $-\mathbb{P}$. □

4 Hypothesis Testing

In this section, we describe the properties of Wald tests for our hypotheses of interest, $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$. We consider behavior under the null and global alternative hypotheses, as well as behavior under natural local alternatives. Because of the foundations provided by the previous sections, our next results follow as straightforward applications of standard arguments. It is necessary, however, to exercise care in specifying the null and alternative hypotheses.

4.1 The Wald Test under Null and Global Alternative Hypotheses

To construct Wald test statistics for our hypotheses of interest, $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$, we define selection matrices $S_1 := \mathbf{I}_{k+1}$ and $S_2 := [\mathbf{0}_k, \mathbf{I}_k]$, where $\mathbf{I}_{k+1}$ is the identity matrix of order $k+1$ and $\mathbf{0}_k$ is the $k \times 1$ vector of zeros. As discussed above, $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$ respectively imply

$$\mathbb{H}_{1o}(\mathbf{g}) : S_1 \boldsymbol{\delta}^* = \mathbf{0}_{k+1} \quad \text{and} \quad \mathbb{H}_{2o}(\mathbf{g}) : S_2 \boldsymbol{\delta}^* = \mathbf{0}_k.$$

The indicated dependence on $\mathbf{g}$ reflects the fact that these hypotheses are implications of $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$. They generally are not identical to $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$, as, e.g., $\mathbb{H}_{1o}(\mathbf{g})$ could hold, even if $\mathbb{H}_{1o}$ fails.

We express the global alternatives as

$$\mathbb{H}_{1A}(\mathbf{g}) : S_1 \boldsymbol{\delta}^* \neq \mathbf{0}_{k+1} \quad \text{and} \quad \mathbb{H}_{2A}(\mathbf{g}) : S_2 \boldsymbol{\delta}^* \neq \mathbf{0}_k.$$

Note that these are not equivalent to $\mathbb{H}_{1A}$ and $\mathbb{H}_{2A}$, respectively, due to the possibility of misspecification of the form of the functional regression under the alternative, as described above. We exhibit the explicit dependence of the global alternatives on $\mathbf{g}$ to reflect this possibility.

Wald statistics for testing $\mathbb{H}_{1o}(\mathbf{g})$ and $\mathbb{H}_{2o}(\mathbf{g})$ based on the FOLS estimator are

$$\mathcal{W}_{j,n} := n \widehat{\boldsymbol{\delta}}_n' S_j' \left[S_j \mathbf{A}^{-1} \widehat{\mathbf{B}}_n \mathbf{A}^{-1} S_j' \right]^{-1} S_j \widehat{\boldsymbol{\delta}}_n, \quad j = 1, 2.$$

Wald statistics for testing $\mathbb{H}_{1o}(\mathbf{g})$ and $\mathbb{H}_{2o}(\mathbf{g})$ based on the TSFOLS estimator and using $\widetilde{\mathbf{B}}_n$ are

$$\widetilde{\mathcal{W}}_{j,n} := n \widetilde{\boldsymbol{\delta}}_n' S_j' \left[S_j \mathbf{A}^{-1} \widetilde{\mathbf{B}}_n \mathbf{A}^{-1} S_j' \right]^{-1} S_j \widetilde{\boldsymbol{\delta}}_n, \quad j = 1, 2.$$

Wald statistics for testing $\mathbb{H}_{1o}(\mathbf{g})$ and $\mathbb{H}_{2o}(\mathbf{g})$ based on the TSFOLS estimator and using $\widetilde{\mathbf{B}}_n^*$ are

$$\mathcal{W}_{j,n}^* := n \widetilde{\boldsymbol{\delta}}_n' S_j' \left[S_j \mathbf{A}_n^{-1} \widetilde{\mathbf{B}}_n^* \mathbf{A}^{-1} S_j' \right]^{-1} S_j \widetilde{\boldsymbol{\delta}}_n, \quad j = 1, 2.$$

The following results are now completely standard. We let $\mathcal{X}_k^2$ denote the standard chi-square distribution with k

degrees of freedom.

Theorem 7. *(i) Given the conditions of Theorems 2 and 5, for $j = 1, 2$,*

(a) under $\mathbb{H}_{jo}(\mathbf{g})$, $\mathcal{W}_{j,n} \overset{A}{\sim} \mathcal{X}^2_{k+2-j}$;

(b) under $\mathbb{H}_{jA}(\mathbf{g})$, $\mathbb{P}[\mathcal{W}_{j,n} \geq c_n] \to 1$ for any sequence $\{c_n\}$ s.t. $c_n = o(n)$;

(ii) Given the conditions of Theorems 4(i) and 6(i), for $j = 1, 2$,

(a) under $\mathbb{H}_{jo}(\mathbf{g})$, $\widetilde{\mathcal{W}}_{j,n} \overset{A}{\sim} \mathcal{X}^2_{k+2-j}$;

(b) under $\mathbb{H}_{jA}(\mathbf{g})$, $\mathbb{P}[\widetilde{\mathcal{W}}_{j,n} \geq c_n] \to 1$ for any sequence $\{c_n\}$ s.t. $c_n = o(n)$; and

(iii) Given the conditions of Theorem 4(ii) and 6(ii), for $j = 1, 2$,

(a) under $\mathbb{H}_{jo}(\mathbf{g})$, $\mathcal{W}^*_{j,n} \overset{A}{\sim} \mathcal{X}^2_{k+2-j}$;

(b) under $\mathbb{H}_{jA}(\mathbf{g})$, $\mathbb{P}[\mathcal{W}^*_{j,n} \geq c_n] \to 1$ for any sequence $\{c_n\}$ s.t. $c_n = o(n)$. □

4.2 The Wald Test under Local Alternatives

We consider local alternatives of the following form: $\{\mu_n\}$ is such that for some $\varsigma \in \mathbb{R}^{1+k}$,

$$\mathbb{H}_{ja}(\mathbf{g}) : \sqrt{n} S_j \boldsymbol{\delta}^*_n \to S_j \varsigma, \quad j = 1, 2,$$

where

$$\boldsymbol{\delta}^*_n := \begin{bmatrix} \delta^*_{0n} \\ \delta^*_n \end{bmatrix} := \begin{bmatrix} E_{\mathbb{Q}}[\mu_n] \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} -E_{\mathbb{Q}}[\,\mathbf{g}\,]' \mathrm{cov}_{\mathbb{Q}}[\mathbf{g}, \mathbf{g}]^{-1} \mathrm{cov}_{\mathbb{Q}}[\mathbf{g}, \mu_n] \\ \mathrm{cov}_{\mathbb{Q}}[\mathbf{g}, \mathbf{g}]^{-1} \mathrm{cov}_{\mathbb{Q}}[\mathbf{g}, \mu_n] \end{bmatrix}.$$

The required evolution of μ_n can arise from evolution of either G_i (becoming G_{in}) or $\mathbb{P}$ (becoming $\mathbb{P}_n$). As the former yields less fundamental and fairly direct modifications to the underlying regularity conditions, we adopt that approach. For brevity, however, we omit restating all the affected conditions (Assumptions 1(*ii*) , 2(*ii*), 3, 5 (which is more easily verified as a weak ULLN for triangular arrays), 6, 7(*ii*, *iii*), 8 (with weak rather than strong convergence to $\mathbf{D}^*$), 9, 10, 11, and 12 (with weak convergence)). Instead, we understand implicitly that any of these conditions referenced in the next result are replaced with their suitable analogs involving G_{in}.

The next results are again standard. We let $\mathcal{X}^2(k, \xi)$ denote the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter ξ. The following noncentrality parameters are relevant for $j = 1, 2$:

$$\xi_j := \varsigma' S'_j [S_j \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1} S'_j]^{-1} S_j \varsigma; \quad \text{and} \quad \xi^*_j := \varsigma' S'_j [S_j \mathbf{A}^{-1} \mathbf{B}^* \mathbf{A}^{-1} S'_j]^{-1} S_j \varsigma.$$

Theorem 8. *(i) Given the conditions of Theorems 2 and 5, for $j = 1, 2$, under $\mathbb{H}_{ja}(\mathbf{g})$, $\mathcal{W}_{j,n} \overset{A}{\sim} \mathcal{X}^2(k + 2 - j, \xi_j)$;*

(ii) Given the conditions of Theorems 4(i) and 6(i), for $j = 1, 2$, under $\mathbb{H}_{ja}(\mathbf{g})$, $\widetilde{\mathcal{W}}_{j,n} \overset{A}{\sim} \mathcal{X}^2(k + 2 - j, \xi_j)$; and

(iii) Given the conditions of Theorems 4(ii) and 6(ii), for $j = 1, 2$, under $\mathbb{H}_{ja}(\mathbf{g})$, $\mathcal{W}^*_{j,n} \overset{A}{\sim} \mathcal{X}^2(k + 2 - j, \xi^*_j)$. □

5 Examples

We illustrate the application of the foregoing results by returning to our examples of Section 2.

Example 1 (Panel Random Effects-Continued): Recall that interest attaches to $G_i(\gamma) = V_i(1)V_i(\gamma)$, and to testing $\mathbb{H}_{1o}$. Because the V_i's are unknown, we use a TSFOLS procedure. Specifically, we work with $\widehat{G}_i(\gamma) = \widehat{V}_i(1)\widehat{V}_i(\gamma)$, where $\widehat{V}_i(\gamma) := \widetilde{V}_i(\gamma, \widehat{\beta}_n) = Y_i(\gamma) - X_i(\gamma)'\widehat{\beta}_n$, and $\widehat{\beta}_n$ is the pooled OLS estimator,

$$\widehat{\beta}_n := \left(\sum_{i=1}^{n}\sum_{\gamma=1}^{T} X_i(\gamma)X_i(\gamma)'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{\gamma=1}^{T} X_i(\gamma)Y_i(\gamma)\right).$$

To determine which asymptotic covariance matrix applies in this case, we investigate $\mathbf{D}^* := \int \widetilde{\mathbf{g}}(\gamma)E_{\mathbb{P}}[\nabla'_{\beta}\widetilde{G}_i(\,\cdot\,,\gamma,\beta^*)]d\mathbb{Q}(\gamma)$. Now, we note that $(\partial/\partial\beta_j)\widetilde{G}_i(\,\cdot\,,\gamma,\beta^*) = [(\partial/\partial\beta_j)\widetilde{V}_i(1,\beta^*)]\widetilde{V}_i(\gamma,\beta^*) + \widetilde{V}_i(1,\beta^*)[(\partial/\partial\beta_j)\widetilde{V}_i(\gamma,\beta^*)] = -X_{ij}(1)V_i(\gamma) - V_i(1)X_{ij}(\gamma)$. Under pure random effects ($\sigma_c^2 = 0$), it then follows that for all $\gamma \in \{2,...,T\}$, $E_{\mathbb{P}}[\nabla'_{\beta}\widetilde{G}_i(\,\cdot\,,\gamma,\beta^*)] = 0$. In this case, the first-stage estimation has no effect on the asymptotic covariance matrix, and we can test for panel random effect assumption using $\widetilde{\mathcal{W}}_{1,n}$ for any desired choice of $\mathbf{g}$ and $\mathbb{Q}$. For example, we may let $\mathbf{g}(\gamma) = g_1(\gamma) = \gamma$. The TSFOLS estimator minimizes

$$\frac{0.5}{n(T-1)}\sum_{i=1}^{n}\sum_{\gamma=2}^{T}\{\widehat{V}_i(1)\widehat{V}_i(\gamma) - \delta_0 - \delta g_1(\gamma)\}^2.$$

Letting $\sum_{\gamma} = \sum_{\gamma=2}^{T}$, the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A} = \frac{1}{(T-1)}\begin{bmatrix} T-1 & \sum_{\gamma} g_1(\gamma) \\ \sum_{\gamma} g_1(\gamma) & \sum_{\gamma} g_1(\gamma)^2 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \frac{1}{(T-1)^2}\sum_{\gamma}\sum_{\widetilde{\gamma}}\begin{bmatrix} \kappa(\gamma,\widetilde{\gamma}) & \kappa(\gamma,\widetilde{\gamma})g_1(\widetilde{\gamma}) \\ g_1(\gamma)\kappa(\gamma,\widetilde{\gamma}) & g_1(\gamma)\kappa(\gamma,\widetilde{\gamma})g_1(\widetilde{\gamma}) \end{bmatrix},$$

where

$$\kappa(\gamma,\widetilde{\gamma}) := \begin{cases} E[C_i^4] + 2\sigma_c^2\sigma_u^2 + \sigma_u^4 - \sigma_c^4, & \text{if } \gamma = \widetilde{\gamma}; \\ E[C_i^4] + \sigma_c^2\sigma_u^2 - \sigma_c^4, & \text{otherwise.} \end{cases}$$

The conditions of Theorem 6(i) apply to deliver the consistency of $\widetilde{\mathbf{B}}_n$ for $\mathbf{B}$. □

Example 2 (Specification Testing - Continued): For specificity, suppose that $d = 2$, $X_i := (X_{i1}, X_{i2})' := (1, X_{2i})'$ and that $E_{\mathbb{P}}[Y_i|X_i] = \pi^* \exp(X_{2i})$. Next, take $f(X,\theta) = \theta_1 + \theta_2 X_2$, so that $\mathbb{M}$ is correctly specified for $E_{\mathbb{P}}[Y_i|\mathbf{X}_i]$ only when $\pi^* = 0$.

Finally, take ψ to be the logistic function, $\psi(z) = 1/[1+\exp(-z)]$, let $\gamma \in \Gamma := [\underline{\gamma}, \bar{\gamma}]$, and let $\mathbb{Q}$ be the uniform distribution on Γ. These specification tests require a first stage estimator, so our results for the TSFOLS estimator will apply. Given the linear structure of $\mathbb{M}$, we take $\widehat{\theta}_n := (\widehat{\theta}_{1n}, \widehat{\theta}_{2n})'$ to be the OLS estimator. We thus work with

$$\widehat{G}_i(\gamma) = [Y_i - \widehat{\theta}_{1n} - \widehat{\theta}_{2n} X_{2i}]\psi(X_{2i}\gamma).$$

The TSFOLS estimator is obtained by choosing $\widetilde{\delta}_{0n}$ and $\widetilde{\delta}_n$ to minimize

$$\frac{1}{2n}\sum_{i=1}^{n}\frac{1}{(\bar{\gamma}-\underline{\gamma})}\int_{\underline{\gamma}}^{\bar{\gamma}}\{\widehat{G}_i(\gamma) - \delta_0 - \mathbf{g}(\gamma)'\delta\}^2 d\gamma,$$

where $\mathbf{g}$ is suitably chosen function.

The theory of the foregoing sections for TSFOLS applies directly. To determine which version of the TSFOLS asymptotic covariance matrix is required, we investigate $\mathbf{D}^* := \int \widetilde{\mathbf{g}}(\gamma) E_{\mathbb{P}}[\nabla_\theta' \widetilde{G}_i(\,\cdot\,,\gamma,\theta^*)]\, d\mathbb{Q}(\gamma) = (\bar{\gamma}-\underline{\gamma})^{-1}\int_{\underline{\gamma}}^{\bar{\gamma}} \widetilde{\mathbf{g}}(\gamma)$ $E_{\mathbb{P}}[(-1,-X_2)\psi(X_2\gamma)]d\gamma$. Inspecting this, we do not see that it vanishes in general, so we must estimate $\mathbf{B}^*$ to compute our test statistic. This estimation involves computation of

$$\widetilde{\mathbf{D}}_n = (\bar{\gamma}-\underline{\gamma})^{-1}\frac{1}{n}\sum_{i=1}^{n}\int_{\underline{\gamma}}^{\bar{\gamma}}\widetilde{\mathbf{g}}(\gamma)(-1,-X_{2i})\psi(X_{2i}\gamma)d\gamma, \qquad \widetilde{\mathbf{K}}_n = (\bar{\gamma}-\underline{\gamma})^{-1}\frac{1}{n}\sum_{i=1}^{n}\int_{\underline{\gamma}}^{\bar{\gamma}} s_i(\,\cdot\,,\widehat{\theta}_n)\widetilde{\varepsilon}_{in}(\,\cdot\,,\gamma)\widetilde{\mathbf{g}}(\gamma)'d\gamma,$$

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} s_i(\,\cdot\,,\widehat{\theta}_n)s_i(\,\cdot\,,\widehat{\theta}_n)', \quad \text{and} \quad \widehat{H}_n = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} -1 \\ -X_{2i}\end{bmatrix}[-1,-X_{2i}],$$

where

$$s_i(\,\cdot\,,\widehat{\theta}_n) = \begin{bmatrix} -1 \\ -X_{2i}\end{bmatrix}[Y_i - \widehat{\theta}_{1n} - \widehat{\theta}_{2n}X_{2i}], \quad \text{and} \quad \widetilde{\varepsilon}_{in}(\,\cdot\,,\gamma) = \widehat{G}_i(\gamma) - \widetilde{\delta}_{0n} - \mathbf{g}(\gamma)'\widetilde{\delta}_n.$$

Here the relevant hypothesis is the hypothesis of correct specification, corresponding to $\mathbb{H}_{1o}$. We thus compute $\mathcal{W}_{1,n}^*$ as specified above.

To examine further features of our test, suppose that we knew that the DGP exhibits conditional heteroskedasticity, such that $U_i = h(X_{2i})\varepsilon_i$, where $U_i := Y_i - E_{\mathbb{P}}[Y_i|X_i]$, where $h(x) = \sin(x)$, and ε_i is IID with $E_{\mathbb{P}}(\varepsilon_i|X_{i2}) = 0$ and $E_{\mathbb{P}}(\varepsilon_i^2|X_{i2}) = 1$, and that $(X_{2i},\varepsilon_i)' \sim$ IID $N((1,0)',\mathbf{I}_2)$. Applying theorem 3 of Bierens (1990) tells us that under $\mathbb{H}_{1o}$, $n^{-1/2}\sum_{i=1}^{n}\widehat{G}_i \Rightarrow \mathcal{Z}$, a zero mean Gaussian process having the covariance structure

$$\kappa(\gamma,\widetilde{\gamma}) = E_{\mathbb{P}}[\sin(X_2)^2(\psi(X_2\gamma) - X'E_{\mathbb{P}}[XX']^{-1}E_{\mathbb{P}}[X\psi(X_2\gamma)])(\psi(X_2\widetilde{\gamma}) - X'E_{\mathbb{P}}[XX']^{-1}E_{\mathbb{P}}[X\psi(X_2\widetilde{\gamma})])].$$

The complexity of this structure makes it difficult to exploit, even under the best circumstances, where we have detailed knowledge of the DGP. In applications, matters are worse as h and the unconditional distribution of X_i are typically unknown a priori. Fortunately, however, our approach here does not require explicitly taking into account the structure of κ, just as tests based on a heteroskedasticity-robust estimator do not require explicitly taking into account the unknown heteroskedasticity.

The tests suggested by Bierens (1990) and SW rely on statistics computed as functionals of

$$\frac{1}{\sqrt{n}}\sum \widehat{G}_i(\gamma) = \frac{1}{\sqrt{n}}\sum [Y_i - \widehat{\theta}_{1n} - \widehat{\theta}_{2n} X_{2i}]\psi(X_{2i}\gamma).$$

These statistics have asymptotic distributions that are generally highly complex, varying for different choices of ψ and for different choice of functional. This distribution typically must be simulated in each case, requiring considerable computational effort in computing the critical values; or a special functional has to be selected to obtain a statistic with asymptotically standard null distribution, as pointed out by Bierens (1990). The benefit of the approach taken here is that our test statistics always have a straightforward asymptotic chi-square distribution regardless of ψ, $\mathbf{g}$, or $\mathbb{Q}$. $\square$

6 Monte Carlo Experiments

In this section, we conduct Monte Carlo experiments using our Wald tests with the DGPs specified in our previous examples. First, we investigate the behavior of functional regression tests for panel data random effects and compare these to a Breusch-Pagan (1979) test. As the panel setting is standard and familiar, these results are intended primarily to illustrate how this familiar setting maps to the functional regression framework, rather than to yield new insights for panel data. Second, we compare the specification tests of Bierens (1990) and SW to our functional regression Wald tests. Here, functional regression offers not only computational convenience, but we also observe some interesting power advantages.

6.1 Example 1: Panel Random Effects

For the panel random effects example, let $d = 2$ and $T = 20$, so that $j \in \{1, 2\}$ and $\gamma \in \{1, 2, ..., T\}$ for $i \in \{1, 2, ..., n\}$. Let $X_{ji}(\gamma)$ be IID $\mathcal{X}_1^2$, and let $U_i(\gamma)$ be such that $U_i(\gamma) + 3 \sim$ IID $\mathcal{X}_3^2$. Thus, for each γ, $E[U_i(\gamma)] = 0$, and the $U_i(\gamma)$'s have a non-normal distribution.

As discussed above, the choice of $\mathbf{g}$ is up to the researcher. Here we consider five different possibilities. The simplest choice omits $\mathbf{g}$ entirely, and simply tests for a zero intercept, coinciding with a standard quasi-maximum likelihood procedure. The remaining choices are linear ($g_1(\gamma) = \gamma$), quadratic ($g_1(\gamma) = \gamma^2$), linear-quadratic ($g_1(\gamma) = \gamma, g_2(\gamma) = \gamma^2$), and geometric ($g_1(\gamma) = 0.5^\gamma$). The latter choice is one a researcher might make if autocorrelation in the $U_i(\gamma)$'s were suspected. We make these choices primarily because of their simplicity. Nevertheless, under the alternative in which $\sigma_c^2 > 0$, μ is just a constant function different from zero. This implies that the functional regression coefficients for the elements of $\mathbf{g}$ will be zero; including $\mathbf{g}$ will thus result in some loss of power. Our experiments with $\mathbf{g}$ included permit us to assess this loss. We denote the Wald statistics for these choices as $\widetilde{\mathcal{W}}_{1,n}$(con), $\widetilde{\mathcal{W}}_{1,n}$(con+lin), $\widetilde{\mathcal{W}}_{1,n}$(con+quad), $\widetilde{\mathcal{W}}_{1,n}$(con+lin+quad), and $\widetilde{\mathcal{W}}_{1,n}(\text{con+}0.5^\gamma)$, respectively.

We also apply the Breusch-Pagan (1979) statistic to test the null of pure random effects structure. This statistic is popularly used to test for unobserved fixed effects, as noted by Wooldridge (2002), and can be written as

$$\mathcal{BP}_n := \left\{ \frac{\sum_{i=1}^{n} \sum_{\gamma=2}^{T} \widehat{V}_i(1)\widehat{V}_i(\gamma)}{\sqrt{\sum_{i=1}^{n} \{\sum_{\gamma=2}^{T} \widehat{V}_i(1)\widehat{V}_i(\gamma)\}^2}} \right\}^2$$

in our context. Under the null, $\sigma_c^2 = 0$ and there is no correlation between $\widehat{G}_i(\gamma)$ and $\widehat{G}_i(\widetilde{\gamma})$ when $\gamma \neq \widetilde{\gamma}$. Thus, $\mathcal{BP}_n$ follows the chi-square distribution with one degree of freedom. On the other hand, the alternative $\sigma_c^2 > 0$ leads to serial correlation, so that $\mathcal{BP}_n$ yields a consistent test.

Tables 1 and 2 display the simulation results for level (10,000 replications) and power (5,000 replications), respectively. We examine power patterns by varying the sample size and the values of σ_c^2 for the alternatives. As expected, the levels of the Wald statistics are well behaved. $\mathcal{BP}_n$ also shows good level behavior. Both $\widetilde{\mathcal{W}}_{1,n}(\text{con})$ and $\mathcal{BP}_n$ have comparable power, with $\widetilde{\mathcal{W}}_{1,n}(\text{con})$ having perhaps a small advantage. As expected, the inclusion of the additional regressors generally leads to modest losses in power, with (as expected) greater losses for $\widetilde{\mathcal{W}}_{1,n}(\text{con+lin+quad})$, which uses three degrees of freedom, than for the others, which use only two degrees of freedom. Although these power losses are modest, these results underscore the importance of using knowledge about the alternative to arrive at a parsimonious functional regression.

6.2 Example 2: Specification Testing

To test the hypotheses $\mathbb{H}_{1o}(\mathbf{g})$ vs. $\mathbb{H}_{1A}(\mathbf{g})$ for the specification tests of Example 2, we again consider the case of functional regression with a constant only, together with the linear, quadratic, and linear-quadratic cases. We denote the Wald statistics for these cases as $\mathcal{W}^*_{1,n}(\text{con})$, $\mathcal{W}^*_{1,n}(\text{con+lin})$, $\mathcal{W}^*_{1,n}(\text{con+quad})$, and $\mathcal{W}^*_{1,n}(\text{con+lin+quad})$, respectively. As in Example 2, the associated integrals are computed using Gauss-Legendre quadrature, now letting $\Gamma = [\underline{\gamma}, \bar{\gamma}] = [-0.5, 0.5]$ with ψ the logistic function, as before.

In addition, we compute test statistics suggested by Bierens (1990) and SW, letting $\mathcal{B}_n$ and $\mathcal{SW}_n$ denote the Bierens and SW test statistics, respectively. For $\mathcal{B}_n$, we follow theorem 4 of Bierens (1990) and let $\gamma = 1$, $\rho = 0.5$, and $t_0 = 1/4$. These parameters must be selected by the researcher before conducting the Bierens test and are those used by Bierens (1990, table 1) for his own Monte Carlo experiments. For comparability, we again take ψ to be the logistic function. Because of the particular structure imposed here, $\mathcal{B}_n$ is distributed asymptotically as $\mathcal{X}_1^2$ under the null.

SW give a simple consistent test procedure using critical values based on the law of the iterated logarithm (LIL) bound. This is quite conservative, as SW point out. We follow their theorem 5.6(a) and let the associated norm be the uniform norm, with ψ again chosen to be the logistic function. SW's LIL procedure yields a test for which the level declines to zero as n increases. For comparability, we scale the LIL-based critical value to yield a level of 5% for $n = 100$. For $n = 100$, the ratio between the LIL-based critical value and the quantile yielding a 5% empirical rejection is 2.2405. We then multiply the other LIL-based critical values for the different sample sizes by this ratio.

Tables 3 and 4 present simulation results for level (10,000 replications) and power (5,000 replications). In Table 3, we see that the Wald tests and $\mathcal{B}_n$ have approximately correct levels. As the sample size increases, the levels appear to converge to their nominal values. As expected, the level for $\mathcal{SW}_n$ decreases with n.

In Table 4, we examine power by varying the sample size and the coefficient π^* (recall that above we specified $E[Y_i|X_i] = \pi^* \exp(X_{2i})$). First, we again see very strong performance for tests based on $\mathcal{W}^*_{1,n}(\text{con})$. Nevertheless, jointly including linear and quadratic functions of γ in the functional regression (using $\mathcal{W}^*_{1,n}(\text{con+lin+quad})$) is now seen to pay off, especially for all but the smaller values of π^*, with relative improvement most noticeable for the smaller sample sizes. We note that results for $\mathcal{W}^*_{1,n}(\text{con+lin})$ and $\mathcal{W}^*_{1,n}(\text{con+quad})$ are similar to each other and are not as good as those for $\mathcal{W}^*_{1,n}(\text{con+lin+quad})$.

Interestingly, we find that $\mathcal{W}^*_{1,n}(\text{con})$ strongly dominates $\mathcal{B}_n$, especially for smaller values of π^*. For $n \geq 100$ (where levels are comparable) we also see the conservative $\mathcal{SW}_n$ test dominating $\mathcal{B}_n$. For these sample sizes, $\mathcal{SW}_n$ performs comparably to $\mathcal{W}^*_{1,n}(\text{con})$ and $\mathcal{W}^*_{1,n}(\text{con+lin+quad})$. Nevertheless, the utility of the $\mathcal{SW}_n$ statistic is limited by the need to find a practical way to control its level.

Overall, these results demonstrate the appeal of the functional regression Wald tests for specification testing. Not only are they easy to apply because of their standard chi-square asymptotic distribution, but they can have power as good or better than previous procedures, such as tests based on $\mathcal{B}_n$ or $\mathcal{SW}_n$.

7 Conclusion

In this paper, we study functional regression and its properties in testing the hypothesis of a constant zero mean function or an unknown constant non-zero mean function by applying the approach in Cho, Phillips, and Seo (2020). As we show, the associated Wald test statistics have standard chi-square limiting null distributions, standard non-central chi-square distributions for local alternatives converging to zero at a $\sqrt{n}$ rate, and are consistent against global alternatives. These properties permit the construction of straightforward tests of the hypotheses of interest.

As we discuss, panel data can be viewed as functional data; we illustrate this with a running example focusing on a test of random effects structure. In particular, we develop new alternatives to tests for regression misspecification, both of which involve nuisance parameters identified only under the alternative. We find that our procedures can have power better than existing methods that do not exploit this covariance structure, like the specification testing procedures of Bierens (1982, 1990) or SW. Interestingly, we find that functional regression tests including only a constant have remarkably good power, even when the functional mean depends non-trivially on its parameter. This suggests that any battery of tests for a zero mean function should include tests based on the intercept only, and that tests including additional functions of the parameter should be judiciously constructed.

Finally, we note that functional regression tests may have utility in a variety of disparate contexts involving hypothesis testing with multiple statistics. For example, Tippet (1931), Fisher (1932), Pearson (1950), Lancaster (1961), van Zwet and Oosterhoff (1967), Westberg (1985), and the references therein consider combining a finite number of multiple statistics using a specified weighting method or a Bayes method. Our approach accommodates such methods, allowing dependence among multiple statistics. It further allows not just a finite number of tests, but allows the tests to be indexed by elements of a multidimensional continuum.

8 Appendix: Proofs

Proof of Theorem 1: The given consistency easily follows by applying the DCT given Assumptions 3, 4, and 5. We note that Assumption 3 implies that

$$\left|\sum_{i=1}^{n} G_i\right| \le \sum_{i=1}^{n} G_i^2 \le \sum_{i=1}^{n} M_i^2 < \infty \text{ a.s.} - \mathbb{P} \text{ and } \left|\sum_{i=1}^{n} G_i g_j\right| \le \sum_{i=1}^{n} G_i^2 g_j^2 \le \sum_{i=1}^{n} M_i^2 g_j^2$$

for every j, so that

$$\int \left| n^{-1}\sum_{i=1}^{n} G_i\right| d\mathbb{Q} < n^{-1}\sum_{i=1}^{n} M_i^2 < \infty \text{ and } \int \left| n^{-1}\sum_{i=1}^{n} G_i g_j\right| d\mathbb{Q} \le n^{-1}\sum_{i=1}^{n} M_i^2 \int g_j^2 d\mathbb{Q} < \infty$$

a.s. $-\mathbb{P}$, as $g_j \in L_2(\mathbb{Q})$ by 4(ii). This implies that we can first let n tend to infinity before integrating the associated random functions, so that

$$\begin{bmatrix} n^{-1}\sum \int G_i - \int \mu \\ n^{-1}\sum \int G_i \mathbf{g} - \int \mu \mathbf{g} \end{bmatrix} = \begin{bmatrix} \int n^{-1}\sum G_i - \int \mu \\ \int n^{-1}\sum G_i \mathbf{g} - \int \mu \mathbf{g} \end{bmatrix} \to \begin{bmatrix} 0 \\ \mathbf{0} \end{bmatrix} \text{ a.s.} - \mathbb{P},$$

where the given convergence follows from 5. Thus, we obtain that

$$\widehat{\boldsymbol{\delta}}_n := \begin{bmatrix} \widehat{\delta}_{0n} \\ \widehat{\delta}_n \end{bmatrix} := \begin{bmatrix} 1 & \int \mathbf{g} \\ \int \mathbf{g} & \int \mathbf{g}\,\mathbf{g}' \end{bmatrix}^{-1} \begin{bmatrix} n^{-1}\sum \int G_i \\ n^{-1}\sum \int G_i\, \mathbf{g} \end{bmatrix} \to \begin{bmatrix} 1 & \int \mathbf{g} \\ \int \mathbf{g} & \int \mathbf{g}\,\mathbf{g}' \end{bmatrix}^{-1} \begin{bmatrix} \int \mu \\ \int \mu \mathbf{g} \end{bmatrix} =: \begin{bmatrix} \delta_0^* \\ \delta^* \end{bmatrix} =: \boldsymbol{\delta}^*$$

a.s. $-\mathbb{P}$. ■

Proof of Theorem 2: From the note that

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) = \begin{bmatrix} 1 & \int \mathbf{g} \\ \int \mathbf{g} & \int \mathbf{g}\,\mathbf{g}' \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum \int (G_i - \mu) \\ n^{-1/2}\sum \int (G_i - \mu)\, \mathbf{g} \end{bmatrix},$$

the desired result follows if

$$\begin{bmatrix} n^{-1/2}\sum \int (G_i - \mu) \\ n^{-1/2}\sum \int (G_i - \mu)\, \mathbf{g} \end{bmatrix} \overset{\text{A}}{\sim} N\left(\mathbf{0}, \mathbf{B}\right), \tag{4}$$

the desired result follows. Assumption 6(*ii*) implies that $n^{-1/2}\sum(G_i-\mu)\Rightarrow\mathcal{G}$, so that we obtain $n^{-1/2}\sum\int(G_i-\mu)\Rightarrow\int\mathcal{G}$, and for each $j\in\{1,2,\ldots,k\}$, $\int(G_i-\mu)g_j\Rightarrow\int\mathcal{G}g_j$ by the continuous mapping theorem. Also, we note that $\int\mathcal{G}$ and $\int\mathcal{G}g_j$ ($j\in\{1,2,\ldots,k\}$) are the integrals of Gaussian processes, so that they are normally distributed with

$$\int\mathcal{G}\sim N\left(0,\int\int\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\right)\quad\text{and}\quad\int\mathcal{G}g_j\sim N\left(0,\int\int g_j(\gamma)\kappa(\gamma,\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\right),\tag{5}$$

where the given variances are computed by applying theorem 2 of Grenander (1981, p. 48). Given this, the positive definite matrix $\mathbf{B}$ in Assumption 6(*iii*) enables us to apply the Cramér-Wold's device, which we omit for brevity. This completes the proof. ■

Proof of Theorem 3: The given consistency can be achieved in a parallel manner to that of Theorem 1. We note that Assumption 7(*ii*) implies that

$$\left|\sum_{i=1}^{n}\widetilde{G}_i\right|\leq\sum_{i=1}^{n}\widetilde{G}_i^2\leq\sum_{i=1}^{n}M_i^2<\infty\ \text{ a.s.}\ -\mathbb{P}\ \text{ and }\ \left|\sum_{i=1}^{n}\widetilde{G}_ig_j\right|\leq\sum_{i=1}^{n}\widetilde{G}_i^2g_j^2\leq\sum_{i=1}^{n}M_i^2g_j^2$$

for every j, so that

$$\int\left|n^{-1}\sum_{i=1}^{n}\widetilde{G}_i\right|d\mathbb{Q}<n^{-1}\sum_{i=1}^{n}M_i^2<\infty\text{ and }\int\left|n^{-1}\sum_{i=1}^{n}\widetilde{G}_ig_j\right|d\mathbb{Q}\leq n^{-1}\sum_{i=1}^{n}M_i^2\int g_j^2d\mathbb{Q}<\infty$$

a.s. $-\mathbb{P}$, as $g_j\in L_2(\mathbb{Q})$ by Assumption 4(*ii*). This implies that we can apply DCT, so that

$$\begin{bmatrix} n^{-1}\sum\int\widehat{G}_i-\int\mu \\ n^{-1}\sum\int\widehat{G}_i\mathbf{g}-\int\mu\mathbf{g}\end{bmatrix}=\begin{bmatrix}\int n^{-1}\sum\widehat{G}_i-\int\mu \\ \int n^{-1}\sum\widehat{G}_i\mathbf{g}-\int\mu\mathbf{g}\end{bmatrix}\to\begin{bmatrix}0\\ \mathbf{0}\end{bmatrix}\ \text{ a.s. }-\mathbb{P}.$$

The given convergence mainly follows from the facts that: (a)

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{G}_i(\gamma)-\mu\right|\leq\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{G}_i(\gamma)-\frac{1}{n}\sum_{i=1}^{n}\widetilde{G}_i(\gamma,\theta_*)\right|+\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{G}_i(\gamma,\theta_*)-\mu\right|;$$

(b) the second element in the RHS converges to zero a.s. $-\mathbb{P}$ by Assumption 5; and (c) applying the mean-value theorem implies that

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{G}_i(\gamma)-\frac{1}{n}\sum_{i=1}^{n}\widetilde{G}_i(\gamma,\theta^*)\right|=\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta\widetilde{G}_i(\gamma,\bar{\theta}_{n,\gamma})(\widehat{\theta}_n-\theta^*)\right|,$$

where the RHS converges to zero a.s. $-\mathbb{P}$ by and Assumptions 7(*iii*) and 8(*i*). Thus, we obtain that

$$\widetilde{\boldsymbol{\delta}}_n:=\begin{bmatrix}\widetilde{\delta}_{0n}\\ \widetilde{\delta}_n\end{bmatrix}:=\begin{bmatrix}1 & \int\mathbf{g}\\ \int\mathbf{g} & \int\mathbf{g}\,\mathbf{g}'\end{bmatrix}^{-1}\begin{bmatrix}n^{-1}\sum\int\widehat{G}_i\\ n^{-1}\sum\int\widehat{G}_i\,\mathbf{g}\end{bmatrix}\to\begin{bmatrix}1 & \int\mathbf{g}\\ \int\mathbf{g} & \int\mathbf{g}\,\mathbf{g}'\end{bmatrix}^{-1}\begin{bmatrix}\int\mu\\ \int\mu\mathbf{g}\end{bmatrix}=:\begin{bmatrix}\delta_0^*\\ \delta^*\end{bmatrix}=:\boldsymbol{\delta}^*$$

a.s. $-\mathbb{P}$. This completes the proof. ■

Proof of Theorem 4: We explicitly prove only 5(*ii*). The proof for 5(*i*) is quite similar.

(*ii*) From the given fact that

$$\sqrt{n}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) = \sqrt{n}\begin{bmatrix} \widetilde{\delta}_{0n} - \delta_0^* \\ \widetilde{\delta}_n - \delta^* \end{bmatrix} = \begin{bmatrix} 1 & \int \mathbf{g} \\ \int \mathbf{g} & \int \mathbf{g}\,\mathbf{g}' \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum \int (\widehat{G}_i - \mu) \\ n^{-1/2}\sum \int (\widehat{G}_i - \mu)\,\mathbf{g} \end{bmatrix},$$

the desired result follows if

$$\frac{1}{\sqrt{n}}\sum \begin{bmatrix} \int (\widehat{G}_i - \mu) \\ \int (\widehat{G}_i - \mu)\,\mathbf{g} \end{bmatrix} \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{B}^*). \tag{6}$$

Given this, we note that applying the mean-value theorem in (2) and Assumption 9 yields that

$$\frac{1}{\sqrt{n}}\int \sum_{i=1}^{n} \widetilde{\mathbf{g}}(\widehat{G}_i - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}(G_i - \mu) + \frac{1}{n}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}[\nabla_\theta' \widetilde{G}_i(\bar{\theta}_{n,\gamma})]\sqrt{n}(\widehat{\theta}_n - \theta^*) \Rightarrow \int \widetilde{\mathbf{g}}\mathcal{G} - \mathbf{D}^* H^{*-1}\mathcal{Z}_0 \tag{7}$$

because (*i*) $n^{-1/2}\sum \int (G_i - \mu) \Rightarrow \int \mathcal{G}$, and for each $j \in \{1, 2, \ldots, k\}$, $\int (G_i - \mu) g_j \Rightarrow \int \mathcal{G} g_j$ by the continuous mapping theorem; and (*ii*) for $j = 1, 2, \ldots, k$ and $\widetilde{j} = 1, 2, \ldots, m$,

$$\sup_{\gamma,\theta}\left| n^{-1}\sum_{i=1}^{n} \frac{\partial}{\partial \theta_{\widetilde{j}}} \widetilde{G}_i(\gamma, \theta)\right| \le \left(n^{-1}\sum_{i=1}^{n} M_i^2\right)^{1/2} < \infty \;\text{ a.s.} \; -\mathbb{P}, \quad \text{and}$$

$$\sup_{\gamma,\theta}\left| n^{-1}\sum_{i=1}^{n} \frac{\partial}{\partial \theta_{\widetilde{j}}} \widetilde{G}_i(\gamma, \theta) g_j(\gamma)\right| \le \left(n^{-1}\sum_{i=1}^{n} M_i^2\right)^{1/2} \times \left(n^{-1}\sum_{i=1}^{n} M_i^2\right)^{1/2} < \infty \;\text{ a.s.} \; -\mathbb{P}$$

by Assumption 8, so that we can let n tend to infinity first before computing the associated integrals by the DCT, implying that

$$n^{-1}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}[\nabla_\theta' \widetilde{G}_i(\bar{\theta}_{n,\gamma})]d\mathbb{Q} \to \int \widetilde{\mathbf{g}} E_{\mathbb{P}}[\nabla_\theta' \widetilde{G}_i(\bar{\theta}_{n,\gamma})]d\mathbb{Q},$$

which we defined as $\mathbf{D}^*$. Given this, we note that (5) and the joint convergence condition in Assumption 9 imply that $\int \widetilde{\mathbf{g}}\mathcal{G} - \mathbf{D}^* H^{*-1}\mathcal{Z}_0$ is also a normal random variable having the covariance matrix $\mathbf{B}^*$, obtained by applying theorem 2 of Grenander (1981, p. 48). Given this, the positive definite matrix $\mathbf{B}^*$ in Assumption 9(*ii*) enables us to apply the Cramér-Wold device, which we omit for brevity. This completes the proof. ■

Proof of Theorem 5: To show this, we examine the asymptotic limit of each element in $\widehat{\mathbf{B}}_n$. First, we consider the first row and first column element in $\widehat{\mathbf{B}}_n$. Note that

$$\begin{aligned}\frac{1}{n}\sum\int\int&\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\ &=\frac{1}{n}\sum\int\int\varepsilon_i(\gamma)\varepsilon_i(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\ &\quad+\frac{2}{n}\sum\int\varepsilon_i(\gamma)d\mathbb{Q}(\gamma)\ \left\{\int\mu(\widetilde{\gamma})-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}+\left\{\int\mu(\gamma)-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\gamma)d\mathbb{Q}(\gamma)\right\}^2,\end{aligned}$$

using the fact that $\widehat{\varepsilon}_{in}=\varepsilon_i+\{\mu(\gamma)-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\gamma)\}$. Further, by the FOC for the FOLS estimator, $n^{-1}\sum\int\{G_i(\gamma)-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\gamma)\}d\mathbb{Q}(\gamma)\equiv 0$, so that

$$\frac{1}{n}\sum\int\int\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})=\frac{1}{n}\sum\int\int\varepsilon_i(\gamma)\varepsilon_i(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})-\left\{\frac{1}{n}\sum\int\varepsilon_i(\gamma)d\mathbb{Q}(\gamma)\right\}^2.$$

Given this, using Cauchy-Schwarz inequality we obtain that

$$\sup_{\gamma,\widetilde{\gamma}}\left|\frac{1}{n}\sum G_i(\gamma)G_i(\widetilde{\gamma})\right|\le\sup_{\gamma,\widetilde{\gamma}}\left|n^{-1}\sum G_i(\gamma)^2\right|^{1/2}\left|n^{-1}\sum G_i(\widetilde{\gamma})^2\right|^{1/2}\le n^{-1}\sum_{i=1}^n M_i^2\text{ a.s. }-\mathbb{P}$$

by Assumption 3, and the RHS is finite a.s. $-\mathbb{P}$. Thus, we can first let n tends to infinity before computing the associated integrals. The given SULLNs in Assumptions 5 and 10 imply that

$$\int\int n^{-1}\sum G_i(\gamma)G_i(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\to\int\int E_{\mathbb{P}}[G_i(\gamma)G_i(\widetilde{\gamma})]d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\quad\text{and}$$

$$\int n^{-1}\sum G_i(\gamma)d\mathbb{Q}(\gamma)\to\int\mu(\gamma)d\mathbb{Q}(\gamma)\text{ a.s. }-\mathbb{P},$$

so that we obtain

$$n^{-1}\sum\int\int\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\to\int\int\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\tag{8}$$

a.s. $-\mathbb{P}$. Second, we consider the first row and $(j+1)$-th column element of $\widehat{\mathbf{B}}_n$, where $j=1,2,\ldots,k$. We note that

$$\begin{aligned}\frac{1}{n}\sum\int\int\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})&=\frac{1}{n}\sum\int\int\varepsilon_i(\gamma)\varepsilon_i(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\ &-2\left\{\frac{1}{n}\sum\int\varepsilon_i(\gamma)d\mathbb{Q}(\gamma)\right\}\left\{\frac{1}{n}\sum\int\varepsilon_i(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}+\left\{\frac{1}{n}\sum\int\varepsilon_i(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}^2\end{aligned}$$

by the FOC for the FOLS estimator, $n^{-1}\sum\{\int[G_i(\gamma)-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\gamma)]g_j(\gamma)\}d\mathbb{Q}(\gamma)=0$. Given this, the Cauchy-Schwarz inequality and Assumption 4(*ii*) imply that

$$\left|n^{-1}\sum G_i(\gamma)G_i(\widetilde{\gamma})g_j(\widetilde{\gamma})\right|\le\left|n^{-1}\sum M_i^2\right|\times|g_j(\widetilde{\gamma})|,\quad\left|n^{-1}\sum G_i(\widetilde{\gamma})g_j(\widetilde{\gamma})\right|\le\left|n^{-1}\sum M_i^2\right|^{1/2}\times|g_j(\widetilde{\gamma})|,\text{ and}$$

$$\left|n^{-1}\sum G_i(\gamma)g_j(\widetilde{\gamma})\right| \leq \left|n^{-1}\sum M_i^2\right|^{1/2} \times |g_j(\widetilde{\gamma})|$$

uniformly in γ and $\widetilde{\gamma}$. Note that when the RHS's of these inequalities are viewed as functions of $\widetilde{\gamma}$, they all are in $L_1(\mathbb{Q})$ a.s. $-\mathbb{P}$. These imply that we can apply the DCT, so that

$$\frac{1}{n}\sum\int\int[G_i(\gamma)-\mu(\gamma)][G_i(\widetilde{\gamma})-\mu(\widetilde{\gamma})]g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \to \int\int\kappa(\gamma,\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \tag{9}$$

a.s. $-\mathbb{P}$. Third, we consider the $(j+1)$-th row and $(\widetilde{j}+1)$-th column element of $\widehat{\mathbf{B}}_n$. Note that

$$\begin{aligned}&\frac{1}{n}\sum\int\int g_j(\gamma)\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})g_{\widetilde{j}}(\gamma)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\ &\quad=\frac{1}{n}\sum\int\int g_j(\gamma)\varepsilon_i(\gamma)\varepsilon_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) - \left\{\frac{1}{n}\sum\int g_j(\gamma)\varepsilon_i(\gamma)d\mathbb{Q}(\gamma)\right\}\left\{\frac{1}{n}\sum\int\varepsilon_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}\end{aligned}$$

using the fact that $n^{-1}\sum\{\int[G_i(\gamma)-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\gamma)]g_j(\gamma)\}d\mathbb{Q}(\gamma)=0$ and $n^{-1}\sum\{\int[G_i(\widetilde{\gamma})-\widehat{\delta}_{0n}-\widehat{\delta}_n'\mathbf{g}(\widetilde{\gamma})]g_{\widetilde{j}}(\widetilde{\gamma})\}d\mathbb{Q}(\gamma)=0$. Also, by exploiting Cauchy-Schwarz inequality iteratively, we can obtain that

$$\left|\frac{1}{n}\sum g_j(\gamma)G_i(\gamma)G_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})\right| \leq \left(n^{-1}\sum M_i^2\right)\times|g_j(\gamma)|\times|g_{\widetilde{j}}(\widetilde{\gamma})| \text{ and}$$

$$\left|\frac{1}{n}\sum g_j(\gamma)G_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})\right| \leq \left(n^{-1}\sum M_i^2\right)^{1/2}\times|g_j(\gamma)|\times|g_{\widetilde{j}}(\widetilde{\gamma})|$$

uniformly in γ and $\widetilde{\gamma}$. Note that the RHSs of these inequalities are in $L_1(\mathbb{Q}\times\mathbb{Q})$ a.s. $-\mathbb{P}$ when they are viewed as functions of γ and $\widetilde{\gamma}$ by Assumption 4(*ii*). This implies that we can apply the DCT. By applying Assumptions 5, 10, and Theorem 1, it follows that

$$\frac{1}{n}\sum\int\int g_j(\gamma)[G_i(\gamma)-\mu(\gamma)][G_i(\widetilde{\gamma})-\mu(\widetilde{\gamma})]g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \to \int\int g_j(\gamma)\kappa(\gamma,\widetilde{\gamma})g_{\widetilde{j}}(\gamma)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \text{ a.s. } -\mathbb{P}. \tag{10}$$

Finally, collecting all the elements in (8), (9), and (10) for $j,\widetilde{j}=1,2,\ldots,k$, we obtain that the asymptotic limit of $\widehat{\mathbf{B}}_n$ is identical to $\mathbf{B}$. This completes the proof. ■

Proof of Theorem 6: (*i*) The proof is almost identical to the proof of Theorem 5. We examine the asymptotic limit of each element in $\widetilde{\mathbf{B}}_n$. First, we consider the first row and first column element in $\widehat{\mathbf{B}}_n$. Note that

$$\frac{1}{n}\sum\int\int\widetilde{\varepsilon}_{in}(\gamma)\widetilde{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) = \frac{1}{n}\sum\int\int\ddot{\varepsilon}_{in}(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) - \left\{\frac{1}{n}\sum\int\ddot{\varepsilon}_{in}(\gamma)d\mathbb{Q}(\gamma)\right\}^2,$$

using the facts that $\widetilde{\varepsilon}_{in}=\ddot{\varepsilon}_{in}+\{\mu(\gamma)-\widetilde{\delta}_{0n}-\widetilde{\delta}_n'\mathbf{g}(\gamma)\}$ and the FOC that $n^{-1}\sum\int\{\widehat{G}_i(\gamma)-\widetilde{\delta}_{0n}-\widetilde{\delta}_n'\mathbf{g}(\gamma)\}d\mathbb{Q}(\gamma)=0$, where $\ddot{\varepsilon}_{in}:=\widehat{G}_i-\mu$. Given this, we already proved in the proof of Theorem 3 that $n^{-1}\sum\int\ddot{\varepsilon}_{in}(\gamma)\to 0$ a.s. $-\mathbb{P}$. Also, Assumption 7(*iii*) enables us to apply the DCT, so that we can first let n tend to infinity before computing the associated integral. Note that

$$\begin{aligned}&\frac{1}{n}\sum\int\int\ddot{\varepsilon}_{in}(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\&\quad=\frac{1}{n}\sum\int\int\widetilde{G}_i(\gamma,\widehat{\theta}_n)\widetilde{G}_i(\widetilde{\gamma},\widehat{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})-\frac{2}{n}\sum\int\widetilde{G}_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma)\int\mu(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})+\left(\int\mu(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right)^2.\end{aligned}$$

We examine each element in the RHS. First,

$$\sup_{\gamma,\widetilde{\gamma},\theta}\left|n^{-1}\sum\widetilde{G}_i(\gamma,\widehat{\theta}_n)\widetilde{G}_i(\widetilde{\gamma},\widehat{\theta}_n)-E_{\mathbb{P}}[\widetilde{G}_i(\gamma,\theta^*)\widetilde{G}_i(\widetilde{\gamma},\theta^*)]\right|\to 0\ \text{ a.s. }-\mathbb{P}$$

by Assumption 12(*i*), Theorem 3, and the continuity of G_i with respect to θ, implying that

$$n^{-1}\sum\int\int\widetilde{G}_i(\gamma,\widehat{\theta}_n)\widetilde{G}_i(\widetilde{\gamma},\widehat{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\to\int\int E_{\mathbb{P}}[\widetilde{G}_i(\gamma,\theta^*)\widetilde{G}_i(\widetilde{\gamma},\theta^*)]d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\text{ a.s. }-\mathbb{P}.$$

Also, from the fact that $n^{-1}\sum\int\ddot{\varepsilon}_i(\gamma)\to 0$ a.s. $-\mathbb{P}$, $n^{-1}\sum\int\int\widehat{G}_i(\gamma)\mu(\widetilde{\gamma})\to\left(\int\mu\right)^2$ a.s. $-\mathbb{P}$, so that it follows that

$$\frac{1}{n}\sum\int\int\ddot{\varepsilon}_{in}(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\to\int\int\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\ \text{ a.s. }-\mathbb{P}.\tag{11}$$

Second, we consider the first row and $(j+1)$-th column element of $\widetilde{\mathbf{B}}_n$, where $j=1,2,\ldots,k$. Note that

$$\begin{aligned}\frac{1}{n}\sum\int\int\widetilde{\varepsilon}_{in}(\gamma)\widetilde{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})&=\frac{1}{n}\sum\int\int\ddot{\varepsilon}_i(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\&-2\left\{\frac{1}{n}\sum\int\ddot{\varepsilon}_{in}(\gamma)d\mathbb{Q}(\gamma)\right\}\left\{\frac{1}{n}\sum\int\ddot{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}+\left\{\frac{1}{n}\sum\int\ddot{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}^2,\end{aligned}$$

and we already saw that $n^{-1}\sum\int\ddot{\varepsilon}_{in}\to 0$ a.s. $-\mathbb{P}$ and $n^{-1}\sum\int\ddot{\varepsilon}_i g_j\to 0$ a.s. $-\mathbb{P}$ in the proof of Theorem 3. Also, note that

$$\begin{aligned}\frac{1}{n}\sum\int\int\ddot{\varepsilon}_{in}(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})&=\frac{1}{n}\sum\int\int\widetilde{G}_i(\gamma,\widehat{\theta}_n)\widetilde{G}_i(\widetilde{\gamma},\widehat{\theta}_n)g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\&-\frac{1}{n}\sum\int\widetilde{G}_i(\gamma,\widehat{\theta}_n)\int\mu(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})-\frac{1}{n}\sum\int\mu(\gamma)\int\widetilde{G}_i(\widetilde{\gamma},\widehat{\theta}_n)g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\&+\int\mu(\gamma)d\mathbb{Q}(\gamma)\int\mu(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma}).\end{aligned}$$

Given this, from the facts that $n^{-1}\sum\int\ddot{\varepsilon}_{in}\to 0$ a.s. $-\mathbb{P}$ and that $n^{-1}\sum\int\ddot{\varepsilon}_{in}g_j\to 0$ a.s. $-\mathbb{P}$, it follows that $n^{-1}\sum\int\widetilde{G}_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma)\to\int\mu(\gamma)d\mathbb{Q}(\gamma)$ a.s. $-\mathbb{P}$ and that $n^{-1}\sum\int\widetilde{G}_i(\widetilde{\gamma},\widehat{\theta}_n)g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\to\int\mu(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})$

a.s. $-\mathbb{P}$ respectively. Further, using the Cauchy-Schwarz inequality, Assumption 4(*ii*), and Assumption 7(*iii*) shows that

$$\left| n^{-1} \sum \widetilde{G}_i(\gamma, \theta) \widetilde{G}_i(\widetilde{\gamma}, \theta) g_j(\widetilde{\gamma}) \right| \leq \left| n^{-1} \sum M_i^2 \right| \times |g_j(\widetilde{\gamma})|$$

uniformly in γ, $\widetilde{\gamma}$, and θ. Note that the RHS of this inequality is in $L_1(\mathbb{Q})$ a.s. $-\mathbb{P}$ when viewed as a function of $\widetilde{\gamma}$ by Assumption 4(*ii*). This implies that we can apply the DCT, so that Assumption 12(*i*) implies that

$$n^{-1} \sum \int \int \ddot{\varepsilon}_{in}(\gamma) \ddot{\varepsilon}_{in}(\widetilde{\gamma}) g_j(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \to \int \int \kappa(\gamma, \widetilde{\gamma}) g_j(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \ \text{ a.s. } - \mathbb{P}. \tag{12}$$

Third, we consider the $(j+1)$-th row and $(\widetilde{j}+1)$-th column element of $\widetilde{\mathbf{B}}_n$. We note that the FOLS FOC $n^{-1}\sum\{\int [G_i(\gamma) - \widehat{\delta}_{0n} - \widehat{\delta}_n' \mathbf{g}(\gamma)] g_j(\gamma)\} d\mathbb{Q}(\gamma) = 0$ and $n^{-1}\sum\{\int [G_i(\widetilde{\gamma}) - \widehat{\delta}_{0n} - \widehat{\delta}_n' \mathbf{g}(\widetilde{\gamma})] g_{\widetilde{j}}(\widetilde{\gamma})\} d\mathbb{Q}(\widetilde{\gamma}) = 0$ imply

$$\begin{aligned}
&\frac{1}{n} \sum \int \int g_j(\gamma) \widetilde{\varepsilon}_{in}(\gamma) \widetilde{\varepsilon}_{in}(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \\
&\quad = \frac{1}{n} \sum \int \int g_j(\gamma) \ddot{\varepsilon}_i(\gamma) \ddot{\varepsilon}_i(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) - \left\{ \frac{1}{n} \sum \int g_j(\gamma) \ddot{\varepsilon}_i(\gamma) d\mathbb{Q}(\gamma) \right\} \left\{ \frac{1}{n} \sum \int \ddot{\varepsilon}_i(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\widetilde{\gamma}) \right\} \\
&\quad = \frac{1}{n} \sum \int \int g_j(\gamma) \ddot{\varepsilon}_i(\gamma) \ddot{\varepsilon}_i(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) + o_{a.s.}(1),
\end{aligned}$$

as $n^{-1} \sum \int \ddot{\varepsilon}_i(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) \to 0$ a.s. $-\mathbb{P}$. Also, note that

$$\begin{aligned}
&\frac{1}{n} \sum \int \int g_j(\gamma) \ddot{\varepsilon}_{in}(\gamma) \ddot{\varepsilon}_{in}(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \\
&\quad = \frac{1}{n} \sum \int \int g_j(\gamma) \widetilde{G}_i(\gamma, \widehat{\theta}_n) \widetilde{G}_i(\widetilde{\gamma}, \widehat{\theta}_n) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) - \int g_j(\gamma) \mu(\gamma) d\mathbb{Q}(\gamma) \int \mu(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\widetilde{\gamma}) + o_{a.s.}(1),
\end{aligned}$$

because $n^{-1}\sum \int \ddot{\varepsilon}_{in} g_j \to 0$ a.s. $-\mathbb{P}$ and $n^{-1}\sum \int \ddot{\varepsilon}_{in} g_j \to 0$ a.s. $-\mathbb{P}$ imply that $n^{-1}\sum \int \widetilde{G}_i(\gamma, \widehat{\theta}_n) g_j(\gamma) d\mathbb{Q}(\gamma) \to \int \mu(\gamma) g_j(\gamma) d\mathbb{Q}(\gamma)$ a.s. $-\mathbb{P}$ and $n^{-1} \sum \int \widetilde{G}_i(\widetilde{\gamma}, \widehat{\theta}_n) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\widetilde{\gamma}) \to \int \mu(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\widetilde{\gamma})$ a.s. $-\mathbb{P}$. Furthermore, exploiting the Cauchy-Schwarz inequality iteratively, we can obtain that

$$\left| n^{-1} \sum g_j(\gamma) \widetilde{G}_i(\gamma, \theta) \widetilde{G}_i(\widetilde{\gamma}, \theta) g_{\widetilde{j}}(\widetilde{\gamma}) \right| \leq \left(n^{-1} \sum M_i^2 \right) \times |g_j(\gamma)| \times |g_{\widetilde{j}}(\widetilde{\gamma})|$$

uniformly in γ, $\widetilde{\gamma}$, and θ. Note that the RHS of this inequalities is in $L_1(\mathbb{Q} \times \mathbb{Q})$ a.s. $-\mathbb{P}$, when it is viewed as a function of γ and $\widetilde{\gamma}$. This also implies that we can apply the DCT. From Assumption 12(*i*), it now follows that

$$\frac{1}{n} \sum \int \int g_j(\gamma) \widetilde{\varepsilon}_{in}(\gamma) \widetilde{\varepsilon}_{in}(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \to \int \int g_j(\gamma) \kappa(\gamma, \widetilde{\gamma}) g_{\widetilde{j}}(\gamma) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \ \text{ a.s. } - \mathbb{P}. \tag{13}$$

Finally, collecting all the elements in (11), (12), and (13) for $j, \widetilde{j} = 1, 2, \ldots, k$, we obtain that the asymptotic limit of $\widetilde{\mathbf{B}}_n$ is identical to $\mathbf{B}$.

(ii) Given Theorem 6(i), the definition of $\widetilde{\mathbf{B}}_n^*$, and the conditions in Assumption 8(iii), the desired result follows if $\widetilde{\mathbf{D}}_n \to \mathbf{D}^*$ and $\widetilde{\mathbf{K}}_n \to \mathbf{K}^*$ a.s. $-\mathbb{P}$. We already saw in the proof of Theorem 4(ii) that $\widetilde{\mathbf{D}}_n \to \mathbf{D}^*$ a.s. $-\mathbb{P}$. Therefore, we only prove here that $\widetilde{\mathbf{K}}_n \to \mathbf{K}^*$ a.s. $-\mathbb{P}$. Note that

$$\widetilde{\mathbf{K}}_n = \frac{1}{n}\sum \int s_i(\widehat{\theta}_n)\ddot{\varepsilon}_{in}(\gamma)\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma) + \frac{1}{n}\sum s_i(\widehat{\theta}_n)\int \{\mu(\gamma) - \widetilde{\delta}_{0n} - \mathbf{g}(\gamma)'\widetilde{\delta}_n\}\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma), \tag{14}$$

and first consider the second element. First, $n^{-1}\sum s_i(\widehat{\theta}_n) - n^{-1}\sum s_i(\theta^*) = o_{a.s.}(1)$ because s_i is continuous with respect to θ, and $\widehat{\theta}_n \to \theta^*$ a.s. $-\mathbb{P}$ by Assumption 8(i). Further, Assumptions 8(iii) and 9(i) imply that $\sum s_i(\theta^*) = o_{a.s.}(n)$, so that $n^{-1}\sum s_i(\widehat{\theta}_n) \to 0$ a.s. $-\mathbb{P}$. Next, we already saw that $n^{-1}\sum\int\{\widehat{G}_i(\gamma) - \widetilde{\delta}_{0n} - \widetilde{\delta}_n'\mathbf{g}(\gamma)\}\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma) = 0$ by the FOC for the TSFOLS estimator, and that $n^{-1}\sum\int \widehat{G}_i(\gamma)\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma) \to \int \mu(\gamma)\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma)$ in the proof of Theorem 6(i). Therefore,

$$\int \{\mu(\gamma) - \widetilde{\delta}_{0n} - \mathbf{g}(\gamma)'\widetilde{\delta}_n\}\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma) \to 0 \;\text{ a.s. } -\mathbb{P}.$$

Third, we consider the first element in (14), and for this we verify that we can apply the DCT. From the definition of $\ddot{\varepsilon}_{in}$, note that for each $j = 1, 2, \cdots, m$ and $\widetilde{j} = 1, 2, \cdots, k+1$,

$$\begin{aligned}\frac{1}{n}\sum \left| s_{ij}(\widehat{\theta}_n)\ddot{\varepsilon}_{in}(\gamma)\widetilde{\mathbf{g}}_{\widetilde{j}}(\gamma)\right| &\le \left\{\frac{1}{n}\sum s_{ij}(\widehat{\theta}_n)^2\right\}^{1/2}\left(\left\{\frac{1}{n}\sum \widehat{G}_i(\gamma)^2\right\}^{1/2} + |\mu(\gamma)|\right) \times |\widetilde{\mathbf{g}}_{\widetilde{j}}(\gamma)| \\ &\le \left\{\frac{1}{n}\sum s_{ij}(\widehat{\theta}_n)^2\right\}^{1/2}\left(\left\{\frac{1}{n}\sum M_i^2\right\}^{1/2} + E[M_i^2]\right) \times |\widetilde{\mathbf{g}}_{\widetilde{j}}(\gamma)|\end{aligned}$$

by Assumption 7(iii). Given this, $\widehat{I}_n$ is finite a.s. $-\mathbb{P}$ and converges to I^* a.s. $-\mathbb{P}$ by Assumption 8(iii), implying that for each $j = 1, 2, \cdots, m$, $n^{-1}\sum s_{ij}(\widehat{\theta}_n)^2$ is finite a.s. $-\mathbb{P}$. Therefore, the RHS must be in $L_1(\mathbb{Q})$, when viewed as a function of γ. Therefore, we can apply the DCT. Given this,

$$\frac{1}{n}\sum s_i(\widehat{\theta}_n)\ddot{\varepsilon}_{in}(\gamma) = \frac{1}{n}\sum s_i(\widehat{\theta}_n)\widehat{G}_i(\gamma) - \mu(\gamma)\frac{1}{n}\sum s_i(\widehat{\theta}_n)$$

by the definition of $\ddot{\varepsilon}_{in}$; and Assumption 7(iii) and $\sum s_i(\widehat{\theta}_n) = o_{a.s.}(n)$ imply that $\mu(\gamma)\sum s_i(\widehat{\theta}_n) = o_{a.s.}(n)$ uniformly in γ. Further, Assumption 12(ii) and the continuity of s_i and G_i with respect to θ by Assumptions 8($iii.a$) and 7(iii) respectively implies that for each γ,

$$\frac{1}{n}\sum s_i(\widehat{\theta}_n)\widehat{G}_i(\gamma) = E_{\mathbb{P}}[s_i(\theta^*)G_i(\gamma, \theta^*)] + o_{a.s.}(1)$$

because $\widehat{\theta}_n \to \theta^*$ a.s. $-\mathbb{P}$ by Assumption 8(i). We note that $E_{\mathbb{P}}[s_i(\theta^*)G_i(\gamma, \theta^*)] = \kappa_0(\gamma)$ from the IID condition and the condition in Assumption 8($iii.a$) that $\sqrt{n}s_n^* = n^{-1/2}\sum s_i(\,\cdot\,, \theta^*) + o_{\mathbb{P}}(1)$. Therefore, $n^{-1}\sum\int s_i(\widehat{\theta}_n)\ddot{\varepsilon}_{in}(\gamma)\widetilde{\mathbf{g}}(\gamma)' d\mathbb{Q}(\gamma)$

$\to \int \kappa_0(\gamma)\widetilde{\mathbf{g}}(\gamma)'d\mathbb{Q}(\gamma)$. Finally, collecting all these together implies that

$$\widetilde{\mathbf{K}}_n = \int \frac{1}{n}\sum s_i \ddot{\varepsilon}_{in}(\gamma)\widetilde{\mathbf{g}}(\gamma)'d\mathbb{Q}(\gamma) + o_{a.s.}(1) = \int \kappa_0(\gamma)\widetilde{\mathbf{g}}(\gamma)'d\mathbb{Q}(\gamma) + o_{a.s.}(1),$$

and this completes the proof. ■

Proof of Theorem 7: (*i*) $\sqrt{n}S_j(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{\Gamma}_j)$ by Theorem 2, where $\mathbf{\Gamma}_j := S_j\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}S_j'$, so that it follows that $\mathbf{\Gamma}^{-1/2}\sqrt{n}S_j(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{I}_{k+2-j})$. Because $\widehat{\mathbf{B}}_n \to \mathbf{B}$ a.s. $-\mathbb{P}$ as given in Theorem 5, $\widehat{\mathbf{\Gamma}}_{nj} \to \mathbf{\Gamma}_j$ a.s. $-\mathbb{P}$ by proposition 2.30 of White (2001), where $\widehat{\mathbf{\Gamma}}_{nj} := S_j\mathbf{A}^{-1}\widehat{\mathbf{B}}_n\mathbf{A}^{-1}S_j'$. Therefore,

$$\mathcal{M}_{j,n} := n(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*)'S_j'\widehat{\mathbf{\Gamma}}_n^{-1}S_j(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{\text{A}}{\sim} \mathcal{X}^2_{k+2-j}$$

by theorem 4.30 of White (2001). Given this, we note that

$$\mathcal{W}_{j,n} = \mathcal{M}_{j,n} + 2n\boldsymbol{\delta}^{*\prime}S_j'\widehat{\mathbf{\Gamma}}_n^{-1}S_j(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) + n\boldsymbol{\delta}^{*\prime}S_j'\widehat{\mathbf{\Gamma}}_n^{-1}S_j\boldsymbol{\delta}^*.$$

Therefore, $\mathcal{M}_{j,n} = \mathcal{W}_{j,n} = O_{\mathbb{P}}(1)$ under $\mathbb{H}_{jo}$, so that $\mathcal{W}_{j,n} \overset{\text{A}}{\sim} \mathcal{X}^2_{k+2-j}$; and $\mathcal{W}_{j,n} = O_{\mathbb{P}}(1) + O_{\mathbb{P}}(\sqrt{n}) + O(n)$ under $\mathbb{H}_{jA}(\mathbf{g})$, implying the desired result.

(*ii*) $\sqrt{n}S_j(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{\Gamma}_j)$ by Theorem 4(*i*), and $\widetilde{\mathbf{B}}_n \to \mathbf{B}$ a.s. $-\mathbb{P}$ from Theorem 6(*i*). The rest is identical to the proof of Theorem 7(*i*).

(*iii*) $\sqrt{n}S_j(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{\Gamma}_j^*)$ by Theorem 4(*ii*), where $\mathbf{\Gamma}_j^* := S_j\mathbf{A}^{-1}\mathbf{B}^*\mathbf{A}^{-1}S_j'$, and $\widetilde{\mathbf{B}}_n^* \to \mathbf{B}^*$ a.s. $-\mathbb{P}$ from Theorem 6(*ii*). The rest is identical to the proof of Theorem 7(*i*). ■

Proof of Theorem 8: (*i*) $\sqrt{n}S_j(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n^*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{\Gamma}_j)$ by applying Theorem 2, where $\mathbf{\Gamma}_j$ is defined in the proof of Theorem 7(*i*), so that $\mathbf{\Gamma}^{-1/2}\sqrt{n}S_j(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}^*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{I}_{k+2-j})$. Given this, $\sqrt{n}S_j\boldsymbol{\delta}_n^* \to S_j\varsigma$ under $\mathbb{H}_{ja}(\mathbf{g})$, which implies that $\sqrt{n}S_j\widehat{\boldsymbol{\delta}}_n \overset{\text{A}}{\sim} N(S_j\varsigma, \mathbf{\Gamma}_j)$. Further, from the fact that $\widehat{\mathbf{B}}_n \to \mathbf{B}$ a.s. $-\mathbb{P}$ as given in Theorem 5, it follows that $\widehat{\mathbf{\Gamma}}_{nj} \to \mathbf{\Gamma}_j$ a.s. $-\mathbb{P}$ by proposition 2.30 of White (2001), where $\widehat{\mathbf{\Gamma}}_{nj}$ is defined in the proof of Theorem 7(*i*). Therefore, $\mathcal{W}_{j,n} \overset{\text{A}}{\sim} \mathcal{X}^2(k+2-j, \xi_j)$ by lemma 8.2 of White (1994), implying the desired result.

(*ii*) $\sqrt{n}S_j\widetilde{\boldsymbol{\delta}}_n \overset{\text{A}}{\sim} N(S_j\varsigma, \mathbf{\Gamma}_j)$ by Theorem 4(*i*), and $\widetilde{\mathbf{B}}_n \to \mathbf{B}$ a.s. $-\mathbb{P}$ from Theorem 6(*i*). The rest is identical to the proof of Theorem 8(*i*).

(*iii*) $\sqrt{n}S_j\widetilde{\boldsymbol{\delta}}_n \overset{\text{A}}{\sim} N(S_j\varsigma, \mathbf{\Gamma}_j^*)$ by Theorem 4(*ii*), and $\widetilde{\mathbf{B}}_n^* \to \mathbf{B}^*$ a.s. $-\mathbb{P}$ from Theorem 6(*ii*), where $\mathbf{\Gamma}_j^*$ is defined in the proof of Theorem 7(*iii*). The rest is identical to the proof of Theorem 8(*i*). ■

References

ANDREWS, D. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

ANDREWS, D. (2001): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

ANDREWS, D. AND PLOBERGER, W. (1994): "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

BAEK, Y., CHO, J.S., AND PHILLIPS, P. C. B. (2015): "Testing Linearity Using Power Transforms of Regressors," *Journal of Econometrics*, 187, 376–384.

BICKEL, P. AND WICHURA, M. (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications," *The Annals of Mathematical Statistics*, 42, 1656–1670.

BIERENS, H. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

BIERENS, H. AND PLOBERGER, W. (1997): "Asymptoticc Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1152.

BILLINGSLEY, P. (1968, 1999): *Convergence of Probability Measures*. New York: Wiley.

BREUSCH, T. AND PAGAN, A. (1979): "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287–1294.

CHANG, Y., HU, B., AND PARK, J. Y. (2019): "Econometric Analysis of Functional Dynamics in the Presence of Persistence," Discussion Paper, Department of Economics, Indiana University.

CHO, J.S. AND WHITE, H. (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

CHO, J.S. AND WHITE, H. (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

CHO, J. S. AND PHILLIPS, P. C. B. (2018): "Sequentially Testing Polynomial Model Hypotheses Using Power Transforms of Regressors," *Journal of Applied Econometrics*, 33, 141–159.

Cho, J.S., Phillips, P.C.B., and Seo, J. (2020): Parametric Conditional Mean Inference with Functional Data Applied to Lifetime Income Curves," Discussion Paper, School of Economics, Yonsei University.

CRAMBES, C., GANNOUN, A., AND HENCHIRI, Y. (2013): “Support Vector Machine Quantile Regression Approach for Functional Data: Simulation and Application Studies,” *Journal of Multivariate Analysis*, 121, 50–68.

DAVIES, R. (1977): “Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative,” *Biometrika*, 64, 247–254.

DAVIES, R. (1987): “Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative,” *Biometrika*, 74, 33–43.

FISHER, R. (1932): *Statistical Methods for Research Workers*. Edinburgh and London: Oliver and Boyd.

GRENANDER, U. (1981): *Abstract Inference*. New York: John Wiley and Sons.

HUBER, P. (1967): “The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions,” *Proceedings of the 5th Berkeley Symposium*, 1, 221–233.

JAIN, N. AND MARCUS, M. (1975): “Central Limit Theorem for $C(S)$–valued Random Variables,” *Journal of Functional Analysis*, 19, 216–231.

JENNRICH, R. (1969): “Asymptotic Properties of Nonlinear Least Squares Estimators,” *Annals of Mathematical Statistics*, 40, 634–644.

LANCASTER, H. (1961): “The Combination of Probabilities: An Application of Orthonomal Functions,” *Australian Jounal of Statistics*, 3, 20–33.

LE CAM, L. (1953): “On Some Asymptotic Properties of Maximum Likelihood Estimates and Relates Bayes Estimates,” *University of California Publications in Statistics*, 1, 277-330.

Li, D., Robinson, P. M., and Shang, H. L. (2020): “Long-Range Dependent Curve Time Series,” *Journal of the American Statistical Association*, 115, 957–971.

NEWEY, W. (1991): “Converence in Probability and Stochastic Equicontinuity,” *Econometrica*, 59, 1161–1167.

NEYMAN, J. (1959): “Optimal Asymptotic Tests of Composite Statistical Hypotheses,” in U. Grenander ed., *Probability and Statistics*, *The Harald Cramér Volume.* 213–234. New York: Wiley.

PEARSON, E. (1950): “On Questions Raised by the Combination of Tests Based on Discontinuous Distributions,” *Biometrika*, 37, 383–398.

PHILLIPS, P. C. B. AND JIANG, L. (2019): “Parametric Autoregression in Function Space,” Working Paper, Singapore Mangement University.

PÖTSCHER, B. AND PRUCHA, I. (1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica*, 57, 675–683.

SHORACK, R. AND WELLNER, J. (1986): *Empirical Processes with Applications to Statistics*. New Tork: Wiley.

STINCHCOMBE, M. AND WHITE, H. (1992): "Some Measurability Results for Extrema of Random Functions over Random Sets," *Review of Economic Studies*, 59, 495–514.

STINCHCOMBE, M. AND WHITE, H. (1998): "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory*, 14, 295–324.

TIPPET, L. (1931): *The Methods of Statistics: An Introduction Mainly for Workers in the Biological Sciences*. London: Williams and Norgate.

VAN DER VAART, A. AND WELLNER, J. (1996): *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.

VAN ZWET AND OOSTERHOFF, J. (1967): "On the Combination of Independent Test Statistics," *The Annals of Mathematical Statistics*, 28, 659–680.

WESTBERG, M. (1985): "Combining Independent Statistical Tests," *The Statistian*, 34, 287–296.

WHITE, H. (1980): "Using Least Squares to Approximate Unknown Pegression Function," *International Economic Review*, 21, 149–170.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

WHITE, H. (1994): *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

WHITE, H. (2001): *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

WHITE, H. (2006): "Approximate Nonlinear Forecasting Methods," in G. Elliott, C. Granger, and A. Timmermann eds., *Handbook of Economics Forecasting*. Vol. 1, 460–512. New York: Elsevier.

WOOLDRIDGE, J. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

ZHANG, J. AND CHEN, J. (2007): "Statistical Inferences for Functional Data," *Annals of Statistics*, 35, 1052–1079.

Table 1: LEVELS OF THE WALD AND BREUSCH AND PAGAN TESTS

NUMBER OF REPLICATIONS: 10,000

Statistics	Levels \ n	25	50	100	200	400	600	800
	1%	1.04	0.83	0.82	0.90	0.96	1.05	0.92
$\widetilde{\mathcal{W}}_{1,n}$(con)	5%	5.74	4.85	5.13	4.75	4.95	5.31	4.95
	10%	11.18	10.74	10.57	9.85	9.79	10.61	9.89
	1%	1.51	0.96	0.99	1.00	1.07	1.13	0.98
$\widetilde{\mathcal{W}}_{1,n}$(con+lin)	5%	6.64	5.32	5.02	5.10	5.00	5.15	4.77
	10%	12.41	11.19	10.53	9.98	10.23	10.32	10.15
	1%	1.68	1.09	0.99	0.93	0.95	0.90	1.07
$\widetilde{\mathcal{W}}_{1,n}$(con+quad)	5%	6.65	5.46	5.07	5.18	4.81	4.91	5.05
	10%	12.75	11.21	10.81	10.44	9.86	9.97	10.16
	1%	2.16	1.38	1.14	0.82	0.79	1.08	0.99
$\widetilde{\mathcal{W}}_{1,n}$(con+lin+quad)	5%	8.09	6.19	5.29	4.87	4.74	4.95	4.73
	10%	15.40	12.45	11.07	10.21	9.73	10.11	9.96
	1%	1.60	1.14	0.91	0.98	0.79	0.94	0.98
$\widetilde{\mathcal{W}}_{1,n}(\text{con} + 0.5^{\gamma})$	5%	6.71	5.58	5.39	5.29	5.23	4.84	5.31
	10%	13.01	11.64	10.81	10.43	10.39	9.89	10.34
	1%	0.31	0.63	0.72	0.81	1.00	1.03	0.81
$\mathcal{BP}_n$	5%	3.77	4.31	4.82	5.04	4.99	4.88	4.74
	10%	9.60	10.07	9.92	9.96	9.66	9.82	10.19

Table 2: POWERS OF THE WALD AND BREUSCH AND PAGAN TESTS (NOMINAL LEVEL: 5%)

NUMBER OF REPLICATIONS: 5,000

Statistics	$\sigma_c^2 \setminus n$	25	50	100	200	400	600	800
	0.10	6.82	7.48	9.66	14.50	24.40	33.00	42.98
	0.20	9.64	11.74	19.54	35.70	62.38	78.74	88.92
$\widetilde{\mathcal{W}}_{1,n}$(con)	0.30	11.90	19.08	32.28	56.98	86.04	96.70	99.16
	0.40	14.64	25.70	48.02	75.44	96.66	99.54	99.88
	0.50	20.28	34.94	59.90	86.28	99.20	99.96	100.0
	0.10	7.06	6.74	7.40	11.00	18.16	24.70	33.34
	0.20	8.84	9.42	15.46	26.60	51.70	69.48	81.76
$\widetilde{\mathcal{W}}_{1,n}$(con+lin)	0.30	11.26	14.80	26.52	47.52	78.04	92.32	97.40
	0.40	13.80	20.28	38.50	66.40	92.02	98.76	99.82
	0.50	15.74	27.32	49.10	79.86	98.06	99.76	100.0
	0.10	7.28	6.88	7.64	10.40	17.50	25.48	33.90
	0.20	9.08	9.78	15.76	27.62	49.90	67.28	81.60
$\widetilde{\mathcal{W}}_{1,n}$(con+quad)	0.30	11.42	14.58	25.04	47.50	79.12	93.02	97.74
	0.40	12.68	19.30	36.96	67.22	92.52	98.70	98.84
	0.50	16.92	26.86	50.00	80.02	97.92	99.80	100.0
	0.10	8.78	6.56	7.64	9.62	14.84	21.00	27.54
	0.20	9.72	9.44	14.00	23.20	45.34	62.10	76.26
$\widetilde{\mathcal{W}}_{1,n}$(con+lin+quad)	0.30	11.52	14.12	21.22	41.46	72.72	89.54	96.34
	0.40	13.70	18.74	31.30	58.94	90.12	97.88	99.72
	0.50	16.38	24.02	42.88	74.40	96.48	99.62	99.98
	0.10	7.74	6.38	7.70	11.88	16.76	24.48	33.68
	0.20	9.06	9.82	14.36	28.80	51.02	68.86	81.68
$\widetilde{\mathcal{W}}_{1,n}$(con+$0.5^{\gamma}$)	0.30	10.90	15.68	26.12	48.28	78.82	92.08	97.92
	0.40	13.68	21.94	37.30	65.88	92.94	98.46	99.76
	0.50	15.38	26.46	48.98	79.76	97.96	99.80	100.0
	0.10	4.28	5.16	8.28	14.28	23.16	32.84	41.52
	0.20	6.84	10.52	19.02	36.18	59.34	78.38	89.34
$\mathcal{BP}_n$	0.30	8.64	17.00	32.52	58.34	86.90	96.10	99.08
	0.40	11.48	23.86	45.58	74.76	96.10	99.68	99.96
	0.50	15.72	29.94	58.14	86.42	99.16	99.96	100.0

Table 3: LEVELS OF THE WALD, BIERENS, AND SW TESTS

NUMBER OF REPLICATIONS: 10,000

Statistics	Levels \ n	25	50	100	200	400	600	800
	1%	1.72	1.15	1.02	0.97	1.20	0.89	1.03
$\mathcal{W}_{1,n}^*$(con)	5%	6.64	5.68	5.36	5.31	5.25	4.96	5.05
	10%	12.44	11.51	10.64	10.31	10.36	10.26	10.02
	1%	1.45	0.77	0.59	0.67	0.60	0.78	0.62
$\mathcal{W}_{1,n}^*$(con+lin)	5%	6.87	4.36	3.96	4.10	4.42	4.20	4.30
	10%	13.37	9.79	9.01	8.91	9.27	9.61	9.41
	1%	1.24	0.70	0.56	0.58	0.57	0.53	0.79
$\mathcal{W}_{1,n}^*$(con+quad)	5%	6.39	4.35	4.03	3.79	4.02	3.83	4.34
	10%	12.83	9.78	8.93	9.22	8.71	8.76	9.47
	1%	2.38	1.26	0.87	0.66	0.58	0.67	0.56
$\mathcal{W}_{1,n}^*$(con+lin+quad)	5%	8.52	5.69	4.40	3.93	3.92	4.06	3.97
	10%	15.82	11.76	9.48	8.57	8.64	9.18	8.91
	1%	0.85	0.77	0.52	0.84	1.03	0.86	0.88
$\mathcal{B}_n$	5%	5.79	4.68	4.71	5.12	5.12	5.07	5.09
	10%	12.86	11.05	10.69	10.48	10.56	10.56	10.19
$\mathcal{SW}_n$		11.42	7.42	5.00	3.91	3.54	3.36	3.28

Table 4: POWERS OF THE WALD, BIERENS, AND SW TESTS (NOMINAL LEVEL: 5%)

NUMBER OF REPLICATIONS: 5,000

Statistics	π^* \ n	25	50	100	200	400	600	800
	0.10	49.02	76.34	95.16	99.04	99.76	99.90	100.0
	0.20	64.14	85.32	95.48	99.02	99.90	100.0	100.0
$\mathcal{W}_{1,n}^*$(con)	0.30	70.28	87.00	95.92	99.36	99.92	100.0	100.0
	0.40	70.44	87.64	96.14	98.96	99.88	100.0	100.0
	0.50	71.92	87.26	96.00	99.08	99.82	100.0	100.0
	0.10	17.04	28.56	60.12	92.16	99.86	100.0	100.0
	0.20	32.08	58.30	90.26	99.52	100.0	100.0	100.0
$\mathcal{W}_{1,n}^*$(con+lin)	0.30	44.76	73.58	94.90	99.68	100.0	100.0	100.0
	0.40	53.98	79.82	96.00	99.86	100.0	100.0	100.0
	0.50	59.56	81.66	96.50	99.90	100.0	100.0	100.0
	0.10	16.20	28.56	59.66	92.20	99.92	100.0	100.0
	0.20	32.50	56.50	90.88	99.80	100.0	100.0	100.0
$\mathcal{W}_{1,n}^*$(con+quad)	0.30	45.34	73.40	96.20	99.92	100.0	100.0	100.0
	0.40	52.78	81.02	97.22	100.0	100.0	100.0	100.0
	0.50	60.32	82.24	97.70	100.0	100.0	100.0	100.0
	0.10	16.50	28.51	57.50	92.33	99.86	100.0	100.0
	0.20	33.40	63.43	94.20	99.97	100.0	100.0	100.0
$\mathcal{W}_{1,n}^*$(con+lin+quad)	0.30	49.92	83.59	99.49	99.98	100.0	100.0	100.0
	0.40	63.24	92.90	99.78	100.0	100.0	100.0	100.0
	0.50	73.36	96.30	99.78	100.0	100.0	100.0	100.0
	0.10	18.82	40.02	70.88	92.18	98.64	99.56	99.74
	0.20	38.42	67.60	87.34	95.74	99.04	99.72	99.90
$\mathcal{B}_n$	0.30	52.30	77.30	89.62	96.40	99.36	99.78	99.92
	0.40	58.12	80.26	90.48	96.90	99.18	99.82	99.98
	0.50	64.30	82.58	91.20	96.58	99.18	99.86	99.96
	0.10	26.30	38.02	65.08	91.78	99.82	100.0	100.0
	0.20	46.78	69.72	93.50	99.76	100.0	100.0	100.0
$\mathcal{SW}_n$	0.30	60.98	82.48	96.94	99.94	100.0	100.0	100.0
	0.40	69.86	87.66	98.06	99.92	100.0	100.0	100.0
	0.50	75.28	90.52	98.26	99.92	100.0	100.0	100.0