Parametric Inference on the Mean of Functional Data Applied to Lifetime Income Curves

JIN SEO CHO

Beijing Institute of Technology and Yonsei University

Email: jinseocho@yonsei.ac.kr

PETER C. B. PHILLIPS

Yale University, University of Auckland, University of Southampton & Singapore Management University

Email: peter.phillips@yale.edu

JUWON SEO

Department of Economics, National University of Singapore, AS2 #05-28, 1 Arts Link, Singapore 117570

Email: ecssj@nus.edu.sg

This version: October 31, 2019

Abstract

We propose a framework for estimation of the conditional mean function in a parametric model with function space covariates. The approach employs a functional mean squared error objective criterion and allows for possible model misspecification. Under regulatory conditions, consistency and asymptotic normality are established. The analysis extends to situations where the asymptotic properties are influenced by estimation errors arising from the presence of nuisance parameters. Wald, Lagrange multiplier, and quasi-likelihood ratio statistics are studied and asymptotic theory is provided. These procedures enable inference about curve shapes in the observed functional data. Several model specifications where our results are useful are analyzed, including random coefficient models, distributional mixtures, and copula mixture models. Simulations exploring the finite sample properties of our methods are provided. An empirical application conducts lifetime income path comparisons across different demographic groups according to years of work experience. Gender and education levels produce differences in mean income paths corroborating earlier research. However, the mean income paths are found to be proportional so that, upon rescaling, the paths match over gender and across education levels.

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

Subject Class: C11, C12, C80.

Acknowledgements: The authors acknowledge helpful discussions with Kees Jan van Garderen, Kevin Sheppard, Richard Smith, Liangjun Su, Ying Wang, and participants of ANZESG (Wellington, 2019) and SETA (Osaka, 2019).

Cho acknowledges research support from an Isaac Manasseh Meyer Fellowship of the National University of Singapore, Phillips acknowledges research support from a Kelly Fellowship at the University of Auckland and the NSF under Grant No. SES 18-50860 and Seo acknowledges research support from AcRF Tier 1.

1 Introduction

Functional data analysis (FDA) has been attracting increasing attention in the statistical and econometric literatures, partly because of the growing availability of very large cross section and spatio-temporal datasets, partly because of the inherent interest in studying function space, curve, or surface realizations of data, and partly because of its potential for wide applicability with economic, financial, and scientific data. In place of individual point observations, functional data lead naturally to the analysis of continuous phenomenon such as time series curves that record trend or growth trajectories and do so under assumptions that can allow great generality. Ramsay and Dalzell (1991), Rice and Silverman (1991), Ramsay and Silverman (1997), Bosq (2000), and Horvath and Kokoska (2012) are now classical references on FDA. We also refer to Ramsay and Silverman (2002), Cai and Hall (2006), Ferraty and Vieu (2006), Cardot, Crambes, Kneip, and Sarda (2007), Hall and Horowitz (2007), Zhang and Chen (2007), Müller, Sen, and Stadtmüller (2011), Cao, Yang, and Todem (2012), and Müller (2012) for some recent research.

Theoretical developments in FDA often focus on nonparametric model analyses such as functional regression models, so that the objects of interest in estimation and inference are typically nonparametric functions or operators acting on a function space in which the data are defined. Such objects can sometimes be difficult to interpret in practice. Rather than pursue a nonparametric approach, the present study works with a parametric formulation that aims to be amenable to implementation and interpretation in application.

For this goal, we propose a novel and efficient framework for the estimation and inference of the (conditional) mean function of functional data. Our approach is different from many other recent FDA studies in that we assume a parametric model for the mean function, one that may possibly be misspecified, whereas the observations remain nonparametric random elements in a measurable function space. We study the influence of potential misspecification of the mean function on our estimator by examining it in parallel to the analysis of a quasi-maximum likelihood estimator, much as in White's (1982, 1994) investigations with finite dimensional data.

Our approach has several advantages. First, it allows us to construct simple statistical tests for the (conditional) mean functions with the estimated parameters using our asymptotic analysis. In the nonparametric context, inference about the slope function in functional regression often involves technically challenging issues arising from well-known ill-posed functional inversion problems. In our framework on the other hand, the relevant null hypothesis can be easily tested by estimating vector valued parameters, enabling straightforward inference on the mean function. Our approach also provides a convenient way to study the derivatives of population mean functions. Derivative functions such as growth rates of income, wealth, or employment are often of significant interest in economic applications. With nonparametric approaches numerical differentiation methods are typically employed, whereas parametric model estimation enables estimation and testing for exact analytic derivatives. For example, statistical analysis of functional data such as shift registration alignment requires the estimation of the mean function and its exact derivatives, for which we can conveniently apply the methods developed herein.

Some well-known statistical methods that are now used in econometrics provide further motivation for the present parametric functional data framework. For instance, there is a large literature in meta analysis on the combination of independent test statistics, some methods using weighting techniques, in which p-curve analysis may be used.

These methods originated in the work of Fisher (1932) (see also Pearson, 1950; Lancaster, 1961; van Zwet and Oosterhoff, 1967; Westberg, 1985, among others) and are now employed to assess selective reporting in empirical work in terms of publication bias and p-hacking - e.g., see Simonsohn, Nelson and Simmons (2014). Such methods can be viewed in terms of functional data analysis with a particular form of weight function. Other relevant literatures are the unidentified model analysis of Davies (1977, 1987) and the minimum distance tests in Pollard (1980). Both provide motivation for the present approach in which the tests are interpreted as statistics obtained from functional observations – see Section 5 of this study for an example. There is now a substantial literature extending the inferential methods of Davies (1977, 1987) and Pollard (1980) and our approach can be applied to similar problems with various identification features or empirical distributions (e.g., Hansen, 1996; Andrews, 2001; Baek, Cho, and Phillips 2015; Bierens, 1990; Cho, Park, and Phillips 2018; Cho and White, 2007, 2010, 2018a; Stinchcombe and White, 1998, and the references therein). Each observation that underlies the statistics in these studies is a functional observation, making the present approach applicable.

The literature on FDA is growing and the direction of the present research contributes to that expanding literature. To mention a few recent developments, we draw attention to the following studies. Crambes, Gannoun, and Henchiri (2013) examine the estimation of a quantile regression function with a functional covariate by means of a support vector machine (SVM) learning method. Here the dependent variable is a real random variable and the explanatory variable is a functional observation. The authors first apply a linear integral operator to the functional data, converting the functional observation to a random variable, thereby enabling estimation of the conditional quantile functional (i.e., the quantile function between the dependent variable and the transformed observations) using SVM methodology. Li, Robinson, and Shang (2019) study time series of function space curves under long range dependence, establishing limit theory for sample averages, estimating the covariance kernel function of the functional data via functional principal component analysis, and using orthonormal functions to span the dominant subspace of the curves. Chang, Hu, and Park (2019) consider estimation of a functional autoregressive model with serially correlated functional data and establish consistency and asymptotic normality of the autoregressive operator estimator. Phillips and Jiang (2019) study parametric autoregression with function valued time series in stationary and nonstationary cases, establish asymptotic theory of estimation and inference, and apply the methods to analyze household Engel curves among ageing seniors in a wide panel dataset. These papers all relate to the current study in terms of the use of functional data but differ from its focus on estimation and inference of a parametric nonlinear conditional mean function involving functional observations and given vector valued explanatory variables, allowing for possible misspecification.

The paper's organization is as follows. Section 2 describes the data, defines the model, and sets up the estimation criterion in terms of a functional mean squared error (FMSE). Section 3 proposes a functional least squares (FLS) estimator for the parametric conditional mean function. Consistency and asymptotic normality of the FLS estimator is established allowing for possible nuisance effects. Asymptotic covariance matrix estimation is also discussed. Section 4 provides a general framework for hypothesis testing, with extensions of Wald, Lagrange multiplier (LM), and quasi-likelihood ratio (QLR) test statistics and asymptotics to the functional data environment. In Section 5, several model specifications where our results are useful are analyzed, including random coefficient models, distributional mixtures, and copula mixture models. Finite sample simulations are also provided. Section 6 reports an empirical application of the methods to lifetime income path comparisons across different demographic groups. Conclusions are given in Section 7 and proofs are collected in the Appendix.

2 Setup

We are interested in studying data that comprise a set of observable random variables and observable 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 $k \in \mathbb{N}_+$. For example, in our empirical work the function $g_i(\,\cdot\,)$ is used to represent an observable curve that shows an individual's income profile over their lifetime working years or some relevant subset of those years, such as those that follow 10 or more years work experience, signifying maturity in the labor force. The vector x_i embodies relevant individual characteristics. The primary object of interest is then the conditional mean function of the income profile curve given the observable characteristics.

The following conditions are used to provide a framework for analysis.

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

The argument space Γ is the space where the functional observations are defined for a fixed $\omega \in \Omega$. For 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 spaces are judiciously chosen by the researcher to match their[1] specific modeling interests. These may in turn be influenced directly by properties associated with estimators of the conditional mean function and test statistics used for inference. As we detail below, the powers of test statistics devised for inference on the conditional mean function are affected by the selection of the adjunct probability measure. This fact may be used by investigators to guide the choice of $\mathbb{Q}$ judiciously so that tests employed may be conveniently computed and have power against what are viewed as realistic alternatives.

The adjunct probability space and adjunct probability measure are not directly part of the stochastic aspects of the data. But the measurability conditions are useful in defining some of the integrals that are introduced below. If g_i is

[1] Throughout this paper gender-neutral pronouns 'their' and 'they' are used in place of gender specific singular pronouns, a usage dating back to Middle English, employed by many notable English writers including Shakespeare, Byron, and Austen, and increasingly endorsed in 21st century literary usage in place of gender specific pronouns.

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 explicitly considering the properties of the function. It will be convenient to proceed in this development by requiring the continuity condition to hold. Also, in the notation (1) and elsewhere, the argument ω is often suppressed for convenience.

We further suppose that the primary subject of interest is the conditional mean function of g_i, which is defined by the integral[2]

$$\mu(\gamma, x) := \int g(\gamma) d\mathbb{P}(g(\gamma)|x), \tag{2}$$

where $\mathbb{P}(\,\cdot\,|x)$ is the conditional probability measure of $g_i(\gamma)$ given $x_i = x$. For each $\gamma \in \Gamma$, we treat the function $g(\gamma)$ 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 the function $g_i(\,\cdot\,)$ is constant a.s., we can view it as a simple 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 a function ρ. Specifically, for each x,

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

The following conditions are assumed for $\mathcal{M}$ and ρ:[3]

Assumption 2. (Model): *(i) For each $\theta \in \Theta$, $\rho(\,\cdot\,, \theta, \,\cdot\,) : \Gamma \times \Omega \mapsto \mathbb{R}$ is $\mathcal{F} \bigotimes \mathcal{G}-$ measurable, 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) the optimizer θ_* is unique and lies in the interior of the parameter space 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) criterion for estimation of the parametric mean function is defined by the functional $q(\,\cdot\,)$ in analogy to the usual mean squared error (MSE) in least squares estimation. As in standard analysis (*ii*) requires a unique optimizer θ_* in the interior of Θ, thereby avoiding possible non-identified model issues from consideration in this development.

Several additional technical conditions on g_i and $\rho(\cdot, \cdot, x_i)$ are given in the following assumption to assist in subsequent derivations. In what follows, for a probability measure $\mathbb{P}$ on Ω, let $L^\ell(\mathbb{P}) := \{f : \int_\Omega |f(\omega)|^\ell d\mathbb{P}(\omega) < \infty\}$, for $\ell = 1$ and 2.

[2]Here and throughout the rest of the paper, unless otherwise noted, we use notations such as (2) to denote integrals over the whole probability space, so that $\int g(\gamma) d\mathbb{P}(g(\gamma)|x) = \int_\Omega g(\omega, \gamma) d\mathbb{P}(g(\gamma)|x)$.

[3]Note that this framework is significantly different from Bugni, Hall, Horowitz, and Neumann (2009) which is concerned with parametric specifications of functional observations. In particular, we let g_i be a random function with no further specification, although its mean function is parametrically specified.

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

(i) $\sup_{\gamma\in\Gamma} |g_i(\gamma)| \leq m_i$ *a.s.* $-\mathbb{P}$*;*

(ii) $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\rho(\gamma,\theta,x_i)| \leq m_i$ *a.s.* $-\mathbb{P}$*;*

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

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

Assumption 3 is imposed to ensure by domination the existence of $q(\cdot)$ and a global minimum of this functional in conjunction with Assumption 2(*iv*). The moment conditions (*iii*) and (*iv*) also ensure that first and second order conditions apply for minimization of $q(\cdot)$.

In practice, of course, the functional form of μ is unknown and our model class $\mathcal{M}$ 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 below provides a useful decomposition of the functional $q(\theta)$ that characterizes the implications of correct specification and misspecification on the minimizer of $q(\,\cdot\,)$.

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 is uniquely 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 an additional term that reflects the error impact of model misspecification. In this case, θ_* should be understood as a parameter value of θ which minimizes the sum of two squared errors, one being the mean squared error obtained if the model had been correctly specified, the other being the squared error component arising from model misspecification. In general, 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. Accordingly, we henceforth assume that $\mathcal{M}$ is possibly misspecified and the asymptotic properties of our proposed estimator are studied under this setting. It is convenient to employ a slight abuse of notation 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 $\int\int k(x,y)dF(x,y)$ as $\int g(x)dF$ and $\int\int k(x,y)dF$, respectively.

3 Functional Least Squares (FLS)

3.1 FLS Estimation without Nuisance Effects

We consider estimation of θ based on the FMSE criterion. The *functional least squares* (FLS) parametric estimator is defined as the extremum estimator

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

The quantity $q_n(\,\cdot\,)$ in this criterion is the sample analogue of $q(\,\cdot\,)$ and is called the *functional sample mean squared error* (FSMSE). Under the regularity conditions given above, we can show that the FSMSE converges uniformly to $q(\,\cdot\,)$. It then follows that the FLS estimator is consistent for θ_* and is asymptotically normally distributed around θ_*. The following theorem establishes consistency.

Theorem 2. *Given Assumptions 1, 2, and 3, as $n \to \infty$,*

(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 a suitable 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 establishing the SULLN required here we repeatedly invoke the dominated convergence theorem (DCT) to interchange the order of discrete summation and integral operators as $n \to \infty$, for which the moment conditions of Assumption 3 are sufficient. The consistency of the FLS estimator follows directly from the fact that the FSMSE converges to FMSE uniformly on Θ a.s.$-\,\mathbb{P}$ whenever, as in Assumption 2(*iv*), the optimizer $\theta_* := \arg\min_{\theta\in\Theta} q(\theta)$ of the limiting functional $q(\theta)$ is unique.

Asymptotic normality of the FLS estimator is obtained for this function space setting in parallel to standard derivations for least squares estimation. We begin by observing that by standard Taylor expansion and for some $\bar{\theta}_n$ between $\widehat{\theta}_n$ and θ_*,

$$\sqrt{n}(\widehat{\theta}_n-\theta_*) = A_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int[g_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma), \tag{3}$$

where

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

Regular asymptotic behavior 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 to establishing asymptotic normality. The following conditions are sufficient for this purpose.

Assumption 4. (Hessian Matrix): *A is positive definite, where $A := \int\int\nabla_\theta\rho(\gamma,\theta_*,x)\nabla'_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) - \int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla^2_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)$.* □

Assumption 5. (CLT Conditions): *(i) For each j and $j' = 1,2,\ldots,d$, $\int\int\int(\partial/\partial\theta_j)\rho(\gamma,\theta_*,x)\cdot\kappa(\gamma,\widetilde{\gamma}|x)\cdot$*

$(\partial/\partial\theta_{j'})\rho(\widetilde{\gamma},\theta_*,x)d\,\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})<\infty$, *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) B *is positive definite, where* $B:=\int\int\int\nabla_\theta\rho(\gamma,\theta_*,x)\kappa(\gamma,\widetilde{\gamma}|x)\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$. □

The matrix A is the probability limit of A_n, and B serves as the limiting covariance matrix of the score function 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*), and the conditional covariance kernel $\kappa(\,\cdot\,,\,\cdot\,|x)$ in Assumption 5(*i*) contributes to the asymptotic covariance matrix of the FLS estimator by way of the matrix B. Importantly, the presence of the covariance kernel $\kappa(\,\cdot\,,\,\cdot\,|x)$ in B imparts the role of the true probability measure $\mathbb{P}$ in the limit distribution of the FLS estimator. For a different $\mathbb{P}$, a different functional form is obtained for $\kappa(\,\cdot\,,\,\cdot\,|x)$, leading to a different variance matrix B. In addition, B also depends on the parametric specification $\rho_i(\cdot,\theta)$, implying that different limit distributions are to be expected for different models.

The following theorem establishes asymptotic normality of the FLS estimator, as implied by the regularity conditions imposed so far:

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

The result follows in a straightforward manner by applying a strong law and a Linderberg CLT in conjunction with the Cramér-Wold device to the components of (3), proceeding in parallel to usual derivations for nonlinear least squares estimation but here within this function space data setting.

The asymptotic covariance matrix exhibits a sandwich form, as is usual. On the other hand, if $\mathcal{M}$ is correctly specified and if the covariance kernel $\kappa(\,\cdot\,,\,\cdot\,|x)$ can be written in functional diagonal form by involving the Dirac delta function, the information matrix equality holds. Note that if $\mathcal{M}$ is correctly specified,

$$A=\int\int\nabla_\theta\rho(\gamma,\theta_*,x)\nabla'_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma).$$

Furthermore, if we let $\delta(\cdot)$ be the Dirac delta function and suppose that for some $\sigma^2>0$, $\kappa(\gamma,\gamma'|x)=\sigma^2\delta(\gamma-\gamma')$, it now follows that

$$\begin{aligned}\int\int\nabla_\theta\rho(\gamma,\theta_*,x)\kappa(\gamma,\widetilde{\gamma}|x)\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})&=\sigma^2\int\int\nabla_\theta\rho(\gamma,\theta_*,x)\delta(\gamma-\widetilde{\gamma})\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\\&=\sigma^2\int\nabla_\theta\rho(\gamma,\theta_*,x)\nabla'_\theta\rho(\gamma,\theta_*,x)d\mathbb{Q}(\gamma),\end{aligned}$$

so that $B=\sigma^2A$, implying that the asymptotic covariance matrix reduces to σ^2A^{-1}. These additional conditions enable us to achieve the information equality, thereby motivating model specification testing via the information matrix equality test (e.g., White, 1982; Cho and White, 2014; Cho and Phillips, 2018b).

3.2 FLS Estimation with Nuisance Effects

Functional data analysis often involves nuisance effects by the very nature of the data. In particular, when data are constructed using aligned discrete observations, such a construction naturally introduces nuisance effects. To examine this extension, we characterize the 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 the product space $\Gamma \times \Xi$. 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 the data structure generalizes that assumed in Section 3.1 because for some known ξ_*, we can let $g_i(\,\cdot\,)$ be identical to $\widetilde{g}_i(\,\cdot\,, \xi_*)$ here. Therefore, the data analysis given in this section is also applicable to $\{g_i(\,\cdot\,), x_i\}_{i=1}^n$. Notwithstanding this specialization, the asymptotic influence of the nuisance effects on the FLS estimator is not negligible in general and typically modifies the limit behavior of the FLS estimator. The results of Section 3.1 are therefore extended here to accommodate the effects conveyed by nuisance parameter estimation.

Functional data are often influenced by nuisance effects, as described in the Introduction. First, when a model is unidentified under the null (e.g., Davies, 1977, 1987), a functional data set with nuisance effects can be collected by letting each individual observation be the score function defined on the set of unidentified parameters with the other parameters being evaluated at the parameter estimates obtained using the null model. In such cases the null parameter estimates play the role of $\widehat{\xi}_n$, and γ can be treated as generic notation for the unidentified parameters with $\widetilde{g}_i(\cdot, \cdot)$ being the score function. Second, functional observations are often constructed by local polynomial kernel or sieve estimation using discrete observations (e.g., Zhang and Chen, 2007; Chen, 2007). In these cases functional observations are influenced by the kernel or sieve estimation error that we capture by $\widehat{\xi}_n$ here, thereby potentially modifying the large sample properties of the FLS estimator. Specifically, using their optimal bandwidth, Zhang and Chen (2007) show that estimated functional observations obtained by local polynomial kernel estimation uniformly converge to continuous functional observations at the rate $n^{(p+1)\delta/(2p+3)}$, where p is the degree of polynomial function and δ is a positive number such that $n_t \geq Cn^{\delta}$ uniformly in t for some positive number C, and n_t is the number of discrete observations underlying the t-th functional observation. If a sufficiently large p is selected such that $(p + 1)\delta/(2p + 3) > 1/2$, the functional approximation obtained by local polynomial kernel estimation is super-consistent for the latent functional observation, so that the nuisance effect in the approximated functional observation can be ignored in the limit theory when applying FLS estimation because the convergence rate of the FLS estimator is $\sqrt{n}$, as given in Theorem 3. On the other hand, if $(p + 1)\delta/(2p + 3) = 1/2$, the nuisance effect can affect the limit distribution of the FLS estimator, as detailed below. In what follows, we suppose that the same nuisance effect is present for individual functional observations and examine how this effect modifies the large sample behavior of the FLS estimator.

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

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

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

We now proceed to examine how the estimation error imbedded in $\widehat{\xi}_n$ changes the asymptotic behavior of the FLS estimator. To tackle this issue we start by extending the previous regularity conditions for Theorems 2 and 3 to cope with the presence of nuisance effects. 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* $(\Omega \times \Gamma, \mathcal{F} \bigotimes \mathcal{G}, \mathbb{P} \times \mathbb{Q})$ *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) $\sup_{(\gamma,\xi)\in\Gamma\times\Xi} |\widetilde{g}_i(\gamma, \xi)| \leq m_i$ *a.s.* $-\mathbb{P}$*;*

(ii) $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\rho_i(\gamma, \theta)| \leq m_i$ *a.s.* $-\mathbb{P}$*;*

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

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

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

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

$$\begin{aligned}
&\frac{1}{n}\int \sum_{i=1}^n \{\widehat{g}_i(\gamma) - \rho_i(\gamma, \theta_*)\}\nabla_\theta \rho_i(\gamma, \theta_*) d\mathbb{Q}(\gamma) \\
=&\frac{1}{n}\int \sum_{i=1}^n \{\widetilde{g}_i(\gamma, \xi_*) - \rho_i(\gamma, \theta_*)\}\nabla_\theta \rho_i(\gamma, \theta_*) d\mathbb{Q}(\gamma) + \frac{1}{n}\sum_{i=1}^n \int \nabla_\theta \rho_i(\gamma, \theta_*)[\nabla_\xi' \widetilde{g}_i(\gamma, \bar{\xi}_{n,\gamma})] d\mathbb{Q} \cdot (\widehat{\xi}_n - \xi_*). \qquad (4)
\end{aligned}$$

The left side of (4) is the first-order derivative of $\widehat{q}_n(\,\cdot\,)$ with respect to θ evaluated at θ_*, whereas the right side is the Taylor expansion with respect to ξ around ξ_*. Assumption 7(*iii*) enables use of the mean-value theorem. Proving consistency of $\widetilde{\theta}_n$ for θ_* then involves showing that the quantities on the right side of (4) vanish as $n \to \infty$. Since the first component in (4) trivially vanishes by applying the proof of Theorem 3, the result involves showing that the second term 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 consistency of the TSFLS estimator. For this purpose we impose the following high level condition concerning $\widehat{\xi}_n$.

Assumption 8. (E-Estimator-1)*: 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*.* □

Consistency of the TSFLS estimator is then immediate under these conditions.

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.*$-\mathbb{P}$*.* □

The assumptions for Theorem 3 are insufficient to deliver the limit distribution of the TSFLS estimator as they do not address the asymptotic properties of $\widehat{\xi}_n$. To establish asymptotic normality we impose the following additional conditions.

Assumption 8. (E-Estimator-2)*: (ii) there exists a finite nonstochastic* $s \times s$ *positive definite matrix* H *and a sequence of random vectors* $\{s_{*n}\}$ *measurable* $- \mathcal{F}$ *for which* $\sqrt{n}(\widehat{\xi}_n - \xi_*) = -H^{-1}\sqrt{n}s_{*n} + o_{\mathbb{P}}(1)$*;*

(iii) for $i = 1, 2, \ldots$*, there exists* $s_i : \Omega \times \Xi \mapsto \mathbb{R}^s$ *such that:*

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

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

(iii.c) for some $m_i \in L^2(\mathbb{P})$*,* $|s_i(\omega, \,\cdot\,)| \le m_i(\omega)$*; 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*) assume that $\widehat{\xi}_n - \xi_*$ is asymptotically equivalent to the product of the nonstochastic matrix H and the score s_{*n}. Many estimators satisfy this characteristic asymptotically, including least squares, generalized method of moments, and (quasi-)maximum likelihood estimators. In order to retain generality in the analysis, we do not specify here how H and s_{*n} are obtained from primitive model formulations. Treating s_{*n} as being formally defined in Assumption 8, we now use $s_i(\xi)$ to denote $s_i(\omega, \xi)$, suppressing the argument ω for notational ease.

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* B *be as defined earlier in Assumption 5. Let*

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

and assume the following:

(i) C *is positive definite;*

(ii) B_* *is positive definite, 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 characterizes the key components needed for the limiting covariance matrix of the TSFLS estimator in the presence of nuisance effects. It generalizes Assumption 5 to accommodate the additional estimator $\widehat{\xi}_n$. The matrix C is employed to capture the asymptotic covariance matrix between the score vectors of the estimates $\widehat{\xi}_n$ and $\widetilde{\theta}_n$, thereby providing a channel for the nuisance effects to be conveyed to the limit distribution of the TSFLS estimator. With this framework for the nuisance effects in hand, asymptotic normality of the TSFLS estimator is established in the following theorem.

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

As the result makes clear, nuisance effects are not asymptotically negligible. In particular, the asymptotic variance matrix of $\widetilde{\theta}_n$ is changed from $A^{-1}BA^{-1}$ to $A^{-1}B_*A^{-1}$, thereby modifying the limit variability of $\widetilde{\theta}_n$. A further impact of the presence of nuisance effects is to introduce changes in the appropriate construction of test statistic formulae based on the FLS and TSFLS estimators, as shown below.

Before moving to the next section, we note that estimating the unconditional mean function of functional data can be conducted in parallel to the estimation of the conditional mean function. In view of this similarity we do not discuss estimation of the unconditional mean function here but provide that 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 these matrices are used to construct various test statistics. This section examines how these covariance matrices may be estimated consistently.

First, we discuss the case with no nuisance effects. The covariance matrix is given as $A^{-1}BA^{-1}$ in Theorem 3. The domination conditions in Assumption 10 enable application of the SULLN to the estimators of A and B given by $\widehat{A}_n$ and $\widehat{B}_n$ in Theorem 6 below. Then, $\widehat{A}_n^{-1}\widehat{B}_n\widehat{A}_n^{-1}$ provides a consistent estimator of the covariance matrix.

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) $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |\varepsilon_i(\gamma, \theta)| \leq m_i^{1/2}$ *a.s.* $-\mathbb{P}$*;*

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

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

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*

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

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

Next consider models with nuisance effects. Since B_* involves component matrices in its definition, further conditions are needed to ensure consistent estimation. The following conditions are used 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. $-\mathbb{P}$. $\square$

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) $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi} |\varepsilon_i(\gamma,\theta,\xi)| \le m_i^{1/2}$ *a.s.* $-\mathbb{P}$*;*

(ii) for $j = 1,2,\ldots,d$, $\sup_{(\gamma,\theta)\in\Gamma\times\Theta} |(\partial/\partial\theta_j)\rho_i(\gamma,\theta)| \le m_i^{1/2}$ *a.s.* $-\mathbb{P}$*;*

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

(iv) for each $j = 1,2,\ldots,s$, $\sup_{(\gamma,\xi)\in\Gamma\times\Xi} |(\partial/\partial\xi_j)\widetilde{G}_i(\gamma,\xi)| \le m_i$ *a.s.* $-\mathbb{P}$*; and*

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

Under Assumption 11, the two submatrices H and J appearing in B_* can be consistently estimated by using $\widehat{H}_n$ and $\widehat{J}_n$. In general, these estimators are obtained by preliminary estimation of $\widehat{\xi}_n$ and can be easily computed using standard methods. For example, if $\widehat{\xi}_n$ is a (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 $\widehat{\xi}_n$ in the usual fashion. 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 for Ξ as well. Furthermore, Assumptions 12(*iii* and *iv*) require the SULLN to hold on other random functions that we use to provide consistent estimation of the component matrices K and M of B_* that appear in Assumption 9.

The following Theorem provides consistent estimators for A and B_* under these regularity conditions.

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*

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

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

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

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

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

From these results it follows that 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 limit distribution of the 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 Inference on the Mean Function

Parametric specification of the conditional mean of functional data is particularly useful in inference. Instead of conducting inference over Γ, we can test a relevant hypothesis by estimating the unknown parameter θ_* directly. In what follows we extend the standard analysis of Wald, Lagrange multiplier (LM), and quasi likelihood-ratio (QLR) test statistics to perform inference on the functional mean. Specifically, suppose the following hypotheses on the mean function are to be tested:

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

We assume that the function $h(\,\cdot\,)$ is given 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. □

Define the constrained FLS (CFLS) and constrained two-stage FLS (CTSFLS) estimators as

$$\ddot{\theta}_n^\flat := \operatorname*{arg\,min}_{\theta \in \Theta;\ h(\theta)=0} q_n(\theta) \quad \text{and} \quad \ddot{\theta}_n^\sharp := \operatorname*{arg\,min}_{\theta \in \Theta;\ h(\theta)=0} \widehat{q}_n(\theta).$$

These restricted estimates are used in developing the test statistics. In view of the constrained optimization, the criteria $q_n(\ddot{\theta}_n^\flat)$ and $q_n(\ddot{\theta}_n^\sharp)$ cannot be smaller than $q_n(\widehat{\theta}_n)$ and $\widehat{q}_n(\widetilde{\theta}_n)$. In a similar way, we define $\theta_\dagger$ to be the minimizer of $q(\theta)$ under the same restriction, i.e., $\theta_\dagger := \arg\min_{\theta \in \Theta;\ h(\theta)=0} q(\theta)$. The following Lemma establishes consistency of the CFLS and CTSFLS estimators and is repeatedly used in the limit theory of the test statistics.

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

We construct the usual Wald (1943) 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 $\widehat{D}_n := D(\widehat{\theta}_n)$, $\widetilde{D}_n := D(\widetilde{\theta}_n)$ and all other notation ise the same as in Section 3. The statistic $\mathcal{W}_n^\flat$ is used for models without nuisance effects, whereas $\mathcal{W}_n^\sharp$ is used for models with nuisance effects. The next result provides limit theory for these Wald tests.

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 a noncentral chi-square variable with degrees of freedom a and noncentrality parameter b;

(i.b) for any sequence $c_n \to \infty$ 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,

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

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

The null limit distribution of the Wald test statistic is chi-squared with degrees of freedom r, where r is the rank of D_* as given in Assumption 13. The result follows easily 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 with probability one. This proves the consistency of the Wald test statistic.

4.2 Lagrange Multiplier (LM) Test

The LM test statistic is 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_*. 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 both consistent for B under $\mathbb{H}_o$.

Asymptotic theory of the LM statistic relies on the first-order derivatives of q_n and $\widehat{q}_n$ evaluated at the constrained estimates $\ddot{\theta}_n^\flat$ and $\ddot{\theta}_n^\sharp$. Under regularity conditions we have

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

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

as shown in Lemma 3(*iii*) in the Appendix. The following theorem then holds.

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

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

(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 $c_n \to \infty$ *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 limit behavior of the LM statistics under the null and alternative hypotheses. The same null limit distributions apply as for the Wald statistic, because under $\mathbb{H}_o$, both $\nabla_\theta q_n(\ddot{\theta}_n^\flat)$ and $\nabla_\theta q_n(\ddot{\theta}_n^\sharp)$ are asymptotically normal with mean zero and covariance matrices that are consistently estimated by the weight matrices employed in construction of the LM statistics.

4.3 Quasi Likelihood Ratio (QLR) Test

The QLR statistics are defined for the models without and with nuisance effects as

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

Approximating $q_n(\,\cdot\,)$ (resp. $\widehat{q}_n(\,\cdot\,)$) via a second-order Taylor 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 is normal under $\mathbb{H}_o$. When $\mathbb{H}_o$ is not true, this quantity is not bounded in probability, thereby distinguishing the null and alternative. But the QLR statistics do not have limiting chi-square distributions. Instead, their limit behavior under $\mathbb{H}_o$ and $\mathbb{H}_a$ are given in the following result.

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 $c_n \to \infty$ *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 $c_n \to \infty$ *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 differ from standard chi-squared theory because the asymptotic covariance matrices of the FLS and TSFLS estimators differ from the limits of the Hessian matrices of FSMSE and the information matrix equality fails. Importantly, $\widehat{B}_n$ and $\widetilde{B}_n$ still play a critical role in applying the QLR test statistics. Even though computation of the QLR statistics does not rely on these matrices, they are needed to obtain critical values of the QLR tests.

5 Applications and Simulations

This section explains how the theoretical results detailed in Sections 3 and 4 relate to standard analysis in applications. Three examples are given showing how the limit theory is applied. In the first application, we study methods of determining whether the coefficient of a linear regression model is random or constant. The second application considers homogeneity testing in finite mixture models. The third application extends this example to examine testing for heterogeneity in copula mixtures representing dependence structures between two variables. These applications are all considered within our functional data framework. Monte Carlo experiments are conducted to assess the adequacy of the limit theory.

5.1 Inference on Random Coefficient

Standard regression models typically assume the coefficients of explanatory variables are fixed parameters. When such assumptions are violated, statistical inference can be misleading due to biases in estimating the standard errors. Inference using the random coefficient model becomes an alternative approach and is particularly useful in modeling conditional heteroskedasticity or time varying coefficients in the time series models. Many studies in the literature consider testing for random coefficients in regression (e.g., see Hsiao, 1974; Breusch and Pagan, 1979; Ramanathan and Rajarshi, 1992; Swamy and Tavlas, 1995; Akharif, Fihri, Hallin, and Mellouk 2018).

Consider a simple linear regression model given by

$$y_i = x_i'\beta_i + \delta_*^{1/2}\varepsilon_i, \tag{5}$$

where $x_i := (1, z_i)'$ and $z_i \in \mathbb{R}$ is an explanatory variable. 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 function $\Omega(\,\cdot\,)$ is assumed to be positive definite uniformly on the space Γ to which the true parameter γ_* belongs. Accordingly, when the variance of ν_i is positive, the coefficient $\boldsymbol{\beta}_i$ is constant if and only if $\pi_* = 0$.

This type of random coefficient model is commonly studied and applied in the literature. For example, Andrews (2001) developed limit theory for models similar to the random coefficient model where parameters may lie on the boundary of the parameter space. Rosenberg (1973) and Engle and Watson (1985) extended the random coefficient model to include conditional heteroskedastic processes in time series settings; and a number of empirical studies exploited features of the random coefficient model relevant to investigating conditional heterogeneity in time series data (e.g., Swamy and Tinsley, 1980; Stock and Watson, 1998).

In what follows we relate the random coefficient model to FDA model analysis. We first note that substituting the expression of β_i into (5) yields the following conditional heteroskedasticity model

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

where $\psi_* := (\psi_{1*}, \psi_{2*})'$ and $u_i := \pi_*^{1/2}x_i'\Omega^{1/2}(\gamma_*)\nu_i + \delta_*^{1/2}\varepsilon_i$, which leads to the explicit conditional variance function

$$\mathrm{var}(u_i|x_i) = \delta_* + \pi_* x_i'\Omega(\gamma_*)x_i = \delta_* + \pi_*[1 + \exp(\gamma_*)z_i^2] \tag{6}$$

when the variance matrix function $\Omega(\gamma_*)$ has the form

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

Let $\Gamma = \{\gamma : \gamma \in [0,1]\}$ and suppose that the researcher estimates the unknown parameter values by maximum likelihood under Gaussian assumptions, so that the data are generated according to $(\varepsilon_i, \nu_i')'|z_i \sim N(0, I_3)$ and $z_i \sim$ IID $N(0,1)$.

We propose to test the random coefficient property within the FDA model setting. For this purpose, we reformulate the given DGP in the FLS framework. Then, if the regression equation (5) is estimated by maximum likelihood, the following likelihood function is maximized:

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

with respect to the parameters. If we suppose that the coefficient is not random, we have the following first-order derivative of the likelihood function under the null. That is, if $\pi_* = 0$, we have

$$\frac{\partial}{\partial \pi} L_n(\psi_*, \delta_*, \gamma_*, \pi_*) = \frac{1}{2\delta_*^2}\sum_{i=1}^{n}[1 + \exp(\gamma_*)z_i^2]\{(y_i - x_i'\psi_*)^2 - \delta_*\}, \tag{7}$$

and $\mathbb{E}[(y_i - x_i'\psi_*)^2|z_i] = \delta_*$, so that the conditional mean of (7) is zero irrespective of $\gamma \in \Gamma = [0,1]$. On the other hand, if the coefficient is random, the population mean of (7) is obtained as follows:

$$\frac{n\pi_*}{2\delta_*^2}\mathbb{E}\{[1 + \exp(\gamma_*)z_i^2][1 + \exp(\gamma_*)z_i^2]\},$$

using (6), thereby motivating the following random function as a candidate function for $\widetilde{g}_i$:

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

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

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

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

$$\rho_i(\gamma, \theta) := \theta_0(1 + \exp(\gamma)z_i^2) + \theta_1(z_i^2 + \exp(\gamma)z_i^4)$$

to test the following hypotheses:

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

Here, we note that

$$\theta_{0*} = -\pi_* \exp(\gamma_*)\mathbb{E}[z_i^2] \quad \text{and} \quad \theta_{1*} := \pi_* \exp(\gamma_*),$$

so that if $\pi_* = 0$, $\mathbb{H}_o$ holds, whereas $\pi_* \neq 0$ under $\mathbb{H}_a$.

In this FDA framework, we conduct simulations by applying the theorems for the models with nuisance effects in Section 3. We let $(\psi_{1*}, \psi_{2*}, \gamma_*, \delta_*) = (1, 1, 0.5, 1)$ under both the null and alternative hypotheses, and set $\pi_* = 0$ under the null. For the alternative DGP, we consider various values for π_*, viz., 0.10, 0.20, 0.30, 0.40, and 0.50. Next, let $z_i \sim$ IID $U[0, 1]$ and $(\nu_i', \varepsilon_i)' \sim$ IID $N(0, I_3)$, so that the model for $\mu(\cdot)$ is correctly specified. In addition, we let the adjunct probability measure be uniformly distributed over $[0, 1]$ with $\Gamma = [0, 1]$. For the Wald and QLR test statistics, we use $\widetilde{A}_n$ and $\widetilde{B}_n$ when computing the test statistics, whereas $\widetilde{A}$ and $\ddot{B}_n^\sharp$ are used in calculating the LM test statistic.

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

Table 5.1 displays the empirical rejection ratios obtained for the Wald, LM and QLR statistics in testing the reformulated model and hypotheses. The results show that the null rejection rates are close to nominal levels for all three test statistics when the sample size is large. The QLR test statistic tends to be slightly oversized when the sample sizes are small, whereas both the Wald and LM statistics perform better in size control. For power analysis, we consider models with $\pi_* = 0.10$, 0.20, 0.30, 0.40, and 0.50, fixing the nominal significance level at 5%. Evidently, the rejection rates turn out to be dependent on sample size for each value of π_* in all three test statistics. As expected, the empirical rejection rates increase as π_* or n increase. Overall, the QLR test shows better performance than the Wald and LM tests.

5.2 Distribution Specification Tests

Finite mixture models are popular for constructing more flexible distribution functions and for modeling clustered data. They can also be used for certain types of distribution specification tests. To fix ideas, 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 from 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 weights $\pi_i \in [0, 1]$ with $\sum_{i=1}^{K} \pi_i = 1$, the corresponding finite mixture model (Everitt and Hand, 1981; McLachlan and Peel, 2004; Schlattmann, 2009) is defined as the weighted sum

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

Various types of parametric distributions have been explored in making such constructions in the literature. For example, mixtures of normals, binomials, gammas and von Mises distributions are popular in applications. Amongst many references in the literature to the use of mixtures we mention Chernoff and Lander (1995), Liang and Rathouz (1999), Chen, Chen, and Kalbfleisch (2001), Cho and White (2007, 2010), Fu, Chen, and Yi (2008), Chen and Li (2009), Ning, Gupta, Yu, and Zhang (2009), Niu, Li, and Zhang (2011), and Wong and Li (2014).

We focus here on inference concerning sample homogeneity as an application of distribution specification tests.

For a specific example 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, assume $\underline{\gamma} > 1$ and $\bar{\gamma} < \infty$. A primary concern in this mixture distribution is to test whether $\pi_* = 0$ for if this hypothesis holds the observations come from a homogeneous population that follows a standard exponential distribution. Davies (1977) applied Neyman's (1959) C(α) test principle and derived a maximal test statistic defined as $\sup_{\gamma\in\Gamma} n^{-1/2}\sum_{i=1}^{n} g_i(\gamma)$, where

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

Note that in this formulation $g_i(\cdot)$ is a random function defined on Γ, so that we may treat it as a functional observation and apply the theory of the present study.

Instead of basing the test on the supremum over γ, we consider the Wald, LM and QLR test statistics introduced in Section 4 treating $g_i : \Gamma \mapsto \mathbb{R}$ as a random function. Importantly, critical values of the standard C(α) test statistic are obtained by explicitly exploiting the functional form of g_i in the construction of that statistic. By contrast, our test statistics do not exploit this feature. In consequence, our statistics lead to tests that typically have lower power than that of the C(α) statistic at least in the specific direction of the alternative implied by g_i. Notwithstanding this apparent disadvantage, lower power in this context should not necessarily discourage use of our tests because assuming knowledge of the functional form of $g_i(\cdot)$ raises specificity and consequently reduces the range of applicability of the maximal test statistic.

For the simulation exercise that follows 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 $(0,0)$. The null and alternative hypotheses are specified as

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

Table 5.2 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 a Newton-Raphson iteration and the associated integrals are computed by Gauss-Legendre numerical quadrature. 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 the adjunct probability measure be the probability measure uniformly distributed on Γ and consider sample sizes $n = 25, 50, 100, 300$, and 500. The nominal levels are fixed at 1%, 5%, and 10%. In the level panel of Table 5.2, we observe that the rejection rates of the three test statistics all approach nominal levels as the sample size increases. Under the alternative, power is computed through 5,000 replications with the same sample sizes, but with the nominal level fixed at 0.05. In particular, we examine power of the tests by letting π_* vary over the range $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. As expected, the rejection rates tend to be larger as we move π_* further from zero; and when the sample size increases, rejection rates approach unity for fixed π_*.

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

5.3 Inference on the Homogeneity of Dependence Structure

As a further application we develop tests for heterogeneity in dependence structures by applying the mixture model assumption. For this purpose, suppose a researcher observes IID observations of multiple variables. Empirical interest often lies in determining whether the dependence structure among these variables is homogeneous. Even though the univariate marginals for each variate may remain constant over the whole population, observations can still be heterogeneous due to different dependence structures.

In what follows we provide test statistics to detect violations of homogeneity using finite copula mixtures. The Sklar theorem (1959) is useful for this purpose as it conveniently separates information on the univariate marginals from the joint distribution by means of the copula function. There is now a vast literature demonstrating the use of copulas for studying dependence structures (e.g., Joe, 1997; Nelsen, 2007; Joe, 2014; and the references therein) and many studies applying mixture copula models (e.g., Dias and Embrechts, 2004; Chen and Fan, 2006; Hu, 2006; Lai, Chen, and Gerlach 2009; Diks, Panchenko, and van Dijik 2010; Zimmer, 2010; Kosmidis and Karlis, 2016; Loaiza-Maya, Smith, and Maneesoonthorn 2018). But methods of inference concerning homogeneity in dependence structures based on finite mixture copula models has, to the best of our knowledge, so far not been addressed.

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 component densities can be considered. But we focus on the simple protypical model above for brevity. We suppose that IID pairs $\{(x_i, y_i)\}_{i=1}^n$ have marginal distributions given as F_X and F_Y, respectively. Inference concerning the homogeneity of the dependence structure can naturally be cconducted by examining the null hypothesis that $\pi_* = 0$ (or $\pi_* = 1$) with function 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

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

Here, g_i is derived by applying the C(α) test principle for testing $\pi_* = 1$.

A practical challenge arises from the fact that the univariate marginals F_X and F_Y are typically unknown to 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 estimates of the marginal distributions in a fashion similar to the inference function for marginals (IFM) approach (Joe and Jianmeng, 1996; Joe, 2001). To apply the results in Section 4.2, we construct the functional data as follows

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

A leading example is the case when $c_1(\cdot)$ is the density implied by the independence copula. If so, the functional form of $g_i(\cdot)$ is further simplified as

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

For our simulation experiments we use the Farlie-Gumbel-Morgenstern (FGM) copula for $c_2(\cdot)$, which enables a closed form solution for the relevant integrals. The population mean function is then straightforwardly derived as $\mu(\pi_*, \gamma_{2*}) := \frac{1}{3}\pi_*\gamma_{2*}$, leading to a simple linear 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:

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

Our simulations are conducted in the following manner. First, we generate random samples using the FGM copula with marginals $N(0, 1)$ 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 maximum likelihood 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 μ and variance σ^2, and $(\widehat{\mu}_{x,n}, \widehat{\sigma}^2_{x,n})$ and $(\widehat{\mu}_{y,n}, \widehat{\sigma}^2_{y,n})$ are the corresponding maximum likelihood estimates obtained from x and y samples. Second, we let the copula parameter γ_{2*} be fixed at 0.9, and the adjunct probability measure is assumed to be uniformly distributed on $\Gamma := [0, 1]$. With this framework, we conduct independent experiments with 5,000 replications using data samples with $n =$ 25, 50, 100, 300, 500, and $1,000$.

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

Table 5.3 reports the empirical rejection rates, giving size and power of our tests for dependence. As before, the rejection rates given in the size panel are computed with fixed nominal levels of 1%, 5%, and 10%. Table 5.3 suggests that when the null hypothesis is true, the rejection rates approach nominal levels as the sample size increases. More specifically, the rejection rates of the Wald and LM test statistics are close to the nominal levels even with sample sizes

as small as $n = 50$ or 100, whereas those of the QLR test are somewhat oversized at these nominal levels and sample sizes. For power analysis we let π_* be 0.1, 0.2, 0.3, 0.4, and 0.5 and fix the nominal level of significance to 5%. A clear tendency for the rejection rates to rise with π_* is apparent; and when π_* is constant, the rejection rates of the three test statistics all approach unity as the sample size increases.

6 Empirical Analysis on the Income Path

Lifetime earning trajectories have attracted great interest among labor economists, leading to the early earnings function pioneered by Mincer (1958, 1974), which revealed that earnings typically rise at a diminishing rate over a lifetime, justifying the use a quadratic form over work experience years in regression specifications. Since then, quadratic specifications with respect to years of work experience have been a popular component of empirical wage equations in the literature (e.g., Bhuller, 2017; Barth, Davis, and Freeman, 2018; Magnac, Pistolesi, and Roux, 2018 for recent studies). On the other hand, Murphy and Welch (1990) explored cubic and quartic specifications for wage equations, showing that a restricted quartic specification provided smaller MSE compared to quadratic specifications. Later, Katz and Murphy (1992), Autor, Katz, and Krueger (1998), and Lemieux (2006) adopted quartic specifications in their empirical work. Cho and Phillips (2018a) examined functional form specifications of the wage equation with respect to work experience years using sequential testing and found that functional form can be sensitive to the presence of other explanatory variables in the regression. In another study, Heckman, Lochner, and Todd (2006) showed that the quadratic model is empirically misspecified for recent wage data, but that relaxation of the quadratic model to a quartic specification does not dramatically change the empirical economic implications of the quadratic model.

This section applies the techniques of the current work to study the mean log income path (MLIP) as a function of work experience. Using several different formulations for the MLIP, we examine whether the overall shape of the MLIP differs in significant ways between genders and amongst education levels. In cases where the difference is significant, we further study how gender and education affect the MLIP, providing some empirical insights to enhance understanding of how different income profiles arise according to gender and education.

Our analysis differs from existing work based upon the estimation of Mincer (1958, 1974) wage equations in two ways. First, our approach uses functional observations that have never before, to our knowledge, been considered in this literature. Second, although we estimate the MLIP to identify gender and education effects in parallel to methodological developments in the wage equation literature, our empirical results here are obtained by rephrasing the methodology into a form suited to functional data to single out how gender and education influence the overall shape of the MLIP. As it turns out, shape is not affected by gender and educational differences if each individual's income path is properly scaled by the individual's integrated log income path (LIP) over the work experience years, implying that different gender and education levels lead to different income paths for individuals, but individual LIPs are proportional to each other between genders and education levels. In short, appropriately scaled LIPs shown to be unaffected by different gender and education levels.

For the empirical analysis we analyze income data obtained from the Continuous Work History Sample (CWHS) database that provides the income variable as annual labor income before taxes. The data include 39 years of income records of full time white male and female workers in the U.S. who were born between 1960 and 1962 and had tax records for at least 39 years. Based on these observations we construct each individual's LIP using the local polynomial kernel (Zhang and Chen, 2007) defined over work experience years from 0 to 40 years and then subdivide the entire sample into different groups based on individual gender and education level. For the latter the subdivisions are no college education, Bachelor's degree, Master's degree and Doctoral degree. For the education level groups (in the order given above), we have 673, 2,828, 539, and 323 income path samples for males and 837, 1,624, 469, and 418 income path samples for females.

We conduct our empirical analysis by controlling for worker job mobility. Earlier literature has noted that Mincer's (1958, 1974) quadratic equation gives a good local approximation of the wage equation with respect to the work experience years. For example, Light and Ureta (1995) observe that the empirical wage profiles are more heterogenous in early career experience than those of workers reaching ages in the forties and fifties, in addition to their more rapidly increasing wage levels during this period of the life cycle. This phenomenon in the data is explained through more frequent job interruptions over the early career experience years, particularly for female workers. The finding is used to link gender wage gaps to the work experience variable. In addition to the work of Light and Ureta (1995), many other studies point out that wage profiles during early career years differ considerably from those of generations with longer work experience. Geographic changes and job mobility dominate among the young generation and young workers may have different perspectives on lifetime wage profiles from older workers when firms bargain over wage-employment packages with labor unions (e.g. Mincer and Jovanovic, 1981; Huizinga, 1990). These characteristics of the labor market motivate the treatment of individual income profiles as a composite of many different income profiles.

We, therefore, separately examine the income profiles defined over the entire working lifetime to profiles based on more mature work experience years. Specifically, we proceed by first focusing on the income paths over the full 0 – 40 work experience years and examine how different gender and education levels affect the income paths. Next, we consider income paths over the 10 – 40 years work experience years and examine gender and education effects on these paths. In doing so, the first 10 years of the income paths are removed from the original paths to accommodate the arguments made by Mincer and Jovanovic (1981). These authors found empirical evidence from the NLS and MID panel data that differences in job mobility during the first 10 years of work experience do not predict long-run differences in earnings, implying that income paths during the first 10 years are likely generated by a mechanism that differs from that determining income paths in mature career experience years. As will become clear, we use this empirical separation as a platform to highlight different gender and education effects on income paths.

6.1 Inference on the Mean of Log Income Path on the Whole Work Experience Years

This section reports estimates of the MLIP over the entire working lifetime based on quadratic, cubic, and quartic models. We also estimate the restricted model posited by Murphy and Welch (1990), compare the results, and discuss inferences on the MLIP that are implied by these estimates.

<<<<<< Insert Figure 1. >>>>>

Figure 1 shows the estimated shapes of the MLIPs with respect to work experience over 0 – 40 years for groups classified according to gender and education levels, which are implied by the quadratic, cubic, and quartic models. The red lines in the Figure denote pointwise MLIPs and the dashed lines present the 80% bootstrap confidence bands of the pointwise MLIPs. Along with the pointwise MLIPs are shown the MLIP curves corresponding to the quadratic, cubic, and quartic models. The quartic model, for example, is specified for each group as

$$\rho(\gamma, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = \theta_1 + \theta_2\gamma + \theta_3\gamma^2 + \theta_4\gamma^3 + \theta_5\gamma^4.$$

After estimating the unknown parameters by FLS the fitted curves $\rho(\cdot, \widehat{\theta}_{1n}, \widehat{\theta}_{2n}, \widehat{\theta}_{3n}, \widehat{\theta}_{4n}, \widehat{\theta}_{5n})$ are shown for the entire working lifetime in the Figure. Although all estimated MLIPs lie inside the 80% confidence bands of the pointwise MLIPs, the fitted MLIP curves are evidently different from the pointwise MLIPs, implying that it is difficult to estimate the MLIPs by quadratic, cubic, and quartic model specification uniformly over the full range of years of work experience. In particular, the pointwise MLIPs over the first 10 work experience years are clearly different from those implied by the model estimates. This may indicate, for example, that high job mobility during the first decade of a working lifetime produces income profiles different from those over the remaining years of work experience, corroborating the argument of Mincer and Jovanovic (1981). In addition to the MLIPs implied by the model estimates, we also estimate the mean function implied by the restricted quartic model given by equation (18) in Murphy and Welch (1990):

$$\rho(\gamma, \theta_1, \theta_2, \theta_3, \alpha) = \theta_1 + \theta_2\gamma + (\theta_3 + \theta_2\alpha)\gamma^2 + 2\alpha\theta_3\gamma^3 + \alpha^2\theta_3\gamma^4 \tag{8}$$

that is commonly estimated in the empirical literature. Henceforth we denote this model as the quartic(r) model.

<<<<<< Insert Table 6.4. >>>>>

The top panel of Table 6.4 reports the estimated FMSEs obtained by the quadratic, cubic, quartic, and restricted quartic models in parallel to Tables 2, 6, and 8 in Murphy and Welch (1990). As expected, the quartic specification provides the smallest FMSE and the quadratic specification yields the largest FMSE among the three specifications. We further observe that the FMSE substantially drops as the degree of the polynomial model increases, matching the improved fit of the MLIPs. The mean paths implied by the cubic and quartic functions are found to be statistically distinct.

The lower panel of Table 6.4 reports the FMSEs using the data obtained by scaling the original income paths. That is, these functional observations are obtained by dividing each individual original LIP with by integral of the corresponding LIP over the entire working lifetime. The MLIPs are obtained using the same methodology as before for each group, and the corresponding FMSEs are shown in the lower panel of the Table. This normalization helps to exclude possible absolute effects of the level of lifetime income on inference concerning the MLIPs. Nonetheless, as the results in Table 6.4 show, the outcomes are qualitatively the same as for the top panel, so that the higher the degree of the polynomial model, the better the approximation. Scaling does not affect this outcome.

Although unreported here, we tested model adequacy by comparing the quadratic versus cubic models and the cubic versus quartic models. The QLR test statistic was employed in these comparisons using the FMSEs in Table 6.4. The test trivially rejected the null hypothesis for every group classified by gender and education levels. Given the statistically substantial drops of the FMSEs as the degree of polynomial model increases, this rejection of the null hypothesis is expected and again reveals that higher-degree polynomial functions better approximate the MLIPs than lower-degree functions.

We next examine the gender and education effects on the MLIP. For this purpose, we extend the previous model structures using a dummy variable for the gender of each individual. Thus, for each education level, the model for the curves becomes

$$\rho_i(\gamma, \theta_1^F, \theta_2^F, \theta_3^F, \theta_1^M, \theta_2^M, \theta_3^M) = (\theta_1^F + \theta_2^F\gamma + \theta_3^F\gamma^2)d_i + (\theta_1^M + \theta_2^M\gamma + \theta_3^M\gamma^2)(1 - d_i) \tag{9}$$

extending the previous quadratic model using $d_i = 1$ for female gender and $d_i = 0$ otherwise. We estimate the unknown parameters by FLS and test whether the corresponding parameters are equal across genders using the Wald, LM and QLR test statistics. So the hypotheses of interest are

$$\mathbb{H}_o : \theta_{1*}^F = \theta_{1*}^M,\ \theta_{2*}^F = \theta_{2*}^M,\ \text{and}\ \theta_{3*}^F = \theta_{3*}^F, \quad \text{versus} \quad \mathbb{H}_a : \theta_{1*}^F \neq \theta_{1*}^M,\ \theta_{2*}^F \neq \theta_{2*}^M,\ \text{or}\ \theta_{3*}^F \neq \theta_{3*}^F$$

and failure to reject $\mathbb{H}_o$ provides evidence that the MLIPs do not differ between genders. Similar extensions were made and tests conducted for the cubic, quartic, and quartic(r) models.

<<<<<< Insert Table 6.5. >>>>>

The results are shown in Table 6.5, which provides empirical evidence for gender effects on the MLIPs. The top panel of Table 6.5 reports test results using the original LIP samples. For the groups with college education (Bachelor, Master, Ph.D), the null hypothesis of mean equivalence is rejected at both 1% and 5% levels, indicating that the average income paths are very different across different genders. This finding is consistent with empirical results in the literature. For example, Wiswall and Zafar (2017) find evidence of a gender effect on income paths and associate this difference with differing job demands between different genders. More specifically, they suggest that job flexibility and job stability may be more important factors for women in job choice, whereas men place a relatively higher preference on earnings, thereby producing a gender effect. These findings are nuanced for the group without college education. There the difference in the mean functions is significant for the LM test statistic, but the QLR test does not reject the null hypothesis of mean equality at the 1% and 5% levels. At the same levels, the Wald test rejects the null for the cubic and quartic specifications but not for the quadratic specification. Notwithstanding this outcome, overall it is evident that the gender effect is dominant for each education level and model specification.

In the lower panel of Table 6.5, we report results for the scaled LIP samples. The overall results from the top panel of Table 6.5 are evident also in the scaled samples. Thus, the gender effect is evident for the MLIPs, although it is less significant with the QLR statistic in the no-college education group and the differences in the MLIPs are less significant in Master and Ph.D groups, suggesting that the MLIPs become less differentiated across gender as educational qualifications rise to these levels.

Next, we examine the education effect on the MLIP across different education levels within the same gender. A pairwise comparison is made between the groups without college education and with a Bachelor degree and similar pairwise comparisons are made for those between the Bachelor and Masters level education, and Masters and Doctoral degrees.

<<<<<< Insert Table 6.6. >>>>>

Table 6.6 reports the associated test outcomes. In the top panel of Table 6.6 we report all the results from the original LIP samples. At the 1% and 5% significance levels, the differences in the MLIPs turn out to be highly significant across different education levels for both male and female workers, implying that education levels have a big impact on the MLIPs, just as Mincer's (1958, 1974) wage equation implies. Since education levels are captured by dummy variables, we cannot be assured of linearity in the income path with respect to education years as Mincer's (1958, 1974) wage equation implies. Nevertheless, our finding is consistent with that wage equation.

The lower panel of Table 6.6 reports results using the scaled LIP samples. The overall education effect is evident for the scaled LIP samples just as before. But some qualitatively different results from the original samples are also apparent. In particular, in the male group, the mean difference is significant in the comparison between groups with no-college education and a Bachelor degree, but this difference diminishes for comparisons between groups with higher education at Masters and Doctoral levels. Thus, as the education level rises, the effects of education on the MLIP diminish for male workers. On the other hand, the female group comparisons always yield p-values close to zero in all cases, implying that the female worker LIP is strictly affected by education at all levels and undiminished at the higher levels compared with male workers.

In sum, gender and education effects on the MLIP are evident in the LIP samples over the entire working lifetime. Different genders and education levels yield different MLIPs irrespective of whether each individual's LIP sample is scaled by aggregating over the full working cycle. Although there are some nuanced differences for different genders and education levels, the overall gender and education effects are evident in both cases.

6.2 Inference on the Mean of Log Income Path on the Post Work Experience Years

In this section we estimate the MLIP curves over the mature work experience years using the quadratic, cubic, quartic, and quartic(r) models as in Section 6.1. When the LIP samples are collected from income profiles with low job mobility, different gender and education effects are to be expected for the MLIP, as Mincer and Jovanovic (1981) point out.

<<<<<< Insert Figure 2. >>>>>

Using the same samples and the same methodology as before, we estimate the shapes of the MLIPs on the work experience years from 10 to 40 years. For the same groups classified by gender and education levels, Figure 2 displays the estimated MLIPs implied by the quadratic, cubic, and quartic models along the pointwise estimated MLIP. As seen in Figure 2, the MLIPs implied by the cubic and quartic models are close to the overall pointwise MLIP. On the other hand, the quadratic model yields almost a linear MLIP over the mature work experience years and relatively large differences exist between the pointwise MLIP and the MLIP implied by the quadratic model.

<<<<<< Insert Table 6.7. >>>>>

As before we proceed in parallel to Section 6.1 by first reporting in Table 6.7 the FMSEs obtained from the quadratic, cubic, quartic, and quartic(r) models. The format of Table 6.7 is the same as that of Table 6.4, but the estimated FMSE results exhibit different patterns from those in Table 6.4. For example, the top panel of Table 6.7 shows that the FMSE drops substantially as polynomial degree increases from quadratic to cubic for both genders and for all levels of education; but the FMSEs are more or less similar among the cubic, quartic, and quartic(r) models. So, extending model specification from cubic to quartic has little effect on model fit compared with increasing polynomial degree from quadratic to cubic. This outcome is concordant with Figure 2, in which the MLIP implied by the quadratic model differs from the pointwise MLIP and the MLIPs implied by the cubic and quartic models.

A similar pattern is observed in the bottom panel of Table 6.7, where FMSEs obtained from the scaled LIP samples are obtained. The scaled LIP paths are constructed differently from those in Section 6.1. Instead of dividing each individual's LIP with integrated LIP over lifetime work experience years, individual LIP is scaled by integrated LIP over mature work experience years. But irrespective of whether or not the LIPs are scaled, the same pattern is observed for the FMEs as in the top panel of Table 6.7.

Notwithstanding these results, it does not follow that higher than quadratic polynomial degrees are unhelpful in reducing the FMSEs. Although we do not report the results here, QLR tests of the quadratic specification versus the cubic specification and cubic versus quartic were conducted and these tests all rejected the null as before, implying that higher degree specifications deliver statistically more satisfactory approximations for the MLIPs.

<<<<<< Insert Table 6.8. >>>>>

We next examine gender effects on the MLIPs over mature work experience years. As before, we apply Wald, LM, and QLR test statistics and report test outcomes in the top panel of Table 6.8 in parallel with Table 6.5. As seen in the top panel of Table 6.8, gender effects on the MLIP are evident in the groups with college educations. In particular, with each of the four model specifications, the hypothesis of constant MLIP between gender is rejected for groups with college education. In contrast, for the no-college education group it is difficult to reject constancy of the MLIP curves between gender, implying that female and male workers with low education levels face almost the same income profiles.

In the lower panel of Table 6.8, we report results using the scaled LIP samples. The figures in the table are obtained by conducting the same inferential procedures as those for the top panel. As is apparent in the table, no strong statistical evidence emerges indicating gender effects on the scaled MLIPs. Thus, if the LIP is associated with low job mobility, individual LIPs differ proportionately between gender so that gender influences are eliminated if proportionality is properly accounted for. This finding differs from that obtained from the LIPs over lifetime work experience years. If each individual's LIP is scaled by integrated LIP over the lifetime work years, the scaled LIP *is* affected by income profiles with high job mobility, leading to different MLIPs between different genders even when the LIPs are restricted to the mature work experience years.

<<<<<< Insert Table 6.9. >>>>>>

We close these empirics with an exploration of the effects of education within the same gender. Pairwise comparison results are reported in Table 6.9 in the same manner as before. The top panel gives test outcomes in parallel to those of Table 6.6. Using the original LIP samples, all test statistics produce qualitatively the same conclusions as those the top panel of Table 6.6, so that education effects on the MLIP are found to be present even after income profiles are constructed from workers with low job mobility.

However, when individual LIPs are scaled by integrated mature work experience levels, different results are obtained. The test outcomes are shown in the lower panel of Table 6.9. Just as in Table 6.8, no strong statistical evidence is found supporting different MLIPs among different education levels, giving the same conclusion as for the gender effect so that for the scaled LIPs education effects disappear within the same gender. Thus, if income profiles are generated from workers with low job mobility, individual LIPs differ proportionally among different education levels. So when these proportional differences are properly accounted for in the LIP samples, the MLIPs match across education levels.

In summary, if the LIP samples are constructed from workers with low job mobility, the empirical analysis produces different conclustions depending on whether the LIP is scaled or not. If the LIP samples are unscaled, results are overall similar to those obtained from the LIPs over the entire lifetime work experience years, so that the MLIPs obtained from these LIP samples do differ between gender and across education levels. But when the LIPs are scaled by respective integrated mature work experience years, the estimated MLIPs match, thereby implying that different gender and education levels affect the LIPs proportionally without changing the overall shape of the MLIP.

7 Conclusion

We develop a methodology of estimation and inference for parametric conditional mean functions that involve functional data. The approach is based on functional least squares (FLS) regression. Consistency and asymptotic normality are established under regularity conditions that allow for the possible presence of nuisance parameters and consequential effects on the limit theory. New Wald, Lagrange multiplier (LM), and quasi-likelihood ratio (QLR) tests are developed for this functional data context that enable inference about curve shapes in the observed data. Various examples where this methodology is useful include inference about the form of random coefficient and distributional specification in mixtures and copulas. The empirical functional form of income paths over work experience years is studied using these methods to examine the mean log income path within the framework of Mincer's (1958, 1974) wage equation. The findings show that gender and education levels produce differences in mean income paths but that these mean paths are proportional to each other. It follows that upon rescaling the income paths by integrated work experience years associated with low job mobility, the mean income paths match over gender and across education levels.

The present study suggests potentially fruitful further developments in functional data inference. This paper assumes that explanatory variables are observed as vector valued whereas the dependent variable is observed in the form of random functions. Other cases of interest arise where both dependent and explanatory variables may take function space form or where the explanatory variable involves functional data and the dependent variable is scalar valued. See the model analyses in James, Wang, and Zhu (2009), Crambes, Gannoun, and Henchiri (2013), Petersen and Müller (2016), and Happ and Greven (2018) for some recent examples. Also, when different estimation methods than FLS are employed, new possibilities for parametric inference become possible. For instance, using penalized profile likelihood function estimation with functional data makes it possible to identify latent structures in functional curve data in a similar fashion to the classification Lasso method of Su, Shi, and Phillips (2016).

Further extensions of the present methods can be made to dependent data. Recently, functional time series analysis has attracted much attention in the literature. For instance, functional data such as the distribution of cross-sectional earnings or near-continuous recording of intraday stock returns potentially lead to autocorrelated dependence structures. Various methods of using such data can be found in Chang, Kim, and Park (2016), Kim and Park (2017), Beare, Seo, and Seo (2017), Hörmann, Kokoszka, and Nisol (2018), Seo and Beare (2019), Li, Robinson, and Shang (2019), and Chang, Hu, and Park (2019). From the perspective of parametric conditional mean function estimation, the presence of temporal dependence inevitably leads to limit theory for FLS estimation and statistical tests that differs from the present study and needs to be accounted for in time series inference. These are some of the many aspects of conditional mean function inference that can be addressed in future research.

Appendices

These Appendices are organized as follows. Preliminary lemmas and proofs are given in Appendix A. Proofs of the results in the main text are given in Appendix B. The final Appendix C provides additional discussion and results on the estimation of the unconditional mean function.

A Preliminary Lemmas

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

Lemma 2. *Given 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*

$$\nabla_\theta\int h(\gamma,\theta)d\mathbb{Q}(\gamma)=\int\nabla_\theta h(\gamma,\theta)d\mathbb{Q}(\gamma),\tag{10}$$

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

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

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

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

Lemma 4. *Given Assumptions 1, 2, 3, and 5, then $B = \widetilde{B}$, 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 $\varepsilon(\gamma, \theta) := g(\gamma) - \rho(\gamma, \theta, x)$. $\square$

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.$-\mathbb{P}$.* $\square$

Proof of Lemma 2: From the differentiability condition, the given function is Lipschitz continuous, so that $|h(\gamma, \theta) - h(\gamma, \theta')| \leq m(\gamma)\|\theta - \theta'\|$, 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, the following bound holds

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

which 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 the dominated convergence theorem (DCT). The left and right sides of this equality are respectively identical to the left and right sides of (10). This completes the proof. $\blacksquare$

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 $|g_i(\,\cdot\,)| \leq m_i$ and $\sup_{(\gamma,\theta)}|(\partial/\partial\theta_j)\rho_i(\gamma, \theta)| \leq m_i$ 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 the Cauchy-Schwarz's inequality. Therefore, applying Lemma 2 yields

$$\nabla_\theta \int\int \mu(\gamma,x)\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) = \int\int \mu(\gamma,x)\nabla_\theta\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma). \tag{11}$$

Furthermore, $\sup_{(\gamma,\theta)} |\rho_i(\gamma,\theta)| \leq m_i$ 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.$-\,\mathbb{P}$, again by Cauchy-Schwarz. Applying Lemma 2 entails that

$$\nabla_\theta \int\int \rho^2(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) = 2\int\int \rho(\gamma,\theta,x)\nabla_\theta\rho(\gamma,\theta,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma). \tag{12}$$

The desired result follows by combining (11) and (12).

(*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. Applying Lemma 2 leads to the desired result as for (*i*).

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

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

Furthermore, for each j, the Cauchy-Schwarz 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 $\sup_\gamma |g_i(\gamma)| \leq m_i \in L^2(\mathbb{P})$ as shown in (*i* and *ii*) using Assumption 3. Hence, applying Lemma 2 leads to

$$\nabla_\theta \frac{1}{n}\sum_{i=1}^n \int g_i(\gamma)\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma) = \frac{1}{n}\sum_{i=1}^n \int g_i(\gamma)\nabla_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma) \tag{13}$$

a.s.$-\,\mathbb{P}$. Finally, combining (12) and (13) 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 and $\sup_{\theta\in\Theta} |(\partial/\partial\theta_j)\rho_i(\,\cdot\,,\theta)| \in L^2(\mathbb{Q})$ as shown in (*ii*). Further, 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.$-\,\mathbb{P}$ again by applying Cauchy-Schwarz 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,

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

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

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

using the fact that $\int\{g(\gamma)-\mu(\gamma,x)\}\{\mu(\widetilde{\gamma},x)-\rho(\widetilde{\gamma},\theta_*,x)\}d\mathbb{P}(g(\gamma),g(\widetilde{\gamma})|x)=0$. Hence

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

Now note that by definition in Assumption 5(ii), $B := \int\int\int \nabla_\theta\rho(\gamma,\theta_*,x)\kappa(\gamma,\widetilde{\gamma}|x)\nabla_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$ 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,

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

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

$$\nabla_\theta\frac{1}{n}\sum_{i=1}^{n}\int \widehat{g}_i(\gamma)\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma) = \frac{1}{n}\sum_{i=1}^{n}\int \widehat{g}_i(\gamma)\nabla_\theta\rho_i(\gamma,\theta)d\mathbb{Q}(\gamma) \tag{14}$$

a.s.$-\mathbb{P}$. Finally, combining (12) and (14) 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, thereby completing the proof. ■

B Proofs of the Main Results

Proof of Theorem 1: Note that

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

Further, $\int\int\{g(\gamma)-\mu(\gamma,x)\}^2 d\mathbb{P}d\mathbb{Q}(\gamma) = \int \mathrm{var}_{\mathbb{P}}[g_i(\gamma)|x]d\mathbb{P}(x)d\mathbb{Q}$, 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 now follows. ■

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

$$q_n(\theta) = \frac{1}{n}\sum_{i=1}^n\int g_i^2(\gamma)d\mathbb{Q}(\gamma) - \frac{2}{n}\sum_{i=1}^n\int\left(g_i(\gamma)\rho_i(\gamma,\theta)\right)d\mathbb{Q} + \frac{1}{n}\sum_{i=1}^n\int\rho_i^2(\gamma,\theta)d\mathbb{Q} \quad \text{and} \tag{15}$$

$$q(\theta) = \int E_{\mathbb{P}}[g_i^2(\gamma)]d\mathbb{Q}(\gamma) - 2\int\mathbb{E}_{\mathbb{P}}[\mu_i(\gamma)\rho_i(\gamma,\theta)]d\mathbb{Q} + \int\mathbb{E}_{\mathbb{P}}[\rho_i^2(\gamma,\theta)]d\mathbb{Q}. \tag{16}$$

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 (16), and we can interchange the integral and sample average operators in computing (15) by virtue of the DCT, as shown in the proof of Lemma 3.

We now examine the limit of each element in the right sides of (15) and (16). 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 the DCT, giving

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

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

$$\sup_{\theta\in\Theta}\int\left|\left(\frac{1}{n}\sum_{i=1}^{n} g_i(\gamma)\rho_i(\gamma,\theta) - \mathbb{E}[\mu_i(\gamma)\rho_i(\gamma,\theta)]\right)\right| d\mathbb{Q}(\gamma)$$
$$\leq \int\sup_{\theta\in\Theta}\left|\left(\frac{1}{n}\sum_{i=1}^{n} g_i(\gamma)\rho_i(\gamma,\theta) - \mathbb{E}[\mu_i(\gamma)\rho_i(\gamma,\theta)]\right)\right| d\mathbb{Q}(\gamma) \to 0 \tag{17}$$

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

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,

$$\sup_{\theta\in\Theta}\int\left|\left(\frac{1}{n}\sum_{i=1}^{n}\rho_i^2(\gamma,\theta) - \mathbb{E}[\rho_i^2(\gamma,\theta)]\right)\right| d\mathbb{Q}(\gamma) \leq \int\sup_{\theta\in\Theta}\left|\left(\frac{1}{n}\sum_{i=1}^{n}\rho_i^2(\gamma,\theta) - \mathbb{E}[\rho_i^2(\gamma,\theta)]\right)\right| d\mathbb{Q}(\gamma) \to 0 \tag{18}$$

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

Finally, combining the above three facts gives

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

as desired.

(*ii*) This result 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

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

so that

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

where

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

We now examine each element in (19), starting with the matrix A_n and writing

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

First, as already seen in the proof of Lemma 3(*iv*), 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 and interchanging the integral and sample average operators in the limit by the DCT yield

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

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 and the DCT yield

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

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, in the same manner

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

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

Next consider the right side of (19). By virtue of Assumption 1, we can apply the multivariate Lindeberg CLT giving

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int[g_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\overset{\text{A}}{\sim}N(0,B)\tag{20}$$

because the common mean of the components is $\int\int\{g(\gamma)-\rho(\gamma,\theta_*,x)\}\nabla_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(g(\gamma),x)d\mathbb{Q}(\gamma)=\int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}\,(\gamma)=\nabla_\theta q(\theta_*)=0$ by Lemma 3(*i*) and for each j and j', $\int\int\int(\partial/\partial\theta_j)\rho(\gamma,\theta_*,x)\cdot\kappa(\gamma,\widetilde{\gamma}|x)\cdot(\partial/\partial\theta_{j'})\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}\,(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})<\infty$, leading to a common positive definite variance matrix B by Assumption 5. From these properties and the IID condition of Assumption 1(ii), the right side of (19) 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 DCT. From the definitions of $\widehat{q}_n(\,\cdot\,)$ and $q(\,\cdot\,)$, for each θ,

$$\widehat{q}_n(\theta)=\int\frac{1}{n}\sum_{i=1}^{n}\widehat{g}_i^2(\gamma)d\mathbb{Q}(\gamma)-2\int\frac{1}{n}\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)\rho_i(\gamma,\theta)\}\,d\mathbb{Q}+\int\frac{1}{n}\sum_{i=1}^{n}\rho_i^2(\gamma,\theta)d\mathbb{Q},\tag{21}$$

interchanging the integral and finite sample averaging operators. We compare this expression with each element on the right side of (16). First, Assumption 7(i) implies that

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

so that appliction of the DCT gives

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

because: (a)

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

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

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

Thus, 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). The desired result follows from properties (a), (b), and (c).

Next, compare the second elements in (21) and (16). First,

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

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, (18) 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 (21) is identical to the third element in the right side of (16). The desired result follows from these three properties.

(*ii*) This result 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

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

and then

$$\widehat{A}_n\sqrt{n}(\widetilde{\theta}_n-\theta_*)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\int[\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta_*)]\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\right\},\tag{22}$$

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 (22). 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 the DCT.

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, Assumption 7(*i*) and 4(*i*). Therefore,

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

by the DCT 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 the DCT imply that

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

These results give $\widehat{A}_n \to A$ a.s.$-\,\mathbb{P}$.

Next, we examine the right side of (22). Applying the mean-value theorem, we obtain the representation given in (4), viz., that for some $\bar{\xi}_{n,\gamma}$ between $\widehat{\xi}_n$ and ξ_*,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\int\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)\\ &=\frac{1}{\sqrt{n}}\int\sum_{i=1}^{n}\{g_i(\gamma,\xi_*)-\rho_i(\gamma,\theta_*)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)+\frac{1}{n}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\theta_*)\cdot\nabla'_\xi g_i(\gamma,\bar{\xi}_{n,\gamma})d\mathbb{Q}\cdot\sqrt{n}(\widehat{\xi}_n-\xi_*).\end{aligned} \tag{23}$$

From the proof of Theorem 3 we have

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

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 DCT 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*) and 9,

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

Therefore,

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

Furthermore, 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 $n^{-1/2}\sum_{i=1}^{n}\int\nabla_\theta\rho_i(\gamma,\theta_*)\cdot\nabla'_\xi g_i(\gamma,\bar{\xi}_{n,\gamma})d\mathbb{Q}(\widehat{\xi}_n-\xi_*)$ is obtained as $-MH^{-1}K$ by Assumption 9, implying that the asymptotic covariance matrix of $n^{-1/2}\int\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)-\mu_i(\gamma)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma)$ is given by $B_* := B - MH^{-1}K - K'H^{-1}M' + MH^{-1}JH^{-1\prime}M'$, which is positive definite by Assumption 9. We therefore obtain

$$\frac{1}{\sqrt{n}}\int\sum_{i=1}^{n}\{\widehat{g}_i(\gamma)-\mu_i(\gamma)\}\nabla_\theta\rho_i(\gamma,\theta_*)d\mathbb{Q}(\gamma) \overset{\mathrm{A}}{\sim} N(0,B_*). \tag{24}$$

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 (22) and (24). This completes the proof. ■

Proof of Theorem 6: We first consider the consistency 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 DCT, it follows that

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

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

$$\frac{1}{n}\sum_{i=1}^{n}\int\varepsilon_i(\gamma,\widehat{\theta}_n)\cdot\nabla^2_\theta\rho_i(\gamma,\widehat{\theta}_n)d\mathbb{Q}(\gamma) \to \int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla^2_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma) \;\; a.s. - \mathbb{P}. \tag{26}$$

Here, we used the fact that $\int g(\gamma)d\mathbb{P}(g(\gamma)|x) = \mu(\gamma,x)$. Now, (25) and (26) imply that $\widehat{A}_n \to A \;\; a.s. - \mathbb{P}$.

We next examine the consistency 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 DCT, 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, $\widehat{B}_n \to \int\int \nabla_\theta\rho(\gamma,\theta_*,x)\kappa(\gamma,\widetilde{\gamma}|x)\nabla'_\theta\rho(\widetilde{\gamma},\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})\;\; a.s. - \mathbb{P}$, corresponding to the definition of B. This completes the proof. ■

Proof of Theorem 7: We start by showing consistency of $\widetilde{A}_n$. First, (25) implies that

$$\frac{1}{n}\sum_{i=1}^{n}\int \nabla_\theta\rho_i(\gamma,\widetilde{\theta}_n)\nabla'_\theta\rho_i(\gamma,\widetilde{\theta}_n)d\mathbb{Q}(\gamma) \to \int\int\nabla_\theta\rho(\gamma,\theta_*,x)\nabla'_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma). \tag{27}$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\int\widetilde{\varepsilon}_i(\gamma,\widetilde{\theta}_n)\cdot\nabla^2_\theta\rho_i(\gamma,\widetilde{\theta}_n) \to \int\int\{\mu(\gamma,x)-\rho(\gamma,\theta_*,x)\}\nabla^2_\theta\rho(\gamma,\theta_*,x)d\mathbb{P}(x)d\mathbb{Q}(\gamma)\;\; a.s. - \mathbb{P} \tag{28}$$

noting that $\int g(\gamma)d\mathbb{P}(g(\gamma)|x) = \mu(\gamma,x)$. Now, (27) and (28) imply that $\widetilde{A}_n \to A$ a.s.$-\,\mathbb{P}$.

We next show the consistency 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.$-\,\mathbb{P}$, then the consistency of $\widetilde{B}_n$ follows from Assumption 11.

(*i*) Proving that $\bar{B}_n \to B$ a.s.$-\,\mathbb{P}$ is almost identical to proving that of $\widehat{B}_n \to B$. Note that for each j and $j' = 1,2,\ldots,d$, $\sup_{(\gamma,\widetilde{\gamma},\theta,\xi)\in\Gamma\times\Gamma\Theta}|(\partial/\partial\theta_j)\rho_i(\gamma,\theta)\varepsilon_i(\gamma,\theta,\xi)\varepsilon_i(\widetilde{\gamma},\theta,\xi)(\partial/\partial\theta_{j'})\rho_i(\widetilde{\gamma},\theta)| \le m_i^2 \in L^1(\mathbb{P})$ 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 DCT, 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 have already seen that the right side is identical to B. Therefore, $\bar{B}_n\to B$ a.s.$-\,\mathbb{P}$.

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

$$\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}\left|\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta\rho_i(\gamma,\theta)\nabla_\xi'\widetilde{g}_i(\gamma,\xi)-\mathbb{E}_\mathbb{P}[\nabla_\theta\rho_i(\gamma,\theta)\nabla_\xi'\widetilde{g}_i(\gamma,\xi)]\right|\to 0\;\; a.s.-\mathbb{P} \tag{29}$$

by the SULLN. Therefore, applying the DCT 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*) Next we show $\widehat{K}_n\to K$ a.s.$-\,\mathbb{P}$. Note that $\sup_{(\gamma,\theta,\xi)\in\Gamma\times\Theta\times\Xi}|s_i(\xi)\{\widetilde{g}_i(\gamma,\xi)-\rho_i(\gamma,\theta\}\nabla_\theta'\rho_i(\gamma,\theta)|\le m_i^2\in L^2(\mathbb{P})$ 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 DCT implies that

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

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

The consistency of $\widetilde{B}_n$ for B_* now follows as a consequence of (*i*, *ii*, and *iii*) in 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 converge 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 converge to $\theta_\dagger$ a.s.$-\,\mathbb{P}$, as $\theta_\dagger$ is constrained by the same constraint $h(\theta)=0$. The remainder of the proof is identical to the proof of Lemma 1(i). ■

Proof of Theorem 8: (i) By virtue of the mean-value theorem, 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 $\sqrt{n}\{h(\widehat{\theta}_n)-h(\theta_*)\} \overset{\text{A}}{\sim} N(0, D_*A^{-1}BA^{-1}D_*')$. Further, $\widehat{B}_n \to B$ a.s.$-\mathbb{P}$ by Theorem 6, so that $n\{h(\widehat{\theta}_n)-h(\theta_*)\}'\{D_*A^{-1}\widehat{B}_nA^{-1}D_*'\}^{-1}\{h(\widehat{\theta}_n)-h(\theta_*)\} \overset{\text{A}}{\sim} \mathcal{X}^2(r,0)$. Therefore, under $\mathbb{H}_0$, $\mathcal{W}_n^\flat \overset{\text{A}}{\sim} \mathcal{X}^2(r,0)$. 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 proofs are almost identical to Theorem 8(i): under $\mathbb{H}_o$, $\sqrt{n}\{h(\widetilde{\theta}_n)-h(\theta_*)\} \overset{\text{A}}{\sim} N(0, D_*A^{-1}B_*A^{-1}D_*')$ by the mean-value theorem and Theorems 4 and 5; and $\widetilde{B}_n \to B_*$ a.s.$-\mathbb{P}$ by Theorem 7, so that $n\{h(\widetilde{\theta}_n) - h(\theta_*)\}'\{D_*A^{-1}\ \widetilde{B}_nA^{-1}D_*'\}^{-1}\{h(\widetilde{\theta}_n)-h(\theta_*)\} \overset{\text{A}}{\sim} \mathcal{X}^2(r,0)$ under $\mathbb{H}_o$, implying that $\mathcal{W}_n^\sharp \overset{\text{A}}{\sim} \mathcal{X}^2(r,0)$ 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

$$\begin{aligned}\sqrt{n}(\ddot{\theta}_n^\flat-\theta_*) = &-\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}(J-\ddot{D}_n^{\flat\prime}\{\ddot{E}_n^\flat\}^{-1}\ddot{D}_n^\flat\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1})\sqrt{n}\psi_n\\ &-\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}\ddot{D}_n^{\flat\prime}\{\ddot{E}_n^\flat\}^{-1}\sqrt{n}h(\theta_*) + o_{\mathbb{P}}(1) \qquad \text{and}\end{aligned} \tag{30}$$

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

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 result, we 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) The proofs are almost identical to those 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

$$\begin{aligned}\sqrt{n}(\ddot{\theta}_n^\sharp - \theta_*) = &- \{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}(J - \ddot{D}_n^{\sharp\prime}\{\ddot{E}_n^\sharp\}^{-1}\ddot{D}_n^\sharp\{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1})\sqrt{n}\widehat{\psi}_n \\ &- \{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}\ddot{D}_n^{\sharp\prime}\{\ddot{E}_n^\sharp\}^{-1}\sqrt{n}h(\theta_*) + o_{\mathbb{P}}(1) \qquad \text{and} \end{aligned} \tag{31}$$

$$\sqrt{n}\ddot{\lambda}_n^\sharp = \{\ddot{E}_n^\sharp\}^{-1}\ddot{D}_n^\sharp\{\nabla_\theta^2 \widehat{q}_n(\bar{\theta}_n^\sharp)\}^{-1}\sqrt{n}\psi_n + \{\ddot{E}_n^\sharp\}^{-1}\sqrt{n}h(\theta_*), \tag{32}$$

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 (24). 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 $\nabla_\theta q(\theta_\dagger) = \nabla_\theta q(\theta_*) + \nabla_\theta^2 q(\theta_\S)(\theta_\dagger - \theta_*)$, so that $\sqrt{n}\ddot{\lambda}_n^\sharp = O_{\mathbb{P}}(\sqrt{n})$ though not $o_{\mathbb{P}}(\sqrt{n})$. The desired result then follows. ■

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

$$q_n(\ddot{\theta}_n^\flat) = q_n(\widehat{\theta}_n) + \frac{1}{2}(\ddot{\theta}_n^\flat - \widehat{\theta}_n)'\{\nabla_\theta^2 q_n(\widetilde{\theta}_n^\flat)\}(\ddot{\theta}_n^\flat - \widehat{\theta}_n). \tag{33}$$

Furthermore, it follows that

$$\sqrt{n}(\ddot{\theta}_n^\flat - \widehat{\theta}_n) = \{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}\ddot{D}_n^{\flat\prime}\{\ddot{E}_n^\flat\}^{-1}\sqrt{n}[\ddot{D}_n^\flat\{\nabla_\theta^2 q_n(\bar{\theta}_n^\flat)\}^{-1}\psi_n - h(\theta_*)] + o_{\mathbb{P}}(1) \tag{34}$$

from (30) and (19). As given in the proof of Theorem 9(*i*), $\ddot{\theta}_n^\flat \to \theta_*$ a.s.$-\mathbb{P}$, $\widetilde{\theta}_n^\flat \to \theta_*$ a.s.$-\mathbb{P}$ under $\mathbb{H}_o$, 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.$-\mathbb{P}$ and $\nabla_\theta^2 q_n(\widetilde{\theta}_n^\flat) \to 2A$ 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 employ all these results in (33), $n\{q_n(\ddot{\theta}_n^\flat) - q_n(\widehat{\theta}_n)\} \Rightarrow W'(D_*A^{-1}D_*')^{-1}W$ under $\mathbb{H}_o$. Meanwhile, $\sqrt{n}(\ddot{\theta}_n^\flat - \widehat{\theta}_*)$ is not bounded in probability under $\mathbb{H}_a$ mainly because $\sqrt{n}h(\theta_*) = O(\sqrt{n})$ though not $o(\sqrt{n})$ in (34); $\ddot{D}_n^\flat \to D_\dagger := D(\theta_\dagger)$ 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 $\nabla_\theta q(\ddot{\theta}^\flat) = \nabla_\theta q(\theta_*) + \nabla_\theta^2 q(\theta_\S)(\ddot{\theta}^\flat - \theta_*)$. The desired result then follows.

(*ii*) The proofs follow those of Theorem 10(*i*). By the mean-value theorem and the first-order condition for $\widetilde{\theta}_n$, for some $\widetilde{\theta}_n^\sharp$ between $\widetilde{\theta}_n$ and $\ddot{\theta}_n^\sharp$,

$$q_n(\ddot{\theta}_n^\sharp) = q_n(\widetilde{\theta}_n) + \frac{1}{2}(\ddot{\theta}_n^\sharp - \widetilde{\theta}_n)'\{\nabla_\theta^2 q_n(\widetilde{\theta}_n^\sharp)\}(\ddot{\theta}_n^\sharp - \widetilde{\theta}_n). \tag{35}$$

Furthermore, it follows that

$$\sqrt{n}(\ddot{\theta}_n^\sharp - \widetilde{\theta}_n) = \{\nabla_\theta^2 q_n(\bar{\theta}_n^\sharp)\}^{-1}\ddot{D}_n^{\sharp\prime}\{\ddot{E}_n^\sharp\}^{-1}\sqrt{n}[\ddot{D}_n^\sharp\{\nabla_\theta^2 q_n(\bar{\theta}_n^\sharp)\}^{-1}\widehat{\psi}_n - h(\theta_*)] + o_{\mathbb{P}}(1) \tag{36}$$

by (31) and (22). As given in the proof of Theorem 9(*ii*), we have $\ddot{\theta}_n^\sharp \to \theta_*$ a.s.$-\mathbb{P}$, $\widetilde{\theta}_n^\sharp \to \theta_*$ 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 $\nabla_\theta^2 q_n(\widetilde{\theta}_n^\sharp) \to 2A$ 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 use these results in (35), $n\{q_n(\ddot{\theta}_n^\sharp) - q_n(\widetilde{\theta}_n)\} \Rightarrow W_*'(D_*A^{-1}D_*')^{-1}W_*$ under $\mathbb{H}_o$. Meanwhile, $\sqrt{n}(\ddot{\theta}_n^\sharp - \widetilde{\theta}_n)$ is not bounded in probability under $\mathbb{H}_a$ mainly because $\sqrt{n}h(\theta_*) = O(\sqrt{n})$ though not $o(\sqrt{n})$ in (36); $\ddot{D}_n^\sharp \to D_\dagger := D(\theta_\dagger)$ a.s.$-\mathbb{P}$; and $\nabla_\theta^2 q_n(\bar{\theta}_n^\sharp) \to \nabla_\theta^2 q(\theta_\S)$ a.s.$-\mathbb{P}$ for some $\theta_\S$ such that $\nabla_\theta q(\ddot{\theta}^\sharp) = \nabla_\theta q(\theta_*) + \nabla_\theta^2 q(\theta_\S)(\ddot{\theta}^\sharp - \theta_*)$. The desired result follows. ■

C Estimating the Population Mean Function

This section explores estimation of the population mean function of $g_i(\,\cdot\,)$. For each γ the mean quantity

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

is no longer a function of x and for each γ we may denote $\mu(\gamma)$ as $\mathbb{E}[g_i(\gamma)]$. Estimation and inference 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 FLS estimation, i.e.,

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

which again is designed to estimate

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

consistently. Note that this FLS estimator is special case of the analysis in Section 3, so that consistency and asymptotic normality of $\ddot{\theta}_n$ can be achieved under milder conditions than those of Section 3. We first collect the conditions together 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 and lies in the interior of Θ;

(v) A_0 is positive definite, 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 $m_i \in L^2(\mathbb{P})$, $\sup_{(\gamma,\xi)\in\Gamma\times\Xi} |\widetilde{g}_i(\gamma, \xi)| \leq m_i$ a.s. $-\mathbb{P}$ and $\sup_j \sup_{(\gamma,\xi)\in\Gamma\times\Xi} |(\partial/\partial\xi_j)\widetilde{g}_i(\gamma, \xi)| \leq m_i$ a.s. $-\mathbb{P}$;

(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

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

$K_0 := \int \mathbb{E}_\mathbb{P}[s_i(\xi_*)\{\widetilde{g}_i(\gamma, \xi_*) - \rho(\gamma, \ddot{\theta}_*)\}]d\mathbb{Q}(\gamma)$, $B_0 := \int\int \nabla_\theta \rho(\gamma, \ddot{\theta}_*)\kappa(\gamma, \widetilde{\gamma})\nabla'_\theta \rho(\widetilde{\gamma}, \ddot{\theta}_*)d\mathbb{Q}(\gamma)d\mathbb{Q}(\widetilde{\gamma})$, 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) $B_\dagger$ is positive definite, 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*) correspond to Assumption 2 and, 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. Note that the integrands of A_0 are non-stochastic, whereas A has 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 gives the limit behavior 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, $\sqrt{n}(\ddot{\theta}_n - \ddot{\theta}_*) \overset{\text{A}}{\sim} N(A_0^{-1}B_0A_0^{-1})$. □

In view of the parallel structure of $\ddot{\theta}_n$ to $\widetilde{\theta}_n$, Corollary (C) can be established by repeating the arguments given in earlier sections and the proof is therefore omitted.

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*, vol. 4. Armsterdam: North Holland, pp. 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.

AUTOR, D. H., KATZ, L. F., AND KRUEGER, A. B. (1998): "Computing Inequality: Have Computers Changed the Labor Market?," *Quarterly journal of economics*, 113, 1169–1213.

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

BARTH, E., DAVIS, J., AND FREEMAN, R. B. (2018): "Augmenting the Human Capital Earnings Equation with Measures of Where People Work," *Journal of Labor Economics*, 36, 71–97.

BEARE, B.K., SEO, J., AND SEO, W.K. (2017): "Cointegrated Linear Processes in Hilbert Space," *Journal of Time Series Analysis*, 38, 1010–1027.

BEARE, B.K. AND SEO, W.K. (2020): "Representation of I(1) and I(2) autoregressive Hilbertian processes," *Econometric Theory*, (forthcoming).

BHULLER, M., MAGNE, M., AND SALVANES, K. G. (2017): "Life-cycle Earnings, Education Premiums, and Internal Rates of Return," *Journal of Labor Economics*, 35, 993–1030.

BICKEL, P. AND WICHURA, M. (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications," *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): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1152.

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

BOSQ, P. (2000): *Linear Processes in Function Space (Lecture notes in statistics; 149)*. New York: Springer.

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.

CHANG, Y., KIM, C.S., AND PARK, J.Y. (2016):“Nonstationarity in Time Series of State Densities,” *Journal of Econometrics*, 192, 152–167.

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

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. (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” in J. Heckmand and E. Learner eds. *Handbook of Econometrics*, vol. 6B. Amsterdam: North Hollad, pp. 5549–5632.

CHEN, X. AND FAN, Y. (2006): “ Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models under Copula Misspecification,” *Journal of Econometrics*, 135, pp. 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. S., 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. S. AND PHILLIPS, P. C. B. (2018a): “Sequentially Testing Polynomial Model Hypotheses Using Power Transforms of Regressors,” *Journal of Applied Econometrics*, 33, 141–159.

CHO, J. S. AND PHILLIPS, P. C. B. (2018b): "Pythagorean Generalization of Testing the Equality of Two Symmetric Positive Definite Matrices," *Journal of Econometrics*, 202, 45–56.

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

CHO, J. S. AND WHITE, H. (2007): "Testing the Equality of Two Positive-Definite Matrices with Application to Information Matrix Testing," *Advances in Econometrics: Essays in Honor of Peter C. B. Phillips*. Vol. 33. Eds. Yoosoon Chang, Thomas B. Fomby, and Joon Y. Park (2014), 491–556. West Yorkshire, UK: Emerald Group Publishing Limited.

CHO, J. S., HUANG, M., AND WHITE, H. (2008): "Testing the Constant Mean Function of Functional Data," Discussion Paper, School of Economics, Yonsei University.

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

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

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

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

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.

HAPP, C. AND GREVEN, S. (2018): "Multivariate Fuctional Principal Component Analysis for Data Observed on Different (Dimensional) Domains," *Journal of the American Statistical Association*, 113, 649–659.

HECKMAN, J. J., LOCHNER, L. J., AND TODD, P. E. (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in E. Hanushek and F. Welch eds. *Handbook of the Economics of Education*, 1, 307–458.

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

HÖRMANN, S., KOKOSZKA, P., AND NISOL, G. (2018): "Testing for Periodicity in Functional Time Series," *Annals of Statistics*, 46, 2960–2984.

HORVATH, L. AND KOKOSZKA, P. (2012): *Inference for functional data with applications* (Vol. 200). Springer Science & Business Media.

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.

HUIZINGA, F. (1990): "An Overlapping Generations Model of Wage Determination," *Scandinavian Journal of Economics*, 92, 81–98.

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

JAMES, G., WANG, J., AND ZHU, J. (2009): "Functional Linear Regression That's Interpretable," *Annals of Statistics*, 37, 2083–2108.

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.

KATZ, L.F. AND MURPHY K. M. (1992): "Changes in Relative Wages, 1963?1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35–78.

KIM, J. AND PARK, J.Y. (2017): "Asymptotics for Recurrent Diffusions with Application to High Frequency Regression," *Journal of Econometrics*, 196, 37–54.

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.

LEMIEUX, T. (2006): "The "Mincer Equation" Thirty Years After *Schooling*, *Expereience*, *and Earnings*," in Shoshana Brossbard ed. *Jacob Mincer: A Pioneer of Modern Labor Economics*. New York: Springer. pp. 127–145.

LI, D., ROBINSON, P. M., AND SHANG, H. L. (2019): "Long-Range Dependent Curve Time Series," *Journal of the American Statistical Association*, forthcoming.

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

LIGHT, A. AND URETA, M. (1995): "Early-Career Work Experience and Gender Wage Differentials," *Journal of Labor Economics*, 13, 121–154.

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

MAGNAC, T., PISTOLESI. N., AND ROUX. S. (2018): "Post-Schooling Human Capital Investments and the Life Cycle of Earnings." *Journal of Political Economy*, 126, 1219–1249.

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

MINCER, J (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302

MINCER, J. (1974): *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

MINCER, J. AND JOVANOVIC, B. (1981): "Labor Mobility and Wages," in S. Rosen ed. *Studies in Labor Markets*, Chapter 5. Chicago: University of Chicago Press, pp. 21–64.

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.

MURPHY, K. M. AND WELCH, F. (1990): "Empirical Age-Earnings Profiles," *Journal of Labor Economics*, 8, 202–229

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.

PETERSEN, A. AND MÜLLER, H. G. (2016): "Functional Data Analysis for Density Functions by Transformation to a Hilbert Space," *Annals of Statistics*, 44(1), 183–218.

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

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

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

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

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.

SEO, W. K. AND BEARE, B. K. (2019): "Cointegrated Linear Processes in Bayes Hilbert space," *Statistics & Probability Letters*, 147, 90–95.

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

SIMONSOHN, S. U., NELSON, L. D. AND SIMMONS, J. P. (2014): "P-Curve: A Key to the File-Drawer," *Journal of Experimental Psychology: General*, 143, 534–547.

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.

STOCK, J. AND WATSON, M. (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

SU, L., SHI, Z., AND PHILLIPS, P. C. B. (2016): "Identifying Latent Structures in Panel Data," *Econometrica*, 84, 2215–2264.

SWAMY, P. AND TINSLEY, P. (1980): "Linear Prediction and Estimation methods for Regression Models with Stationary Stochastic Coefficients," *Journal of Econometrics*, 12, 103–142.

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.

WISWALL, M. AND ZAFAR, B. (2017): "Preference for the Workplace, Investment in Human Capital, and Gender," *Quarterly Journal of Economics*, 133(1), 457–507.

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," *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	1,000	2,000
	1%	5.14	5.30	4.38	2.78	1.94	1.50	0.84
$\mathcal{W}_n^\sharp$	5%	10.12	9.66	9.14	7.28	6.22	5.42	5.20
	10%	13.98	14.00	13.74	12.26	10.74	10.40	10.38
	1%	0.36	0.72	1.26	1.92	1.44	1.32	0.82
$\mathcal{LM}_n^\sharp$	5%	3.10	5.08	5.56	6.20	5.50	5.36	5.14
	10%	7.58	10.84	10.40	11.14	10.04	10.08	10.16
	1%	3.92	2.22	1.66	0.96	1.16	1.14	1.04
$\mathcal{QLR}_n^\sharp$	5%	12.70	8.26	6.62	5.48	5.26	5.62	5.02
	10%	19.72	14.44	12.30	10.80	9.92	11.22	10.24

Power of the Test Statistics (Level of Significance: 5%)

Statistics	$\pi_*\backslash n$	25	50	100	300	500	1,000	2,000
	0.10	9.54	9.04	6.36	3.52	3.66	4.54	9.06
	0.20	9.52	7.54	4.80	2.98	4.72	10.98	29.42
$\mathcal{W}_n^\sharp$	0.30	9.12	7.22	4.30	3.92	8.52	21.32	54.26
	0.40	9.38	7.26	4.36	5.48	11.66	33.22	72.82
	0.50	9.82	6.82	4.40	7.78	16.42	43.00	85.02
	0.10	3.10	4.50	3.52	2.72	3.22	4.06	8.88
	0.20	3.24	3.40	2.80	2.38	3.98	10.08	28.60
$\mathcal{LM}_n^\sharp$	0.30	3.26	3.30	2.58	3.16	7.48	20.06	53.66
	0.40	3.28	2.82	2.48	4.22	10.44	31.96	72.14
	0.50	3.28	2.54	2.60	6.62	15.04	41.68	84.50
	0.10	13.04	10.08	6.76	5.90	6.50	10.12	17.90
	0.20	14.16	9.80	7.20	7.14	11.56	22.98	47.76
$\mathcal{QLR}_n^\sharp$	0.30	13.94	10.54	6.90	10.00	18.10	39.00	72.28
	0.40	14.08	9.24	7.34	13.02	23.92	53.18	87.18
	0.50	13.52	8.48	7.10	15.92	30.18	63.66	93.88

Table 5.1: 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 $z_i \sim$ IID $U[0,1]$ and $u_i \sim$ IID $N(0,1)$, 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 $x_i := (1, z_i)'$, $z_i \sim$ IID $U[0,1]$, and $(\nu_i', \varepsilon_i)' \sim$ IID $N(0, I_3)$, where $\psi_* = (1,1)'$, $\gamma_* = 0.5$, and $\delta_* = 1$. Model: $\rho(\gamma, \theta_1, \theta_2) = \theta_1 + \theta_2 \exp(\gamma)$, $\gamma \in [0,1]$. Number of Experiments: 5,000.

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 5.2: 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 $\mathrm{Exp}(1)$. Alternative DGP: $x_i \sim$ IID $\pi_*\mathrm{Exp}(1) + (1-\pi_*)\mathrm{Exp}(2)$. Model: $\rho(\gamma, \theta_1, \theta_2) = \theta_1 + \theta_2(\gamma - 1)/(2\gamma - 1)^{1/2}$, $\gamma \in [1.5, 2.5]$. 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
$\mathcal{W}_n^\sharp$	0.10	7.46	6.72	6.06	8.10	10.62
	0.20	8.58	8.62	9.20	18.04	27.58
	0.30	9.32	11.56	15.58	35.86	52.46
	0.40	11.70	15.16	24.52	55.62	77.20
	0.50	13.74	20.66	33.06	74.44	93.24
$\mathcal{LM}_n^\sharp$	0.10	4.56	5.24	5.42	7.56	10.20
	0.20	5.60	6.38	7.94	17.00	25.98
	0.30	6.32	9.38	13.66	33.80	49.42
	0.40	8.30	12.86	21.88	52.92	74.28
	0.50	9.62	17.20	30.58	71.38	88.84
$\mathcal{QLR}_n^\sharp$	0.10	10.34	8.98	9.00	9.78	12.02
	0.20	11.56	10.44	10.94	20.36	30.10
	0.30	12.80	14.18	18.32	39.02	55.56
	0.40	15.54	18.36	27.16	59.06	79.78
	0.50	17.76	23.68	36.58	77.66	94.42

Table 5.3: 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 $N(0,5)$. Here, $c_1(\cdot)$ and $c_2(\cdot)$ are the independence and FGM copulas, respectively. Model: $\rho(\gamma, \theta_1, \theta_2) = \theta_1 + \theta_2\gamma$, $\gamma \in [0,1]$. Number of Experiments: 5,000.

Estimated FMSE of the mean of the log income paths								
	Male				Female			
	Quadratic	Cubic	Quartic	Quartic(r)	Quadratic	Cubic	Quartic	Quartic(r)
w/o Degree	6.08	5.94	5.93	5.93	5.77	5.66	5.64	5.64
Bachelor	5.58	5.52	5.47	5.55	5.32	5.27	5.22	5.28
Master	5.55	5.47	5.42	5.57	5.08	5.02	4.97	5.08
Ph.D	5.57	5.48	5.41	5.59	5.50	5.44	5.39	5.51

Estimated FMSE of the scaled mean of the log income paths								
	Male				Female			
	Quadratic	Cubic	Quartic	Quartic(r)	Quadratic	Cubic	Quartic	Quartic(r)
w/o Degree	3.20	2.92	2.90	2.90	3.03	2.81	2.78	2.78
Bachelor	3.15	3.04	2.96	3.10	3.10	3.01	2.94	3.04
Master	3.34	3.21	3.12	3.37	3.17	3.08	2.99	3.18
Ph.D	3.30	3.14	3.02	3.34	3.49	3.41	3.32	3.51

Table 6.4: FUNCTIONAL MEAN SQUARED ERRORS (FMSE) USING FUNCTION DATA OVER 0 TO 40 WORK EXPERIENCE YEARS. This table shows the estimated FMSEs of the log income paths under the quadratic, cubic, quartic and restricted quartic specifications for each group of the workers classified according to their education levels and genders. Here, the restricted quartic specification means the model specified by Murphy and Welch (1990) given in (8).

Inference results on the mean of log income paths across different genders

	Education Level	Wald	LM	QLR
Quadratic	w/o Degree	5.08	11.45**	11.51
	Bachelor	140.41**	124.26**	1426.65**
	Master	54.40**	29.02**	345.38**
	Ph.D	40.32**	42.77**	459.46**
Cubic	w/o Degree	20.02**	41.57**	28.50
	Bachelor	174.84**	130.21**	143.77**
	Master	40.69**	32.37**	364.48**
	Ph.D	60.51**	48.23**	468.17**
Quartic	w/o Degree	24.88**	42.31**	30.16
	Bachelor	290.97**	153.36**	1439.01**
	Master	63.05**	34.45**	365.54**
	Ph.D	83.61**	55.99**	470.26**
Quartic(r)	w/o Degree	13.32**	19.73**	18.17
	Bachelor	150.16**	132.58**	1414.82**
	Master	33.72**	30.13**	352.93**
	Ph.D	43.57**	47.33**	457.76**

Inference results on the mean of the scaled log income paths across different genders

	Education Level	Wald	LM	QLR
Quadratic	w/o Degree	17.23**	16.55**	9.94
	Bachelor	21.46**	41.00**	73.40**
	Master	6.65	13.33**	19.12
	Ph.D	0.71	1.17	3.14
Cubic	w/o Degree	13.53**	27.18**	26.89*
	Bachelor	26.58**	54.48**	84.47**
	Master	7.69	18.07**	38.22*
	Ph.D	6.08	13.14*	11.86
Quartic	w/o Degree	26.149**	29.77**	28.55*
	Bachelor	53.80**	57.17**	85.76**
	Master	13.18*	17.58**	39.28*
	Ph.D	9.97	17.55**	13.95
Quartic(r)	w/o Degree	12.86**	16.57**	16.56
	Bachelor	24.93**	39.59**	61.57**
	Master	7.72	12.18**	26.68*
	Ph.D	0.6346	0.83	1.45

Table 6.5: INFERENCE RESULTS USING FUNCTION DATA OVER 0 TO 40 WORK EXPERIENCE YEARS. This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of equal mean of log income paths across different genders. The figures attached by '*' and '**' indicate the rejection of the null hypothesis at the 5% and 1% significance level, respectively.

Inference results on the mean of log income paths across different education levels

		Male			Female		
		Wald	LM	QLR	Wald	LM	QLR
Quadratic	w/o Degree vs. Bachelor	1908.46**	1374.39**	24804.82**	1589.52**	1129.27**	16617.82**
	Bachelor vs. Master	385.68**	173.95**	2621.40**	300.29**	191.17**	2153.12**
	Master vs. Ph.D	56.11**	51.10**	333.48**	73.14**	42.08**	194.69**
Cubic	w/o Degree vs. Bachelor	2178.91**	1510.92	2544.11**	2013.79**	1348.72**	16973.37**
	Bachelor vs. Master	234.09**	181.54**	2803.90**	250.68**	197.24**	2215.04**
	Master vs. Ph.D	64.47**	48.02**	335.42**	50.23**	41.50**	203.51**
Quartic	w/o Degree vs. Bachelor	2654.04**	12678.00**	13178.66**	1060.14**	502.77**	8578.08**
	Bachelor vs. Master	399.12**	204.82**	2806.21**	380.31**	229.95**	2215.59**
	Master vs. Ph.D	100.55**	58.26**	336.26**	78.14**	45.57**	203.62**
Murphy R	w/o Degree vs. Bachelor	1177.80**	1063.50**	12937.50**	535.29**	463.89**	8477.60**
	Bachelor vs. Master	192.97**	180.37**	2732.06**	216.33**	204.09**	2174.74**
	Master vs. Ph.D	62.83**	54.63**	333.22**	50.57**	42.72**	202.33**

Inference results on the scaled mean of log income paths across different education levels

		Male			Female		
		Wald	LM	QLR	Wald	LM	QLR
Quadratic	w/o Degree vs. Bachelor	93.09**	181.28**	257.58**	36.40**	89.41**	98.14**
	Bachelor vs. Master	44.20**	59.41**	6.78	24.62**	40.87**	47.08**
	Master vs. Ph.D	8.91*	9.75*	20.65	27.97**	26.33**	67.23**
Cubic	w/o Degree vs. Bachelor	353.74**	827.37**	541.89**	119.76**	276.20**	227.33**
	Bachelor vs. Master	36.632**	86.33**	189.23**	26.99**	55.92**	109.00**
	Master vs. Ph.D	9.29	16.25**	22.59	19.03**	26.89**	76.05**
Quartic	w/o Degree vs. Bachelor	719.97**	946.54**	1346.18**	332.41**	541.68**	574.77**
	Bachelor vs. Master	87.17**	81.95**	191.55**	58.52**	60.14**	109.55**
	Master vs. Ph.D	17.85*	18.36**	23.43	28.35**	27.27**	76.15**
Murphy R	w/o Degree vs. Bachelor	381.66**	601.30**	789.26**	159.01**	247.771**	318.7432**
	Bachelor vs. Master	36.31**	53.38**	117.40**	24.85**	37.37**	68.70**
	Master vs. Ph.D	9.30*	9.67*	20.39*	18.34**	26.08**	74.87**

Table 6.6: INFERENCE RESULTS USING FUNCTION DATA OVER 0 TO 40 WORK EXPERIENCE YEARS. This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of equal mean of log income paths across different education levels. The figures attached by '*' and '**' indicate the rejection of the null hypothesis at the 5% and 1% significance level, respectively.

Estimated FMSE of the mean of the log income paths

	Male				Female			
	Quadratic	Cubic	Quartic	Quartic(r)	Quadratic	Cubic	Quartic	Quartic(r)
w/o Degree	5.16	5.13	5.13	5.13	4.83	4.80	4.80	4.80
Bachelor	4.74	4.70	4.70	4.70	4.48	4.45	4.44	4.44
Master	4.71	4.68	4.68	4.68	4.30	4.27	4.27	4.26
Ph.D	4.74	4.71	4.71	4.71	4.63	4.61	4.60	4.60

Estimated FMSE of the scaled mean of the log income paths

	Male				Female			
	Quadratic	Cubic	Quartic	Quartic(r)	Quadratic	Cubic	Quartic	Quartic(r)
w/o Degree	2.32	2.24	2.24	2.24	2.27	2.21	2.21	2.20
Bachelor	2.27	2.20	2.19	2.19	2.29	2.23	2.22	2.21
Master	2.24	2.18	2.17	2.17	2.20	2.15	2.15	2.15
Ph.D	2.19	2.11	2.11	2.11	2.24	2.19	2.18	2.18

Table 6.7: FUNCTIONAL MEAN SQUARED ERRORS (FMSE) USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS. This table shows the estimated FMSEs of the log income paths under the quadratic, cubic, quartic and restricted quartic specifications for each group of the workers classified according to their education levels and genders. Here, the restricted quartic specification means the model specified by Murphy and Welch (1990) given in (8).

Inference results on the mean of log income paths across different genders

	Education Level	Wald	LM	QLR
Quadratic	w/o Degree	1.20	0.48	6.85
	Bachelor	164.48**	132.47**	1243.35**
	Master	37.62**	31.31**	316.41**
	Ph.D	34.79**	34.94**	347.86**
Cubic	w/o Degree	8.85	10.29*	8.16
	Bachelor	435.48**	146.02**	1244.39**
	Master	98.55**	34.18**	316.64**
	Ph.D	131.26**	45.65**	349.11**
Quartic	w/o Degree	16.38**	12.11*	8.76
	Bachelor	592.75**	163.49**	1244.44**
	Master	131.89**	36.13**	316.72**
	Ph.D	199.89**	48.03**	349.29**
Quartic(r)	w/o Degree	1.51	1.44	7.25
	Bachelor	172.20**	147.59**	1243.31**
	Master	39.37**	34.25**	316.40**
	Ph.D	45.41**	45.78**	349.04**

Inference results on the mean of the scaled log income paths across different genders

	Education Level	Wald	LM	QLR
Quadratic	w/o Degree	2.57	2.21	0.36
	Bachelor	0.56	1.35	2.22
	Master	0.11	0.18	0.41
	Ph.D	0.31	3.18	1.42
Cubic	w/o Degree	3.68	9.53	1.67
	Bachelor	4.26	13.08*	3.26
	Master	0.83	2.65	0.64
	Ph.D	2.70	8.22	2.67
Quartic	w/o Degree	14.42*	11.03	2.28
	Bachelor	12.23*	14.65*	3.30
	Master	4.53	2.75	0.72
	Ph.D	10.67	7.98	2.84
Quartic(r)	w/o Degree	1.80	2.06	0.77
	Bachelor	0.59	1.28	2.18
	Master	0.10	0.15	0.40
	Ph.D	3.68	5.43	2.60

Table 6.8: INFERENCE RESULTS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS. This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of equal mean of log income paths across different genders. The figures attached by '*' and '**' indicate the rejection of the null hypothesis at the 5% and 1% significance level, respectively.

Inference results on the mean of log income paths across different education levels

		Male			Female		
		Wald	LM	QLR	Wald	LM	QLR
Quadratic	w/o Degree vs. Bachelor	1934.80**	1452.74**	21226.31**	1543.13**	1256.10**	13911.35**
	Bachelor vs. Master	218.02**	175.93**	2363.97**	235.96**	190.21**	1858.30**
	Master vs. Ph.D	52.65**	43.93**	287.84**	51.89**	39.87**	198.45**
Cubic	w/o Degree vs. Bachelor	6746.19**	1562.15**	8675.22**	5245.07**	1461.64**	13911.62**
	Bachelor vs. Master	822.53**	202.54**	966.42**	760.05**	220.86**	1859.11**
	Master vs. Ph.D	147.68**	54.71**	157.25**	92.39**	42.28**	198.51**
Quartic	w/o Degree vs. Bachelor	1916.50**	1536.87**	2126.46**	6696.87**	1508.22**	1391.18**
	Bachelor vs. Master	217.74**	192.22**	2364.33**	960.79**	252.28**	1859.29**
	Master vs. Ph.D	67.39**	52.15**	288.62**	96.64**	43.91**	198.53**
Murphy R	w/o Degree vs. Bachelor	1916.50**	1256.10**	21226.46**	547.36**	494.71**	7014.03**
	Bachelor vs. Master	217.74**	190.21**	2364.33**	238.53**	210.17**	1858.61**
	Master vs. Ph.D	67.39**	52.15**	288.62**	53.44**	41.19**	198.54**

Inference results on the scaled mean of log income paths across different education levels

		Male			Female		
		Wald	LM	QLR	Wald	LM	QLR
Quadratic	w/o Degree vs. Bachelor	2.23	2.81	4.30	1.51	1.24	1.84
	Bachelor vs. Master	2.93	3.89	13.01	2.39	3.58	11.17
	Master vs. Ph.D	2.84	3.34	7.31	3.38	4.48	12.53
Cubic	w/o Degree vs. Bachelor	5.12	3.78	4.98	1.89	1.39	2.11
	Bachelor vs. Master	8.71	5.33	14.06	7.73	4.45	11.98
	Master vs. Ph.D	4.58	3.46	8.03	4.81	5.33	12.59
Quartic	w/o Degree vs. Bachelor	7.15	4.21	5.08	13.15	7.12	2.30
	Bachelor vs. Master	19.39**	5.88	4.11	17.15**	4.52	12.16
	Master vs. Ph.D	11.11	3.82	8.15	20.12**	6.41	12.61
Murphy R	w/o Degree vs. Bachelor	2.12	2.96	4.46	1.02	1.41	2.02
	Bachelor vs. Master	5.21	3.88	13.37	3.85	3.63	11.48
	Master vs. Ph.D	8.34	3.94	8.09	2.96	5.06	12.63

Table 6.9: INFERENCE RESULTS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS. This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of equal mean of log income paths across different education levels. The figures attached by '*' and '**' indicate the rejection of the null hypothesis at the 5% and 1% significance level, respectively.

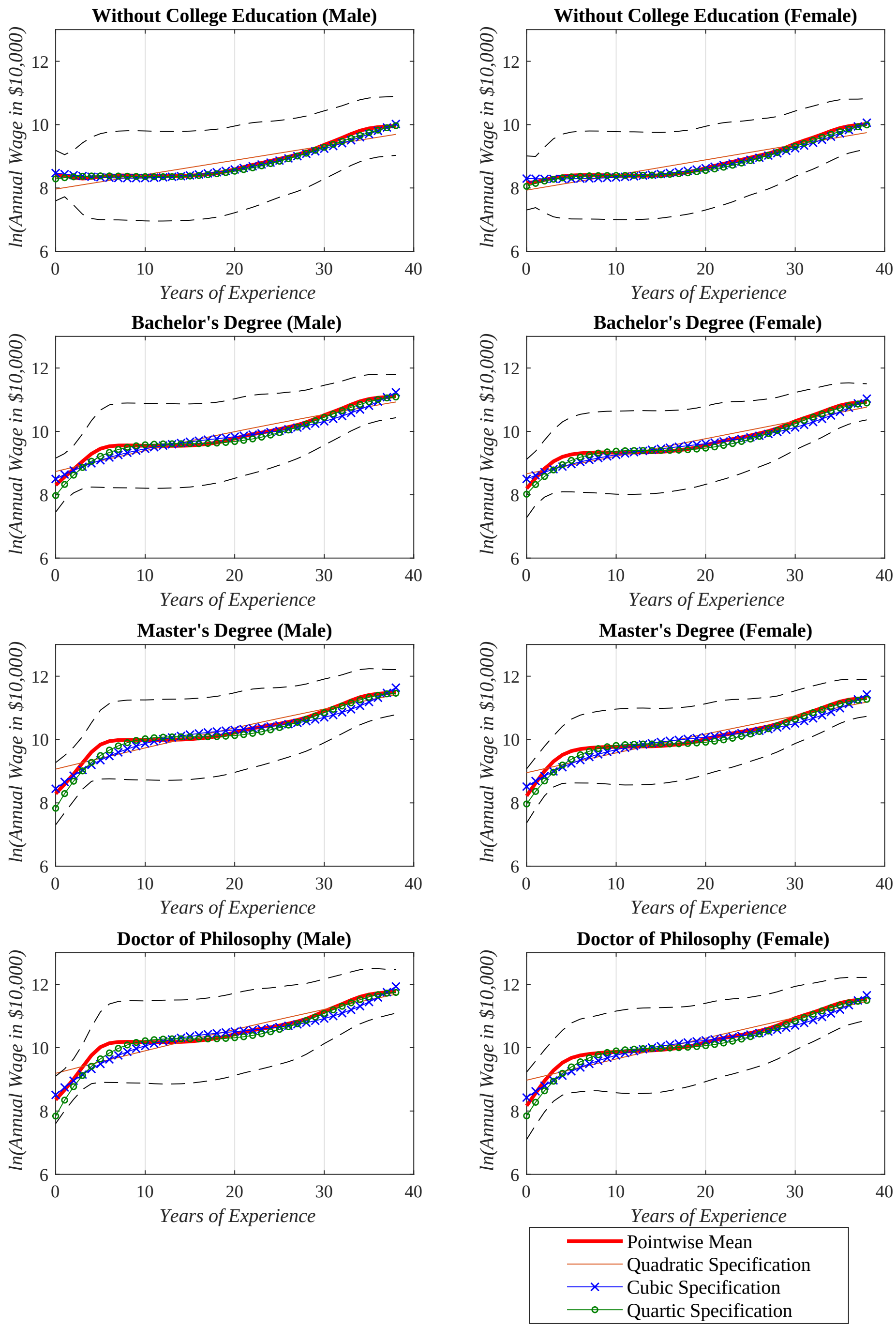

Figure 1: ESTIMATED MEAN CURVES OVER 0 TO 40 WORK EXPERIENCE YEARS. The red line corresponds to the pointwise mean of the individual log income paths, and the dotted lines correspond to its 80% bootstrap confidence bands. The mean estimates of log income paths under the quadratic, cubic and quartic specifications are displayed in brown, blue, and green lines.

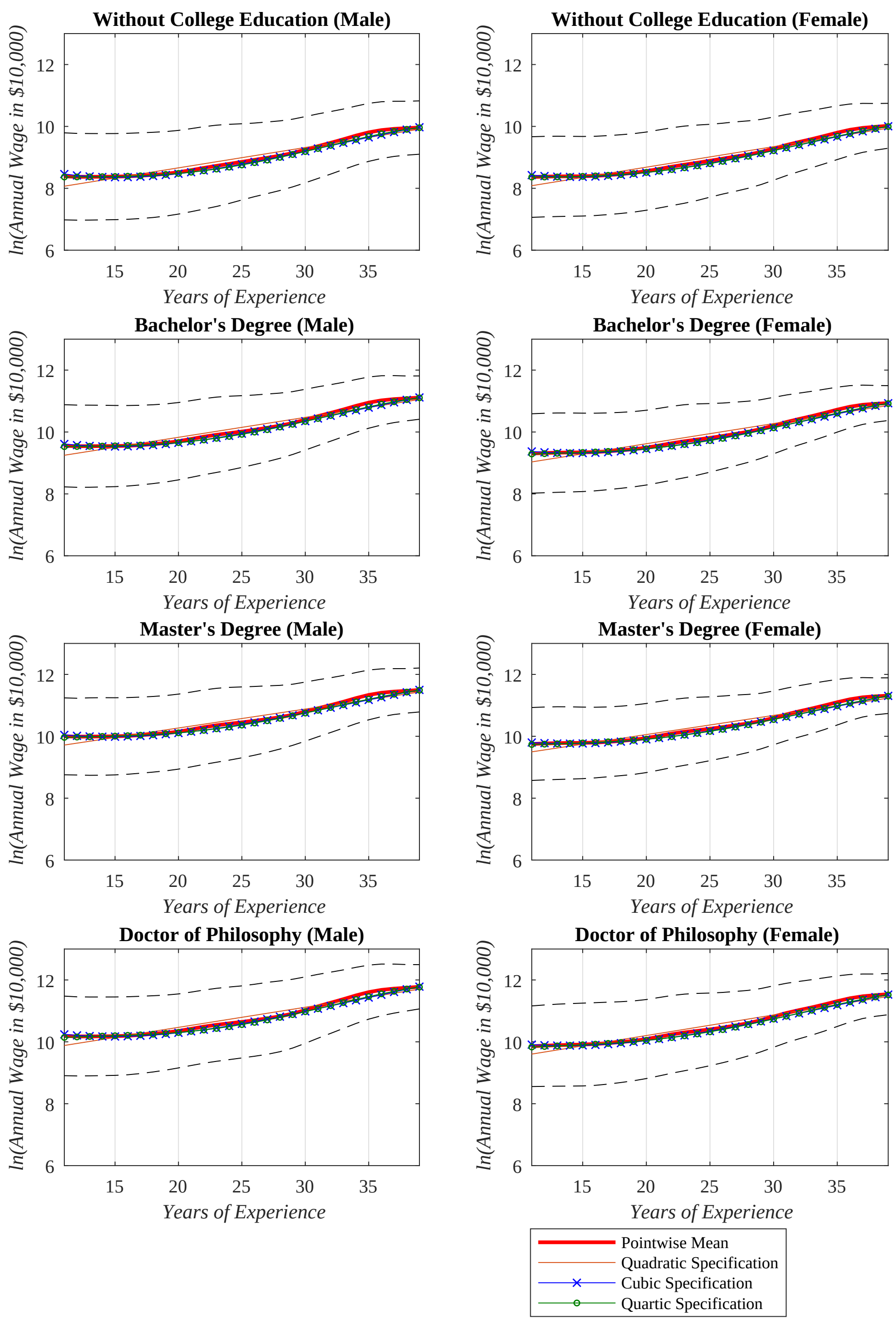

Figure 2: ESTIMATED MEAN CURVES OVER 10 TO 40 WORK EXPERIENCE YEARS. The red line corresponds to the pointwise mean of the individual log income paths, and the dotted lines correspond to its 80% bootstrap confidence bands. The mean estimates of log income paths under the quadratic, cubic and quartic specifications are displayed in brown, blue, and green lines.