Parametric Inference on the Mean of Functional Data

JIN SEO CHO

School of Economics, Yonsei University

50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Korea

Email: jinseocho@yonsei.ac.kr

JUWON SEO

Department of Economics, National University of Singapore

AS2 #05-28, 1 Arts Link, Singapore 117570

Email: ecssj@nus.edu.sg

This version: Jan 6th, 2019

Abstract

We consider estimating the population mean of functional data by minimizing the functional mean squared error. We assume a possibly misspecified parametric model for the population mean function and present an appropriate set of regularity conditions under which the consistency and asymptotic normality of our estimator are achieved. We also discuss the situations where the asymptotic properties of our estimator are influenced by the estimation errors of nuisance parameters. Based on the results, we further study statistical inferences by extending the standard Wald, Lagrange multiplier, and quasi-likelihood ratio tests to the framework of our functional data analysis. The asymptotic behaviors of the tests are derived under the null and altenative hypotheses. To illustrate the use of our methodology, we apply our tests to a certain class of distribution specification tests and random coefficient tests.

Key Words: Functional data; Mean function; Wald test; Lagrange multiplier test; Quasi-likelihood ratio test.

Subject Class: C11, C12, C80.

1 Introduction

Functional data analysis (FDA) has obtained increasing attention in the literature of data analysis due to its wide applicability, in which random functional realization is regarded as an individual observation rather than an individual data point, leading to the analysis of continuous phenomenon such as time trends or change of rates under weaker assumptions than traditional analysis. See Ramsay and Dalzell (1991), Rice and Silverman (1991), Ramsay and Silverman (1997) for the classical reference of FDA. We also refer to Ramsay and Silverman (2002), Cai and Hall (2006), Ferraty and Vieu (2006), Cardot et al. (2007), Hall and Horowitz (2007), Zhang and Chen (2007), Müller et al. (2011), Cao et al. (2012), and Müller (2012) for some recent developments.

The theoretical development in FDA tends to center mainly around nonparametric model analyses with a focus on the functional regression models, leading to nonparametric functions as the estimated coefficients or estimated mean functions. Due to this, it is practically difficult to obtain convincing and intuitive interpretations from the estimates in many instances.

In this paper, we aim to seek a suitable methodology that provides more interpretable and meaningful analysis. For this goal, we propose a novel framework for the estimation and inference of the (conditional) mean function of functional data. Our approach is significantly different from other previous studies in that we assume a parametric model for the mean function which can be possibly misspecified, whereas the observations remain as nonparametric random elements in a measurable function space. We study the influence of the misspecification of the mean function on our estimator by examining it in a parallel to the analysis for the quasi-maximum likelihood estimation in White (1982, 1994).

Our approach has several advantages. Firstly, it allows us to construct simple statistical tests for the (conditional) mean functions using the estimated parameters based on our analysis. In the nonparametric context, the inferences on the slope function in functional regression are often technically challenging and accompanied with many issues due to the well known ill-posed problem. On the other hand, the relevant null hypothesis can be easily tested by estimating scalar valued parameters in our framework, enabling us to infer on the mean function straightforwardly. Furthermore, our approach provides a convenient way to study the derivatives of population mean functions. Particularly in economic literature, a variety of rates (e.g., growth rates of GDP or growth rates of income) are of significant interest. Compared to the nonparametric approach using numerical differentiation, our parametric model estimation enables us to analytically estimate the exact derivative. For example, statistical analysis for FDA such as shift registration alignment requires the estimation of the mean function and its derivative, for which we can conveniently apply our analysis.

Interestingly enough, some well-known econometric analysis motivate the analysis framework of our functional data analysis. For instance, there is a large literature that considers combining multiple independent statistics by their weighting method considering power and convenient use of test statistics (e.g., Fisher, 1932; Pearson, 1950; Lancaster, 1961; van Zwet and Oosterhoff, 1967; Westberg, 1985, among others). Given that almost surely constant random functions can be regarded as random variables, all these can be viewed as a special case of the functional data analysis of here with a particular form of weighting functions, thereby facilitating our analysis as a generalized approach of the existing various results on random variables in the literature. As another literature relevant to FDA, the unidentified model analysis in Davies (1977, 1987) and the minimum distance testing in Pollard (1980) strongly motivate our approach of here, in which we may interpret the classical test statistics, say likelihood-ratio test statistic, as statistics obtained from functional observations. There is an affluent literature extending Davies's (1977, 1987) and Pollard's (1980) model inference, and we can apply our approach to similar problems with different identification features or empirical distributions (e.g., Hansen, 1996; Andrews, 2001; Baek et al., 2015; Bierens, 1990; Cho et al., 2018; Cho and White, 2007, 2010, 2017; Stinchcombe and White, 1998, and the references therein). Each observation forming their statistics is a functional observation, so that we can apply our approach of here to their model analysis.

The paper is organized as follows. In Section 2, we formally define the data and model. We also define the functional mean squared error (FMSE) as our estimation criterion. In Section 3, we propose the functional least squares (FLS) estimator for parametric conditional mean function. Asymptotic consistency and normality of the parametric FLS estimator is proved under the two circumstances where we have nuisance effect or no nuisance effect. We also discuss how to estimate the covariance matrices that appear in the limit distribution. In Section 4, we provide a general framework for inferences on a range of parametric hypotheses by extending the standard Wald, Lagrange multiplier (LM), and quasi-likelihood ratio (QLR) tests to FDA. The asymptotic behaviors of the tests are derived under the null and alternative hypotheses. In Section 5, we illustrate use of our methodology by applying it to a certain class of distribution specification tests and random coefficient tests. Finally, we conclude and discuss directions for future research in Section 6. All the mathematical proofs are collected to Appendix.

2 Setup

Suppose that we are interested in examining a data set composed of a set of random variables and random functions, which are given as

$$\{(g_i(\,\cdot\,), x_i')' : g_i : \Gamma \mapsto \mathbb{R} \quad \text{and} \quad x_i \in \mathbb{R}^k\}_{i=1}^n, \tag{1}$$

where n is the sample size and k is an element of $\mathbb{N}$. More formally, we impose the following conditions on the data:

Assumption 1. (Data): *(i)* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a complete probability space and* (Γ, ρ) *is a compact metric space;*

(ii) $\{(g_i(\,\cdot\,,\,\cdot\,), x_i')' : g_i : \Omega \times \Gamma \mapsto \mathbb{R}$ *and* $x_i : \Omega \mapsto \mathbb{R}^k\}_{i=1}^n$ *is a set of identically and independently distributed (IID) observations such that for each* $\gamma \in \Gamma$, $\{(g_i(\,\cdot\,, \gamma), x_i')'\}$ *is measurable* $-\mathcal{F}$*, and* $g_i(\omega, \cdot\,) \in \mathcal{C}^{(0)}(\Gamma)$ *for all* $\omega \in F \in \mathcal{F}$ *with* $\mathbb{P}(F) = 1$*, where* $\mathcal{C}^{(\ell)}(\,\cdot\,)$ *denotes the space of* ℓ*-times continuously differentiable functions;*

(iii) $(\Gamma, \mathcal{G}, \mathbb{Q})$ *and* $(\Omega \times \Gamma, \mathcal{F} \bigotimes \mathcal{G}, \mathbb{P} \cdot \mathbb{Q})$ *are complete probability spaces, and* $g_i(\,\cdot\,,\,\cdot\,)$ *is measurable* $-\mathcal{F} \bigotimes \mathcal{G}$. □

The metric space Γ is the space where the functional observations are defined for a fixed $\omega \in \Omega$. For our convenience, we define the functional observations g_i on the product space of $\Omega \times \Gamma$ rather than interpreting them as elements in Hilbert Space. In Assumption 1(*iii*), the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is combined with $(\Gamma, \mathcal{G}, \mathbb{Q})$ to form the product probability space $(\Omega \times \Gamma, \mathcal{F} \times \mathcal{G}, \mathbb{P} \times \mathbb{Q})$. We call $(\Gamma, \mathcal{G}, \mathbb{Q})$ and $\mathbb{Q}$ the *adjunct probability space* and *adjunct* *probability measure*, respectively. We suppose that these are judiciously chosen by the researcher to the extent of his or her interest. They are not part of the stochastic aspects of data, but the measurability condition is useful to the integrals introduced below. If g_i is a continuous in γ almost surely (a.s.) with respect to $\mathbb{P}$, the joint measurability condition in Assumption 1(*iii*) trivially holds by lemma 2.15 of Stinchcombe and White (1992). Otherwise, the measurability condition has to be verified by considering the function properties. For the sake of convenience, we proceed with our discussions by ensuring the continuity condition. The notation in (1) should be understood as the one that suppressed ω component for our convenience.

We further suppose that our subject of interests is the conditional mean function of g_i defined as

$$\mu(\gamma, x) := \int g(\gamma) d\mathbb{P}(g(\gamma)|x)$$

where $\mathbb{P}(\,\cdot\,|x)$ is the conditional probability measure of $g_i(\gamma)$ on $x_i = x$. For each $\gamma \in \Gamma$, we treat $g(\gamma)$ as a random variable and compute its conditional mean μ. Therefore, if we let $E_{\mathbb{P}}$ denote the expectation operator associated with the probability measure $\mathbb{P}$, $\mu(\gamma, x)$ can be expressed as $E_{\mathbb{P}}[g_i(\gamma)|x_i = x]$. If $g_i(\,\cdot\,)$ is constant a.s., we can view it as a random variable, so that $E_{\mathbb{P}}[g_i(\gamma)|x_i = x]$ becomes the conventional conditional mean of $g_i(\,\cdot\,)$. For a parametric specification of the conditional mean, we define $\mathcal{M}$ to be a collection of parametric models specified by ρ. That is,

$$\mathcal{M} := \{\rho(\,\cdot\,, \theta, x) : \Gamma \mapsto \mathbb{R} | \theta \in \Theta \subset \mathbb{R}^d\}.$$

The following conditions are assumed for $\mathcal{M}$ and ρ:[1]

Assumption 2. (Model): *(i) For each $\theta \in \Theta$, $\rho(\,\cdot\,, \theta,\,\cdot\,) : \Gamma \times \Omega \mapsto \mathbb{R}$ is measurable $-\mathcal{F} \bigotimes \mathcal{G}$, where the parameter space Θ is a compact and convex set in $\mathbb{R}^d$ for $d \in \mathbb{N}$;*

(ii) for each $\gamma \in \Gamma$, $\rho(\gamma,\,\cdot\,, \omega) : \Theta \mapsto \mathbb{R} \in \mathcal{C}^{(2)}(\Theta)$ for all $\omega \in F \in \mathcal{F}$ with $\mathbb{P}(F) = 1$;

(iii) for each $\theta \in \Theta$, $\rho(\,\cdot\,, \theta, \omega) \in \mathcal{C}^{(0)}(\Gamma)$ for all $\omega \in F \in \mathcal{F}$ with $\mathbb{P}(F) = 1$; and

(iv) θ_* is unique and interior to Θ, where $\theta_* := \arg\min_{\theta \in \Theta} q(\theta)$ and $q(\theta) := \int \int \{g(\gamma) - \rho(\gamma, \theta, x)\}^2 d\mathbb{P}(g(\gamma), x)$ $d\mathbb{Q}(\gamma)$. □

Based on the specification of the mean function μ, the functional mean squared error (FMSE) for parametric mean function is defined as $q(\,\cdot\,)$ in analogy to the MSE for the least squares estimation. As in the standard analysis, we assume that there is a unique θ_* in the interior of Θ and avoid possibly non-identified model issues from our consideration. To provide more discussion in a formal manner, we make several technical assumptions on the bounds of g_i and ρ_i.

Assumption 3. (Moments): *For some $m_i \in L^2(\mathbb{P})$,*

(i) $\sup_{\gamma \in \Gamma} |g_i(\gamma)| \leq m_i$*;*

(ii) $\sup_{(\gamma,\theta) \in \Gamma \times \Theta} |\rho(\gamma, \theta, x_i)| \leq m_i$*;*

[1] Note that our framework is significantly different from Bugni, Hall, Horowitz, and Neumann (2009) which concerns parametric specifications of functional observations. On the other hand, we let g_i be a random function with no further specification, although its mean function is parametrically specified.

(iii) for each $j = 1, 2, \ldots, d$, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |(\partial/\partial\theta_j)\rho(\gamma, \theta, x_i)| \leq m_i$;

(iv) for each j *and* $j' = 1, 2, \ldots, d$, $\sup_{\theta\in\Theta} \left|(\partial^2/\partial\theta_j\partial\theta_{j'})\rho(\,\cdot\,, \theta, x_i)\right| \leq m_i$. □

Assumption 3 is imposed to ensure the existence of the global minimum (or minima) of q.

In practice, the functional form of μ is unknown, and our model $\mathcal{M}$ may not contain a parameter value θ such that $\rho(\,\cdot\,, \theta, x) = \mu(\,\cdot\,, x)$ for all x. We say $\mathcal{M}$ is *correctly specified* if there is $\theta_0 \in \Theta$ such that $\mu(\,\cdot\,, x_i) = \rho(\,\cdot\,, \theta_0, x_i)$ a.s.$-\ \mathbb{P} \cdot \mathbb{Q}$. Otherwise, we say that $\mathcal{M}$ is *misspecified*. Theorem 1 sorts out the implication on how to interpret the minimizer of $q(\,\cdot\,)$ under correct specification and misspecification.

Theorem 1. *Given Assumptions 1, 2, and 3, we have*

$$q(\theta) = \int\int \mathrm{var}_{\mathbb{P}}[g_i(\gamma)|x] d\mathbb{P}(x) d\mathbb{Q}(\gamma) + \int\int \{\mu(\gamma, x) - \rho(\gamma, \theta, x)\}^2 d\mathbb{P}(x) d\mathbb{Q}(\gamma),$$

where for each γ, $\mathrm{var}_{\mathbb{P}}[g_i(\gamma)|x] := \int\{g(\gamma) - \mu(\gamma, x)\}^2 d\mathbb{P}(g(\gamma)|x)$. □

When $\mathcal{M}$ is correctly specified, we have $q(\theta_0) = \int \mathrm{var}_{\mathbb{P}}[g_i(\gamma)|x] d\mathbb{P}(x) d\mathbb{Q}(\gamma)$ with $\theta_* = \theta_0$, so that the mean function μ can be identified. In this case, the FMSE cannot be smaller than $\int \mathrm{var}_{\mathbb{P}}[g_i(\gamma)] d\mathbb{Q}(\gamma)$. On the other hand, when $\mathcal{M}$ is not correctly specified, θ_0 cannot be identified by minimizing $q(\,\cdot\,)$. This is because the FMSE is affected by additional error that reflects the effect from model misspecification. In this case, θ_* should be understood as a parameter value of θ which minimizes the sum of two squared errors: one is the mean squared error obtained if the model had been correctly specified, and another is the squared error component arising from model misspecification. In general situations, we may not presume that the model $\mathcal{M}$ is correctly specified unless additional information on the parametric form of the mean function is provided. Henceforward, we assume that $\mathcal{M}$ is possibly misspecified and study the asymptotic properties of our proposed estimator detailed below. For convenience, we abuse our notations and use $\mu_i(\,\cdot\,)$ and $\rho_i(\gamma, \theta)$ to denote $\mu(\,\cdot\,, x_i)$ and $\rho(\gamma, \theta, x_i)$, respectively. We also abbreviate $\int g(x) dF(x)$ and $\int\int k(x, y) dF(x, y)$ into $\int g(x) dF$ and $\int\int k(x, y) dF$, respectively. For a probability measure $\mathbb{P}$ on Ω, we let $L^\ell(\mathbb{P}) := \{f : \int_\Omega |f(\omega)|^\ell d\mathbb{P}(\omega) < \infty\}$, for $\ell = 1$ and 2.

3 Functional Least Squares (FLS)

3.1 FLS Estimator without nuisance effect

In this section, we discuss on how to estimate θ_* based on the FMSE criterion. For this purpose, we let the *functional least squares* (FLS) estimator for the parametric estimation be defined as

$$\widehat{\theta}_n := \underset{\theta\in\Theta}{\arg\min}\, q_n(\theta), \quad \text{where} \quad q_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \int \{g_i(\gamma) - \rho_i(\gamma, \theta)\}^2 d\mathbb{Q}.$$

Note that $q_n(\,\cdot\,)$ is a sample analogue of $q(\,\cdot\,)$ which we call the *functional sample mean squared error* (FSMSE). Under the regularity conditions given above, we can show that the FSMSE uniformly converges to $q(\,\cdot\,)$. Furthermore, it follows from this that the FLS estimator is consistent for θ_* and normally distributed around θ_*. The following theorem first establishes the consistence:

Theorem 2. *Given Assumptions 1, 2, and 3, as n tends to infinity,*

(i) $\sup_{\theta\in\Theta} |q_n(\theta) - q(\theta)| \to 0$ *a.s.*$-\mathbb{P}$*;*

(ii) $\widehat{\theta}_n \to \theta_*$ a.s. $-\mathbb{P}$. □

The uniform consistency of the FSMSE is verified by applying various versions of the strong uniform law of large numbers (SULLN). For example, we can apply the SULLN of Andrews (1992) or Newey (1991) under Assumptions 2 and 3. In proving the SULLN, we repeatedly invoke the dominated convergence theorem (DCT) to interchange the order of discrete summation and integral operators. Note that the moment conditions in Assumption 3 are sufficient for this interchange. The consistency of the FLS estimator follows from the fact that the FSMSE converges to FMSE uniformly on Θ a.s.$-\mathbb{P}$, whenever θ_* is unique.

The asymptotic normality of the FLS estimator is obtained in parallel to the standard least squares estimator. For the asymptotic normality, we begin with observing that for some $\bar{\theta}_n$ between $\widehat{\theta}_n$ and θ_*,

$$\sqrt{n}(\widehat{\theta}_n - \theta_*) = A_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int [g_i(\gamma) - \rho_i(\gamma, \theta_*)]\nabla_\theta \rho_i(\gamma, \theta_*) d\mathbb{Q}(\gamma), \tag{2}$$

where

$$A_n := \frac{1}{n}\sum_{i=1}^{n}\int \left\{\nabla_\theta \rho_i(\gamma, \bar{\theta}_n)\nabla_\theta' \rho_i(\gamma, \bar{\theta}_n) - [g_i(\gamma) - \rho_i(\gamma, \bar{\theta}_n)]\nabla_\theta^2 \rho_i(\gamma, \bar{\theta}_n)\right\} d\mathbb{Q}(\gamma).$$

Here, the asymptotic regular behaviors of A_n and $n^{-1/2}\sum_{i=1}^{n}\int [g_i(\gamma) - \rho_i(\gamma, \theta_*)]\ \nabla_\theta \rho_i(\gamma, \theta_*) d\mathbb{Q}(\gamma)$ are central in establishing the asymptotic normality. In the following, we first formally state our assumptions:

Assumption 4. (Hessian Matrix): $\lambda_{\min}(A) > 0$*, for* $A := \int\int \nabla_\theta \rho(\gamma, \theta_*, x)\nabla_\theta' \rho(\gamma, \theta_*, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma) - \int\int \{\mu(\gamma, x) - \rho(\gamma, \theta_*, x)\}\nabla_\theta^2 \rho(\gamma, \theta_*, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma)$ *where* $\lambda_{\min}(A)$ *is the smallest eigenvalue of* A*.* □

Assumption 5. (CLT): *(i) For each* j *and* $j' = 1, 2, \ldots, d$*,* $\int\int (\partial/\partial\theta_j)\rho(\gamma, \theta_*, x) \cdot \kappa(\gamma, \widetilde{\gamma}|x) \cdot (\partial/\partial\theta_{j'})\rho(\widetilde{\gamma}, \theta_*, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) < \infty$*, where* $\kappa(\gamma, \widetilde{\gamma}|x) := \int \{g(\gamma) - \rho(\gamma, \theta_*, x)\}\{g(\widetilde{\gamma}) - \rho(\widetilde{\gamma}, \theta_*, x)\} d\mathbb{P}(g(\gamma), g(\widetilde{\gamma})|x)$*; and*

(ii) $\lambda_{\min}(B) > 0$*, for* $B := \int\int \nabla_\theta \rho(\gamma, \theta_*, x)\kappa(\gamma, \widetilde{\gamma}|x)\nabla_\theta' \rho(\widetilde{\gamma}, \theta_*, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma})$*.* □

The matrix A is provided as the probability limit of A_n, and B serves as the covariance matrix in the limit distribution of the FLS estimator. Note that the maximum eigenvalues of A and B are both finite by Assumptions 3 and 5(*ii*), respectively. We, therefore, only require that the minimum eigenvalues of A and B be strictly positive as stated in Assumptions 4 and 5(*ii*). Here, the conditional covariance kernel $\kappa(\,\cdot\,, \cdot\,|x)$ in Assumption 5(*i*) contributes to the asymptotic covariance matrix of the FLS estimator through the matrix B. Note that the covariance kernel also signifies the role of $\mathbb{P}$ for the FLS estimator. For a different $\mathbb{P}$, a different functional form is obtained for $\kappa(\,\cdot\,, \cdot\,|x)$, leading to different B. In addition, B also depends on the parametric specification $\rho_i(\cdot, \theta)$, implying that different limit distributions are expected for different models.

The following theorem now states the asymptotic normality implied by the regularity conditions imposed so far:

Theorem 3. *Given Assumptions 1, 2, 3, 4, and 5,* $\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{\text{A}}{\sim} N\left(0, A^{-1}BA^{-1}\right)$. □

Theorem 3 is proved in the Appendix by applying Crámer-Wold's device and employing the multivariate CLT to (2) in parallel to the asymptotic normality of the standard least squares estimator.

3.2 FLS Estimator with nuisance effect

Functional data analysis often involves nuisance effect by its nature. In particular, when data are constructed using aligned discrete observations, such a construction naturally brings up nuisance effect. For this examination, we characterize our functional data as

$$\widetilde{g}_i : \Gamma \times \Xi \mapsto \mathbb{R},$$

where Γ is the same as before, and Ξ is a compact parameter space for a nuisance parameter ξ_*, so that the functional observations are defined on $\Gamma \times \Xi$. We assume that the nuisance parameter ξ_* is identifiable and can be consistently estimated by $\widehat{\xi}_n$ obtained in a preliminary stage before estimating θ_*, from which our functional observations are constructed as $\widehat{g}_i(\,\cdot\,) \equiv \widetilde{g}_i(\,\cdot\,, \widehat{\xi}_n)$. This assumption on data structure generalizes that assumed in Section 3.1 because for some known ξ_*, we can let $g_i(\,\cdot\,)$ be identical to $\widetilde{g}_i(\,\cdot\,, \xi_*)$ of here. Therefore, the data analysis given in this section is also applicable to $\{g_i(\,\cdot\,), x_i\}_{i=1}^n$. Nevertheless, the asymptotic influence of the nuisance effect to the FLS estimator is not negligible in general, modifying the limit behavior of the FLS estimator. In this section, we, therefore, extend the results in Section 3.1 to accommodate the nuisance effect conveyed by the nuisance parameter estimation.

To fix the idea, we let our data set be given as $\{(\widehat{g}_i(\,\cdot\,), x_i')'\}_{i=1}^n$. After replacing g_i with $\widehat{g}_i$, we obtain our FLS estimator by minimizing the functional mean squared error as before:

$$\widetilde{\theta}_n := \underset{\theta \in \Theta}{\arg\min}\, \widehat{q}_n(\theta), \qquad \text{where} \qquad \widehat{q}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n} \int \{\widehat{g}_i(\gamma) - \rho_i(\gamma, \theta)\}^2 d\mathbb{Q}.$$

Henceforward, we refer to $\widetilde{\theta}_n$ the two-stage FLS (TSFLS) estimator for the parametric estimation in the mean function.

The main concern of this section is to examine how the parameter estimation error imbedded in $\widehat{\xi}_n$ changes the asymptotic behavior of the FLS estimator. We tackle this aspect by extending the previous regularity conditions for Theorems 2 and 3 to cope with the nuisance effect. We modify Assumptions 1 and 3 in the following:

Assumption 6. (Data): *(i) Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* (Γ, ρ) *be a complete probability space and a compact metric space respectively.* $\Gamma \subset \mathbb{R}^d$ $(d \in \mathbb{N})$ *and* $\Xi \subset \mathbb{R}^s$ $(s \in \mathbb{N})$ *are compact;*

(ii) $\{(\widetilde{g}_i(\,\cdot\,), x_i')' : \widetilde{g}_i : \Omega \times \Gamma \times \Xi \mapsto \mathbb{R}$ *and* $x_i : \Omega \mapsto \mathbb{R}^k\}_{i=1}^n$ $(k \in \mathbb{N})$ *is a set of IID observations such that*

(ii.a) for each $(\gamma, \xi) \in \Gamma \times \Xi$, $(\widetilde{g}_i(\,\cdot\,), x_i')'$ *is measurable* $-\mathcal{F}$*;*

(ii.b) for each $\xi \in \Xi$, $\widetilde{g}_i(\omega, \,\cdot\,, \xi) \in \mathcal{C}^{(0)}(\Gamma)$ *for all* $\omega \in F \in \mathcal{F}$ *with* $\mathbb{P}(F) = 1$*;*

(ii.c) for each $(\omega, \gamma) \in \Omega \times \Gamma$, $\widetilde{g}_i(\omega, \gamma, \,\cdot\,)$ *is in* $\mathcal{C}^{(1)}(\Xi)$ *for all* $\omega \in F \in \mathcal{F}$ *with* $\mathbb{P}(F) = 1$*;*

(iii) $(\Gamma, \mathcal{G}, \mathbb{Q})$ *and* $(\Omega \times \Gamma, \mathcal{F} \bigotimes \mathcal{G}, \mathbb{P}\cdot\mathbb{Q})$ *are complete probability spaces and for* $i = 1, 2, \ldots$ *and* $\xi \in \Xi$, $\widetilde{g}_i(\,\cdot\,, \,\cdot\,, \xi)$ *is measurable* $-\mathcal{F} \bigotimes \mathcal{G}$. □

Assumption 7. (E-Moments)*: For some* $m_i \in L^2(\mathbb{P})$,

(i) $\sup_{(\gamma,\xi)\in\Gamma\times\Xi} |\widetilde{g}_i(\gamma,\xi)| \leq m_i$*;*

(ii) $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\rho_i(\gamma,\theta)| \leq m_i$*;*

(iii) $\sup_j \sup_{(\gamma,\xi)\in\Gamma\times\Xi} |(\partial/\partial\xi_j)\widetilde{g}_i(\gamma,\xi)| \leq m_i$*;*

(iv) for each $j = 1, 2, \ldots, d$, $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi} |(\partial/\partial\theta_j)\rho_i(\gamma,\theta,\xi)| \leq m_i$*;*

(v) for each j *and* $j' = 1, 2, \ldots, d$, $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi} |(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\gamma,\theta,\xi)| \leq m_i$. □

The consistency of the TSFLS estimator can be verified by investigating the limit behavior of the first-order condition for the TSFLS estimator. For this purpose, note that for some $\bar{\xi}_{n,\gamma}$ between $\widehat{\xi}_n$ and ξ_*,

$$\frac{1}{n}\int\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)$$
$$=\frac{1}{n}\int\sum_{i=1}^{n}\{\widetilde{g}_i(\gamma,\xi_*)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)+\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\theta_*)[\nabla'_\xi\widetilde{g}_i(\gamma,\bar{\xi}_{n,\gamma})]d\mathbb{Q}\cdot(\widehat{\xi}_n-\xi_*). \quad (3)$$

The left side of (3) is the first-order derivative of $\widehat{q}_n(\,\cdot\,)$ with respect to θ evaluated at θ_*, whereas the right side is its Taylor expansion with respect to ξ around ξ_*. Here, Assumption 7(*iii*) makes it straightforward to apply the mean-value theorem and derive the right side from the left side. The main part of proving the consistency of $\widetilde{\theta}_n$ for θ_* is in showing that the quantities in the right side of (3) vanish to zero as n tends to infinity. Since the first component in (3) is trivially shown to converge to zero by applying the proof of Theorem 3, the main proof of this is in showing that the second term in the right side converges to zero in probability. Note that the second component is $O_{\mathbb{P}}(\widehat{\xi}_n - \xi_*)$ because the sample average of the integrals in the second term is $O_{\mathbb{P}}(1)$ by Assumption 7, so that if the deviation $(\widehat{\xi}_n - \xi_*)$ is asymptotically negligible, the first-order condition asymptotically holds at θ_*, leading to the consistency of the TSFLS estimator. For this goal, we impose the following condition:

Assumption 8. (E-Estimator)*: There exists a sequence of measurable functions* $\{\widehat{\xi}_n : \Omega \mapsto \Xi\}$ *such that*

(i) $\widehat{\xi}_n \to \xi_*$ *a.s.*$-\mathbb{P}$*, where* ξ_* *is an interior element in* Ξ. □

The consistency of the TSFLS estimator straightforwardly follows from the conditions we have imposed so far, and we contain it in the following theorem:

Theorem 4. *Given Assumptions 2, 6, 7, and 8(i), as* n *tends to infinity,*

(i) $\sup_{\theta\in\Theta} |\widehat{q}_n(\theta) - q(\theta)| \to 0$ *a.s.*$-\mathbb{P}$*; and*

(ii) $\widetilde{\theta}_n \to \theta_*$ *a.s.*$-\mathbb{P}$. □

Next, we establish the asymptotic normality of the TSFLS. The assumptions for Theorem 3 are not sufficient to derive the asymptotic distribution of the TSFLS estimator as they do not handle the limit distribution of $\widehat{\xi}_n$. To employ Crámer-Wold's device for the multivariate CLT, we further impose the conditions stated in the following assumptions:

Assumption 8. (E-Estimator–continued)*: (ii) for a nonstochastic finite* $s \times s$ *matrix* H *such that* $\lambda_{\min}(H) > 0$ *and a sequence of random vectors* $\{s_{*n}\}$ *measurable* $-\mathcal{F}$, $\sqrt{n}(\widehat{\xi}_n - \xi_*) = -H^{-1}\sqrt{n}s_{*n} + o_{\mathbb{P}}(1)$*; and* *(iii) for $i = 1, 2, \ldots$, there is $s_i : \Omega \times \Xi \mapsto \mathbb{R}^s$ such that*

(iii.a) for each $\xi \in \Xi$, $s_i(\,\cdot\,, \xi)$ is measurable $-\mathcal{F}$ and IID;

(iii.b) $s_i(\omega, \,\cdot\,)$ is continuous for all $\omega \in F \subset \mathcal{F}$, $\mathbb{P}(F) = 1$;

(iii.c) for some $m_i \in L^2(\mathbb{P})$, $|s_i(\omega, \,\cdot\,)| \leq m_i(\omega)$; and

(iii.d) $\sqrt{n} s_{*n} = n^{-1/2} \sum_{i=1}^n s_i(\,\cdot\,, \xi_*) + o_{\mathbb{P}}(1)$ such that for each $j = 1, 2, \ldots, s$, $\mathbb{E}_{\mathbb{P}}[s_{ji}(\,\cdot\,, \xi_*)^2] < \infty$, where $s_{ji}(\,\cdot\,, \xi_*)$ is the j-th row element of $s_i(\,\cdot\,, \xi_*)$. □

Assumptions 8(*ii* and *iii*) presume that the distribution of $\widehat{\xi}_n$ is asymptotically equivalent to the product of the nonstochastic matrix H and the score s_{*n}. We can apply many estimators to fit this formation such as least squares, generalized method of moments, (quasi-)maximum likelihood estimators, and etc. We do not specify how H and s_{*n} are obtained not to lose generality of the current analysis. As s_{*n} is formally defined in Assumption 8, we now use $s_i(\xi)$ to denote $s_i(\omega, \xi)$, suppressing ω.

Assumption 9. (E-CLT)*: Let $J := \mathbb{E}[s_i(\xi_*) s_i(\xi_*)']$, $K := \int \mathbb{E}_{\mathbb{P}}[s_i(\xi_*)\{\widetilde{g}_i(\gamma, \xi_*) - \rho_i(\gamma, \theta_*)\} d\mathbb{Q}(\gamma)$, and*

$$C := \begin{bmatrix} J & K' \\ K & B \end{bmatrix}.$$

The matrix B is equivalently defined as in Assumption 5.

(i) C is positive definite;

(ii) $\lambda_{\min}(B_*) > 0$, where $B_* := B - MH^{-1}K - K'H^{-1\prime}M' + MH^{-1}JH^{-1\prime}M'$ and $M := \int \mathbb{E}_{\mathbb{P}}[\nabla_\theta \rho(\gamma, \theta_*, x_i)$ $\nabla'_\xi \widetilde{g}_i(\gamma, \xi_*)] d\mathbb{Q}(\gamma)$. □

Assumption 9 imposes the properties of the covariance matrix for the TSFLS estimator to derive its limit distribution. It generalizes Assumption 5 to handle the existing estimator $\widehat{\xi}_n$. Note that C is employed to capture the asymptotic covariance matrix of the score for $(\widehat{\xi}'_n, \widetilde{\theta}'_n)'$ that is a channel through which the nuisance effect is conveyed to the limit distribution of the TSFLS estimator.

Finally, we now establish the asymptotic normality of TSFLS in the following theorem:

Theorem 5. *Given Assumptions 2, 4, 6, 7, 8, and 9, $\sqrt{n}(\widetilde{\theta}_n - \theta_*) \overset{\text{A}}{\sim} N(0, A^{-1}B_*A^{-1})$.* □

Note that the nuisance effect is not asymptotically negligible as expected. That is, the asymptotic covariance matrix of $\widetilde{\theta}_n$ is modified from $A^{-1}BA^{-1}$ to $A^{-1}B_*A^{-1}$, rendering the test statistic formulas defined below be different for the FLS and TSFLS estimators.

Before moving to the next section, we note that estimating the unconditional mean function of functional data can be estimated in parallel to estimating the conditional mean function. Due to their similarity in structure and contents, we do not specifically discuss on estimating the unconditional mean function here but contain its discussion in the Appendix.

3.3 Estimation of the Covariance of FLS

The role of the covariance matrices in Theorems 3 and 5 is important as we can exploit them to construct statistical tests for a certain form of hypotheses. In this section, we discuss how to estimate the covariances consistently.

First of all, we discuss the case with no nuisance effect. The covariance matrix is given as $A^{-1}BA^{-1}$ in Theorem 3. Assumption 10 allows us to apply the SULLN to the estimator of A and B, which are denoted as $\widehat{A}_n$ and $\widehat{B}_n$, respectively in Theorem 6. Then, $\widehat{A}_n^{-1}\widehat{B}_n\widehat{A}_n^{-1}$ provides a consistent estimator by applying the Slutsky theorem for random matrices.

Assumption 10. (SULLN*): *Let $\varepsilon_i(\gamma,\theta) := g_i(\gamma) - \rho_i(\gamma,\theta)$. For some $m_i \in L^2(\mathbb{P})$,*

(i) $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\varepsilon_i(\gamma,\theta)| \leq m_i^{1/2}$;

(ii) for $j = 1,2,\ldots,d$, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |(\partial/\partial\theta_j)\rho_i(\gamma,\theta)| \leq m_i^{1/2}$; and

(iii) for each j, $j' = 1,2,\ldots,d$, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\gamma,\theta)| \leq m_i$. □

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

$$\widehat{A}_n := \frac{1}{n}\sum_{i=1}^{n}\int \nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)\nabla_\theta'\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma) - \frac{1}{n}\sum_{i=1}^{n}\int \varepsilon_i(\gamma,\widehat{\theta}_n)\nabla_\theta^2\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma) \text{ and}$$

$$\widehat{B}_n := \frac{1}{n}\sum_{i=1}^{n}\int\int \nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)\varepsilon_i(\gamma,\widehat{\theta}_n)\varepsilon_i(\widetilde{\gamma},\widehat{\theta}_n)\nabla_\theta'\rho_i(\widetilde{\gamma},\widehat{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}).$$ □

Secondly, we consider the models with nuisance effect. Since B_* involves more matrices in its expression, we need further conditions for estimating it consistently. The following conditions are imposed for this purpose:

Assumption 11. (E-Covariance): *(i) For a sequence of measurable functions $\{\widehat{J}_n : \Omega \mapsto \mathbb{R}^{s\times s}\}$, $\widehat{J}_n \to J$ a.s.$-\mathbb{P}$; and*

(ii) for a sequence of measurable functions $\{\widehat{H}_n : \Omega \mapsto \mathbb{R}^{s\times s}\}$, $\widehat{H}_n \to H$ a.s. $-\mathbb{P}$. □

Assumption 12. (SULLN**): *Let $\varepsilon_i(\gamma,\theta,\xi) := g_i(\gamma,\xi) - \rho_i(\gamma,\theta)$. For some $m_i \in L^2(\mathbb{P})$,*

(i) $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi} |\varepsilon_i(\gamma,\theta,\xi)| \leq m_i^{1/2}$;

(ii) for $j = 1,2,\ldots,d$, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |(\partial/\partial\theta_j)\rho_i(\gamma,\theta)| \leq m_i^{1/2}$;

(iii) for each j and $j' = 1,2,\ldots,d$, $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi} |(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\gamma,\theta)| \leq m_i$;

(iv) for each $j = 1,2,\ldots,s$, $\sup_{(\gamma,\xi)\in\Gamma\times\Xi} |(\partial/\partial\xi_j)\widetilde{G}y_i(\gamma,\xi)| \leq m_i$; and

(v) for each $j = 1,2,\ldots,s$, $\sup_{\xi\in\Xi} |s_{ji}(\xi)| \leq m_i$, where $s_{ji}(\xi)$ is the j-th row element of $s_i(\xi)$. □

Under Assumption 11, the two submatrices H and J in B_* can be consistently estimated by exploiting $\widehat{H}_n$ and $\widehat{J}_n$. In general, these estimators are determined by the preliminary estimation procedure of $\widehat{\xi}_n$ and can be easily computed using standard estimation methods. For example, if $\widehat{\xi}_n$ is (quasi-)maximum likelihood estimator, $\widehat{H}_n$ and $\widehat{J}_n$ may be identified as the Hessian matrix of the quasi-likelihood function and the sample average of the products of the first-order derivatives evaluated at $\widehat{\xi}_n$, respectively. Note that Assumptions 12(*ii* and *iii*) are stronger than Assumption 10 because the SULLN is required to hold not only for the parameter space $\Gamma\times\Theta$ but also for Ξ, as well.

Furthermore, Assumptions 12(*iii* and *iv*) require the SULLN to hold on some other random functions, which we use to have consistent estimations for K and M, respectively.

The following Theorem 7 provides consistent estimators for A and B_* under the regularity conditions thus far:

Theorem 7. *Let $\widetilde{\varepsilon}_{in}(\gamma, \theta) := \varepsilon(\gamma, \theta, \widehat{\xi}_n)$. Given Assumptions 2, 6, 8, 9, 11, and 12, $\widetilde{A}_n \to A$ a.s.$-\mathbb{P}$ and $\widetilde{B}_n \to B_*$ a.s.$-\mathbb{P}$, where*

$$\widetilde{A}_n := \frac{1}{n}\sum_{i=1}^{n}\int \nabla_\theta \rho_i(\gamma, \widetilde{\theta}_n)\nabla'_\theta \rho_i(\gamma, \widetilde{\theta}_n) d\mathbb{Q}(\gamma) - \frac{1}{n}\sum_{i=1}^{n}\int \widetilde{\varepsilon}_{in}(\gamma, \widetilde{\theta}_n)\nabla^2_\theta \rho_i(\gamma, \widetilde{\theta}_n) d\mathbb{Q}(\gamma);$$

$$\widetilde{B}_n := \bar{B}_n - \widehat{M}_n \widehat{H}_n^{-1}\widehat{K}_n - \widehat{K}'_n \widehat{H}_n^{-1'}\widehat{M}'_n + \widehat{M}_n \widehat{H}_n^{-1}\widehat{J}_n \widehat{H}_n^{-1'}\widehat{M}'_n;$$

$$\bar{B}_n := \frac{1}{n}\sum_{i=1}^{n}\int\int \nabla_\theta \rho(\gamma, \widetilde{\theta}_n)\widetilde{\varepsilon}_{in}(\gamma, \widetilde{\theta}_n)\widetilde{\varepsilon}_{in}(\widetilde{\gamma}, \widetilde{\theta}_n)\nabla'_\theta \rho(\widetilde{\gamma}, \widetilde{\theta}_n) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma});$$

$$\widehat{M}_n := \frac{1}{n}\sum_{i=1}^{n}\int \nabla_\theta \rho_i(\gamma, \widetilde{\theta}_n)\nabla'_\xi \widetilde{g}_i(\gamma, \widehat{\xi}_n) d\mathbb{Q}(\gamma); \quad and$$

$$\widehat{K}_n := \frac{1}{n}\sum_{i=1}^{n}\int s_i(\widetilde{\theta}_n)\widetilde{\varepsilon}_{in}(\gamma, \widetilde{\theta}_n)\nabla'_\theta \rho_i(\gamma, \widetilde{\theta}_n) d\mathbb{Q}(\gamma). \qquad \square$$

Accordingly, the covariance matrix in Theorem 5 can be consistently estimated by $\widetilde{A}_n^{-1}\bar{B}_n\widetilde{A}_n^{-1}$. If $M = 0$, the limiting distribution of TFLS estimator is identical to that of FLS estimator. Hence, both estimators $\widetilde{A}_n^{-1}\widetilde{B}_n\widetilde{A}^{-1}$ and $\widetilde{A}_n^{-1}\bar{B}_n\widetilde{A}^{-1}$ are consistent.

4 Inferences on the Mean Function

The parametric specification for the conditional mean of functional data is particularly useful when we are to make inference on the mean function. Instead of making inference over Γ, we can test a relevant hypothesis by estimating the unknown parameter θ_*. Henceforward, we extend the standard analysis of Wald, Lagrange multiplier (LM), and quasi likelihood-ratio (QLR) tests to make the inference on functional mean. Suppose that we wish to test the following hypotheses on the mean function:

$$\mathbb{H}_o : h(\theta_*) = 0 \;\text{ versus }\; \mathbb{H}_a : h(\theta_*) \neq 0.$$

We assume that the function $h(\,\cdot\,)$ is known and satisfies Assumption 13:

Assumption 13. (Hypothesis)*:* *(i) $h : \Theta \mapsto \mathbb{R}^r$ is in $\mathcal{C}^{(1)}(\Theta)$ with $r \in \mathbb{N}$ and $r \leq d$; and*

(ii) $D(\theta_*) := \nabla'_\theta h(\theta_*)$ has full rank r. $\square$

Next, we define the constrained FLS (CFLS) and constrained two-stage FLS (CTSFLS) estimators as

$$\ddot{\theta}^\flat_n := \arg\min_{\theta\in\Theta} q_n(\theta) \text{ such that } h(\theta) = 0 \quad \text{and} \quad \ddot{\theta}^\sharp_n := \arg\min_{\theta\in\Theta} \widehat{q}_n(\theta) \text{ such that } h(\theta) = 0.$$

We below use these estimators to define our test statistics later. Since the CFLS and CTSFLS estimators are constrained by the restriction $h(\theta) = 0$, $q_n(\ddot{\theta}_n^\flat)$ and $q_n(\ddot{\theta}_n^\sharp)$ cannot be smaller than $q_n(\widehat{\theta}_n)$ and $\widehat{q}_n(\widetilde{\theta}_n)$, respectively. In a similar way, we define $\theta_\dagger$ to be the minimizer of $q(\theta)$ under the same restriction. That is, $\theta_\dagger := \arg\min_{\theta\in\Theta} q(\theta)$ such that $h(\theta) = 0$. The following Lemma 1 establishes the consistency of the CFLS and CTSFLS estimators and will be repeatedly used to derive the limits of our test statistics in this section.

Lemma 1. *(i) Given Assumptions 1, 2, 3, and 13, $\ddot{\theta}_n^\flat \to \theta_\dagger$ a.s.$-\mathbb{P}$, and $\theta_\dagger = \theta_*$ under $\mathbb{H}_o$; and*

(ii) Given Assumptions 2, 6, 7, 8, and 13, $\ddot{\theta}_n^\sharp \to \theta_\dagger$ a.s.$-\mathbb{P}$, and $\theta_\dagger = \theta_*$ under $\mathbb{H}_o$. □

4.1 Wald test

Following Wald's (1943) test principle, we construct the Wald statistic as follows:

$$\mathcal{W}_n^\flat := nh(\widehat{\theta}_n)'\{\widehat{D}_n\widehat{A}_n^{-1}\widehat{B}_n\widehat{A}_n^{-1}\widehat{D}_n'\}^{-1}h(\widehat{\theta}_n); \quad \text{and} \quad \mathcal{W}_n^\sharp := nh(\widetilde{\theta}_n)'\{\widetilde{D}_n\widetilde{A}_n^{-1}\widetilde{B}_n\widetilde{A}_n^{-1}\widetilde{D}_n'\}^{-1}h(\widetilde{\theta}_n)$$

where $\widehat{D}_n := D(\widehat{\theta}_n)$ and $\widetilde{D}_n := D(\widetilde{\theta}_n)$. All the other notations are the same as in Section 3. The statistic $\mathcal{W}_n^\flat$ is considered for the model without the nuisance effect, whereas $\mathcal{W}_n^\sharp$ is used for the model with the nuisance effect.

Theorem 8 summarizes the asymptotic behavior of the Wald statistics:

Theorem 8. *(i) Given Assumptions 1, 2, 4, 5, 10 and 13,*

(i.a) $\mathcal{W}_n^\flat \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$ under $\mathbb{H}_o$, where $\mathcal{X}^2(a, b)$ denotes the non-central chi-square variable with degree of freedom a and non-centrality parameter b; and

(i.b) for any sequence $\{c_n\}$ such that $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}[\mathcal{W}_n^\flat \geq c_n] = 1$ under $\mathbb{H}_a$; and

(ii) Given Assumptions 2, 4, 6, 8, 9, 11, 12, and 13,

(i.a) $\mathcal{W}_n^\sharp \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$ under $\mathbb{H}_o$; and

(i.b) for any sequence $\{c_n\}$ such that $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}(\mathcal{W}_n^\sharp \geq c_n) = 1$ under $\mathbb{H}_a$. □

That is, the null limit distribution of the Wald test statistic is chi-squared with the degrees of freedom r (where r is the number of restrictions). The result mainly follows from the fact that $h(\widehat{\theta}_n)$ and $h(\widetilde{\theta}_n)$ are asymptotically normal. On the other hand, when the null hypothesis is false, the Wald statistics diverge to infinity with probability one. This proves the consistency of the Wald test.

4.2 Lagrange Multiplier (LM) test

Consider the LM test statistic defined as follows:

$$\mathcal{LM}_n^\flat := \frac{n}{4}\nabla_\theta' q_n(\ddot{\theta}_n^\flat)\widehat{A}_n^{-1}\ddot{D}_n^{\flat\prime}\{\ddot{D}_n^\flat\widehat{A}_n^{-1}\widehat{B}_n\widehat{A}_n^{-1}\ddot{D}_n^{\flat\prime}\}^{-1}\ddot{D}_n^\flat\widehat{A}_n^{-1}\nabla_\theta q_n(\ddot{\theta}_n^\flat) \quad \text{and}$$

$$\mathcal{LM}_n^\sharp := \frac{n}{4}\nabla_\theta'\widehat{q}_n(\ddot{\theta}_n^\sharp)\widetilde{A}_n^{-1}\ddot{D}_n^{\sharp\prime}\{\ddot{D}_n^\sharp\widetilde{A}_n^{-1}\widetilde{B}_n\widetilde{A}_n^{-1}\ddot{D}_n^{\sharp\prime}\}^{-1}\ddot{D}_n^\sharp\widetilde{A}_n^{-1}\nabla_\theta\widehat{q}_n(\ddot{\theta}_n^\sharp),$$

where $\ddot{D}_n^\flat := D(\ddot{\theta}_n^\flat)$ and $\ddot{D}_n^\sharp := D(\ddot{\theta}_n^\sharp)$. Here, $\widehat{B}_n$ and $\widetilde{B}_n$ can be replaced by other consistent estimators for B and B_*, respecively. For example, if we let

$$\ddot{B}_n^\flat := \frac{1}{n}\sum_{i=1}^n \int\int \nabla_\theta \rho_i(\gamma, \ddot{\theta}_n^\flat)\{g_i(\gamma) - \rho_i(\gamma, \ddot{\theta}_n^\flat)\}\{g_i(\widetilde{\gamma}) - \rho_i(\widetilde{\gamma}, \ddot{\theta}_n^\flat)\}\nabla'_\theta \rho_i(\gamma, \ddot{\theta}_n^\flat) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}) \quad \text{and}$$

$$\ddot{B}_n^\sharp := \frac{1}{n}\sum_{i=1}^n \int\int \nabla_\theta \rho_i(\gamma, \ddot{\theta}_n^\sharp)\{g_i(\gamma, \widehat{\xi}_n) - \rho_i(\gamma, \ddot{\theta}_n^\sharp)\}\{g_i(\widetilde{\gamma}, \widehat{\xi}_n) - \rho_i(\widetilde{\gamma}, \ddot{\theta}_n^\sharp)\}\nabla'_\theta \rho_i(\gamma, \ddot{\theta}_n^\sharp) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma}),$$

it is clear that $\ddot{B}_n^\flat$ and $\ddot{B}_n^\sharp$ are consistent estimator for B and B_*, respectively under $\mathbb{H}_o$.

The key argument in establishing the asymptotic behaviors of the LM test statistic is the first-order derivatives of q_n and $\widehat{q}_n$ evaluated at the estimated parameter obtained by imposing the null restriction. Under a set of regularity conditions, we can show that

$$\nabla_\theta q_n(\ddot{\theta}_n^\flat) = -\frac{2}{n}\sum_{i=1}^n \int \{g_i(\gamma) - \rho_i(\gamma, \ddot{\theta}_n^\flat)\}\nabla_\theta \rho_i(\gamma, \ddot{\theta}_n^\flat) d\mathbb{Q}(\gamma) \ \text{ a.s.} - \mathbb{P} \quad \text{and}$$

$$\nabla_\theta \widehat{q}_n(\ddot{\theta}_n^\sharp) = -\frac{2}{n}\sum_{i=1}^n \int \{g_i(\gamma, \widehat{\xi}_n) - \rho_i(\gamma, \ddot{\theta}_n^\sharp)\}\nabla_\theta \rho_i(\gamma, \ddot{\theta}_n^\sharp) d\mathbb{Q}(\gamma) \ \text{ a.s.} - \mathbb{P}$$

(see the Lemma 3(*iii*) in the Appendix). Based on this result, the following theorem follows:

Theorem 9. *(i)Given Assumptions 1, 2, 4, 5, 10 and 13,*

(i.a) $\mathcal{LM}_n^\flat \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$ *under* $\mathbb{H}_o$*; and*

(i.b) for any sequence c_n *such that* $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}(\mathcal{LM}_n^\flat \geq c_n) = 1$ *under* $\mathbb{H}_a$*; and*

(ii) Given Assumptions 2, 4, 6, 8, 9, 11, 12, and 13,

(ii.a) $\mathcal{LM}_n^\sharp \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$ *under* $\mathbb{H}_o$*; and*

(ii.b) for any sequence c_n *such that* $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}(\mathcal{LM}_n^\sharp \geq c_n) = 1$ *under* $\mathbb{H}_a$. □

Theorem 9 delivers the asymptotic behaviors of the LM statistics under the null hypothesis and alternative hypotheses. The LM statistics have the same null limit distributions as the Wald statistics, which is the chi-square distribution with degrees of freedom r. This is mainly because under $\mathbb{H}_o$, both $\nabla_\theta q_n(\ddot{\theta}_n^\flat)$ and $\nabla_\theta q_n(\ddot{\theta}_n^\sharp)$ asymptotically follow the normal distributions with mean zero and the covariance matrices that can be consistently estimated by the weight matrices used for the LM statistics.

4.3 Quasi Likelihood Ratio (QLR) test

Lastly, we examine the QLR statistics defined as

$$\mathcal{QLR}_n^\flat := n\{q_n(\ddot{\theta}_n^\flat) - q_n(\widehat{\theta}_n)\} \quad \text{and} \quad \mathcal{QLR}_n^\sharp := n\{\widehat{q}_n(\ddot{\theta}_n^\sharp) - \widehat{q}_n(\widetilde{\theta}_n)\}$$

for the models without and with nuisance effect, respectively.

Approximating $q_n(\,\cdot\,)$ (resp. $\widehat{q}_n(\,\cdot\,)$) via a second-order Taylor's expansion yields the asymptotic distribution of $\sqrt{n}(\ddot{\theta}_n^{\flat} - \widehat{\theta}_n)$ (resp. $\sqrt{n}(\ddot{\theta}_n^{\sharp} - \widehat{\theta}_n)$), which turns out to be a normal distribution under $\mathbb{H}_o$. When $\mathbb{H}_o$ is not true, this quantity is not bounded in probability, so that we can distinguish the null from the alternative. Nevertheless, the QLR statistics do not follow chi-square distributions unlike the previous cases. The following theorem provides their specific asymptotic behaviors under $\mathbb{H}_o$ and $\mathbb{H}_a$:

Theorem 10. *(i) Given Assumptions 1, 2, 4, 5, 10 and 13,*

(i.a) $\mathcal{QLR}_n^{\flat} \overset{\text{A}}{\sim} W'\{D_* A^{-1} D_*'\}^{-1} W$ *under* $\mathbb{H}_o$*, where* $D_* := D(\theta_*)$*, and* $W \sim N(0, D_* A^{-1} B A^{-1} D_*')$*; and*

(i.b) for any sequence $\{c_n\}$ *such that* $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}(\,\mathcal{QLR}_n^{\flat} \geq c_n) = 1$ *under* $\mathbb{H}_a$*; and*

(ii) Given Assumptions 2, 4, 6, 8, 9, 11, 12, and 13; and

(ii.a) $\mathcal{QLR}_n^{\sharp} \overset{\text{A}}{\sim} W_*'\{D_* A^{-1} D_*'\}^{-1} W_*$ *under* $\mathbb{H}_o$*, where* $W_* \sim N(0, D_* A^{-1} B_* A^{-1} D_*')$*; and*

(ii.b) for any sequence $\{c_n\}$ *such that* $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}(\mathcal{QLR}_n^{\sharp} \geq c_n) = 1$ *under* $\mathbb{H}_a$. □

The null limit distributions of the QLR test statistics are different from standard chi-squared distribution, because the asymptotic covariance of the FLS and TSFLS estimators are different from the limits of the Hessian matrices of FSMSE. That is, the information matrix equality does not holds as for the standard quasi-maximum likelihood estimator. As another remark relevant to Theorem 10, the roles of $\widehat{B}_n$ and $\widetilde{B}_n$ are still critical in applying the QLR test statistics. Note that the QLR statistics do not directly compute them, but they are necessary in obtaining the critical values of the QLR test statistics.

5 Applications and Simulations

In this section, we explain how the theoretical results detailed in Sections 3 and 4 can relate to standard statistical model analyses. We display two examples in which our theorems can be applied. In the first application, we consider the finite mixture models to test whether our observations are from homogeneous or heterogeneous population. In the second application, we study an inference on whether the coefficient of a linear regression model is random or a constant. After analyzing these problems in our framework, we present our Monte Carlo experiments conducted to affirm our theory on the FLS and TSFLS estimators.

5.1 Distribution Specification Tests

Let $f_i(\,\cdot\,;\theta_{i*})$ be a component density function for $i = 1, ..., K$. The component densities can be chosen from the same or different families of distributions. Accordingly, the parameter vector $\theta_* = (\theta_{1*}, \theta_{2*}, ..., \theta_{K*})'$ is defined on the product of each parameter space, $\Theta_1 \times \Theta_2 \times ... \times \Theta_K$. For $\pi_i \in [0, 1]$ with $\sum_{i=1}^{K} \pi_i = 1$, a finite mixture model (Everitt and Hand, 1981; McLachlan and Peel, 2004; Schlattmann, 2009) is defined as

$$f(\,\cdot\,;\pi_1, ..., \pi_K, \theta) = \sum_{i=1}^{K} \pi_i f_i(\,\cdot\,;\theta_i).$$

Such finite mixture models have been popular for constructing a more flexible distribution function or modeling cluster. They are also used for a certain type of distribution specification tests. Here, we focus on the inferences for the homogeneity of samples based on the distribution specification tests. Various types of parametric distributions have been explored in this context such as the mixtures of normal distributions, binomial distributions, gamma distributions and von Mises distributions. We refer to Chernoff and Lander (1995), Liang and Rathouz (1999), Chen et al. (2001), Cho and White (2007, 2010), Fu et al. (2008), Chen and Li (2009), Ning et al. (2009), Niu et al. (2011) and Wong and Li (2014) for the relevant reference.

As a specific example, we consider the following mixture of exponential distributions:

$$f(x;\pi_*,\gamma_*) = (1-\pi_*)\exp(-x) + \pi_*\gamma_*\exp(-\gamma_* x),$$

where $\gamma_* \in \Gamma := [\underline{\gamma}, \bar{\gamma}]$. For simplicity, we assume $\underline{\gamma} > 1$ and $\bar{\gamma} < \infty$. Our main concern here is to test whether $\pi_* = 0$. If $\pi_* = 0$, our observations come from homogeneious population which follows the standard exponential distribution. The standard C(α) test by Neyman (1959) and Davies (1977) can be applied with the test statistic $\sup_{\gamma\in\Gamma} n^{-1/2}\sum_{i=1}^{n} g_i(\gamma)$ employing g_i defined as

$$g_i(\gamma) := \frac{(2\gamma-1)^{1/2}}{\gamma-1}\left\{\gamma\exp[(1-\gamma)x_i] - 1\right\}.$$

Instead of taking the supremum over γ, we consider the Wald, LM and QLR test statistics introduced in Section 4 regarding $g_i : \Gamma \mapsto \mathbb{R}$ as a random function. The critical values of the standard C(α) test statistic are obtained by explicitly exploiting the functional form of g_i, whereas our test statistics do not exploit this feature to apply our test statistics, leading to lower powers for our test statistics than that of the C(α) test statistic. Nevertheless, this lower power should not discourage use of our tests, because assuming the knowledge on the functional form of g_i can be restrictive, as its applicability can be limited.

For our simulations, we specify ρ as follows:

$$\rho(\gamma,\theta_1,\theta_2) := \theta_1 + \theta_2\frac{(\gamma-1)}{(2\gamma-1)^{1/2}}.$$

Note that under the DGP described above, μ is computed as $\mu(\gamma) = \pi_*(\gamma-1)/(2\gamma-1)^{1/2}$, implying that $(\theta_{1*},\theta_{2*}) = (0,0)$ under $\mathbb{H}_o$. Otherwise, $(\theta_{1*},\theta_{2*})$ differs from $(0,0)$. We, hereby, specify the null hypothesis and alternative hypothesis as follows:

$$\mathbb{H}_o : (\theta_{1*},\theta_{2*}) = (0,0) \quad \text{versus} \quad \mathbb{H}_a : (\theta_{1*},\theta_{2*}) \neq (0,0).$$

Table 5.1 displays the size and power of the Wald, LM and QLR test statistics studied in Section 3. Throughout the experiment, the FLS estimator is estimated by Newton-Raphson's iterative procedure, and the associated integrals are computed by Gauss-Legendre numerical approximation method. Γ is chosen to be the interval $[1.5, 2.5]$, and this interval is arbitrarily selected to accommodate the fact that the researcher may not have information on the underlying

DGP. We also let $n = 25, 50, 100, 300$ and 500. The nominal levels are fixed at 1%, 5%, and 10%. In the level panel, we observe that the rejection rates of the three tests approach the nominal levels as the sample size increases. Next, the power is computed through 5,000 replications with the same sample sizes as above, but the nominal level is fixed at 0.05. In particular, we examine the power of the tests by letting π_* vary over the range $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. As we expect, the rejection rates tend to be larger as we move π_* further from zero. As the sample size increases, we also observe that the rejection rates approach unity for a fixed π_*.

<<<<<< Insert Table 5.1. >>>>>

5.2 Inference on the Heterogeneity of Dependence Structure

As another experiment, we apply a similar technique to test for the heterogeneity of dependence structure by applying the mixture model assumption. For this purpose, we suppose that the researcher observes IID observations of more than a single variable. Our interest lies in determining whether the dependence structure of the multivariate variables are homogeneous or not. Even when the univariate margin of each variable remains the same over the whole population, the observations can be heterogeneous as they have a different dependence structure.

We hereby propose test statistics to detect the violation of the homogeneity using the finite copula mixture models. The Sklar's theorem (1959) is useful for this purpose as it allows us to separate the information on the univariate margins from the joint distribution, specifying the associated copula function. There is emerging literature demonstrating the use of copulas for studying dependence structure (Joe, 1997; Nelsen, 2007; Joe, 2014). Besides, there are also a number of papers discussing the applications of the mixture copula models (see Dias and Embrechts, 2004; Chen and Fan, 2006; Hu, 2006; Lai et al., 2009; Diks et al., 2010; Zimmer, 2010; Kosmidis and Karlis, 2016; Loaiza-Maya et al., 2018). To our best knowledge, however, inference on the homogeneity based on the finite mixture of copula models has not been addressed to date.

We proceed by considering a mixture of two distinct bivariate copula component densities c_1 and c_2. That is, for $(u, v) \in [0, 1]^2$,

$$c(u, v; \pi_*, \gamma_{1*}, \gamma_{2*}) = (1 - \pi_*)c_1(u, v; \gamma_{1*}) + \pi_* c_2(u, v; \gamma_{2*})$$

with $\pi_* \in [0, 1]$. More generally, each component density can be of any dimension, and a mixture with more number of component densities can be considered. Nevertheless, we focus on the simple model described above solely for brevity. We further suppose that the IID pairs of data $\{(x_i, y_i)\}_{i=1}^n$ have marginal distributions given as F_X and F_Y, respectively. The inference on the homogeneity of dependence structure can be naturally conducted by examining the null hypothesis that $\pi_* = 0$ (or $\pi_* = 1$) with g_i given by

$$g_i(U_i, V_i; \gamma_1, \gamma_2) := \frac{c_2(U_i, V_i; \gamma_2) - c_1(U_i, V_i; \gamma_1)}{c_1(U_i, V_i; \gamma_1)\sqrt{c^*(U_i, V_i; \gamma_1, \gamma_2) - 1}}$$

where $U_i := F_X(x_i)$, $V_i := F_Y(y_i)$, and

$$c^*(u, v; \gamma_1, \gamma_2) := \int_0^1 \int_0^1 \frac{c_2(u, v; \gamma_2)^2}{c_1(u, v; \gamma_1)} dudv.$$

A practical challenge here arises from the fact that the univariate margins F_X and F_Y are unknown to the researchers. In the first stage estimation, we, therefore, approximate U_i and V_i by $\widehat{U}_i = \widehat{F}_X(x_i)$ and $\widehat{V}_i = \widehat{F}_Y(y_i)$, respectively using the estimates of the marginal distributions, in a similar fashion to the inference functions for margins (IFM) approach (Joe and Jianmeng, 1996; Joe, 2001). Then, the theorems in Section 4.2 can be applied with the functional data are constructed by

$$\widehat{g}_i(\gamma_1, \gamma_2) := \frac{c_2(\widehat{U}_i, \widehat{V}_i; \gamma_2) - c_1(\widehat{U}_i, \widehat{V}_i; \gamma_1)}{c_1(\widehat{U}_i, \widehat{V}_i; \gamma_1)\sqrt{c^*(\widehat{U}_i, \widehat{V}_i; \gamma_1, \gamma_2) - 1}}.$$

A leading example would be the case when c_1 is the density of the independence copula. This provides a novel class of tests for the independence between x_i and y_i. When c_1 is the independence copula density, the functional form of g_i is further simplified as follows:

$$\widehat{g}_i(\gamma_1, \gamma_2) := \frac{c_2(\widehat{U}_i, \widehat{V}_i; \gamma_2) - 1}{\sqrt{c^*(\widehat{U}_i, \widehat{V}_i; \gamma_1, \gamma_2) - 1}}, \qquad \text{where} \qquad c^*(u, v; \gamma_2) = \int_0^1 \int_0^1 c_2(u, v; \gamma_2)^2 dudv.$$

For our experiments, we particularly assume the Farlie-Gumbel-Morgenstern (FGM) copula for c_2, and this lets us have a closed form solution to the relevant integrals. From this, the population mean function can be straightforwardly derived as $\mu(\pi_*, \gamma_{2*}) := \frac{1}{3}\pi_*\gamma_{2*}$, enabling us to specify a simple linear model as a model for the mean function, viz.,

$$\rho(\gamma, \theta_1, \theta_2) := \theta_1 + \theta_2\gamma,$$

where $\gamma \in \Gamma := [0, 1]$. Note that if $\pi_* = 0$, $\mu(\pi_*, \cdot) \equiv 0$, and $\rho(\cdot, \theta_{1*}, \theta_{2*}) \equiv 0$ if and only if $(\theta_{1*}, \theta_{2*})' = (0, 0)'$. We, therefore, specify the null and alternative hypotheses as follows:

$$\mathbb{H}_o : (\theta_{1*}, \theta_{2*}) = (0, 0) \qquad \text{versus} \qquad \mathbb{H}_a : (\theta_{1*}, \theta_{2*}) \neq (0, 0).$$

Our simulations are conducted in the following plan. Firstly, we generate random samples using the FGM copula with margins $N(0, 1)$ and $N(0, 5)$ for x_i and y_i, respectively, and we estimate the means and variances of x_i and y_i by the maximum likelihood estimation in the first stage to obtain $\widehat{U}_i := \Phi(x_i, \widehat{\mu}_{x,n}, \widehat{\sigma}^2_{x,n})$ and $\widehat{V}_i := \Phi(y_i, \widehat{\mu}_{y,n}, \widehat{\sigma}^2_{y,n})$, where $\Phi(\cdot, \mu, \sigma^2)$ denotes the normal distribution function with mean and variance μ and σ^2, respectively, and $(\widehat{\mu}_{x,n}, \widehat{\sigma}^2_{x,n})$ and $(\widehat{\mu}_{y,n}, \widehat{\sigma}^2_{y,n})$ are the maximum likelihood estimators obtained from x and y samples, respectively. Secondly, we let the copula parameter γ_2 be fixed at 0.9. Finally, we repeat independent experiments 10,000 times using the samples with $n =$ 25, 50, 100, 300, 500, and $1,000$.

<<<<<< Insert Table 5.2. >>>>>>

In Table 5.2, we report the empirical rejection rates of our independence tests. As before, the rejection rates for the size panel are computed with fixed nominal levels at 1%, 5%, and 10%. Table 5.2 suggests that when the null hypothesis is true, the rejection rates get closer to the nominal levels as the sample size increases. More specifically, the rejection rates of the Wald and LM tests are close to the nominal levels even with small sample sizes such as $n = 50$ or 100. For the power panel, we let π_* be 0.1, 0.2, 0.3, 0.4, and 0.5 after we fix the nominal level to be 5%. Here, we observe a clear tendency that the rejection rates increase when the value of π_* increases. We also observe that holding π_* constant, the rejection rates of the three tests approach unity, as the sample size increases.

5.3 Inference on Random Coefficient

Standard regression models typically suppose fixed coefficients of explanatory variables. When such assumptions violate, the statistical inferences may be misleading due to the biases in estimating the standard errors, and inference via the random coefficient model becomes a more proper approach, in which the coefficients of the explanatory variables are assumed nonconstant. The random coefficient model is particularly useful in modelling conditional heteroskedasticity or time varying coefficients in the time series models. In particular, there are a number of papers testing the randomness of the coefficient in a regression model (see Hsiao, 1974; Breusch and Pagan, 1979; Ramanathan and Rajarshi, 1992; Swamy and Tavlas, 1995; Akharif et al., 2018).

Consider a simple linear regression model given by

$$y_i = x_i' \beta_i + \delta_*^{1/2} \varepsilon_i, \tag{4}$$

where $z_i \in \mathbb{R}$ is an explanatory variable and $x_i = (1, z_i)'$. We assume that the coefficient β_i potentially contains a random element such that

$$\beta_i := (\psi_{1*}, \psi_{2*})' + \pi_*^{1/2} \Omega^{1/2}(\gamma_*) \nu_i.$$

Here, $(\psi_{1*}, \psi_{2*})'$ is a constant vector and $\nu_i \in \mathbb{R}^2$ is a random component. The matrix $\Omega(\,\cdot\,)$ is assumed to be positive definite uniformly on Γ to which γ_* belongs. Accordingly, when the variance of ν_i is positive, the coefficient $\boldsymbol{\beta}_i$ constant if and only if $\pi_* = 0$.

Random coefficient model is popularly investigated in the literature. For example, Andrews (2001) studies a similar type of model to the random coefficient model. Rosenberg (1973) and Engle and Watson (1985) also extend the random coefficient model to conditional heteroskeastic process in time series context. In addition to these studies, a number of empirical studies exploit the feature of the random coefficient model pertinent to examining the conditional heteroskedatic feature of data.

We relate the random coefficient model to the FDA model analysis. We first note that substituting the expression of β_i into (4) yields the following conditional heteroskedasticity model:

$$y_i = x_i' \psi_* + u_i,$$

where $\psi_* := (\psi_{1*}, \psi_{2*})'$ and $u_i := \pi_*^{1/2} x_i' \Omega^{1/2}(\gamma_*)\nu_i + \delta_*^{1/2}\varepsilon_i$, leading to that

$$\mathrm{var}(u_i|x_i) = \delta_* + \pi_* x_i'\Omega(\gamma_*)x_i = \delta_* + \pi_*[1 + \exp(\gamma_*)x_i^2] \tag{5}$$

by assuming that

$$\Omega(\gamma_*) := \begin{bmatrix} 1 & 0 \\ 0 & \exp(\gamma_*) \end{bmatrix}.$$

We also let $\Gamma = \{\gamma : \gamma \in [0, 1]\}$. Next, we further suppose that the researcher estimates the unknown parameter values by the maximum likelihood estimation. For this process, we further suppose that data are generated according to $(\varepsilon_i, \nu_i')'|x_i \sim N(0, I_3)$ and $z_i \sim$ IID $N(0, 1)$. Thirdly, we test for the random coefficient model property by the FDA framework. For this purpose, we rephrase the given DGP condition to the FLS model framework. Note that if the regression equation (4) is estimated by the maximum likelihood estimation, the following likelihood function is maximized:

$$L_n(\psi, \delta, \gamma, \pi) = -\frac{1}{2}\sum_{i=1}^{n} \ln(\delta + \pi x_i'\Omega(\gamma)x_i) - \frac{1}{2}\sum_{i=1}^{n} \frac{(y_i - x_i'\psi)^2}{(\delta + \pi x_i'\Omega(\gamma)x_i)}.$$

Here, if we suppose that the coefficient is not random, we obtain the following first-order derivative of the likelihood function:

$$\frac{\partial}{\partial \pi} L_n(\psi, \delta, \gamma, \pi) = \frac{1}{2\delta_*^2}\sum_{i=1}^{n}[1 + \exp(\gamma_*)x_i^2]\{(y_i - x_i'\psi_*)^2 - \delta_*\}. \tag{6}$$

Note that $E[(y_i - x_i'\psi_*)^2|x_i] = \delta_*$, so that the conditional mean of (6) equal to zero irrespective of $\gamma \in \Gamma$. On the other hand, if the coefficient is random, the population mean of (6) is obtained as follows:

$$\frac{n\pi_*}{2\delta^2} E\{[1 + \exp(\gamma_*)x_i^2][1 + \exp(\gamma_*)x_i^2]\}$$

using (5), motivating the following random function as a candidate function for $\widetilde{g}_i$:

$$\widetilde{g}_i(\gamma, \psi, \delta) := \{1 + \exp(\gamma)x_i^2\}\{(y_i - x_i'\psi)^2 - \delta\},$$

where we can estimate the unknown parameters ψ_* and δ_* by

$$\widehat{\psi}_n := \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i \quad \text{and} \quad \widehat{\delta}_n := \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\widehat{\psi}_n)^2,$$

respectively. Accordingly, our functional observations can be constructed as $\widehat{g}_i(\gamma) := \widetilde{g}_i(\gamma, \widehat{\delta}_n, \widehat{\psi}_n)$, and we specify the mean function as follows:

$$\rho(\gamma, \theta) := \theta_0 + \theta_1 \exp(\gamma)$$

to test the following hypotheses:

$$\mathbb{H}_o : (\theta_{0*}, \theta_{1*})' = 0 \quad \text{versus} \quad \mathbb{H}_a : (\theta_{0*}, \theta_{1*})' \neq 0.$$

Here, note that

$$\sum_{i=1}^{n} \widehat{g}_i(\gamma) = \sum_{i=1}^{n} x_i^2 \{(y_i - x_i' \widehat{\psi}_n)^2 - \widehat{\delta}_n\} \exp(\gamma),$$

so that intercept term of the sample average of $\{\widehat{g}_i(\cdot)\}$ is identical to zero. This fact conversely implies that the estimated TSFLS estimator for the intercept is identical to zero, viz., $\widetilde{\theta}_{0n} \equiv 0$. Due to this fact, the three test statistics follow the chi-squared distribution with degree of freedom unity.

We conduct our simulations by applying the theorems for the models with nuisance effect and specifying the following additional features: first, we let $(\psi_{1*}, \psi_{2*}, \gamma_*, \delta_*) = (1, 1, 1, 1)$ under both the null and alternative hypotheses, and we let $\pi_* = 0$ under the null condition. For the alternative DGP, we consider various values for π_*, viz., 0.01, 0.02, 0.03, 0.04, and 0.05. Next, we let $(z_i, \nu_i', \varepsilon_i)' \sim$ IID $N(0, I_4)$ so that our model for μ is correctly specified under $\mathbb{H}_o$.

<<<<<< Insert Table 5.3. >>>>>

Table 5.3 displays the empirical rejection rations obtained from the Wald, LM and QLR test statistics to test the rephrased hypothesis. Here are a few things observed in the level panel. First, the null rejection rates are close to the nominal levels for all the three test statistics, when the sample size is large enough. Second, the Wald and LM tests tend to be slightly oversized when the sample sizes are small, while the QLR test performs better. For the power of the test statistics, we consider the models with $\pi_* = 0.1, 0.2, 0.3, 0.4, 0.5$ fixing the significance level at 5%. Apparently, the rejection rates are dependent on the sample size for each choice and the value of of π_* in all the three test statistics. That is, the empirical rejection rates increase as π_* or n increases.

6 Conclusion

In this paper, we consider independently and identically distributed functional data and provide a methodology to estimate the population mean function of the functional data. For this, we suppose a parametric nonlinear model specification for the population mean function and also a set of regularity conditions to estimate this parametric model consistently.

Specifically, we minimize the functional mean squared error with respect to unknown parameters, and this is defined as functional least squares estimator. We show that this estimator is consistent for the parameter minimizing the mean squared error of the difference between the parametric model and the population mean function with respect to adjunct probability measure, which is selected by researchers' interests. Further, we also consider the asymptotic distribution of the functional least squares estimator and provide a certain regularity conditions under which the estimator follows a normal distribution. These regularity conditions are fairly mild for further empirical applications.

Next, the asymptotic distribution of the functional nonlinear least squares estimator is exploited to infer the unknown population mean function. For this, we define Wald, Lagrange multiplier, and quasi-likelihood ratio statistics by following the standard principle for these statistics under numerical data considerations. We consider a null hypothesis specified by a parametric restriction and obtain the null distribution of these statistics and show that they are consistent under the alternative. In particular, these statistics have (noncentral) chi-square distribution under the null, so that their application is straightforward.

Finally, we verify the theory of this paper by Monte Carlo experiments, and for this, the mixture of exponential distributions considered in Davies (1977) and the random coefficient model are reexamined. One of the goals of this experiment is to examine the performances of the statistics defined in this paper. Our experiments show that Wald, Lagrange multiplier, and quasi-likelihood ratio statistics follow the asymptotic distributions as stated in the paper under the null, and also they are consistent under the alternative.

Appendices

A Preliminary Lemmas

Before proving the claims in the text, we first provide some supplementary lemmas to be used for later purpose.

Lemma 2. *Given that there is a measurable function* $h(\,\cdot\,,\theta):\Gamma\mapsto\mathbb{R}$ *on* $(\Gamma,\mathcal{G},\mathbb{Q})$ *for each* $\theta\in\Theta$*, if for each* $\gamma\in\Gamma$*,* $h(\gamma,\,\cdot\,)\in\mathcal{C}^{(1)}(\Theta)$ *and for each* $j\in\{1,2,\ldots,d\}$*,* $\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)h(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})$*, then*

$$\nabla_\theta\int h(\gamma,\theta)d\mathbb{Q}(\gamma)=\int\nabla_\theta h(\gamma,\theta)d\mathbb{Q}(\gamma), \tag{7}$$

where Θ *is a compact and convex set in* $\mathbb{R}^d$ *and* $d\in\mathbb{N}$ *as in the text.* □

Lemma 3. *Given Assumptions 1, 2, 3, and 4, for each* $\theta\in\Theta$*,*

(i) $\nabla_\theta q(\theta)=-2\int\int\{\mu(\gamma,x)-\rho(\gamma,\theta,x)\}\nabla_\theta\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)$*;*

(ii) $\nabla_\theta^2 q(\theta)=2\int\int\left\{\nabla_\theta\rho(\gamma,\theta,x)\nabla_\theta'\rho(\gamma,\theta,x)-\{\mu(\gamma,x)-\rho(\gamma,\theta,x)\}\nabla_\theta^2\rho(\gamma,\theta,x)\right\}d\mathbb{P}(x)d\mathbb{Q}(\gamma)$*;*

(iii) $\nabla_\theta q_n(\theta)=-2n^{-1}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\theta)\}\nabla_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)$ *a.s.*$-\mathbb{P}$*; and*

(iv) $\nabla_\theta^2 q_n(\theta)=2n^{-1}\sum_{i=1}^n\int\nabla_\theta\rho_i(\gamma,\theta)\nabla_\theta'\rho_i(\gamma,\theta)-\{g_i(\gamma)-\rho_i(\gamma,\theta)\}\nabla_\theta^2\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)$ *a.s.*$-\mathbb{P}$. □

Lemma 4. *Given Assumptions 1,* 2*, 3, and 5,* $B=\widetilde{B}$*, where*

$$\widetilde{B}:=\int\int\int\nabla_\theta\rho(\gamma,\theta_*,x)\varepsilon(\gamma,\theta_*)\varepsilon(\widetilde{\gamma},\theta_*)\nabla_\theta'\rho(\gamma,\theta_*,x)d\mathbb{P}(g(\gamma),g(\widetilde{\gamma}),x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$$

and $\varepsilon(\gamma,\theta):=g(\gamma)-\rho(\gamma,\theta,x)$*.* □

Lemma 5. *Given Assumptions 2, 6, 7, and 8, for each* $\theta\in\Theta$*,* *(i)* $\nabla_\theta q(\theta) = -2\int\int\{\mu(\gamma,x)-\rho(\gamma,\theta,x)\}\nabla_\theta\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)$*;*

(ii) $\nabla^2_\theta q(\theta) = 2\int\int\left\{\nabla_\theta\rho(\gamma,\theta,x)\nabla'_\theta\rho(\gamma,\theta,x)-\{\mu(\gamma,x)-\rho(\gamma,\theta,x)\}\nabla^2_\theta\rho(\gamma,\theta,x)\right\}d\mathbb{P}(x)d\mathbb{Q}(\gamma)$*;*

(iii) $\nabla_\theta\widehat{q}_n(\theta) = -2n^{-1}\sum_{i=1}^n\int\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta)\}\nabla_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)$ *a.s.*$-\mathbb{P}$*; and*

(iv) $\nabla^2_\theta\widehat{q}_n(\theta) = 2n^{-1}\sum_{i=1}^n\int\nabla_\theta\rho_i(\gamma,\theta)\nabla'_\theta\rho_i(\gamma,\theta)-\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta)\}\nabla^2_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)$ *a.s.*$-\mathbb{P}$. □

Proof of Lemma 2: From the differentiability condition, the given function is a Lipschitz continuous function, so that $|h(\gamma,\theta)-h(\gamma,\theta')|\le m(\gamma)\|\theta-\theta'\|$, where for each $\gamma\in\Gamma$, we let $m(\gamma):=\sup_{j\in\{1,2,\ldots,d\}}\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)h(\gamma,\theta)|$. Therefore,

$$\frac{1}{\|\theta-\theta'\|}\left|\int h(\gamma,\theta')d\mathbb{Q}(\gamma)-\int h(\gamma,\theta)d\mathbb{Q}(\gamma)\right|\le\int m(\gamma)d\mathbb{Q}(\gamma)<\infty.$$

This further implies that

$$\lim_{\theta'\to\theta}\frac{1}{\|\theta-\theta'\|}\left[\int h(\gamma,\theta')d\mathbb{Q}(\gamma)-\int h(\gamma,\theta)d\mathbb{Q}(\gamma)\right]=\int\lim_{\theta'\to\theta}\frac{1}{\|\theta-\theta'\|}\left[h(\gamma,\theta')-h(\gamma,\theta)\right]d\mathbb{Q}(\gamma)$$

by LDCT. The left and right sides of this equality are respectively identical to the left and right sides of (7). This completes the proof. ■

Proof of Lemma 3: (*i*) The left side is expanded as

$$\nabla_\theta\int\int\{g(\gamma)-\rho(\gamma,\theta,x)\}^2d\mathbb{P}(g(\gamma),x)d\mathbb{Q}(\gamma)=\nabla_\theta\int\int\{-2\mu(\gamma,x)\rho(\gamma,\theta,x)+\rho^2(\gamma,\theta,x)\}d\mathbb{P}(x)d\mathbb{Q}(\gamma)$$

using the fact that $\mu(\gamma,x)=\int g(\gamma)d\mathbb{P}(g(\gamma)|x)$. Given this, $\mu_i(\,\cdot\,)\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$ and for each $j\in\{1,2,\ldots,d\}$, $\sup_\theta|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$ because $|g_i(\,\cdot\,)|\le m_i$ and $\sup_{(\gamma,\theta)}|(\partial/\partial\theta_j)\rho_i(\gamma,\theta)|\le m_i$ by Assumptions 3, so that $\sup_{\theta\in\Theta}|\mu_i(\,\cdot\,)\ (\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ by Cauchy-Schwarz's inequality. Therefore, applying Lemma 2 yields that

$$\nabla_\theta\int\int\mu(\gamma,x)\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)=\int\int\mu(\gamma,x)\nabla_\theta\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma). \tag{8}$$

Furthermore, $\sup_{(\gamma,\theta)}|\rho_i(\gamma,\theta)|\le m_i$ by Assumption 3, so that $\sup_\theta|\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$, implying that $\sup_{\theta\in\Theta}|\rho_i(\,\cdot\,,\theta)(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ by Cauchy-Schwarz's inequality. Applying Lemma 2 once again entails that

$$\nabla_\theta\int\int\rho^2(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)=2\int\int\rho(\gamma,\theta,x)\nabla_\theta\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma). \tag{9}$$

Therefore, the desired result follows by combining (8) and (9).

(*ii*) Note that $\mu_i(\,\cdot\,)\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$, $\sup_{\theta\in\Theta}|\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$ as shown in (*i*). Furthermore, for each $j,j'\in\{1,2,\ldots,d\}$, $\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$ and $\sup_{\theta\in\Theta}|(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$ by Assumption 3. This implies that

$$\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)(\partial/\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})\ \text{ a.s.}-\mathbb{P},$$

$$\sup_{\theta\in\Theta}|\mu(\,\cdot\,)(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})\ \text{ a.s.}-\mathbb{P},\quad\text{and}\quad\sup_{\theta\in\Theta}|\rho_i(\,\cdot\,,\theta)(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})\ \text{ a.s.}-\mathbb{P}$$

by Cauchy-Schwarz's inequality. Applying Lemma 2 leads to the desired result as for (*i*).

(*iii*) The left side is expanded as follows:

$$\nabla_\theta\frac{1}{n}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\theta)\}^2d\mathbb{Q}(\gamma)=\nabla_\theta\frac{1}{n}\sum_{i=1}^n\int\{-2g_i(\gamma)\rho_i(\gamma,\theta)+\rho_i^2(\gamma,\theta)\}d\mathbb{Q}(\gamma).$$

Furthermore, for each j, Cauchy-Schwarz's inequality leads to $\sup_{\theta\in\Theta}|g_i(\,\cdot\,)(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ because $\sup_\theta|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ and $\sup_\gamma|g_i(\gamma)|\le m_i\in L^2(\mathbb{P})$ as shown in (*i* and *ii*) using Assumption 3. Hence, applying Lemma 2 leads to that

$$\nabla_\theta\frac{1}{n}\sum_{i=1}^n\int g_i(\gamma)\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)=\frac{1}{n}\sum_{i=1}^n\int g_i(\gamma)\nabla_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)\tag{10}$$

a.s.$-\mathbb{P}$. Finally, combining (9) and (10) yields the desired result.

(*iv*) Given the left side, for each $j,j'\in\{1,2,\ldots,d\}$, $\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)(\partial/\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ by Cauchy-Schwarz's inequality and $\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})$ as shown in (*ii*). Furthermore, note that for each $j,j'\in\{1,2,\ldots,d\}$, $\sup_{\theta\in\Theta}|\{g_i(\,\cdot\,)-\rho_i(\,\cdot\,,\theta)\}(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ by applying Cauchy-Schwarz's inequality because $g_i(\,\cdot\,)\in L^2(\mathbb{Q})$ a.s.$-\mathbb{P}$,

$$\sup_{\theta}|\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})\ \text{ a.s. }-\mathbb{P},\quad\text{and}\quad\sup_{\theta\in\Theta}|(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,,\theta)|\in L^2(\mathbb{Q})\ \text{ a.s.}-\mathbb{P}$$

as shown in (*i*, *ii*, and *iii*). Therefore,

$$\begin{aligned}&\nabla_\theta\frac{1}{n}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\theta)\}\nabla_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)\\&\qquad=\frac{1}{n}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\theta)\}\nabla_\theta^2\rho_i(\gamma,\theta)-\nabla_\theta\rho_i(\gamma,\theta)\nabla'_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma)\quad\text{a.s.}-\mathbb{P}\end{aligned}$$

by applying Lemma 2. This completes the proof. ■

Proof of Lemma 4: We first focus to the internal integral with respect to $\mathbb{P}$. Note that

$$\begin{aligned}&\int\varepsilon(\gamma,\theta_*)\varepsilon(\widetilde{\gamma},\theta_*)d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x)=\int\{g(\gamma)-\rho(\gamma,\theta_*,x)\}\{g(\widetilde{\gamma})-\rho(\widetilde{\gamma},\theta_*,x)\}d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x)\\&\quad=\int\{g(\gamma)-\mu(\gamma,x)\}\{g(\widetilde{\gamma})-\mu(\widetilde{\gamma},x)\}d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x)+\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\{\mu(\widetilde{\gamma},x)-\rho(\widetilde{\gamma},\theta_*,x)\}\\&\quad=\kappa(\gamma,\widetilde{\gamma}|x)+\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\{\mu(\widetilde{\gamma},x)-\rho(\widetilde{\gamma},\theta_*,x)\}\end{aligned}$$

using that $\int \{g(\gamma) - \mu(\gamma, x)\}\{\mu(\widetilde{\gamma}, x) - \rho(\widetilde{\gamma}, \theta_*, x)\} d\mathbb{P}(g(\gamma), g(\widetilde{\gamma})|x) = 0$. Therefore, we can obtain that

$$\widetilde{B} = \int\int\int \nabla_\theta \rho(\gamma, \theta_*, x)\kappa(\gamma, \widetilde{\gamma}|x)\nabla'_\theta \rho(\widetilde{\gamma}, \theta_*, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma})$$
$$+ \int \nabla_\theta \rho(\gamma, \theta_*, x)\{\mu(\gamma, x) - \rho(\gamma, \theta_*, x)\} d\mathbb{Q}(\gamma) \int \{\mu(\widetilde{\gamma}, x) - \rho(\widetilde{\gamma}, \theta_*, x)\}\nabla'_\theta \rho(\widetilde{\gamma}, \theta_*, x) d\mathbb{Q}(\widetilde{\gamma}).$$

Now note that $B := \int\int\int \nabla_\theta \rho(\gamma, \theta_*, x)\kappa(\gamma, \widetilde{\gamma}|x)\nabla_\theta \rho(\widetilde{\gamma}, \theta_*, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma) d\mathbb{Q}(\widetilde{\gamma})$ and $\int \nabla_\theta \rho(\gamma, \theta_*, x)\{\mu(\gamma, x) - \rho(\gamma, \theta_*, x)\} d\mathbb{Q}(\gamma) = 0$ by the definition of θ_* and Lemma 3(*i*). This completes the proof. ■

Proof of Lemma 5: (*i* and *ii*) Assumptions 6 and 7 imply Assumptions 1 and 3. Therefore, the proofs of Lemma 3(*i* and *ii*) are sufficient for the proofs of Lemma 5(*i* and *ii*).

(*iii*) Note that $\nabla_\theta \sum_{i=1}^n \int \{\widehat{g}_i(\gamma) - \rho_i(\gamma, \theta)\}^2 d\mathbb{P} d\mathbb{Q}(\gamma) = \nabla_\theta \sum_{i=1}^n \int \{-2\widehat{g}_i(\gamma)\rho_i(\gamma, \theta) + \rho^2(\gamma, \theta)\} d\mathbb{Q}(\gamma)$ a.s.$-\,\mathbb{P}$. Furthermore, for each j, $\sup_{\theta\in\Theta} |\widehat{g}_i(\,\cdot\,) \cdot (\partial/\partial\theta_j)\rho_i(\,\cdot\,, \theta)| \in L^1(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ by Cauchy-Schwarz's inequality,

$$\sup_\theta |(\partial/\partial\theta_j)\rho_i(\,\cdot\,, \theta)| \in L^2(\mathbb{Q})$$

as proved in the proof of Lemma 3, and $\widetilde{g}_i \in L^2(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ by Assumption 7(*i*). Hence, applying Lemma 2 leads to that

$$\nabla_\theta \frac{1}{n}\sum_{i=1}^n \int \widehat{g}_i(\gamma)\rho_i(\gamma, \theta) d\mathbb{Q}(\gamma) = \frac{1}{n}\sum_{i=1}^n \int \widehat{g}_i(\gamma)\nabla_\theta \rho_i(\gamma, \theta) d\mathbb{Q}(\gamma) \tag{11}$$

a.s.$-\,\mathbb{P}$. Finally, combining (9) and (11) yields the desired result.

(*iv*) Given the left side, for each j and j', $\sup_{\theta\in\Theta} |(\partial/\partial\theta_j)\rho_i(\,\cdot\,, \theta) \cdot (\partial/\partial\theta_{j'})\rho_i(\,\cdot\,, \theta)| \in L^1(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ as proved in the proof of Lemma 3. Furthermore, for each j and j', $\sup_{\theta\in\Theta} |\{\widehat{g}_i(\,\cdot\,) - \rho_i(\,\cdot\,, \theta)\}(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,, \theta)| \in L^1(\mathbb{Q})$ a.s.$-\,\mathbb{P}$, because $\widetilde{g}_i \in L^2(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ by Assumption 7(*i*), and $\sup_{\theta\in\Theta} |\rho_i(\,\cdot\,, \theta)| \in L^2(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ and $\sup_{\theta\in\Theta} |(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\,\cdot\,, \theta)| \in L^2(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ as shown in the proof of Lemma 3. Therefore,

$$\nabla_\theta \frac{1}{n}\sum_{i=1}^n \int \{\widehat{g}_i(\gamma) - \rho_i(\gamma, \theta)\}\nabla_\theta \rho_i(\gamma, \theta) d\mathbb{Q}(\gamma)$$
$$= \frac{1}{n}\sum_{i=1}^n \int \{\widehat{g}_i(\gamma) - \rho_i(\gamma, \theta)\}\nabla^2_\theta \rho_i(\gamma, \theta) - \nabla_\theta \rho_i(\gamma, \theta)\nabla'_\theta \rho_i(\gamma, \theta) d\mathbb{Q}(\gamma)$$

a.s.$-\,\mathbb{P}$ by applying Lemma 2. This completes the proof. ■

B Proofs

Proof of Theorem 1: Note that

$$\int\int\{g(\gamma)-\rho(\gamma,\theta,x)\}^2 d\mathbb{P}(g(\gamma),x)d\mathbb{Q}(\gamma)=\int\int\{g(\gamma)-\mu(\gamma,x)\}^2 d\mathbb{P}(g(\gamma),x)d\mathbb{Q}(\gamma)$$
$$-2\int\int\{g(\gamma)-\mu(\gamma,x)\}\{\rho(\gamma,\theta,x)-\mu(\gamma,x)\}d\mathbb{P}(g(\gamma),x)d\mathbb{Q}(\gamma)+\int\int\{\rho(\gamma,\theta,x)-\mu(\gamma,x)\}^2 d\mathbb{P}(x)d\mathbb{Q}(\gamma).$$

Note that $\int\int\{g(\gamma)-\mu(\gamma,x)\}^2 d\mathbb{P}d\mathbb{Q}(\gamma)=\int \mathrm{var}_{\mathbb{P}}[g_i(\gamma)|x]d\mathbb{P}(x)d\mathbb{Q}$, and $\int\int\{g(\gamma)-\mu(\gamma,x)\}d\mathbb{P}(g(\gamma)|x)\{\mu(\gamma,x)-\rho(\gamma,\theta,x)\}d\mathbb{P}(x)d\mathbb{Q}(\gamma)=0$ because $\int\{g(\gamma)-\mu(\gamma,x)\}d\mathbb{P}(g(\gamma)|x)=0$. The desired result follows from these. ■

Proof of Theorem 2: (*i*) The desired result follows by applying the SULLN and LDCT. Specifically, from the definitions of $q_n(\,\cdot\,)$ and $q(\,\cdot\,)$, for each θ,

$$q_n(\theta)=\frac{1}{n}\sum_{i=1}^n\int g_i^2(\gamma)d\mathbb{Q}(\gamma)-\frac{2}{n}\sum_{i=1}^n\int\left(g_i(\gamma)\rho_i(\gamma,\theta)\right)d\mathbb{Q}+\frac{1}{n}\sum_{i=1}^n\int\rho_i^2(\gamma,\theta)d\mathbb{Q}\quad\text{and}\tag{12}$$

$$q(\theta)=\int E_{\mathbb{P}}[g_i^2(\gamma)]d\mathbb{Q}(\gamma)-2\int\mathbb{E}_{\mathbb{P}}[\mu_i(\gamma)\rho_i(\gamma,\theta)]d\mathbb{Q}+\int\mathbb{E}_{\mathbb{P}}[\rho_i^2(\gamma,\theta)]d\mathbb{Q}.\tag{13}$$

Here, we use the fact that $\mathbb{E}_{\mathbb{P}}[g_i(\gamma)\rho_i(\gamma.\theta)]=\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[g_i(\gamma)|x]\rho_i(\gamma.\theta)]=\mathbb{E}_{\mathbb{P}}[\mu_i(\gamma)\rho_i(\gamma.\theta)]$ in deriving (13), and we can interchange the integral and sample average operators in computing (12) by applying the LDCT as shown in the proof of Lemma 3.

We now examine the limit of each element in the right sides of (12) and (13). First, Assumption 3(*i*) implies that

$$\int\left|\frac{1}{n}\sum_{i=1}^n g_i^2(\gamma)-E_{\mathbb{P}}[g_i^2(\gamma)]\right|d\mathbb{Q}(\gamma)\le\frac{1}{n}\sum_{i=1}^n m_i^2+E_{\mathbb{P}}[m_i^2]<\infty\text{ a.s.}-\mathbb{P},$$

so that we can apply LDCT, which implies that

$$\int\left|\frac{1}{n}\sum_{i=1}^n g_i^2(\gamma)-E_{\mathbb{P}}[g_i^2(\gamma)]\right|d\mathbb{Q}(\gamma)\to 0\text{ a.s.}-\mathbb{P}$$

by LDCT.

Next, Assumptions 3(*i* and *ii*) imply that $\sup_{\theta\in\Theta}|\rho_i(\,\cdot\,,\theta)g_i(\,\cdot\,)|\le m_i^2\in L^1(\mathbb{P})$, so that

$$\sup_{\theta\in\Theta}\int\left|\left(\frac{1}{n}\sum_{i=1}^n g_i(\gamma)\rho_i(\gamma,\theta)-\mathbb{E}[\mu_i(\gamma)\rho_i(\gamma,\theta)]\right)\right|d\mathbb{Q}(\gamma)$$
$$\le\int\sup_{\theta\in\Theta}\left|\left(\frac{1}{n}\sum_{i=1}^n g_i(\gamma)\rho_i(\gamma,\theta)-\mathbb{E}[\mu_i(\gamma)\rho_i(\gamma,\theta)]\right)\right|d\mathbb{Q}(\gamma)\to 0\tag{14}$$

a.s.$-\,\mathbb{P}$ by the LDCT.

Third, from the fact that $\sup_{\theta\in\Theta}|\rho_i(\,\cdot\,,\theta)| \in L^2(\mathbb{Q})$ a.s.$-\,\mathbb{P}$ as shown in the proof of Lemma 3,

$$\sup_{\theta\in\Theta}\int\left|\left(\frac{1}{n}\sum_{i=1}^n \rho_i^2(\gamma,\theta)-\mathbb{E}[\rho_i^2(\gamma,\theta)]\right)\right|d\mathbb{Q}(\gamma)\le\int\sup_{\theta\in\Theta}\left|\left(\frac{1}{n}\sum_{i=1}^n \rho_i^2(\gamma,\theta)-\mathbb{E}[\rho_i^2(\gamma,\theta)]\right)\right|d\mathbb{Q}(\gamma)\to 0 \tag{15}$$

a.s.$-\,\mathbb{P}$.

Finally, we combine the above three facts. That is,

$$\begin{aligned}\sup_{\theta\in\Theta}|q_n(\theta)-q(\theta)| \le \int\left|\frac{1}{n}\sum_{i=1}^n g_i^2(\gamma)-E_{\mathbb{P}}[g_i^2(\gamma)]\right|d\mathbb{Q}(\gamma)&\\ +2\sup_{\theta\in\Theta}\left|\int\frac{1}{n}\sum_{i=1}^n g_i(\gamma)\rho_i(\gamma,\theta)-\mathbb{E}_{\mathbb{P}}[\mu_i(\gamma)\rho_i(\gamma,\theta)]d\mathbb{Q}(\gamma)\right|&\\ +\sup_{\theta\in\Theta}\int\left|\left(\frac{1}{n}\sum_{i=1}^n \rho_i^2(\gamma,\theta)-\mathbb{E}[\rho_i^2(\gamma,\theta)]\right)\right|d\mathbb{Q}(\gamma)\to 0 \text{ a.s. } -\mathbb{P}&\end{aligned}$$

as desired.

(*ii*) This follows from the definition of $\widehat{\theta}_n$ and Theorem 2(*i*), given the fact that for each $\gamma\in\Gamma$, $\rho_i(\gamma,\cdot)$ is in $\mathcal{C}^{(2)}(\Theta)$ a.s.$-\,\mathbb{P}$. ■

Proof of Theorem 3: We first note that $n^{-1}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\widehat{\theta}_n)\}\nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma)=0$ by Lemma 3(*iii*) and the definition of $\widehat{\theta}_n$. We apply the mean-value theorem to the element in the integral by Lemma 3(*iv*), so that for some $\bar{\theta}_n$ between θ_* and $\widehat{\theta}_n$, it follows that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\widehat{\theta}_n)\}\nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma)=\frac{1}{n}\sum_{i=1}^n\int\{g_i(\gamma)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)&\\ +\frac{1}{n}\sum_{i=1}^n\int\{-\nabla_\theta\rho_i(\gamma,\bar{\theta}_n)\nabla_\theta'\rho_i(\gamma,\bar{\theta}_n)+[g_i(\gamma)-\rho_i(\gamma,\bar{\theta}_n)]\nabla_\theta^2\rho_i(\gamma,\bar{\theta}_n)\}d\mathbb{Q}(\widehat{\theta}_n-\theta_*),&\end{aligned}$$

so that

$$A_n\sqrt{n}(\widehat{\theta}_n-\theta_*)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int[g_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma), \tag{16}$$

where

$$A_n:=\left\{\frac{1}{n}\sum_{i=1}^n\int\left\{\nabla_\theta\rho_i(\gamma,\bar{\theta}_n)\nabla_\theta'\rho_i(\gamma,\bar{\theta}_n)-[g_i(\gamma)-\rho_i(\gamma,\bar{\theta}_n)]\nabla_\theta^2\rho_i(\gamma,\bar{\theta}_n)\right\}d\mathbb{Q}(\gamma)\right\}.$$

Here, if we apply the LDCT, it is possible to interchange the integral and sample average operators:

$$A_n=\frac{1}{n}\sum_{i=1}^n\left\{\int\left\{\nabla_\theta\rho_i(\gamma,\bar{\theta}_n)\nabla_\theta'\rho_i(\gamma,\bar{\theta}_n)\right\}d\mathbb{Q}(\gamma)-\int\left\{[g_i(\gamma)-\rho_i(\gamma,\bar{\theta}_n)]\nabla_\theta^2\rho_i(\gamma,\bar{\theta}_n)\right\}d\mathbb{Q}(\gamma)\right\}.$$

We now examine each element in (16). First, we already saw in the proof of Lemma 3(*iv*) that for each $j,j'\in$

$\{1,2,\ldots,d\}$, $|(\partial/\partial\theta_j)\rho_i(\cdot,\theta)\cdot(\partial/\partial\theta_{j'})\rho_i(\cdot,\theta)| \in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$, so that Theorem 2 yields that

$$\int \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta \rho_i(\gamma,\bar{\theta}_n)\nabla'_\theta \rho_i(\gamma,\bar{\theta}_n) d\mathbb{Q}(\gamma) \to \int\int \nabla_\theta \rho(\gamma,\theta_*,x)\nabla'_\theta \rho(\gamma,\theta_*,x) d\mathbb{P}(x)d\mathbb{Q}(\gamma) \text{ a.s.} - \mathbb{P}.$$

Second, we also note that for each $i, j \in \{1,2,\ldots,d\}$, $\sup_{\theta\in\Theta} |\rho_i(\cdot,\theta)(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\cdot,\theta)| \in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ as shown in the proof of Lemma 3. Therefore, Theorem 2 yields that

$$\int \frac{1}{n}\sum_{i=1}^{n} \rho_i(\gamma,\bar{\theta}_n)\nabla^2_\theta \rho_i(\gamma,\bar{\theta}_n) d\mathbb{Q}(\gamma) \to \int\int \rho(\gamma,\theta_*,x)\nabla^2_\theta \rho(\gamma,\theta_*,x) d\mathbb{P}(x)d\mathbb{Q}(\gamma) \quad \text{a.s.} - \mathbb{P}.$$

Third, for each $j, j' \in \{1,2,\ldots,d\}$, $n^{-1}\sum_{i=1}^{n}\sup_{\theta\in\Theta} |g_i(\cdot)\cdot(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\cdot,\theta)| \in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$ as shown in the proof of Lemma 3. Therefore,

$$\int \frac{1}{n}\sum_{i=1}^{n} g_i(\gamma)\nabla^2_\theta \rho_i(\gamma,\bar{\theta}_n) d\mathbb{Q}(\gamma) \to \int\int \mu(\gamma,x)\nabla^2_\theta \rho(\gamma,\theta_*,x) d\mathbb{P}(x)d\mathbb{Q}(\gamma) \quad \text{a.s.} - \mathbb{P}.$$

From these three facts, $A_n \to A$ a.s.$-\mathbb{P}$.

Fourth, we consider the right side of (16). Note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int [g_i(\gamma) - \rho_i(\gamma,\theta_*)]\nabla_\theta \rho_i(\gamma,\theta_*) d\mathbb{Q}(\gamma) \overset{\text{A}}{\sim} N(0,B) \tag{17}$$

because $\int\int \{g(\gamma)-\rho(\gamma,\theta_*,x)\}\nabla_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(g(\gamma),x)d\mathbb{Q}(\gamma) = \int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) = \nabla_\theta q(\theta_*) = 0$ by Lemma 3(*i*) and for each j and j', $\int\int(\partial/\partial\theta_j)\rho(\gamma,\theta_*,x)\cdot\kappa(\gamma,\widetilde{\gamma}|x)\cdot(\partial/\partial\theta_{j'})\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) < \infty$ by Assumption 5. From this, the right side of (16) is asymptotically distributed as $N(0,B)$. Therefore, $\sqrt{n}(\widehat{\theta}_n - \theta_*) \overset{\text{A}}{\sim} N(0, A^{-1}BA^{-1})$. This completes the proof. ■

Proof of Theorem 4: (*i*) The desired result can be obtained by following the proof of Theorem 2. Specifically, we can apply the SULLN and LDCT. From the definitions of $\widehat{q}_n(\cdot)$ and $q(\cdot)$, for each θ,

$$\widehat{q}_n(\theta) = \int \frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i^2(\gamma)d\mathbb{Q}(\gamma) - 2\int\frac{1}{n}\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)\rho_i(\gamma,\theta)\}\,d\mathbb{Q} + \int\frac{1}{n}\sum_{i=1}^{n}\rho_i^2(\gamma,\theta)d\mathbb{Q}. \tag{18}$$

Here, we interchanged the integral and sample average operators by applying the LDCT. We compare this with each element in the right side of (13). First, Assumption 7(*i*) implies that

$$\int\left|\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i^2(\gamma) - E_{\mathbb{P}}[g_i^2(\gamma)]\right|d\mathbb{Q}(\gamma) \le \frac{1}{n}\sum_{i=1}^{n} m_i^2 + E_{\mathbb{P}}[m_i^2] < \infty \text{ a.s.} - \mathbb{P},$$

so that we can apply LDCT, which implies that

$$\int \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i^2(\gamma) - E_{\mathbb{P}}[g_i^2(\gamma)] \right| d\mathbb{Q}(\gamma) \to 0 \text{ a.s.} - \mathbb{P}$$

because (a)

$$\int \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i^2(\gamma) - E_{\mathbb{P}}[g_i^2(\gamma)] \right| d\mathbb{Q}(\gamma)$$
$$\leq \int \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i^2(\gamma) - \frac{1}{n} \sum_{i=1}^{n} g_i^2(\gamma) \right| d\mathbb{Q}(\gamma) + \int \left| \frac{1}{n} \sum_{i=1}^{n} g_i^2(\gamma) - E_{\mathbb{P}}[g_i^2(\gamma)] \right| d\mathbb{Q}(\gamma);$$

(b) the proof of Theorem 2(i) implies that $\int \left| \frac{1}{n} \sum_{i=1}^{n} g_i^2(\gamma) - E_{\mathbb{P}}[g_i^2(\gamma)] \right| d\mathbb{Q}(\gamma) \to 0$ a.s.$-\mathbb{P}$; and (c) finally, by applying the mean-value theorem, for some $\bar{\xi}_{n,\gamma}$ between ξ_* and $\widehat{\xi}_n$,

$$\int \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i^2(\gamma) - \frac{1}{n} \sum_{i=1}^{n} g_i^2(\gamma) \right| d\mathbb{Q}(\gamma) = 2 \int \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i(\gamma) \nabla'_{\xi} \widetilde{g}_i(\gamma, \bar{\xi}_{n,\gamma}) \right| d\mathbb{Q} \cdot \left| \widehat{\xi}_n - \xi_* \right|.$$

Given this, for each $j = 1, 2, \ldots, s$, Assumption 7(iii) implies that $\int \left| \frac{1}{n} \sum_{i=1}^{n} \widehat{g}_i(\gamma) \cdot (\partial / \partial \xi_j) g_i(\gamma, \bar{\xi}_{n,\gamma}) \right| d\mathbb{Q} \leq n^{-1} \sum_{i=1}^{n} m_i^2$, and $\widehat{\xi}_n \to \xi_*$ a.s.$-\mathbb{P}$ by Assumption 8(i). Therefore, the desired result follows from the facts (a), (b), and (c).

Next, we compare the second elements in (18) and (13). First,

$$\sup_{\theta \in \Theta} \left| \int \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{g}_i(\gamma) \rho_i(\gamma, \theta) - \mathbb{E}_{\mathbb{P}}[\mu_i(\gamma) \rho_i(\gamma, \theta)] \right) d\mathbb{Q} \right|$$
$$= \int \sup_{\theta \in \Theta} \left| \left(\frac{1}{n} \sum_{i=1}^{n} g_i(\gamma) \rho_i(\gamma, \theta) - \mathbb{E}_{\mathbb{P}}[g_i(\gamma) \rho_i(\gamma, \theta)] \right) \right| d\mathbb{Q} + o_{\mathbb{P}}(1)$$

because applying the mean-value theorem implies that for some $\bar{\xi}_{\gamma,n}$ between ξ_* and $\widehat{\xi}_n$,

$$\sup_{\theta \in \Theta} \left| \int \left(\frac{1}{n} \sum_{i=1}^{n} \{ \widehat{g}_i(\gamma) - g_i(\gamma) \} \right) \rho_i(\gamma, \theta) d\mathbb{Q} \right| = \sup_{\theta \in \Theta} \left| \int \frac{1}{n} \sum_{i=1}^{n} \nabla'_{\xi} \widetilde{g}_i(\gamma, \bar{\xi}_{\gamma,n}) \cdot \rho_i(\gamma, \theta) d\mathbb{Q} \cdot (\widehat{\xi}_n - \xi_*) \right|,$$

so that for each $j = 1, 2, \ldots, s$,

$$\sup_{\theta \in \Theta} \left| \int \frac{1}{n} \sum_{i=1}^{n} (\partial / \partial \xi_j) \widetilde{g}_i(\gamma, \xi_{\gamma,n}) \cdot \rho_i(\gamma, \theta) d\mathbb{Q} \cdot (\widehat{\xi}_{j,n} - \xi_{j*}) \right|$$
$$\leq \left(\frac{1}{n} \sum_{i=1}^{n} m_i^2 \right)^{1/2} \sup_{\theta \in \Theta} \left| \int \frac{1}{n} \sum_{i=1}^{n} \rho_i^2(\gamma, \theta) d\mathbb{Q} \right| \cdot \left| \widehat{\xi}_{j,n} - \xi_{j*} \right| \to 0$$

a.s.$-\ \mathbb{P}$ by Assumptions 7(*ii* and *iii*) and 8(*i*). Now, (15) implies that

$$\sup_{\theta\in\Theta}\left|\int\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{g}_i(\gamma)\rho_i(\gamma,\theta)-\mathbb{E}_{\mathbb{P}}[\mu_i(\gamma)\rho_i(\gamma,\theta)]\right)d\mathbb{Q}\right|\to 0\ a.s.\ -\mathbb{P}.$$

Finally, the third component in the right side of (18) is identical to the third element in the right side of (13). Therefore, the desired result follows from these three facts.

(*ii*) This follows from the definition of $\widetilde{\theta}_n$ and Theorem 4(*i*), given the fact that for each $\gamma\in\Gamma$, $\rho_i(\gamma,\cdot)$ is in $\mathcal{C}^{(2)}(\Theta)$ a.s.$-\ \mathbb{P}$. ■

Proof of Theorem 5: We first note that $n^{-1}\sum_{i=1}^{n}\int\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\widetilde{\theta}_n)\}\nabla_\theta\rho_i(\gamma,\widetilde{\theta}_n)d\mathbb{Q}(\gamma)\equiv 0$ by Lemma 5(*iii*) and the definition of $\widetilde{\theta}_n$. We apply the mean-value theorem to the element in the integral by Lemma 3(*iv*), so that for some $\bar{\theta}_n$ between θ_* and $\widetilde{\theta}_n$, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\int\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\widehat{\theta}_n)\}\nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\int\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\int\{-\nabla_\theta\rho_i(\gamma,\bar{\theta}_n)\nabla'_\theta\rho_i(\gamma,\bar{\theta}_n)+[g_i(\gamma)-\rho_i(\gamma,\bar{\theta}_n)]\nabla^2_\theta\rho_i(\gamma,\bar{\theta}_n)\}d\mathbb{Q}(\widetilde{\theta}_n-\theta_*),$$

so that

$$\widehat{A}_n\sqrt{n}(\widetilde{\theta}_n-\theta_*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\int[\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\right\},\tag{19}$$

where

$$\widehat{A}_n:=\left\{\int\{\nabla_\theta\rho_i(\gamma,\bar{\theta}_n)\nabla'_\theta\rho_i(\gamma,\bar{\theta}_n)-\frac{1}{n}\sum_{i=1}^{n}[\widehat{g}_i(\gamma)-\rho_i(\gamma,\bar{\theta}_n)]\nabla^2_\theta\rho_i(\gamma,\bar{\theta}_n)\}d\mathbb{Q}(\gamma)\right\}.$$

We examine each element in (19). First, for each $j,j'\in\{1,2,\ldots,d\}$, $|(\partial/\partial\theta_j)\rho_i(\cdot,\theta)\cdot(\partial/\partial\theta_{j'})\rho_i(\cdot,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\ \mathbb{P}$ as shown in the proof of Lemma 3, implying that

$$\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\bar{\theta}_n)\nabla'_\theta\rho_i(\gamma,\bar{\theta}_n)d\mathbb{Q}(\gamma)\to\int\nabla_\theta\rho(\gamma,\theta_*,x)\nabla'_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)\ \ a.s.-\ \mathbb{P}.$$

by Theorem 4(*ii*) and LDCT.

Second, for each $i,j\in\{1,2,\ldots,d\}$, $\sup_{\theta\in\Theta}|\rho_i(\cdot,\theta)(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\cdot,\theta)|\in L^1(\mathbb{Q})$ a.s.$-\ \mathbb{P}$ by Cauchy-Schwarz's inequality, Assumption 7(*i*) and 4(*i*). Therefore,

$$\int\frac{1}{n}\sum_{i=1}^{n}\rho_i(\gamma,\bar{\theta}_n)\nabla^2_\theta\rho_i(\gamma,\bar{\theta}_n)d\mathbb{Q}(\gamma)\to\int\int\rho(\gamma,\theta_*,x)\nabla^2_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)\ a.s.-\ \mathbb{P}$$

by LDCT and Theorem 4(*ii*).

Third, for each $j,j'\in\{1,2,\ldots,d\}$, $n^{-1}\sum_{i=1}^{n}\sup_{\xi\in\Xi}\sup_{\theta\in\Theta}|\widetilde{g}_i(\cdot,\cdot,\xi)\cdot(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\cdot,\theta)|\in L^1\ (\mathbb{Q})$

a.s.$-\mathbb{P}$ as shown in the proof of Lemma 3. Thus, Assumption 3 and LDCT and imply that

$$\frac{1}{n}\int\sum_{i=1}^{n}G_i y(\gamma)\nabla_\theta^2\rho_i(\gamma,\bar{\theta}_n)d\mathbb{Q}(\gamma)\to\int\int\mu(\gamma,x)\nabla_\theta^2\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)\ \ a.s.-\mathbb{P}.$$

From these three facts, we now obtain that $\widehat{A}_n\to A$ a.s.$-\mathbb{P}$.

Fourth, we examine the right side of (19). Applying the mean-value theorem, we obtain (3): for some $\bar{\xi}_{n,\gamma}$ between $\widehat{\xi}_n$ and ξ_*,

$$\begin{aligned}\frac{1}{\sqrt{n}}\int\sum_{i=1}^{n}&\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\\&=\frac{1}{\sqrt{n}}\int\sum_{i=1}^{n}\{g_i(\gamma,\xi_*)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\\&\qquad+\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\theta_*)\cdot\nabla_\xi' g_i(\gamma,\bar{\xi}_{n,\gamma})d\mathbb{Q}\cdot\sqrt{n}(\widehat{\xi}_n-\xi_*).\end{aligned}\tag{20}$$

Given this, we have already seen in the proof of Theorem 3 that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int[g_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\overset{\mathrm{A}}{\sim}N(0,B).$$

We also note that for $j=1,\ldots,d$ and $j'=1,\ldots,s$, Assumptions 3(*iii*) and 7(*iii*) imply that $\sup_{(\theta,\xi)}|(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)\cdot(\partial/\partial\xi_{j'})\widetilde{g}_i(\,\cdot\,,\xi)|\in L^1(\mathbb{Q})$ a.s.$-\mathbb{P}$, so that applying the LDCT shows that

$$\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\theta_*)\nabla_\xi' g_i(\gamma,\bar{\xi}_{n,\gamma})d\mathbb{Q}(\gamma)\to M:=\int\mathbb{E}_{\mathbb{P}}[\nabla_\theta\rho(\gamma,\theta_*,x_i)\nabla_\xi' g_i(\gamma,\xi_*)]d\mathbb{Q}(\gamma)\ \ a.s.\ -\ \mathbb{P}$$

a.s.$-\mathbb{P}$. In addition, if we combine Assumptions 8(*ii* and *iii*) with 9,

$$\sqrt{n}(\widehat{\xi}_n-\xi_*)=-H^{-1}\sqrt{n}s_{*n}+o_{\mathbb{P}}(1)\overset{\mathrm{A}}{\sim}N(0,H^{-1}JH^{-1\prime}).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\theta_*)\nabla_\xi' g_i(\gamma,\bar{\xi}_{n,\gamma})d\mathbb{Q}\sqrt{n}(\widehat{\xi}_n-\xi_*)\overset{\mathrm{A}}{\sim}N(0,MH^{-1}JH^{-1\prime}M^{-1\prime}).$$

Therefore, if we further apply Cráer-Wold's device to obtain the asymptotic covariance between $n^{-1/2}\sum_{i=1}^n\int[g_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)$ and $n^{-1/2}\sum_{i=1}^n\int\nabla_\theta\rho_i(\gamma,\theta_*)\cdot\nabla_\xi' g_i(\gamma,\bar{\xi}_{n,\gamma})d\mathbb{Q}(\widehat{\xi}_n-\xi_*)$ that is identical to $-MH^{-1}K$ by Assumption 9, we obtain that

$$\frac{1}{\sqrt{n}}\int\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)-\mu_i(\gamma)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\overset{\mathrm{A}}{\sim}N(0,B_*).\tag{21}$$

Finally, A^{-1} exists by Assumption 4(*ii*), so that $\sqrt{n}(\widetilde{\theta}_n-\theta_*)\overset{\mathrm{A}}{\sim}N(0,A^{-1}B_*A^{-1})$ by (19) and (21). This completes the proof. ■

Proof of Theorem 6: We first examine the consistence of $\widehat{A}_n$. Note that for each j and $j' = 1, 2, \ldots, d$, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta}$ $|(\partial/\partial\theta_j)\rho_i(\gamma,\theta)(\partial/\partial\theta_{j'})\rho_i(\gamma,\theta)| \leq m_i \in L^2(\mathbb{P})$ by Assumption 10(*ii*), so that

$$\sup_{(\gamma,\theta)\in\Gamma\times\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}(\nabla_\theta\rho_i(\gamma,\theta)\nabla'_\theta\rho_i(\gamma,\theta) - \mathbb{E}_\mathbb{P}[\nabla_\theta\rho_i(\gamma,\theta)\nabla'_\theta\rho_i(\gamma,\theta)])\right| \to 0 \ \ a.s. - \ \mathbb{P}$$

by the SULLN. Therefore, applying the LDCT, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)\nabla'_\theta\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma) \to \int\int\nabla_\theta\rho(\gamma,\theta_*,x)\nabla'_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) \tag{22}$$

using the fact that $\widehat{\theta}_n \to \theta_*$ a.s.$-\,\mathbb{P}$. We also note that $\sup_{(\gamma,\theta)\in\Gamma\times\Theta}|\varepsilon_i(\gamma,\theta)(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\gamma,\theta)| \leq m_i^2 \in L^1(\mathbb{P})$ by Assumptions 10(*i* and *iii*). Therefore,

$$\sup_{(\gamma,\theta)\in\Gamma\times\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i(\gamma,\theta)\cdot\nabla^2_\theta\rho_i(\gamma,\theta) - \mathbb{E}_\mathbb{P}[\varepsilon_i(\gamma,\theta)\cdot\nabla^2_\theta\rho_i(\gamma,\theta)])\right| \to 0 \ \ a.s. - \ \mathbb{P}$$

by the SULLN. Therefore, the LDCT and the fact that $\widehat{\theta}_n \to \theta_*$ a.s.$-\,\mathbb{P}$ imply that

$$\frac{1}{n}\sum_{i=1}^{n}\int\varepsilon_i(\gamma,\widehat{\theta}_n)\cdot\nabla^2_\theta\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma) \to \int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla^2_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) \ \ a.s. - \ \mathbb{P}. \tag{23}$$

Here, we used the fact that $\int g(\gamma)d\mathbb{P}(g(\gamma)|x) = \mu(\gamma,x)$. Now, (22) and (23) imply that $\widehat{A}_n \to A$.

We next examine the consistence of $\widehat{B}_n$. Note that for each j and $j' = 1, 2, \ldots, d$, $\sup_{(\gamma,\widetilde{\gamma},\theta)\in\Gamma\times\Gamma\Theta}|(\partial/\partial\theta_j)\rho_i(\gamma,$ $\theta)\varepsilon_i(\gamma,\theta)\varepsilon_i(\widetilde{\gamma},\theta)(\partial/\partial\theta_{j'})\rho_i(\widetilde{\gamma},\theta)| \leq m_i^2 \in L^1(\mathbb{P})$ by Assumptions 10(*i* and *ii*), so that

$$\sup_{(\gamma,\widetilde{\gamma},\theta)\in\Gamma\times\Gamma\times\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}(\nabla_\theta\rho_i(\gamma,\theta)\varepsilon_i(\gamma,\theta)\varepsilon_i(\widetilde{\gamma},\theta)\nabla'_\theta\rho_i(\widetilde{\gamma},\theta) - \mathbb{E}_\mathbb{P}[\nabla_\theta\rho_i(\gamma,\theta)\varepsilon_i(\gamma,\theta)\varepsilon_i(\widetilde{\gamma},\theta)\nabla'_\theta\rho_i(\widetilde{\gamma},\theta)])\right| \to 0$$

a.s.$-\,\mathbb{P}$ by the SULLN. Therefore, applying the LDCT, it follows that

$$\begin{aligned}\widehat{B}_n := \frac{1}{n}\sum_{i=1}^{n}\int\int&\nabla_\theta\rho_i(\gamma,\widehat{\theta}_n)\varepsilon_i(\gamma,\widehat{\theta}_n)\varepsilon_i(\widetilde{\gamma},\widehat{\theta}_n)\nabla'_\theta\rho_i(\widetilde{\gamma},\widehat{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\ &\to \int\int\nabla_\theta\rho(\gamma,\theta_*,x)\int\varepsilon(\gamma,\theta_*)\varepsilon(\widetilde{\gamma},\theta_*)d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x)\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \ \ a.s. - \ \mathbb{P}\end{aligned}$$

using the fact that $\widehat{\theta}_n \to \theta_*$ a.s.$-\,\mathbb{P}$. Here,

$$\int\varepsilon(\gamma,\theta_*)\varepsilon(\widetilde{\gamma},\theta_*)d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x) = \int\{g(\gamma)-\rho(\gamma,\theta_*,x)\}\{g(\widetilde{\gamma})-\rho(\widetilde{\gamma},\theta_*,x)\}d\mathbb{P}(|x) =: \kappa(\gamma,\widetilde{\gamma}|x).$$

Therefore, $\widehat{B}_n \to \int\int\nabla_\theta\rho(\gamma,\theta_*,x)\kappa(\gamma,\widetilde{\gamma}|x)\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$ that is the definition of B. This completes the proof. ■

Proof of Theorem 7: We first examine the consistence of $\widetilde{A}_n$. First of all, (22) implies that

$$\frac{1}{n}\sum_{i=1}^{n}\int \nabla_\theta \rho_i(\gamma,\widetilde{\theta}_n)\nabla'_\theta \rho_i(\gamma,\widetilde{\theta}_n)d\mathbb{Q}(\gamma) \to \int\int \nabla_\theta \rho(\gamma,\theta_*,x)\nabla'_\theta \rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma). \tag{24}$$

Next, Assumptions 12(*i* and *iii*) imply that $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}|\varepsilon_i(\gamma,\theta,\xi)\cdot(\partial^2/\partial\theta_j\partial\theta_{j'})\rho_i(\gamma,\theta)| \leq m_i^2 \in L^1(\mathbb{P})$ for each j and $j'=1,2,\ldots,d$. Therefore, the SULLN implies that

$$\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(\gamma,\theta,\xi)\cdot\nabla^2_\theta\rho_i(\gamma,\theta) - \mathbb{E}_\mathbb{P}[\varepsilon_i(\gamma,\theta,\xi)\cdot\nabla^2_\theta\rho_i(\gamma,\theta)]\right| \to 0 \;\; a.s.-\mathbb{P}$$

by the LDCT. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\int \widetilde{\varepsilon}_i(\gamma,\widetilde{\theta}_n)\cdot\nabla^2_\theta\rho_i(\gamma,\widetilde{\theta}_n) \to \int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla^2_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) \;\; a.s.-\mathbb{P} \tag{25}$$

by noting that $\int g(\gamma)d\mathbb{P}(g(\gamma)|x)=\mu(\gamma,x)$. Now, (24) and (25) imply that $\widetilde{A}_n \to A$ a.s.$-\mathbb{P}$.

We next show the consistence of $\widetilde{B}_n$. From the definition of $\widetilde{B}_n$, if we show that (*i*) $\bar{B}_n \to B$ a.s.$-\mathbb{P}$; (*ii*) $\widehat{M}_n \to M$ a.s.$-\mathbb{P}$; and (*iii*) $\widehat{K}_n \to K$ a.s.$-\mathbb{P}$, the consistence of $\widetilde{B}_n$ follows from Assumption 11.

(*i*) Proving that $\bar{B}_n \to B$ a.s.$-\mathbb{P}$ is almost identical to that of $\widehat{B}_n \to B$. Note that for each j and $j'=1,2,\ldots,d$, $\sup_{(\gamma,\widetilde{\gamma},\theta,\xi)\in\Gamma\times\Gamma\Theta}|(\partial/\partial\theta_j)\rho_i(\gamma,\theta)\varepsilon_i(\gamma,\theta,\xi)\varepsilon_i(\widetilde{\gamma},\theta,\xi)(\partial/\partial\theta_{j'})\rho_i(\widetilde{\gamma},\theta)| \leq m_i^2 \in L^1(\mathbb{P})$ by Assumptions 12(*i* and *ii*), so that

$$\sup_{(\gamma,\widetilde{\gamma},\theta,\xi)}\left|\frac{1}{n}\sum_{i=1}^{n}(\nabla_\theta\rho_i(\gamma,\theta)\varepsilon_i(\gamma,\theta,\xi)\varepsilon_i(\widetilde{\gamma},\theta,\xi)\nabla'_\theta\rho_i(\widetilde{\gamma},\theta) - \mathbb{E}_\mathbb{P}[\nabla_\theta\rho_i(\gamma,\theta)\varepsilon_i(\gamma,\theta,\xi)\varepsilon_i(\widetilde{\gamma},\theta,\xi)\nabla'_\theta\rho_i(\widetilde{\gamma},\theta)])\right| \to 0$$

a.s.$-\mathbb{P}$ by the SULLN. Therefore, applying the LDCT, it follows that

$$\begin{aligned}\bar{B}_n &:= \frac{1}{n}\sum_{i=1}^{n}\int\int \nabla_\theta\rho_i(\gamma,\widetilde{\theta}_n)\widetilde{\varepsilon}_{in}(\gamma,\widetilde{\theta}_n)\widetilde{\varepsilon}_{in}(\widetilde{\gamma},\widetilde{\theta}_n)\nabla'_\theta\rho_i(\widetilde{\gamma},\widetilde{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\ &\to \int\int \nabla_\theta\rho(\gamma,\theta_*,x)\int \varepsilon(\gamma,\theta_*,\xi_*)\varepsilon(\widetilde{\gamma},\theta_*,\xi_*)d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x)\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma}) \;\; a.s.-\mathbb{P}\end{aligned}$$

using the fact that $(\widehat{\xi}_n,\widetilde{\theta}_n) \to (\xi_*,\theta_*)$ a.s.$-\mathbb{P}$. In the proof of Theorem 6, we already seen that the right side is identical to B. Therefore, $\bar{B}_n \to B$ a.s.$-\mathbb{P}$.

(*ii*) Now, we show that $\widehat{M}_n \to M$ a.s.$-\mathbb{P}$. Note that Assumptions 12(*ii* and *iv*) imply that for each $j=1,2,\ldots,d$ and $j'=1,2,\ldots,s$, $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}|(\partial/\partial\theta_j)\rho_i(\gamma,\theta)\cdot(\partial/\partial\xi_{j'})\widetilde{G}y_i(\gamma,\xi)| \leq m_i^2 \in L^1(\mathbb{P})$. Therefore,

$$\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}\left|\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta\rho_i(\gamma,\theta)\nabla'_\xi\widetilde{g}_i(\gamma,\xi) - \mathbb{E}_\mathbb{P}[\nabla_\theta\rho_i(\gamma,\theta)\nabla'_\xi\widetilde{g}_i(\gamma,\xi)]\right| \to 0 \;\; a.s.-\mathbb{P} \tag{26}$$

by the SULLN. Therefore, applying the LDCT implies that

$$\widehat{M}_n := \frac{1}{n}\sum_{i=1}^{n}\int \nabla_\theta \rho_i(\gamma,\widetilde{\theta}_n)\nabla'_\xi \widetilde{g}_i(\gamma,\widehat{\xi}_n)d\mathbb{Q}(\gamma) \to \int E_{\mathbb{P}}[\nabla_\theta \rho_i(\gamma,\theta_*)\nabla'_\xi \widetilde{g}_i(\gamma,\xi_*)]d\mathbb{Q}(\gamma) \;\; a.s. - \mathbb{P}$$

using the fact that $(\widehat{\xi}_n, \widetilde{\theta}_n) \to (\xi_*, \theta_*)$ a.s.$-\mathbb{P}$. Note that the right side is M, implying that $\widehat{M}_n \to M$ a.s.$-\mathbb{P}$.

(*iii*) Third, we now show that $\widehat{K}_n \to K$ a.s.$-\mathbb{P}$. Note that $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi} |s_i(\xi)\{\widetilde{g}_i(\gamma,\xi) - \rho_i(\gamma,\theta\}\nabla'_\theta \rho_i(\gamma,\theta)| \leq m_i^2 \in L^2(\mathbb{P})$ by Assumptions 12(*i*, *ii*, and *v*). Therefore, the SULLN implies that

$$\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}\left|\frac{1}{n}\sum_{i=1}^{n} s_i(\xi)\{\widetilde{g}_i(\gamma,\xi)-\rho_i(\gamma,\theta)\}\nabla'_\theta\rho_i(\gamma,\theta) - \mathbb{E}_{\mathbb{P}}[s_i(\xi)\{\widetilde{g}_i(\gamma,\xi)-\rho_i(\gamma,\theta)\}\nabla'_\theta\rho_i(\gamma,\theta)]\right| \to 0$$

a.s.$-\mathbb{P}$. Applying the LDCT implies that

$$\widehat{K}_n := \frac{1}{n}\sum_{i=1}^{n}\int s_i(\widehat{\xi}_n)\{\widetilde{g}_i(\gamma,\widehat{\xi}_n) - \rho_i(\gamma,\widetilde{\theta}_n)\}\nabla'_\theta\rho_i(\gamma,\widetilde{\theta}_n)d\mathbb{Q}$$
$$\to \int E_{\mathbb{P}}[s_i(\xi_*)\{\widetilde{g}_i(\gamma,\xi_*) - \rho_i(\gamma,\theta_*\}\nabla'_\theta\rho_i(\gamma,\theta_*)]d\mathbb{Q}(\gamma) =: K \;\; a.s. - \mathbb{P}$$

using the fact that $(\widehat{\xi}_n, \widetilde{\theta}_n) \to (\xi_*, \theta_*)$ a.s.$-\mathbb{P}$.

Finally, the desired consistence of $\widetilde{B}_n$ for B_* follows from the consequences from (*i*, *ii*, and *iii*) of this proof. ■

Proof of Lemma 1: (*i*) First, Theorem 2(*i*) implies that $\sup_{\theta\in\Theta}|q_n(\theta) - q(\theta)\,| \to 0$ a.s$-\mathbb{P}$, so that the CFLS estimator $\ddot{\theta}_n^\flat$ must be converging to $\theta_\dagger$ a.s.$-\mathbb{P}$, as $\theta_\dagger$ is constrained by the same constraint $h(\theta) = 0$. Second, θ_* is the global minimizer of $q(\,\cdot\,)$ and also satisfies that $h(\theta_*) = 0$ under $\mathbb{H}_o$. Therefore, $\theta_* = \theta_\dagger$, as desired.

(*ii*) Theorem 4(*i*) implies that $\sup_{\theta\in\Theta}|\widehat{q}_n(\theta) - q(\theta)| \to 0$ a.s$-\mathbb{P}$, so that the CTSFLS estimator $\ddot{\theta}_n^\sharp$ must be converging to $\theta_\dagger$ a.s.$-\mathbb{P}$, as $\theta_\dagger$ is constrained by the same constraint $h(\theta) = 0$. The remaining is identical to the proof of Lemma 1(*i*). ■

Proof of Theorem 8: (*i*) By the mean-value theorem, we obtain that for some $\bar{\theta}_n$ between $\widehat{\theta}_n$ and θ_*, $\sqrt{n}\{h(\widehat{\theta}_n) - h(\theta_*)\} = D(\bar{\theta}_n)\sqrt{n}(\widehat{\theta}_n - \theta_*)$, so that Theorems 2 and 3 imply that $\sqrt{n}\{h(\widehat{\theta}_n) - h(\theta_*)\} \overset{A}{\sim} N(0, D_*A^{-1}BA^{-1}D'_*)$. Furthermore, $\widehat{B}_n \to B$ a.s.$-\mathbb{P}$ by Theorem 6, so that $n\{h(\widehat{\theta}_n) - h(\theta_*)\}'\{D_*A^{-1}\widehat{B}_nA^{-1}D'_*\}^{-1}\{h(\widehat{\theta}_n) - h(\theta_*)\} \overset{A}{\sim} \mathcal{X}^2(r,0)$. Therefore, under $\mathbb{H}_0$, $\mathcal{W}_n^\flat \overset{A}{\sim} \mathcal{X}^2(r,0)$. Meanwhile, $nh(\widehat{\theta}_n)'\,\{D_*A^{-1}\widehat{B}_nA^{-1}D'_*\}^{-1}h(\widehat{\theta}_n) = O_{\mathbb{P}}(n)$ but not $o_{\mathbb{P}}(n)$ because $h(\widehat{\theta}_n) \to h(\theta_*) \neq 0$ under $\mathbb{H}_a$, so that the desired result follows.

(*ii*) The proof is almost identical to Theorem 8(*i*): under $\mathbb{H}_o$, $\sqrt{n}\{h(\widetilde{\theta}_n) - h(\theta_*)\} \overset{A}{\sim} N(0, D_*A^{-1}B_*A^{-1}D'_*)$ by the mean-value theorem, Theorems 4 and 5; and $\widetilde{B}_n \to B_*$ a.s.$-\mathbb{P}$ by Theorem 7, so that $n\{h(\widetilde{\theta}_n) - h(\theta_*)\}'\{D_*A^{-1}\widetilde{B}_nA^{-1}D'_*\}^{-1}\{h(\widetilde{\theta}_n) - h(\theta_*)\} \overset{A}{\sim} \mathcal{X}^2(r,0)$ under $\mathbb{H}_o$, implying that $\mathcal{W}_n^\sharp \overset{A}{\sim} \mathcal{X}^2(r,0)$ under $\mathbb{H}_o$. Meanwhile, $nh(\widetilde{\theta}_n)'\{D_*A^{-1}\widetilde{B}_nA^{-1}D'_*\}^{-1}h(\widetilde{\theta}_n) = O_{\mathbb{P}}(n)$ but not $o_{\mathbb{P}}(n)$ because $h(\widetilde{\theta}_n) \to h(\theta_*) \neq 0$ under $\mathbb{H}_a$, as desired. ■

Proof of Theorem 9: (*i*) The CFLS estimator can be obtained by minimizing the Lagrange function: $\mathcal{L}_n(\theta,\lambda) := q_n(\theta) - \lambda' h(\theta)$, whose first-order conditions can be given as $\nabla_\theta q_n(\ddot{\theta}_n^\flat) - \ddot{\lambda}_n^{\flat\prime}\ddot{D}_n^\flat \equiv 0$ and $h(\ddot{\theta}_n^\flat) \equiv 0$, where $(\ddot{\theta}_n^{\flat\prime}, \ddot{\lambda}_n^{\flat\prime})'$

is the solution for the first-order conditions. In addition, for some $\bar{\theta}_n^\flat$ between $\ddot{\theta}_n^\flat$ and θ_*,

$$\nabla_\theta q_n(\ddot{\theta}_n^\flat) = \nabla_\theta q_n(\theta_*) + \nabla_\theta^2 q_n(\bar{\theta}_n^\flat)(\ddot{\theta}_n^\flat - \theta_*) \text{ and } h(\ddot{\theta}_n^\flat) = h(\theta_*) + \ddot{D}_n^\flat(\ddot{\theta}_n^\flat - \theta_*) + o_{\mathbb{P}}(1).$$

Plugging these into the first-order conditions, we obtain that

$$\begin{aligned}\sqrt{n}(\ddot{\theta}_n^\flat - \theta_*) =& - \{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}(J - \ddot{D}_n^{\flat\prime}\{\ddot{E}_n^\flat\}^{-1}\ddot{D}_n^\flat\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1})\sqrt{n}\psi_n \\ & - \{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}\ddot{D}_n^{\flat\prime}\{\ddot{E}_n^\flat\}^{-1}\sqrt{n}h(\theta_*) + o_{\mathbb{P}}(1) \qquad \text{and} \end{aligned} \tag{27}$$

$$\sqrt{n}\ddot{\lambda}_n^\flat = \{\ddot{E}_n^\flat\}^{-1}\ddot{D}_n^\flat\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}\sqrt{n}\psi_n + \{\ddot{E}_n^\flat\}^{-1}\sqrt{n}h(\theta_*),$$

where $\ddot{E}_n^\flat := -\ddot{D}_n^\flat\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}\ddot{D}_n^{\flat\prime}$, and $\psi_n := \nabla_\theta q_n(\theta_*)$.

Given this, applying Theorem 3 implies that $\sqrt{n}\psi_n \overset{\text{A}}{\sim} N(0, 4B)$. Furthermore, under $\mathbb{H}_o$, $\ddot{\theta}_n^\flat \to \theta_*$ a.s.$-\mathbb{P}$, and $\widehat{\theta}_n \to \theta_*$ a.s.$-\mathbb{P}$, so that $\ddot{E}_n^\flat \to -\frac{1}{2}D_*A^{-1}D_*'$ a.s.$-\mathbb{P}$. Therefore, $\sqrt{n}\ddot{\lambda}_n^\flat \overset{\text{A}}{\sim} N[0, 4(D_*A^{-1}D_*')^{-1}D_*A^{-1}BA^{-1}D_*'(D_*A^{-1}D_*')^{-1}]$, implying that $\frac{n}{4}\ddot{\lambda}_n^{\flat\prime}D_*A^{-1}D_*'(D_*A^{-1}BA^{-1}D_*')^{-1}D_*A^{-1}D_*'\ddot{\lambda}_n^\flat \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$. Given this, we can also obtain that $\mathcal{LM}_n^\flat \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$ from Theorem 6, $\nabla_\theta q_n(\ddot{\theta}_n^\flat) \equiv \ddot{\lambda}_n^{\flat\prime}\ddot{D}_n^\flat$, and the fact that $\ddot{\theta}_n^\flat \to \theta_*$ a.s.$-\mathbb{P}$ under $\mathbb{H}_o$. Meanwhile, $\sqrt{n}h(\theta_*) = O_{\mathbb{P}}(\sqrt{n})$ though not $o_{\mathbb{P}}(\sqrt{n})$ under $\mathbb{H}_a$; $\ddot{D}_n^\flat \to D_\dagger := D(\theta_\dagger)$ a.s.$-\mathbb{P}$; and $\nabla_\theta^2 q_n(\bar{\theta}_n^\flat) \to \nabla_\theta^2 q(\theta_\S)$ a.s.$-\mathbb{P}$ for some $\theta_\S$ such that $\nabla_\theta q(\theta_\dagger) = \nabla_\theta q(\theta_*) + \nabla_\theta^2 q(\theta_\S)(\theta_\dagger - \theta_*)$, so that $\sqrt{n}\ddot{\lambda}_n^\flat = O_{\mathbb{P}}(\sqrt{n})$ though not $o_{\mathbb{P}}(\sqrt{n})$. Therefore, the desired result follows.

(*ii*) This proof is almost identical to the proof of Theorem 9(*i*). The TSCFLS estimator can be obtained by minimizing the Lagrange function: $\widehat{\mathcal{L}}_n(\theta, \lambda) := \widehat{q}_n(\theta) - \lambda' h(\theta)$, whose first-order conditions are given as $\nabla_\theta \widehat{q}_n(\ddot{\theta}_n^\sharp) - \ddot{\lambda}_n^{\sharp\prime}\ddot{D}_n^\sharp \equiv 0$ and $h(\ddot{\theta}_n^\sharp) \equiv 0$, where $(\ddot{\theta}_n^{\sharp\prime}, \ddot{\lambda}_n^{\sharp\prime})'$ is the solution for the first-order conditions. In addition, for some $\bar{\theta}_n^\sharp$ between $\ddot{\theta}_n^\sharp$ and θ_*,

$$\nabla_\theta \widehat{q}_n(\ddot{\theta}_n^\sharp) = \nabla_\theta \widehat{q}_n(\theta_*) + \nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)(\ddot{\theta}_n^\sharp - \theta_*) \text{ and } h(\ddot{\theta}_n^\sharp) = h(\theta_*) + \ddot{D}_n^\sharp(\ddot{\theta}_n^\sharp - \theta_*) + o_{\mathbb{P}}(1).$$

By plugging these into the first-order conditions, we obtain that

$$\begin{aligned}\sqrt{n}(\ddot{\theta}_n^\sharp - \theta_*) =& - \{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}(J - \ddot{D}_n^{\sharp\prime}\{\ddot{E}_n^\sharp\}^{-1}\ddot{D}_n^\sharp\{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1})\sqrt{n}\widehat{\psi}_n \\ & - \{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}\ddot{D}_n^{\sharp\prime}\{\ddot{E}_n^\sharp\}^{-1}\sqrt{n}h(\theta_*) + o_{\mathbb{P}}(1) \qquad \text{and} \end{aligned} \tag{28}$$

$$\sqrt{n}\ddot{\lambda}_n^\sharp = \{\ddot{E}_n^\sharp\}^{-1}\ddot{D}_n^\sharp\{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}\sqrt{n}\psi_n + \{\ddot{E}_n^\sharp\}^{-1}\sqrt{n}h(\theta_*), \tag{29}$$

where $\ddot{E}_n^\sharp := -\ddot{D}_n\{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}\ddot{D}_n^{\sharp\prime}$, and $\widehat{\psi}_n := \nabla_\theta \widehat{q}_n(\theta_*)$.

Given this, $\sqrt{n}\psi_n \overset{\text{A}}{\sim} N(0, 4B_*)$ by applying (21). Further, $\ddot{\theta}_n^\sharp \to \theta_*$ a.s.$-\mathbb{P}$ under $\mathbb{H}_o$, and $\widetilde{\theta}_n \to \theta_*$ a.s.$-\mathbb{P}$, so that $\ddot{E}_n^\sharp \to -\frac{1}{2}D_*A^{-1}D_*'$ a.s.$-\mathbb{P}$ and $\sqrt{n}\ddot{\lambda}_n^\sharp \overset{\text{A}}{\sim} N[0, 4(D_*A^{-1}D_*')^{-1}D_*A^{-1}B_*A^{-1}D_*'(D_*A^{-1}D_*')^{-1}]$, implying that $\frac{n}{4}\ddot{\lambda}_n^{\sharp\prime}D_*A^{-1}D_*'(D_*A^{-1}B_*A^{-1}D_*')^{-1}D_*A^{-1}D_*'\ddot{\lambda}_n^\sharp \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$. Therefore, $\mathcal{LM}_n^\sharp \overset{\text{A}}{\sim} \mathcal{X}^2(r, 0)$ by Theorem 7, $\nabla_\theta \widehat{q}_n(\ddot{\theta}_n^\sharp) \equiv \ddot{\lambda}_n^{\sharp\prime}\ddot{D}_n^\sharp$, and the fact that $\ddot{\theta}_n^\sharp \to \theta_*$ a.s.$-\mathbb{P}$ under $\mathbb{H}_o$. Meanwhile, $\sqrt{n}h(\theta_*) = O_{\mathbb{P}}(\sqrt{n})$ though not $o_{\mathbb{P}}(\sqrt{n})$ under $\mathbb{H}_a$; $\ddot{D}_n^\sharp \to D_\dagger := D(\theta_\dagger)$ a.s.$-\mathbb{P}$; and $\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp) \to \nabla_\theta^2 q(\theta_\S)$ a.s.$-\mathbb{P}$ for some $\theta_\S$ such that

$\nabla_\theta q(\theta_\dagger) = \nabla_\theta q(\theta_*) + \nabla^2_\theta q(\theta_\S)(\theta_\dagger - \theta_*)$, so that $\sqrt{n}\ddot{\lambda}^\sharp_n = O_{\mathbb{P}}(\sqrt{n})$ though not $o_{\mathbb{P}}(\sqrt{n})$. This leads to the desired result follows. ■

Proof of Theorem 10: (*i*) By the mean-value theorem and the first-order condition for $\widehat{\theta}_n$, note that for some $\widetilde{\theta}^\flat_n$ between $\widehat{\theta}_n$ and $\ddot{\theta}^\flat_n$,

$$q_n(\ddot{\theta}^\flat_n) = q_n(\widehat{\theta}_n) + \frac{1}{2}(\ddot{\theta}^\flat_n - \widehat{\theta}_n)'\{\nabla^2_\theta q_n(\widetilde{\theta}^\flat_n)\}(\ddot{\theta}^\flat_n - \widehat{\theta}_n). \tag{30}$$

Furthermore, it follows that

$$\sqrt{n}(\ddot{\theta}^\flat_n - \widehat{\theta}_n) = \{\nabla^2_\theta q_n(\bar{\theta}^\flat_n)\}^{-1}\ddot{D}^{\flat\prime}_n\{\ddot{E}^\flat_n\}^{-1}\sqrt{n}[\ddot{D}^\flat_n\{\nabla^2_\theta q_n(\bar{\theta}^\flat_n)\}^{-1}\psi_n - h(\theta_*)] + o_{\mathbb{P}}(1) \tag{31}$$

from (27) and (16). As given in the proof of Theorem 9(*i*), $\ddot{\theta}^\flat_n \to \theta_*$ a.s.$-\,\mathbb{P}$, $\widetilde{\theta}^\flat_n \to \theta_*$ a.s.$-\,\mathbb{P}$ under $\mathbb{H}_o$, and $\widehat{\theta}_n \to \theta_*$ a.s.$-\,\mathbb{P}$, so that $\ddot{E}^\flat_n \to -\frac{1}{2}D_*A^{-1}D'_*$ a.s.$-\,\mathbb{P}$ and $\nabla^2_\theta q_n(\widetilde{\theta}^\flat_n) \to 2A$ a.s.$-\,\mathbb{P}$. In addition, $\sqrt{n}\psi_n \overset{\text{A}}{\sim} N(0, 4B)$ as given in the proof of Theorem 3. Therefore, if we plug all these into (30), $n\{q_n(\ddot{\theta}^\flat_n) - q_n(\widehat{\theta}_n)\} \Rightarrow W'(D_*A^{-1}D'_*)^{-1}W$ under $\mathbb{H}_o$. Meanwhile, $\sqrt{n}(\ddot{\theta}^\flat_n - \widehat{\theta}_*)$ is not bounded in probability under $\mathbb{H}_a$ mainly because $\sqrt{n}h(\theta_*) = O(\sqrt{n})$ though not $o(\sqrt{n})$ in (31); $\ddot{D}^\flat_n \to D_\dagger := D(\theta_\dagger)$ a.s.$-\,\mathbb{P}$; and $\nabla^2_\theta q_n(\bar{\theta}^\flat_n) \to \nabla^2_\theta q(\theta_\S)$ a.s.$-\,\mathbb{P}$ for some $\theta_\S$ such that $\nabla_\theta q(\ddot{\theta}^\flat) = \nabla_\theta q(\theta_*) + \nabla^2_\theta q(\theta_\S)(\ddot{\theta}^\flat - \theta_*)$. Thus, the desired result follows.

(*ii*) The proof is almost identical to the proof of Theorem 10(*i*). By the mean-value theorem and the first-order condition for $\widetilde{\theta}_n$, for some $\widetilde{\theta}^\sharp_n$ between $\widetilde{\theta}_n$ and $\ddot{\theta}^\sharp_n$,

$$q_n(\ddot{\theta}^\sharp_n) = q_n(\widetilde{\theta}_n) + \frac{1}{2}(\ddot{\theta}^\sharp_n - \widetilde{\theta}_n)'\{\nabla^2_\theta q_n(\widetilde{\theta}^\sharp_n)\}(\ddot{\theta}^\sharp_n - \widetilde{\theta}_n). \tag{32}$$

Furthermore, it follows that

$$\sqrt{n}(\ddot{\theta}^\sharp_n - \widetilde{\theta}_n) = \{\nabla^2_\theta q_n(\bar{\theta}^\sharp_n)\}^{-1}\ddot{D}^{\sharp\prime}_n\{\ddot{E}^\sharp_n\}^{-1}\sqrt{n}[\ddot{D}^\sharp_n\{\nabla^2_\theta q_n(\bar{\theta}^\sharp_n)\}^{-1}\widehat{\psi}_n - h(\theta_*)] + o_{\mathbb{P}}(1) \tag{33}$$

by (28) and (19). As given in the proof of Theorem 9(*ii*), it also follows that $\ddot{\theta}^\sharp_n \to \theta_*$ a.s.$-\,\mathbb{P}$, $\widetilde{\theta}^\sharp_n \to \theta_*$ a.s.$-\,\mathbb{P}$ under $\mathbb{H}_o$, and $\widetilde{\theta}_n \to \theta_*$ a.s.$-\,\mathbb{P}$, so that $\ddot{E}^\sharp_n \to -\frac{1}{2}D_*A^{-1}D'_*$ a.s.$-\,\mathbb{P}$ and $\nabla^2_\theta q_n(\widetilde{\theta}^\sharp_n) \to 2A$ a.s.$-\,\mathbb{P}$. Further, $\sqrt{n}\widehat{\psi}_n \overset{\text{A}}{\sim} N(0, 4B_*)$ as given in the proof of Theorem 5. Thus, if we plug all these into (32), $n\{q_n(\ddot{\theta}^\sharp_n) - q_n(\widetilde{\theta}_n)\} \Rightarrow W'_*(D_*A^{-1}D'_*)^{-1}W_*$ under $\mathbb{H}_o$. Meanwhile, $\sqrt{n}(\ddot{\theta}^\sharp_n - \widetilde{\theta}_n)$ is not bounded in probability under $\mathbb{H}_a$ mainly because $\sqrt{n}h(\theta_*) = O(\sqrt{n})$ though not $o(\sqrt{n})$ in (33); $\ddot{D}^\sharp_n \to D_\dagger := D(\theta_\dagger)$ a.s.$-\,\mathbb{P}$; and $\nabla^2_\theta q_n(\bar{\theta}^\sharp_n) \to \nabla^2_\theta q(\theta_\S)$ a.s.$-\,\mathbb{P}$ for some $\theta_\S$ such that $\nabla_\theta q(\ddot{\theta}^\sharp) = \nabla_\theta q(\theta_*) + \nabla^2_\theta q(\theta_\S)(\ddot{\theta}^\sharp - \theta_*)$. Therefore, the desired result follows. ■

C Estimating the Population Mean Function

In this section, we discuss on estimating the population mean function of $g_i(\,\cdot\,)$: for each γ,

$$\mu(\gamma) := \int g(\gamma)d\mathbb{P}(g(\gamma))$$

that is not a function of x any longer, so that for each γ, we may denote it as $\mathbb{E}[g_i(\gamma)]$. Estimating and inferring on $\mu(\,\cdot\,)$ cannot be made by $\mathcal{M}$ due to the presence of x_i in $\rho_i(\,\cdot\,,\theta)$. We, therefore, suppose another model without x_i as follows:

$$\mathcal{M}_0 := \{\rho(\,\cdot\,,\theta) : \Gamma \mapsto \mathbb{R} | \theta \in \Theta \in \mathbb{R}^d\}$$

and estimate $\mu(\,\cdot\,)$ by the FLS estimation:

$$\ddot{\theta}_n := \arg\min_{\theta\in\Theta} \ddot{q}_n(\theta), \quad \text{where} \quad \ddot{q}_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}\int \{\widetilde{g}_i(\gamma,\widehat{\xi}_n) - \rho(\gamma,\theta)\}^2 d\mathbb{Q}(\gamma)$$

that again is designed to estimate

$$\ddot{\theta} := \arg\min_{\theta\in\Theta} \ddot{q}(\theta), \quad \text{where} \quad \ddot{q}(\theta) := \int\int \{g(\gamma) - \rho(\gamma,\theta)\}^2 d\mathbb{P}(g(\gamma) d\mathbb{Q}(\gamma),$$

consistently. Note that this FLS estimator is special case of Section 3, so that the consistence and asymptotic normality of $\ddot{\theta}_n$ can be achieved under milder conditions than those in Section 3. We first collect those conditions as follows:

Assumption 14. *(i) For each $\theta \in \Theta$, $\rho(\,\cdot\,,\theta) : \Gamma \mapsto \mathbb{R}$ is measurable $-\mathcal{G}$;*

(ii) for each $\gamma \in \Gamma$, $\rho(\gamma,\,\cdot\,) : \Theta \mapsto \mathbb{R}$ is in $\mathcal{C}^{(2)}(\Theta)$;

(iii) Θ is a compact and convex set in $\mathbb{R}^d$ ($d \in \mathbb{N}$);

(iv) $\ddot{\theta}_*$ is unique in and interior to Θ;

(v) $\lambda_{\min}(A_0) > 0$, where $A_0 := \int \nabla_\theta\rho(\gamma,\ddot{\theta}_*)\nabla'_\theta\rho(\gamma,\ddot{\theta}_*)d\mathbb{Q}(\gamma) - \int\{\mu(\gamma) - \rho(\gamma,\theta_*)\}\nabla^2_\theta\rho(\gamma,\ddot{\theta}_*)d\mathbb{Q}(\gamma)$;

(vi) for some $m_i \in L^2(\mathbb{P})$, $\sup_{(\gamma,\xi)\in\Gamma\times\Xi}|\widetilde{g}_i(\gamma,\xi)| \leq m_i$ and $\sup_j \sup_{(\gamma,\xi)\in\Gamma\times\Xi}|(\partial/\partial\xi_j)\widetilde{g}_i(\gamma,\xi)| \leq m_i$;

(vii) $\sup_{\theta\in\Theta}|\rho(\,\cdot\,,\theta)| \in L^2(\mathbb{Q})$ and for each j and $j' = 1, 2, \ldots, d$, $\sup_{\theta\in\Theta}|(\partial/\partial\theta_j)\rho(\,\cdot\,,\theta)| \in L^2(\mathbb{Q})$ and $\sup_{\theta\in\Theta}|(\partial^2/\partial\theta_j\partial\theta_{j'})\rho(\,\cdot\,,\theta)| \in L^2(\mathbb{Q})$;

(viii) C_0 is positive definite, where

$$C_0 := \begin{bmatrix} J & K'_0 \\ K_0 & B_0 \end{bmatrix},$$

$K_0 := \int \mathbb{E}_\mathbb{P}[s_i(\xi_*)\{\widetilde{g}_i(\gamma,\xi_*) - \rho(\gamma,\ddot{\theta}_*)\}]d\mathbb{Q}(\gamma)$, $B_0 := \int\int \nabla_\theta\rho(\gamma,\ddot{\theta}_*)\kappa(\gamma,\widetilde{\gamma})\nabla'_\theta\rho(\tilde{\gamma},\ddot{\theta}_*)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$, and $\kappa(\gamma,\widetilde{\gamma}) := \int(g(\gamma) - \rho(\gamma,\ddot{\theta}_*))(g(\widetilde{\gamma}) - \rho(\widetilde{\gamma},\ddot{\theta}_*))d\mathbb{P}(g(\gamma), g(\widetilde{\gamma}))$; and

(ix) $\lambda_{\min}(B_\dagger) > 0$, where $B_\dagger := B_0 - M_0H^{-1}K_0 - K_0'^{-1}M_0 + M_0H^{-1}JH^{-1}M'_0$ and $M_0 := \int \nabla_\theta\rho(\gamma,\ddot{\theta}_*)$ $\mathbb{E}_\mathbb{P}[\nabla'_\xi\widetilde{g}_i(\gamma,\xi_*)]d\mathbb{Q}(\gamma)$. □

There is a correspondence between Assumption 14 and the earlier conditions. Assumptions 14(*i*, *ii*, *iii*, and *iv*) corresponds to Assumption 2. Due to the absence of x_i in the model, the conditions in Assumption 2 are appropriately modified. Assumption 14(*iv*) also corresponds to Assumption 4. Also, note that the integrands of A_0 are non-stochastic, whereas A has the stochastic integrands. This difference again stems from the absence of x_i from the model $\mathcal{M}_0$. Assumptions 14(*vi* and *vii*) correspond to Assumption 7. Assumption 14(*vi*) is the same as Assumptions 7(*i* and *iii*), but Assumption 14(*vii*) is milder than Assumption 7(*ii*, *iv*, and *v*). We do not need the stochastic bound conditions for the model and its derivatives due to the absence of x_i. Finally, note that Assumptions 14(*viii* and *ix*) correspond to Assumption 9.

The following corollary provides the asymptotic behaviors of $\ddot{q}_n(\cdot)$ and $\ddot{\theta}_n$ using this condition additional to the conditions for $\widehat{\xi}_n$:

Corollary (A). *Given Assumptions 6, 8, and 14,*

(i) $\sup_{\theta\in\Theta}|\ddot{q}_n(\theta)-\ddot{q}(\theta)|\to 0$ *a.s.*$-\mathbb{P}$*;*

(ii) $\ddot{\theta}_n\to\ddot{\theta}_*$ *a.s.*$-\mathbb{P}$*;*

(iii) $\sqrt{n}(\ddot{\theta}_n-\ddot{\theta}_*)\overset{\text{A}}{\sim}N(A_0^{-1}B_\dagger A_0^{-1})$*; and*

(iv) if ξ_* *is known,* $\sqrt{n}(\ddot{\theta}_n-\ddot{\theta}_*)\overset{\text{A}}{\sim}N(A_0^{-1}B_0A_0^{-1})$. □

Due to the parallel structure between $\ddot{\theta}_n$ to $\widetilde{\theta}_n$, Corollary (C) can be proved by iterating the proofs in the earlier sections. Therefore, we omit the proof.

References

AKHARIF, A., FIHRI, M., HALLIN, M., AND MELLOUK, A. (2018): "Optimal Pseudo-Gaussian and Rank-Based Random Coefficient Detection in Multiple Regression," Discussion Papers ECARES, 2018-39, ULB–Universite Libre de Bruxelles.

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

ANDREWS, D. (1992): "Generic Uniform Converence," *Econometric Theory*, 8, 241–157.

ANDREWS, D. (1993): "An Introduction to Econometric Applications of Empirical Process Theory for Dependent Random Vairables," *Econometric Reviews*, 12, 183–216.

ANDREWS, D. (1994): "Empirical Process Methods in Econometrics," in R. Engle and D. McFadden, eds. *Handbook of Econometrics*, 4, 2247–2294.

ANDREWS, D. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 543–563.

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

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

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

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

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

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

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

BUGNI, F., HALL, P., HOROWITZ, P., AND NEUMANN, G. (2009): "Goodness-of-Fit Tests for Functional Data," *Econometric Journal*, 12, S1–S18.

CAI, T. AND HALL, P. (2006): "Prediction in Functional Linear Regression," *Annals of Statistics*, 34, 2159–2179.

CAO, G., YANG, L., AND TODEM, D. (2012): "Simultaneous Inference for the Mean Function Based on Dense Functional Data," *Journal of Nonparametric Statistics*, 24, 359–377.

CARDOT, H., CRAMBES, C., KNEIP, A., AND SARDA, P. (2007): "Smoothing Splines Estimators in Functional Linear Regression with Errors-in-Variables," *Computational Statistics & Data Analysis*, 51, 4832–4848.

CHARNIGO, R. AND SUN, J. (2008): "Testing Homogeneity in Discrete Mixtures," *Journal of Statistical Planning and Inference*, 138, 1368–1388.

CHEN, H., CHEN, J., AND KALBFLEISCH, J. D. (2001): "A Modified Likelihood Ratio Test for Homogeneity in Finite Mixture Models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 19–29.

CHEN, X. AND FAN, Y. (2006): " Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models under Copula Misspecification," *Journal of Econometrics*, 135, 125–154.

CHEN, J. AND LI, P. (2009): "Hypothesis Test for Normal Mixture Models: The EM Approach," *Annals of Statistics*, 37, 2523–2542.

CHERNOFF, H. AND LANDER, E. (1995): "Asymptotic Distribution of the Likelihood Ratio Test That a Mixture of Two Binomials is a Single Binomial," *Journal of Statistical Planning and Inference*, 43, 19–40.

CHO, J., PARK, M., AND PHILLIPS, P.C.B. (2018): "Practical Kolmogorov-Smirnov Testing by Minimum Distance Applied to Measure Top Income Shares in Korea," *Journal of Business & Economic Statistics*, 36, 523–537.

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

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

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

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

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

DIAS, A. AND EMBRECHTS, P. (2004): "Dynamic Copula Models for Multivariate High-Frequency Data in Finance," Discussion Papers 04-01, Warwick Business School, Finance Group.

DIKS, C., PANCHENKO, V., AND VAN DIJK, D. (2010): "Out-of-Sample Comparison of Copula Specifications in Multivariate Density Forecasts," *Journal of Economic Dynamics and Control*, 34, 1596–1609.

DOOB, J. (1953): *Stochastic Processes*. New York: Wiley.

DOUKHAN, P., MASSART, P., AND RIO, E. (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques*, 31, 393–427.

ENGLE, R. AND WATSON, M. (1985): "The Kalman Filter: Applications to Forecasting and Rational Expectations Models," in T. Bewley, eds. *Advances in Econometrics: Fifth World Congress*, 1, 245–283. New York: Cambridge University Press.

EVERITT B. AND HAND D. (1981): *Finite Mixture Distributions*. Netherlands: Springer.

FERRATY, F. AND VIEU, P. (2006): *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer-Verlag.

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

FU, Y., CHEN, J., AND LI, P. (2008): " Modified Likelihood Ratio Test for Homogeneity in a Mixture of von Mises Distributions," *Journal of Statistical Planning and Inference*, 138, 667–681.

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

HALL, P. AND HOROWITZ, J. (2007): "Methodology and Convergence Rates for Functional Linear Regression," *Annals of Statistics*, 35, 70–91.

HANSEN, B. (1996): "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

HILDRETH, C. AND HOUCK, J. (1968): "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association*, 63, 584–595.

HSIAO, C. (1974): "Statistical Inference for a Model with Both Random Cross-Sectional and Time Effects," *International Economic Review*, 15, 12–30.

HU, L. (2006): "Dependence Patterns across Financial Markets: a Mixed Copula Approach," *Applied Financial Economics*, 16, 717–729.

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

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

JOE, H. AND XU, J. (1996): "The Estimation Method of Inference Functions for Margins for Multivariate Models", Technical Report 166, Department of Statistics, University of British Columbia.

JOE, H. (1997): *Multivariate Models and Multivariate Dependence Concepts*. New York: Chapman & Hall/CRC.

JOE, H. (2001): *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability, 73. New York: Chapman & Hall/CRC.

JOE, H. (2014): *Dependence Modeling with Copulas*. New York: Chapman and Hall/CRC.

KOSMIDIS, I. AND KARLIS, D. (2016): "Model-Based Clustering Using Copulas with Applications," *Statistics and Computing*, 26, 1079–1099.

LAI, Y., CHEN, C. W., AND GERLACH, R. (2009): "Optimal Dynamic Hedging via Copula-Threshold-GARCH Models," *Mathematics and Computers in Simulation*, 79, 2609–2624.

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

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

LIANG, K. AND RATHOUZ, P. (1999): "Hypothesis Testing under Mixture Models: Application to Genetic Linkage Analysis," *Biometrics*, 55, 65–74.

LOAIZA-MAYA, R., SMITH, M., AND MANEESOONTHORN, W. (2018): "Time Series copulas for Heteroskedastic Data," *Journal of Applied Econometrics*, 33, 332–354.

MCLACHLAN, G. AND PEEL, D. (2004): *Finite Mixture Models*. New York: John Wiley & Sons.

MÜLLER, H. G. (2012): *Nonparametric Regression Analysis of longitudinal Data*, 46. New York: Springer-Verlag.

MÜLLER, H. G., SEN, R., & STADTMÜLLER, U. (2011): "Functional Data Analysis for Volatility," *Journal of Econometrics*, 165, 233–245.

NELSEN, R. (2007): *An Introduction to Copulas*. New York: Springer-Verlag.

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

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

NING, W., GUPTA, A., YU, C., AND ZHANG, S. (2009): "A Moment-Based Test for Homogeneity in Finite Mixture Models," *Communications in Statistics-Theory and Methods*, 38, 1371–1382.

NIU, X., LI, P., AND ZHANG, P. (2011): "Testing Homogeneity in a Multivariate Mixture Model," *Canadian Journal of Statistics*, 218–238.

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

PITMAN, E. (1949): *Lecture Notes on Nonparametric Statistics*. New York: Columbia University.

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

POLLARD, D. (1980): "Minimum Distance Method of Testing," *Metrika*, 27, 43–70.

RAMANATHAN, T. AND RAJARSHI, M. (1992): "Rank Tests for Testing Randomness of a Regression Coefficient in a Linear Regression Model," *Metrika*, 39, 113–124.

RAMSAY, J. AND DALZELL, C. (1991): "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 539–572.

RAMSAY, J. AND SILVERMAN, B. (1997): *The Analysis of Functional Data*. New York: Springer.

RAMSAY, J. AND SILVERMAN, B. (2002): *Applied Functional Data Analysis: Methods And Case Studies*. New York: Springer.

RANGA RAO, R. (1962): "Relations Between Weak and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659–680.

RICE, J. AND SILVERMAN, B. (1991): "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves," *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 233–243.

ROSENBERG, B. (1973): "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, 2, 399–428.

SCHLATTMANN, P. (2009): *Medical Aapplications of Finite Mixture Models*. Berlin Heidelberg: Springer-Verlag.

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

SKLAR, M. (1959): "Fonctions de Répartition à *n* Dimensions et Leurs Marges," *Publications de l'Institut Statistique de l'Universite de Paris*, 8, 229–231.

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

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

SWAMY, P. AND TAVLAS, G. (1995): "Random Coefficient Models: Theory and Applications," *Journal of Economic Surveys*, 9, 165–196.

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

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

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

WALD, A. (1943): "Tests if Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426–486.

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

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

WHITE, H. (1987): "Specification Testing in Dynamic Models," in T. Bewley ed., *Advances in Econometrics - Fifth World Congress*. 1, 1–58. New York: Cambridge University Press.

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

WONG, T. AND LI, W. (2014): "Test for Homogeneity in Gamma Mixture Models Using Likelihood Ratio," *Computational Statistics & Data Analysis*, 70, 127–137.

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

ZIMMER, D. M. (2010): "The Role of Copulas in the Housing Crisis," *Review of Economics and Statistics*, 94, 607–620.

Size of the Test Statistics						
Statistics	Levels$\backslash n$	25	50	100	300	500
	1%	6.36	3.29	2.10	1.38	1.15
$\mathcal{W}_n^\flat$	5%	12.26	8.30	7.01	5.25	5.45
	10%	17.50	13.71	12.12	10.48	10.51
	1%	6.36	3.29	2.10	1.38	1.15
$\mathcal{LM}_n^\flat$	5%	12.26	8.30	7.01	5.25	5.45
	10%	17.50	13.71	12.12	10.48	10.51
	1%	3.84	2.27	1.82	1.01	1.27
$\mathcal{QLR}_n^\flat$	5%	9.38	7.13	6.66	5.21	5.48
	10%	14.60	12.55	12.13	9.76	10.91

Power of the Test Statistics (Level of Significance: 5%)						
Statistics	$\pi_*\backslash n$	25	50	100	300	500
	0.10	13.76	9.56	10.16	13.46	20.20
	0.20	16.20	17.02	19.86	43.28	65.94
$\mathcal{W}_n^\flat$	0.30	20.42	24.78	37.38	79.32	96.04
	0.40	29.08	38.96	61.68	97.38	99.86
	0.50	37.08	54.30	80.58	99.94	100.0
	0.10	13.76	9.56	10.16	13.46	20.20
	0.20	16.20	17.02	19.86	43.28	65.94
$\mathcal{LM}_n^\flat$	0.30	20.42	24.78	37.38	79.32	96.04
	0.40	29.08	38.96	61.68	97.38	99.86
	0.50	37.08	54.30	80.58	99.94	100.0
	0.10	9.34	8.84*	10.40	16.84	25.42
	0.20	12.94	16.52	22.76	53.38	75.22
$\mathcal{QLR}_n^\flat$	0.30	17.88	26.76	44.34	86.88	98.38
	0.40	27.06	44.34	71.06	99.10	100.0
	0.50	37.34	61.82	87.70	99.96	100.0

Table 5.1: SIZE AND POWER OF THE TEST STATISTICS FOR HOMOGENEITY (IN PERCENT). This table shows the empirical rejection rates of the Wald, LM, and QLR test statistics. Null DGP: $x_i \sim$ IID $\text{Exp}(1)$. Alternative DGP: $x_i \sim$ IID $\pi_*\text{Exp}(1) + (1-\pi_*)\text{Exp}(2)$. Model: $\rho(\,\cdot\,, \theta_1, \theta_2) = \theta_1 + \theta_2(\gamma-1)/(2\gamma-1)^{1/2}$, $\gamma \in [1.5, 2.5]$. Number of Experiments: 5,000.

Size of the Test Statistics						
Statistics	Levels$\backslash n$	25	50	100	300	500
$\mathcal{W}_n^\sharp$	1%	2.45	1.64	1.40	0.97	0.99
	5%	8.28	6.02	5.96	4.87	4.77
	10%	14.03	11.30	10.92	9.73	9.77
$\mathcal{LM}_n^\sharp$	1%	0.66	0.84	1.00	0.82	0.88
	5%	5.14	4.62	5.18	4.48	4.45
	10%	10.62	9.60	9.83	9.20	9.23
$\mathcal{QLR}_n^\sharp$	1%	4.43	2.76	2.14	1.55	1.25
	5%	11.50	7.86	7.35	6.19	5.79
	10%	17.15	13.38	12.82	11.43	11.31

Power of the Test Statistics (Level of Significance: 5%)						
Statistics	$\pi_*\backslash n$	25	50	100	300	500
$\mathcal{W}_n^\sharp$	0.1	7.46	6.72	6.06	8.10	10.62
	0.2	8.58	8.62	9.20	18.04	27.58
	0.3	9.32	11.56	15.58	35.86	52.46
	0.4	11.70	15.16	24.52	55.62	77.20
	0.5	13.74	20.66	33.06	74.44	93.24
$\mathcal{LM}_n^\sharp$	0.1	4.56	5.24	5.42	7.56	10.20
	0.2	5.60	6.38	7.94	17.00	25.98
	0.3	6.32	9.38	13.66	33.80	49.42
	0.4	8.30	12.86	21.88	52.92	74.28
	0.5	9.62	17.20	30.58	71.38	88.84
$\mathcal{QLR}_n^\sharp$	0.1	10.34	8.98	9.00	9.78	12.02
	0.2	11.56	10.44	10.94	20.36	30.10
	0.3	12.80	14.18	18.32	39.02	55.56
	0.4	15.54	18.36	27.16	59.06	79.78
	0.5	17.76	23.68	36.58	77.66	94.42

Table 5.2: SIZE AND POWER OF THE INDEPENDENCE TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald, LM, and QLR test statistics. DGP: $c(u_i, v_i; \pi_*, \gamma_{1*}, \gamma_{2*}) := (1-\pi_*)c_1(u_i, v_i; \gamma_{1*}) + \pi_* c_2(u_i, v_i; \gamma_{2*})$, where $u_i := \Phi(x_i, 0, 1)$ and $v_i := \Phi(y_i; 0.5)$ such that $x_i \sim$ IID $N(0,1)$ and $y_i \sim$ IID $N(0,5)$. Model: $\rho(\,\cdot\,, \theta_1, \theta_2) = \theta_1 + \theta_2\gamma$, $\gamma \in [???]$. Number of Experiments: 10,000.

Size of the Test Statistics							
Statistics	Levels$\backslash n$	25	50	100	300	500	1,000
$\mathcal{W}_n^\sharp$	1%	3.24	3.47	3.04	2.70	2.04	1.77
	5%	12.04	11.74	9.93	8.15	6.73	6.01
	10%	20.81	19.15	16.53	13.66	12.08	11.10
$\mathcal{LM}_n^\sharp$	1%	1.16	2.27	2.56	2.59	1.99	1.81
	5%	8.77	10.16	9.34	7.95	6.63	6.23
	10%	17.88	18.08	15.95	13.55	11.96	10.63
$\mathcal{QLR}_n^\sharp$	1%	2.03	1.83	1.15	0.93	0.85	0.98
	5%	4.61	5.22	5.05	5.23	4.72	5.63
	10%	8.98	10.18	10.43	10.50	9.43	10.25

Power of the Test Statistics (Level of Significance: 5%)							
Statistics	$\pi_*\backslash n$	25	50	100	300	500	1,000
$\mathcal{W}_n^\sharp$	0.01	11.84	9.46	7.90	5.84	6.32	7.58
	0.02	12.02	9.64	8.62	8.06	11.00	20.15
	0.03	12.70	10.44	9.02	12.70	19.72	41.66
	0.04	11.88	11.38	10.56	18.64	31.28	60.26
	0.05	12.46	11.42	12.26	25.44	43.44	78.42
$\mathcal{LM}_n^\sharp$	0.01	8.34	8.02	7.42	5.74	6.20	7.35
	0.02	8.94	8.30	7.92	7.78	10.78	19.93
	0.03	9.40	8.96	8.46	12.36	19.54	41.67
	0.04	10.24	8.98	9.92	18.18	31.04	59.91
	0.05	10.64	9.84	10.34	24.94	43.22	78.44
$\mathcal{QLR}_n^\sharp$	0.01	5.54	6.26	7.16	6.86	8.80	10.76
	0.02	7.56	8.02	9.74	13.14	17.14	28.82
	0.03	7.84	10.16	12.22	21.66	30.28	51.65
	0.04	8.58	11.96	16.46	30.34	44.42	71.72
	0.05	9.72	14.28	21.10	39.08	57.04	86.22

Table 5.3: SIZE AND POWER OF THE RANDOM COEFFICIENT TESTS (IN PERCENT). This table shows the empirical rejection rates of the Wald, LM, and QLR test statistics. Null DGP: $y_i = x_i'\psi_* + u_i$ such that $x_i := (1, z_i)'$ and $(z_i, u_i)' \sim$ IID $N(0, I_2)$, where $\psi_* = (1, 1)'$. Alternative DGP: $y_i = x_i'\psi_* + u_i$ such that $u_i = \pi_*^{1/2} x_i'\Omega^{1/2}(\gamma_*)\nu_i + \delta_*^{1/2}\varepsilon_i$ such that $x_i := (1, z_i)'$ and $(z_i, \nu_i', \varepsilon_i)' \sim$ IID $N(0, I_4)$, where $\psi_* = (1, 1)'$, $\gamma_* = 1$, and $\delta_* = 1$. Model: $\rho(\,\cdot\,, \theta_1, \theta_2) = \theta_1 + \theta_2 \exp(\gamma)$, $\gamma \in [0, 1]$. Number of Experiments: 5,000.