Testing a Constant Mean Function Using Functional Regression

JIN SEO CHO
School of Economics, Yonsei University, Seodaemun-gu, Seoul 03722, Korea
Email: jinseocho@yonsei.ac.kr

MENG HUANG
Quantitative Analytics & Model Development Group, PNC, Washington D.C., US
Email: 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: April 2023

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 (2022) 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 using the neural network model, that involves nuisance parameters identified only under the alternative. Our Monte Carlo simulations affirm the theory of the current study. Finally, we apply our methodology to the probit models for voter turnout that are estimated by Wolfinger and Rosenstone (1980); Nagler (1991) and test whether the models are correctly specified or not.

Key Words: Davies Test; Functional Data; Misspecification; Nuisance Parameters; Wald Test; Voting Turnout.

Subject Classification: C11, C12, C80, D72.

Acknowledgements: The third author (Halbert White) passed away while the earlier version was circulating as a discussion paper. We are grateful for the constructive comments of the Editor, the Associate Editor, and six anonymous referees. We also have benefited from discussions with Peter C.B. Phillips, Juwon Seo, and seminar participants at Yonsei University and Economics Joint Conference. Cho also acknowledges financial support from the Yonsei University Research Grant of 2022.

1 Introduction

A considerable variety of useful testing procedures involve "nuisance" parameters when a standard neural network model is employed. Examples are those considered in the work of Davies (1977, 1987); Bierens (1982, 1990); Bierens and Ploberger (1997); Andrews (1994); Stinchcombe and White (1998); Cho, Ishida, and White (2011); Cho and Ishida (2012); White and Cho (2012); Baek, Cho, and Phillips (2015); Cho and Phillips (2018), among others. 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 null limit distributions of the resulting tests are highly context-specific, requiring special-purpose computations to obtain suitable critical values.

Functional data analysis is getting popular. We mention a few recent developments. Crambes, Gannoun, and Henchiri (2013) estimates 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 applies the local polynomial kernel estimation to discrete data and next estimates the functional coefficients of the covariates consisting 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) examines 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) estimates a functional autoregressive model with serially correlated functional data, establishing a limit theory of their estimator. Phillips and Jiang (2019) studies parametric autoregression with function-valued time series in stationary and nonstationary cases and also establishes a limit theory of their estimation. Finally, when the conditioning variable is a random variable, Cho et al. (2022) examines estimating the conditional mean of functional data, that is nonlinear with respect to unknown parameters, and Cho, Phillips, and Seo (2023) extends this to estimating the quantile functions of functional data.

In this paper, we consider a different approach, useful in this context, that yields statistics having standard chi-square null limit 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); Stinchcombe and White (1998). Note that the tests of Bierens (1982, 1990); Stinchcombe and White (1998) do not take account of correlations among the elements of the Gaussian process underlying the test 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 et al. (2022). Specifically, Cho et al. (2022) examines 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. More 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. As an appealing feature of using functional regression, the resulting tests have standard chi-square limit distributions under the null. Wald, Lagrange multiplier, and quasi-likelihood ratio versions of these tests are also available. For concreteness and conciseness, our focus here is on the use of Wald test.

Although functional regression is of theoretical interest in its own right, our focus here is further 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, we can also view the functions of interest arising in the analysis of models involving nuisance parameters identified only under the alternative 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); Stinchcombe and White (1998).

The plan of this paper is as follows. In Section 2, we motivate and formally describe the data generating process (DGP) underlying functional regression, illustrating 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 on the functional regression. We provide conditions under which these estimators are consistent and asymptotically normal, and we provide consistent estimators of their asymptotic covariance matrices. In Section 4, we specify the null hypotheses of interest and introduce Wald tests. As we show, the tests 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 a $\sqrt{n}$ rate, where n is the sample size. Section 5 applies the theory to obtain tests for our panel data and specification testing examples. Section 6 provides a Monte Carlo analysis and an empirical application of the FOLS estimator, where we study the finite and large sample properties of the tests developed in Section 5. We also apply our methodology to test whether popularly assumed models for voting turnout are correctly specified or not by using 1984 presidential election data of the U.S. Section 7 contains a summary and concluding remarks.

2 The Data Generating Process and Functional Regression

In this section, we motivate and formally describe the DGP condition 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).* □

Often in prior literature, 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 functionals of a DGP 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 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 $\omega \in \Omega$. More specifically, as an example, for a random variable X_i, if we let $G_i(\omega, \gamma)$ be $\ell(X_i(\omega), \gamma)$ that is the log-likelihood of X_i at γ, $\ell(X_i, \cdot)$ becomes a function defined on Γ. We view $\ell(X_i, \cdot)$ as a functional observation whose functional form is different from one individual to another individual, so that our data set is now a collection of realized random functions. As we illustrate using running examples, this view is useful when developing test statistics for model specification. In particular, if the model is not identified, it could be more convenient for the purpose of a model specification testing to treat $\ell(X_t, \cdot)$ as an observation instead of examining X_i as an observation. We also note that the notion of a random process is the same as that of the random function, and the IID condition is not essential. 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 et al. (2022). Note that we do not impose any functional restriction on the functional data, whereas Cho et al. (2022) restricts its interest to continuous functional data. One of the examples we discuss below assumes functional data that are discrete at equally distanced points.

We illustrate 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, 2010, ch.10.4).

Example 1 (Panel Random Effects): Let $\gamma \in \Gamma := \{1, 2, ..., T\}$, and suppose data are generated as $Y_i(\gamma) = X_i(\gamma)^\top \beta_0 + V_i(\gamma)$, $i = 1, 2, ...$, where $\beta_0 \in \mathbb{R}^d$ and $V_i(\gamma) := C_i + U_i(\gamma)$. We assume that $(Y_i, X_i^\top)^\top : \Omega \times \Gamma \mapsto \mathbb{R}^{1+d}$ are IID. $U_i : \Omega \times \Gamma \mapsto \mathbb{R}$ and $C_i : \Omega \mapsto \mathbb{R}$ are unobserved. Let $\mathbf{X}_i := (X_i(1), X_i(2), \cdots, X_i(T))^\top$, $\mathbf{V}_i := (V_i(1), V_i(2), \cdots, V_i(T))^\top$, and assume that $\mathbf{\Sigma} := E[\mathbf{V}_i \mathbf{V}_i^\top]$ is finite and positive definite, with $\text{rank}(E[\mathbf{X}_i^\top \mathbf{\Sigma}^{-1} \mathbf{X}_i]) = d$. The data exhibit *random effects* structure when, for $i = 1, 2, ...$,

1. $U_i(\gamma)$ are 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.

We can easily define functional observations from this. Now let $G_i(\gamma) = V_i(1)V_i(\gamma)$ that is a discrete function on Γ, and it has different implications for the estimation of β_0. 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); Cho and Ishida (2012); 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).

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

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

and $\psi : \mathbb{R} \to \mathbb{R}$ is a given function that is typically the activation function used for the neural network model. For example, Bierens (1990) specifies $\psi = \exp$; Stinchcombe and White (1998) considers 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 Stinchcombe and White (1998) show, when $\mathbb{M}$ is correctly specified (so that there exists $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ such that $E[Y_i|X_i] = f(X_i, \boldsymbol{\theta}_0)$), provided that $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ (as holds for the NLS estimator as well as for linear exponential family-based QML estimators generally), then $E[G_i(\gamma)] = 0 \quad$ for all $\gamma \in \Gamma$; whereas when $\mathbb{M}$ is *not* correctly specified and ψ is GCR (*e.g.*, $\psi = \exp$ is GCR), Stinchcombe and White (1998) shows $E[G_i(\gamma)] \neq 0$ for almost all $\gamma \in \Gamma$. Bierens (1990) and Stinchcombe and White (1998) exploit this property to construct tests for model misspecification. Their test statistics are based on

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

If the model is correctly specified, $Z_n(\cdot)$ obeys the functional central limit theorem under some regularity conditions, but it is not the case if $E[G_i(\gamma)] \neq 0$ for some γ. $\square$

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, or we can view μ in a Bayesian model perspective. For instance, when $\Gamma = [0, 1]$, the researcher can choose a beta distribution with specific parameter values to focus on a particular region of $[0, 1]$. Alternatively, if the researcher has no particular interest in any specific region of the unit interval, the standard uniform distribution can be selected as the adjunct probability measure. 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;*

[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 $\Phi(\mathbf{X}_i) := [\tan^{-1}(X_{1i}),\ \tan^{-1}(X_{2i}),\ \cdots,\ \tan^{-1}(X_{di})]^\top$. We leave this implicit here.

(ii) For $i = 1, 2, ..., G_i$ is measurable $-\mathcal{F} \otimes \mathcal{G}$. □

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 $G \in \mathcal{G}$. Here, 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 $\omega \in F$, $\mathbb{P}[F] = 1$, the measutability condition holds. More generally, the desired condition holds by applying the theory of separable stochastic process developed by Doob (*e.g.*, Dellacherie and Meyer, 1978). If there is a countable dense subset of Γ that determines the class of function $G_i(\omega, \cdot)$, the measurability of G_i follows.

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. 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 $M_i \in L_2(\mathbb{P})$, $\sup_{\gamma \in \Gamma} |G_i(\gamma)| \leq M_i$, $i = 1, 2, ...$* □

Assumption 3 is equivalent to supposing that $\sup_{\gamma \in \Gamma} |G_i(\gamma)|$ has a finite second moment. From this, it follows that μ as defined above exists and is measurable $-\mathcal{G}$, and that $\mu \in L_2(\mathbb{Q})$, meaning that μ is a deterministic function defined on Γ such that $\int_\Gamma |\mu(\gamma)|^2 d\mathbb{Q}(\gamma) < \infty$. Note that 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 $\int \mathrm{var}_{\mathbb{P}}[G_i(\gamma)] d\mathbb{Q}(\gamma) = \mathrm{var}_{\mathbb{P}}[G_i(\gamma_0)]$. 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

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

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

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 $\mathbf{g} := (g_1, g_2, ..., g_k)^\top$. 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*

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

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 $m(\,\cdot\,, \boldsymbol{\xi}) = \delta_0 + \mathbf{g}(\,\cdot\,)^\top \boldsymbol{\delta}$. Thus, m belongs to the linear model

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

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 $\mathbf{g}(\gamma) = \gamma$, that yields a linear functional regression. The model structure of $\mathcal{A}(\mathbf{g})$ is simpler than that considered by Cho et al. (2022) as their model consists of random functions and is nonlinearly parameterized. Due to this simple structure, we can 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

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

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

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

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

The first-order conditions for the optimum are

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

These yield convenient expressions for δ_0^* and $\boldsymbol{\delta}^*$, analogous to the standard regression approximation case:

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

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

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

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

It is readily verified that if $\mu=0\ a.s.-\mathbb{Q}$ ($\mathbb{H}_{1o}$ holds) then $\boldsymbol{\xi}^*=\left[0,\mathbf{0}^\top\right]^\top$. If instead, for unknown constant c, $\mu=c\ \ a.s.-\mathbb{Q}$ ($\mathbb{H}_{2o}$ holds) then $\boldsymbol{\xi}^*=\left[c,\mathbf{0}^\top\right]^\top$. Thus, δ_0^* and $\boldsymbol{\delta}^*$ 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 $\boldsymbol{\delta}^*$ need not equal zero, as $\mathrm{cov}_{\mathbb{Q}}[\mathbf{g},\mu]$ is not necessarily $\mathbf{0}$ under $\mathbb{H}_{1A}$. Similarly, if $\mathbb{H}_{2o}$ does not hold, then $\boldsymbol{\delta}^*$ need not equal zero. This behavior gives our tests their power. We emphasize that in these cases, the optimizer $m(\,\cdot\,,\boldsymbol{\xi}^*)$ generally does not coincide with μ, as $m(\,\cdot\,,\boldsymbol{\xi}^*)$ is essentially a misspecified approximation to μ under the specified alternatives.

3 Functional Ordinary Least Squares Estimation

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

$$Q_n(\boldsymbol{\xi}) := \frac{1}{n}\sum_{i=1}^{n}\int\{G_i(\gamma)-\delta_0-\mathbf{g}(\gamma)^\top\boldsymbol{\delta}\}^2 d\mathbb{Q}(\gamma),$$

where $\boldsymbol{\xi} := (\delta_0, \boldsymbol{\delta}^\top)^\top$. The resulting estimator is the *functional ordinary least squares* (FOLS) estimator, denoted as $\widehat{\boldsymbol{\xi}}_n := (\widehat{\delta}_{0n}, \widehat{\boldsymbol{\delta}}_n^\top)^\top$. This has the convenient representation

$$\widehat{\boldsymbol{\xi}}_n := \begin{bmatrix} \widehat{\delta}_{0n} \\ \widehat{\boldsymbol{\delta}}_n \end{bmatrix} := \begin{bmatrix} 1 & \int \mathbf{g}(\gamma)d\mathbb{Q}(\gamma) \\ \int \mathbf{g}(\gamma)d\mathbb{Q}(\gamma) & \int \mathbf{g}(\gamma)\,\mathbf{g}(\gamma)^\top d\mathbb{Q}(\gamma) \end{bmatrix}^{-1} \begin{bmatrix} n^{-1}\sum_{i=1}^n \int G_i(\gamma)d\mathbb{Q}(\gamma) \\ n^{-1}\sum_{i=1}^n \int \mathbf{g}(\gamma)G_i(\gamma)d\mathbb{Q}(\gamma) \end{bmatrix}.$$

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)**.** $\sup_{\gamma\in\Gamma}|n^{-1}\sum_{i=1}^n G_i(\gamma)-\mu(\gamma)| \to 0$ *a.s.* □

Given the domination condition of Assumption 3, this holds under mild additional conditions on $\{G_i\}$. For example, 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); 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 $\widehat{\boldsymbol{\delta}}_n$. 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 1–5, $\widehat{\boldsymbol{\xi}}_n \to \boldsymbol{\xi}^*$ a.s.* □

3.2 Asymptotic Normality of FOLS Estimator

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 zero-mean 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 extensive literature providing primitive conditions for the FCLT. Billingsley (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); van der 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 1–6,* $\sqrt{n}(\widehat{\boldsymbol{\xi}}_n-\boldsymbol{\xi}^*)\overset{A}{\sim}N(\mathbf{0},\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1})$. □

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

[2] Jain and Marcus (1975) provides 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., so that for all $\gamma,\widetilde{\gamma}\in\Gamma$, $|G_i(\gamma)-G_i(\widetilde{\gamma})|\leq K_i\rho(\gamma,\widetilde{\gamma})$ 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.

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 Estimator

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 are relevant instances. To handle these cases in a general way, it suffices to assume that

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

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

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

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

$$\widetilde{\boldsymbol{\xi}}_n := \begin{bmatrix} \widetilde{\delta}_{0n} \\ \widetilde{\boldsymbol{\delta}}_n \end{bmatrix} := \begin{bmatrix} 1 & \int \mathbf{g}(\gamma)d\mathbb{Q}(\gamma) \\ \int \mathbf{g}(\gamma)d\mathbb{Q}(\gamma) & \int \mathbf{g}(\gamma)\,\mathbf{g}(\gamma)^\top d\mathbb{Q}(\gamma) \end{bmatrix}^{-1} \begin{bmatrix} n^{-1}\sum_{i=1}^n \int \widehat{G}_i(\gamma)d\mathbb{Q}(\gamma) \\ n^{-1}\sum_{i=1}^n \int \widehat{G}_i(\gamma)\mathbf{g}(\gamma)d\mathbb{Q}(\gamma) \end{bmatrix}.$$

When $\widehat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}^*$ 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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{bmatrix} \int(\widehat{G}_i(\gamma)-\mu(\gamma))d\mathbb{Q}(\gamma) \\ \int \mathbf{g}(\gamma)(\widehat{G}_i(\gamma)-\mu(\gamma))d\mathbb{Q}(\gamma) \end{bmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int \widetilde{\mathbf{g}}(\gamma)(\widehat{G}_i(\gamma)-\mu(\gamma))d\mathbb{Q}(\gamma),$$

where $\widetilde{\mathbf{g}} := (1,\mathbf{g}^\top)^\top$. 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 $\boldsymbol{\theta}^*$ (interior to $\boldsymbol{\Theta}$) gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}(\gamma)[\widetilde{G}_i(\,\cdot\,,\gamma,\widehat{\boldsymbol{\theta}}_n)-\mu(\gamma)]d\mathbb{Q}(\gamma)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}(\gamma)[\widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta}^*)-\mu(\gamma)]d\mathbb{Q}(\gamma)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}(\gamma)[\nabla_{\boldsymbol{\theta}}^{\top}\widetilde{G}_i(\,\cdot\,,\gamma,\bar{\boldsymbol{\theta}}_{n,\gamma})]d\mathbb{Q}(\gamma)\;\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}^*), \tag{2}$$

where the mean value $\bar{\boldsymbol{\theta}}_{n,\gamma}$ lies between $\widehat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}^*$ and, as indicated, depends on γ. With $G_i(\,\cdot\,,\gamma):=\widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta}^*)$, 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:

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

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

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

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

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

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

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

and an asymptotic normality result follows straightforwardly under some 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* $\boldsymbol{\Theta}\subset\mathbb{R}^m, m\in\mathbb{N}$, *be compact;*

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

(iii) $\boldsymbol{\Theta}$ *is convect and for each* $(\omega,\gamma)\in\Omega\times\Gamma$,

- *(iii.a)* $\widetilde{G}_i(\omega,\gamma,\,\cdot\,)$ *is continuously differentiable on* $\boldsymbol{\Theta}$*;*
- *(iii.b)* $\sup_{(\gamma,\boldsymbol{\theta})\in\Gamma\times\boldsymbol{\Theta}}|\widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta})|\leq M_i$,*; and*
- *(iii.c)* $\sup_{j=1,\dots,m}\sup_{(\gamma,\boldsymbol{\theta})\in\Gamma\times\boldsymbol{\Theta}}|(\partial/\partial\theta_j)\widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta})|\leq M_i$, $i=1,2,\dots$. □

Assumptions 7(*i* and *ii*) ensure that Assumptions 1 and 2 hold for $G_i(\,\cdot\,,\gamma):=\widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta}^*$ 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* $\boldsymbol{\theta}^*\in\boldsymbol{\Theta}$ *and a sequence of measurable functions* $\{\widehat{\boldsymbol{\theta}}_n:\Omega\mapsto\boldsymbol{\Theta}\}$ *such that*

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

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

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

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 $\sqrt{n}\boldsymbol{s}_n^*$ and $n^{-1/2}\sum_{i=1}^n (G_i - \mu)$. This condition implies Assumption 6.

Assumption 9 (Joint Convergence-A)**.**

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

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

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

$$E_{\mathbb{P}}[\boldsymbol{\mathcal{Z}}(\gamma)\boldsymbol{\mathcal{Z}}(\widetilde{\gamma})^\top] = \begin{bmatrix} \mathbf{I}^* & \boldsymbol{\kappa}_0(\widetilde{\gamma}) \\ \boldsymbol{\kappa}_0(\gamma)^\top & \kappa(\gamma,\widetilde{\gamma}) \end{bmatrix},$$

where $\mathbf{I}^*$ *is a nonstochastic finite symmetric positive semi-definite* $m \times m$ *matrix;* $\boldsymbol{\kappa}_0 : \Gamma \mapsto \mathbb{R}^m$ *belongs to* $L_2(\mathbb{Q})$*; and* κ *is as in Assumption 6; and*

(ii) $\lambda_{\min}(\mathbf{B}^*) > 0$*, where we let* $\mathbf{B}^* := \mathbf{B} - \mathbf{D}^*\mathbf{H}^{*-1}\mathbf{K}^* - \mathbf{K}^{*\top}\mathbf{H}^{*-1\top}\mathbf{D}^{*\top} + \mathbf{D}^*\mathbf{H}^{*-1}\mathbf{I}^*\mathbf{H}^{*-1\top}\mathbf{D}^{*\top}$ *and* $\mathbf{K}^* := \int \boldsymbol{\kappa}_0(\gamma)\widetilde{\mathbf{g}}(\gamma)^\top d\mathbb{Q}(\gamma)$. □

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 3–5, 7, and 8(i) for* $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta}^*)$, $\widetilde{\boldsymbol{\xi}}_n \to \boldsymbol{\xi}^*$ *a.s.* □

The asymptotic normality result for the TSFOLS estimator is

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

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

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

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 $\mathbf{A}$ in this context, as it is known.

Let the functional regression estimated residuals $\widehat{\varepsilon}_{in} : \Omega \times \Gamma \mapsto \mathbb{R}$ be defined by $\widehat{\varepsilon}_{in}(\,\cdot\,, \gamma) := G_i(\,\cdot\,, \gamma) - \widehat{\delta}_{0n} - \mathbf{g}(\gamma)^\top \widehat{\boldsymbol{\delta}}_n$. For convenience, we write $\widehat{\varepsilon}_{in}(\gamma)$ as a shorthand for $\widehat{\varepsilon}_{in}(\,\cdot\,, \gamma)$. We consider estimators of the form

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

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

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

If we take all of these conditions together, Assumptions 1–6, 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 1–6, and 10,* $\widehat{\mathbf{B}}_n \to \mathbf{B}$ $a.s.$ □

For the TSFOLS estimator, we use the second-stage residuals $\widetilde{\varepsilon}_{in} : \Omega \times \Gamma \mapsto \mathbb{R}$ defined by $\widetilde{\varepsilon}_{in}(\,\cdot\,, \gamma) := \widehat{G}_i(\,\cdot\,, \gamma) - \widetilde{\delta}_{0n} - \mathbf{g}(\gamma)^\top \widetilde{\boldsymbol{\delta}}_n$. 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 as $\widetilde{\mathbf{B}}_n$. Otherwise, we construct the estimator

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

where we let

$$\widetilde{\mathbf{D}}_n := \frac{1}{n}\sum_{i=1}^{n}\int \widetilde{\mathbf{g}}(\gamma)\,\nabla_{\boldsymbol{\theta}}^\top \widetilde{G}_i(\,\cdot\,, \gamma, \widehat{\boldsymbol{\theta}}_n)\, d\mathbb{Q}(\gamma),$$

$$\widetilde{\mathbf{K}}_n := \frac{1}{n}\sum_{i=1}^{n}\int \boldsymbol{s}_i(\,\cdot\,, \widehat{\boldsymbol{\theta}}_n)\widetilde{\varepsilon}_{in}(\,\cdot\,, \gamma)\widetilde{\mathbf{g}}(\gamma)^\top\, d\mathbb{Q}(\gamma), \quad \text{and} \quad \widehat{\mathbf{I}}_n := \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i(\,\cdot\,, \widehat{\boldsymbol{\theta}}_n)\boldsymbol{s}_i(\,\cdot\,, \widehat{\boldsymbol{\theta}}_n)^\top$$

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

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* $\boldsymbol{s}_i : \Omega \times \boldsymbol{\Theta} \mapsto \mathbb{R}^m$ *such that*

(i.a) $\boldsymbol{s}_i(\,\cdot\,, \boldsymbol{\theta})$ *is measurable*$-\mathcal{F}$ *for each* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *and* $\boldsymbol{s}_i(\omega, \cdot\,)$ *is continuous on* $\boldsymbol{\Theta}$ *for all* $\omega \in F \in \mathcal{F}$, $\mathbb{P}(F) = 1$; $\sqrt{n}\boldsymbol{s}_n^* = n^{-1/2}\sum_{i=1}^n \boldsymbol{s}_i(\,\cdot\,, \boldsymbol{\theta}^*) + o_{\mathbb{P}}(1)$; *and*

(i.b) $\widehat{\mathbf{I}}_n \to \mathbf{I}^*$ $a.s.$; *and*

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

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

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

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

Note that Assumption 12(*i*) implies Assumption 10, because $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta}^*)$. 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 3–5 for* $G_i(\,\cdot\,,\gamma) := \widetilde{G}_i(\,\cdot\,,\gamma,\boldsymbol{\theta}^*)$, *7, 8(i), and 12(i),* $\widetilde{\mathbf{B}}_n \to \mathbf{B}$ $a.s.$*;*

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

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 tests for our hypotheses of interest, $\mathbb{H}_{1o}$ and $\mathbb{H}_{2o}$, we define selection matrices $\mathbf{S}_1 := \mathbf{I}_{k+1}$ and $\mathbf{S}_2 := [\mathbf{0}_k, \mathbf{I}_k]$, 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

$$\mathbb{H}_{1o}(\mathbf{g}) : \mathbf{S}_1\boldsymbol{\xi}^* = \mathbf{0}_{k+1} \quad \text{and} \quad \mathbb{H}_{2o}(\mathbf{g}) : \mathbf{S}_2\boldsymbol{\xi}^* = \mathbf{0}_k.$$

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

$$\mathbb{H}_{1A}(\mathbf{g}) : \mathbf{S}_1\boldsymbol{\xi}^* \neq \mathbf{0}_{k+1} \quad \text{and} \quad \mathbb{H}_{2A}(\mathbf{g}) : \mathbf{S}_2\boldsymbol{\xi}^* \neq \mathbf{0}_k.$$

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

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

Wald tests 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

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

Wald tests 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

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

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

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

(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.* $c_n = o(n)$. □

According to Theorem 7, the Wald tests are subject to both type I and type II errors. As the sample size n approaches infinity, Theorem 7 shows that the Wald tests become consistent against the alternatives, effectively removing the risk of type II error in the long run. However, the risk of type I error cannot be eliminated as long as a finite critical value is used for comparison in the Wald tests. Therefore, when n is finite, the researcher must weigh the trade-off between type I and type II errors in order to optimize the cost of making an error using the Wald tests.

4.2 The Wald Test under Local Alternatives

We consider local alternatives of the following form: $\{\mu_n\}$ is such that for some ς_j,

$$\mathbb{H}_{ja}(\mathbf{g}) : \sqrt{n}\mathbf{S}_j\boldsymbol{\xi}^*_n \to \varsigma_j, \quad j = 1, 2,$$

where

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

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 (that is more easily verified as a weak ULLN for triangular arrays), 6, 7(*ii*, *iii*), 8 (with weak rather than strong convergence to $\mathbf{D}^*$), and 9–12 (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 $\mathcal{X}^2(k, \tau)$ denote the noncentral chi-square distribution with k degrees of freedom and noncentrality parameter τ. The following noncentrality parameters are relevant for $j = 1, 2$:

$$\tau_j := \varsigma_j^\top [\mathbf{S}_j\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{S}_j^\top]^{-1}\varsigma_j; \quad \text{and} \quad \tau^*_j := \varsigma_j^\top [\mathbf{S}_j\mathbf{A}^{-1}\mathbf{B}^*\mathbf{A}^{-1}\mathbf{S}_j^\top]^{-1}\varsigma_j.$$

Theorem 8.

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

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

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

5 Examples

We illustrate the application of the foregoing results by returning to our examples in 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 $\widehat{V}_i(\gamma) := \widetilde{V}_i(\gamma, \widehat{\beta}_n) = Y_i(\gamma) - X_i(\gamma)^\top \widehat{\beta}_n$, and $\widehat{\beta}_n$ is the pooled OLS estimator,

$$\widehat{\beta}_n := \left(\sum_{i=1}^{n} \sum_{\gamma=1}^{T} X_i(\gamma) X_i(\gamma)^\top \right)^{-1} \left(\sum_{i=1}^{n} \sum_{\gamma=1}^{T} X_i(\gamma) Y_i(\gamma) \right).$$

To determine which asymptotic covariance matrix applies in this case, we investigate $\mathbf{D}^* := \int \widetilde{\mathbf{g}}(\gamma) E_{\mathbb{P}} [\nabla_\beta^\top \widetilde{G}_i(\cdot, \gamma, \beta^*)] d\mathbb{Q}(\gamma)$. Now, we note that $(\partial/\partial\beta_j)\widetilde{G}_i(\cdot, \gamma, \beta^*) = -X_{ij}(1)V_i(\gamma) - V_i(1)X_{ij}(\gamma)$. Under pure random effects ($\sigma_c^2 = 0$), it then follows that for all $\gamma \in \{2, ..., T\}$, $E_{\mathbb{P}}[\nabla_\beta^\top \widetilde{G}_i(\,\cdot\,, \gamma, \beta^*)] = 0$. In this case, the first-stage estimation does not affect 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

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

The matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

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

$$\mathbf{B} = \frac{1}{(T-1)^2} \sum_{\gamma=2}^{T} \sum_{\widetilde{\gamma}=2}^{T} \left[\begin{array}{cc} \kappa(\gamma, \widetilde{\gamma}) & \kappa(\gamma, \widetilde{\gamma}) g_1(\widetilde{\gamma}) \\ g_1(\gamma)\kappa(\gamma, \widetilde{\gamma}) & g_1(\gamma)\kappa(\gamma, \widetilde{\gamma}) g_1(\widetilde{\gamma}) \end{array} \right],$$

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 $d = 2$, $X_i := (X_{i1}, X_{i2})^\top := (1, X_{2i})^\top$ and that $E_{\mathbb{P}}[Y_i|X_i] = \pi^* \exp(X_{2i})$. Next, take $f(X, \boldsymbol{\theta}) = \theta_1 + \theta_2 X_2$, 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 $\widehat{\boldsymbol{\theta}}_n := (\widehat{\theta}_{1n}, \widehat{\theta}_{2n})^\top$ 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 $\widetilde{\boldsymbol{\delta}}_n$ to minimize

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

where $\mathbf{g}$ is a 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 $\mathbf{D}^* := \int \widetilde{\mathbf{g}}(\gamma) E_{\mathbb{P}}[\nabla_{\boldsymbol{\theta}}^\top \widetilde{G}_i(\,\cdot\,, \gamma, \boldsymbol{\theta}^*)]\, d\mathbb{Q}(\gamma) = (\bar{\gamma} - \underline{\gamma})^{-1} \int_{\underline{\gamma}}^{\bar{\gamma}} \widetilde{\mathbf{g}}(\gamma)\, E_{\mathbb{P}}[(-1, -X_2)\psi(X_2\gamma)] d\gamma$. 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 the computation of

$$\widetilde{\mathbf{D}}_n = (\bar{\gamma} - \underline{\gamma})^{-1} \frac{1}{n} \sum_{i=1}^{n} \int_{\underline{\gamma}}^{\bar{\gamma}} \widetilde{\mathbf{g}}(\gamma)(-1, -X_{2i})\psi(X_{2i}\gamma) d\gamma,$$

$$\widetilde{\mathbf{K}}_n = (\bar{\gamma} - \underline{\gamma})^{-1} \frac{1}{n} \sum_{i=1}^{n} \int_{\underline{\gamma}}^{\bar{\gamma}} \boldsymbol{s}_i(\,\cdot\,, \widehat{\boldsymbol{\theta}}_n) \widetilde{\varepsilon}_{in}(\,\cdot\,, \gamma) \widetilde{\mathbf{g}}(\gamma)^\top d\gamma,$$

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

where

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

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 $(X_{2i}, \varepsilon_i)^\top \sim$ IID $N((1,0)^\top, \mathbf{I}_2)$. 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

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

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 Stinchcombe and White (1998) rely on statistics computed as functionals of

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

These statistics have asymptotic distributions that are generally highly complex, varying for different choices of ψ and for different choices 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 $\mathbb{Q}$. □

6 Numerical Analysis

In this section, we conduct Monte Carlo experiments using our Wald tests with the DGPs specified in the previous examples. First, we investigate the behavior of functional regression tests for panel data random effects and compare these to BP-test Breusch and Pagan (1979). 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 yielding new insights for panel data. Second, we compare the specification tests of Bierens (1990); Stinchcombe and White (1998) to our functional regression Wald tests. Here, functional regression offers not only computational convenience but also interesting power advantages. Finally, we apply our methodology to empirical data by testing whether popularly assumed models for voting turnout are correctly specified or not.

6.1 Simulation 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 QML 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 tests 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 $\widetilde{\mathcal{W}}_{1,n}$(con+$0.5^\gamma$), respectively.

We also apply BP-test to test the null of pure random effects structure. This statistic is popularly used to test for unobserved fixed effects, as noted by Wooldridge (2010) 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 tests 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 $\widetilde{\mathcal{W}}_{1,n}(\text{con+lin+quad})$, that uses three degrees of freedom, than for the others, that 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 Simulation 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 tests 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); Stinchcombe and White (1998), letting $\mathcal{B}_n$ and $\mathcal{SW}_n$ denote the tests in Bierens (1990) and Stinchcombe and White (1998), 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.

Stinchcombe and White (1998) gives a simple consistent test procedure using critical values based on the law of the iterated logarithm (LIL) bound. This is quite conservative, as Stinchcombe and White (1998) points 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. The LIL procedure in Stinchcombe and White (1998) 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 $\mathcal{SW}_n$.

6.3 Empirical Application

In the political science literature, a classic research topic is to explain voting turnout. For example, Wolfinger and Rosenstone (1980) estimates a probit model using the 1972 presidential election in the U.S. In addition, many empirical studies are conducted to explain voting turnout using empirical data (*e.g.*, Nagler, 1991, 1994; Bénabou, 2000; Besley and Case, 2003; Berry, DeMeritt, and Esarley, 2010).

The empirical model in Wolfinger and Rosenstone (1980) is influential to the literature as it shows the influence of education on voting turnout. Specifically, after letting the education level (*Education*), the squared-education level (*Education2*), age (*Age*), squared-age (*Age2*), a dummy for the South (*South*), a dummy for the presence of a gubernatorial election in the state (*Gubernatorial Election*), and the number of days before the election that registration close (*Closing Date*) be the explanatory variables, it estimates a linear probit model and claims that the registration requirement by the voting law most severely affects the least educated group. *Closing Date* measures the voting law requirement. The estimation shows that if *Closing Date* is hypothetically set to zero, the average turnout rate increases for the least educated group, Meanwhile, the most educated voters' turnout increase is the smallest. Using the notion of the voting cost, they explain that more educated voters pay less a cost for understanding the implications of complex and abstract political issues. This finding is a stylized fact in political science literature.

Nagler (1991) criticizes that the empirical result in Wolfinger and Rosenstone (1980) is an output of using the probit model. The probit probability is highly affected if the explanatory variable is around zero, so that if *Education* is near zero, the predicted probability increase is greater than other voters with higher education. Nagler (1991) remedies this problem by estimating another probit model with two additional explanatory variables: *Closing Date* $\times$ *Education* and *Closing Date* $\times$ *Education*^{2}. By including them, Nagler (1991) captures the interactive effects of *Closing Date* and *Education* on the turnout. Using the 1972 and 1984 presidential election data in the U.S., Nagler (1991) rejects the empirical result in Wolfinger and Rosenstone (1980).

Nevertheless, both models specified by Wolfinger and Rosenstone (1980) and Nagler (1991) could be misspecified. We, therefore, test whether the models are correctly specified or not using a specification testing procedure that applies the methodology in Section 6.2. Specifically, we follow the following procedure: First, we estimate the two models by the QML estimation by supposing that their probit models are possibly misspecified and using the 1984 presidential election data provided by Altman and McDonald (2003). We let $\mathbf{X}_i$ be the conditioning variable for Y_i such that $Y_i = 1$ if the i-th individual votes, and $Y_i = 0$, otherwise. Note that if for some $\boldsymbol{\theta}_*$, $\mathbb{E}[Y_t|\mathbf{X}_t] = \Phi(\mathbf{X}_t^\top \boldsymbol{\theta}_*)$, then the model is correctly specified, so that estimating $\boldsymbol{\theta}_*$ by the QML estimation is identical to maximum likelihood (ML) estimation, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a standard normal random variable. Otherwise, the probit model is misspecified and the QML estimation differs from the ML estimation, so that the information matrix equality does not hold. For such a case, $\boldsymbol{\theta}_*$ is the probability limit of the QML estimator, and it follows that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*) \overset{\text{A}}{\sim} N(\mathbf{0}, \mathbf{H}^{-1}\mathbf{I}\mathbf{H}^{-1})$, where $\mathbf{H}$ (resp. $\mathbf{I}$) is the limit of negative Hessian log-likelihood function (resp. the covariance matrix of the quasi-score) that is evaluated at the limit of the QML estimator. Nevertheless, we also note that the QML estimation can still estimate $\mathbb{E}[Y_t|\mathbf{X}_t]$ consistently despite the distributional misspecification (*e.g.*, Gourieroux, Monfort, and Trognon, 1984; White, 1994). By this, we can apply the QML estimation to correct model specification for $\mathbb{E}[Y_t|\mathbf{X}_t]$.

Second, we apply the testing methodology in Section 6.2 to the two probit models specified by Wolfinger and Rosenstone (1980) and Nagler (1991) and test correct model specification for $\mathbb{E}[Y_t|\mathbf{X}_t]$. Specifically, we first let

$$\widehat{G}_i(\gamma) := [Y_i - \Phi(\mathbf{X}_i^\top \boldsymbol{\theta}_n)]\psi(Z_i\gamma) \quad \text{and} \quad \widehat{Q}_n(\boldsymbol{\xi}) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(\bar{\gamma} - \underline{\gamma})}\int_{\underline{\gamma}}^{\bar{\gamma}} \{\widehat{G}_i(\gamma) - \delta_0 - g(\gamma)\delta\}^2 d\gamma,$$

where $\Gamma := [\underline{\gamma}, \bar{\gamma}] = [-1, 1]$, and Z_i is one of the conditioning variables. We choose one out of *Closing Date*, *Education*, and *Age* for Z_t. The other conditioning variables in $\mathbf{X}_i$ are dummies, so that they are not effective for the goal of our testing. We further select either γ or γ^2 for $g(\gamma)$, and either exponential or logistic function for ψ. After estimating δ_0 and δ by FOLS, we apply the Wald test in Section 6.2.

We report the empirical estimation and inference outputs in Tables 5 and 6. We summarize them as follows. First, the QML estimates are contained in Table 5. The probit models without and with interactive terms are the models specified by Wolfinger and Rosenstone (1980) and Nagler (1991), respectively. Berry et al. (2010) also estimates the same models, and all the parameter estimates in Table 5 are the same as theirs. The only difference is the p-values of the t-tests given in parentheses. They are computed by the robust standard errors.

Second, the testing results are contained in Table 6 which shows the Wald tests and their respective p-values in parentheses. As we can see from the table, the inference results are different between the probit models with and without the interactive terms. For the probit model without the products, the p-values of the Wald tests are significantly close to zero only when $Z_i =$ *Education*. In particular, when $g(\gamma) = \gamma^2$, the p-values are close to zero for both exponential and logistic functions. This fact strongly implies that the probit model in Wolfinger and Rosenstone (1980) is misspecified for the voter turnout data. For the probit model with the products in Nagler (1991), the p-values of the Wald tests are significantly different from zero only when $Z_i =$ *Age*. In particular, if $Z_i =$ *Closing Date*, the p-values are close to zero when the exponential function is selected for $\psi(\cdot)$. This fact strongly suggests that the probit model posited by Nagler (1991) for the voter turnout data is also misspecified.

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 (2022). As we show, the associated Wald test statistics have standard chi-square null limit 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.

We demonstrate the use of the Wald tests using examples. Panel data can be viewed as functional data; we illustrate this with a running example focusing on a test of random effects structure. As another running example, we develop new alternatives to tests for regression misspecification using the neural network model. 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 Stinchcombe and White (1998). 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 empirically test whether popularly assumed models for voting turnout are correctly specified or not by using the 1984 presidential election data of the U.S., finding that the probit models in Wolfinger and Rosenstone (1980); Nagler (1991) are misspecified.

Functional regression tests may have utility in a variety of disparate contexts. For example, methods that involve combining a finite number of multiple statistics using a specified weighting method or a Bayes method have been considered in the literature (Tippett, 1931; Fisher, 1932; Pearson, 1950; Lancaster, 1961; van Zwet and Oosterhoff, 1967; Westberg, 1985), and our approach accommodates such methods involving hypothesis testing with multiple statistics in a unified framework as described in Cho et al. (2022). It further allows not just a finite number of tests but allows the tests to be indexed by elements of a multidimensional continuum. As examples, the variability of multiple trajectories of growth functions can be analyzed in this framework (*e.g.*, Contreras-Reyes, Quintero, and Wiff, 2018) and the model specification test using extreme learning machine can also be viewed in the framework of the current study (*e.g.*, Cho and White, 2011; Shin and Cho, 2013). In addition, the framework can be used to analyze popularly examined nonstationary panel data in economics by treating each individual's time-series process as an individual functional observation (*e.g.*, Cho et al., 2022, 2023), which expands the applicability of the current study.

8 Appendix: Proofs

We introduce some mathematical notation used throughout this section. First, for notational simplicity, 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. Second, when there is no possible ambiguity, we may further abbreviate these to $\int g$ and $\int\int h$.

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

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

for every j, so that

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

a.s., 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

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

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

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

a.s. ■

Proof of Theorem 2: From the note that

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

the desired result follows if

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

the desired result follows. Assumption 6(ii) implies that $n^{-1/2}\sum_{i=1}^n (G_i - \mu) \Rightarrow \mathcal{G}$, so that we obtain $n^{-1/2}\sum_{i=1}^n \int (G_i - \mu) \Rightarrow \int \mathcal{G}$, 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

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

$$\int \mathcal{G} g_j \sim N\left(0, \int\int g_j(\gamma)\kappa(\gamma, \widetilde{\gamma}) g_j(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma})\right), \tag{6}$$

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, that 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

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

for every j, so that

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

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

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

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

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

(b) the second element on the right-hand side (RHS) converges to zero a.s. by Assumption 5; and (c) applying the mean-value theorem implies that

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

where the RHS converges to zero a.s. by Assumptions 7(*iii*) and 8(*i*). Thus, we obtain that

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

a.s. This completes the proof. ■

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

(ii) From the given fact that

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

the desired result follows if

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{bmatrix} \int (\widehat{G}_i - \mu) \\ \int (\widehat{G}_i - \mu)\,\mathbf{g} \end{bmatrix} \overset{\text{A}}{\sim} N\left(\mathbf{0}, \mathbf{B}^*\right). \tag{7}$$

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

$$\begin{aligned} \frac{1}{\sqrt{n}}\int \sum_{i=1}^n \widetilde{\mathbf{g}}(\widehat{G}_i - \mu) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int \widetilde{\mathbf{g}}(G_i - \mu) + \frac{1}{n}\sum_{i=1}^n \int \widetilde{\mathbf{g}}[\nabla_{\boldsymbol{\theta}}^\top \widetilde{G}_i(\bar{\boldsymbol{\theta}}_{n,\gamma})]\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \\ &\Rightarrow \int \widetilde{\mathbf{g}}\mathcal{G} - \mathbf{D}^*\mathbf{H}^{*-1}\boldsymbol{\mathcal{Z}}_0 \end{aligned} \tag{8}$$

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

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

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

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

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

that we defined as $\mathbf{D}^*$. Given this, we note that (6) and the joint convergence condition in Assumption 9 imply that $\int \widetilde{\mathbf{g}}\mathcal{G} - \mathbf{D}^*\mathbf{H}^{*-1}\boldsymbol{\mathcal{Z}}_0$ 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, that 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

$$\frac{1}{n}\sum_{i=1}^{n}\int\int\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\int\int\varepsilon_i(\gamma)\varepsilon_i(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$$
$$+\frac{2}{n}\sum_{i=1}^{n}\int\varepsilon_i(\gamma)d\mathbb{Q}(\gamma)\left\{\int\mu(\widetilde{\gamma})-\widehat{\delta}_{0n}-\widehat{\boldsymbol{\delta}}_n^{\top}\mathbf{g}(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}$$
$$+\left\{\int\mu(\gamma)-\widehat{\delta}_{0n}-\widehat{\boldsymbol{\delta}}_n^{\top}\mathbf{g}(\gamma)d\mathbb{Q}(\gamma)\right\}^2,$$

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

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

Given this, using Cauchy-Schwarz inequality we obtain that

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

by Assumption 3, and the RHS is finite a.s. Thus, we can first let n tend to infinity before computing the associated integrals. The given SULLNs in Assumptions 5 and 10 imply that

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

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

so that we obtain

$$n^{-1}\sum_{i=1}^{n}\int\int\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\to\int\int\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\tag{9}$$

a.s. 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

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

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

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

$$\left|\frac{1}{n}\sum_{i=1}^{n}G_i(\gamma)g_j(\widetilde{\gamma})\right| \le \left|\frac{1}{n}\sum_{i=1}^{n}M_i^2\right|^{1/2}\times|g_j(\widetilde{\gamma})|$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\int\int[G_i(\gamma)-\mu(\gamma)][G_i(\widetilde{\gamma})-\mu(\widetilde{\gamma})]g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \to \int\int \kappa(\gamma,\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \quad (10)$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\int\int g_j(\gamma)\widehat{\varepsilon}_{in}(\gamma)\widehat{\varepsilon}_{in}(\widetilde{\gamma})g_{\widetilde{j}}(\gamma)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) = \frac{1}{n}\sum_{i=1}^{n}\int\int g_j(\gamma)\varepsilon_i(\gamma)\varepsilon_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$$
$$-\left\{\frac{1}{n}\sum_{i=1}^{n}\int g_j(\gamma)\varepsilon_i(\gamma)d\mathbb{Q}(\gamma)\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}\int \varepsilon_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\right\}$$

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

$$\left|\frac{1}{n}\sum_{i=1}^{n}g_j(\gamma)G_i(\gamma)G_i(\widetilde{\gamma})g_{\widetilde{j}}(\widetilde{\gamma})\right| \le \left(\frac{1}{n}\sum_{i=1}^{n}M_i^2\right)\times|g_j(\gamma)|\times|g_{\widetilde{j}}(\widetilde{\gamma})| \text{ and}$$

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

uniformly in γ and $\widetilde{\gamma}$. Note that the RHSs of these inequalities are in $L_1(\mathbb{Q}\times\mathbb{Q})$ a.s. 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

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\int\int g_j(\gamma)[G_i(\gamma)-\mu(\gamma)][G_i(\widetilde{\gamma})-\mu(\widetilde{\gamma})]g_{\widetilde{j}}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})&\\ \to \int\int g_j(\gamma)\kappa(\gamma,\widetilde{\gamma})g_{\widetilde{j}}(\gamma)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \text{ a.s.}&\qquad (11)\end{aligned}$$

Finally, collecting all the elements in (9), (10), and (11) 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

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

using the facts that $\widetilde{\varepsilon}_{in}=\ddot{\varepsilon}_{in}+\{\mu(\gamma)-\widetilde{\delta}_{0n}-\widetilde{\boldsymbol{\delta}}_n^\top\mathbf{g}(\gamma)\}$ and the FOC that $n^{-1}\sum_{i=1}^n\int\{\widehat{G}_i(\gamma)-\widetilde{\delta}_{0n}-\widetilde{\boldsymbol{\delta}}_n^\top\mathbf{g}(\gamma)\}d\mathbb{Q}(\gamma)=0$, where $\ddot{\varepsilon}_{in}:=\widehat{G}_i-\mu$. Given this, we already proved in the proof of Theorem 3 that $n^{-1}\sum_{i=1}^n\int\ddot{\varepsilon}_{in}(\gamma)\to 0$ a.s. 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

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

We examine each element in the RHS. First,

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

by Assumption 12(i), Theorem 3, and the continuity of G_i with respect to $\boldsymbol{\theta}$, implying that

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

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

$$\frac{1}{n}\sum_{i=1}^{n}\int\int\ddot{\varepsilon}_{in}(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\to\int\int\kappa(\gamma,\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \text{ a.s.} \tag{12}$$

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

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

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

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

Given this, from the facts that $n^{-1}\sum_{i=1}^{n}\int\ddot{\varepsilon}_{in}\to 0$ a.s. and that $n^{-1}\sum_{i=1}^{n}\int\ddot{\varepsilon}_{in}g_j\to 0$ a.s., it follows that $n^{-1}\sum_{i=1}^{n}\int\widetilde{G}_i(\gamma,\widehat{\boldsymbol{\theta}}_n)d\mathbb{Q}(\gamma)\to\int\mu(\gamma)d\mathbb{Q}(\gamma)$ a.s. and that $n^{-1}\sum_{i=1}^{n}\int\widetilde{G}_i(\widetilde{\gamma},\widehat{\boldsymbol{\theta}}_n)g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})\to\int\mu(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\widetilde{\gamma})$ a.s. respectively. Further, using the Cauchy-Schwarz inequality, Assumption 4(ii), and

Assumption 7(*iii*) shows that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\widetilde{G}_i(\gamma,\boldsymbol{\theta})\widetilde{G}_i(\widetilde{\gamma},\boldsymbol{\theta})g_j(\widetilde{\gamma})\right| \leq \left|\frac{1}{n}\sum_{i=1}^{n}M_i^2\right| \times |g_j(\widetilde{\gamma})|$$

uniformly in γ, $\widetilde{\gamma}$, and $\boldsymbol{\theta}$. Note that the RHS of this inequality is in $L_1(\mathbb{Q})$ a.s. 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

$$\frac{1}{n}\sum_{i=1}^{n}\int\int\ddot{\varepsilon}_{in}(\gamma)\ddot{\varepsilon}_{in}(\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \to \int\int\kappa(\gamma,\widetilde{\gamma})g_j(\widetilde{\gamma})d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \text{ a.s.} \tag{13}$$

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 $n^{-1}\sum_{i=1}^{n}\{\int[G_i(\gamma)-\widehat{\delta}_{0n}-\widehat{\boldsymbol{\delta}}_n^\top\mathbf{g}(\gamma)]g_j(\gamma)\}d\mathbb{Q}(\gamma)=0$ and $n^{-1}\sum_{i=1}^{n}\{\int[G_i(\widetilde{\gamma})-\widehat{\delta}_{0n}-\widehat{\boldsymbol{\delta}}_n^\top\mathbf{g}(\widetilde{\gamma})]g_{\widetilde{j}}(\widetilde{\gamma})\}d\mathbb{Q}(\widetilde{\gamma})$ $=0$ imply

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

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

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

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

$$\left| n^{-1} \sum_{i=1}^{n} g_j(\gamma) \widetilde{G}_i(\gamma, \boldsymbol{\theta}) \widetilde{G}_i(\widetilde{\gamma}, \boldsymbol{\theta}) g_{\widetilde{j}}(\widetilde{\gamma}) \right| \le \left(n^{-1} \sum_{i=1}^{n} M_i^2 \right) \times |g_j(\gamma)| \times |g_{\widetilde{j}}(\widetilde{\gamma})|$$

uniformly in γ, $\widetilde{\gamma}$, and $\boldsymbol{\theta}$. Note that the RHS of this inequality is in $L_1(\mathbb{Q} \times \mathbb{Q})$ a.s., 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

$$\frac{1}{n} \sum_{i=1}^{n} \int \int g_j(\gamma) \widetilde{\varepsilon}_{in}(\gamma) \widetilde{\varepsilon}_{in}(\widetilde{\gamma}) g_{\widetilde{j}}(\widetilde{\gamma}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \to \int \int g_j(\gamma) \kappa(\gamma, \widetilde{\gamma}) g_{\widetilde{j}}(\gamma) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \ \text{ a.s.} \tag{14}$$

Finally, collecting all the elements in (12)–(14) 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. We already saw in the proof of Theorem 4(ii) that $\widetilde{\mathbf{D}}_n \to \mathbf{D}^*$ a.s. Therefore, we only prove here that $\widetilde{\mathbf{K}}_n \to \mathbf{K}^*$ a.s. Note that

$$\widetilde{\mathbf{K}}_n = \frac{1}{n} \sum_{i=1}^{n} \int \boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n) \ddot{\varepsilon}_{in}(\gamma) \mathbf{g}(\gamma)^\top d\mathbb{Q}(\gamma) + \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n) \int \{\mu(\gamma) - \widetilde{\delta}_{0n} - \mathbf{g}(\gamma)^\top \widetilde{\boldsymbol{\delta}}_n\} \mathbf{g}(\gamma)^\top d\mathbb{Q}(\gamma), \tag{15}$$

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

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

Third, we consider the first element in (15), 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 $\widetilde{j} = 1, 2, \cdots, k+1$,

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

by Assumption 7(*iii*). Given this, $\widehat{\mathbf{I}}_n$ is finite a.s. and converges to $\mathbf{I}^*$ a.s. by Assumption 8(*iii*), implying that for each $j = 1, 2, \cdots, m$, $n^{-1}\sum_{i=1}^{n}\boldsymbol{s}_{ij}(\widehat{\boldsymbol{\theta}}_n)^2$ is finite a.s. 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,

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n)\ddot{\varepsilon}_{in}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n)\widehat{G}_i(\gamma) - \mu(\gamma)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n)$$

by the definition of $\ddot{\varepsilon}_{in}$; and Assumption 7(*iii*) and $\sum_{i=1}^{n}\boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n) = o_{a.s.}(n)$ imply that $\mu(\gamma)\sum_{i=1}^{n}\boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n) = o_{a.s.}(n)$ uniformly in γ. Furthermore, Assumption 12(*ii*) and the continuity of $\boldsymbol{s}_i$ and G_i with respect to $\boldsymbol{\theta}$ by Assumptions 8(*iii.a*) and 7(*iii*) imply that for each γ,

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

because $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^*$ a.s. by Assumption 8(*i*). We note that $E_{\mathbb{P}}[\boldsymbol{s}_i(\boldsymbol{\theta}^*)G_i(\gamma,\boldsymbol{\theta}^*)] = \boldsymbol{\kappa}_0(\gamma)$ from the IID condition and the condition in Assumption 8(*iii.a*) that $\sqrt{n}\boldsymbol{s}_n^* = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{s}_i(\,\cdot\,,\boldsymbol{\theta}^*) + o_{\mathbb{P}}(1)$. Therefore, $n^{-1}\sum_{i=1}^{n}\int\boldsymbol{s}_i(\widehat{\boldsymbol{\theta}}_n)\ddot{\varepsilon}_{in}(\gamma)\mathbf{g}(\gamma)^{\top}d\mathbb{Q}(\gamma) \to \int\boldsymbol{\kappa}_0(\gamma)\mathbf{g}(\gamma)^{\top}d\mathbb{Q}(\gamma)$. Finally, collecting all these together implies that

$$\widetilde{\mathbf{K}}_n = \int\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i\ddot{\varepsilon}_{in}(\gamma)\widetilde{\mathbf{g}}(\gamma)^{\top}d\mathbb{Q}(\gamma) + o_{a.s.}(1) = \int\boldsymbol{\kappa}_0(\gamma)\widetilde{\mathbf{g}}(\gamma)^{\top}d\mathbb{Q}(\gamma) + o_{a.s.}(1),$$

and this completes the proof. ■

Proof of Theorem 7: (*i*) $\sqrt{n}\mathbf{S}_j(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}^*) \overset{\text{A}}{\sim} N(\mathbf{0},\boldsymbol{\Gamma}_j)$ by Theorem 2, where $\boldsymbol{\Gamma}_j := \mathbf{S}_j\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{S}_j^{\top}$, so that it follows that $\boldsymbol{\Gamma}_j^{-1/2}\sqrt{n}\mathbf{S}_j(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}^*) \overset{\text{A}}{\sim} N(\mathbf{0},\mathbf{I}_{k+2-j})$. Because $\widehat{\mathbf{B}}_n \to \mathbf{B}$ a.s. as given in Theorem 5,

$\widehat{\mathbf{\Gamma}}_{nj} \to \mathbf{\Gamma}_j$ a.s. by proposition 2.30 of White (2001), where $\widehat{\mathbf{\Gamma}}_{nj} := \mathbf{S}_j\mathbf{A}^{-1}\widehat{\mathbf{B}}_n\mathbf{A}^{-1}\mathbf{S}_j^\top$. Therefore,

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

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

$$\mathcal{W}_{j,n} = \mathcal{M}_{j,n} + 2n\boldsymbol{\xi}^{*\top}\mathbf{S}_j^\top\widehat{\mathbf{\Gamma}}_n^{-1}\mathbf{S}_j(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}^*) + n\boldsymbol{\xi}^{*\top}\mathbf{S}_j^\top\widehat{\mathbf{\Gamma}}_n^{-1}\mathbf{S}_j\boldsymbol{\xi}^*.$$

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

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

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

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

(*iii*) $\sqrt{n}\mathbf{S}_j\widetilde{\boldsymbol{\xi}}_n \overset{\text{A}}{\sim} N(\varsigma_j, \mathbf{\Gamma}_j^*)$ by Theorem 4(*ii*), and $\widetilde{\mathbf{B}}_n^* \to \mathbf{B}^*$ a.s. 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

ALTMAN, M. AND M. MCDONALD (2003): "Replication with Attention to Accuracy," *Political Analysis*, 11, 302–307.

ANDREWS, D. W. (1994): "Empirical Process Methods in Econometrics," Armsterdam: Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 37, 2247–2294.

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

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

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

BÉNABOU, R. (2000): "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, 90, 96–129.

BERRY, W., J. DEMERITT, AND J. ESARLEY (2010): "Testing for Interaction in Binary Logit and Probit Models: Is a Product Terms Essential?" *American Journal of Political Science*, 54, 248–266.

BESLEY, T. AND A. CASE (2003): "Political Institutions and Policy Choices: Evidence from the United States," *Journal of Economic Literature*, 41, 7–73.

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

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

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

BIERENS, H. J. AND W. PLOBERGER (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

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

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

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

CHO, J. S. AND I. ISHIDA (2012): "Testing for the effects of omitted power transformations," *Economics Letters*, 117, 287–290.

CHO, J. S., I. ISHIDA, AND H. WHITE (2011): "Revisiting Tests for Neglected Nonlinearity Using Artificial Neural Networks," *Neural Computation*, 23, 1133–1186.

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

CHO, J. S., P. C. B. PHILLIPS, AND J. SEO (2022): "Parametric Conditional Mean Inference with Functional Data Applied to Lifetime Income Curves," *International Economic Review*, 170, 391–456.

——— (2023): "Functional Data Inference in a Parametric Quantile Model Applied to Lifetime Income Curves," Working Paper, School of Economics, Yonsei University.

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

——— (2010): "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

——— (2011): "Testing Correct Model Specification Using Extreme Learning Machines," *Neurocomputing*, 74, 2552–2565, advances in Extreme Learning Machine: Theory and Applications Biological Inspired Systems. Computational and Ambient Intelligence.

CONTRERAS-REYES, J. E., F. O. L. QUINTERO, AND R. WIFF (2018): "Bayesian Modeling of Individual Growth Variability Using Back-Calculation: Application to Pink Cusk-Eel (Genypterus Blacodes) off Chile," *Ecological Modelling*, 385, 145–153.

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

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

——— (1987): “Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternatives,” *Biometrika*, 74, 33–43.

DELLACHERIE, C. AND P.-A. MEYER (1978): “Stochastic processes,” Armsterdam: North-Holland, vol. 29 of *North-Holland Mathematics Studies*, chap. IV, 83–154.

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

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): “Pseudo Maximum Likelihood Methods: Applications to Poisson Models,” *Econometrica*, 52, 701–20.

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

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

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

JENNRICH, R. I. (1969): “Asymptotic Properties of Non-Linear Least Squares Estimators,” *The Annals of Mathematical Statistics*, 40, 633–643.

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

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

LI, D., P. M. ROBINSON, AND H. L. SHANG (2020): “Long-Range Dependent Curve Time Series,” *Journal of the American Statistical Association*, 115, 957–971.

NAGLER, J. (1991): “The Effect of Registration Laws and Education on U.S. Voter Turnout,” *American Political Science Review*, 85, 1393–1405.

——— (1994): “Scobit: An Alternative Estimator to Logit and Probit,” *American Journal of Political Science*, 38, 230–255.

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167.

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

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

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

SHIN, K. AND J. S. CHO (2013): "Testing for Neglected Nonlinearity Using Extreme Learning Machines," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 21, 117–129.

SHORACK, G. R. AND J. A. WELLNER (1986): *Empirical Processes with Applications to Statistics*, New York: John Wiley & Sons.

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

——— (1998): "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory*, 14, 295–325.

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

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

VAN ZWET, W. R. AND J. OOSTERHOFF (1967): "On the Combination of Independent Test Statistics," *The Annals of Mathematical Statistics*, 38, 659–680.

WESTBERG, M. (1985): "Combining Independent Statistical Tests," *Journal of the Royal Statistical Society. Series D (The Statistician)*, 34, 287–296.

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

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

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

WHITE, H. AND J. S. CHO (2012): "Higher-Order Approximations for Testing Neglected Nonlinearity," *Neural Computation*, 24, 273–287.

WOLFINGER, R. AND S. ROSENSTONE (1980): *Who Votes?*, New Haven: Yale University Press.

WOOLDRIDGE, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, vol. 1, Cambridge: The MIT Press.

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

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
$\widetilde{\mathcal{W}}_{1,n}(\text{con})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+lin})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+quad})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+lin+quad})$	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 1: EMPIRICAL LEVELS OF THE WALD AND BP TESTS (NUMBER OF REPLICATIONS: 10,000) This table supposes the panel data example in Section 6.1 and shows the empirical levels of the Wald tests employing various functional regressors for the levels of significance 1%, 5%, and 10%. It further compares the empirical levels of the Wald tests with BP-test in Breusch and Pagan (1979).

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
$\widetilde{\mathcal{W}}_{1,n}(\text{con})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+lin})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+quad})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+lin+quad})$	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
$\widetilde{\mathcal{W}}_{1,n}(\text{con+}0.5^{\gamma})$	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 2: POWERS OF THE WALD AND BP TESTS (NUMBER OF REPLICATIONS: 5,000, NOMINAL LEVEL: 5%) This table supposes the panel data example in Section 6.1 and shows the empirical powers of the Wald tests employing various functional regressors. It further compares the empirical powers of the Wald tests with BP-test in Breusch and Pagan (1979).

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
$\mathcal{W}_{1,n}^{*}(\text{con})$	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
$\mathcal{W}_{1,n}^{*}(\text{con+lin})$	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
$\mathcal{W}_{1,n}^{*}(\text{con+quad})$	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
$\mathcal{W}_{1,n}^{*}(\text{con+lin+quad})$	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 3: EMPIRICAL LEVELS OF THE TESTS (NUMBER OF REPLICATIONS: 10,000) This table supposes the specification test example in Section 6.2 and shows the empirical levels of the Wald tests employing various functional regressors for the levels of significance 1%, 5%, and 10%. It further compares the empirical levels of the Wald tests with the tests in Bierens (1990) and Stinchcombe and White (1998).

Statistics	$\pi^* \setminus n$	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
$\mathcal{W}^*_{1,n}(\text{con})$	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
$\mathcal{W}^*_{1,n}(\text{con+lin})$	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
$\mathcal{W}^*_{1,n}(\text{con+quad})$	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
$\mathcal{W}^*_{1,n}(\text{con+lin+quad})$	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

Table 4: EMPIRICAL POWERS OF THE TESTS (NUMBER OF REPLICATIONS: 5,000, NOMINAL LEVEL: 5%) This table supposes the specification test example in Section 6.2 and shows the empirical powers of the Wald tests employing various functional regressors. It further compares the empirical powers of the Wald tests with the tests in Bierens (1990) and Stinchcombe and White (1998).

Statistics \ Probit Models	w/ Products	w/o Products
Constant	-2.7431	-2.5229
	(0.0000)	(0.0000)
Closing Date	0.0006	-0.0078
	(0.8685)	(0.0000)
Education	0.2645	0.1818
	(0.0000)	(0.0000)
Education^2	0.0050	0.0123
	(0.2433)	(0.0000)
Age	0.06965	0.0697
	(0.0000)	(0.0000)
Age^2	-0.0005	-0.0005
	(0.0000)	(0.0000)
South	-0.1154	-0.1159
	(0.0000)	(0.0000)
Gubernatorial Election	0.0034	0.0034
	(0.7670)	(0.7666)
Closing Date $\times$ *Education*	-0.0031	
	(0.0399)	
Closing Date $\times$ *Education*^{2}	0.0002	
	(0.0075)	
Sample Size	99,676	99,676
Log-Likelihood	-55,815.28	-55,818.03

Table 5: EMPIRICAL MODEL ESTIMATIONS. Two probit models for voting turnout are estimated using 1984 presidential data of U.S. provided by Altman and McDonald (2003). The probit model without products is the one specified by Wolfinger and Rosenstone (1980), whereas the probit model with products is the one specified by Nagler (1991). The figures in parentheses stand for the p-values. The p-values of the parameter estimates are computed by heteroskedasticity robust consistent standard errors.

Probit Models	$\psi(\cdot)$	$g(\gamma) \setminus Z_t$	*Closing Date*	*Education*	*Age*
w/o Products	Exponential	γ	5.7287	12.0473	4.6968
			(0.0570)	(0.0024)	(0.0955)
		γ^2	6.5866	13.8395	5.2719
			(0.0371)	(0.0010)	(0.0717)
	Logistic	γ	3.4938	1.2557	3.7292
			(0.1743)	(0.5337)	(0.1550)
		γ^2	4.4657	10.4633	3.9043
			(0.1072)	(0.0053)	(0.1420)
w/ Products	Exponential	γ	10.2576	11.9407	6.0734
			(0.0059)	(0.0026)	(0.0480)
		γ^2	10.9863	13.7240	5.1820
			(0.0041)	(0.0010)	(0.0749)
	Logistic	γ	3.3729	1.2217	3.7026
			(0.1852)	(0.5429)	(0.1570)
		γ^2	4.3062	10.2814	3.8766
			(0.1161)	(0.0059)	(0.1440)

Table 6: WALD TESTS AND p-VALUES. Wald tests are provided as specification tests by applying the functional regression in Section 6.2. We let $\gamma \in \Gamma := [-1, 1]$. Two analytic functions are employed for $\psi(\cdot)$: exponential and logistic function; two functional regressors are employed for $g(\gamma)$: γ and γ^2; and finally, three variables are employed for Z_t: *Closing Date*, *Education*, and *Age*. The figures in parentheses stand for the p-values of the Wald tests computed by chi-square distribution with two degrees of freedom.