Functional Data Inference in a Parametric Quantile Model applied to Lifetime Income Curves

JIN SEO CHO
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
National University of Singapore
Email: ecssj@nus.edu.sg

This version: February 2023

Abstract

A parametric quantile function estimation procedure is developed for functional data. The approach involves minimizing the sum of integrated functional distances that measure the functional gap between each functional observation and the quantile curve in terms of the check function. The procedure is validated under both correctly specified and misspecified models by allowing for the presence of nuisance parameter estimation effects. Testing methodology is developed using Wald, Lagrange multiplier, and quasi-likelihood ratio procedures in this functional data setting. Finite sample performance is assessed using simulations and the methodology is applied to study how lifetime income paths differ between genders and among different education levels using continuous work history samples. The methodology enables the analysis of full career income paths with temporal and possibly persistent dependence structures embodied in the observations.The results capture both gender and education effects but these empirical differences are shown to be mitigated upon rescaling to take account of lifetime experience and job mobility.

Key Words: Functional data; quantile function; nuisance effects; quantile inference; lifetime income path; gender and education effects.

Subject Class: C12, C21, C31, C80.

Acknowledgements: Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 at Yale University and a Kelly Fellowship at the University of Auckland.

1 Introduction

With the continuing growth and availability of vast datasets in economics and finance, the use of functional data in applied econometric work is becoming increasingly popular. These developments are facilitated by methodological extensions of existing econometric tools of estimation and inference to a functional data environment, which in turn relies on early statistical research, including Ramsay and Dalzell (1991), Rice and Silverman (1991), Ramsay and Silverman (1997), Bosq (2000), and Horvath and Kokoszka (2012) among many others.

In the econometric literature various empirical features of functional data have been studied, including quantile curve properties. To mention a few: Li, Robinson, and Shang (2020) use functional principal component analysis to estimate the long run covariance function of functional data with long run dependence; Chang, Hu, and Park (2019) focus on the serial correlation between functional observations; Crambes, Gannoun, and Henchiri (2013) study estimation of the quantile function when the dependent variable is a random variable but the explanatory variables involve functional observations, which are transformed to random variables by integration using support vector machines to estimate the quantile function between the dependent variable and transformed observations; Phillips and Jiang (2019) develop parametric autoregressive methods with function valued time series, establish asymptotic theory allowing for nonstationarity, and apply the methods to household Engel curves; Cho, Phillips, and Seo (2022) explore conditional mean estimation and inference with functional data in a parametric model context, develop asymptotic theory, and apply the methodology to lifetime income profiles. Readers are referred to the latter paper for further discussion of the existing literature.

Quantile regression is commonly used to provide useful additional information about how the generating mechanism may be influenced at different quantiles. This device has been heavily used in empirical work with time series and cross section data but may also be employed when the data involve observable curves or functions, just as is the case in estimating moments such as the population mean of a function or curve. Development of this framework is one of the goals of the present paper. Functional quantile regression helps to provide a deeper analysis of the mechanisms that influence the characteristics of observed curves, such as the lifetime income profiles studied in Cho, Phillips, and Seo (2022), revealing how such factors as gender and education may affect the income profile at various quantiles in the population.

Quantile function estimation presents new econometric challenges and many advantages. Existing studies in the literature assume a scalar-valued random response variable that is determined by certain inner products of function-valued covariates with regression coefficient functions that may be quantile dependent (e.g., Cardot et al., 2005; Ferraty et al., 2005; Chen and Müller, 2012; Kato, 2012; Crambes et al., 2013; Li et al., 2022). This model enables inference concerning the constancy of the coefficient function across quantiles. Our approach differs from this research by allowing the response variable itself to be function-valued, so that the model explains how the dependent function is determined as an element in a function space and how this determination may be influenced at different quantiles. Each quantile curve is formulated in a

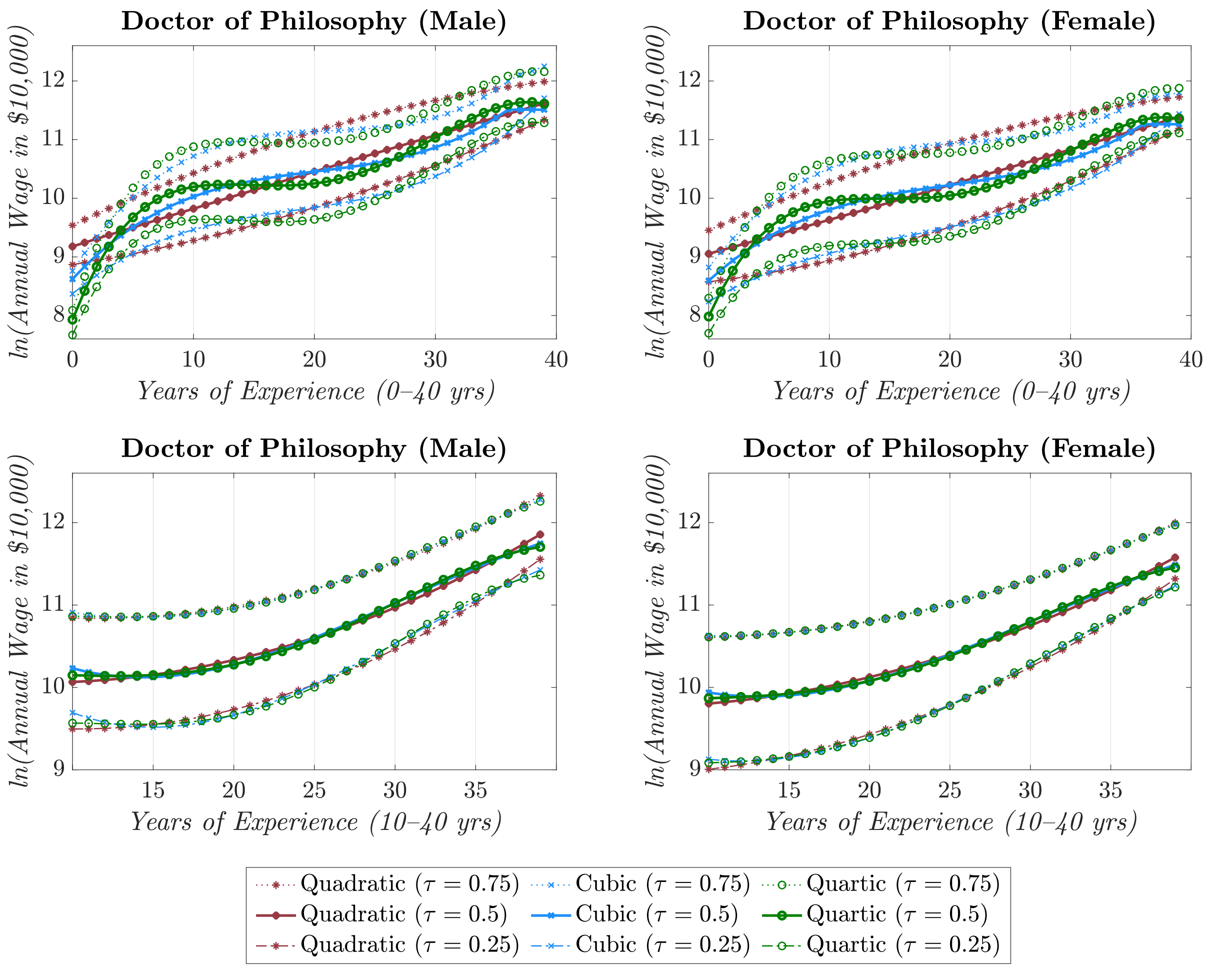

Figure 1: Estimated quantile curves at levels $\tau \in \{0.25, 0.50, 0.75\}$ for 0-40 year and 10-40 year working careers of men and women with doctoral level education.

parametric regression form wherein the parametric coefficients provide meaningful measures of how certain covariates influence the response curves at each quantile. This methodology enables analysis of data, such as full career income paths, with temporal and possibly persistent dependence structures embodied in the observations themselves.

The goal of the present study is therefore to develop a methodology for modeling such quantile curves as function-valued regressions with an asymptotic theory of estimation and inference that is useful to applied researchers interested in understanding the features of functional data. To illustrate this methodology we examine worker's log income profiles as curves that evolve over time and reflect the influence on income of such facets as gender and educational qualifications in combination with years of accumulated career experience. Figure 1 provides an empirical example studied later in the paper in terms of the lifetime log income paths (LIPs) over 0–40 years and over 10–40 years of work experience for female and male white workers in the US each with doctoral education levels and each born between 1960 and 1962. The lines in the figure show fitted quantile time curves obtained by quadratic (red), cubic (blue), and quartic (green) parametric specifications of these functions. The three lines at the top and the lines at the bottom of the figure are the fitted quantile functions at levels $\tau = 0.75$ and $\tau = 0.25$. The three middle lines are the fitted quantile functions at the median level $\tau = 0.5$. These fitted curves show evidence of differences between genders. For each τ, the male quantile function is located above the female quantile function; and the difference between the female quantile functions at $\tau = 0.25, 0.75$ are wider at lower years of experience but more narrower at higher years of experience than the corresponding male quantile functions. These measures speak to labor market differences between male and female workers at the higher educational attainment. More generally, the curves provide a convenient high dimensional summary of career income profiles for workers in various quantiles according to categories that can be used for inference once appropriate methodology for dealing with functional data of this type is developed. Notably greater nonlinearity is apparent in the fitted curves of the lifetime LIPs over 0–40 years than those over 10–40 years, reflecting the effects of early career differences. The empirical investigation reported in Section 8 provides detailed classifications and inferences about these curves according to gender and multiple education levels.

The simplest approach to estimation and inference is to use a parametric model for the quantile function that can flexibly capture the quantile levels as a function defined over the domain of the data. If the functional data are continuously distributed with a cumulative distribution function (CDF) at each domain level, the true quantile function is also continuous; and, when correctly specified parametrically, the quantile function can be consistently estimated and predicted using just a finite number of unknown coefficients of relevant covariates. Correct model specification is particularly difficult in this setting because there is always a positive probability of quantile crossings and hence misspecification, a difficulty that applies even in linear quantile regressions for simple random variables, as discussed in Phillips (2015). The present paper allows for the quantile function model to be misspecified, instead of enforcing fully orderly quantile behavior. In such cases the estimated quantile function is viewed as an approximation for the quantile levels and asymptotic properties of the estimated parameters are developed under potential model misspecification, parallel to quasi-maximum likelihood estimation of the conditional mean function using a misspecified model, as in White (1982).

This paper studies quantile function estimation in settings where parametric misspecification is allowed and its asymptotic implications are examined. Functional data can be affected by parameter estimation errors that produce nuisance effects which can impact the quantile function asymptotics, which are examined separately. When several percentiles are considered, multiple quantile functions have to be estimated, the estimates are asymptotically related, and their large sample behavior is investigated, allowing for both potential nuisance effects and misspecification. Wald, Lagrange multiplier (LM), and likelihood-ratio (LR) tests are constructed for inference about the parametric curves and simulations are performed to assess finite sample performance. Finally, the testing methodology is applied to the empirical income profiles of white male and female workers born in the U.S. between 1960 and 1962, exploring how the fitted quantile functions are affected at different percentile levels by gender and education, shedding light on disparities such as those displayed in Figure 1. It is shown that when log income profiles are rescaled in a manner that accounts for each individual's integrated log income path over their work experience years both gender and education effect differences diminish, thereby implying that these two factors influence the log income profile proportionally across quantiles. Larger differences in the quantile curves are found when the first ten years of working careers are included in the observations.[1]

The organization of the paper follows. Section 2 motivates the use of functional data and examines both correctly specified and misspecified quantile function model estimation, and analyzes the relation between correctly specified and misspecified models. Section 3 introduces models for functional data with nuisance effects. Section 4 examines consistent covariance matrix estimation of the estimated parameters, Section 5 considers multiple quantile level estimation, and Section 6 develops inferential methods for quantile curve functions. Section 7 reports simulation findings and Section 8 applies the methodology to worker income profiles. Section 9 concludes. An Online Supplement contains proofs, technical material, and additional empirical results. For notation we use $\mathcal{L}_{ip}(\cdot)$ and $\mathcal{C}^{(\ell)}(\cdot)$ to denote spaces of Lipschitz continuous functions and ℓ-times continuously differentiable functions defined on their respective arguments, and $\lambda_{\min}(A)$ denotes the smallest eigenvalue of the square matrix A. Other notation is standard.

2 Estimation without Nuisance Effects

We begin by describing the methodological framework. Let $G(\cdot)$ $(\in \mathbb{R})$ be a continuous random function defined on a set Γ that is a compact and convex subset of $\mathbb{R}^g$ $(g \in \mathbb{N})$. Let $x_\tau(\gamma)$ be the quantile level associated with a percentile $\tau \in (0,1)$ so that $x_\tau(\gamma) := \inf\{x \in \mathbb{R} : F_\gamma(x) \geq \tau\}$, where for each $\gamma \in \Gamma$, $F_\gamma(\cdot)$ is the cumulative distribution function (CDF) of $G(\gamma)$. Our goal is to estimate $x_\tau(\gamma)$ consistently uniformly in γ and to make inferences from the functional observations about γ and the functional form $G(\cdot)$.

A number of empirical examples motivate this methodology and help to shape the framework of the present study. An early example is apparent in Mincer (1974) and Mincer and Jovanovic (1981) who modeled the functional form of labor income career profiles as potentially quadratic in career years. Cho, Phillips, and Seo (2022) recently extended that investigation using continuous work history sample (CWHS) data. In that research, for each individual an annual labor income profile before taxes was interpolated using local polynomial kernel estimation, producing a continuous income path. These lifetime income paths were compared across demographic groups according to the gender, years of work experience, and education of each worker, revealing that the mean income paths were largely proportional over gender and education levels. This investigation is more directly executed in the present study by treating the underlying path data as a continuous random function, represented by $G(\cdot)$, over career years and its quantile function $x_\tau(\cdot)$ may

[1] The Online Supplement provides a detailed comparison between the quantile curve findings for 0-40 working career and 10-40 working career cycles and for all tertiary education levels. See Figs. A.1 and A.2 and attendant discussion in the Online Supplement.

then be estimated. For each γ, the mean function $\mathbb{E}[G(\gamma)]$ can be represented as the integral $\int_0^1 x_\tau(\gamma)d\tau$, so that different gender and education effects on the income profiles can be identified at different τ percentile levels in a more direct manner. This approach is illustrated in Figure 1, as discussed above. Section 8 reports a more detailed analysis of these data, classifying according to various education levels and investigating how gender and education influence the quantile functions using the inferential methods developed below.

This type of quantile functional data analysis is by no means limited to labor income profiles. Measurement of any economic variable over time is a fundamental step in evaluating the evolution and impact of prevailing economic conditions. Quantile function evolution over time at different τ levels using functional observations enables estimation and inference concerning the impact of relevant covariates on the shapes of these quantile curves. As another example, many government economic policies are implemented for a redistributive purpose. Minimum wage legislation, capital gains taxes, and progressive income taxes are all intended to impact particular groups differentially rather than uniformly across the entire economy. For instance, if $G(\cdot)$ denotes the income process after taxes over time and a capital gains tax is levied within the sample period, it may be difficult to detect the treatment effect of the capital gains tax by estimating just the mean value $\mathbb{E}[G(\cdot)]$. Instead, estimation of the time profile of a bottom or top percentile x-% of the income distribution after taxes might be much more useful in detecting the relevant treatment effect, viz., $x_\tau(\cdot)$.

Our approach also contributes to the literature by enabling estimation of and inference concerning possibly misspecified parametric models for the quantile function $x_\tau(\cdot)$. Davies (1977, 1987) discussed testing hypotheses where a nuisance parameter is not identified and the resulting methodology has been applied in several econometric model contexts. For example, Andrews (1993) used this approach in developing testing methodology for structural break analysis without knowedge of the structural break point and where the break point is treated as a nuisance parameter that is unidentified under the null of no structural break. That paper showed how the appropriately standardized score function converges weakly to a functional of a Gaussian process defined on the unit interval under the null. For such a case, the individual quasi-score obtained with respect to the identified parameter can be treated as $G(\cdot)$ on the space of the nuisance parameter that is unidentified under the null, and the quantile function $x_\tau(\cdot)$ can be developed for use in this and other structural break tests.

To provide a formal framework for quantile functional data analysis we specifically suppose the following data generating process (DGP) condition for continuous functional random observations.

Assumption 1. *(i)* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a complete probability space and* Γ *is a compact metric space; (ii)* $\{G_i : \Gamma \mapsto \mathbb{R}\}_{i=1}^n$ *is a set of identically and independently distributed (iid) observations such that for each* $\gamma \in \Gamma$, $\{G_i(\gamma)\}$ *is* $\mathcal{F}$*-measurable, and* $G_i(\cdot) \in \mathcal{L}_{ip}(\Gamma)$ *for all* $\omega \in F \in \mathcal{F}$ *with* $\mathbb{P}(F) = 1$*; (iii)* $(\Gamma, \mathcal{G}, \mathbb{Q})$ *and* $(\Omega \times \Gamma, \mathcal{F} \otimes \mathcal{G}, \mathbb{P} \cdot \mathbb{Q})$ *are complete probability spaces, and* $g_i(\cdot, \cdot)$ *is* $\mathcal{F} \otimes \mathcal{G}$*-measurable ; (iv) for each* γ, $F_\gamma(\cdot) \in \mathcal{C}^{(1)}(\mathbb{R})$*, and* $f_\gamma(\cdot)$ *is uniformly bounded, where* $F_\gamma(\cdot)$ *and* $f_\gamma(\cdot)$ *are the CDF and probability density function (PDF) of* $G_i(\gamma)$*, respectively.* □

The DGP condition in Assumption 1 extends that of Cho, Phillips, and Seo (2022). Here, $\mathbb{Q}$ is an adjunct probability measure that augments $\mathbb{P}$ and is a probability measure selected by the investigator and attached to the space $(\Gamma, \mathcal{G})$ to complete the probabilistic structure of a parametric curve representation and assist in developing parameter estimation.

The standard quantile regression (QR) framework of Koenker and Bassett (1978) is now extended to this formal setting to accommodate quantile function regression. Specifically, for each $\tau \in (0, 1)$, we let the check function be defined as $\xi_\tau(u) := u\,(\tau - \mathbb{1}\{u \leq 0\})$ and further let

$$q_\tau(\gamma, u) := \mathbb{E}[\xi_\tau(G(\gamma) - u)] = (\tau - 1)\int_{-\infty}^{u}(g - u)dF_\gamma(g) + \tau \int_{u}^{\infty}(g - u)dF_\gamma(g).$$

Note that for each γ, $q_\tau(\gamma, \cdot)$ is minimized at $x_\tau(\gamma)$. Furthermore, if $u = x_\tau(\gamma)$, it follows that for each γ,

$$q_\tau(\gamma, x_\tau(\gamma)) = \mathbb{E}[\xi_\tau(G(\gamma) - x_\tau(\gamma))] = \tau \int_{-\infty}^{\infty}(g - x_\tau(\gamma))dF_\gamma(g) - \int_{-\infty}^{x_\tau(\gamma)}(g - x_\tau(\gamma))dF_\gamma(g).$$

Let $d_\tau(\gamma, u) := \mathbb{E}[\xi_\tau(G(\gamma)-u)]-\mathbb{E}[\xi_\tau(G(\gamma)-x_\tau(\gamma))]$, so that $q_\tau(\gamma, u) = d_\tau(\gamma, u)+\mathbb{E}[\xi_\tau(G(\gamma)-x_\tau(\gamma))]$. Here, $d_\tau(\cdot, u)$ is the only term associated with u on the right side, so that optimization of $q_\tau(\gamma, \cdot)$ can be equivalently conducted by optimizing $d_\tau(\gamma, \cdot)$. Furthermore, we can view $d_\tau(\cdot, u)$ in a different way by associating it with a model for $x_\tau(\gamma)$. For this purpose, we provide the following lemma.

Lemma 1. *Given Assumption 1, for each* $u \in \mathbb{R}$, $d_\tau(\gamma, u) = \int_{\min[u, x_\tau(\gamma)]}^{\max[u, x_\tau(\gamma)]} |F_\gamma(g) - F_\gamma(x_\tau(\gamma))|dg.$ □

Lemma 1 implies that $d_\tau(\gamma, u) \geq 0$ uniformly in u, and $d_\tau(\gamma, u) = 0$ if and only if $u = x_\tau(\gamma)$. Therefore, $d_\tau(\gamma, \cdot)$ is minimized by letting $u = x_\tau(\gamma)$.

We next suppose a parametric model for the quantile function as a function of the parameter γ and relate this model to Lemma 1. More specifically, suppose that an empirical investigator specifies a particular model for u in Lemma 1 to further minimize $d_\tau(\gamma, \cdot)$. We write this model for $x_\tau(\cdot)$ in the general form

$$\mathcal{M}_\tau := \{\rho_\tau(\cdot, \theta_\tau) : \theta_\tau \in \Theta_\tau\}, \tag{1}$$

where the parameter space Θ_τ is a compact and convex subset in $\mathbb{R}^{c_\tau}$ ($c_\tau \in \mathbb{N}$). That is, the researcher chooses a specific parametric functional form $\rho_\tau(\cdot, \theta_\tau)$ to model the quantile function $x_\tau(\cdot)$. Then $\mathcal{M}_\tau$ is *correctly specified*, if there exists a $\theta_\tau^0 \in \Theta_\tau$ such that $\rho_\tau(\cdot, \theta_\tau^0) = x_\tau(\cdot)$. Otherwise, $\mathcal{M}_\tau$ is misspecified. In the current study the model $\mathcal{M}_\tau$ may be misspecified or correctly specified for the true functional quantile $x_\tau(\cdot)$ and asymptotic theory of estimation and inference is developed for both cases. To fix ideas, a simple linear specification of ρ_τ such as (10) is used later in simulations and has the form $\rho_\tau(\gamma, \theta_\tau) = \theta_{\tau1} + \theta_{\tau2}\gamma$, with parameter vector $\theta_\tau = (\theta_{\tau1}, \theta_{\tau2})' \in \Theta_\tau \subset \mathbb{R}^2$, dimension $c_\tau = 2$, and $\gamma \in \Gamma \subset \mathbb{R}$. Various fitted quadratic and higher order polynomial quantile curves are alternatives to this linear specification. Some examples are shown in the empirical illustration of Fig. 1 and more are considered in the application of Section 8. The model formulation given in (1) provides a convenient parametric formulation of such functional information that enables empirical comparisons of the quantile curves obtained from various parametric model specifications in applications.

We now provide different views on $d_\tau(\cdot, u)$ by associating Lemma 1 with $\mathcal{M}_\tau$. First, if we combine $\mathcal{M}_\tau$ with $d_\tau(\gamma, \cdot)$ by integrating the latter with respect to γ weighted by the adjunct probability $\mathbb{Q}(\gamma)$, it follows that for each $\theta_\tau \in \Theta_\tau$,

$$d_\tau(\theta_\tau) := \int_\gamma d_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau)) d\mathbb{Q}(\gamma) = \int_\gamma \int_{\min[\rho_\tau(\gamma,\theta_\tau), x_\tau(\gamma)]}^{\max[\rho_\tau(\gamma,\theta_\tau), x_\tau(\gamma)]} |F_\gamma(g) - F_\gamma(x_\tau(\gamma))| dg d\mathbb{Q}(\gamma),$$

where the equality follows from Lemma 1. Note that $\mathbb{Q}(\cdot)$ is the probability measure defined on Γ that is selected by the investigator to suit the particular application in hand[2], and so the functional form of $d_\tau(\cdot)$ also depends on $\mathbb{Q}(\cdot)$. If $\rho_\tau(\cdot, \theta_\tau)$ differs from $x_\tau(\cdot)$, for each γ, $d_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau))$ becomes a distance bigger than zero, letting $d_\tau(\theta_\tau)$ be an average of the non-zero distances weighted by the adjunct probability measure $\mathbb{Q}$. Thus, the minimum value of $d_\tau(\cdot)$ can be viewed as the minimized weighted average of the distances. We now let $\theta_\tau^* := \arg\min_{\theta_\tau \in \Theta_\tau} d_\tau(\theta_\tau)$. If $\mathcal{M}_\tau$ is correctly specified, $\theta_\tau^* = \theta_\tau^0$ by noting that $d_\tau(\theta_\tau^0) = 0$ from the definition of $d_\tau(\cdot)$. Otherwise, we can view θ_τ^* as the parameter value that minimizes the quasi-check function, just as in quasi-maximum likelihood estimation. Second, our earlier discussion on $d_\tau(\cdot)$ can be extended by combining $d_\tau(\cdot)$ with $\mathcal{M}_\tau$, leading to estimation of θ_τ^* in a straightforward manner. Note that if we let $m_\tau(\gamma, u) := \mathbb{E}[\xi_\tau(G(\gamma) - u)] - \mathbb{E}[\xi_\tau(G(\gamma) - \rho_\tau(\gamma, \theta_\tau^*))]$, it follows that

$$q_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau)) = m_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau)) + \mathbb{E}[\xi_\tau(G(\gamma) - \rho_\tau(\gamma, \theta_\tau^*))].$$

By using this relationship and further letting $m_\tau(\theta_\tau) := \int_\gamma m_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau)) d\mathbb{Q}(\gamma)$, we define

$$q_\tau(\theta_\tau) := \int_\gamma q_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau)) d\mathbb{Q}(\gamma) = m_\tau(\theta_\tau) + \int_\gamma \mathbb{E}[\xi_\tau(G(\gamma) - \rho_\tau(\gamma, \theta_\tau^*))] d\mathbb{Q}(\gamma).$$

As for the optimization of $q_\tau(\gamma, \cdot)$, θ_τ is associated with only $m_\tau(\cdot)$ on the right side, so that we can obtain θ_τ^* by optimizing $q_\tau(\cdot)$ instead of $m_\tau(\cdot)$. That is, $\theta_\tau^* = \arg\min_{\theta_\tau \in \Theta_\tau} q_\tau(\theta_\tau)$.

We therefore estimate the unknown parameter θ_τ^* by first estimating $q_\tau(\cdot)$ and proceeding to minimize the function with respect to θ_τ. Specifically, for each $\theta_\tau \in \Theta_\tau$, define

$$q_{\tau n}(\theta_\tau) := \int_\gamma n^{-1} \sum_{i=1}^n \xi_\tau(G_i(\gamma) - \rho_\tau(\gamma, \theta_\tau)) d\mathbb{Q}(\gamma) \tag{2}$$

and let $\widehat{\theta}_{\tau n} := \arg\min_{\theta_\tau \in \Theta_\tau} q_{\tau n}(\theta_\tau)$. We call $\widehat{\theta}_{\tau n}$ the *functional quantile regression (FQR)* estimator if $\mathcal{M}_\tau$ is correctly specified; otherwise, $\widehat{\theta}_{\tau n}$ will be called the *quasi-functional quantile regression (QFQR)* estima-

[2]For instance, in the application associated with Fig 1, it might be of empirical interest to place greater emphasis on lower or higher income levels.

tor. Note that the sample average of the check functions in $q_{\tau n}(\cdot)$ is employed to estimate $q_\tau(\gamma, \rho_\tau(\gamma, \theta_\tau))$ consistently. Under the regularity conditions given in Assumptions 1, 2, and 3 below, $q_{\tau n}(\cdot)$ is consistent for $q_\tau(\cdot)$. Therefore, if θ_τ^* is unique and $q_\tau(\cdot)$ is continuous on Θ_τ, the estimator $\widehat{\theta}_{\tau n}$ is consistent for θ_τ^* under some general regularity conditions on the model.

Assumption 2. *(i) For each $\theta_\tau \in \Theta_\tau$, $\rho_\tau(\,\cdot\,, \theta_\tau) : \Gamma \mapsto \mathbb{R}$ is $\mathcal{G}$–measurable, where Θ_τ is a compact and convex set in $\mathbb{R}^{c_\tau}$ $(c_\tau \in \mathbb{N})$; (ii) for each $\gamma \in \Gamma$, $\rho_\tau(\gamma, \,\cdot\,) \in \mathcal{C}^{(2)}(\Theta_\tau)$; (iii) for each $\theta_\tau \in \Theta_\tau$, $\rho_\tau(\,\cdot\,, \theta_\tau) \in \mathcal{L}_{ip}(\Gamma)$; and (iv) if we let $q_\tau(\theta_\tau) := \int_\gamma \int \xi_\tau\{g(\gamma) - \rho_\tau(\gamma, \theta_\tau)\} d\mathbb{P}(g(\gamma)) d\mathbb{Q}(\gamma)$, $\theta_\tau^* := \arg\min_{\theta_\tau} q_\tau(\theta_\tau)$ is unique and interior to Θ_τ.* □

Assumption 3. *For some $M_i \in L^2(\mathbb{P})$ and $M < \infty$, (i) $\sup_\gamma |G_i(\gamma)| \leq M_i$; (ii) $\sup_{(\gamma,\theta_\tau)} |\rho_\tau(\gamma, \theta_\tau)| \leq M$; (iii) for each $j = 1, 2, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau)} |(\partial/\partial\theta_{\tau j})\rho_\tau(\gamma, \theta_\tau)| \leq M$; and (iv) for each j and $j' = 1, 2, \ldots, c_\tau$, $\sup_{\theta_\tau} \left|(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau j'})\rho_\tau(\,\cdot\,, \theta_\tau)\right| \leq M$.* □

Assumption 2 gives conditions on the model $\mathcal{M}_\tau$ and Assumption 3 provides bound conditions for the functional observations and the model function that ensure regular behavior for (Q)FQR estimation. Under these conditions the (Q)FQR estimator is shown to be consistent for θ_τ^* and asymptotically normally distributed.

2.1 Estimation under possible misspecification

To analyze the QFQR estimator we use an asymptotic approximation of the functional quantile estimator. For each γ let the PDF of $G_i(\gamma)$ be $f_\gamma(\cdot)$ and apply the approximation approach in Oberhofer and Haupt (2016, p. 710) to obtain the following representation

$$\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^*) = -A_\tau^{*-1} \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^*)\} - \tau\right) d\mathbb{Q}(\gamma) + o_{\mathbb{P}}(1), \tag{3}$$

where $\nabla_{\theta_\tau} = \partial/\partial\theta_\tau$ and

$$A_\tau^* := \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*) f_\gamma(\rho_\tau(\gamma, \theta_\tau^*)) \nabla'_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*) d\mathbb{Q}(\gamma).$$

Although Oberhofer and Haupt (2016) assume a correctly specified model to make use of the results of Knight (1998), this approximation remains valid even when $\mathcal{M}_\tau$ is misspecified.

The approximation implies that the limit behavior of $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^*)$ is determined by the two factors in the leading term on the right side of (3). The matrix A_τ^* in the first factor involves only non random model components. For regular behavior of $\widehat{\theta}_{\tau n}$ it is necessary for A_τ^* to be positive definite, as assumed in Assumption 4 below. The limit distribution of the QFQR estimator is determined mainly by the other components on the right side of (3). From Assumptions 2 and 3 it trivially follows that $q_\tau(\cdot)$ satisfies the first-order condition $\nabla_{\theta_\tau} q_\tau(\theta^*_\tau) = 0$ at θ^*_τ, implying that

$$\mathbb{E}\left[\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^*_\tau)(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta^*_\tau)\} - \tau)d\mathbb{Q}(\gamma)\right] = 0.$$

Hence, letting $J_{\tau i} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^*_\tau)\,(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta^*_\tau)\} - \tau)\,d\mathbb{Q}(\gamma)$ and with $B^*_\tau := \mathbb{E}[J_{\tau i}J'_{\tau i}]$ positive definite, standard multivariate central limit theory (CLT) using the Cramér-Wold device yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} J_{\tau i} \overset{\mathrm{A}}{\sim} \mathcal{N}(0, B^*_\tau).$$

As $\mathcal{M}_\tau$ is possibly misspecified, for each γ, $\mathbb{E}[\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta^*_\tau)\}]$ is not necessarily identical to τ, although the first-order condition still has to hold.

Limit theory for the QFQR estimator is given below in Theorem 1, based on the following regularity conditions that are used in deriving the limit distribution under possible misspecification of $\mathcal{M}_\tau$.

Assumption 4. *(i) $\lambda_{\min}(A^*_\tau) > 0$, where $A^*_\tau := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^*_\tau) f_\gamma(\rho_\tau(\gamma,\theta^*_\tau))\nabla'_{\theta_\tau}\rho_\tau(\gamma,\theta^*_\tau)d\mathbb{Q}(\gamma)$; and (ii) $\lambda_{\min}(B^*_\tau) > 0$ where $B^*_\tau := \mathbb{E}[J_{\tau i}J'_{\tau i}]$ and $J_{\tau i} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^*_\tau)(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta^*_\tau)\} - \tau)d\mathbb{Q}(\gamma)$.* □

Using these and the earlier conditions, asymptotic theory for the estimator $\widehat{\theta}_{\tau n}$ is as follows.

Theorem 1. *Given Assumptions 1, 2, 3, and 4, if $\mathcal{M}_\tau$ is misspecified, $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta^*_\tau) \overset{\mathrm{A}}{\sim} \mathcal{N}(0, C^*_\tau)$, where $C^*_\tau := A^{*-1}_\tau B^*_\tau A^{*-1}_\tau$.* □

The distribution of the QFQR estimator is therefore asymptotically normal with a variance matrix that has a sandwich-form, as expected in a misspecified case. The matrix B^*_τ is the variance matrix of $J_{\tau i}$ and must be estimated consistently to enable inference, which is discussed later in Section 4.

2.2 Estimation under correct specification

When $\mathcal{M}_\tau$ is correctly specified the asymptotic theory given in Theorem 1 remains applicable and relevant to FQR estimation. But it is useful to provide an explicit derivation and representation of the limit theory, which can be obtained using functional central limit theory (FCLT).

For each γ, let $F_\gamma(\cdot)$ be the marginal CDF of $G_i(\gamma)$ so that $F_\gamma(\rho_\tau(\gamma,\theta^0_\tau)) = \tau$. It follows that $(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta^0_\tau)\} - \tau) = (\mathbb{1}\{U_i(\gamma) \le \tau\} - \tau)$, where for each γ, $U_i(\gamma) := F_\gamma(G_i(\gamma))$. Note that $U_i(\gamma)$ is the probability integral transformation (PIT) of $G_i(\gamma)$, so that for each γ, $U_i(\gamma)$ follows a standard uniform distribution, implying that $\mathbb{E}[\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta^0_\tau)\}] = \tau$. Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{U_i(\gamma) \le \tau\} \overset{\mathbb{P}}{\to} \tau \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{1}\{U_i(\gamma) \le \tau\} - \tau) \overset{\mathrm{A}}{\sim} \mathcal{N}(0, \tau(1-\tau))$$

by the law of large numbers (LLN) and CLT, respectively. The limit theory is strengthened by use of the functional limit result of the following lemma.

Lemma 2. *Given Assumptions 1 and 2, if $\mathcal{M}_\tau$ is correctly specified, $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot,\theta_\tau^0)\} - \tau) \Rightarrow \mathcal{G}_\tau(\cdot)$, where $\mathcal{G}_\tau(\cdot)$ is a zero-mean Gaussian process such that for each γ and $\gamma' \in \Gamma$, $\mathbb{E}[\mathcal{G}_\tau(\gamma)\mathcal{G}_\tau(\gamma')]$ $= \kappa_\tau(\gamma,\gamma') := \mathbb{E}\left[\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\gamma') \leq \tau\}\right] - \tau^2$.* □

Therefore, from Lemma 2, it follows that

$$\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0) n^{-1/2}\sum_{i=1}^n \left(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^0)\} - \tau\right) d\mathbb{Q}(\gamma) \Rightarrow \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathcal{G}_\tau(\gamma) d\mathbb{Q}(\gamma),$$

by applying continuous mapping. Note that the weak limit function has the same normal distribution as given in Theorem 1. Lemma 2 is straightforwardly proved if $n^{-1/2}\sum_{i=1}^n \left(\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot,\theta_\tau^0)\} - \tau\right)$ is stochastically equicontinuous, thereby enabling use of the FCLT (e.g., Billingsley, 1999; Pollard, 1984). Andrews (1994) provides sufficient conditions for stochastic equicontinuity for various types of functions that apply Ossiander's L^2 entropy condition. Indeed, in the present case it is sufficient to apply example 2 in Andrews (1994, p. 2279) to show that the random function in Lemma 2 satisfies Ossiander's L^2 entropy condition.

Limit theory for the FQR estimator is given in Theorem 2 based on the following regularity conditions.

Assumption 5. *(i) $\lambda_{\min}(A_\tau^0) > 0$, where $A_\tau^0 := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0) f_\gamma(\rho_\tau(\gamma,\theta_\tau^0))\nabla'_{\theta_\tau^0}\rho_\tau(\gamma,\theta_\tau^0)d\mathbb{Q}(\gamma)$ and θ_τ^0 is such that $\rho_\tau(\cdot,\theta_\tau^0) = x_\tau(\cdot)$ and for each γ, $x_\tau(\gamma)$ denotes the τ-th quantile level of $G_i(\gamma)$; and (ii) $\lambda_{\min}(B_\tau^0) > 0$ where we let $B_\tau^0 := \int_{\gamma'}\int_\gamma \nabla_\theta\rho_\tau(\gamma,\theta_\tau^0)\kappa_\tau(\gamma,\gamma')\nabla'_\theta\rho_\tau(\widetilde{\gamma},\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')$ and $\kappa_\tau(\gamma,\gamma') :=$ $\mathbb{E}[\mathbb{1}\{F_\gamma(G_i(\gamma)) \leq \tau\}\mathbb{1}\{F_{\gamma'}(G_i(\gamma')) \leq \tau\}] - \tau^2$.* □

Theorem 2. *Given Assumptions 1, 2, 3, and 5, if $\mathcal{M}_\tau$ is correctly specified, $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^0) \overset{\text{A}}{\sim} \mathcal{N}(0, C_\tau^0)$, where $C_\tau^0 := A_\tau^{0\,-1}B_\tau^0 A_\tau^{0\,-1}$.* □

The FQR limit theory is obtained in a different way from that of QFQR, even though the former specializes to give the result under correct specification. Since $\theta_\tau^* = \theta_\tau^0$ under correction model specification it follows that $A_\tau^* = A_\tau^0$. Further, the matrix B_τ^0 is obtained from the covariance kernel of $\mathcal{G}_\tau(\cdot)$, implying that B_τ^0 can be consistently estimated by first estimating the kernel function $\kappa_\tau(\cdot,\cdot)$. This approach is discussed later in Section 4.

3 Estimation with Nuisance Effects

For practical application it is useful to extend the functional data structure by allowance for estimation errors in the observations. More specifically, functional data are often affected by parameter estimation errors, as illustrated in the examples of Cho, Phillips, and Seo (2022). The limit theory in both FQR and QFQR estimation is affected by such estimation errors.

To allow for such measurement errors, let $\widehat{G}(\cdot)$ $(\in \mathbb{R})$ be a continuous random function defined on the same Γ as before but such that $\widehat{G}(\cdot) := G(\cdot,\widehat{\pi}_n)$, where $\widehat{\pi}_n$ $(\in \mathbb{R}^s)$ denotes a set of nuisance parameters such that for positive definite matrices $P^* \in \mathbb{R}^{s\times s}$ and $H^* \in \mathbb{R}^{s\times s}$, $\widehat{\pi}_n$ is a consistent estimator of some $\pi^* \in \Pi \subset \mathbb{R}^s$ ($s \in \mathbb{N}$) with

$$\sqrt{n}(\widehat{\pi}_n - \pi^*) = -P^{*-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i + o_{\mathbb{P}}(1) \overset{\text{A}}{\sim} \mathcal{N}\left(0, P^{*-1}H^*P^{*-1}\right).$$

Many standard procedures have estimators with this general property and associated asymptotic normal limit theory, including least squares, two-stage least squares, maximum likelihood, and GMM; and functional observations are often generated in such a manner with parameter estimation errors, as discussed in Cho, Phillips, and Seo (2022). Functional observations of the type in Section 2 can be regarded as data with no nuisance effects induced by parameter estimation by letting $G_i(\cdot) = G_i(\cdot, \pi^*)$.

Before proceeding, we augment the regularity and probability space condition to include nuisance effects.

Assumption 6. *(i) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, $\Gamma \subset \mathbb{R}^g$ ($g \in \mathbb{N}$) be a compact metric space, and $\Pi \subset \mathbb{R}^s$ ($s \in \mathbb{N}$) be compact; (ii) $\{\widetilde{G}_i : \Gamma \times \Pi \mapsto \mathbb{R}\}_{i=1}^n$ is a set of iid observations such that (ii.a) for each $(\gamma, \pi) \in \Gamma \times \Pi$, $\widetilde{G}_i(\gamma, \pi)$ is $\mathcal{F}$-measurable; (ii.b) for each $\pi \in \Pi$, $\widetilde{G}_i(\,\cdot\,, \pi) \in \mathcal{L}_{ip}(\Gamma)$ for all $\omega \in F \in \mathcal{F}$ with $\mathbb{P}(F) = 1$; (ii.c) for each $\gamma \in \Gamma$, $\widetilde{G}_i(\gamma, \,\cdot\,)$ is in $\mathcal{C}^{(1)}(\Pi)$ for all $\omega \in F \in \mathcal{F}$ with $\mathbb{P}(F) = 1$; (iii) $(\Gamma, \mathcal{G}, \mathbb{Q})$ and $(\Omega \times \Gamma, \mathcal{F} \otimes \mathcal{G}, \mathbb{P} \cdot \mathbb{Q})$ are complete probability spaces and for $i = 1, 2, \ldots$ and $\pi \in \Pi$, $\widetilde{G}_i(\,\cdot\,, \pi)$ is $\mathcal{F} \otimes \mathcal{G}$-measurable; and (iv) for each γ, $F_\gamma(\cdot) \in \mathcal{C}^{(1)}(\mathbb{R})$, and $f_\gamma(\cdot) \in \mathcal{L}_{ip}(\mathbb{R})$ and uniformly bounded, where for some π^*, $F_\gamma(\cdot)$ and $f_\gamma(\cdot)$ are the CDF and PDF of $\widetilde{G}_i(\gamma, \pi^*)$, respectively.* □

Assumption 7. *There exists a sequence of measurable functions $\{\widehat{\pi}_n : \Omega \mapsto \Pi\}$ such that (i) $\widehat{\pi}_n \to \pi^*$ a.s.$-\mathbb{P}$, where π^* is an interior element in Π; (ii) for a nonstochastic finite $s \times s$ matrix P^* such that $\lambda_{\min}(P^*) > 0$ and a sequence of $\mathcal{F}$-measurable random vectors $\{S_{n*}\}$, $\sqrt{n}(\widehat{\pi}_n - \pi^*) = -P^{*-1}n^{1/2}S_n^* + o_{\mathbb{P}}(1)$; and (iii) for $i = 1, 2, \ldots$, there is $S_i : \Omega \mapsto \mathbb{R}^s$ such that (iii.a) S_i is $\mathcal{F}$-measurable and iid; (iii.b) $\sqrt{n}S_n^* = n^{-1/2}\sum_{i=1}^n S_i + o_{\mathbb{P}}(1)$; and (iii.b) for some $M_i \in L^2(\mathbb{P})$ and for each $j = 1, \ldots, s$, $|S_{ij}| \leq M_i$, where S_{ij} is the j-th row element of S_i.* □

Assumption 8. *For some $M_i \in L^2(\mathbb{P})$ and $M < \infty$, (i) $\sup_{(\gamma,\pi)} |\widetilde{G}_i(\gamma, \pi)| \leq M_i$; (ii) $\sup_j \sup_{(\gamma,\pi)} |(\partial/\partial\pi_j)\widetilde{G}_i(\gamma, \pi)| \leq M_i$; (iii) $\sup_{(\gamma,\theta_\tau)} |\rho_\tau(\gamma, \theta_\tau)| \leq M$; (iv) for each $j = 1, 2, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau,\pi)} |(\partial/\partial\theta_{\tau j})\rho_\tau(\gamma, \theta_\tau, \pi)| \leq M$; (v) for each $j, j' = 1, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau,\pi)} |(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau j'})\rho_\tau(\gamma, \theta_\tau, \pi)| \leq M$; and (vi) for each $j = 1, 2, \ldots, s$, $\mathbb{E}[(\partial/\partial\pi_j)\widetilde{G}_i(\cdot, \pi^*)] \in \mathcal{L}_{ip}(\Gamma)$.* □

Assumptions 6 and 7 are included to allow for functional data featuring nuisance effects, characterizing the implications of the parametric estimates $\widehat{\pi}_n$ embodied in the functional observations; and Assumption 8 generalizes the bound condition of Assumption 3 to accommodate nuisance parameter estimation.

The quantile function can be estimated in parallel to (Q)FQR estimation using the same model $\mathcal{M}_\tau$, allowing for possible misspecification. In particular, the quantile function is obtained by minimizing the following function, which embodies the estimated curves $\widehat{G}(\cdot) := G(\cdot, \widehat{\pi}_n)$, so that for each $\theta_\tau \in \Theta_\tau$,

$$\widehat{q}_{\tau n}(\theta_\tau) := \int_\gamma \frac{1}{n}\sum_{i=1}^{n} \xi_\tau\{\widehat{G}_i(\gamma) - \rho_\tau(\gamma, \theta_\tau)\} d\mathbb{Q}(\gamma),$$

giving $\widetilde{\theta}_{\tau n} := \arg\min_{\theta_\tau \in \Theta_\tau} \widehat{q}_{\tau n}(\theta_\tau)$. We call $\widetilde{\theta}_{\tau n}$ the *two-stage functional quantile regression (TSFQR)* estimator in the case of a correctly specified model $\mathcal{M}_\tau$; otherwise, $\widetilde{\theta}_{\tau n}$ is called the *two-stage quasi-functional quantile regression (TSQFQR)* estimator. The only difference between $\widehat{q}_{\tau n}(\cdot)$ and $q_{\tau n}(\cdot)$ is in the fact that $\widehat{q}_{\tau n}(\cdot)$ is obtained using the functional observations $\widehat{G}(\cdot) := G(\cdot, \widehat{\pi}_n)$ that embody nuisance effects.

3.1 Estimation under possible model misspecification

As in Section 2.1, suppose that the model $\mathcal{M}_\tau$ may be misspecified. The asymptotic properties of the TSQFQR estimator can be derived in a similar fashion to those of the QFQR estimator. In particular, approximating the TSQFQR estimation error as in (3) by using the Oberhofer and Haupt (2016, p. 710) approach, it follows that for each γ,

$$\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^*) = -A_\tau^{*-1} \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^*) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma, \theta_\tau^*)\} - \tau\Big) d\mathbb{Q}(\gamma) + o_{\mathbb{P}}(1). \quad (4)$$

As before the limit theory is determined by the two factors in the leading component on the right side of (4). The matrix A_τ^* is the same as in QFQR estimation but without the nuisance effect; and the second factor determines the limit distribution theory. In view of Assumptions 2 and 8, first-order conditions hold for $q_\tau(\cdot)$ at θ_τ^*, so that $\nabla_{\theta_\tau} q_\tau(\theta_\tau^*) = 0$ and

$$\mathbb{E}\left[\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^*) \Big(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma, \theta_\tau^*)\} - \tau\Big) d\mathbb{Q}(\gamma)\right] = 0.$$

Next, let $\widehat{J}_{\tau i} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^*)(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma, \theta_\tau^*)\} - \tau) d\mathbb{Q}(\gamma)$ and suppose that its asymptotic covariance matrix, denoted by $\widetilde{B}_\tau^*$, is positive definite. Then by standard multivariate central limit theory

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widehat{J}_{\tau i} \overset{\mathrm{A}}{\sim} \mathcal{N}(0, \widetilde{B}_\tau^*).$$

In the present case, $\mathbb{E}[\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma, \theta_\tau^*)\}]$ is not necessarily identical to τ for each γ because $\mathcal{M}_\tau$ may be misspecified. Nevertheless, its integral is necessarily associated with τ by virtue of the first-order conditions for θ_τ^*; and this property leads to the asymptotic normal distribution even when $\mathcal{M}_\tau$ is misspecified.

The limit distributions with and without nuisance effects differ, mainly because of differences between the variance matrices B_τ^* and $\widetilde{B}_\tau^*$. Applying the mean-value theorem, for each γ and some $\bar{\pi}_{\gamma n}$ between π^*

and $\widehat{\pi}_n$, we have

$$\sum_{i=1}^{n} \widehat{J}_{\tau i} = \int_{\gamma} \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*) \sum_{i=1}^{n} (\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) + \nabla'_\pi \widetilde{G}_i(\gamma, \bar{\pi}_{\gamma n})(\widehat{\pi}_n - \pi^*) \leq \rho_\tau(\gamma, \theta_\tau^*)\} - \tau) d\mathbb{Q}(\gamma).$$

If we further let $K_\tau^* := \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*) f_\gamma(\rho_\tau(\gamma, \theta_\tau^*)) \mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma, \pi^*)] d\mathbb{Q}(\gamma)$, we can rewrite this equation as follows:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{J}_{\tau i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (J_{\tau i} - K_\tau^* P^{*-1} S_i) + o_{\mathbb{P}}(1)$$

by applying Assumptions 6, 7, and 8 which suffice to apply the uniform law of large numbers (ULLN). More specifically, $n^{-1} \sum_{i=1}^{n} \nabla_\pi \widetilde{G}_i(\cdot, \pi^*) \stackrel{\mathbb{P}}{\to} \mathbb{E}[\nabla_\pi \widetilde{G}_i(\cdot, \pi^*)]$ uniformly on Γ. Using this property, if the multivariate CLT applies, the asymptotic covariance matrix of $n^{-1/2} \sum_{i=1}^{n} \widehat{J}_{\tau i}$ is obtained as

$$\widetilde{B}_\tau^* := B_\tau^* - \mathbb{E}[J_{\tau i} S_i'] P^{*-1} K_\tau^{*\prime} - K_\tau^* P^{*-1} \mathbb{E}[S_i J_{\tau i}'] + K_\tau^* P^{*-1} H^* P^{*-1} K_\tau^{*\prime}$$

by noting that $B_\tau^* := \mathbb{E}[J_{\tau i} J_{\tau i}']$, where $H^* := \mathbb{E}[S_i S_i']$. That is, $\mathbb{E}[\widehat{J}_{\tau i} \widehat{J}_{\tau i}'] = \widetilde{B}_\tau^* + o_{\mathbb{P}}(1)$. Note that $\widetilde{B}_\tau^*$ differs from B_τ^* mainly due to the nuisance effects. In the absence of nuisance effects simply set $S_i \equiv 0$, which leads to $\widetilde{B}_\tau^* := B_\tau^*$.

The limit theory for TSQFQR estimation is based on the following additional regularity conditions, which ensure that the limit distribution of the TSQFQR estimator is non-degenerate.

Assumption 9. *(i) $\lambda_{\min}(A_\tau^*) > 0$; (ii) $\lambda_{\min}(L_\tau^*) > 0$; and (iii) $\lambda_{\min}(\widetilde{B}_\tau^*) > 0$, where*

$$L_\tau^* := \begin{bmatrix} H^* & V_\tau^{*\prime} \\ V_\tau^* & B_\tau^* \end{bmatrix},$$

$H^* := \mathbb{E}[S_i S_i']$, and $V_\tau^* := \mathbb{E}[J_{\tau i} S_i']$. □

By virtue of Assumption 9, S_i, $J_{\tau i}$ and $\widehat{J}_{\tau i}$ all have positive definite variance matrices.

Theorem 3. *Given Assumptions 2, 6, 7, 8, and 9, if $\mathcal{M}_\tau$ is misspecified, $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^*) \stackrel{\mathrm{A}}{\sim} \mathcal{N}(0, \widetilde{C}_\tau^*)$, where $\widetilde{C}_\tau^* := A_\tau^{*-1} \widetilde{B}_\tau^* A_\tau^{*-1}$.* □

3.2 Estimation under correct model specification

As in Section 2.2, the limit theory of the TSQFQR estimator continues to apply for TSFQR estimation but is convenient to analyze using functional central limit theory in a similar fashion as for FQR estimation.

Specifically, applying the mean-value theorem, for each γ and some $\bar{\pi}_{\gamma n}$ between π^* and $\widehat{\pi}_n$, it follows that

$$\begin{aligned}&\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\sum_{i=1}^n \Big(\mathbb{1}\{\widehat{G}_i(\gamma)\le \rho_\tau(\gamma,\theta_\tau^0)\}-\tau\Big)\,d\mathbb{Q}(\gamma) \qquad (5)\\ &\quad = \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)+\nabla_\pi'\widetilde{G}_i(\gamma,\bar{\pi}_{\gamma n})(\widehat{\pi}_n-\pi^*)\le \rho_\tau(\gamma,\theta_\tau^0)\}-\tau\Big)\,d\mathbb{Q}(\gamma).\end{aligned}$$

The component (5) differs from the corresponding component in (3) because nuisance effects are accommodated in the indicator function. Here, for each τ, $n^{-1}\sum_{i=1}^n \mathbb{1}\{\widehat{G}_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^0)\}\overset{\mathbb{P}}{\to}\tau$ by the LLN from the fact that $\mathcal{M}_\tau$ is correctly specified, and for each γ, $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{\widehat{G}_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^0)\}-\tau)$ is asymptotically normal by the CLT, just as in FQR estimation. Nevertheless, the limit distribution of the TSFQR estimator differs from FQR estimation because of nuisance parameter estimation effects. To analyze these effects we first provide the limit behavior of $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{\widehat{G}_i(\cdot)\le\rho_\tau(\cdot,\theta_\tau^0)\}-\tau)$.

Lemma 3. *Given Assumptions 2, 6, 7, and 8, if $\mathcal{M}_\tau$ is correctly specified, $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{\widetilde{G}_i(\cdot,\pi^*)+\nabla_\pi'\widetilde{G}_i(\cdot,\bar{\pi}_{(\cdot)n})\,(\widehat{\pi}_n-\pi^*)\le\rho_\tau(\cdot,\theta_\tau^0)\}-\tau)\Rightarrow\widetilde{\mathcal{G}}_\tau(\cdot)$, where $\nabla_\pi'=\partial/\partial\pi'$ and $\widetilde{\mathcal{G}}_\tau(\cdot)$ is a zero-mean Gaussian process such that for each γ and γ', $\mathbb{E}[\widetilde{\mathcal{G}}_\tau(\gamma)\widetilde{\mathcal{G}}_\tau(\gamma')]=\widetilde{\kappa}_\tau(\gamma,\gamma')$ with*

$$\begin{aligned}\widetilde{\kappa}_\tau(\gamma,\gamma') :=&\kappa_\tau(\gamma,\gamma') - f_\gamma(\rho_\tau(\gamma,\theta_\tau^0))\mathbb{E}[\nabla_\pi'\widetilde{G}_i(\gamma,\pi^*)]P^{*-1}\mathbb{E}[S_i(\mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*)<\rho_\tau(\gamma',\theta_\tau^0)\}-\tau)]\\ &- f_{\gamma'}(\rho_\tau(\gamma',\theta_\tau^0))\mathbb{E}[\nabla_\pi'\widetilde{G}_i(\gamma',\pi^*)]P^{*-1}\mathbb{E}[S_i(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)<\rho_\tau(\gamma,\theta_\tau^0)\}-\tau)]\\ &+ f_{\gamma'}(\rho_\tau(\gamma',\theta_\tau^0))\mathbb{E}[\nabla_\pi'\widetilde{G}_i(\gamma',\pi^*)]P^{*-1}H^*P^{*-1}\mathbb{E}[\nabla_\pi\widetilde{G}_i(\gamma,\pi^*)]f_\gamma(\rho_\tau(\gamma,\theta_\tau^0)),\end{aligned}$$

and for each γ, $f_\gamma(\cdot)$ is the marginal PDF of $\widetilde{G}_i(\gamma,\pi^*)$, as before. □

This limit theory involves the Gaussian stochastic process $\widetilde{\mathcal{G}}_\tau(\cdot)$ as for $\mathcal{G}_\tau(\cdot)$ given in Lemma 2, but the covariance kernels differ. If π^* were known, there would be no need to approximate $\widehat{G}_i(\cdot)$ by the mean-value theorem with respect to π, so that $S_i\equiv 0$ and the covariance kernel $\widetilde{\kappa}_\tau(\cdot,\cdot)$ would be identical in form to that of $\kappa_\tau(\cdot,\cdot)$. Furthermore, if $\theta_\tau^0=\theta_\tau^*$, the asymptotic variance matrix of $\widehat{J}_{\tau i}$ is identical to $\int_\gamma\int_{\gamma'}\nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)\widetilde{\kappa}_\tau(\gamma,\gamma')\nabla_{\theta_\tau}'\rho_\tau\ (\gamma,\theta_\tau^*)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')$, so that $\widetilde{B}_\tau^*$ can be estimated by first estimating the covariance kernel $\widetilde{\kappa}_\tau(\cdot,\cdot)$ as detailed in Section 4.

Lemma 3 is established by showing stochastic equicontinuity, as in Lemma 2. In particular, we derive the covariance kernel of $\widetilde{\mathcal{G}}_\tau(\cdot)$ by separating the nuisance effects from $\widehat{G}_i(\cdot)$ in the indicator function. Setting $\widehat{\mu}_{ni}(\gamma):=\nabla_\pi'G_i(\gamma,\bar{\pi}_{\gamma n})(\widehat{\pi}_n-\pi^*)$, we have

$$\begin{aligned}&\mathbb{1}\{G_i(\gamma,\pi^*)+\nabla_\pi'G_i(\gamma,\bar{\pi}_{\gamma n})(\widehat{\pi}_n-\pi^*)\le\rho_\tau(\gamma,\theta_\tau^0)\}-\tau=\mathbb{1}\{G_i(\gamma,\pi^*)\le\rho_\tau(\gamma,\theta_\tau^0)\}-\tau\\ &\quad+\mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^0)<G_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^0)-\widehat{\mu}_{ni}(\gamma)\}-\mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^0)-\widehat{\mu}_{ni}(\gamma)<G_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^0)\}. \qquad (6)\end{aligned}$$

Lemma 2 applies to the first term on the right side, converging weakly to $\mathcal{G}_\tau(\cdot)$. But the second and third terms of (6) still affect the weak limit of the left side, making the covariance kernel of $\widetilde{\mathcal{G}}_\tau(\cdot)$ different from

$\kappa_\tau(\cdot,\cdot)$, leading to $\widetilde{\kappa}_\tau(\cdot,\cdot)$.

To obtain an explicit limit distribution of the TSFQR estimator when $\mathcal{M}_\tau$ is correctly specified the following regularity conditions are employed. These conditions match those in Assumption 9 and are employed for the same reason.

Assumption 10. *(i) $\lambda_{\min}(A^0_\tau) > 0$; (ii) $\lambda_{\min}(L^0_\tau) > 0$; and (iii) $\lambda_{\min}(\widetilde{B}^0_\tau) > 0$, where*

$$L^0_\tau := \begin{bmatrix} H^* & V^{0\prime}_\tau \\ V^0_\tau & B^0_\tau \end{bmatrix},$$

$V^0_\tau := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^0_\tau)\mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \leq \rho_\tau(\gamma,\theta^0_\tau)\} - \tau)\cdot S'_i]d\mathbb{Q}(\gamma)$, $\widetilde{B}^0_\tau := \int_\gamma\int_{\gamma'} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^0_\tau)\widetilde{\kappa}_\tau(\gamma,\gamma')$ $\nabla'_{\theta_\tau}\rho_\tau(\gamma,\theta^0_\tau)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')$, and B^0_τ is defined in Assumption 5. □

Assumption 10 implies that S_i, $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^0_\tau)(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta^0_\tau)\} - \tau)d\mathbb{Q}(\gamma)$ and $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta^0_\tau)$ $(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma,\theta^0_\tau)\} - \tau)d\mathbb{Q}(\gamma)$ all have positive definite variance matrices.

Theorem 4. *Given Assumptions 2, 6, 7, 8, and 10, if $\mathcal{M}_\tau$ is correctly specified, $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta^0_\tau) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^0_\tau)$, where $\widetilde{C}^0_\tau := A^{0\,-1}_\tau \widetilde{B}^0_\tau A^{0\,-1}_\tau$.* □

Even when $\mathcal{M}_\tau$ is correctly specified, the limit distribution of the TSFQR estimator differs from that of the FQR estimator in Theorem 1. The variance matrices $\widetilde{B}^0_\tau$ and B^0_τ differ due to the presence of the parameter estimation error introduced by $\widehat{\pi}_n$. Without nuisance parameter estimation, $S_i \equiv 0$, so that $\widetilde{B}^0_\tau = B^0_\tau$, leading to the same distribution for both $\widehat{\theta}_{\tau n}$ and $\widetilde{\theta}_{\tau n}$.

4 Asymptotic Variance Matrix Estimation

Just as in standard quantile regression the limit distributions of the (Q)FQR and TS(Q)FQR estimators and particularly their asymptotic variance matrices play a central role in performing inference with functional data. A key step in the construction of suitable statistics for testing and confidence interval construction is therefore consistent estimation of the matrices B^*_τ and $\widetilde{B}^*_\tau$, which is now discussed. These variance matrix estimates may be employed in conjunction with the limiting normal distributions of the parameter estimates to conduct inference in the usual manner.

Estimation of the variance matrices in the general case, allowing for model misspecification, is conveniently done by employing a plug-in approach. In this case, first let

$$J_{\tau ni} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})\left(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\widehat{\theta}_{\tau n})\} - \tau\right)d\mathbb{Q}(\gamma) \quad \text{and}$$

$$\widehat{J}_{\tau ni} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\left(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\} - \tau\right)d\mathbb{Q}(\gamma)$$

and estimate B_τ^* by $\widehat{B}_{\tau n} := n^{-1}\sum_{i=1}^n J_{\tau ni}J_{\tau ni}'$ and $\widetilde{B}_\tau^*$ by

$$\widetilde{B}_{\tau n} := \frac{1}{n}\sum_{i=1}^n \widehat{J}_{\tau ni}\widehat{J}_{\tau ni}' - \widehat{V}_{\tau n}\widehat{P}_n^{-1}\widehat{K}_{\tau n}' - \widehat{K}_{\tau n}\widehat{P}_n^{-1}\widehat{V}_{\tau n}' + \widehat{K}_{\tau n}\widehat{P}_n^{-1}\widehat{H}_n\widehat{P}_n^{-1}\widehat{K}_{\tau n}',$$

where

$$\widehat{V}_{\tau n} := \frac{1}{n}\sum_{i=1}^n \widehat{J}_{\tau ni}S_i' \quad \text{and} \quad \widehat{K}_{\tau n} := \frac{1}{n}\sum_{i=1}^n \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\widehat{f}_{\gamma n}(\rho_\tau(\gamma,\widetilde{\theta}_{\tau n}))\nabla_\pi'\widetilde{G}_i(\gamma,\widehat{\pi}_n)d\mathbb{Q}(\gamma).$$

Here, $\widehat{H}_n$, $\widehat{P}_n$, and $\widehat{f}_{\gamma n}(\cdot)$ are consistent estimators respectively for H^*, P^*, and $f_\gamma(\cdot)$, as is assumed by the following condition.

Assumption 11. *(i) For a sequence of measurable random variables $\{\widehat{H}_n \in \mathbb{R}^{s\times s}\}$, $\widehat{H}_n \xrightarrow{\mathbb{P}} H^*$; (ii) for a sequence of measurable random variables $\{\widehat{P}_n \in \mathbb{R}^{s\times s}\}$, $\widehat{P}_n \xrightarrow{\mathbb{P}} P^*$; and (iii) for a sequence of measurable functions $\{\widehat{f}_{\gamma n}(\cdot) : \mathbb{R} \mapsto \mathbb{R}\}$, $\widehat{f}_{\gamma n}(\cdot) \xrightarrow{\mathbb{P}} f_\gamma(\cdot)$ uniformly in γ.* □

Given these regularity conditions together with earlier assumptions, it is straightforward to show that $\widehat{V}_{\tau n}$ and $\widehat{K}_{\tau n}$ are consistent for V_τ and $K_\tau^* = \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)f_\gamma(\rho_\tau(\gamma,\theta_\tau^*))\mathbb{E}[\nabla_\pi'\widetilde{G}_i(\gamma,\pi^*)]d\mathbb{Q}(\gamma)$, as defined earlier. Hence, if both $n^{-1}\sum_{i=1}^n J_{\tau ni}J_{\tau ni}'$ and $n^{-1}\sum_{i=1}^n \widehat{J}_{\tau ni}\widehat{J}_{\tau ni}'$ are consistent for B_τ^*, it follows that $\widehat{B}_{\tau n}$ and $\widetilde{B}_{\tau n}$ are consistent for B_τ^* and $\widetilde{B}_\tau^*$, as in Theorem 5 below.

If $\mathcal{M}_\tau$ is correctly specified, then instead of estimating $J_{\tau i}$ and $\widehat{J}_{\tau i}$, the covariance matrices can be estimated from explicit forms of the covariance kernels in the functional law, which can then be used to estimate B_τ^0 and $\widetilde{B}_\tau^0$. With this approach we first estimate $\kappa_\tau(\gamma,\gamma')$ and $\widetilde{\kappa}_\tau(\gamma,\gamma')$ by

$$\widehat{\kappa}_{\tau n}(\gamma,\gamma') := \frac{1}{n}\sum_{i=1}^n \left(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\widehat{\theta}_{\tau n})\} - \tau\right)\left(\mathbb{1}\{G_i(\gamma') \le \rho_\tau(\gamma',\widehat{\theta}_{\tau n})\} - \tau\right) \quad \text{and}$$

$$\widetilde{\kappa}_{\tau n}(\gamma,\gamma') := \frac{1}{n}\sum_{i=1}^n \left(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\} - \tau\right)\left(\mathbb{1}\{\widehat{G}_i(\gamma') \le \rho_\tau(\gamma',\widetilde{\theta}_{\tau n})\} - \tau\right).$$

Then, for each $\gamma \in \Gamma$, we let $\zeta_\tau(\gamma) := \mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le \rho_\tau(\gamma,\theta_\tau^0)\} - \tau)S_i']$, which is estimated by its sample analog

$$\widetilde{\zeta}_{\tau n}(\gamma) := \frac{1}{n}\sum_{i=1}^n (\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\} - \tau)S_i'.$$

These estimators are pointwise consistent for their respective target quantities by the LLN and continuous mapping. This property can be strengthened by applying the ULLN. Indeed, under Assumptions 1, 2, and 3, it follows that $\widehat{\kappa}_{\tau n}(\cdot,\cdot)$ is consistent for $\kappa_\tau(\cdot,\cdot)$ uniformly on $\Gamma\times\Gamma$ by the ULLN. Likewise, $\widetilde{\kappa}_{\tau n}(\cdot,\cdot)$ and $\widetilde{\zeta}_{\tau n}(\cdot)$ turn out to be consistent for $\kappa_\tau(\cdot,\cdot)$ and $\zeta_\tau(\cdot)$, respectively under Assumptions 2, 6, 7, and 8.

Therefore, if we let

$$\widehat{B}^{\sharp}_{\tau n} := \int_{\gamma}\int_{\gamma'} \nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})\widehat{\kappa}_{\tau n}(\gamma,\gamma')\nabla'_{\theta_\tau}\rho_\tau(\gamma',\widehat{\theta}_{\tau n})d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma') \quad \text{and}$$

$$\begin{aligned}\widetilde{B}^{\sharp}_{\tau n} := &\int_{\gamma}\int_{\gamma'} \nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\widetilde{\kappa}_{\tau n}(\gamma,\gamma')\nabla'_{\theta_\tau}\rho_\tau(\gamma',\widetilde{\theta}_{\tau n})d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma') + \widehat{K}_{\tau n}\widehat{P}_n^{-1}\widehat{H}_n\widehat{P}_n^{-1}\widehat{K}'_{\tau n}\\ &- \int_{\gamma}\nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\widetilde{\zeta}_{\tau n}(\gamma)d\mathbb{Q}(\gamma)\widehat{P}_n^{-1}\widehat{K}'_{\tau n} - \widehat{K}_{\tau n}\widehat{P}_n^{-1}\int_{\gamma}\widetilde{\zeta}_{\tau n}(\gamma)'\nabla'_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})d\mathbb{Q}(\gamma)\end{aligned}$$

be variance matrix estimators for B^0_τ and $\widetilde{B}^0_\tau$, these matrices are consistent for B^0_τ and $\widetilde{B}^0_\tau$. Note that $\widehat{B}^{\sharp}_{\tau n}$ and $\widetilde{B}^{\sharp}_{\tau n}$ are simply sample analogs of B^0_τ and $\widetilde{B}^0_\tau$, so that consistency of these estimates follows directly from the consistency of $\widehat{\kappa}_{\tau n}(\cdot,\cdot)$, $\widetilde{\kappa}_{\tau n}(\cdot,\cdot)$, and $\widetilde{\zeta}_{\tau n}(\cdot)$. The formal result is given in the following theorem.

Theorem 5. *(i) Given Assumption 1, 2, 3, (i.a) if $\mathcal{M}_\tau$ is misspecified and Assumption 4 holds, $\widehat{B}_{\tau n} \xrightarrow{\mathbb{P}} B^*_\tau$; and (i.b) if $\mathcal{M}_\tau$ is correctly specified and Assumption 5 holds, $\widehat{B}^{\sharp}_{\tau n} \xrightarrow{\mathbb{P}} B^0_\tau$;*

(ii) Given Assumption 2, 6, 7, 8, 11, (ii.a) if $\mathcal{M}_\tau$ is misspecified and Assumption 9 holds, $\widetilde{B}_{\tau n} \xrightarrow{\mathbb{P}} \widetilde{B}^*_\tau$; (ii.b) if $\mathcal{M}_\tau$ is correctly specified and Assumption 10 holds, $\widetilde{B}^{\sharp}_{\tau n} \xrightarrow{\mathbb{P}} \widetilde{B}^0_\tau$. □

Two remarks are in order. First, $\widehat{B}_{\tau n}$ (resp. $\widetilde{B}_{\tau n}$) is numerically identical to $\widehat{B}^{\sharp}_{\tau n}$ (resp. $\widetilde{B}^{\sharp}_{\tau n}$), but the definition of $\widehat{B}^{\sharp}_{\tau n}$ (resp. $\widetilde{B}^{\sharp}_{\tau n}$) is conceptually different from that of $\widehat{B}_{\tau n}$ (resp. $\widetilde{B}_{\tau n}$), just as in the conceptual difference between B^*_τ and B^0_τ (resp. $\widetilde{B}^*_\tau$ and $\widetilde{B}^0_\tau$). Second, estimating $\zeta_\tau(\cdot)$ can be further involved when the score S_i depends on the nuisance parameter estimator π^*. For such a case, we can consistently estimate $\zeta_\tau(\cdot)$ by using $\widehat{\pi}_n$ in the construction of S_i. Consistency follows by applying the continuous mapping theorem under some mild regularity conditions.

5 Joint Estimation of Multiple Quantile Functions

The framework is now extended to allow for multiple quantile curve estimation from the same data, either $\{G_i(\cdot)\}_{i=1}^n$ or, in the case of nuisance effects, $\{\widehat{G}_i(\cdot)\}_{i=1}^n$. This extension allows for joint estimation and inference concerning quantile functions for multiple quantiles $\tau := (\tau_1, \tau_2, \ldots, \tau_p)'$. It is convenient to focus on $\widehat{G}_i(\cdot)$ and cover the case of functional data without nuisance effects in corollaries to the development.

The multiple quantile functions are estimated using the TS(Q)FQR procedure. For each $j = 1, 2, \ldots, p$, suppose model $\mathcal{M}_{\tau_j}$ is specified and the parameters $\theta^*_{\tau_j}$ are estimated by $\widetilde{\theta}_{\tau_j n}$ as before, letting $\widetilde{\theta}_n := (\widetilde{\theta}'_{\tau_1 n}, \widetilde{\theta}'_{\tau_2 n}, \ldots, \widetilde{\theta}'_{\tau_p n})'$ be the combined vector of separate TS(Q)FQR parametric estimators. We call $\widetilde{\theta}_n$ the *multiple two-stage functional quantile regression (MTSFQR)* estimator if, for each $j = 1, 2, \ldots, p$, the model $\mathcal{M}_{\tau_j}$ is correctly specified. Otherwise, we call $\widetilde{\theta}_n$ the *multiple two-stage quasi-functional quantile regression (MTSQFQR)* estimator. To develop MTS(Q)FQR asymptotics we define $\mathcal{M} := \bigcup_{j=1}^p \mathcal{M}_{\tau_j}$ as the multiple quantile function model and employ the following regularity conditions.

Assumption 12. *(i) For each* $\tau_j \in \{\tau_1, \ldots, \tau_p\}$ *and* $\theta_{\tau_j} \in \Theta_{\tau_j}$, $\rho_{\tau_j}(\,\cdot\,, \theta_{\tau_j}) : \Gamma \mapsto \mathbb{R}$ *is measurable* $-\mathcal{G}$, *where* Θ_{τ_j} *is a compact and convex set in* $\mathbb{R}^{c_{\tau_j}}$ $(c_{\tau_j} \in \mathbb{N})$*; (ii) for each* $\tau_j \in \{\tau_1, \ldots, \tau_p\}$ *and* $\gamma \in \Gamma$, $\rho_{\tau_j}(\gamma, \,\cdot\,) \in \mathcal{C}^{(2)}(\Theta_{\tau_j})$*; (iii) for each* $\tau_j \in \{\tau_1, \ldots, \tau_p\}$ *and* $\theta_{\tau_j} \in \Theta_{\tau_j}$, $\rho_{\tau_j}(\,\cdot\,, \theta_{\tau_j}) \in \mathcal{L}_{ip}(\Gamma)$*; and (iv) for each* $\tau_j \in \{\tau_1, \ldots, \tau_p\}$*, if we let* $q_{\tau_j}(\theta_{\tau_j}) := \int_\gamma \int \xi_{\tau_j}\{g(\gamma) - \rho_{\tau_j}(\gamma, \theta_{\tau_j})\} d\mathbb{P}(g(\gamma)) d\mathbb{Q}(\gamma)$, $\theta^*_{\tau_j} := \arg\min_{\theta_{\tau_j}} q_{\tau_j}(\theta_{\tau_j})$ *is unique and interior to* Θ_{τ_j}.

Assumption 13. *For some* $M_i \in L^2(\mathbb{P})$ *and* $M < \infty$*, (i)* $\sup_\gamma |G_i(\gamma)| \leq M_i$*; (ii) for each* $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |\rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$*; (iii) for each* $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$ *and* $j = 1, \ldots, c_{\tau_\ell}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |(\partial/\partial\theta_{\tau_\ell j})\, \rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$*; and (iv) for each* $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$ *and* $j, j' = 1, \ldots, c_{\tau_\ell}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |(\partial^2/\partial\theta_{\tau_\ell j}\partial\theta_{\tau_\ell j'}) \rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$. □

Assumption 14. *For some* $M_i \in L^2(\mathbb{P})$ *and* $M < \infty$*, (i)* $\sup_{(\gamma,\pi)} |\widetilde{G}_i(\gamma, \pi)| \leq M_i$*; (ii)* $\sup_j \sup_{(\gamma,\pi)} |(\partial/\partial\pi_j)\widetilde{G}_i(\gamma, \pi)| \leq M_i$*; (iii) for each* $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |\rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$*; (iv) for each* $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$ *and* $j = 1, \ldots, c_{\tau_\ell}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |(\partial/\partial\theta_{\tau_\ell j})\rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$*; (v) for each* $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$ *and* $j, j' = 1, \ldots, c_{\tau_\ell}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |(\partial^2/\partial\theta_{\tau_\ell j}\partial\theta_{\tau_\ell j'})\rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$*; and (vi) for each* $j = 1, 2, \ldots, s$, $\mathbb{E}[(\partial/\partial\pi_j)\widetilde{G}_i(\cdot, \pi^*)] \in \mathcal{L}_{ip}(\Gamma)$. □

Assumption 12 extends Assumption 2 to allow for the presence of multiple quantile functions. Importantly, the dimensions c_{τ_j} of the parametric components of the individual quantile function models $\mathcal{M}_j$ may differ, thereby allowing for different parametric model specifications at different quantile levels. Assumptions 13 and 14 extend Assumptions 3 and 8, ensuring that the regular bound conditions in Assumptions 3 and 8 apply to the case of multiple quantile function estimation.

5.1 Estimation under possible misspecification

Given a finite quantile number p, consistency of the full MTSQFQR estimator $\widetilde{\theta}_n$ follows directly from the consistency of the individual TSQFQR estimators $\widetilde{\theta}_{\tau_j n}$ of $\theta^*_{\tau_j}$ by joint convergence, so that $\widetilde{\theta}_n \overset{\mathbb{P}}{\to} \theta^* := (\theta^{*\prime}_{\tau_1}, \theta^{*\prime}_{\tau_2}, \ldots, \theta^{*\prime}_{\tau_p})'$. On the other hand, the joint limit distribution theory for $\widetilde{\theta}_n$ does not follow directly from the limit theory of the individual component estimators. Joint asymptotics are obtained by working explicitly with the full vector $\widetilde{\theta}_n$ and combining the individual asymptotic approximations in (4) as follows

$$\sqrt{n}(\widetilde{\theta}_n - \theta^*) = -A^{*-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{J}_i + o_{\mathbb{P}}(1), \tag{7}$$

where $\widehat{J}_i := [\widehat{J}_{\tau_1 i}, \widehat{J}_{\tau_2 i}, \ldots, \widehat{J}_{\tau_p i}]$, $A^* := \mathrm{diag}[A^*_{\tau_1}, A^*_{\tau_2}, \ldots, A^*_{\tau_p}]$, and for each $j = 1, 2, \ldots, p$, $A^*_{\tau_j} := \int_\gamma \nabla_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^*_{\tau_j}) f_\gamma(\rho_{\tau_j}(\gamma, \theta^*_{\tau_j})) \nabla'_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^*_{\tau_j}) d\mathbb{Q}(\gamma)$ as before. Here A^* is a $c \times c$ block diagonal square matrix and $\widehat{J}_i$ is a $c \times 1$ vector, where $c = \sum_{j=1}^{p} c_{\tau_j}$.

The limit distribution of the MTSQFQR estimator is obtained by examining the limit behaviors of the two factors in the leading term on the right side of (7). First, like the individual matrices $A^*_{\tau_j}$, A^* does not contain any stochastic components. In addition, if for each $j = 1, 2, \dots, p$, $A_{\tau_j^*}$ is positive definite, then so is A^* by construction. Thus, the limit distribution theory is effectively determined by the second factor in (7). Theorem 3 establishes that for each $j = 1, 2, \dots, p$, $\sqrt{n}(\widehat{\theta}_{\tau_j n} - \theta^*_{\tau_j}) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^*_{\tau_j})$, where $\widetilde{C}^*_{\tau_j} := A^{*-1}_{\tau_j} \widetilde{B}^*_{\tau_j} A^{*-1}_{\tau_j}$. Using a standard asymptotic argument for arbitrary linear combinations of all these components to produce a multivariate CLT, it then follows that $\sqrt{n}(\widetilde{\theta}_n - \theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^*)$, where $\widetilde{C}^* := A^{*-1}\widetilde{B}^* A^{*-1}$. The central matrix $\widetilde{B}^*$ in this sandwich form is a $c \times c$ matrix with submatrix in the j-th block row and j'-th block column matrix given by $\widetilde{B}^*_{\tau_j \tau_{j'}}$, which has the form

$$\widetilde{B}^*_{\tau_j \tau_{j'}} := B^*_{\tau_j \tau_{j'}} - \mathbb{E}[J_{\tau_j i} S_i'] P^{*-1} K^{*\prime}_{\tau_{j'}} - K^*_{\tau_j} P^{*-1} \mathbb{E}[S_i J'_{\tau_{j'} i}] + K^*_{\tau_j} P^{*-1} H^* P^{*-1} K^{*\prime}_{\tau_{j'}}$$

for each j and $j' = 1, 2, \dots, p$, where $B^*_{\tau_j \tau_{j'}} := \mathbb{E}[J_{\tau_j i} J'_{\tau_{j'} i}]$, and as before, for each $j = 1, 2, \dots, p$, $J_{\tau_j i} := \int_\gamma \nabla_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^*_{\tau_j})(\mathbb{1}\{G_i(\gamma) \leq \rho_{\tau_j}(\gamma, \theta^*_{\tau_j})\} - \tau_j) d\mathbb{Q}(\gamma)$. The limit distribution theory for the full MTSQFQR estimator $\widetilde{\theta}_n$ is then obtained based on the following regularity conditions:

Assumption 15. *(i)* $\lambda_{\min}(A^*) > 0$*; (ii)* $\lambda_{\min}(L^*) > 0$*; and (iii)* $\lambda_{\min}(\widetilde{B}^*) > 0$*, where*

$$L^* := \begin{bmatrix} H^* & V^{*\prime} \\ V^* & B^* \end{bmatrix},$$

$V^* := \mathbb{E}[J_i S_i']$, $J_i := [J'_{\tau_1 i}, J'_{\tau_2 i}, \dots, J'_{\tau_p i}]'$, *and* $B^* := \mathbb{E}[J_i J_i']$. □

It is now straightforward to derive the limit distribution of the MTSQFQR estimator using (7).

Theorem 6. *Given Assumptions 6, 7, 12, 14, and 15, if $\mathcal{M}$ is misspecified,* $\sqrt{n}(\widetilde{\theta}_n - \theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^*)$. □

Several remarks are warranted concerning Theorem 6. First, Theorem 6 specializes to Theorem 3 when $p = 1$ and if $j = j'$, $\widetilde{B}^*_{\tau_j \tau_{j'}}$ is identical to $\widetilde{B}^*_{\tau_j}$ in Section 4. Second, $\widetilde{\theta}_n$ can be obtained by minimizing a weighted sum of the check functions. That is, if $\{w_j\}$ is a set of strictly positive weights such that $\sum_{j=1}^p w_j \equiv 1$, $\widetilde{\theta}_n := \arg\min_{\theta \in \otimes_{j=1}^p \Theta_{\tau_j}} \widehat{q}_n(\theta)$, where $\widehat{q}_n(\theta) := \sum_{j=1}^p w_j \widehat{q}_{\tau_j n}(\theta_{\tau_j})$. Given that θ_{τ_j} is associated with only $\widehat{q}_{\tau_j n}(\cdot)$, minimizing $\widehat{q}_n(\cdot)$ is equivalent to minimizing the individual $\widehat{q}_{\tau_j n}(\cdot)$ and collecting the individual estimators to form $\widetilde{\theta}_n$. Third, if the functional data do not involve nuisance effects, we can estimate the unknown parameter θ^* by $\widehat{\theta}_n := \arg\min_{\theta \in \otimes_{j=1}^p \Theta_{\tau_j}} q_n(\theta)$, where $q_n(\theta) := \sum_{j=1}^p w_j q_{\tau_j n}(\theta_{\tau_j})$ for the same weights $\{w_j\}$. By applying this approach to derive the limit distribution of $\widetilde{\theta}_n$, it follows that $\sqrt{n}(\widehat{\theta}_n - \theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, A^{*-1} B^* A^{*-1})$, where B^* is a $c \times c$ matrix whose submatrix in the j-th block row and j'-th block column is $B^*_{\tau_j \tau_{j'}}$.

5.2 Estimation under correct specification

Theorem 6 may be used to deliver the limit theory of the MTSFQR estimator. But it is also useful to develop the asymptotics using functional limit law arguments. For this purpose we define some notation.

Let $\nabla_\theta \rho(\gamma, \theta^0) := \mathrm{diag}[\nabla_{\theta_{\tau_1}} \rho_{\tau_1}(\gamma, \theta^0_{\tau_1}), \nabla_{\theta_{\tau_2}} \rho_{\tau_2}(\gamma, \theta^0_{\tau_2}), \ldots, \nabla_{\theta_{\tau_p}} \rho_{\tau_p}(\gamma, \theta^0_{\tau_p})]$ and

$$\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho(\gamma, \theta^0)\} := [\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_{\tau_1}(\gamma, \theta^0_{\tau_1})\}, \cdots, \mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_{\tau_p}(\gamma, \theta^0_{\tau_p})\}]'.$$

These quantities are defined with some abuse of notation: $\nabla_\theta \rho(\gamma, \theta^0)$ is a $c \times p$ block diagonal matrix with $c_{\tau_j} \times 1$ column vectors in its diagonal blocks instead of a column vector; and $\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho(\gamma, \theta^0)\}$ is a $p \times 1$ column vector instead of a scalar.

The joint limit distribution of the MTSFQR estimator is efficiently obtained by collecting the asymptotic approximations in (4) into a vector. Specifically, (4) can be rewritten as

$$\sqrt{n}(\widetilde{\theta}_n - \theta^0) = -A^{0\,-1} \left(\int_\gamma \nabla_\theta \rho(\gamma, \theta^0) \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho(\gamma, \theta^0)\} - \tau \right) d\mathbb{Q}(\gamma) \right) + o_{\mathbb{P}}(1), \tag{8}$$

where $A^0 := \mathrm{diag}[A^0_{\tau_1}, A^0_{\tau_2}, \ldots, A^0_{\tau_p}]$. So the limit distribution of the MTSFQR estimator is determined by the two factors in the leading term on the right side of (8). The matrix A^0 is square and nonrandom just as A^*. It follows that the second factor is the main determinant of the limit distribution. Lemma 3 shows that each component in $n^{-1/2} \sum_{i=1}^n (\mathbb{1}\{\widehat{G}_i(\cdot) \leq \rho(\cdot, \theta^0)\} - \tau)$ weakly converges to a Gaussian stochastic process and the following lemma proves that the full vector converges weakly to a vector Gaussian process.

Lemma 4. *Given Assumptions 6, 7, 12, and 14,* $n^{-1/2} \sum_{i=1}^n (\mathbb{1}\{\widehat{G}_i(\cdot) \leq \rho(\cdot, \theta^0)\} - \tau) \Rightarrow \widetilde{\mathcal{G}}(\cdot) := [\widetilde{\mathcal{G}}_{\tau_1}(\cdot), \widetilde{\mathcal{G}}_{\tau_2}(\cdot), \cdots, \widetilde{\mathcal{G}}_{\tau_p}(\cdot)]'$*, where* $\widetilde{\mathcal{G}}(\cdot)$ *is a mean-zero Gaussian process such that for* j *and* $j' = 1, 2, \ldots, p$*, and* γ *and* $\gamma' \in \Gamma$*, the covariance kernel is*

$$\begin{aligned}
&\mathbb{E}[\widetilde{\mathcal{G}}_{\tau_j}(\gamma)\widetilde{\mathcal{G}}_{\tau_{j'}}(\gamma')] = \widetilde{\kappa}_{\tau_j\tau_{j'}}(\gamma, \gamma') \\
&\quad := \kappa_{\tau_j\tau_{j'}}(\gamma, \gamma') - f_\gamma(\rho_{\tau_j}(\gamma, \theta^0_{\tau_j}))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma, \pi^*)]P^{*-1}\mathbb{E}[S_i(\mathbb{1}\{\widetilde{G}_i(\gamma', \pi^*) \leq \rho_{\tau_{j'}}(\gamma', \theta^0_{\tau_{j'}})\} - \tau_{j'})] \\
&\qquad - f_{\gamma'}(\rho_{\tau_{j'}}(\gamma', \theta^0_{\tau_{j'}}))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma', \pi^*)]P^{*-1}\mathbb{E}[S_i(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \leq \rho_{\tau_j}(\gamma, \theta^0_{\tau_j})\} - \tau_j)] \\
&\qquad + f_\gamma(\rho_{\tau_j}(\gamma, \theta^0_{\tau_j}))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma, \pi^*)]P^{*-1}H^*P^{*-1}\mathbb{E}[\nabla_\pi \widetilde{G}_i(\gamma', \pi^*)]f_{\gamma'}(\rho_{\tau_{j'}}(\gamma, \theta^0_{\tau_{j'}})),
\end{aligned}$$

and $\kappa_{\tau_j\tau_{j'}}(\gamma, \gamma') := \mathbb{E}[\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \leq \rho_{\tau_j}(\gamma, \theta^0_{\tau_j})\}\mathbb{1}\{\widetilde{G}_i(\gamma', \pi^*) \leq \rho_{\tau_{j'}}(\gamma', \theta^0_{\tau_{j'}})\}] - \tau_j\tau_{j'}$. □

Lemma 4 extends Lemma 3 and specializes to it when $p = 1$. Further, when $j = j'$, $\kappa_{\tau_j\tau_{j'}}(\cdot, \cdot)$ and $\widetilde{\kappa}_{\tau_j\tau_{j'}}(\cdot, \cdot)$ are identical to $\kappa_{\tau_j}(\cdot, \cdot)$ and $\widetilde{\kappa}_{\tau_j}(\cdot, \cdot)$. Lemma 4 is proved by showing that the Cramér-Wold device in Wooldridge and White (1988, proposition 4.1) holds for $\widetilde{\mathcal{G}}(\cdot)$, which leads directly to the multivariate functional limit law and the convergence

$$\sqrt{n}(\widetilde{\theta}_n - \theta^0) \Rightarrow -A^{0\,-1} \int_\gamma \nabla_\theta \rho(\gamma, \theta^0) \widetilde{\mathcal{G}}(\gamma) d\mathbb{Q}(\gamma) \sim \mathcal{N}(0, \widetilde{C}^0)$$

follows by continuous mapping, where $\widetilde{C}^0 := A^{0\,-1}\widetilde{B}^0 A^{0\,-1}$ and $\widetilde{B}^0$ is a $c \times c$ matrix with j-th block row and j'-th block column matrix

$$\widetilde{B}^0_{\tau_j\tau_{j'}} := \int_\gamma \int_{\gamma'} \nabla_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^0_{\tau_j}) \widetilde{\kappa}_{\tau_j\tau_{j'}}(\gamma, \gamma') \nabla'_{\theta_{\tau_{j'}}} \rho_{\tau_{j'}}(\gamma, \theta^0_{\tau_{j'}}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\gamma').$$

The limit distribution of the MTSFQR estimator is obtained under the following regularity conditions.

Assumption 16. *(i) $\lambda_{\min}(A^0) > 0$; (ii) $\lambda_{\min}(L^0) > 0$; and (iii) $\lambda_{\min}(\widetilde{B}^0) > 0$, where*

$$L^0 := \begin{bmatrix} H^* & V^{0\prime} \\ V^0 & B^0 \end{bmatrix},$$

$V^0 := [V^{0\prime}_{\tau_1}, V^{0\prime}_{\tau_2}, \ldots, V^{0\prime}_{\tau_p}]'$, B^0 is a $c \times c$ matrix such that for each j and $j' = 1, 2, \ldots p$, its j-th block row and j'-th block column matrix is $B^0_{\tau_j\tau_{j'}} := \int_{\gamma'} \int_\gamma \nabla_\theta \rho_{\tau_j}(\gamma, \theta^0_{\tau_j}) \kappa_{\tau_j\tau_{j'}}(\gamma, \gamma') \nabla'_\theta \rho_{\tau_{j'}}(\widetilde{\gamma}, \theta^0_{\tau_{j'}}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\gamma')$, and $\widetilde{B}^0$ is defined as just above this assumption. □

By Assumption 16, S_i, $\int_\gamma \nabla_\theta \rho(\gamma, \theta^0)(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le \rho(\gamma, \theta^0)\} - \tau) d\mathbb{Q}(\gamma)$ and $\int_\gamma \nabla_\theta \rho(\gamma, \theta^0)(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho(\gamma, \theta^0)\} - \tau) d\mathbb{Q}(\gamma)$ have positive definite variance matrices. In the diagonal block with $j = j'$ the matrices $B^0_{\tau_j\tau_{j'}}$ and $\widetilde{B}^0_{\tau_j\tau_{j'}}$ are identical to $B^0_{\tau_j}$ and $\widetilde{B}^0_{\tau_j}$ as defined earlier in Section 3.2, Assumptions 5 and 10. Under Assumption 16 and earlier conditions, the limit distribution of the MTSFQR estimator is non-degenerate and given in the following theorem.

Theorem 7. *Under Assumptions 6, 7, 12, 14, and 16, if $\mathcal{M}$ is correctly specified, $\sqrt{n}(\widetilde{\theta}_n - \theta^0) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^0)$.* □

The limit distribution in Theorem 7 relates closely to the misspecifed case. Specifically, when $\theta^* = \theta^0$, we have $\widetilde{B}^* = \widetilde{B}^0$ and $A^* = A^0$, so that $\widetilde{C}^* = \widetilde{C}^0$. Further, when there are no nuisance effects, the limit distribution of a multiple quantile version of the FQR estimator studied in Section 2.2 can be deduced from Theorem 7. In particular, by setting $S_i \equiv 0$ in the definition of $\widetilde{\kappa}_{\tau_j\tau_{j'}}(\gamma, \gamma')$, which leads to $\widetilde{\kappa}_{\tau_j\tau_{j'}}(\cdot, \cdot) = \kappa_{\tau_j\tau_{j'}}(\cdot, \cdot)$ and $\widetilde{B}^0 = B^0$, we have $\sqrt{n}(\widehat{\theta}_n - \theta^0) \overset{\text{A}}{\sim} \mathcal{N}(0, C^0)$, where $C^0 := A^{0\,-1} B^0 A^{0\,-1}$. The $c \times c$ matrix B^0 is defined in Assumption 16.

5.3 Variance matrix estimation

The limit theory of Sections 5.1 and 5.2 enables hypothesis testing on the unknown model parameters once the relevant asymptotic variance matrices are estimated. The approach follows Section 4 closely and is only briefly detailed here.

If the model $\mathcal{M}$ is misspecified, let $J_{ni} := [J'_{\tau_1 ni}, J'_{\tau_2 ni}, \ldots, J'_{\tau_p ni}]$ and $\widehat{J}_{ni} := [\widehat{J}'_{\tau_1 ni}, \widehat{J}'_{\tau_2 ni}, \ldots, \widehat{J}'_{\tau_p ni}]$. Define the estimates

$$\widehat{B}_n := \frac{1}{n} \sum_{i=1}^n J_{ni} J'_{ni} \quad \text{and} \quad \widetilde{B}_n := \frac{1}{n} \sum_{i=1}^n \widehat{J}_{ni} \widehat{J}'_{ni} - \widehat{V}_n \widehat{P}_n^{-1} \widehat{K}'_n - \widehat{K}_n \widehat{P}_n^{-1} \widehat{V}'_n + \widehat{K}_n \widehat{P}_n^{-1} \widehat{H}_n \widehat{P}_n^{-1} \widehat{K}'_n,$$

where

$$\widehat{V}_n := \frac{1}{n}\sum_{i=1}^{n}\widehat{J}_{ni}S_i', \quad \widehat{K}_n := \frac{1}{n}\sum_{i=1}^{n}\int_{\gamma}\nabla_\theta\rho(\gamma,\widetilde{\theta}_n)\widehat{f}_{\gamma n}(\rho(\gamma,\widetilde{\theta}_n))\nabla_\pi'\widetilde{G}_i(\gamma,\widehat{\pi}_n)d\mathbb{Q}(\gamma),$$

and $\widehat{f}_{\gamma n}(\rho(\gamma,\widetilde{\theta}_n)) := \mathrm{diag}[\widehat{f}_{\gamma n}(\rho_{\tau_1}(\gamma,\widetilde{\theta}_{\tau_1 n}))\cdot I_{c_{\tau_1}},\ldots,\widehat{f}_{\gamma n}(\rho_{\tau_p}(\gamma,\widetilde{\theta}_{\tau_p n}))\cdot I_{c_{\tau_p}}]$. It immediately follows that $\widehat{B}_n \overset{\mathbb{P}}{\to} B^*$ under Assumptions 1, 12, 13, and 15 by applying Theorem 5; and similarly $\widetilde{B}_n \overset{\mathbb{P}}{\to} \widetilde{B}^*$ under Assumptions 1, 12, 13, and 16.

If $\mathcal{M}$ is correctly specified, the variance matrices B^0 and $\widetilde{B}^0$ can be estimated by first estimating the unknown covariance kernel functions $\kappa(\cdot,\cdot)$ and $\widetilde{\kappa}(\cdot,\cdot)$. Let

$$\widehat{\kappa}_n(\gamma,\gamma') := \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}\{G_i(\gamma)\leq\rho(\gamma,\widehat{\theta}_n)\}-\tau\right)\left(\mathbb{1}\{G_i(\gamma')\leq\rho(\gamma',\widehat{\theta}_n)\}-\tau\right),$$

$$\widetilde{\kappa}_n(\gamma,\gamma') := \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}\{\widehat{G}_i(\gamma)\leq\rho(\gamma,\widetilde{\theta}_n)\}-\tau\right)\left(\mathbb{1}\{\widehat{G}_i(\gamma')\leq\rho(\gamma',\widetilde{\theta}_n)\}-\tau\right),$$

and $\widetilde{\zeta}_n(\gamma) := n^{-1}\sum_{i=1}^{n}(\mathbb{1}\{\widehat{G}_i(\gamma)\leq\rho(\gamma,\widetilde{\theta}_n)\}-\tau)S_i'$. Note that $\widehat{\kappa}_n(\cdot,\cdot)$ and $\widetilde{\kappa}_n(\cdot,\cdot)$ are $p\times p$ matrices of functions, and they are consistent for $\kappa(\cdot,\cdot)$ uniformly on $\Gamma\times\Gamma$ under mild regularity conditions as in the univariate case, where $\kappa(\cdot,\cdot)$ is a $p\times p$ matrix with j-th row and j'-th column blocks being $\kappa_{\tau_j\tau_{j'}}(\cdot,\cdot)$. Likewise, $\widetilde{\zeta}_n(\cdot)$ turns out to be consistent for $\zeta(\cdot) := \mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\cdot,\pi^*)\leq\rho(\cdot,\theta^*)\}-\tau)S_i']$ uniformly on Γ by applying the ULLN. Using $\widehat{\kappa}_n(\cdot,\cdot)$ and $\widetilde{\kappa}_n(\cdot,\cdot)$, we estimate B^0 and $\widetilde{B}^0$ by plug-in, giving

$$\widehat{B}_n^\sharp := \int_\gamma\int_{\gamma'}\nabla_\theta\rho(\gamma,\widehat{\theta}_n)\widehat{\kappa}_n(\gamma,\gamma')\nabla_\theta'\rho(\gamma',\widehat{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma') \quad \text{and}$$

$$\begin{aligned}\widetilde{B}_n^\sharp := \int_\gamma\int_{\gamma'}\nabla_\theta\rho(\gamma,\widetilde{\theta}_n)\widetilde{\kappa}_n(\gamma,\gamma')\nabla_\theta'\rho(\gamma',\widetilde{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma') - \int_\gamma\nabla_\theta\rho(\gamma,\widetilde{\theta}_n)\widetilde{\zeta}_n(\gamma)d\mathbb{Q}(\gamma)\widehat{P}_n^{-1}\widehat{K}_n' \\ -\widehat{K}_n\widehat{P}_n^{-1}\int_\gamma\widetilde{\zeta}_n(\gamma)'\nabla_\theta'\rho(\gamma,\widetilde{\theta}_n)d\mathbb{Q}(\gamma)+\widehat{K}_n\widehat{P}_n^{-1}\widehat{H}_n\widehat{P}_n^{-1}\widehat{K}_n'.\end{aligned}$$

$\widehat{B}_n^\sharp$ and $\widetilde{B}_n^\sharp$ are $c\times c$ matrices with (j,j') block submatrices that estimate $B^0_{\tau_j\tau_{j'}}$ and $\widetilde{B}^0_{\tau_j\tau_{j'}}$, for j and $j'=1,2,\ldots,p$. It immediately follows that $\widehat{B}_n^\sharp \overset{\mathbb{P}}{\to} B^0$ under Assumptions 1, 12, 13, and 15 by applying Theorem 5. Similarly $\widetilde{B}_n^\sharp \overset{\mathbb{P}}{\to} \widetilde{B}^0$ under Assumptions 1, 12, 13, and 16. The result is given in the following Corollary whose proof is almost identical to that of Theorem 5 and is omitted.

Corollary 1. *(i) Given Assumption 1, 12, and 13, (i.a) if $\mathcal{M}$ is misspecified and Assumption 15 holds, $\widehat{B}_n \overset{\mathbb{P}}{\to} B^*$; (i.b) if $\mathcal{M}$ is correctly specified and Assumption 16 holds, $\widehat{B}_n^\sharp \overset{\mathbb{P}}{\to} B^0$;*

(ii) Given Assumption 6, 7, 8, 11, 12, and 14, (ii.a) if $\mathcal{M}$ is misspecified and Assumption 15 holds, $\widetilde{B}_n \overset{\mathbb{P}}{\to} \widetilde{B}^*$; (ii.b) if $\mathcal{M}_\tau$ is correctly specified and Assumption 16 holds, $\widetilde{B}_n^\sharp \overset{\mathbb{P}}{\to} \widetilde{B}^0$. □

6 Multiple Quantile Function Inference

Once consistent estimates of the relevant variance matrix estimates are obtained, tests and confidence intervals may be conducted in the usual manner making use of the limit distribution theory for multiple quantile estimates. Test procedures on the model parameters have the same basis irrespective of whether $\mathcal{M}$ is correctly specified. So the following development provides test methodology for the misspecified model case. Suppose interest centers on the following null and alternative hypotheses

$$\mathbb{H}_o : R(\theta^*) = 0 \quad \text{versus} \quad \mathbb{H}_a : R(\theta^*) \neq 0, \tag{9}$$

where $R : \otimes_{j=1}^{p} \Theta_{\tau_j} \mapsto \mathbb{R}^r$ ($r \in \mathbb{N}$) is continuously differentiable such that for each $j = 1, 2, \ldots, p$, $r_{\tau_j} := \text{rank}[\nabla'_{\theta_{\tau_j}} R(\theta^*)] \leq c_{\tau_j}$ and $r = \text{rank}[\nabla'_{\theta} R(\theta^*)]$.

Inference concerning hypotheses such as (9) on the unknown parameters of multiple quantile functions can be made in parallel to the standard testing procedures based on maximum likelihood (ML) estimation. First, define the following constrained estimators of the parameters

$$\bar{\theta}_n := \underset{\theta \in \otimes_{j=1}^{p} \Theta_{\tau_j}}{\arg\min}\ q_n(\theta) \quad \text{such that} \quad R(\theta) = 0 \quad \text{and} \quad \ddot{\theta}_n := \underset{\theta \in \otimes_{j=1}^{p} \Theta_{\tau_j}}{\arg\min}\ \widehat{q}_n(\theta) \quad \text{such that} \quad R(\theta) = 0.$$

We call $\bar{\theta}_n$ the *constrained multiple quasi-functional quantile regression (CMQFQR)* estimator and $\ddot{\theta}_n$ the *constrained multiple two-stage quasi-functional quantile regression (CMTSQFQR)* estimator. These estimators involve constrained Lagrangian estimation, analogous to constrained ML estimation. Although the functions $q_n(\cdot)$ and $\widehat{q}_n(\cdot)$ are not continuously differentiable, they are asymptotically differentiable, thereby enabling development of a suitable limit theory for testing and inference. The following additional regularity conditions are used to to derive the limit distributions of $\bar{\theta}_n$ and $\ddot{\theta}_n$.

Assumption 17. *(i)* $\lambda_{\min}(A^*) > 0$*; and (ii)* $\lambda_{\min}(B^*) > 0$. □

Assumption 18. *(i)* $R : \Theta \mapsto \mathbb{R}^r$ *is in* $\mathcal{C}^{(1)}(\Theta)$ *with* $r \in \mathbb{N}$ *and for each* $j = 1, 2, \ldots, p$, $r_{\tau_j} \leq c_{\tau_j}$*; and (ii)* $D(\theta^*) := \nabla'_{\theta} R(\theta^*) \in \mathbb{R}^{r \times c}$ *has full rank* r *where* ∇_{θ} *is the* $c \times 1$ *gradient operator.* □

Here if $\mathcal{M}$ is correctly specified, then the quantile functions cannot cross, i.e., there is no probability of crossing (see Phillips, 2015). This implies that the null condition becomes valid when it is consistent with the crossing condition. For example, if $\rho(\gamma, \theta_{\tau}^0) = \theta_{1\tau}^0 + \theta_{2\tau}^0 \gamma$ is a correct quantile function and $\tau_1 < \tau_2$, then for each γ, it has to follow that $\theta_{1\tau_1}^0 + \theta_{2\tau_1}^0 \gamma < \theta_{1\tau_2}^0 + \theta_{2\tau_2}^0 \gamma$. If a null hypothesis is stated that violates this inequality, the null can be trivially rejected.

The limit distributions of the CMQFQR and CMTSQFQR estimators are given in the following lemma.

Lemma 5. *(i) Given Assumptions 1, 12, 13, 17, and 18,* $\sqrt{n}(\bar{\theta}_n - \theta^*) + \sqrt{n}(\Omega A^*)^{-1} D^{*\prime} [D^*(\Omega A^*)^{-1} D^{*\prime}]^{-1} R(\theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \{I + (\Omega A^*)^{-1} D^{*\prime} [D^*(\Omega A^*)^{-1} D^{*\prime}]^{-1} D^*\}\, C^* \{I + D^{*\prime} [D^*(\Omega A^*)^{-1} D^{*\prime}]^{-1} D^*(\Omega A^*)^{-1}\})$*, where* $\Omega := \text{diag}[\omega_1 I_{c_1}, \omega_2 I_{c_2}, \ldots, \omega_p I_{c_p}]$, $D^* := \nabla'_{\theta} R(\theta^*)$*, and* $C^* := A^{*-1} B^* A^{*-1}$*; and* *(ii) Given Assumptions 6, 7, 12, 14, 15, and 18,* $\sqrt{n}(\ddot{\theta}_n-\theta^*)+\sqrt{n}(\Omega A^*)^{-1}D^{*\prime}[D^*(\Omega A^*)^{-1}D^{*\prime}]^{-1}R(\theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0,\{I+(\Omega A^*)^{-1}D^{*\prime}[D^*(\Omega A^*)^{-1}D^{*\prime}]^{-1}D^*\}\widetilde{C}^*\{I+D^{*\prime}[D^*(\Omega A^*)^{-1}D^{*\prime}]^{-1}D^*(\Omega A^*)^{-1}\})$. □

Under $\mathbb{H}_o$ $\sqrt{n}(\bar{\theta}_n-\theta^*)$ and $\sqrt{n}(\ddot{\theta}_n-\theta^*)$ are asymptotically distributed and centred at zero, whereas they are not bounded under $\mathbb{H}_a$. They are therefore useful in forming the tests discussed below. Note that the CMQFQR and CMTSQFQR limit distributions are influenced by the selection of the weights Ω, with different distributions for different Ω.

Tests are formed using standard Wald, LM, and LR test principles. The Wald tests use the unconstrained MQFQR and MTSQFQR estimators giving

$$\bar{\mathcal{W}}_n := nR(\widehat{\theta}_n)'\{\widehat{D}_n\widehat{C}_n\widehat{D}_n'\}^{-1}R(\widehat{\theta}_n) \quad \text{and} \quad \ddot{\mathcal{W}}_n := nR(\widetilde{\theta}_n)'\{\widetilde{D}_n\widetilde{C}_n\widetilde{D}_n'\}^{-1}R(\widetilde{\theta}_n),$$

where $\widehat{D}_n := \nabla_\theta' R(\widehat{\theta}_n)$, $\widehat{C}_n := \widehat{A}_n^{-1}\widehat{B}_n\widehat{A}_n^{-1}$, $\widetilde{D}_n := \nabla_\theta' R(\widetilde{\theta}_n)$, and $\widetilde{C}_n := \widetilde{A}_n^{-1}\widetilde{B}_n\widetilde{A}_n^{-1}$. Here, we let $\widehat{A}_n := \text{diag}[\widehat{A}_{\tau_1 n}, \widehat{A}_{\tau_2 n}, \ldots, \widehat{A}_{\tau_p n}]$, $\widetilde{A}_n := \text{diag}[\widetilde{A}_{\tau_1 n}, \widetilde{A}_{\tau_2 n}, \ldots, \widetilde{A}_{\tau_p n}]$, and for each $j = 1, 2, \ldots, p$,

$$\widehat{A}_{\tau_j n} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau_j n})\widehat{f}_{\gamma n}(\rho_\tau(\gamma,\widehat{\theta}_{\tau_j n}))\nabla_{\theta_\tau}'\rho_\tau(\gamma,\widehat{\theta}_{\tau_j n})d\mathbb{Q}(\gamma) \quad \text{and}$$

$$\widetilde{A}_{\tau_j n} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau_j n})\widehat{f}_{\gamma n}(\rho_\tau(\gamma,\widetilde{\theta}_{\tau_j n}))\nabla_{\theta_\tau}'\rho_\tau(\gamma,\widetilde{\theta}_{\tau_j n})d\mathbb{Q}(\gamma).$$

Under Assumption 11 and given consistency of $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ for θ^*, both $\widehat{A}_n$ and $\widetilde{A}_n$ are consistent for A^*. The Wald tests assess the magnitudes of $R(\widehat{\theta}_n)$ and $R(\widetilde{\theta}_n)$ in suitable metrics and unless $R(\theta^*) = 0$, the tests are not bounded in probability.

The LM tests are constructed as

$$\bar{\mathcal{LM}}_n := n\bar{Q}_n'\bar{A}_n^{-1}\bar{D}_n'\{\bar{D}_n\bar{C}_n\bar{D}_n'\}^{-1}\bar{D}_n\bar{A}_n^{-1}\bar{Q}_n \quad \text{and} \quad \ddot{\mathcal{LM}}_n := n\ddot{Q}_n'\ddot{A}_n^{-1}\ddot{D}_n'\{\ddot{D}_n\ddot{C}_n\ddot{D}_n'\}^{-1}\ddot{D}_n\ddot{A}_n^{-1}\ddot{Q}_n,$$

where we let $\bar{D}_n := \nabla_\theta' R(\bar{\theta}_n)$, $\bar{C}_n := \bar{A}_n^{-1}\bar{B}_n\bar{A}_n^{-1}$, $\ddot{D}_n := \nabla_\theta' R(\ddot{\theta}_n)$, $\ddot{C}_n := \ddot{A}_n^{-1}\ddot{B}_n\ddot{A}_n^{-1}$, $\bar{B}_n := n^{-1}\sum_{i=1}^n \bar{J}_{ni}\bar{J}_{ni}'$ and $\ddot{B}_n := n^{-1}\sum_{i=1}^n \ddot{J}_{ni}\ddot{J}_{ni}'$ with $\bar{J}_{ni} := [\bar{J}_{\tau_1 ni}', \bar{J}_{\tau_2 ni}', \ldots, \bar{J}_{\tau_p ni}']$, $\ddot{J}_{ni} := [\ddot{J}_{\tau_1 ni}', \ddot{J}_{\tau_2 ni}', \ldots, \ddot{J}_{\tau_p ni}']$, $\bar{Q}_n := n^{-1}\sum_{i=1}^n \bar{J}_{ni}$, $\ddot{Q}_n := n^{-1}\sum_{i=1}^n \ddot{J}_{ni}$,

$$\bar{J}_{\tau_j ni} := \int_\gamma \nabla_{\theta_{\tau_j}}\rho_{\tau_j}(\gamma,\bar{\theta}_{\tau n})\left(\mathbb{1}\{G_i(\gamma) \le \rho_{\tau_j}(\gamma,\bar{\theta}_{\tau_j n})\} - \tau\right)d\mathbb{Q}(\gamma), \quad \text{and}$$

$$\ddot{J}_{\tau ni} := \int_\gamma \nabla_{\theta_{\tau_j}}\rho_{\tau_j}(\gamma,\ddot{\theta}_{\tau_j n})(\mathbb{1}\{\ddot{G}_i(\gamma) \le \rho_{\tau_j}(\gamma,\ddot{\theta}_{\tau_j n})\} - \tau)d\mathbb{Q}(\gamma).$$

Further, define $\bar{A}_n := \text{diag}[\bar{A}_{\tau_1 n}, \bar{A}_{\tau_2 n}, \ldots, \bar{A}_{\tau_p n}]$, $\ddot{A}_n := \text{diag}[\ddot{A}_{\tau_1 n}, \ddot{A}_{\tau_2 n}, \ldots, \ddot{A}_{\tau_p n}]$, where for each

$j = 1, 2, \ldots, p$,

$$\bar{A}_{\tau_j n} := \int_{\gamma} \nabla_{\theta_\tau} \rho_\tau(\gamma, \bar{\theta}_{\tau_j n}) \widehat{f}_{\gamma n}(\rho_\tau(\gamma, \bar{\theta}_{\tau_j n})) \nabla'_{\theta_\tau} \rho_\tau(\gamma, \bar{\theta}_{\tau_j n}) d\mathbb{Q}(\gamma) \quad \text{and}$$

$$\ddot{A}_{\tau_j n} := \int_{\gamma} \nabla_{\theta_\tau} \rho_\tau(\gamma, \ddot{\theta}_{\tau_j n}) \widehat{f}_{\gamma n}(\rho_\tau(\gamma, \ddot{\theta}_{\tau_j n})) \nabla'_{\theta_\tau} \rho_\tau(\gamma, \ddot{\theta}_{\tau_j n}) d\mathbb{Q}(\gamma).$$

Given Assumption 11 and the consistency of $\bar{\theta}_n$ and $\ddot{\theta}_n$ for θ^* under the null, it follows that both $\bar{A}_n$ and $\ddot{A}_n$ are consistent for A^* under the null. Note that $\bar{J}_{ni}$ and $\ddot{J}_{ni}$ are defined in parallel to $\widehat{J}_{ni}$ and $\widetilde{J}_{ni}$. The only difference is that $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ in $\widehat{J}_{ni}$ and $\widetilde{J}_{ni}$ are replaced by $\bar{\theta}_n$ and $\ddot{\theta}_n$. Although Lemma 5 implies that the limit distributions of $\sqrt{n}(\bar{\theta}_n - \theta^*)$ and $\sqrt{n}(\ddot{\theta}_n - \theta^*)$ are influenced by the selection of Ω, the LM test statistics are defined without direct association with Ω. If $\mathbb{H}_o$ holds, both CMQFQR and CMTSQR estimators converge to θ^*, so that both $\bar{Q}_n$ and $\ddot{Q}_n$ converge to zero. Otherwise, neither $\bar{Q}_n$ nor $\ddot{Q}_n$ converges to zero, thereby giving the LM test discriminatory power under $\mathbb{H}_a$.

To construct the LR test the quantile functions are estimated under both hypotheses. Under the null $\mathbb{H}_o$, $\bar{\theta}_n$ and $\widehat{\theta}_n$ (resp. $\ddot{\theta}_n$ and $\widetilde{\theta}_n$) both converge to θ^*, so that the distance between $q_n(\bar{\theta}_n)$ and $q_n(\widehat{\theta}_n)$ (resp. $\widehat{q}_n(\ddot{\theta}_n)$ and $\widehat{q}_n(\widetilde{\theta}_n)$) converges to zero. But neither $\bar{\theta}_n$ nor $\ddot{\theta}_n$ converges to θ^* under $\mathbb{H}_a$, so that the distance between $q_n(\bar{\theta}_n)$ and $q_n(\widehat{\theta}_n)$ and the distance between $\widehat{q}_n(\ddot{\theta}_n)$ and $\widehat{q}_n(\widetilde{\theta}_n)$ is non zero asymptotically. These distances then form the basis of the following quasi-LR (QLR) tests

$$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n := 2n\{q_n(\bar{\theta}_n) - q_n(\widehat{\theta}_n)\} \quad \text{and} \quad \mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n := 2n\{\widehat{q}_n(\ddot{\theta}_n) - \widehat{q}_n(\widetilde{\theta}_n)\}.$$

The QLR statistics are nonnegative because both $\bar{\theta}_n$ and $\ddot{\theta}_n$ minimize the objective functions subject to the restrictions $R(\theta) = 0$, whereas both $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ minimize the same objective functions without constraint.

The limit distribution theory of the three tests under $\mathbb{H}_o$ and $\mathbb{H}_a$ are given in the following result.

Theorem 8. *For any sequence c_n such that $c_n = o(n)$,*

(i) If Assumptions 1, 11, 12, 13, 17, and 18 hold, (i.a) $\bar{\mathcal{W}}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty} \mathbb{P}(\bar{\mathcal{W}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; (i.b) $\mathcal{L}\bar{\mathcal{M}}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty} \mathbb{P}(\mathcal{L}\bar{\mathcal{M}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; (i.c) $\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n \overset{\text{A}}{\sim} W'(D^*(\Omega A^*)^{-1}D^{*\prime})^{-1}W$ under $\mathbb{H}_o$ and $\lim_{n\to\infty} \mathbb{P}(\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n \geq c_n) = 1$ under $\mathbb{H}_a$, where $W \sim \mathcal{N}(0, D^* C^* D^{*\prime})$;

(ii) If Assumptions 6, 7, 11, 12, 14, 15, and 18 hold, (ii.a) $\ddot{\mathcal{W}}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty} \mathbb{P}(\ddot{\mathcal{W}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; (ii.b) $\mathcal{L}\ddot{\mathcal{M}}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty} \mathbb{P}(\mathcal{L}\ddot{\mathcal{M}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; (iii.c) $\mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n \overset{\text{A}}{\sim} \widetilde{W}'(D^*(\Omega A^*)^{-1}D^{*\prime})^{-1}\widetilde{W}$ under $\mathbb{H}_o$ and $\lim_{n\to\infty} \mathbb{P}(\mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n \geq c_n) = 1$ under $\mathbb{H}_a$, where $\widetilde{W} \sim \mathcal{N}(0, D^* \widetilde{C}^* D^{*\prime})$. □

According to Theorem 8, the Wald, LM, and QLR statistics are each bounded in probability under $\mathbb{H}_o$ but unbounded under $\mathbb{H}_a$, so that the tests are consistent under the alternative hypothesis. The null limit distributions of the Wald and LM stastistics are equivalent and chi-squared with degrees of freedom given by the number of restrictions. The null limit distribution of the QLR test is given by a quadratic form in Gaussian variates, and thus a weighted sum of chi-square distributions. The limit theory in this case is influenced by the selection of Ω under both $\mathbb{H}_o$ and $\mathbb{H}_a$. When proving Theorem 8, we focus on the proof of Theorem 8 (*ii*), because the proof of Theorem 8 (*i*) is almost identical to that for Theorem 8 (*ii*).

7 Simulations

Simulations were conducted to assess the finite sample performance of FQR estimation and inference in relation to the asymptotic theory or and affirm the theoretical results in the earlier sections. In the following experiments, functional data were generated according to the regularity conditions in Section 6.

Let $\{G_i : \Gamma \mapsto \mathbb{R} : i = 1, 2, \ldots, n\}$ be data of iid functional observations, where $G_i(\gamma) := X_i + X_i\gamma$, $X_i = Z_i - 1/2$, $Z_i \sim_{iid} U[0,1]$, and $\gamma \in \Gamma := [1/2, 1]$, so that for each γ, $G_i(\gamma) \sim U[-(1+\gamma)/2, (1+\gamma)/2]$. Here, $G_i(\cdot)$ is viewed as a continuous functional observation with intercept X_i and linear coefficient X_i. Accordingly, for each $\tau \in (0,1)$, the quantile function of $G_i(\cdot)$ is obtained as $(\tau - 1/2) + (\tau - 1/2)\gamma$.

Next suppose that the following linear model is specified for the quantile function of $G_i(\cdot)$: for each $\tau \in (0,1)$,

$$\mathcal{M}_\tau := \{\rho_\tau(\gamma, \theta_\tau) := \theta_\tau + \theta_\tau\gamma, \theta_\tau \in \Theta := [-1/2, 1/2]\}. \tag{10}$$

Note that $\mathcal{M}_\tau$ is correctly specified for the quantile function of $G_i(\cdot)$ and setting $\theta^*_\tau = (\tau - 1/2)$, $\rho_\tau(\cdot, \theta^*_\tau)$ is identical to the quantile function of $G_i(\cdot)$.

With this DGP and modeling framework the simulation plan is as follows. First the unknown parameters are estimated by minimizing the sample average of the check functions. From the definition of the check function, we have for each i

$$\int_\Gamma \xi_\tau(G_i(\gamma) - \rho_\tau(\gamma, \theta_\tau))d\gamma = \int_\Gamma \tau(G_i(\gamma) - \rho_\tau(\gamma, \theta_\tau)) - (G_i(\gamma) - \rho_\tau(\gamma, \theta_\tau)\mathbb{1}\{G_i(\gamma) \geq \rho_\tau(\gamma, \theta_\tau)\}d\gamma,$$

and the M(Q)FQR estimator is obtained by minimizing $q_n(\cdot)$, where for each $\theta := (\theta_{\tau_1}, \theta_{\tau_2})$ with $(\tau_1, \tau_2) = (1/3, 2/3)$, $q_n(\theta) := \frac{1}{2}q_{\tau_1 n}(\theta_{\tau_1}) + \frac{1}{2}q_{\tau_2 n}(\theta_{\tau_2})$, and for each τ_j,

$$q_{\tau_j n}(\theta_{\tau_j}) := \int_\gamma \frac{1}{n}\sum_{i=1}^n \xi_{\tau_j}(G_i(\gamma) - \rho_{\tau_j}(\gamma, \theta_{\tau_j}))d\gamma.$$

The adjunct probability measure $\mathbb{Q}(\cdot)$ is assumed to be the uniform distribution on Γ, and we let $\omega_1 = \omega_2 = 1/2$. The MQFQR estimator $\widehat{\theta}_n := (\widehat{\theta}_{\tau_1 n}, \widehat{\theta}_{\tau_2 n})$ is obtained by a grid search in which Θ is partitioned into an equispaced grid of distance $1/250$ and $\widehat{\theta}_n$ is chosen as the parameter value that minimizes $q_n(\cdot)$ on this grid.

Wald, LM, and QLR tests are constructed for the following hypotheses: $\mathbb{H}_o$: $[\theta^*_{\tau_1}, \theta^*_{\tau_2}]' = r$ and

$\mathbb{H}_a : \left[\theta^*_{\tau_1}, \theta^*_{\tau_2}\right]' \neq r$, where $r := (\tau_1 - \frac{1}{2}, \tau_2 - \frac{1}{2})'$. The test statistics are

$$\bar{\mathcal{W}}_n = n(\widehat{\theta}_n - r)'\widehat{C}_n^{-1}(\widehat{\theta}_n - r), \quad \mathcal{L}\bar{\mathcal{M}}_n = n\bar{Q}_n' A^{*-1}\bar{C}_n^{-1}A^{*-1}\bar{Q}_n, \quad \text{and} \quad \mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n = 2n\{q_n(\bar{\theta}_n) - q_n(\widehat{\theta}_n)\},$$

where $\widehat{C}_n := A^{*-1}\widehat{B}_n A^{*-1}$, $\bar{C}_n := A^{*-1}\bar{B}_n A^{*-1}$, $\bar{\theta}_n = r$, and $A^* := \mathrm{diag}[A^*_{\tau_1}, A^*_{\tau_2}] = \mathrm{diag}[7/8, 7/8]$ with

$$\widehat{J}_{\tau_j ni} := \int_\gamma (1+\gamma)(\mathbb{1}\{G_i(\gamma) \leq \rho_{\tau_j}(\gamma, \widehat{\theta}_{\tau_j n})\} - \tau_j)d\gamma \quad \text{and}$$

$$\bar{J}_{\tau_j ni} := \int_\gamma (1+\gamma)(\mathbb{1}\{G_i(\gamma) \leq \rho_{\tau_j}(\gamma, \bar{\theta}_{\tau_j n})\} - \tau_j)d\gamma,$$

for each $j = 1, 2$. In this calculation A^* is computed by assuming that the density function of $G_i(\gamma)$ is known. That is, from the definition of $A^*_{\tau_j}$, it follows that $A^*_{\tau_j} = \int_\gamma (1+\gamma)^2 f_\gamma(\rho_{\tau_j}(\gamma, \theta^*_{\tau_j}))d\gamma$, and the DGP condition on $G_i(\gamma)$ implies that $f_\gamma(\rho_{\tau_j}(\gamma, \theta^*_{\tau_j})) = 1/(1+\gamma)$. The density function $f_\gamma(\rho_{\tau_j}(\gamma, \theta^*_{\tau_j}))$ can be straightforwardly estimated in practice for given τ by kernel density estimation. For the simulation we use a Gaussian kernel with Scott's rule of thumb for the bandwidth and further estimate $A^*_{\tau_j}$ by numerically integrating $(1+\gamma)^2 \widehat{f}_\gamma(\rho_{\tau_j}(\gamma, \widehat{\theta}_{\tau_j n}))$ with respect to γ.

Power is analyzed by modifying the DGP. Using the same definition of $G_i(\gamma)$, viz., $G_i(\gamma) = X_i + X_i\gamma$, let

$$X_i := \mathbb{1}\{W_i > \alpha/\sqrt{n}\}Z_i + \mathbb{1}\{W_i \leq \alpha/\sqrt{n}\}Z_i^{1/2} - \frac{1}{2},$$

where $W_i \sim_{iid} U[0,1]$. When $\alpha = 0$, the functional observation $G_i(\cdot)$ follows the same probability law as in the earlier DGP. But when $\alpha = \sqrt{n}$ the probability law of $G_i(\cdot)$ differs from the earlier DGP and follows a fixed alternative $\mathbb{H}_a$. Further, setting a fixed $\alpha > 0$ produces a Pitman-type local alternative for examining the local power of the three tests.

Simulations are conducted according to the above scheme with 5,000 replications under the null, alternative, and local alternative hypotheses. The results are reported in Table 1. The table has three panels giving the findings obtained under each hypothesis, which are summarized as follows.

First, null behavior of the three tests is generally well approximated by the limit distribution, which in this case is χ^2_2 under the null, corroborating Theorem 8 (*i*). When the sample size is small, the null distribution of the Wald test statistic differs slightly from the asymptotic but the differences disappear as the sample size increases. For sample sizes greater than 100, the null limit distribution provides a good approximation in all cases, with the LM and QLR test results showing the best conformity to the asymptotic distribution.

Second, power properties are studied by letting $\alpha = \sqrt{n}$. As the second panel of Table 1 shows, all three tests have increasing power as the sample size increases with empirical rejection rates rapidly approaching 100%, corroborating the asymptotic theory in Theorem 8 (*ii*). Notably, the finite sample power of the QLR test statistic exceeds that of the other two test statistics.

Local power properties are studied with $\alpha = 5$ and are shown in the third panel of Table 1. As the sample size increases, the empirical rejection rates of the three tests all converge to levels between the nominal size and 100%, showing evidence of stable local power. Again, the local power of the QLR test exceeds that of the other tests. Both Wald and LM tests exhibit similar local power patterns.

8 Empirical Application

This section reports an application of functional quantile methodology to study lifetime log income paths (LIPs). Specific attention is given to analyzing empirical differences in the income paths for different genders and education levels. The LIP quantile curves are parameterized as polynomial functions. In previous work Cho, Phillips, and Seo (2022) estimated conditional mean functions of the LIPs using the functional data classified by gender and education levels, drawing inferences about the gaps in these mean functions. The present work extends the scope of that analysis by examining the nature of the discrepancy in the quantile function curves, as estimated using the (Q)FQR techniques developed in the present paper.

In a similar context, García, Hernández, and López-Nicolás (2001), Sakellariou (2004a,b), Gardeazabal and Ugidos (2005), and Nicodemo (2009) studied gender gaps in various countries by employing standard quantile regression methods. Unlike those studies, functional data analysis is used here to explore the nature of gaps in the full income paths. The lifetime incomes of individuals are tracked over their careers and used as functional observations given their individual characteristics of gender and educational background. This methodology has the advantage that it considers full career income paths with their temporal and (persistent) dependence structures embodied in the observations, thereby enabling inference about a cross section of lifetime income paths without the complications of addressing potential complications in the internal temporal dynamics of those curves.

Data is drawn from the Continuous Work History Sample (CWHS) in the US. The same data set was used in Cho, Phillips, and Seo (2022) and contains 39 years of annual labor income before taxes of full-time U.S. white male and female workers born between 1960 and 1962. We divide the entire sample into different groups based on gender and education levels. Four education levels are considered: no college education, bachelor degree, master degree, and doctoral degree. According to this subdivision, the sample contains 673, 2,828, 539, and 323 individual income paths for males, and 837, 1,624, 469, and 418 individual income paths for females.

Data analysis is performed for two career paths, first over the full 0–40 years of work experience and second over the 10–40 year cycle of work experience. In each case, the study examines how gender and education level affect the income paths. This division in the analysis makes allowance for the fact that job mobility is typically higher in the first 10 years of work experience, as discovered in the early nonlinear regression study of Mincer and Jovanovic (1981) which gave empirical evidence of differences in job mobility during the first 10 years of work experience and showed that early career profiles are not necessarily good predictors of longer run difference in earnings. Huizinga (1990) and Light and Ureta (1995) provide similar supportive evidence of these differences.

Polynomial function specifications have been widely used to study the shape of lifetime income paths. The quadratic specification of Mincer (1958, 1974) has been the most popular specification used to model earnings over work experience (Bhuller, Mogstad, and Salvanes, 2017; Barth, Davis, and Freeman, 2018; Magnac, Pistolesi, and Roux, 2018). But Katz and Murphy (1992), Autor, Katz, and Krueger (1998), and Lemieux (2006) adopted quartic specifications in their empirical work and Cho and Phillips (2018) developed sequential testing methods to assess evidence for different functional form specifications of the wage equation with respect to work experience years. The present study uses quadratic, cubic, and quartic models in the empirical application.

8.1 Inference on income paths over full lifetime experience

Income profiles are first analyzed over the entire career, taken as work experience from 0 to 40 years. The general quartic model is specified as

$$\rho_\tau(\gamma, \theta_{\tau 1}, \theta_{\tau 2}, \theta_{\tau 3}, \theta_{\tau 4}, \theta_{\tau 5}) = \theta_{\tau 1} + \theta_{\tau 2}\gamma + \theta_{\tau 3}\gamma^2 + \theta_{\tau 4}\gamma^3 + \theta_{\tau 5}\gamma^4.$$

When $\theta_{\tau 5} = 0$ the model is cubic, and if $\theta_{\tau 4} = \theta_{\tau 5} = 0$ the specification is quadratic. For curve estimation the adjunct probability measure $\mathbb{Q}$ is set to be uniform on Γ, so that equal chances are allocated to each γ for possible violation of the null. The probability density function $f_\gamma(\cdot)$ is estimated nonparametrically by kernel density estimation using a Gaussian kernel with Silverman's rule of thumb for the bandwidth setting. Estimated plots of $\rho_\tau(\cdot)$ are provided in the Online Supplement using the functional observations classified by gender and education, along with the estimated errors measured by the quantity $q_{\tau n}(\widehat{\theta}_\tau)$, which captures the value of the criterion function (2) at the estimate $\widehat{\theta}_\tau$.

To infer possible gender effects a dummy variable d_i is introduced to the fitted model with $d_i = 1$ for female and $d_i = 0$ for male. Parameter setting vectors for male and female are θ_τ^M and θ_τ^F: for the quartic model $\theta_\tau^M = (\theta_{\tau 1}^M, \theta_{\tau 2}^M, \theta_{\tau 3}^M, \theta_{\tau 4}^M, \theta_{\tau 5}^M)'$ and $\theta_\tau^F = (\theta_{\tau 1}^F, \theta_{\tau 2}^F, \theta_{\tau 3}^F, \theta_{\tau 4}^F, \theta_{\tau 5}^F)'$; then $(\theta_{\tau 5}^M, \theta_{\tau 5}^F)$ are set to zero for the cubic model; and $(\theta_{\tau 4}^M, \theta_{\tau 4}^F; \theta_{\tau 5}^F, \theta_{\tau 5}^M)$ are set to zero for the quadratic model. So the equations for the male and female group LIPs are specified as $\rho_\tau^M(\gamma, \theta_\tau^M) = \theta_{\tau 1}^M + \theta_{\tau 2}^M\gamma + \theta_{\tau 3}^M\gamma^2 + \theta_{\tau 4}^M\gamma^3 + \theta_{\tau 5}^M\gamma^4$ and $\rho_\tau^F(\gamma, \theta_\tau^F) = \theta_{\tau 1}^F + \theta_{\tau 2}^F\gamma + \theta_{\tau 3}^F\gamma^2 + \theta_{\tau 4}^F\gamma^3 + \theta_{\tau 5}^F\gamma^4$, which are written in combined form as

$$\rho_\tau(\gamma, \theta_\tau^M, \theta_\tau^F) = \rho_\tau^F(\gamma, \theta_\tau^F)d_i + \rho_\tau^M(\gamma, \theta_\tau^M)(1 - d_i),$$

with $d_i = 1$ for female; and $d_i = 0$, otherwise. A primary null of interest is the gender hypothesis $\mathbb{H}_o : \theta_\tau^{F*} = \theta_\tau^{M*}$. Failure to reject $\mathbb{H}_o$ provides evidence that the lifetime income paths do not differ significantly between genders.

Tables 2 and 3 summarize the inferential results for the gender hypothesis on income. First, the LIPs differ significantly between genders in groups with college education. At both 1% and 5% levels the null of equivalence in the LIPs is rejected by the most tests. Several factors may influence these differences in the LIP. For instance, job flexibility and stability may be more important factors for females, whereas a higher emphasis may be placed on earnings for males, as Wiswall and Zafar (2017) demonstrate in their empirical study. Second, differences in the LIPs are less evident in the group without college education, although the results depend on the quantile τ. Specifically. when $\tau = 0.25$, the difference in the LIPs is significant for both Wald and LM tests, but not statistically significant for the QLR test at 1% and 5% levels. At quantile $\tau = 0.75$, the null hypothesis is not rejected for most specifications, implying that the LIPs of the workers earning higher income without college education tend to be closer across genders than for workers earning lower incomes. Third, increases in quantile τ leads to a reversal of inferences between the Master's and Doctoral groups. When $\tau = 0.25$, test outcomes for the Doctoral group exceed those for the Master's group under quadratic, cubic, and quartic specifications; but for quantile $\tau = 0.75$, test outcomes for the Master's group exceed those for the Doctoral group. This result suggests that the gender gap in the LIPs of the Doctoral group tends to shrink at higher income and education levels whereas the gender gap of the Master group expands.

We next examine the education effect on the quantile functions of the LIP within the same gender group. For this, we make pairwise comparisons between the following groups: Bachelor's degree vs. no college education; Bachelor's vs. Master's level education, and Master's vs. Doctoral degrees. Tables 4 and 5 contain the test outcomes for $\tau = 0.25, 0.5$, and 0.75. Table 4 reports the test results using the samples of the original LIPs, while Table 5 is constructed using rescaled samples in which each individual's income path is scaled by the individual's integrated log income path over the work experience years. As expected, for the male and female groups, we find highly significant differences in the quantile functions of the LIP across different education levels. We reject the null hypothesis in most cases at the 1% and 5% significance levels. Interestingly, as Table 5 shows, this difference is less apparent between the Master's and Doctoral male groups, implying that through the rescaling process, the overall shapes of the quantile functions of the LIP are more or less similar between the Master's and Doctoral male groups.

8.2 Inference on income paths for later work years

We repeat the exercises conducted in Section 8.1 using the same samples and group specifications classified by gender and education levels. The LIP domain is now restricted to the 10–40 year cycle to remove early job mobility effects on the test statistics and focus on the later years work cycle.

Tables 6 and 7 report inferential findings on gender effects and are summarized as follows. Table 6 gives test results from the Wald, LM, and QLR statistics. Similar to Table 2, the gender effect on the quantile functions of the LIP is significant for the groups with college education. In particular, for each of the three models, the hypothesis of equal LIPs between genders is rejected for the groups with college education in most tests, but not so for the group without college education (except for the median quantile $\tau = 0.5$ by the Wald and LM tests). In addition, the gender gap becomes less evident in the Doctoral group relative to the Master's group as τ increases. These findings reflect those reported earlier in Section 8.1. Second, the inference results are sharply reversed if the LIPs are rescaled by their respective integrated LIPs over the mature work experience years, 10–40 years. As shown in Table 7, for all τ levels under consideration, there is no strong evidence to conclude that the quantile functions of the rescaled LIPs differ between genders and among the different education levels. Interestingly, this phenomenon is more evident for larger τ: when $\tau = 0.25$, for instance, the LM test rejects the null hypothesis of the equal quantile functions between genders in the Doctoral group at the 5% level of significance, but we fail to reject the null hypothesis by all of the Wald, LM, and QLR tests, when $\tau = 0.75$.

Finally, Tables 8 and 9 report test outcomes of the education effect on the quantile functions of the income paths over the mature career years. These outcomes parallel those of Tables 4 and 5. Table 8 shows that the education effect cannot be ignored for quadratic, cubic, and quartic specifications if the data are not rescaled. But upon rescaling the LIPs, the education effect diminishes. Moreover, as in Table 6, this tendency is clearer for the large value of τ. Indeed, when $\tau = 0.75$, the null hypothesis of equivalence is not rejected by the Wald, LM, and QLR tests for each of the quadratic, cubic, and quartic model specifications.

To sum up, gender and education effects on the quantile functions of the income curves are evident, irrespective of whether full lifetime experience or just mature career years are considered. But when the income paths are rescaled for each individual, the gaps in the quantile functions induced by different genders and education levels become less obvious. This feature is consistent with the findings of Cho, Phillips, and Seo (2022) based on the mean functions of the LIP. Specifically, they provide empirical evidence that the mean functions of the rescaled LIPs do not differ between genders and/or among education levels, although the mean functions of the original LIPs do differ. So these empirical findings are compatible. Our findings further reveal that this tendency is more noticeable at higher quantile values of τ. Thus, the gap in the quantile functions of the rescaled LIPs induced by different genders and education levels becomes smaller for workers with higher income levels in each group.

9 Concluding Remarks

This paper extends standard parametric quantile regression methodology to a functional data setting, providing estimation and inferential techniques that enable evidence based analysis of the quantiles of curve observations. A full asymptotic theory is developed under regularity conditions that enable a wide range of potential applications and allow for misspecified as well as correctly specified parametric model formulations. The methods and limit theory also allow for the functional data to be influenced by nuisance parameter estimation effects which often figure in dataset construction. Taken together, the methods provide a new approach to quantile regression estimation and inference that has many applications. The labor income empirical application given in the paper is one example in which curve data are of interest in economics, particularly in microeconomic analysis, where an assessment of such factors as gender and education level in determining the shape of lifetime income profiles is relevant. The present methods should prove useful in other fields of analysis where curve data appear or can be readily constructed, including multi-country macroeconomic and international trade studies in economics that involve comparisons of various economic indicators such as inflation, unemployment, growth, and merchandise trade data measured over the same time period.

References

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

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Armsterdam: North Holland, Chapter 37, pp. 2247–2294, vol. 4.

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

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

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

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

BOSQ, D. (2000): *Linear Processes in Function Spaces: Theory and Applications*, New York: Springer.

CARDOT, H., C. CRAMBES, AND P. SARDA (2005): "Quantile regression when the covariates are functions," *Nonparametric Statistics*, 17, 841–856.

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

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

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

CHO, J. S., P. C. B. PHILLIPS, AND J. SEO (2022): "Parametric conditional mean inference with functional data applied to lifetime income curves," *International Economic Review*, 63, 391–456.

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

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

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

FERRATY, F., A. RABHI, AND P. VIEU (2005): "Conditional quantiles for dependent functional data with application to the climatic" el niño" phenomenon," *Sankhyā: The Indian Journal of Statistics*, 378–398.

GARCÍA, J., P. J. HERNÁNDEZ, AND A. LÓPEZ-NICOLÁS (2001): "How Wide is the Gap? An Investigation of Gender Wage Differences using Quantile Regression," *Empirical Economics*, 26, 149–167.

GARDEAZABAL, J. AND A. UGIDOS (2005): "Gender Wage Discrimination at Quantiles," *Journal of Population Economics*, 18, 165–179.

HORVATH, L. AND P. KOKOSZKA (2012): *Inference for Functional Data with Applications*, vol. 200, New York: Springer-Verlag.

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

KATO, K. (2012): "Estimation in functional linear quantile regression," *The Annals of Statistics*, 40, 3108–3136.

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

KNIGHT, K. (1998): "Limiting Distributions for L_1 Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755–770.

KOENKER, R. AND G. J. BASSETT (1978): "Regrssion Quantiles," *Econometrica*, 46, 33–50.

LEMIEUX, T. (2006): "The "Mincer Equation" Thirty Years After *Schooling, Experience, and Earnings*," in *Jacob Mincer A Pioneer of Modern Labor Economics*, ed. by S. Grossbard, Boston: Springer, 127–145.

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

LI, M., K. WANG, A. MAITY, AND A.-M. STAICU (2022): "Inference in functional linear quantile regression," *Journal of Multivariate Analysis*, 190, 104985.

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

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

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

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

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

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

NEWEY, W. K. AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. McFadden, Armsterdam: Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 36, 2111–2245.

NICODEMO, C. (2009): "Gender Pay Gap and Quantile Regression in European Families," IZA Discussion Papers 3978, Institute of Labor Economics (IZA).

OBERHOFER, W. AND H. HAUPT (2016): "Asymptotic Theory for Nonlinear Quantile Regression under Weak Dependence," *Journal of the American Statistical Association*, 32, 686–713.

PHILLIPS, P. C. B. (2015): "Halbert White Jr. Memorial JFEC Lecture: Pitfalls and Possibilities in Predictive Regression," *Journal of Financial Econometrics*, 13, 521–555.

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

POLLARD, D. (1984): *Convergence of Stochastic Processes*, New York: Springer-Verlag.

——— (1985): "New Ways to Prove Central Limit Theorem," *Econometric Theory*, 1, 295–314.

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

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

RICE, J. A. AND B. W. SILVERMAN (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.

SAKELLARIOU, C. (2004a): “Gender-earnings Differentials using Quantile Regressions,” *Journal of Labor Research*, 25, 457–468.

——— (2004b): “The use of Quantile Regressions in Estimating Gender Wage Differentials: a Case Study of the Philippines,” *null*, 36, 1001–1007.

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

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

WOOLDRIDGE, J. AND H. WHITE (1988): “Some Invariance Principles and Central Limit Theorems for Dependence Heterogeneous Processess,” *Econometric Theory*, 4, 210–230.

		Size of the Tests					
Statistics	Levels$\backslash n$	50	100	200	300	400	500
	10%	12.38	12.62	10.72	10.98	10.84	10.92
$\bar{\mathcal{W}}_n$	5%	6.76	7.38	6.04	5.40	5.60	5.60
	1%	2.06	2.02	1.68	1.18	1.50	1.32
	10%	11.34	9.82	10.18	9.70	10.44	9.66
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	5.16	5.04	5.18	4.98	5.28	4.66
	1%	1.14	1.08	1.12	0.88	1.14	0.88
	10%	10.14	10.16	9.70	10.38	9.84	10.14
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	5.22	5.32	4.66	4.94	5.38	4.90
	1%	1.16	1.16	1.06	0.92	1.16	0.98
		Power of the Tests					
Statistics	Levels$\backslash n$	20	40	60	80	100	120
	10%	78.70	97.30	99.76	99.96	100.0	100.0
$\bar{\mathcal{W}}_n$	5%	69.96	94.50	99.28	99.94	99.96	100.0
	1%	51.44	86.48	96.52	99.40	99.82	99.98
	10%	74.10	97.00	99.84	100.0	100.0	100.0
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	66.06	94.12	99.54	99.90	99.98	100.0
	1%	35.12	80.54	96.14	99.38	99.92	99.98
	10%	85.92	98.62	99.94	100.0	100.0	100.0
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	77.20	97.30	99.86	100.0	100.0	100.0
	1%	55.68	90.00	98.50	99.86	99.94	100.0
		Local Power of the Tests					
Statistics	Levels$\backslash n$	50	100	200	300	400	500
	10%	83.86	81.64	81.48	79.24	79.96	79.18
$\bar{\mathcal{W}}_n$	5%	75.86	72.86	71.68	69.52	70.22	69.18
	1%	57.92	53.44	50.10	47.80	47.86	46.92
	10%	83.16	80.60	80.12	78.00	78.96	77.74
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	72.34	69.22	69.24	68.66	69.30	67.56
	1%	48.64	46.02	44.70	44.60	45.80	43.36
	10%	89.62	87.24	87.24	85.28	86.36	85.56
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	82.78	79.74	79.44	77.00	77.98	77.10
	1%	63.68	60.14	57.72	56.18	56.74	55.32

Table 1: EMPIRICAL REJECTION RATES OF THE TEST STATISTICS UNDER THE NULL, ALTERNATIVE, AND LOCAL ALTERNATIVE HYPOTHESES (IN PERCENT). This table shows the simulation results for the Wald, LM, and QLR test statistics under the null, alternative, and local alternative hypothesis. We let $G_i(\gamma) := X_i + X_i\gamma$ with $\gamma \in [1/2, 1]$ and $X_i := \mathbb{1}\{W_i > \alpha/\sqrt{n}\}Z_i + \mathbb{1}\{W_i \leq \alpha/\sqrt{n}\}Z_i^{1/2} - 1/2$, where $Z_i \sim_{iid} U[0,1]$ and $W_i \sim_{iid} U[0,1]$. For $\tau_1 = 1/3$ and $\tau_2 = 2/3$, we let $\rho(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ denote the model for the quantile function, where for $j = 1, 2$, $\theta_{\tau_j} \in \Theta := [-1/2, 1/2]$. In addition, the null, alternative, and local alternative DGPs are generated by letting α be 0, $\sqrt{n}$, and 5, respectively, to test $\mathbb{H}_o : \theta^*_{\tau_1} = -1/6$ and $\theta^*_{\tau_2} = 1/6$ against $\mathbb{H}_a : \theta^*_{\tau_1} \neq -1/6$ or $\theta^*_{\tau_2} \neq 1/6$.

Inference results on the quantiles of the log income path across different genders					
		Education Level	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree	14.51**	42.60**	31.44
		Bachelor	102.40**	66.45**	355.28**
		Master	9.26*	5.81	37.16
		Ph.D	38.06**	26.72**	155.59**
	Cubic	w/o Degree	19.13**	63.78**	33.49
		Bachelor	111.79**	80.56**	356.25**
		Master	21.90**	19.37**	59.05*
		Ph.D	59.07**	38.51**	194.44**
	Quartic	w/o Degree	25.59**	66.69**	38.68
		Bachelor	121.49**	84.71**	368.70**
		Master	16.62**	7.51	63.87*
		Ph.D	56.72**	33.07**	222.87**
$\tau = 0.5$	Quadratic	w/o Degree	21.95**	45.63**	31.85
		Bachelor	128.55**	139.76**	777.47**
		Master	34.66**	34.96**	182.31**
		Ph.D	28.92**	31.72**	176.13**
	Cubic	w/o Degree	21.62**	49.36**	32.18
		Bachelor	163.66**	162.02**	820.52**
		Master	44.91**	60.26**	176.89**
		Ph.D	29.93**	40.99**	178.77**
	Quartic	w/o Degree	22.99**	43.89**	30.81
		Bachelor	261.41**	186.51**	814.74**
		Master	90.92**	68.53**	182.80**
		Ph.D	38.47**	38.54**	174.09**
$\tau = 0.75$	Quadratic	w/o Degree	1.77	2.60	2.18
		Bachelor	132.20**	138.11**	742.19**
		Master	77.35**	72.03**	332.41**
		Ph.D	15.88**	19.85**	126.70**
	Cubic	w/o Degree	1.75	3.83	1.98
		Bachelor	180.50**	214.27**	768.28**
		Master	68.28**	76.36**	307.91**
		Ph.D	25.61**	23.20**	129.27**
	Quartic	w/o Degree	5.04	11.71*	9.66
		Bachelor	309.30**	210.61**	712.39**
		Master	110.34**	70.81**	294.06**
		Ph.D	24.56**	18.46**	104.02**

Table 2: INFERENCE RESULTS USING FUNCTION DATA OVER 0 TO 40 WORK EXPERIENCE YEARS (NON-SCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of equal log income paths across different genders at quantiles $\tau = 0.25, 0.50$, and 0.75. The figures with affixes '*' and '**' indicate rejection of the null hypothesis at 5% and 1% significance levels.

		Education Level	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree	11.11*	6.81	7.44
		Bachelor	28.45**	43.32**	59.02**
		Master	22.35**	13.96**	22.53*
		Ph.D	9.83*	5.38	1.51
	Cubic	w/o Degree	10.05*	14.59**	9.30
		Bachelor	23.76**	52.53**	54.27**
		Master	15.75**	19.01**	17.62
		Ph.D	21.34**	20.66**	8.94
	Quartic	w/o Degree	13.67*	12.90*	9.82
		Bachelor	28.23**	30.92**	21.72
		Master	19.11**	7.77	4.68
		Ph.D	17.45**	20.02**	20.09
$\tau = 0.5$	Quadratic	w/o Degree	6.03	9.06*	10.34
		Bachelor	136.95**	165.08**	732.60**
		Master	12.90**	5.40	11.17
		Ph.D	5.64	2.29	3.08
	Cubic	w/o Degree	8.05	17.55**	12.19
		Bachelor	49.71**	88.66**	90.18**
		Master	16.96**	28.95**	33.35*
		Ph.D	13.86**	33.82**	23.54*
	Quartic	w/o Degree	13.68*	16.80**	10.13
		Bachelor	264.71**	174.43**	725.93**
		Master	95.31**	57.52**	160.43**
		Ph.D	43.16**	36.62**	149.01**
$\tau = 0.75$	Quadratic	w/o Degree	8.96*	8.51*	10.38
		Bachelor	36.75**	20.24**	0.33
		Master	9.92*	2.04	3.58
		Ph.D	8.22*	0.59	17.42
	Cubic	w/o Degree	12.02*	14.51**	9.34
		Bachelor	33.25**	59.62**	59.85**
		Master	10.31*	10.83*	20.62*
		Ph.D	8.10	8.70	8.72
	Quartic	w/o Degree	13.19*	15.42**	7.09
		Bachelor	29.98**	66.13**	109.93**
		Master	7.20	19.65**	40.30**
		Ph.D	4.69	17.90**	21.62

Inference results on the quantiles of the rescaled log income path across different genders

Table 3: INFERENCE RESULTS USING FUNCTION DATA OVER 0 TO 40 WORK EXPERIENCE YEARS (RESCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal rescaled log income paths across different genders at quantiles $\tau = 0.25, 0.50$, and 0.75. The figures with affixes '*' and '**' indicate rejection of the null hypothesis at the 5% and 1% significance levels.

			Male			Female		
			Wald	LM	QLR	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree vs. Bachelor	1035.70**	1428.50**	9486.50**	1613.80**	1386.30**	7626.70**
		Bachelor vs. Master	132.97**	75.94**	867.58**	293.63**	118.09**	936.04**
		Master vs. Ph.D	29.32**	30.10**	108.38**	15.63**	13.30**	17.05
	Cubic	w/o Degree vs. Bachelor	1075.40**	1417.70**	9596.00**	2438.90**	1489.40**	7691.70**
		Bachelor vs. Master	223.22**	91.48**	955.75**	462.57**	153.90**	938.21**
		Master vs. Ph.D	39.14**	22.76**	100.45**	21.97**	25.01**	22.79
	Quartic	w/o Degree vs. Bachelor	1192.40**	1540.70**	9761.80**	3139.10**	1604.50**	7877.60**
		Bachelor vs. Master	178.71**	86.05**	974.70**	498.43**	167.94**	943.39**
		Master vs. Ph.D	32.73**	19.18**	114.34**	22.83**	17.40**	19.56
$\tau = 0.5$	Quadratic	w/o Degree vs. Bachelor	1297.40**	972.09**	10469.00**	1198.60**	921.94**	7913.50**
		Bachelor vs. Master	156.98**	125.81**	1243.30**	227.00**	190.56**	1293.40**
		Master vs. Ph.D	18.19**	28.51**	105.82**	24.44**	26.05**	69.72*
	Cubic	w/o Degree vs. Bachelor	1754.40**	1002.70**	10631.00**	1736.70**	995.46**	7899.30**
		Bachelor vs. Master	203.43**	189.45**	1249.80**	305.58**	240.92**	1301.60**
		Master vs. Ph.D	18.32**	33.09**	106.09**	23.04**	30.71**	73.70*
	Quartic	w/o Degree vs. Bachelor	2224.10**	999.07**	10635.00**	2377.70**	1040.40**	7884.90**
		Bachelor vs. Master	312.23**	178.14**	1304.00**	478.45**	250.22**	1349.30**
		Master vs. Ph.D	20.86**	25.18**	93.95*	42.80**	40.11**	72.39*
$\tau = 0.75$	Quadratic	w/o Degree vs. Bachelor	994.63**	552.80**	7437.60**	526.95**	507.70**	5404.20**
		Bachelor vs. Master	144.49**	185.61**	1235.70**	118.55**	178.13**	866.55**
		Master vs. Ph.D	23.51**	23.78**	84.43*	46.36**	37.38**	171.52**
	Cubic	w/o Degree vs. Bachelor	1389.00**	588.99**	7644.60**	657.82**	491.60**	5465.90**
		Bachelor vs. Master	149.57**	172.99**	1215.30**	96.33**	167.14**	855.38**
		Master vs. Ph.D	39.00**	28.27**	93.41*	47.84**	34.34**	182.34**
	Quartic	w/o Degree vs. Bachelor	1702.30**	499.12**	7477.00**	757.48**	499.21**	5326.00**
		Bachelor vs. Master	176.91**	200.45**	1219.10**	157.37**	199.91**	817.24**
		Master vs. Ph.D	59.29**	33.49**	97.74*	64.63**	29.94**	192.86**

Inference results on the quantiles of the log income path across different education levels

Table 4: INFERENCE RESULTS USING DATA OVER 0 TO 40 WORK EXPERIENCE YEARS (NON-SCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal log income paths across different education levels at quantiles $\tau = 0.25, 0.50$, and 0.75. The affixes '*' and '**' signify rejection of the null hypothesis at the 5% and 1% significance levels.

			Male			Female		
			Wald	LM	QLR	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree vs. Bachelor	455.37**	300.87**	422.37**	189.00**	127.96**	173.43**
		Bachelor vs. Master	70.32**	51.36**	92.75**	18.84**	24.70**	43.68**
		Master vs. Ph.D	5.40	6.11	9.34	16.74**	12.13**	22.27*
	Cubic	w/o Degree vs. Bachelor	435.83**	344.16**	489.41**	212.49**	213.42**	229.57**
		Bachelor vs. Master	50.71**	73.04**	119.30**	14.21**	33.20**	60.86**
		Master vs. Ph.D	17.64**	9.72*	9.03	16.39**	15.65**	33.12*
	Quartic	w/o Degree vs. Bachelor	450.76**	170.09**	344.50**	184.43**	126.68**	213.68**
		Bachelor vs. Master	69.94**	49.41**	92.70**	14.14*	17.06**	37.75**
		Master vs. Ph.D	13.01*	13.50*	16.58	17.42**	14.10*	44.45**
$\tau = 0.5$	Quadratic	w/o Degree vs. Bachelor	422.39**	322.04**	370.49**	168.50**	121.72**	107.53**
		Bachelor vs. Master	41.79**	18.01**	24.10*	38.26**	27.49**	43.88**
		Master vs. Ph.D	4.76	4.78	11.54	13.08**	16.28**	39.55**
	Cubic	w/o Degree vs. Bachelor	497.42**	571.74**	543.37**	227.70**	307.39**	179.60**
		Bachelor vs. Master	36.89**	57.85**	100.68**	27.59**	46.96**	68.61*
		Master vs. Ph.D	6.52	13.30**	13.32	16.96**	20.77**	43.34**
	Quartic	w/o Degree vs. Bachelor	446.61**	394.76**	530.48**	191.67**	240.80**	210.42**
		Bachelor vs. Master	41.43**	55.26**	124.62**	27.51**	34.05**	72.13**
		Master vs. Ph.D	7.31	8.83	15.11	18.71**	19.16**	46.75**
$\tau = 0.75$	Quadratic	w/o Degree vs. Bachelor	318.23**	375.63**	204.05**	115.69**	107.56**	55.63**
		Bachelor vs. Master	26.26**	19.86**	17.42*	27.57**	16.38**	4.38
		Master vs. Ph.D	5.99	4.47	9.00	8.25*	18.14**	37.79**
	Cubic	w/o Degree vs. Bachelor	395.43**	966.35**	453.60**	205.53**	447.47**	163.01**
		Bachelor vs. Master	23.93**	29.96**	43.27**	23.39**	29.03**	25.37**
		Master vs. Ph.D	9.42	13.86**	19.42	10.38*	19.11**	33.00**
	Quartic	w/o Degree vs. Bachelor	300.81**	995.12**	587.19**	145.27**	470.57**	210.00**
		Bachelor vs. Master	20.20**	36.11**	77.88**	26.41**	32.67**	43.41**
		Master vs. Ph.D	10.12	15.54**	17.15	12.11*	18.11**	32.89*

Table 5: INFERENCE RESULTS USING DATA OVER 0 TO 40 WORK EXPERIENCE YEARS (RESCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal rescaled log income paths across different education levels at quantiles $\tau = 0.25, 0.50$, and 0.75. The affixes '*' and '**' signify rejection of the null hypothesis at the 5% and 1% significance levels.

Inference results on the quantiles of the log income path across different genders					
		Education Level	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree	0.19	0.80	1.84
		Bachelor	106.73**	81.71**	301.97**
		Master	8.68*	12.29**	49.08*
		Ph.D	32.07**	27.10**	153.83**
	Cubic	w/o Degree	4.18	9.49*	4.98
		Bachelor	140.13**	76.32**	309.77**
		Master	20.97**	8.44	47.55*
		Ph.D	48.66**	30.58**	162.53**
	Quartic	w/o Degree	11.05	10.21	4.33
		Bachelor	138.52**	73.35**	306.74**
		Master	21.43**	8.15	48.28
		Ph.D	59.54**	31.14**	162.70**
$\tau = 0.5$	Quadratic	w/o Degree	10.30*	16.03**	10.25
		Bachelor	56.05**	46.51**	58.34**
		Master	39.70**	46.55**	159.29**
		Ph.D	27.45**	36.85**	145.37**
	Cubic	w/o Degree	13.05*	14.83**	10.08
		Bachelor	255.53**	165.62**	722.14**
		Master	88.96**	58.39**	161.99**
		Ph.D	43.51**	38.69**	152.19**
	Quartic	w/o Degree	9.09	10.43	8.02
		Bachelor	52.69**	70.86**	91.16**
		Master	17.43**	27.64**	40.97*
		Ph.D	15.16**	32.77**	22.75
$\tau = 0.75$	Quadratic	w/o Degree	1.36	1.66	1.44
		Bachelor	140.07**	148.19**	692.16**
		Master	65.99**	61.67**	290.77**
		Ph.D	18.68**	20.45**	98.20**
	Cubic	w/o Degree	2.34	2.42	1.50
		Bachelor	287.81**	156.67**	693.07**
		Master	91.11**	64.60**	293.62**
		Ph.D	24.79**	23.02**	98.64**
	Quartic	w/o Degree	2.76	2.72	1.84
		Bachelor	307.25**	158.45**	691.63**
		Master	98.89**	69.41**	294.82**
		Ph.D	24.74**	23.90**	98.71*

Table 6: INFERENCE RESULTS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS (NON-SCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal log income paths across different genders at quantiles $\tau = 0.25, 0.50$, and 0.75. The affixes '*' and '**' signify rejection of the null hypothesis at the 5% and 1% significance levels.

Inference results on the quantiles of the rescaled log income path across different genders					
		Education Level	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree	2.34	0.40	0.46
		Bachelor	0.52	1.29	0.56
		Master	3.09	0.96	4.63
		Ph.D	5.95	11.47**	12.61
	Cubic	w/o Degree	3.68	4.43	1.83
		Bachelor	2.57	3.09	1.07
		Master	2.74	1.95	6.20
		Ph.D	8.13	13.26*	13.17
	Quartic	w/o Degree	5.08	8.01	4.02
		Bachelor	2.92	2.17	0.88
		Master	5.31	2.49	5.67
		Ph.D	12.44*	11.86*	13.35
$\tau = 0.5$	Quadratic	w/o Degree	1.12	2.17	2.07
		Bachelor	1.78	2.35	5.35
		Master	2.36	2.64	1.38
		Ph.D	3.76	8.13*	6.94
	Cubic	w/o Degree	1.35	6.53	5.29
		Bachelor	4.44	6.29	5.83
		Master	3.78	10.93*	3.38
		Ph.D	4.17	16.84**	8.14
	Quartic	w/o Degree	13.74*	10.39	5.24
		Bachelor	13.27*	8.87	8.10
		Master	3.35	12.00*	2.37
		Ph.D	8.58	15.74**	8.69
$\tau = 0.75$	Quadratic	w/o Degree	1.90	1.54	1.14
		Bachelor	0.30	0.27	1.22
		Master	1.51	1.17	0.98
		Ph.D	1.37	2.12	0.82
	Cubic	w/o Degree	1.68	2.02	1.37
		Bachelor	2.65	2.10	1.09
		Master	1.97	1.77	1.32
		Ph.D	1.43	1.71	0.81
	Quartic	w/o Degree	6.19	4.74	2.47
		Bachelor	3.92	3.91	1.70
		Master	3.33	8.25	1.78
		Ph.D	1.44	2.35	0.78

Table 7: INFERENCE RESULTS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS (RESCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal rescaled log income paths across different genders at quantiles $\tau = 0.25, 0.50$, and 0.75. The affixes '*' and '**' signify rejection of the null hypothesis at the 5% and 1% significance levels.

			Male			Female		
			Wald	LM	QLR	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree vs. Bachelor	1027.90**	1433.40**	8223.50**	1888.40**	1465.30**	6498.80**
		Bachelor vs. Master	138.07**	79.61**	764.64**	348.92**	141.31**	777.12**
		Master vs. Ph.D	29.06**	17.46**	90.89*	18.17**	19.37**	20.25
	Cubic	w/o Degree vs. Bachelor	1534.20**	1469.20**	8301.10**	3359.40**	1530.00**	6566.00**
		Bachelor vs. Master	284.93**	87.00**	766.95**	604.28**	167.03**	784.66**
		Master vs. Ph.D	33.61**	18.39**	94.12*	22.59**	17.08**	19.85
	Quartic	w/o Degree vs. Bachelor	1501.30**	1523.40**	8297.70**	4090.80**	1634.80**	6581.80**
		Bachelor vs. Master	290.83**	88.60**	769.47**	642.94**	177.83**	783.86**
		Master vs. Ph.D	36.78**	19.41**	96.78*	30.86**	18.17**	19.98
$\tau = 0.5$	Quadratic	w/o Degree vs. Bachelor	1408.30**	984.33**	9024.80**	1329.10**	996.06**	6613.20**
		Bachelor vs. Master	166.87**	157.57**	1074.80**	253.49**	214.95**	1154.90**
		Master vs. Ph.D	16.29**	23.97**	90.13*	23.06**	27.53**	67.28*
	Cubic	w/o Degree vs. Bachelor	2631.70**	992.88**	8995.20**	2632.20**	1065.90**	6612.50**
		Bachelor vs. Master	384.62**	178.07**	1093.40**	519.79**	241.20**	1163.00**
		Master vs. Ph.D	21.81**	22.20**	89.81*	37.06**	32.67**	70.32*
	Quartic	w/o Degree vs. Bachelor	2727.20**	1007.50**	8993.80**	2744.20**	1093.70**	6606.60**
		Bachelor vs. Master	398.37**	186.70**	1098.50**	582.02**	258.73**	1162.80**
		Master vs. Ph.D	21.93**	21.99**	88.10*	41.31**	37.38**	68.68*
$\tau = 0.75$	Quadratic	w/o Degree vs. Bachelor	1079.60**	511.69**	6297.30**	527.86**	482.93**	4459.80**
		Bachelor vs. Master	130.01**	181.08**	1075.60**	103.03**	172.33**	713.65**
		Master vs. Ph.D	27.57**	25.36**	76.43*	41.07**	31.76**	171.82**
	Cubic	w/o Degree vs. Bachelor	2069.40**	499.91**	6254.30**	1002.50**	500.93**	4419.20**
		Bachelor vs. Master	183.45**	189.39**	1073.00**	173.75**	192.26**	711.00**
		Master vs. Ph.D	61.46**	30.18**	80.98*	71.59**	29.80**	176.09**
	Quartic	w/o Degree vs. Bachelor	2090.00**	498.15**	6255.30**	1025.90**	503.29**	4407.50**
		Bachelor vs. Master	203.95**	196.07**	1077.10**	184.88**	219.93**	709.44**
		Master vs. Ph.D	73.05**	35.54**	81.53*	73.41**	29.81**	178.75**

Inference results on the quantiles of the log income path across different education levels

Table 8: INFERENCE RESULTS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS (NON-SCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal log income paths across different education levels at quantiles $\tau = 0.25, 0.50$, and 0.75. The affixes '*' and '**' signify rejection of the null hypothesis at the 5% and 1% significance levels.

			Male			Female		
			Wald	LM	QLR	Wald	LM	QLR
$\tau = 0.25$	Quadratic	w/o Degree vs. Bachelor	9.79*	2.07	11.10	3.70	2.89	7.73
		Bachelor vs. Master	11.87**	4.82	19.27*	1.88	1.98	1.93
		Master vs. Ph.D	8.16*	3.56	5.39	3.56	4.01	8.76
	Cubic	w/o Degree vs. Bachelor	9.87*	4.53	15.44	4.62	5.17	11.02
		Bachelor vs. Master	16.51**	5.84	22.78*	2.17	1.41	1.74
		Master vs. Ph.D	8.11	3.84	7.15	4.05	6.38	12.96
	Quartic	w/o Degree vs. Bachelor	12.96*	7.18	14.95	12.82*	5.68	12.31
		Bachelor vs. Master	18.97**	6.52	21.99*	2.62	1.25	1.02
		Master vs. Ph.D	9.25	4.14	7.49	6.55	5.22	10.35
$\tau = 0.5$	Quadratic	w/o Degree vs. Bachelor	2.25	1.11	4.08	0.66	1.42	3.27
		Bachelor vs. Master	2.81	4.74	12.61	9.68*	7.17	28.24**
		Master vs. Ph.D	1.92	3.01	7.14	3.71	7.70	22.18*
	Cubic	w/o Degree vs. Bachelor	2.23	2.15	4.98	1.59	1.83	5.18
		Bachelor vs. Master	3.45	5.81	15.02	11.13*	7.10	27.11*
		Master vs. Ph.D	1.99	3.19	8.67	5.43	8.13	22.49*
	Quartic	w/o Degree vs. Bachelor	3.19	6.06	5.42	3.49	1.82	4.88
		Bachelor vs. Master	4.70	5.20	14.60	19.77**	7.39	26.37*
		Master vs. Ph.D	2.97	5.13	9.20	14.09*	8.12	22.08
$\tau = 0.75$	Quadratic	w/o Degree vs. Bachelor	0.99	1.95	1.99	2.98	2.86	0.28
		Bachelor vs. Master	6.01	4.79	6.11	3.93	2.99	8.40
		Master vs. Ph.D	4.10	3.56	8.53	2.31	2.82	8.19
	Cubic	w/o Degree vs. Bachelor	1.18	3.06	1.89	4.25	3.56	0.12
		Bachelor vs. Master	6.60	5.37	7.48	4.42	4.15	10.28
		Master vs. Ph.D	4.25	3.80	8.32	2.97	3.53	8.69
	Quartic	w/o Degree vs. Bachelor	2.62	4.58	2.61	4.59	4.34	0.33
		Bachelor vs. Master	10.90	6.83	7.90	4.82	5.30	10.76
		Master vs. Ph.D	10.69	5.84	9.15	4.33	3.33	9.27

Inference results on the rescaled quantiles of the rescaled log income path across different education levels

Table 9: INFERENCE RESULTS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS (RESCALED). This table shows the Wald, LM, and QLR test statistics and the inference results for the null hypothesis of the equal rescaled log income paths across different education levels at quantiles $\tau = 0.25, 0.50$, and 0.75. The affixes '*' and '**' signify rejection of the null hypothesis at the 5% and 1% significance levels.

Online Supplement for 'Functional Data Inference in a Parametric Quantile Model'*

by

Jin Seo Cho[a], Peter C. B. Phillips[b,c,d,e] and Juwon Seo[f]

[a]Yonsei University

[b]University of Auckland, [c]Yale University, [d]Singapore Management University,

[e]University of Southampton

[f]National University of Singapore

This Online Supplement is an Appendix that provides proofs of all the results in the paper, including the lemmas, as well as some additional empirical findings. Proofs are given in Section A.1 and the supplementary empirical application is in Section A.2

A Appendix

A.1 Proofs

Proof of Lemma 1: Note that $d_\tau(\gamma, u) := \mathbb{E}[\xi_\tau(G(\gamma) - u)] - \mathbb{E}[\xi_\tau(G(\gamma) - x_\tau(\gamma))] = uF_\gamma(u) - uF_\gamma(x_\tau(\gamma)) + \int_{-\infty}^{x_\tau(\gamma)} gdF_\gamma(g) - \int_{-\infty}^{u} gdF_\gamma(g)$. Applying integration by parts, $x_\tau(\gamma)F_\gamma(x_\tau(\gamma)) = \int_{-\infty}^{x_\tau(\gamma)} F_\gamma(g)dg + \int_{-\infty}^{x_\tau(\gamma)} gdF_\gamma(g)$, and $uF_\gamma(u) = \int_{-\infty}^{u} F_\gamma(g)dg + \int_{-\infty}^{u} gdF_\gamma(g)$, giving $\int_{-\infty}^{x_\tau(\gamma)} g\ dF_\gamma(g) - \int_{-\infty}^{u} gdF_\gamma(g) = x_\tau(\gamma)F_\gamma(x_\tau(\gamma)) - uF_\gamma(u) + \int_{-\infty}^{x_\tau(\gamma)} F_\gamma(g)dg - \int_{-\infty}^{u} F_\gamma(g)dg$. Hence, $d_\tau(\gamma, u) = (x_\tau(\gamma) - u)F_\gamma(x_\tau(\gamma)) + \int_{-\infty}^{x_\tau(\gamma)} F_\gamma(g)dg - \int_{-\infty}^{u} F_\gamma(g)dg$. Further, if $x_\tau(\gamma) > u$, then $d_\tau(\gamma, u) = \int_{u}^{x_\tau(\gamma)} \{F_\gamma(x_\tau(\gamma)) - F_\gamma(g)\}dg$; and if $x_\tau(\gamma) < u$, then $d_\tau(\gamma, u) = \int_{x_\tau(\gamma)}^{u} \{F_\gamma(g) - F_\gamma(x_\tau(\gamma))\}dg$, so that $d_\tau(\gamma, u) := \mathbb{E}[\xi_\tau(G(\gamma) - u)] - \mathbb{E}[\xi_\tau(G(\gamma) - x_\tau(\gamma))] = \int_{\min[u, x_\tau(\gamma)]}^{\max[u, x_\tau(\gamma)]} |F_\gamma(g) - F_\gamma(x_\tau(\gamma))|dg$. This completes the proof. ∎

Proof of Theorem 1: First note that (3) implies that $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^*) = -A_\tau^{*-1} n^{-1/2} \sum_{i=1}^{n} J_{\tau i} + o_{\mathbb{P}}(1)$. Given that θ_τ^* is identified as given in Assumption 2, the first-order condition holds, so that $\mathbb{E}[J_{\tau i}] = 0$. Assumption 4 also implies that $B_\tau^* := \mathbb{E}[J_{\tau i} J_{\tau i}']$ is positive definite. Furthermore, Assumption 3 implies that for each $j = 1, 2, \ldots, c_\tau$, $\mathbb{E}[J_{\tau ij}^2] < \infty$, where $J_{\tau ij}$ is the j-th row element of $J_{\tau i}$. Therefore, $n^{-1/2} \sum_{i=1}^{n} J_{\tau i} \overset{\text{A}}{\sim} \mathcal{N}(0, B_\tau^*)$ by the multivariate CLT, so that $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^*) \overset{\text{A}}{\sim} \mathcal{N}(0, A_\tau^{*-1} B_\tau^* A_\tau^{*-1})$. This completes the proof. ∎

Proof of Lemma 2: For simplicity let $\rho_\tau(\gamma') := \rho_\tau(\gamma', \theta_\tau^0)$ and show stochastic equicontinuity of $n^{-1/2}$

*Phillips acknowledges research support from the NSF under Grant No. SES 18-50860 at Yale University and a Kelly Fellowship at the University of Auckland.

$\sum_{i=1}^{n}(\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot, \theta_\tau^0)\} - \tau)$ using Ossiander's L^2 entropy condition: for some $\nu > 0$ and $C > 0$,

$$\mathbb{E}\left(\sup_{\|\gamma-\gamma'\|<\delta} |\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^0)\} - \tau - (\mathbb{1}\{G_i(\gamma' \leq \rho_\tau(\gamma', \theta_\tau^0))\} - \tau)|^2\right)^{1/2} \leq C\delta^\nu.$$

To verify this, first note that if we let $U_i(\gamma) := F_\gamma(G_i(\gamma))$, where $F_\gamma(\cdot)$ is the CDF of $G_i(\gamma)$, the left side is identical to

$$\mathbb{E}\left(\sup_{\|\gamma-\gamma'\|<\delta} |\mathbb{1}\{U_i(\gamma) - \tau \leq 0\} - \mathbb{1}\{U_i(\gamma') - \tau \leq 0\})|^2\right)^{1/2}$$

by noting that $F_\gamma(\rho_\tau(\gamma, \theta_\tau^0)) = \tau$ and $F_{\gamma'}(\rho_\tau(\gamma', \theta_\tau^0)) = \tau$. Next, apply the proof in Andrews (1994, p. 2779), letting his U_t and $h^*(Z_t, \cdot)$ be 1 and $U_i(\cdot) - \tau$, respectively and note that Assumption 1 implies that $U_i(\cdot)$ is Lipschitz continuous almost surely: for some $C > 0$, $|U_i(\gamma) - U_i(\gamma')| \leq C\|\gamma - \gamma'\|$. Here, we further note that $U_i(\gamma)$ is uniformly distributed over $[0, 1]$, so that its density function is bounded above uniformly on Γ. Therefore, example 3 in Andrews (1994, p. 2779) proves equicontinuity by Ossiander's L^2 entropy condition.

Next derive the covariance structure of the Gaussian stochastic process $\mathcal{G}_\tau(\cdot)$, noting that for each γ and γ',

$$\begin{aligned}
&\mathbb{E}[(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^0)\} - \tau)(\mathbb{1}\{G_i(\gamma') \leq \rho_\tau(\gamma', \theta_\tau^0)\} - \tau)] \\
&\quad = \mathbb{E}[(\mathbb{1}\{U_i(\gamma) \leq \tau\} - \tau)(\mathbb{1}\{U_i(\gamma') \leq \tau\} - \tau)] \\
&\quad = \mathbb{E}[\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\gamma') \leq \tau\}] - \tau\mathbb{E}[\mathbb{1}\{U_i(\gamma) \leq \tau\}] - \tau\mathbb{E}[\mathbb{1}\{U_i(\gamma') \leq \tau\}] + \tau^2 \\
&\quad = \mathbb{E}[\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\gamma') \leq \tau\}] - \tau^2 = \kappa(\gamma, \gamma'),
\end{aligned}$$

where the final equality follows from the fact that $\mathbb{E}[\mathbb{1}\{U_i(\gamma) \leq \tau\}] = \tau$ uniformly on γ. This completes the proof. ■

Proof of Theorem 2: Given Lemma 2, we note by continuous mapping that

$$\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^0)\} - \tau\right) d\mathbb{Q}(\gamma) \Rightarrow \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0)\mathcal{G}_\tau(\gamma)d\mathbb{Q}(\gamma)$$

which follows a normal distribution since $\mathcal{G}_\tau(\cdot)$ is a Gaussian stochastic process. Further note that applying dominated convergence using Assumption 3,

$$\mathbb{E}\left[\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0)\mathcal{G}_\tau(\gamma)d\mathbb{Q}(\gamma)\right] = \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0)\mathbb{E}[\mathcal{G}_\tau(\gamma)]d\mathbb{Q}(\gamma) = 0 \quad \text{and}$$

$$\mathbb{E}\left[\int_{\gamma}\int_{\gamma'} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathcal{G}_\tau(\gamma)\mathcal{G}_\tau(\gamma')\nabla_{\theta_\tau}\rho_\tau(\gamma',\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')\right]$$
$$= \int_{\gamma}\int_{\gamma'} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathbb{E}[\mathcal{G}_\tau(\gamma)\mathcal{G}_\tau(\gamma')]\nabla'_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')$$
$$= \int_{\gamma}\int_{\gamma'} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\kappa_\tau(\gamma,\gamma')\nabla'_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma') =: B_\tau^0,$$

by the definition of $\kappa_\tau(\cdot,\cdot)$. Therefore, $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathcal{G}_\tau(\gamma)d\mathbb{Q}(\gamma) \sim \mathcal{N}(0,B_\tau^0)$ where B_τ^0 is positive definite by Assumption 5. This fact further implies that

$$\sqrt{n}(\widehat{\theta}_{\tau n}-\theta_\tau^0) \Rightarrow -A_\tau^{0\,-1}\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathcal{G}_\tau(\gamma)d\mathbb{Q}(\gamma) \sim \mathcal{N}(0,C_\tau^0),$$

as required. ■

Proof of Theorem 3: If we apply (4) to the misspecified model, it now follows that $\sqrt{n}(\widetilde{\theta}_{\tau n}-\theta_\tau^*) = -A_\tau^{*-1}n^{-1/2}\sum_{i=1}^n \widehat{J}_{\tau i} + o_{\mathbb{P}}(1)$. We focus on $n^{-1/2}\sum_{i=1}^n \widehat{J}_{\tau i}$ to derive the limit distribution. Apply (A.1), as given in the proof of Lemma 3, to the misspecified model giving, for each $\gamma\in\Gamma$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{J}_{\tau i} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\mathbb{1}\{\widehat{G}_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau\Big)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)+\nabla'_\pi\widetilde{G}_i(\gamma,\bar{\pi}_{\gamma n})(\widehat{\pi}_n-\pi^*)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau\Big)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big[\Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau\Big) - f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi\widetilde{G}_i(\gamma,\pi^*)]P^{*-1}S_i\Big] + o_{\mathbb{P}}(1).\end{aligned}$$

Here, we applied the ULLN to obtain $n^{-1}\sum_{i=1}^n \nabla_\pi\widetilde{G}_i(\cdot,\bar{\pi}_{\gamma n}) \overset{\mathbb{P}}{\to} \mathbb{E}[\nabla_\pi\widetilde{G}_i(\cdot,\pi^*)]$ by using Assumption 8. It now follows that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{J}_{\tau i} =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)\Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau\Big)d\mathbb{Q}(\gamma)\\ &- \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi\widetilde{G}_i(\gamma,\pi^*)]d\mathbb{Q}(\gamma)P^{*-1}S_i + o_{\mathbb{P}}(1).\end{aligned}$$

Here, Assumptions 6 and 8 imply that $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi\widetilde{G}_i(\gamma,\pi^*)]d\mathbb{Q}(\gamma)$ is well defined. We further note that $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau)$ and $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi\widetilde{G}_i(\gamma,\pi^*)]d\mathbb{Q}(\gamma)$ are defined as $J_{\tau i}$ and K_τ^*, respectively, so that we can rewrite this equation as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{J}_{\tau i} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (J_{\tau i} - K_\tau^* P^{*-1}S_i) + o_{\mathbb{P}}(1).$$

Given this result, Assumptions 2 and 7 imply that $\mathbb{E}[J_{\tau i}] = 0$ and $\mathbb{E}[S_i] = 0$. Furthermore, Assumption 9 implies that $\widetilde{B}_\tau^* := \mathbb{E}[(J_{\tau i} - K_\tau^* P^{*-1} S_i)(J_{\tau i} - K_\tau^* P^{*-1} S_i)']$ is positive definite, and for each $j = 1, 2, \ldots, c_\tau$, $\mathbb{E}[J_{\tau ij}^2] < \infty$ and $\mathbb{E}[S_{ij}^2] < \infty$ by Assumptions 7 and 8. It now follows by the multivariate CLT that $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}_\tau^*)$, as required. ■

Proof of Lemma 3: We first derive the covariance kernel of $\widetilde{\mathcal{G}}_\tau(\cdot)$. Note that for any c, if $a > 0$, $\mathbb{1}\{x \le c - a\} = \mathbb{1}\{x \le c\} - \mathbb{1}\{x \in (c-a, c]\}$. On the other hand, if $a < 0$, $\mathbb{1}\{x \le c - a\} = \mathbb{1}\{x \le c\} + \mathbb{1}\{x \in (c, c-a]\}$. Therefore, $\mathbb{1}\{x \le c - a\} = \mathbb{1}\{x \le c\} - \mathbb{1}\{c - a < x \le c\} + \mathbb{1}\{c < x \le c - a\}$.

We use this equality to show the given claim. For notational simplicity, let $x_\tau(\gamma)$ and $\widehat{\mu}_{ni}(\gamma)$ denote $\rho_\tau(\gamma, \theta_\tau^0)$ and $\nabla_\pi' G_i(\gamma, \bar{\pi}_{\gamma n})(\widehat{\pi}_n - \pi^*)$, respectively. If we further let x, c, and a be $G_i(\gamma)$, $x_\tau(\gamma)$, and $\widehat{\mu}_{ni}(\gamma)$, respectively, it now follows that $\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) + \nabla_\pi' G_i(\gamma, \bar{\pi}_{\gamma n})(\widehat{\pi}_n - \pi^*) \le \rho_\tau(\gamma, \theta_\tau^0)\} = \mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} + \mathbb{1}\{x_\tau(\gamma) < \widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma) - \widehat{\mu}_{ni}(\gamma)\} - \mathbb{1}\{x_\tau(\gamma) - \widehat{\mu}_{ni}(\gamma) < \widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\}$. Note that Assumption 8 implies that $\nabla_\pi \widetilde{G}_i(\cdot, \cdot) = O_{\mathbb{P}}(1)$ and $(\widehat{\pi}_n - \pi^*) = o_{\mathbb{P}}(1)$, so that $\widehat{\mu}_{ni}(\gamma) = o_{\mathbb{P}}(1)$ uniformly in γ. Therefore,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \in (x_\tau(\gamma), x_\tau(\gamma) - \widehat{\mu}_n(\gamma)]\} - \mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \in (x_\tau(\gamma) - \widehat{\mu}_n(\gamma), x_\tau(\gamma)]\}\\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma) - \widehat{\mu}_{ni}(\gamma)\} - F_\gamma(x_\tau(\gamma))\Big) - \Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} - F_\gamma(x_\tau(\gamma))\Big)\\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} - F_\gamma(x_\tau(\gamma))\Big) + \frac{1}{n} \sum_{i=1}^n f_\gamma(x_\tau(\gamma)) \nabla_\pi' \widetilde{G}_i(\gamma, \pi^*) \sqrt{n}(\widehat{\pi}_n - \pi^*)\\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} - F_\gamma(x_\tau(\gamma))\Big) + o_{\mathbb{P}}(1)\\
&= f_\gamma(x_\tau(\gamma)) \mathbb{E}[\nabla_\pi' \widetilde{G}_i(\gamma, \pi^*)] \sqrt{n}(\widehat{\pi}_n - \pi^*) + o_{\mathbb{P}}(1),
\end{aligned}
$$

where the second equality follows from that $n^{-1/2} \sum_{i=1}^n \mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma) - \widehat{\mu}_{ni}(\gamma)\} = n^{-1/2} \sum_{i=1}^n \mathbb{1}\{ \widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} + n^{-} \sum_{i=1}^n F_\gamma'(x_\tau(\gamma)) \sqrt{n} \widehat{\mu}_{ni}(\gamma) + o_{\mathbb{P}}(1)$ and applying the mean-value theorem at the limit. Note that $F_\gamma'(x_\tau(\gamma)) = f_\gamma(x_\tau(\gamma))$. Therefore,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) + \nabla_\pi' \widetilde{G}_i(\gamma, \bar{\pi}_{\gamma n})(\widehat{\pi}_n - \pi^*) \le \rho_\tau(\gamma, \theta_\tau^0)\} - \tau\Big)\\
&\qquad = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} - \tau\Big) + f_\gamma(x_\tau(\gamma)) \mathbb{E}[\nabla_\pi' \widetilde{G}_i(\gamma, \pi^*)] \sqrt{n}(\widehat{\pi}_n - \pi^*) + o_{\mathbb{P}}(1)\\
&\qquad = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big[\Big(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le x_\tau(\gamma)\} - \tau\Big) - f_\gamma(x_\tau(\gamma)) \mathbb{E}[\nabla_\pi' \widetilde{G}_i(\gamma, \pi^*)] P^{*-1} S_i\Big] + o_{\mathbb{P}}(1) \quad \text{(A.1)}
\end{aligned}
$$

Given this, we compute the covariance kernel using the summand on the right side of (A.1), viz.,

$$
\begin{aligned}
&\mathbb{E}[[(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \tau) + f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma,\pi^*)]P^{*-1}S_i] \\
&\quad \times [(\mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*) \le x_\tau(\gamma')\} - \tau) + f_{\gamma'}(x_\tau(\gamma'))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma',\pi^*)]P^{*-1}S_i]] \\
&\qquad = \mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \tau)(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \tau)] \\
&\qquad\quad - f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma,\pi^*)]P^{*-1}\mathbb{E}[S_i(\mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*) \le x_\tau(\gamma')\} - \tau)] \\
&\qquad\quad - f_{\gamma'}(x_\tau(\gamma'))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma',\pi^*)]P^{*-1}\mathbb{E}[S_i(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \tau)] \\
&\qquad\quad + f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma,\pi^*)]P^{*-1}H^*P^{*-1}\mathbb{E}[\nabla_\pi \widetilde{G}_i(\gamma',\pi^*)]f_{\gamma'}(x_\tau(\gamma)).
\end{aligned}
$$

Observing that $\mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \tau)(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \tau)] = \kappa_\tau(\gamma,\gamma')$, the desired covariance kernel $\widetilde{\kappa}_\tau(\gamma,\gamma')$ is now obtained from this equality.

We next prove that the left side of (A.1) is stochastically equicontinuous. We let $\varsigma(\gamma) := f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla'_\pi \widetilde{G}_i(\gamma,\pi^*)]P^{*-1}$ for notational simplicity and show that the right side of (A.1) satisfies the bound condition to apply Ossiander's L^2 entropy condition: for some C and $\nu > 0$,

$$
\mathbb{E}\left(\sup_{\|\gamma-\gamma'\|<\delta} |(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} + \varsigma(\gamma)S_i) - (\mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*) \le x_\tau(\gamma')\} + \varsigma(\gamma')S_i)|^2\right)^{1/2} \le C\delta^\nu. \tag{A.2}
$$

We here note that

$$
\begin{aligned}
&\sup_{\|\gamma-\gamma'\|<\delta} |(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} + \varsigma(\gamma)S_i) - (\mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*) \le x_\tau(\gamma')\} + \varsigma(\gamma')S_i)|^2 \\
&\qquad \le \sup_{\|\gamma-\gamma'\|<\delta} |\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} - \mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*) \le x_\tau(\gamma')\}|^2 \\
&\qquad\quad + \sup_{\|\gamma-\gamma'\|<\delta} \|\varsigma(\gamma) - \varsigma(\gamma')\| \cdot \|S_i\| + \sup_{\|\gamma-\gamma'\|<\delta} |(\varsigma(\gamma) - \varsigma(\gamma'))S_i)|^2.
\end{aligned}
$$

In the proof of Lemma 2, we already saw that there are C_1 and $\nu_1 > 0$ such that

$$
\mathbb{E}\left(\sup_{\|\gamma-\gamma'\|<\delta} |(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\}) - (\mathbb{1}\{\widetilde{G}_i(\gamma',\pi^*) \le x_\tau(\gamma')\})|^2\right)^{1/2} \le C_1\delta^{\nu_1}.
$$

Next, Assumptions 2, 6, and 8 imply that $\varsigma(\cdot)$ is Lipschitz continuous, because a composition of Lipschitz continuous functions is Lipschitz continuous, and the product of two Lipschitz continuous functions is Lipschitz continuous: for some $m > 0$, $\|\varsigma(\gamma) - \varsigma(\gamma')\| \le m\|\gamma - \gamma'\|$, so that for some C_2 and $\nu_2 > 0$,

$$
\mathbb{E}\left(\sup_{\|\gamma-\gamma'\|<\delta} |(\varsigma(\gamma) - \varsigma(\gamma'))| \cdot |S_i|\right) \le C_2\delta^{\nu_2}
$$

by letting $C_2 := ms \cdot \max_{j=1,\dots,s} \mathbb{E}[|S_{ij}|^2]$ and $\nu_2 = 1$. Note that $\max_{j=1,\dots,s} \mathbb{E}[|S_{ij}|^2] < \infty$ from Assumption 7. We note that

$$|(\varsigma(\gamma) - \varsigma(\gamma'))S_i)|^2 \leq \max_{j=1,\dots,s} \mathbb{E}[S_{ij}|^2] \cdot \|\varsigma(\gamma) - \varsigma(\gamma')\|^2 \leq \max_{j=1,\dots,s} \mathbb{E}[S_{ij}|^2] \cdot m^2 \|\gamma - \gamma'\|^2,$$

so that if we let $C_3 := m^2 \cdot \max_{j=1,\dots,s} \mathbb{E}[|S_{ij}|^2]$ and $\nu_3 = 2$,

$$\mathbb{E}\left(\sup_{\|\gamma-\gamma'\|<\delta} |(\varsigma(\gamma) - \varsigma(\gamma'))S_i|^2 \right) \leq C_3 \delta^{\nu_3}.$$

Therefore, if we let $C := \max[C_1, C_2, C_3]$ and $\nu := \max[\nu_1, \nu_2, \nu_3]$, the desired inequality in (A.2) follows. This shows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) + \nabla'_\pi \widetilde{G}_i(\gamma, \bar{\pi}_{\gamma n})(\widehat{\pi}_n - \pi^*) \leq \rho_\tau(\gamma, \theta^0_\tau)\} - \tau \right)$$

is stochastically equicontinuous, completing the proof. ■

Proof of Theorem 4: Given Lemma 3, we note that

$$\int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma, \theta^0_\tau)\} - \tau \right) d\mathbb{Q}(\gamma) \Rightarrow \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \widetilde{\mathcal{G}}_\tau(\gamma) d\mathbb{Q}(\gamma)$$

by applying the continuous mapping theorem. The final integral follows a normal distribution from the fact that $\widetilde{\mathcal{G}}_\tau(\cdot)$ is a Gaussian stochastic process. By dominated convergence using Assumption 14,

$$\mathbb{E}\left[\int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \widetilde{\mathcal{G}}_\tau(\gamma) d\mathbb{Q}(\gamma) \right] = \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \mathbb{E}[\widetilde{\mathcal{G}}_\tau(\gamma)] d\mathbb{Q}(\gamma) = 0 \quad \text{and}$$

$$\begin{aligned}
\mathbb{E}\left[\int_\gamma \int_{\gamma'} \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \widetilde{\mathcal{G}}_\tau(\gamma) \widetilde{\mathcal{G}}_\tau(\gamma') \nabla_{\theta_\tau} \rho_\tau(\gamma', \theta^0_\tau) d\mathbb{Q}(\gamma) d\mathbb{Q}(\gamma') \right] \\
= \int_\gamma \int_{\gamma'} \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \mathbb{E}[\widetilde{\mathcal{G}}_\tau(\gamma) \widetilde{\mathcal{G}}_\tau(\gamma')] \nabla'_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) d\mathbb{Q}(\gamma) d\mathbb{Q}(\gamma') \\
= \int_\gamma \int_{\gamma'} \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \widetilde{\kappa}_\tau(\gamma, \gamma') \nabla'_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) d\mathbb{Q}(\gamma) d\mathbb{Q}(\gamma')
\end{aligned}$$

which is defined as $\widetilde{B}^0_\tau$. Therefore, $\int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \widetilde{\mathcal{G}}_\tau(\gamma) d\mathbb{Q}(\gamma) \sim \mathcal{N}(0, \widetilde{B}^0_\tau)$, implying that

$$\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta^0_\tau) \Rightarrow -A^{0\,-1}_\tau \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^0_\tau) \widetilde{\mathcal{G}}_\tau(\gamma) d\mathbb{Q}(\gamma) \sim \mathcal{N}(0, \widetilde{C}^0_\tau),$$

giving the desired result. ■

Proof of Theorem 5: (*i*) As the proof of the consistency in (*i*) is almost identical to that of (*ii*), we prove only (*i*).

(*i.a*) If we apply the mean-value theorem to $\rho_\tau(\gamma, \widehat{\theta}_{\tau n})$ and $\nabla_\theta \rho_\tau(\gamma, \widehat{\theta}_{\tau n})$ around the unknown parameter θ^*_τ, for each γ, there are $\bar{\theta}^*_\gamma$ and $\acute{\theta}^*_\gamma$ such that

$$\rho_\tau(\gamma, \widehat{\theta}_{\tau n}) = \rho_\tau(\gamma, \theta^*_\tau) + \nabla'_{\theta_\tau} \rho_\tau(\gamma, \bar{\theta}^*_{\tau\gamma})(\widehat{\theta}_{\tau n} - \theta^*_\tau) \quad \text{and}$$

$$\nabla_{\theta_\tau} \rho_\tau(\gamma, \widehat{\theta}_{\tau n}) = \rho_\tau(\gamma, \theta^*_\tau) + \nabla'_\theta \rho_\tau(\gamma, \acute{\theta}^*_{\tau\gamma})(\widehat{\theta}_{\tau n} - \theta^*_\tau).$$

For notational simplicity, let $\widehat{\nu}_n(\gamma) := -\nabla'_\theta \rho_\tau(\gamma, \bar{\theta}^*_\gamma)(\widehat{\theta}_{\tau n} - \theta^*_\tau)$. Given these expressions, note that

$$\begin{aligned}
\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \widehat{\theta}_{\tau n})\} - \tau &= \mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau)\} - \tau \\
&\quad + \mathbb{1}\{\rho_\tau(\gamma, \theta^*_\tau) < G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau) - \widehat{\nu}_n(\gamma)\} - \mathbb{1}\{\rho_\tau(\gamma, \theta^*_\tau) - \widehat{\nu}_n(\gamma) < G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau)\} \\
&= \mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau)\} - \tau + o_{\mathbb{P}}(1) \qquad \text{(A.3)}
\end{aligned}$$

using the fact that $\mathbb{1}\{\rho_\tau(\gamma, \theta^*_\tau) < G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau) - \widehat{\nu}_n(\gamma)\} = o_{\mathbb{P}}(1)$ and $\mathbb{1}\{\rho_\tau(\gamma, \theta^*_\tau) - \widehat{\nu}_n(\gamma) < G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau)\} = o_{\mathbb{P}}(1)$ from the fact that for each γ, $\widehat{\nu}_n(\gamma) = o_{\mathbb{P}}(1)$. It follows that

$$\begin{aligned}
J_{\tau n i} &:= \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \widehat{\theta}_{\tau n}) \left(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \widehat{\theta}_{\tau n})\} - \tau \right) d\mathbb{Q}(\gamma) + o_{\mathbb{P}}(1) \\
&= \int_\gamma \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta^*_\tau) \left(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta^*_\tau)\} - \tau \right) d\mathbb{Q}(\gamma) + o_{\mathbb{P}}(1) = J_{\tau i} + o_{\mathbb{P}}(1),
\end{aligned}$$

given that for each j and $j' = 1, 2, \ldots, c_\tau$, $|\partial^2/(\partial\theta_{\tau j}\partial\theta_{\tau j'})\rho_\tau(\cdot, \cdot)| \leq M < \infty$ and $|\partial/(\partial\theta_{\tau j})\rho_\tau(\cdot, \cdot)| \leq M < \infty$ from Assumption 8. Thus,

$$\widehat{B}_{\tau n} := \frac{1}{n} \sum_{i=1}^n J_{\tau n i} J'_{\tau n i} = \frac{1}{n} \sum_{i=1}^n J_{\tau i} J'_{\tau i} + o_{\mathbb{P}}(1).$$

We now further note that for each $j = 1, 2, \ldots, c_\tau$, $\mathbb{E}[J^2_{\tau i j}] < \infty$ from Assumption 8, so that it now follows that $\widehat{B}_{\tau n} = \frac{1}{n} \sum_{i=1}^n J_{\tau i} J'_{\tau i} + o_{\mathbb{P}}(1) \xrightarrow{\mathbb{P}} \mathbb{E}[J_{\tau i} J'_{\tau i}] =: B^*_\tau$. Therefore, $\widehat{B}_{\tau n} \xrightarrow{\mathbb{P}} B^*_\tau$.

Proof of consistency of $\widetilde{B}_{\tau n}$ is not detailed because it follows in a similar fashion to the consistency of $\widehat{B}_{\tau n}$. In particular, given the moment conditions in Assumption 8 and the condition for the other consistent estimators for P^*, H^*, and K^*_τ as given in Assumption 11, it follows that $\widehat{J}_{\tau n i} = \widehat{J}_{\tau i} + o_{\mathbb{P}}(1)$, and the result $\widetilde{B}_{\tau n} \xrightarrow{\mathbb{P}} \widetilde{B}^*_\tau$ follows.

(*i.b*) Given the second-order differentiability of $\rho_\tau(\gamma, \cdot)$ in Assumption 2 and Theorem 3(*i*), we apply (A.3) to obtain that

$$\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \widehat{\theta}_{\tau n})\} - \tau = \mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta^0_\tau)\} - \tau + o_{\mathbb{P}}(1),$$

implying that $\widehat{\kappa}_{\tau n}(\gamma,\gamma') = \widehat{\kappa}_{\tau}(\gamma,\gamma') + o_{\mathbb{P}}(1)$, where

$$\widehat{\kappa}_{\tau}(\gamma,\gamma') := \frac{1}{n}\sum_{i=1}^{n}(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^0)\} - \tau)(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^0)\} - \tau).$$

Furthermore, $\nabla_{\theta_\tau}\rho_\tau(\cdot,\widehat{\theta}_{\tau n}) \overset{\mathbb{P}}{\to} \nabla_{\theta_\tau}\rho_\tau(\cdot,\theta_\tau^0)$ from the fact that $\widehat{\theta}_{\tau n} \overset{\mathbb{P}}{\to} \theta_\tau^0$ and the continuity of $\rho_\tau(\cdot,\cdot)$. Therefore, from the definition of $\widehat{B}^{\sharp}_{\tau n}$, if $\widehat{\kappa}_\tau(\cdot,\cdot)$ is consistent for $\kappa_\tau(\cdot,\cdot)$ uniformly on $\Gamma \times \Gamma$, then the desired result follows.

For the proof of the consistency of $\widehat{\kappa}_\tau(\cdot,\cdot)$, we note that

$$\begin{aligned}\widehat{\kappa}_{\tau}(\gamma,\gamma') :=& \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^0)\}\mathbb{1}\{G_i(\gamma') \leq \rho_\tau(\gamma',\theta_\tau^0)\}\\ &- \frac{\tau}{n}\sum_{i=1}^{n}\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^0)\} - \frac{\tau}{n}\sum_{i=1}^{n}\mathbb{1}\{G_i(\gamma') \leq \rho_\tau(\gamma',\theta_\tau^0)\} + \tau^2.\end{aligned}$$

Here, the uniform consistency of $n^{-1}\sum_{i=1}^{n}\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot,\theta_\tau^0)\}$ follows if $\{\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot,\theta_\tau^0)\}\}$ is stochastically equicontinuous as shown in Newey (1991). Note that the proof of Lemma 2 already shows that $\{n^{-1}\sum_{i=1}^{n}\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot,\theta_\tau^0)\}\}$ is stochastically equicontinuous.

We therefore only show that $\{n^{-1}\sum_{i=1}^{n}\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot,\theta_\tau^0)\}\mathbb{1}\{G_i(\cdot') \leq \rho_\tau(\cdot',\theta_\tau^0)\}\}$ is stochastically equicontinuous for the uniform continuity of $\widehat{\kappa}_\tau(\cdot,\cdot)$, where "$(\cdot')$" is used to distinguish it from "$(\cdot)$". For this purpose, we use Ossiander's L^2 entropy condition as in the proof of Lemma 2: if we let $U_i(\gamma) := F_\gamma(G_i(\gamma))$ be the PIT of $G_i(\gamma)$ as in the proof of Lemma 2,

$$\begin{aligned}&\left|\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^0)\}\mathbb{1}\{G_i(\gamma') \leq \rho_\tau(\gamma',\theta_\tau^0)\} - \mathbb{1}\{G_i(\gamma'') \leq \rho_\tau(\gamma'',\theta_\tau^0)\}\mathbb{1}\{G_i(\gamma''') \leq \rho_\tau(\gamma''',\theta_\tau^0)\}\right|\\ &\quad= \left|\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\gamma') \leq \tau\} - \mathbb{1}\{U_i(\gamma'') \leq \tau\}\mathbb{1}\{U_i(\gamma''') \leq \tau\}\right|,\end{aligned}$$

so that Ossiander's L^2 entropy condition requires that there are are $\nu > 0$ and $C > 0$ such that

$$\mathbb{E}\left[\sup_{\|(\gamma,\gamma')-(\gamma'',\gamma''')\|<\delta}\left|\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\gamma') \leq \tau\} - \mathbb{1}\{U_i(\gamma'') \leq \tau\}\mathbb{1}\{U_i(\gamma''') \leq \tau\}\right|^2\right] \leq C^0\delta^\nu. \tag{A.4}$$

We first note that

$$\begin{aligned}&\left|\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\gamma') \leq \tau\} - \mathbb{1}\{U_i(\gamma'') \leq \tau\}\mathbb{1}\{U_i(\gamma''') \leq \tau\}\right|\\ &\quad= |\mathbb{1}\{U_i(\gamma) \leq \tau\}(\mathbb{1}\{U_i(\gamma') \leq \tau\} - \mathbb{1}\{U_i(\gamma'') \leq \tau\})\\ &\qquad\quad+ \mathbb{1}\{U_i(\gamma'') \leq \tau\}(\mathbb{1}\{U_i(\gamma) \leq \tau\} - \mathbb{1}\{U_i(\gamma''') \leq \tau\})|\\ &\quad\leq \left|\mathbb{1}\{U_i(\gamma') \leq \tau\} - \mathbb{1}\{U_i(\gamma'') \leq \tau\}\right| + \left|\mathbb{1}\{U_i(\gamma) \leq \tau\} - \mathbb{1}\{U_i(\gamma''') \leq \tau\}\right|.\end{aligned}$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\|(\gamma,\gamma')-(\gamma'',\gamma''')\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}\mathbb{1}\{U_i(\gamma')\le\tau\}-\mathbb{1}\{U_i(\gamma'')\le\tau\}\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right] \\
&\quad\le 2\mathbb{E}\left[\sup_{\|\gamma-\gamma''\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}-\mathbb{1}\{U_i(\gamma'')\le\tau\}\right|^2\right] \\
&\quad + 2\mathbb{E}\left[\sup_{\|\gamma-\gamma'\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}-\mathbb{1}\{U_i(\gamma')\le\tau\}\right|\cdot\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|\right].
\end{aligned}
$$

We here note that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\|\gamma-\gamma'\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}-\mathbb{1}\{U_i(\gamma')\le\tau\}\right|\cdot\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|\right]^2 \\
&\qquad\le \mathbb{E}\left[\left(\sup_{\|\gamma-\gamma'\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}-\mathbb{1}\{U_i(\gamma')\le\tau\}\right|\right)^2\right] \\
&\qquad\qquad\times\mathbb{E}\left[\left(\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|\right)^2\right]
\end{aligned}
$$

by applying Cauchy-Schwarz. Note that

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|\right\}^2\right] \\
&\qquad\le \mathbb{E}\left[\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right].
\end{aligned}
$$

It follows that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\|\gamma-\gamma'\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}-\mathbb{1}\{U_i(\gamma')\le\tau\}\right|\cdot\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|\right]^2 \\
&\qquad\le \mathbb{E}\left[\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right]^2,
\end{aligned}
$$

implying that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\|(\gamma,\gamma')-(\gamma'',\gamma''')\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}\mathbb{1}\{U_i(\gamma')\le\tau\}-\mathbb{1}\{U_i(\gamma'')\le\tau\}\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right]\\
&\qquad\le 2\mathbb{E}\left[\sup_{\|\gamma-\gamma''\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}-\mathbb{1}\{U_i(\gamma'')\le\tau\}\right|^2\right]\\
&\qquad\qquad\qquad + 2\mathbb{E}\left[\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right]\\
&\qquad\qquad = 4\mathbb{E}\left[\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right].
\end{aligned}
$$

We have already seen in the proof of Lemma 2 that there are $\nu > 0$ and $C > 0$ such that

$$
\mathbb{E}\left[\sup_{\|\gamma''-\gamma'''\|<\delta}\left|\mathbb{1}\{U_i(\gamma'')\le\tau\}-\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right]\le C\delta^{\nu},
$$

so that

$$
\mathbb{E}\left[\sup_{\|(\gamma,\gamma')-(\gamma'',\gamma''')\|<\delta}\left|\mathbb{1}\{U_i(\gamma)\le\tau\}\mathbb{1}\{U_i(\gamma')\le\tau\}-\mathbb{1}\{U_i(\gamma'')\le\tau\}\mathbb{1}\{U_i(\gamma''')\le\tau\}\right|^2\right]\le 4C\delta^{\nu}.
$$

We now let $C^0 := 4C$ in (A.4) for the same ν to complete the proof that $\{n^{-1}\sum_{i=1}^n \mathbb{1}\{G_i(\cdot)\le\rho_\tau(\cdot,\theta_\tau^0)\}\mathbb{1}\{G_i(\cdot')\le\rho_\tau(\cdot',\theta_\tau^0)\}\}$ is stochastically equicontinuous. This completes the proof. ■

Proof of Theorem 6: Theorem 3 (*ii*) implies that for each $j = 1, 2, \ldots, p$, $n^{-1/2}\sum_{i=1}^n \widehat{J}_{\tau_j i} \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{B}^*_{\tau_j})$. In addition, $\widetilde{B}^*_\tau$ is positive definite by Assumption 15. It therefore follows by the Cramér-Wold device that $n^{-1/2}\sum_{i=1}^n \widehat{J}_i \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{B}^*)$, so that $\sqrt{n}(\widetilde{\theta}_n - \theta^*) = -A^{*-1}n^{-1/2}\sum_{i=1}^n \widehat{J}_i + o_{\mathbb{P}}(1) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^*)$. ■

Proof of Lemma 4: To show the claim, for each $j = 1, 2, \ldots, p$, we first let

$$
\widehat{\omega}_{nj}(\gamma) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{1}\{\widehat{G}_i(\gamma)\le\rho_{\tau_j}(\gamma,\theta^0_{\tau_j})\}
$$

for notational simplicity. Second, note that Lemma 3 implies that for any $\epsilon_j > 0$ and $\eta_j > 0$, there exist n_{0j} and $\delta_j > 0$ such that if $n > n_{0j}$,

$$
\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta_j}\left|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\right| > \epsilon_j\right) < \eta_j. \tag{A.5}
$$

Third, applying the Cramér-Wold device gives the desired result. That is, for all $\lambda \in \mathbb{R}^p$ such that $\lambda'\lambda = 1$, if we show that for all $\epsilon > 0$ and $\eta > 0$, there are n_0 and $\delta > 0$ such that if $n > n_0$,

$$\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta}\left|\sum_{j=1}^{p}\lambda_j(\widehat{\omega}_{nj}(\gamma)-\tau_j)-\sum_{j=1}^{p}\lambda_j(\widehat{\omega}_{nj}(\gamma')-\tau_j)\right|>\epsilon\right)<\eta, \tag{A.6}$$

the desired result follows as in Wooldridge and White (1988, proposition 4.1). Here, for each $j = 1, 2, \ldots, p$, λ_j denotes the j-th row element of λ.

To show (A.6), we let $\epsilon > 0$ and $\eta > 0$ and show the stochastic equicontinuity using its definition. If $\lambda_j \neq 0$, we let ϵ_j and η_j be $\epsilon/(p \cdot |\lambda_j|)$ and η/p, respectively. Then, it follows that if $n > n_0 := \max[n_{01}, n_{02}, \ldots, n_{0p}]$,

$$\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta_j}\left|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\right|>\frac{\epsilon}{p\cdot|\lambda_j|}\right)<\frac{\eta}{p}$$

from (A.5). On the other hand, if $\lambda_j = 0$,

$$\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta_j}|\lambda_j|\cdot\left|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\right|>\frac{\epsilon}{p}\right)=0<\frac{\eta}{p}.$$

Therefore,

$$\sum_{j=1}^{p}\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta_j}|\lambda_j|\cdot\left|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\right|>\epsilon\right)<\eta,$$

and we also note that

$$\begin{aligned}\eta &> \sum_{j=1}^{p}\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta_j}|\lambda_j|\cdot\left|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\right|>\epsilon\right)\geq\mathbb{P}\left(\sum_{j=1}^{p}\sup_{\|\gamma-\gamma'\|<\delta_j}|\lambda_j|\cdot\left|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\right|>\epsilon\right)\\ &\geq\mathbb{P}\left(\sum_{j=1}^{p}\sup_{\|\gamma-\gamma'\|<\delta_j}\left|\lambda_j\{\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma')\}\right|>\epsilon\right)\geq\mathbb{P}\left(\sum_{j=1}^{p}\sup_{\|\gamma-\gamma'\|<\delta}\left|\lambda_j(\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\gamma'))\right|>\epsilon\right)\end{aligned}$$

by letting $\delta := \min[\delta_1, \delta_2, \ldots, \delta_p]$. That is, for each $\epsilon > 0$ and $\eta > 0$, there are n_0 and $\delta > 0$ such that

$$\mathbb{P}\left(\sup_{\|\gamma-\gamma'\|<\delta}\left|\sum_{j=1}^{p}\lambda_j(\widehat{\omega}_{nj}(\gamma)-\tau_j)-\sum_{j=1}^{p}\lambda_j(\widehat{\omega}_{nj}(\gamma')-\tau_j)\right|>\epsilon\right)<\eta.$$

This completes the proof. ∎

Proof of Theorem 7: Given Lemma 4, note that

$$\int_{\gamma} \nabla_{\theta}\rho(\gamma,\theta^0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{1}\{\widehat{G}_i(\gamma)\leq \rho(\gamma,\theta^0)\}-\tau\right)d\mathbb{Q}(\gamma)\Rightarrow \int_{\gamma}\nabla_{\theta}\rho(\gamma,\theta^0)\widetilde{\mathcal{G}}(\gamma)d\mathbb{Q}(\gamma)$$

by continuous mapping. Further note that the final integral follows a normal distribution from the fact that for each $j=1,2,\ldots,p$, $\widetilde{\mathcal{G}}_{\tau_j}(\cdot)$ is a Gaussian stochastic process. By dominated convergence theorem using Assumption 16,

$$\mathbb{E}\left[\int_{\gamma}\nabla_{\theta}\rho(\gamma,\theta^0)\widetilde{\mathcal{G}}(\gamma)d\mathbb{Q}(\gamma)\right]=\int_{\gamma}\nabla_{\theta}\rho(\gamma,\theta^0)\mathbb{E}[\widetilde{\mathcal{G}}(\gamma)]d\mathbb{Q}(\gamma)=0,\quad \text{and}$$

defining $\widetilde{\kappa}(\cdot,\cdot):\Gamma\times\Gamma\mapsto\mathbb{R}^{p\times p}$ such that its j-th row and j'-th column element is $\widetilde{\kappa}_{\tau_j,\tau_{j'}}(\cdot,\cdot)$ given in Lemma 4,

$$\begin{aligned}\mathbb{E}&\left[\int_{\gamma}\int_{\gamma'}\nabla_{\theta}\rho(\gamma,\theta^0)\widetilde{\mathcal{G}}(\gamma)\widetilde{\mathcal{G}}(\gamma')\nabla_{\theta}\rho(\gamma',\theta^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')\right]\\ &=\int_{\gamma}\int_{\gamma'}\nabla_{\theta}\rho(\gamma,\theta^0)\mathbb{E}[\widetilde{\mathcal{G}}(\gamma)\widetilde{\mathcal{G}}(\gamma')']\nabla_{\theta}'\rho(\gamma,\theta^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')\\ &=\int_{\gamma}\int_{\gamma'}\nabla_{\theta}\rho(\gamma,\theta^0)\widetilde{\kappa}(\gamma,\gamma')\nabla_{\theta}'\rho(\gamma,\theta^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\gamma')\end{aligned}$$

which yields $\widetilde{B}^0$. Therefore, $\int_{\gamma}\nabla_{\theta}\rho(\gamma,\theta^0)\widetilde{\mathcal{G}}(\gamma)d\mathbb{Q}(\gamma)\sim\mathcal{N}(0,\widetilde{B}^0)$. Given that $\widetilde{B}^0$ is positive definite by Assumption 16, it follows that

$$\sqrt{n}(\widetilde{\theta}_n-\theta^0)\Rightarrow -A^{0\,-1}\int_{\gamma}\nabla_{\theta}\rho(\gamma,\theta^0)\widetilde{\mathcal{G}}(\gamma)d\mathbb{Q}(\gamma)\sim\mathcal{N}(0,\widetilde{C}^0),$$

as desired, completing the proof. ■

Proof of Lemma 5: Since the asymptotic approximation of $\bar{\theta}_n$ is the same as that of $\ddot{\theta}_n$, we prove only (*ii*). Given that $\widehat{q}_n(\cdot)$ is stochastically differentiable in the sense of Pollard (1985, theorem 5), we can construct the Lagrange function to obtain the CTSFQR estimator (see also Newey and McFadden, 1994, section 7). The asymptotic first-order conditions are

$$\Omega\ddot{Q}_n+\ddot{D}_n'\ddot{\lambda}_n=o_{\mathbb{P}}(1)\quad\text{and}\quad R(\ddot{\theta}_n)\equiv 0, \tag{A.7}$$

where $\ddot{\lambda}_n$ stands for the asymptotic Lagrange multiplier. Note further that

$$\Omega\ddot{Q}_n=\Omega\widehat{Q}_n+\Omega A^*(\ddot{\theta}_n-\theta^*)+o_{\mathbb{P}}(1)\quad\text{and}\quad R(\ddot{\theta}_n)=R(\theta^*)+D^*(\widetilde{\theta}_n)(\ddot{\theta}_n-\theta^*)+o_{\mathbb{P}}(1), \tag{A.8}$$

where $\widehat{Q}_n := (n^{-1}\sum_{i=1}^n \widehat{J}_i)$. Solving for $(\ddot{\theta}_n - \theta^*)$ from these two conditions, it now follows that

$$\sqrt{n}(\ddot{\theta}_n - \theta^*) = ((\Omega A^*)^{-1} + (\Omega A^*)^{-1}D^{*\prime}E^{*-1}D^*(\Omega A^*)^{-1})\sqrt{n}\Omega\ddot{Q}_n \\ + ((\Omega A^*)^{-1}D^{*\prime}E^{*-1})\sqrt{n}R(\theta^*) + o_{\mathbb{P}}(\sqrt{n}),$$

where $E^* := -D^*(\Omega A^*)^{-1}D^{*\prime}$ and $\sqrt{n}\Omega\widehat{Q}_n \overset{\text{A}}{\sim} \mathcal{N}(0, \Omega\widetilde{B}^*\Omega)$ by applying Theorem 6. Next,

$$((\Omega A^*)^{-1} + (\Omega A^*)^{-1}D^{*\prime}E^{*-1}D^*(\Omega A^*)^{-1})\sqrt{n}\Omega\ddot{Q}_n \\ \overset{\text{A}}{\sim} \mathcal{N}(0,(\Omega A^*)^{-1}[(\Omega A^*) + D^{*\prime}E^{*-1}D^*](\Omega A^*)^{-1}\Omega\widetilde{B}^*\Omega(\Omega A^*)^{-1}[(\Omega A^*) + D^{*\prime}E^{*-1}D^*](\Omega A^*)^{-1}).$$

Here, we note that Ω and A^* are block diagonal matrices, so that the given asymptotic variance matrix simplifies to

$$[I + (\Omega A^*)^{-1}D^{*\prime}E^{*-1}D^*]\widetilde{C}^*[I + D^{*\prime}E^{*-1}D^*(\Omega A^*)^{-1}],$$

so that $\sqrt{n}\{(\ddot{\theta}_n - \theta^*) - ((\Omega A^*)^{-1}D^{*\prime}E^{*-1})R(\theta^*)\} \overset{\text{A}}{\sim} \mathcal{N}(0, [I + (\Omega A^*)^{-1}D^{*\prime}E^{*-1}D^*]\widetilde{C}^*[I + D^{*\prime}E^{*-1}D^* (\Omega A^*)^{-1}])$. Substituting $-D^*(\Omega A^*)^{-1}D^{*\prime}$ for E^*, the desired result follows. ■

Proof of Theorem 8: (*ii*) Since the proofs of (*i*) are almost identical to those of (*ii*), we prove only (*ii*).

(*ii.a*) Applying the mean-value theorem, $R(\widetilde{\theta}_n) = R(\theta^*) + \nabla'_\theta R(\theta_n^\flat)(\widetilde{\theta}_n - \theta^*)$ for some $\theta_n^\flat$ between $\widetilde{\theta}_n$ and θ^*, and if $\mathbb{H}_o$ is imposed, $\sqrt{n}R(\widetilde{\theta}_n) = \nabla'_\theta R(\theta_n^\flat)\sqrt{n}(\widetilde{\theta}_n - \theta^*)$. Note that $\theta_n^\flat \overset{\mathbb{P}}{\to} \theta^*$, so that $\sqrt{n}R(\widetilde{\theta}_n) = \nabla'_\theta R(\theta^*)\sqrt{n}(\widetilde{\theta}_n - \theta^*) + o_{\mathbb{P}}(1)$. Therefore, $\sqrt{n}R(\widetilde{\theta}_n) = \nabla'_\theta R(\theta^*)\sqrt{n}(\widetilde{\theta}_n - \theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \nabla'_\theta R(\theta^*)\widetilde{C}^*\nabla_\theta R (\theta^*))$ by Theorem 6 (*ii*). Since $\widetilde{D}_n \overset{\mathbb{P}}{\to} \nabla'_\theta R(\theta^*)$ it follows that $\widetilde{D}_n\widetilde{C}_n\widetilde{D}'_n$ consistently estimates the asymptotic variance matrix of $\sqrt{n}R(\widetilde{\theta}_n)$ from the fact that $\widetilde{A}_n$ is consistent for A^*. It therefore follows that $\ddot{\mathcal{W}}_n := nR(\widetilde{\theta}_n)'\{\widetilde{D}_n\widetilde{C}_n\widetilde{D}'_n\}^{-1}R(\widetilde{\theta}_n) \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$.

Under $\mathbb{H}_a$, $\sqrt{n}R(\widetilde{\theta}_n) = \sqrt{n}R(\theta^*) + \nabla'_\theta R(\theta_n^\flat)\sqrt{n}(\widetilde{\theta}_n - \theta^*)$, so that $\sqrt{n}R(\widetilde{\theta}_n) = O_{\mathbb{P}}(\sqrt{n})$ because $\sqrt{n}R(\theta^*) = O(\sqrt{n})$ and $\nabla'_\theta R(\theta_n^\flat)\sqrt{n}(\widetilde{\theta}_n - \theta^*) = O_{\mathbb{P}}(1)$, implying that $\ddot{\mathcal{W}}_n = O_{\mathbb{P}}(n)$. Therefore, if $c_n = o(n)$, $\lim_{n\to\infty} \mathbb{P}(\ddot{\mathcal{W}} \geq c_n) = 1$.

(*ii.b*) Solving for $\ddot{\lambda}_n$ from (A.7) and (A.8), $\sqrt{n}\ddot{\lambda}_n = -(E^{*-1}D^*(\Omega A^*)^{-1})\sqrt{n}\Omega\widehat{Q}_n - E^{*-1}\sqrt{n}R(\theta^*) + o_{\mathbb{P}}(\sqrt{n})$. Given that $\sqrt{n}\Omega\widehat{Q}_n \overset{\text{A}}{\sim} \mathcal{N}(0, \Omega\widetilde{B}^*\Omega)$, it follows that

$$\sqrt{n}\ddot{\lambda}_n + E^{*-1}\sqrt{n}R(\theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}), \tag{A.9}$$

so that, if $\mathbb{H}_o$ holds, $R(\theta^*) = 0$ and

$$n\ddot{\lambda}'_n\{E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}\}^{-1}\ddot{\lambda}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2. \tag{A.10}$$

Note that $\{E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}\}^{-1} = E^*(D^*\widetilde{C}^*D^{*\prime})^{-1}E^* = D^*(\Omega A^*)^{-1}D^{*\prime}(D^*\widetilde{C}^*D^{*\prime})^{-1}D^* (\Omega A^*)^{-1}$

$D^{*\prime}$ using the fact that $E^* := -D^*(\Omega A^*)^{-1}D^{*\prime}$. Therefore,

$$\begin{aligned}n\ddot{\lambda}_n'\{E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}\}^{-1}\ddot{\lambda}_n &= n\ddot{\lambda}_n'D^*(\Omega A^*)^{-1}D^{*\prime}(D^*\widetilde{C}^*D^{*\prime})^{-1}D^*(\Omega A^*)^{-1}D^{*\prime}\ddot{\lambda}_n\\ &= n\ddot{\lambda}_n'\ddot{D}_n(\Omega A^*)^{-1}\ddot{D}_n'(\ddot{D}_n\ddot{C}_n\ddot{D}_n')^{-1}\ddot{D}_n(\Omega A^*)^{-1}\ddot{D}_n\ddot{\lambda}_n + o_{\mathbb{P}}(1)\\ &= n\ddot{Q}_n'A^{*-1}\ddot{D}_n'(\ddot{D}_n\ddot{C}_n\ddot{D}_n')^{-1}\ddot{D}_nA^{*-1}\ddot{Q}_n + o_{\mathbb{P}}(1),\end{aligned}$$

where the penultimate equality follows because $\ddot{D}_n \overset{\mathbb{P}}{\to} D^*$ and $\ddot{B}_n \overset{\mathbb{P}}{\to} B^*$ under $\mathbb{H}_o$, as implied by Lemma 5 and the consistency of $\widetilde{A}_n$ for A^*. The last equality follows from (A.7) and the fact that Ω is a diagonal matrix. Note that this final expression is asymptotically equivalent to the definition of $\mathcal{L}\ddot{\mathcal{M}}_n$. So the desired result now follows from (A.10).

Under $\mathbb{H}_a$, note that $\Omega\ddot{Q}_n + \ddot{D}_n\ddot{\lambda}_n = o_{\mathbb{P}}(1)$ from (A.7) and $\sqrt{n}\ddot{\lambda}_n = O_{\mathbb{P}}(\sqrt{n})$ from (A.9), so that $\sqrt{n}\ddot{Q}_n = O_{\mathbb{P}}(\sqrt{n})$. Furthermore, $\ddot{D}_n = O_{\mathbb{P}}(1)$ and $\ddot{B}_n = O_{\mathbb{P}}(1)$ from Assumption 18, implying that $\mathcal{L}\ddot{\mathcal{M}}_n = O_{\mathbb{P}}(n)$. Therefore, if $c_n = o(n)$, $\lim_{n\to\infty}\mathbb{P}(\mathcal{L}\ddot{\mathcal{M}}_n \geq c_n) = 1$.

(*ii.c*) Given stochastic differentiability of $\widehat{q}_n(\cdot)$ in the sense of Pollard (1985, theorem 5), we can apply a second-order Taylor expansion around $\widetilde{\theta}_n$, so that $2n\{\widehat{q}_n(\ddot{\theta}_n)-\widehat{q}_n(\widetilde{\theta}_n)\} = n(\ddot{\theta}_n-\widetilde{\theta}_n)'\Omega A^*(\ddot{\theta}_n-\widetilde{\theta}_n)+o_{\mathbb{P}}(1)$ using the fact that the stochastic second derivative of $\widehat{q}_n(\cdot)$ is asymptotically equal to ΩA^* at θ^*. The proof of Lemma 5 already showed that $\sqrt{n}(\ddot{\theta}_n - \theta^*) - (\Omega A^*)^{-1}\sqrt{n}\Omega\ddot{Q}_n = \{(\Omega A^*)^{-1}D^*E^{*-1}D^{*\prime}(\Omega A^*)^{-1}\}\sqrt{n}\Omega$ $\ddot{Q}_n + ((\Omega A^*)^{-1}D^*E^{*-1})\sqrt{n}R(\theta^*) + o_{\mathbb{P}}(\sqrt{n})$, from which we further note that $(\Omega A^*)^{-1}\sqrt{n}\Omega\ddot{Q}_n = A^{*-1}$ $\sqrt{n}\ddot{Q}_n = (\widetilde{\theta}_n - \theta^*) + o_{\mathbb{P}}(1)$ as implied by Theorem 6. It follows that

$$\sqrt{n}(\ddot{\theta}_n - \widetilde{\theta}_n) = \{(\Omega A^*)^{-1}D^*E^{*-1}D^{*\prime}(\Omega A^*)^{-1}\}\sqrt{n}\Omega\ddot{Q}_n + ((\Omega A^*)^{-1}D^*E^{*-1})\sqrt{n}R(\theta^*) + o_{\mathbb{P}}(\sqrt{n}).$$

Hence, if $\mathbb{H}_o$ holds,

$$\begin{aligned}2n\{\widehat{q}_n(\ddot{\theta}_n) - \widehat{q}_n(\widetilde{\theta}_n)\} &= n\ddot{Q}_n'\Omega(\Omega A^*)^{-1}D^{*\prime}E^{*-1}\{D^*(\Omega A^*)^{-1}D^{*\prime}\}E^{*-1}D^*(\Omega A^*)^{-1}\Omega\ddot{Q}_n + o_{\mathbb{P}}(1)\\ &= n\ddot{Q}_n'A^{*-1}D^{*\prime}\{D^*(\Omega A^*)^{-1}D^{*\prime}\}^{-1}D^*A^{*-1}\ddot{Q}_n + o_{\mathbb{P}}(1),\end{aligned}$$

since $E^* := -D^*(\Omega A^*)^{-1}\,D^{*\prime}$. We further note that $\sqrt{n}D^*A^{*-1}\ddot{Q}_n \Rightarrow \widetilde{W} \sim \mathcal{N}(0, D^*A^{*-1}\widetilde{B}^*A^{*-1}D^{*\prime})$. It therefore follows that $\mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n := 2n\{\widehat{q}_n(\ddot{\theta}_n) - \widehat{q}_n(\widetilde{\theta}_n)\} \Rightarrow \widetilde{W}'\{D^*(\Omega A^*)^{-1}D^{*\prime}\}^{-1}\widetilde{W}$ under $\mathbb{H}_o$, as desired.

Under $\mathbb{H}_a$, $\sqrt{n}(\ddot{\theta}_n - \widetilde{\theta}_n) = O_{\mathbb{P}}(\sqrt{n})$ since $\{(\Omega A^*)^{-1}D^{*\prime}E^{*-1}D^*(\Omega A^*)^{-1}\}\sqrt{n}\Omega\ddot{Q}_n = O_{\mathbb{P}}(1)$ and $R(\theta^*) \neq 0$, so that $\mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n = O_{\mathbb{P}}(n)$. Therefore, if $c_n = o(n)$, $\lim_{n\to\infty}\mathbb{P}(\mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n \geq c_n) = 1$. This completes the proof. ■

A.2 Supplementary empirical applications

This section provides additional empirical material for Section 8. First, we provide the estimated $\rho_\tau(\cdot)$ for each group classified by gender and education. Using quadratic, cubic and quartic models for $x_\tau(\cdot)$, Figure

A.1 plots the estimated LIPs using work experiences over 0–40 years, and Figure A.2 plots the estimated LIPs using work experience over 10–40 years. The red, blue, and green lines in the figures denote the fitted LIPs obtained by the quadratic, cubic, and quartic specifications, respectively. The (three colored) curves at the top and the curves at the bottom of each figure are the estimated quantile LIPs for $\tau = 0.75$ and $\tau = 0.25$, respectively. The (three colored) curves in the middle of each figure are the median quantile functions for $\tau = 0.5$. As is apparent in the two figures, the shapes of the estimated quantile curves differ between Figures A.1 and A.2. In particular, the curves in Figure A.2 generally have less curvature and are closer to linearity than those of Figure A.1 which show different patterns depending on the polynomial specification. Further, the fitted quantile functions differ among the polynomial function specification. This feature indicates that the overall shape of the quantile function curve requires a reasonable degree of nonlinearity to accommodate the irregular patterns of the first 10 experience years in the income profiles.

Second, we report the estimation errors measured by $q_{\tau n}(\widehat{\theta}_\tau)$ in each group specification, capturing the value of the criterion function (2) at the estimate $\widehat{\theta}_\tau$. Tables A.1 and A.2 display the errors in the estimated LIPs using work experiences over 0–40 years and 10–40 years, respectively. As shown in the tables, the quartic specification provides the smallest $q_{\tau n}(\widehat{\theta}_\tau)$, and the quadratic specification yields the largest $q_{\tau n}(\widehat{\theta}_\tau)$ among the three specifications. Nonetheless, the quadratic, cubic, and quartic models yield similar estimation errors overall. In the lower panel of each table, we also report $q_{\tau n}(\widehat{\theta}_\tau)$ computed using the rescaled income paths that are obtained by dividing each individual LIP with its integral over the entire working experience profile. As in the nonscaled data case, the estimation errors decline as the degree of the polynomial function rises, although the overall results remain similar.

Estimated errors of the quantiles of the original log income path

		Male			Female		
		Quadratic	Cubic	Quartic	Quadratic	Cubic	Quartic
$\tau = 0.25$	w/o Degree	12.00	11.98	11.96	11.13	11.12	11.10
	Bachelor	10.82	10.77	10.67	10.25	10.20	10.10
	Master	10.93	10.70	10.59	9.91	9.80	9.68
	Ph.D	10.62	10.35	10.20	10.65	10.55	10.43
$\tau = 0.5$	w/o Degree	14.23	14.23	14.20	13.92	13.92	13.89
	Bachelor	13.52	13.39	13.27	12.87	12.80	12.67
	Master	13.59	13.28	13.16	12.22	12.06	11.89
	Ph.D	13.59	13.31	13.14	13.56	13.39	13.26
$\tau = 0.75$	w/o Degree	11.04	11.03	11.01	11.06	11.05	11.01
	Bachelor	10.86	10.66	10.61	10.29	10.21	10.11
	Master	10.85	10.58	10.52	9.78	9.66	9.54
	Ph.D	11.40	11.04	11.04	10.80	10.60	10.55

Estimated errors of the quantiles of the rescaled log income path

		Male			Female		
		Quadratic	Cubic	Quartic	Quadratic	Cubic	Quartic
$\tau = 0.25$	w/o Degree	5.87	5.85	5.76	5.63	5.63	5.56
	Bachelor	6.13	6.10	5.86	6.03	6.02	5.85
	Master	6.47	6.35	6.03	6.21	6.16	5.88
	Ph.D	6.45	6.27	5.85	6.58	6.54	6.22
$\tau = 0.5$	w/o Degree	6.83	6.83	6.77	6.65	6.65	6.59
	Bachelor	7.11	7.07	6.80	6.95	6.95	6.76
	Master	7.56	7.38	6.99	7.15	7.07	6.83
	Ph.D	7.53	7.28	6.79	7.50	7.48	7.21
$\tau = 0.75$	w/o Degree	5.20	5.20	5.19	5.04	5.02	5.01
	Bachelor	5.38	5.29	5.10	5.25	5.19	5.05
	Master	5.72	5.50	5.23	5.32	5.20	5.06
	Ph.D	5.70	5.46	5.14	5.62	5.51	5.36

Table A.1: ESTIMATION ERRORS USING FUNCTION DATA OVER 0 TO 40 WORK EXPERIENCE YEARS. This table shows the estimation errors of the original and rescaled log income paths using the quadratic, cubic and quartic models for each group of the workers classified according to their education levels and genders.

Estimated errors of the quantiles of the log income path

		Male			Female		
		Quadratic	Cubic	Quartic	Quadratic	Cubic	Quartic
$\tau = 0.25$	w/o Degree	8.97	8.94	8.94	8.12	8.10	8.10
	Bachelor	7.92	7.89	7.89	7.31	7.29	7.28
	Master	7.91	7.90	7.90	7.07	7.06	7.06
	Ph.D	7.65	7.63	7.62	7.57	7.56	7.55
$\tau = 0.5$	w/o Degree	10.65	10.64	10.63	10.19	10.18	10.18
	Bachelor	9.89	9.87	9.86	9.27	9.25	9.24
	Master	9.89	9.88	9.87	8.74	8.72	8.72
	Ph.D	9.93	9.90	9.89	9.82	9.80	9.80
$\tau = 0.75$	w/o Degree	8.28	8.27	8.27	8.10	8.09	8.09
	Bachelor	7.89	7.88	7.88	7.47	7.46	7.45
	Master	7.84	7.82	7.82	7.13	7.11	7.11
	Ph.D	8.32	8.32	8.32	7.95	7.95	7.95

Estimated errors of the quantiles of the rescaled log income path

		Male			Female		
		Quadratic	Cubic	Quartic	Quadratic	Cubic	Quartic
$\tau = 0.25$	w/o Degree	3.99	3.90	3.89	3.86	3.79	3.78
	Bachelor	3.87	3.81	3.80	3.90	3.85	3.84
	Master	3.79	3.74	3.74	3.81	3.75	3.75
	Ph.D	3.70	3.65	3.63	3.82	3.76	3.75
$\tau = 0.5$	w/o Degree	4.68	4.63	4.61	4.61	4.57	4.56
	Bachelor	4.54	4.49	4.48	4.57	4.54	4.53
	Master	4.43	4.39	4.38	4.44	4.42	4.41
	Ph.D	4.34	4.30	4.28	4.56	4.53	4.52
$\tau = 0.75$	w/o Degree	3.54	3.53	3.51	3.50	3.49	3.48
	Bachelor	3.43	3.42	3.41	3.46	3.46	3.45
	Master	3.35	3.35	3.34	3.32	3.33	3.32
	Ph.D	3.27	3.26	3.25	3.46	3.46	3.45

Table A.2: ESTIMATION ERRORS USING FUNCTION DATA OVER 10 TO 40 WORK EXPERIENCE YEARS. This table shows the estimation errors of the original and rescaled log income paths under the quadratic, cubic, and quartic for each group of the workers classified according to their education levels and genders.

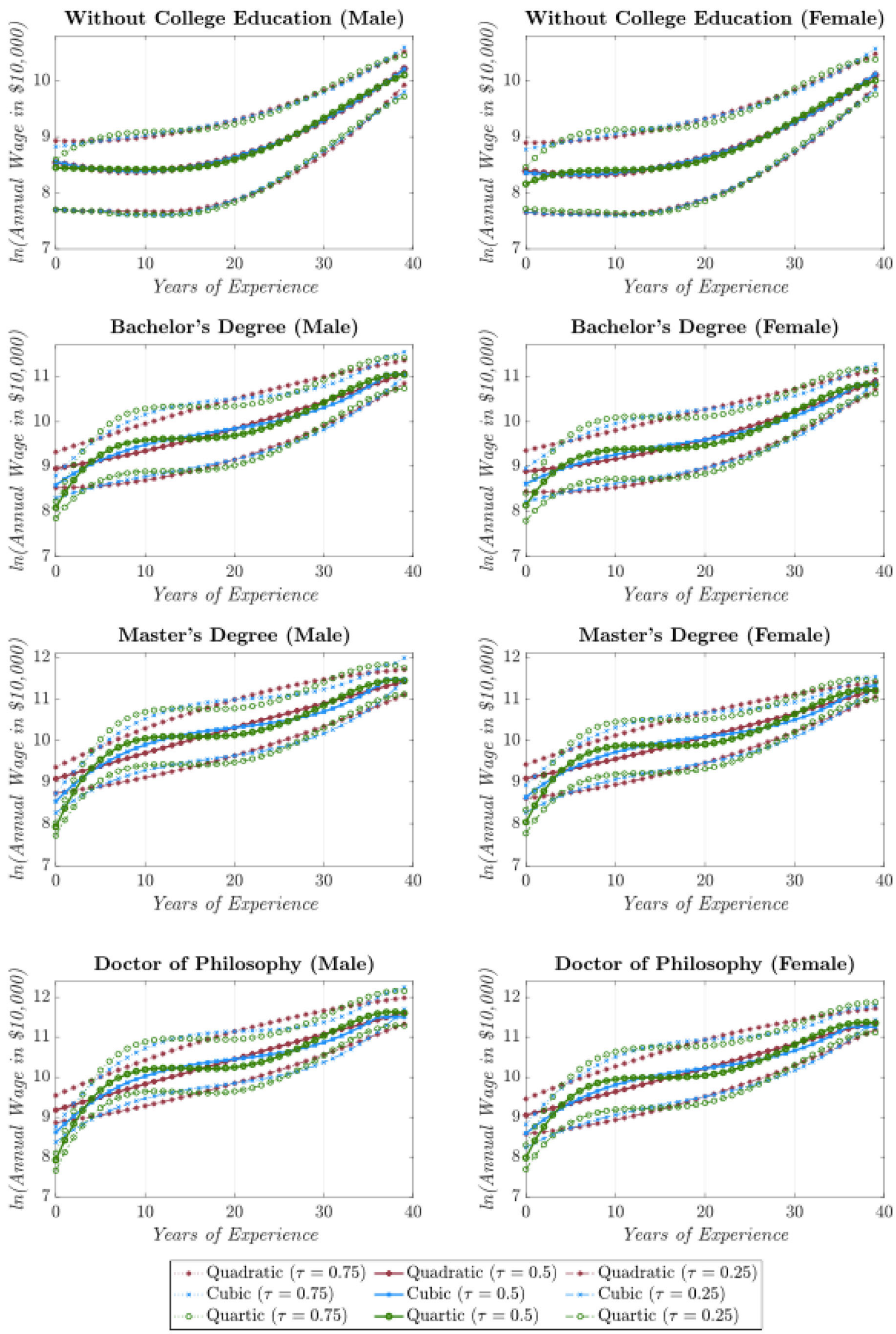

Figure A.1: ESTIMATED QUANTILE FUNCTIONS OVER 0 TO 40 WORK EXPERIENCE YEARS USING THE ORIGINAL LIPS.

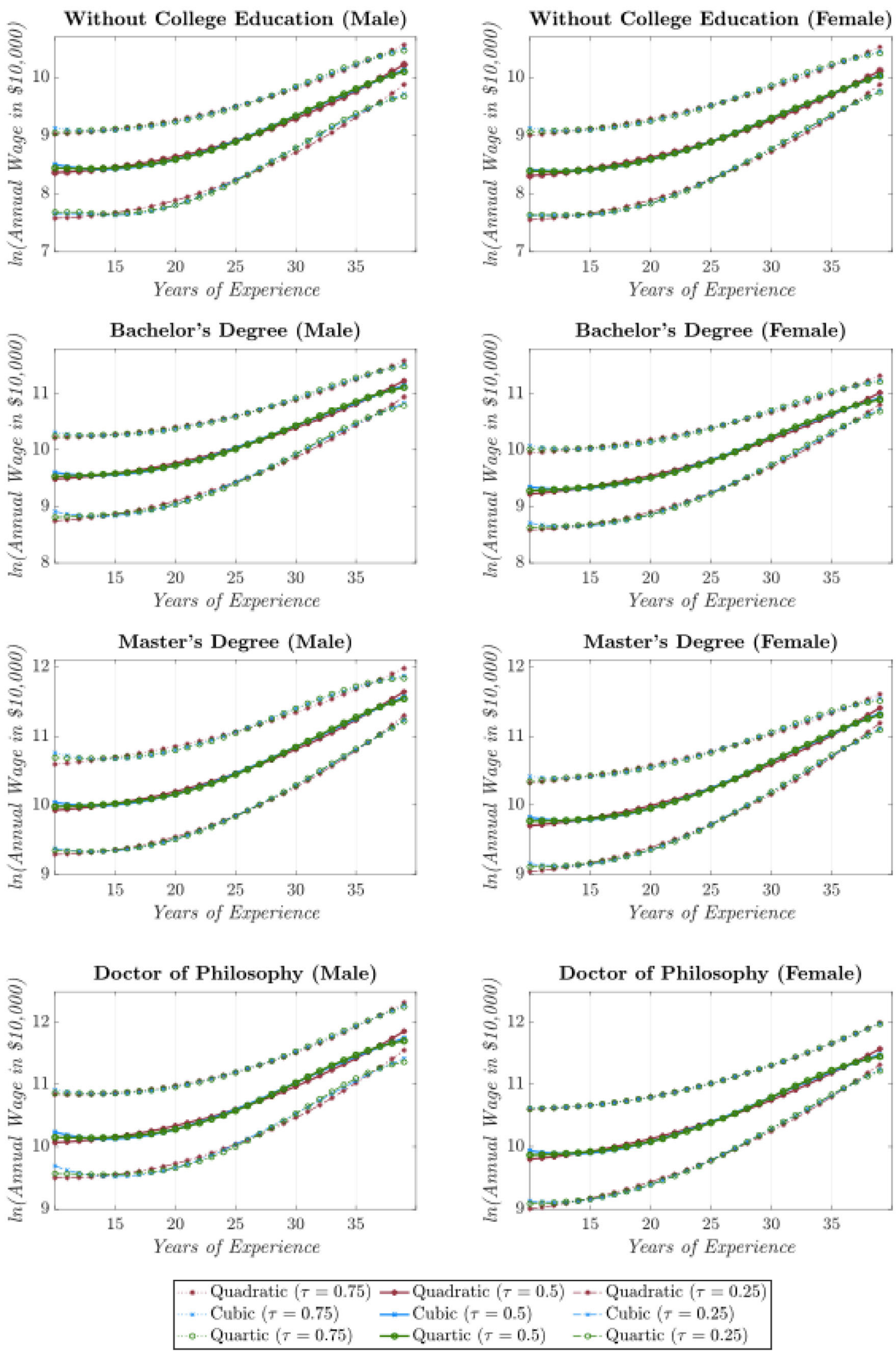

Figure A.2: ESTIMATED QUANTILE FUNCTIONS OVER 10 TO 40 WORK EXPERIENCE YEARS USING THE ORIGINAL LIPS.