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 & Singapore Management University
Email: peter.phillips@yale.edu

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

March 2026

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: The co-editor, Degui Li, and three referees each provided helpful comments for which we are most grateful. Cho acknowledges research support from the Yonsei University Research Grant of 2023; 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; and Seo acknowledges research support from the Tier 1 grant FY2023-FRC1-003.

*: Corresponding author.

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, and related methodology for functional data analysis has been developing accordingly. 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 to estimate autoregressive operator consistently along with the limit distribution of the estimator; Kato (2012) and 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 to estimate the quantile function between the dependent variable and transformed observations; Phillips and Jiang (2016, 2025) develop parametric autoregressive methods with function valued time series, establish asymptotic theory allowing for nonstationarity, and apply the methods to household Engel curves; Cho et al. (2022, hereafter, CPS (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 CPS (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). For example, Cardot et al. (2005) integrate functional data converting it to scalar variates and then estimate the parameters associated with the quantile function by quantile regression. But to our knowledge no attempt has been made in the prior literature to estimate the quantile function with the response variable itself being function valued.

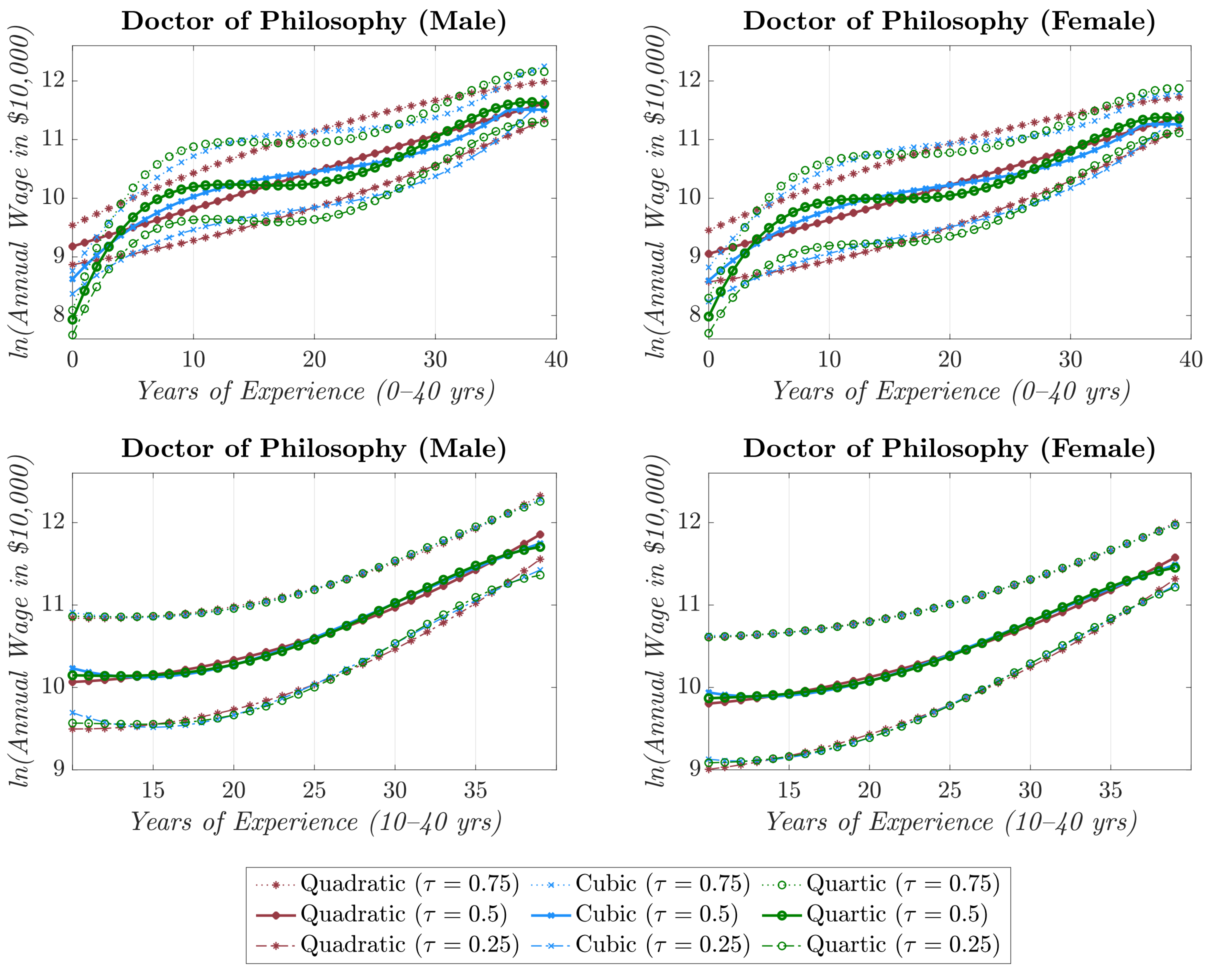

Figure 1: ESTIMATED QUANTILE CURVES AT LEVELS $\tau \in \{0.25, 0.50, 0.75\}$ FOR 0-40 YEAR AND 10-40 YEAR WORKING CAREERS OF MEN AND WOMEN WITH DOCTORAL LEVEL EDUCATION. This figure displays estimated quantile curves using functional datasets, specifically $\{G_i(\gamma) : \gamma \in [0, 40]\}_{i=1}^n$ and $\{G_i(\gamma) : \gamma \in [10, 40]\}_{i=1}^n$. In this context, $G_i(\cdot)$ means the i-th individual's interpolated log annual wage, and the domains [0, 40] and [10, 40] represent the ranges corresponding to careers of 0-40 years and 10-40 years, respectively, with n denoting the sample size.

The goal of the present study is to develop a methodology for modeling such quantile curves directly in function valued regressions with an asymptotic theory of estimation and inference that is useful to applied researchers interested in understanding all the features of functional data. These include those arising from parametric functions and those obtained by interpolating discretely observed data. To do so we develop a novel methodology for quantile regression that extends the methods in CPS (2022). Specifically, each quantile curve is formulated in a parametric form wherein the parametric coefficients measure the extent to which functional form the response curve exhibits at each percentile level. The unknown parameters are estimated using quantile regression with integral transforms of the functional observations. This approach enables us to analyze data with temporal and possibly persistent dependence structures embodied in the observations in a straightforward manner similar to the standard quasi-maximum likelihood estimation.

We illustrate the resulting outcomes of our approach using the log income profiles of workers that are analyzed as curves evolving over time and reflecting 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. Each individual's annual income levels over 30 years are interpolated by local polynomial kernel estimation to apply the so-called "smoothing first, then estimation" principle by following Zhang and Chen (2007), and we estimate the quantile time curves for each percentile. 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 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 highest level of 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 7 provides a detailed study with classifications and inferences about these curves according to gender and multiple education levels.

The methodological goal and empirical implementation are achieved by constructing a model framework for continuous data in parallel to that of customary extremum estimation and inference for simple random variables using methods such as quasi-maximum likelihood estimation. 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. This allowance extends the possibly misspecified linear quantile model assumption made by Kim and White (2003) and Angrist et al. (2006) to the context of functional data which, to the best of our knowledge, is novel in the literature for functional data. 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). While the model might not precisely reveal all the intrinsic properties of the data, it may approximate the data generating conditions sufficiently well when the parametric model is carefully defined to address key properties in the observations. For example, as depicted in Figure 1, we can observe the impact of progressive fine tuning the functional form of the quantile curve on the precision of the specification; and the use of a parametric model formulation naturally improves efficiency in estimation and inference. By contrast the task of pointwise quantile estimation for each element along a quantile curve is an intensive exercise and the determination of the overall shape of the quantile functions is not a straightforward assessment through hypothesis testing. In the modeling framework of the current study, the aim is for the researcher to view functional data analysis in the customary manner of which it is normally conducted for quasi-maximum likelihood estimation and inference.

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 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 potential nuisance effects and misspecification. In pursuit of this goal, it transpires that use of linear operators in functional data modeling plays a central role, enabling analysis in parallel to quasi-maximum likelihood estimation. Indeed, Wald, Lagrange multiplier (LM), and likelihood-ratio (LR) inference about the parametric curves is conducted in a customary manner using quasi-maximum likelihood estimation. The tests are asymptotically chi-squared under the null hypothesis and diverge under the alternative, so that tests can be applied straightforwardly in practical work with functional data. In our empirical work the methodology is applied to the 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, which helps to shed 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]

Functional data analysis of this type complements conventional panel data analysis. The approach has supportive rationales for empirical research. First, constructing panel data from individuals over a long period often introduces data quality problems such as data attrition and unbalanced panel data. Functional observations obtained by general interpolation methods bypass such difficulties by working with continuous functions that underlie conventional panels. For example, methods such as local polynomial, sieve, krigging, and polynomial spline estimations interpolate discrete observations to produce continuous functional obser-

[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 Figure S.2 and S.3 and attendant discussion in the Online Supplement.

vations (e.g., Zhang and Chen, 2007; Chen, 2007; Chilès and Delfiner, 1999; Wang, 2011). Specifically, Zhang and Chen (2007) discuss data conditions involving the cross section sample sizes and time periods of observation, showing that interpolation errors can be ignored asymptotically in parameter estimation and inference when using continuous functions obtained from longitudinal data by local polynomial methods. Second, continuous functional data are useful in characterizing the shape information of longitudinal data more directly and intimately than discrete observations. For example, the shapes of lifetime income curves are of key interest in the human capital literature concerned with the influence of gender, education and other variables over time (Mincer, 1958, 1974; Mincer and Jovanovic, 1981). Continuous sample paths assist in revealing the form of shape responses in monetary income against experience.

The organization of the paper follows. Section 2 motivates the use of functional data. Section 3 introduces models for functional data with nuisance effects and examines misspecified quantile function model estimation. Section 4 considers multiple quantile level estimation, and Section 5 develops inferential methods for quantile curve functions. Section 6 reports simulation findings and Section 7 applies the methodology to worker income profiles. Section 8 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 Environment, Model, and Estimation

We begin by describing the methodological framework. Let $G(\cdot)$ $(\in \mathbb{R})$ be an observable 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 draw inferences from the functional observations about γ and the functional form $G(\cdot)$. The quantile level $x_\tau(\gamma)$ as a function of γ is useful in discovering the stochastic properties of the continuous random function $G(\cdot)$. By estimating $x_\tau(\cdot)$, we can reveal the marginal distribution of $G(\gamma)$ indirectly, which may be difficult to capture just by estimating the mean of $G(\gamma)$. For if $F_\gamma(\cdot)$ is asymmetric or heavy-tailed, the population mean of $G(\gamma)$ may fail to be an informative summary for the distribution. The quantile levels $x_\tau(\gamma)$ help to reveal $F_\gamma(\cdot)$ better. But estimating and drawing inferences about $x_\gamma(\cdot)$ is challenging because γ lies in a continuum, making pointwise estimation of $x_\tau(\gamma)$ for every $\gamma \in \Gamma$ intensive. Our approach achieves the goal by extending the work of CPS (2022) to estimate the mean of $G(\gamma)$ as a function of γ under regularity conditions that enable $x_\tau(\cdot)$ to be efficiently estimated.

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. CPS (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 (e.g., Zhang and Chen, 2007), producing a continuous income path, viz., $G(\cdot)$ in that context. 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. The present study exlores this question more generally by treating the underlying path data as a continuous random function, represented by $G(\cdot)$, over career years and its quantile function $x_\tau(\cdot)$ is estimated directly. 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, for $\tau = 0.25$, 0.50, and 0.75, $x_\tau(\cdot)$ is estimated using continuous income paths belonging to the male and female groups holding doctorate degrees. Section 7 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. Letting Γ be the time domain considered, 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 the 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 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 may 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)* $(\mathbf{\Omega}, \mathcal{F}, \mathbb{P})$ *is a complete probability space and* Γ *is a compact metric space; (ii)* $\{G_i : \mathbf{\Omega} \times \Gamma \mapsto \mathbb{R}\}_{i=1}^n$ *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* $(\mathbf{\Omega} \times \Gamma, \mathcal{F} \otimes \mathcal{G}, \mathbb{P} \cdot \mathbb{Q})$ *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 CPS (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 define the QR distance as

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

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 $x_\tau(\gamma)$ 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. We note that the current notion of model specification generalizes that considered by Kim and White (2003) and Angrist et al. (2006) who assume a possibly misspecified linear model for the quantile of a random variable. Note that our object of interest is not a random variable but a random function indexed by γ.

To fix ideas, a simple linear specification of ρ_τ such as (6) in the Online Supplement and has the form $\rho_\tau(\gamma, \theta_\tau) = \theta_{\tau 1} + \theta_{\tau 2}\gamma$, with parameter vector $\theta_\tau = (\theta_{\tau 1}, \theta_{\tau 2})' \in \Theta_\tau \subset \mathbb{R}^2$, dimension $c_\tau = 2$, and $\gamma \in \Gamma \subset \mathbb{R}$. If $\mathcal{M}_\tau$ is correctly specified, there are $\theta^0_{\tau 1}$ and $\theta^0_{\tau 2}$ such that for each $\gamma \in \Gamma$, $x_\tau(\gamma) = \theta^0_{\tau 1} + \theta^0_{\tau 2}\gamma$, and we can use the parameters to figure out how the marginal distribution of $G(\gamma)$ evolves as γ varies. For example, if $\theta^0_{2\tau} = 0$ for every τ, it means that $F_\gamma(\cdot)$ does not vary as γ varies. Various fitted quadratic and higher order polynomial quantile curves are alternatives to this linear specification. Some examples are shown in the empirical illustration of Figure 1 and more are considered in the application of Section 7.

Two remarks follow. First, although the above illustrations are linear with respect to the parameters, we do not necessarily impose the linearity condition on $\rho_\tau(\gamma, \theta_\tau)$ with respect to θ_τ. For example, Kato (2012) assumes a linear operator to map from functional observation to random variable by quantile regression, and CPS (2022) assume a linear model in the mean regression context. Instead, we treat the linear model as a special case of our model. Oberhofer and Haupt (2016) provide a nonlinear QR estimation theory for a random variable, and we similarly use a theoretical groundwork to estimate the nonlinear parameter in $\mathcal{M}_\tau$. Various models for capturing nonlinear relationships are widely discussed in the literature. Among them, polynomial models are perhaps the most prevalent, as illustrated in the Introduction. In instances where polynomial models suffer from Runge's phenomenon—characterized by undesirable oscillations at the boundaries of the domain that can magnify near the ends of the interpolation points—alternative approaches such as splines or orthogonal series may be employed. Furthermore, while we consider models that are linear with respect to θ_τ, we also allow $\mathcal{M}_\tau$ to incorporate models that are intrinsically nonlinear in their parameters θ_τ. Second, the model formulation given in (1) indeed 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. Otherwise, estimating $x_\tau(\gamma)$ pointwisely for every $\gamma \in \Gamma$ would be an intensive task to the empirical researcher.

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

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

zero, letting $d_\tau(\theta_\tau)$ be an average of the non-zero distances weighted by the adjunct probability measure $\mathbb{Q}(\cdot)$. 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 $d_\tau(\cdot)$, irrespective of $\mathbb{Q}(\cdot)$. 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, $\theta_\tau^* = \arg\min_{\theta_\tau \in \Theta_\tau} q_\tau(\theta_\tau)$. Here, we deliberately define the objective function $q_\tau(\cdot)$ by integral transformation. As it turns out, this enables estimation and inference to be conducted in parallel to standard quasi-maximum likelihood estimation (see White, 1982), and estimators and tests are presented by applying linear transformations to the functional data and model. Mild regularity conditions are provided for the estimators and tests within this framework. Other operators might be used in defining the objective function at the cost of complicating estimation, inference and the corresponding limit theory under possibly stronger conditions than ours.[3]

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

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

and let $\widehat{\theta}_{\tau n} := \arg\min_{\theta_\tau \in \Theta_\tau} q_{\tau n}(\theta_\tau)$, which can be achieved in practice by numerical quadrature. 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 (quasi-FQR)* estimator. 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 the Online Supplement, $q_{\tau n}(\cdot)$ is consistent for $q_\tau(\cdot)$. Therefore, if θ_τ^* is unique and

[3] An even more general approach is to use a sieve approximation with a complete set $k \to \infty$ of orthonormal functions in $L_2[0, 1]$. We leave this as a future research topic.

$q_\tau(\cdot)$ is continuous on Θ_τ, the estimator $\widehat{\theta}_{\tau n}$ is consistent for θ_τ^* under some general regularity conditions on the model.

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 CPS (2022). The limit theory in both FQR and quasi-FQR 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 time domain Γ as before but such that $\widehat{G}(\cdot) := \widetilde{G}(\cdot, \widehat{\pi}_n)$, 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

$$\sqrt{n}(\widehat{\pi}_n - \pi^*) = -P^{*-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_i + o_{\mathbb{P}}(1),$$

where $S_i \in \mathbb{R}^s$ is a random sequence that satisfies central limit theory, viz., $n^{-1/2}\sum_{i=1}^n S_i \overset{\text{A}}{\sim} \mathcal{N}(0, H^*)$. 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 CPS (2022). We refer to these effects as the nuisance effects. Functional observations of the type in Section 2 can be regarded as data with no nuisance effects induced by parameter estimation by letting $G_i(\cdot) = \widetilde{G}_i(\cdot, \pi^*)$.

Several examples of functional data with nuisance effects fall into this framework. In the context of conditional moment specification testing, for instance, Bierens (1990) develops a testing methodology for $\mathbb{E}[(Y_i - m(X_i, \pi^*))\exp(\gamma' X_i)] = 0$ for every $\gamma \in \Gamma$, where $Y_i \in \mathbb{R}$ is a dependent variable and $m(X_i, \pi)$ is a model for $\mathbb{E}[Y_i|X_i]$ such that $X_i \in \mathbb{R}^d$. In this case interest stems from the fact that if for some π^*, $\mathbb{E}[Y_i|X_i] = m(X_i, \pi^*)$, then for every γ around zero, $\mathbb{E}[(Y_i - m(X_i, \pi^*))\exp(\gamma' X_i)] = 0$, as revealed by lemma 1 of Bierens (1990). If we further let $\widetilde{G}_i(\cdot, \pi^*) := (Y_i - m(X_i, \pi^*))\exp((\cdot)' X_i)$, we can view this as a functional observation. Indeed, Bierens (1990) replaces π^* with a nonlinear least squares estimator $\widehat{\pi}_n$ to form functional data $\{\widehat{G}_i(\cdot) := \widetilde{G}_i(\cdot, \widehat{\pi}_n) : i = 1, 2, \ldots, n\}$. As another example, when testing the mixture model hypothesis, the null score function of a single component can be regarded as a functional observation with a nuisance effect. Specifically, if we let the log-likelihood function be $\ell_i(\gamma, \pi, p) := \log(pf(X_i, \pi) + (1-p)f(X_i, \gamma))$, where $f(\cdot, \gamma)$ is a PDF, then γ^* is not identified under the null of $p^* = 1$. So Davies's (1977; 1987) identification problem arises and the null score function implied by $p^* = 1$ is $\widetilde{G}_i(\cdot, \pi^*, p^* = 1) = 1 - f(X_i, \cdot)/f(X_i, \pi^*)$ and $\mathbb{E}[\widetilde{G}_i(\cdot, \pi^*, 1)] \equiv 0$. If π^* is estimated by $\widehat{\pi}_n$, then the maximum likelihood estimator under the null, $\{\widehat{G}_i(\cdot) := \widetilde{G}_i(\cdot, \widehat{\pi}_n) : i = 1, 2, \ldots, n\}$, can be treated as a functional dataset with nuisance effect. In addition to these parametric model implied functional data, we note that functional data are often constructed by applying basis functions such as local polynomial kernel, sieve, polynomial spline, and kriging estimation using discrete observations (for example Zhang and Chen, 2007; Chilès and Delfiner, 1999; Wang, 2011). Such functional data are contaminated by the kernel or sieve estimation errors that can be captured by $\widehat{\pi}_n$ in the present framework. For example, we may suppose interpolating discrete observations using kernel density estimation: for a panel dataset: $\{g_{i,t}\}_{i=1,t=1}^{n,T}$, if we use a kernel function $K(\cdot)$ for interpolation, individual functional observations are obtained as

$$\widetilde{G}_i(\gamma,\pi) := \sum_{t=1}^{T} g_{i,t} K\left(\frac{t/T-\gamma}{\pi}\right) \Big/ \sum_{t=1}^{T} K\left(\frac{t/T-\gamma}{\pi}\right),$$

in which π plays the role of bandwidth. Likewise, many functional data popular in the literature are formed as $\widetilde{G}_i(\gamma,\pi) := \sum_{t=1}^{T} g_{i,t} W_t(\gamma,\pi)$, where $W_t(\gamma,\lambda)$ is the weight function (e.g., Fan and Gijbels, 1996, p. 44). For the above example, we can let $W_t(\gamma,\pi) = K\left((t/T-\gamma)/\pi\right) / \sum_{t=1}^{T} K\left((t/T-\gamma)/\pi\right)$, and under the assumption of the current study, it follows that $\{\widehat{G}_i(\cdot) = \widetilde{G}_i(\cdot,\widehat{\pi}_n) : i = 1,2,\ldots,n\}$.

Two purposes are served by the regularity condition on $\widehat{\pi}_n$. First, the regularity condition applies in many practical cases as the first two examples demonstrate. Second, as it turns out the (quasi-)2FQR estimator becomes asymptotically regular if $\widehat{\pi}_n$ has a $\sqrt{n}$ convergence rate. If $\widehat{\pi}_n$ has a convergence rate faster than $\sqrt{n}$, then the asymptotic distribution of the (quasi-)2FQR estimator becomes identical to that of the (quasi-)FQR estimator. Conversely, if $\widehat{\pi}_n$ has a convergence rate slower than $\sqrt{n}$, a correspondingly slower convergence rate is expected to affect the limit distribution of the (quasi-)2FQR estimator, meaning that the theory for (quasi-)2FQR differs from (quasi-)FQR once the convergence rate of $\widehat{\pi}_n$ equals $\sqrt{n}$. In view of this property we assume a $\sqrt{n}$ convergence rate for $\widehat{\pi}_n$, so that applicability of the (quasi-)2FQR estimator is affected by its asymptotic distribution, thereby differing from that of the (quasi)FQR estimator.

The quantile function can be estimated in parallel to (quasi-)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 $\widehat{G}(\cdot) := \widetilde{G}(\cdot,\widehat{\pi}_n)$, 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 (2FQR)* estimator in the case of a correctly specified model $\mathcal{M}_\tau$; otherwise, $\widetilde{\theta}_{\tau n}$ is called the *quasi-two-stage functional quantile regression (quasi-2FQR)* 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 $\widehat{G}(\cdot) := \widetilde{G}(\cdot,\widehat{\pi}_n)$ that embody nuisance effects.

Before proceeding, we formally collect the regularity conditions we discussed above and other conditions, that enable us to reveal the large sample properties of the (quasi-)FQR estimators.

Assumption 2. *(i) Let $(\mathbf{\Omega},\mathcal{F},\mathbb{P})$ 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) $\{\widetilde{G}_i : \mathbf{\Omega}\times\Gamma\times\Pi \mapsto \mathbb{R}\}_{i=1}^{n}$ 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 $(\mathbf{\Omega}\times\Gamma,\mathcal{F}\otimes\mathcal{G},\mathbb{P}\cdot\mathbb{Q})$ 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 $f_\gamma(\cdot)$ is 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 3. *There exists a sequence of measurable functions $\{\widehat{\pi}_n : \boldsymbol{\Omega} \mapsto \Pi\}$ 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 $s \times s$ symmetric matrix P^* such that $\lambda_{\min}(P^*) > 0$ and a sequence of $\mathcal{F}$-measurable random vectors $\{S_{n*}\}$, $\sqrt{n}(\widehat{\pi}_n - \pi^*) = -P^{*-1}\sqrt{n}S_n^* + o_{\mathbb{P}}(1)$; and (iii) for $i = 1, 2, \ldots$, there is $S_i : \boldsymbol{\Omega} \mapsto \mathbb{R}^s$ 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 4. *(i) For each $\theta_\tau \in \Theta_\tau$, $\rho_\tau(\,\cdot\,, \theta_\tau)$ 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 $\gamma \in \Gamma$, $\rho_\tau(\gamma, \cdot\,) \in \mathcal{C}^{(2)}(\Theta_\tau)$ and for $j, t = 1, 2, \ldots, c_\tau$, $(\partial/\partial\theta_{\tau j})\rho_\tau(\cdot,\cdot) \in \mathcal{C}(\Gamma \times \Theta_\tau)$, and $(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau t})\rho_\tau(\cdot,\cdot) \in \mathcal{C}(\Gamma \times \Theta_\tau)$; (iii) for each $\theta_\tau \in \Theta_\tau$, $\rho_\tau(\,\cdot\,, \theta_\tau) \in \mathcal{L}_{ip}(\Gamma)$; (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; and (v) $f_{(\cdot)}(\rho_\tau(\cdot,\cdot)) \in \mathcal{C}(\Gamma \times \Theta_\tau)$.* □

Assumption 5. *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 $j = 1, 2, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau)} |(\partial/\partial\theta_{\tau j})\rho_\tau(\gamma, \theta_\tau)| \leq M$; (v) for each $j, t = 1, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau)} |(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau t})\rho_\tau(\gamma, \theta_\tau)| \leq M$; and (vi) for each $j = 1, 2, \ldots, s$, $\mathbb{E}[(\partial/\partial\pi_j)\widetilde{G}_i(\cdot, \pi^*)] \in \mathcal{L}_{ip}(\Gamma)$ and for each θ_τ, $f_{(\cdot)}(\rho_\tau(\cdot, \theta_\tau)) \in \mathcal{L}_{ip}(\Gamma)$.* □

Assumptions 2 and 3 allow for functional data featuring nuisance effects, characterizing the implications of the parametric estimates $\widehat{\pi}_n$ embodied in the functional observations described in Assumption 1. Here, the interior element condition on π^* in Assumption 3 can be relaxed to a boundary parameter condition, so that π^* can lie on the boundary of Π. In such a case, $\widehat{\pi}_n$ can equal π^* with a positive probability, so that the functional observation may not be affected by the nuisance effects with positive probability. We leave this investigation as a future research topic. Assumption 4 gives conditions on the model $\mathcal{M}_\tau$. Here, Assumption 4 (v) can be heuristically stated as follows. Given that $G_i(\gamma)$ is continuous with respect to γ, its marginal distribution function is also continuous with respect to γ. Therefore, if we rewrite $f_\gamma(\cdot)$ as $f(\cdot, \gamma)$, then $f(\cdot, \circ)$ is continuous on $\mathbb{R} \times \Gamma$. In addition, if we suppose that $\rho_\tau(\cdot, \circ) \in \mathcal{C}(\Gamma \times \Theta_\tau)$, then it follows that $f(\rho(\cdot, \circ), \cdot)$ is a mapping defined on $\Gamma \times \Theta_\tau$ and belongs to $\mathcal{C}(\Gamma \times \Theta_\tau)$. Finally, Assumption 5 provides bound conditions for the functional observations and the model function that ensure regular behavior for (quasi-)FQR estimation. Under these conditions the (quasi-)FQR estimator is shown to be consistent for θ_τ^* and asymptotically normally distributed.

Before discussing the estimator, it is worth pointing out that the adjunct probability measure $\mathbb{Q}(\cdot)$ can be selected to help reduce the degree of potential misspecification. If $\mathcal{M}_\tau$ is misspecified, then $\rho_\tau(\cdot, \theta_\tau) \neq x_\tau(\cdot)$ for any θ_τ, so that $d_\tau(\theta_\tau) > 0$, and θ_τ^* depends on $\mathbb{Q}(\cdot)$. Therefore, if $\mathbb{Q}(\cdot)$ is extended to encompass more probability measures, say $\mathbb{Q}(\cdot, \delta)$ with δ in a certain parameter space, we can consider minimizing

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

with respect to θ_τ and δ, so that the selected adjunct probability measure places low and high chances on γ with high and low distances between $\rho_\tau(\gamma, \theta_\tau)$ and $x_\tau(\gamma)$, respectively, from which we can select θ_τ^* to reduce the degree of misspecification. Analogously, we may modify (2) to

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

and minimizing $q_{\tau n}(\theta_\tau, \delta)$ with respect to θ_τ and where δ allows the data to reveal the degree of model misspecification. This formulation involves a Bayesian perspective on the adequacy of the model. Specifically, by choosing an adjunct probability measure that reduces the QR distance, we make the empirical model closer to the correctly specified model in a manner that gives greater weight to the correctly specified region of γ than to misspecified ones. A number of adjunct probability measures can be used for this purpose. In the Online Supplement, we report simulations that explore the influence of selecting different adjunct probability measures on estimation and inference. We can also expect to estimate $x_\tau(\cdot)$ nonparametrically by choosing the adjunct probability measure appropriately. For example, we may apply local polynomial estimation by letting $\mathbb{Q}(\gamma, \delta)$ be a localized probability measure distributed around δ, so that the objective function can be defined as

$$q_{\tau n}(\theta_\tau, \delta) := \int_\gamma \frac{1}{n}\sum_{i=1}^{n} \xi_\tau\left(G_i(\gamma) - \sum_{j=0}^{p} \theta_{\tau j}(\gamma - \delta)^j\right) d\mathbb{Q}(\gamma - \delta),$$

where $p \in \mathbb{N}$ and $\theta_\tau := (\theta_{\tau 0}, \theta_{\tau,1}, \ldots, \theta_{\tau,p})'$. Here, $q_{\tau n}(\theta_\tau, \delta)$ is defined in parallel to the objective function for estimating the local polynomial model, and $\mathbb{Q}(\gamma - \delta)$ plays the role of the weights used in the local polynomial model estimation (e.g., Fan and Gijbels, 1996). Minimizing $q_{\tau n}(\cdot, \delta)$ then enables quantile function estimation nonparametrically.

3 Quasi-2FQR Estimation

For manageable discussions, we suppose that the model $\mathcal{M}_\tau$ may be misspecified and derive the asymptotic properties of the quasi-2FQR estimator. From the asymptotic properties, those of 2FQR, quasi-FQR, and FQR estimators follow.

3.1 Large Sample Distribution of the Quasi-2FQR Estimation

To analyze the quasi-2FQR estimator we use an asymptotic approximation of the functional quantile estimator. For each γ let the PDF of $\widetilde{G}_i(\gamma, \phi_*)$ be $f_\gamma(\cdot)$ and we apply the approximation approach in Oberhofer and Haupt (2016, pp. 710–711):

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

using the continuity conditions in Assumption 4, where $\nabla_{\theta_\tau} := \partial/\partial\theta_\tau$ and

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

This inequality is obtained by the directional derivative of $\widehat{q}_{\tau n}(\theta_\tau)$. That is, applying Lemma 2N of Oberhofer and Haupt (2016), we obtain the lower and upper bounds of the first-order directional derivative of $\widehat{q}_{\tau n}(\theta_\tau)$: for w such that $\|w\| = 1$,

$$\widetilde{R}_{ln}(w) \leq \frac{1}{\sqrt{n}}\sum_{i=1}^n w' \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\} - \tau) d\mathbb{Q}(\gamma) \leq \widetilde{R}_{un}(w),$$

where

$$\widetilde{R}_{ln}(w) := -\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_\gamma \mathbb{1}\{\widehat{G}_i(\gamma) = \rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\} |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})| \mathbb{1}\{w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n}) < 0\} d\mathbb{Q}(\gamma),$$

$$\widetilde{R}_{un}(w) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_\gamma \mathbb{1}\{\widehat{G}_i(\gamma) = \rho_\tau(\gamma,\widetilde{\theta}_{\tau n})\} |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n})| \mathbb{1}\{w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_{\tau n}) \geq 0\} d\mathbb{Q}(\gamma).$$

In the Online Supplement, we show that $\widetilde{R}_{ln}(w) = o_{\mathbb{P}}(1)$ and $\widetilde{R}_{un}(w) = o_{\mathbb{P}}(1)$, so that the desired inequality follows.

From this approximation we obtain the following representation

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

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, and further enables analysis of the FQR estimator as a special case of the quasi-FQR estimator, as detailed below.

The approximation implies that the limit behavior of $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^*)$ is determined by the two factors in the leading term on the right side of (3), which is useful in deriving its limit distribution. We note the following three key features from (3). First, the approximation on the right side is analogous to that of standard quasi-maximum likelihood estimation, which is essentially the result of defining the objective function through an integral transformation. Second, the matrix A_τ^* in the first factor involves only non random model components. For regular behavior of $\widetilde{\theta}_{\tau n}$ it is necessary for A_τ^* to be positive definite, as required in the Online Supplement. Third, the limit distribution of the quasi-FQR estimator is determined mainly by the other components on the right side of (3). From Assumptions 4 and 5 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

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

Hence, letting $J_{\tau i} := \int_{\gamma} \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*) \, (\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^*)\} - \tau) \, d\mathbb{Q}(\gamma)$ and with $B_\tau^* := \mathbb{E}[J_{\tau i} J_{\tau i}']$ positive definite, we can apply standard multivariate central limit theory (CLT) to $n^{-1/2} \sum_{i=1}^n J_{\tau i}$. Next, let $\widehat{J}_{\tau i} := \int_{\gamma} \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^*)(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^*)\} - \tau) d\mathbb{Q}(\gamma)$ and suppose that its asymptotic covariance matrix, denoted by $\widetilde{B}_\tau^*$, is positive definite. Then by standard multivariate central limit theory

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

As $\mathcal{M}_\tau$ is possibly misspecified, for each γ, $\mathbb{E}[\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^*)\}]$ is not necessarily identical to τ. 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 $\bar{\pi}_{\gamma n i}$ between π^* and $\widehat{\pi}_n$, we have

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

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 2, 3, and 5 which suffice to apply the uniform law of large numbers (ULLN). More specifically, $n^{-1} \sum_{i=1}^n \nabla_\pi \widetilde{G}_i(\cdot, \pi^*) \overset{\mathbb{P}}{\to} \mathbb{E}[\nabla_\pi \widetilde{G}_i(\cdot, \pi^*)]$ 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 quasi-2FQR estimation is based on the following additional regularity conditions, which ensure that the limit distribution of the quasi-2FQR estimator is non-degenerate.

Assumption 6. *(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 6, S_i, $J_{\tau i}$ and $\widehat{J}_{\tau i}$ all have positive definite variance matrices.

Theorem 1. *Given Assumptions 2, 3, 4, 5, and 6, if $\mathcal{M}_\tau$ is misspecified, $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}_\tau^*)$, where $\widetilde{C}_\tau^* := A_\tau^{*-1}\widetilde{B}_\tau^* A_\tau^{*-1}$.* □

Remarks 1. (a) Theorem 1 implies that the 2FQR estimator is asymptotically normal by noting that $\theta_\tau^* = \theta_\tau^0$. Nevertheless, the asymptotic distribution can be obtained using a stronger statement on the first-order condition. That is, if $\mathcal{M}_\tau$ is correct, it follows that

$$n^{-1/2}\sum_{i=1}^{n}\Big(\mathbb{1}\{\widehat{G}_i(\cdot) \le \rho_\tau(\cdot, \theta_\tau^0)\} - \tau\Big) \Rightarrow \widetilde{\mathcal{G}}_\tau(\cdot),$$

where $\widetilde{\mathcal{G}}_\tau(\cdot)$ is a zero-mean Gaussian process such that for each γ and $\bar{\gamma}$, $\mathbb{E}[\widetilde{\mathcal{G}}_\tau(\gamma)\widetilde{\mathcal{G}}_\tau(\bar{\gamma})] = \widetilde{\kappa}_\tau(\gamma, \bar{\gamma})$ with

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

and $\kappa_\tau(\gamma,\bar{\gamma}) := \mathbb{E}[\mathbb{1}\{F_\gamma(G_i(\gamma)) \le \tau\}\mathbb{1}\{F_\gamma(G_i(\bar{\gamma})) \le \tau\}] - \tau^2$. Using this statement, we can obtain the asymptotic covariance matrix as $\widetilde{C}_\tau^0 := A_\tau^{0\,-1}\widetilde{B}_\tau^0 A_\tau^{0\,-1}$, where

$$A_\tau^0 := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0) f_\gamma(\rho_\tau(\gamma,\theta_\tau^0))\nabla_{\theta_\tau^0}'\rho_\tau(\gamma,\theta_\tau^0)d\mathbb{Q}(\gamma) \quad \text{and}$$

$$\widetilde{B}_\tau^0 := \int_\gamma\int_{\bar{\gamma}} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\widetilde{\kappa}_\tau(\gamma,\bar{\gamma})\nabla_{\theta_\tau}'\rho_\tau(\bar{\gamma},\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\bar{\gamma}),$$

so that $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^0) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}_\tau^0)$. Here, $\widetilde{B}_\tau^0$ is the covariance matrix of $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathcal{G}_\tau(\gamma)d\mathbb{Q}(\gamma)$. This weak convergence is established by proving that $n^{-1/2}\sum_{i=1}^n\big(\mathbb{1}\{G_i(\cdot) \le \rho_\tau(\cdot,\theta_\tau^0)\} - \tau\big)$ is stochastically equicontinuous, thereby enabling use of the FCLT (e.g., Billingsley, 1968, 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 satisfies Ossiander's L^2 entropy condition.

(b) The 2FQR limit theory is useful for developing testing methodologies and is obtained in a different way from that of quasi-2FQR, 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 $\widetilde{B}_\tau^0$ is obtained from the covariance kernel of $\widetilde{\mathcal{G}}_\tau(\cdot)$, implying that $\widetilde{B}_\tau^0$ can be consistently estimated by first estimating the kernel function $\widetilde{\kappa}_\tau(\cdot,\cdot)$. This approach is discussed later in Section 3.2. In addition, we note that the information matrix equality does not hold even when the model is correctly specified, so that specifying the model correctly does not necessarily lead to efficient parameter estimation. The information matrix equality does hold under some strict conditions. The efficiency of the 2FQR estimator depends on the covariance kernel. For example, if $\widetilde{\kappa}_\tau(\gamma,\bar{\gamma}) \propto f_\gamma^{1/2}(\rho_\tau(\gamma,\theta_\tau^0))f_\gamma^{1/2}(\rho_\tau(\bar{\gamma},\theta_\tau^0))\delta(\gamma-\bar{\gamma})$ and $d\mathbb{Q}(\gamma) \propto d\gamma$, then $\widetilde{B}_\tau^0 \propto A_\tau^0$, so that the information matrix equality follows, where $\delta(\cdot)$ is the Dirac delta function.

(c) Theorem 1 also implies that the quasi-FQR estimator is asymptotically normal. Given the absence of the nuisance effects, we can let $S_i \equiv 0$, so that if $\mathcal{M}_\tau$ is misspecified, $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^*) \overset{\text{A}}{\sim} \mathcal{N}(0, C_\tau^*)$, where $C_\tau^* := A_\tau^{*-1}B_\tau^*A_\tau^{*-1}$ and $B_\tau^* := \mathbb{E}[J_{\tau i}J_{\tau i}']$.

(d) From Theorem 1, we can derive the limit distribution of the FQR estimator. Similarly to the quasi-FQR estimator, we can let $S_i \equiv 0$. Then $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^0) \overset{\text{A}}{\sim} \mathcal{N}(0, C_\tau^0)$, where $C_\tau^0 := A_\tau^{0-1}B_\tau^0A_\tau^{0-1}$ and $B_\tau^0 := \int_\gamma\int_{\bar{\gamma}} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\kappa_\tau(\gamma,\bar{\gamma})\nabla_{\theta_\tau}'\rho_\tau(\bar{\gamma},\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\bar{\gamma})$.

(e) When $x_\tau(\cdot)$ is nonlinear and the specified model is incorrect for $x_\tau(\cdot)$, we can expect an approximation error from the model estimation. We can illustrate this using Figure 1. When $[0, 40]$ is used as the domain of LIP, we observe Runge's phenomenon around the domain edges. But if $[10, 40]$ is used as the domain, Runge's phenomenon is negligible. These different results are obtained as the polynomial model is almost correct for the LIP on $[10, 40]$ but misspecified for the LIP on $[0, 40]$. According to Mincer and Jovanovic (1981), labor mobility happens mostly in the labor market during early career years of the worker, which makes Mincer's equation vulnerable to misspecification if the entire career years are considered. In view of this complication, Mincer and Jovanovic (1981) recommended removing the first 10 career years from the whole career cycles in order to achieve satisfactory specification of the Mincer equation. That is, the polynomial model is incorrect for the LIP on $[0, 40]$ but almost correct for the LIP over $[10, 40]$. In other words, approximation error can be substantially reduced by specifying an appropriate domain for modeling.

(f) Runge's phenomenon can arise even when a polynomial model is correctly specified for $x_\tau(\cdot)$. For such a case, it is appropriate to specify a model for $x_\tau(\cdot)$ by using basis functions or Bernstein polynomials. For example, if Bernstein's polynomial is employed for $x_\tau(\cdot)$, then for a series of coefficients $\{\beta_{i\geq 0}\}$ such that for $i \geq 0$, $|\beta_i| > |\beta_{i+1}|$, it follows that $\sum_{n=0}^p \beta_n b_n(\cdot)$ uniformly approximates $x_\tau(\cdot)$ as $p \to \infty$, where $b_n(\cdot)$ is the n-th order Bernstein basis polynomial. Hence, if $\mathcal{M}_\tau$ is specified as $\sum_{n=0}^p \beta_n H_n(\cdot)$ for some finite p, then $\mathcal{M}_\tau$ is misspecified, but Runge's phenomenon can be reduced by increasing p. Note that the results in Theorem 1 are not tied only to polynomial models. The same parametric structure can incorporate spline or orthogonal basis functions without affecting any asymptotic argument.

(g) In the Online Supplement, we provide the regularity conditions specifically designed for the 2FQR, quasi-FQR, and FQR estimators.

(h) A reasonably good and parsimonious approximation of $x_\tau(\cdot)$ can be achieved by applying a sequential specification testing procedure (e.g., Cho and Phillips, 2018). For example, if the empirical researcher advocates a polynomial model for $x_\tau(\cdot)$, we may first evaluate the model. Specifically, if a p-th degree polynomial model is assumed, for specification analysis purposes, we may use the following model formulation

$$\mathcal{M}_{\tau,p} := \left\{ \rho_\tau(\cdot;\alpha,\beta,\theta_\tau) : \rho_\tau(\gamma;\alpha,\beta,\theta_\tau) = \theta_{\tau 0} + \sum_{j=1}^{p} \theta_{\tau j}\gamma^j + \alpha_\tau \gamma^{\beta_\tau} \right\},$$

where $\theta_\tau = (\theta_{\tau 0}, \ldots, \theta_{\tau p})'$. This is a p-th degree polynomial model augmented by a power transformation. After estimating the unknown parameters in $\mathcal{M}_{\tau,p}$, we can test the hypothesis $H_0 : \alpha_{\tau *} = 0$ or $\beta_{\tau *} = 0$ to evaluate the p-th degree polynomial model. If H_0 is rejected, we increase the degree to $p+1$ to specify $\mathcal{M}_{\tau,p+1}$ and test the same hypothesis. In this manner, we can iterate the testing process until H_0 cannot be rejected. If $\mathcal{M}_{\tau,q}$ with $q > p$ is not rejected, we can choose the q-th degree polynomial function as a good approximation for $x_\tau(\cdot)$. Here, the power transformation plays the role of the activation function in the artificial neural network literature, corresponding to the exponential function in the consistent conditional moment testing literature. From this, we can expect to detect model misspecification consistently with omnibus power by testing H_0. Indeed, many specification tests are available in the literature. For example, Escanciano and Goh (2014) and Dong et al. (2019), among others, provide testing methodologies to detect misspecification in linear quantile models. By applying the above test procedure in the functional data context, we can expect to approximate $x_\tau(\cdot)$ within the polynomial model framework without employing too many parameters. $\square$

The limit distribution of the quantile function estimator is delivered by Theorem 1 and given in the following corollary.

Corollary 1. *Given Assumptions 2, 3, 4, 5, and 6, if $\mathcal{M}_\tau$ is misspecified, $\sqrt{n}\{\rho_\tau(\cdot, \widetilde{\theta}_{\tau n}) - \rho_\tau(\cdot, \theta_\tau^*)\} \Rightarrow \mathcal{Z}(\cdot)$, where $\mathcal{Z}(\cdot)$ is a Gaussian process defined on Γ such that for each γ and $\bar{\gamma} \in \Gamma$, $\mathcal{Z}(\gamma) \overset{\text{A}}{\sim} \mathcal{N}(0, \nabla'_{\theta_\tau}\rho(\gamma, \theta_\tau^*) \widetilde{C}_\tau^* \nabla_{\theta_\tau}\rho(\gamma, \theta_\tau^*))$ and $\mathbb{E}[\mathcal{Z}(\gamma)\mathcal{Z}(\bar{\gamma})] = \nabla'_{\theta_\tau}\rho(\gamma, \theta_\tau^*)\widetilde{C}_\tau^*\nabla_{\theta_\tau}\rho(\bar{\gamma}, \theta_\tau^*)$.* $\square$

Corollary 1 is trivially obtained from Theorem 1 to $\rho_\tau(\cdot, \widetilde{\theta}_{\tau n})$ using the Delta method.

3.2 Asymptotic Variance Matrix Estimation

Just as in standard quantile regression the limit distribution of the quasi-2FQR estimator and particularly its asymptotic variance matrix plays 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 matrix $\widetilde{B}_\tau^*$, which is now discussed. The variance matrix estimate 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 matrix in the general case, allowing for model misspecification, is conveniently done by employing a plug-in approach. In this case, first let

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

and estimate $\widetilde{B}^*_\tau$ by

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

where

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

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 7. *(i) For a sequence of measurable random variables $\{\widehat{H}_n \in \mathbb{R}^{s\times s}\}$, $\widehat{H}_n \overset{\mathbb{P}}{\to} H^*$; (ii) for a sequence of measurable random variables $\{\widehat{P}_n \in \mathbb{R}^{s\times s}\}$ that is uniformly positive definite with respect to n, $\widehat{P}_n \overset{\mathbb{P}}{\to} P^*$; and (iii) for a sequence of measurable functions $\{\widehat{f}_{\gamma n}(\cdot) : \mathbb{R} \mapsto \mathbb{R}\}$, $\widehat{f}_{\gamma n}(\cdot) \overset{\mathbb{P}}{\to} f_\gamma(\cdot)$ 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 $n^{-1}\sum_{i=1}^n \widehat{J}_{\tau n i}\widehat{J}'_{\tau n i}$ is consistent for B^*_τ, it follows that $\widetilde{B}_{\tau n}$ is consistent for $\widetilde{B}^*_\tau$, as in the following theorem below.

Theorem 2. *Given Assumption 2, 3, 4, 5, 6, and 7, if $\mathcal{M}_\tau$ is misspecified, $\widetilde{B}_{\tau n} \overset{\mathbb{P}}{\to} \widetilde{B}^*_\tau$.* □

Remarks 2. (a) If $\mathcal{M}_\tau$ is correctly specified, then the asymptotic covariance matrix of the 2FQR estimator $\widetilde{B}^0_\tau$ can be consistently estimated from explicit form of the covariance kernel in the functional law, which can then be used to estimate $\widetilde{B}^0_\tau$. With this approach we first estimate $\widetilde{\kappa}_\tau(\gamma, \bar{\gamma})$ by

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

Then, for each $\gamma \in \Gamma$, we let $\zeta_\tau(\gamma) := \mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \le \rho_\tau(\gamma, \theta^0_\tau)\} - \tau)S'_i]$, 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 under Assumptions 2, 3, 4, and 5, it follows that $\widetilde{\kappa}_{\tau n}(\cdot, \cdot)$ and $\widetilde{\zeta}_{\tau n}(\cdot)$ turn out to be consistent for $\kappa_\tau(\cdot, \cdot)$ and

$\zeta_\tau(\cdot)$, respectively. Therefore, if we let

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

be variance matrix estimator for $\widetilde{B}^0_\tau$, this matrix is consistent for $\widetilde{B}^0_\tau$. Note that $\widetilde{B}^\sharp_{\tau n}$ is simply sample analog of $\widetilde{B}^0_\tau$, so that consistency of the estimate follows directly from the consistency of $\widetilde{\kappa}_{\tau n}(\cdot,\cdot)$ and $\widetilde{\zeta}_{\tau n}(\cdot)$. Here, 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.

(b) If $\mathcal{M}_\tau$ is misspecified and the nuisance effect is absent, we can consistently estimate the asymptotic covariance matrix of the quasi-FQR estimator B^*_τ by $\widehat{B}_{\tau n} := n^{-1}\sum_{i=1}^n J_{\tau n i}J'_{\tau n i}$, where

$$J_{\tau n i} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \widehat{\theta}_{\tau n})\left(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma, \widehat{\theta}_{\tau n})\} - \tau\right)d\mathbb{Q}(\gamma).$$

(c) If $\mathcal{M}_\tau$ is correctly specified and the nuisance effect is absent, we can consistently estimate the asymptotic covariance matrix of the FQR estimator B^*_τ by estimating the kernel function first as for the 2FQR estimator. That is, if we first estimate $\kappa_\tau(\gamma, \bar{\gamma})$ by

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

then, it follows that $\widehat{\kappa}_{\tau n}(\cdot,\cdot)$ is consistent for $\kappa_\tau(\cdot,\cdot)$ uniformly on $\Gamma \times \Gamma$ by the ULLN. Therefore, if we let

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

this matrix is consistent for B^0_τ.

(d) The Online Supplement provides regularity conditions for estimation of the asymptotic covariance matrices of the 2FQR, quasi-FQR, and FQR estimators.

(e) $\widetilde{B}_{\tau n}$ (resp. $\widehat{B}_{\tau n}$) is numerically identical to $\widetilde{B}^\sharp_{\tau n}$ (resp. $\widehat{B}^\sharp_{\tau n}$), but the definition of $\widetilde{B}^\sharp_{\tau n}$ (resp. $\widehat{B}^\sharp_{\tau n}$) is conceptually different from that of $\widetilde{B}_{\tau n}$ (resp. $\widehat{B}_{\tau n}$), just as in the conceptual difference between $\widetilde{B}^*_\tau$ and $\widetilde{B}^0_\tau$ (resp. B^*_τ and B^0_τ). □

4 Quasi-2MFQR Estimation

The framework is now extended to allow for multiple quantile curve estimation from the same data $\{\widehat{G}_i(\cdot)\}_{i=1}^n$. This extension allows for joint estimation and inference concerning quantile functions for multiple quantiles

$\tau := (\tau_1, \tau_2, \ldots, \tau_p)'$.

The multiple quantile functions are estimated using the quasi-2FQR 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 quasi-2FQR parametric estimators. We call $\widetilde{\theta}_n$ the *quasi-two-stage multiple functional quantile regression (quasi-2MFQR)* estimator. To develop quasi-2MFQR 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 8. *(i) For each $\tau_j \in \{\tau_1, \ldots, \tau_p\}$ and $\theta_{\tau_j} \in \Theta_{\tau_j}$, $\rho_{\tau_j}(\,\cdot\,, \theta_{\tau_j})$ is $\mathcal{G}$-measurable, 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}; and (v) $f_{(\cdot)}(\rho_{\tau_j}(\cdot,\cdot)) \in \mathcal{C}(\Gamma \times \Theta_{\tau_j})$.*

Assumption 9. *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 $j, t = 1, \ldots, c_{\tau_\ell}$, $\sup_{(\gamma,\theta_{\tau_\ell})} |(\partial^2/\partial\theta_{\tau_\ell j}\partial\theta_{\tau_\ell t})\rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$; and (vi) for each $j = 1, 2, \ldots, s$, $\mathbb{E}[(\partial/\partial\pi_j)\widetilde{G}_i(\cdot, \pi^*)] \in \mathcal{L}_{ip}(\Gamma)$ and for each $\tau_\ell \in \{\tau_1, \ldots, \tau_p\}$ and for each $\theta_{\tau_\ell} \in \Theta_{\tau_\ell}$, $f_{(\cdot)}(\rho_{\tau_j}(\cdot, \theta_{\tau_\ell})) \in \mathcal{L}_{ip}(\Gamma)$.* □

Assumption 8 extends Assumption 4 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 $\mathcal{M}_{\tau_j}$ may differ, thereby allowing for different parametric model specifications at different quantile levels. Assumption 9 extends Assumption 5, ensuring that the regular bound conditions in Assumption 5 apply to the case of multiple quantile function estimation.

4.1 Large Sample Distribution of the Quasi-2MFQR Estimation

Given a finite quantile number p, consistency of the full quasi-2MFQR estimator $\widetilde{\theta}_n$ follows directly from the consistency of the individual quasi-2FQR 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 (3) as follows

$$\sqrt{n}(\widetilde{\theta}_n - \theta^*) = -A^{*-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{J}_i + o_{\mathbb{P}}(1), \tag{4}$$

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 quasi-2MFQR estimator is obtained by examining the limit behaviors of the two factors in the leading term on the right side of (4). First, like the individual matrices $A^*_{\tau_j}$, A^* does not contain any stochastic components. In addition, if for each $j = 1, 2, \ldots, p$, $A^*_{\tau_j}$ is positive definite, then so is A^* by construction. Thus, the limit distribution theory is effectively determined by the second factor in (4). Theorem 1 establishes that for each $j = 1, 2, \ldots, p$, $\sqrt{n}(\widetilde{\theta}_{\tau_j n} - \theta^*_{\tau_j}) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^*_{\tau_j})$, 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 t-th block column matrix given by $\widetilde{B}^*_{\tau_j \tau_t}$, which has the form

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

for each j and $t = 1, 2, \ldots, p$, where $B^*_{\tau_j\tau_t} := \mathbb{E}[J_{\tau_j i} J'_{\tau_t i}]$, and as before, for each $j = 1, 2, \ldots, p$, $J_{\tau_j i} := \int_\gamma \nabla_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^*_{\tau_j})(\mathbb{1}\{G_i(\gamma) \leq \rho_{\tau_j}(\gamma, \theta^*_{\tau_j})\} - \tau_j) d\mathbb{Q}(\gamma)$. The limit distribution theory for the full quasi-2MFQR estimator $\widetilde{\theta}_n$ is then obtained based on the following regularity conditions:

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

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

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

It is now straightforward to derive the limit distribution of the quasi-2MFQR estimator using (4).

Theorem 3. *Given Assumptions 2, 3, 8, 9, and 10, if $\mathcal{M}$ is misspecified,* $\sqrt{n}(\widetilde{\theta}_n - \theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}^*)$. □

Remarks 3. (a) If $\mathcal{M}$ is correctly specified, then we call $\widetilde{\theta}_n$ the *two-stage multiple quantile regression* (*2MFQR*) estimator. Theorem 3 can be strengthened by developing the asymptotics using functional limit law arguments. For this purpose we let

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

where $\rho_\tau(\gamma, \theta) := [\rho_{\tau_1}(\gamma, \theta_{\tau_1}), \rho_{\tau_2}(\gamma, \theta_{\tau_2}), \ldots, \rho_{\tau_p}(\gamma, \theta_{\tau_p})]'$. Then, it follows that $n^{-1/2}\sum_{i=1}^n (\mathbb{1}\{\widehat{G}_i(\cdot) \leq \rho_\tau(\cdot, \theta^0)\} - \tau) \Rightarrow \widetilde{\mathcal{G}}(\cdot) := [\widetilde{\mathcal{G}}_{\tau_1}(\cdot), \widetilde{\mathcal{G}}_{\tau_2}(\cdot), \cdots, \widetilde{\mathcal{G}}_{\tau_p}(\cdot)]'$ under mild regularity conditions, where $\widetilde{\mathcal{G}}(\cdot)$ is a mean-zero Gaussian process such that for j and $t = 1, 2, \ldots, p$, and γ and $\bar{\gamma} \in \Gamma$, the covariance kernel is

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

and $\kappa_{\tau_j\tau_t}(\gamma,\bar{\gamma}) := \mathbb{E}[\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le \rho_{\tau_j}(\gamma,\theta^0_{\tau_j})\}\mathbb{1}\{\widetilde{G}_i(\bar{\gamma},\pi^*) \le \rho_{\tau_t}(\bar{\gamma},\theta^0_{\tau_t})\}] - \tau_j\tau_t$. The Cramér-Wold device in Wooldridge and White (1988, proposition 4.1) leads directly to the multivariate functional limit law, and from this, it follows that

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

where $\widetilde{C}^0 := A^{0\,-1}\widetilde{B}^0 A^{0\,-1}$, $A^0 := \mathrm{diag}[A^0_{\tau_1}, A^0_{\tau_2},\ldots,A^0_{\tau_p}]$, and $\widetilde{B}^0$ is a $c\times c$ matrix with j-th block row and t-th block column matrix

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

In the diagonal block with $j=t$ the matrices $B^0_{\tau_j\tau_t}$ and $\widetilde{B}^0_{\tau_j\tau_t}$ are identical to $B^0_{\tau_j}$ and $\widetilde{B}^0_{\tau_j}$ as defined earlier in Section 3.2.

For this case, $\theta^* = \theta^0$, we have $\widetilde{B}^* = \widetilde{B}^0$ and $A^* = A^0$, so that $\widetilde{C}^* = \widetilde{C}^0$.

(b) If $\mathcal{M}$ is misspecified and the functional data do not involve nuisance effects, we can estimate the unknown parameter θ^* by $\widehat{\theta}_n := \arg\min_{\theta\in\otimes^p_{j=1}\Theta_{\tau_j}} q_n(\theta)$, which we call the *quasi-multiple functional quantile regression (quasi-MFQR)* estimator, where $q_n(\theta) := \sum^p_{j=1} w_j q_{\tau_j n}(\theta_{\tau_j})$ for the same weights $\{w_j\}$. By applying the approach to derive the limit distribution of $\widetilde{\theta}_n$ in Theorem S.5 of the Online Supplement and by letting $S_i \equiv 0$, 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 t-th block column is $B^*_{\tau_j\tau_t}$.

(c) If $\mathcal{M}$ is correctly specified and the nuisance effect is absent, the limit distribution of $\widehat{\theta}_n$ can be deduced from Theorem S.5 of the Online Supplement. By setting $S_i \equiv 0$ in the definition of $\widetilde{\kappa}_{\tau_j\tau_t}(\gamma,\bar{\gamma})$, which leads to $\widetilde{\kappa}_{\tau_j\tau_t}(\cdot,\cdot) = \kappa_{\tau_j\tau_t}(\cdot,\cdot)$ 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 $C^0 := A^{0\,-1}B^0A^{0\,-1}$, such that B^0 is defined a $c\times c$ matrix with j-th block row and t-th block column matrix $B^0_{\tau_j\tau_t}$.

(d) In the Online Supplement, we provide the regularity conditions specifically designed for the 2MFQR, quasi-MFQR, and MFQR estimators.

(e) $\widetilde{\theta}_n$ 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 $\sum^p_{j=1} w_j \equiv 1$, $\widetilde{\theta}_n := \arg\min_{\theta\in\otimes^p_{j=1}\Theta_{\tau_j}} \widehat{q}_n(\theta)$, where $\widehat{q}_n(\theta) := \sum^p_{j=1} w_j\widehat{q}_{\tau_j n}(\theta_{\tau_j})$. If for each j, w_j is equal to $1/p$, $\widetilde{\theta}_n$ corresponds to the composite quantile regression estimator (see Zou and Yuan, 2008, for example). 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 $\widetilde{\theta}_n$. This fact also implies that the asymptotic distribution of $\widetilde{\theta}_n$ is independent of the weight. One of the conditions leading to this result is that different parameters θ_τ's are assumed for the quantile function for different percentiles τ's. If the model for the quantile function is specified to share a common parameter, so that $x_\tau(\gamma)$ is specified by $\rho(\tau,\gamma,\theta)$ say, a different approach is required from what we investigate here. First, the empirical researcher needs to carefully specify how τ affects the quantile function through the coefficients. Second, a proper objective function needs to be employed to estimate the unknown parameter. The loss function $\widehat{q}_n(\theta)$ is defined as the weighted average of the check functions, and each check function assumes a single percentile τ_j. If the same loss function is defined by the quantile functions with the common parameter, the parameter estimator is not independent of the weights. This means that the loss function needs to be properly defined to estimate the unknown parameter appropriately. We leave this investigation as a future research topic. □

The following corollary states the limit distribution of the multivariate quantile function estimator.

Corollary 2. *Given Assumptions 2, 3, 8, 9, and 10, if $\mathcal{M}$ is misspecified, $\sqrt{n}(\rho_\tau(\cdot, \widetilde{\theta}_n) - \rho_\tau(\cdot, \theta^*)) \Rightarrow \mathcal{Z}(\cdot)$, where $\mathcal{Z}(\cdot)$ is a multivariate Gaussian process defined on Γ such that for each γ and $\bar{\gamma} \in \Gamma$, $\mathcal{Z}(\gamma) \overset{\text{A}}{\sim} \mathcal{N}(0, \nabla'_\theta\rho(\gamma, \theta)\widetilde{C}^*\nabla_\theta\rho(\gamma, \theta^*))$ and $\mathbb{E}[\mathcal{Z}(\gamma)\mathcal{Z}(\bar{\gamma})] = \nabla'_\theta\rho(\gamma, \theta)\widetilde{C}^*\nabla_\theta\rho(\bar{\gamma}, \theta^*)$.* □

Corollary 2 corresponds to Corollary 1, and the proof is omitted in the Online Supplement as it trivially follows from Theorem 3.

4.2 Asymptotic Variance Matrix Estimation

The limit theory of Sections 4.1 enables hypothesis testing on the unknown model parameters once the relevant asymptotic variance matrices are estimated. The approach follows Section 3.2 closely and is only briefly detailed here.

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

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

where

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

and $\widehat{f}_{\gamma n}(\rho_\tau(\gamma, \widetilde{\theta}_n)) := \text{diag}[\widehat{f}_{\gamma n}(\rho_{\tau_1}(\gamma, \widetilde{\theta}_{\tau_1 n})) \cdot I_{c_{\tau_1}}, \ldots, \widehat{f}_{\gamma n}(\rho_{\tau_p}(\gamma, \widetilde{\theta}_{\tau_p n})) \cdot I_{c_{\tau_p}}]$. The following corollary immediately follows.

Corollary 3. *Given Assumption 2, 3, 5, 7, 8, 9, and 10, if $\mathcal{M}$ is misspecified, $\widetilde{B}_n \overset{\mathbb{P}}{\to} \widetilde{B}^*$.* □

Remarks 4. (a) If $\mathcal{M}$ is correctly specified, the variance matrix $\widetilde{B}^0$ can be estimated by first estimating the unknown covariance kernel function $\widetilde{\kappa}(\cdot, \cdot)$. Let

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

and $\widetilde{\zeta}_n(\gamma) := n^{-1}\sum_{i=1}^{n}(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma, \widetilde{\theta}_n)\} - \tau)S'_i$. Note that $\widetilde{\kappa}_n(\cdot, \cdot)$ is $p \times p$ matrix of functions, and it is 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 t-th column blocks being $\kappa_{\tau_j\tau_t}(\cdot,\cdot)$. Likewise, $\widetilde{\zeta}_n(\cdot)$ turns out to be consistent for $\zeta(\cdot) := \mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\cdot,\pi^*) \leq \rho_\tau(\cdot,\theta^0)\} - \tau)S_i']$ uniformly on Γ by applying the ULLN. Using $\widehat{\kappa}_n(\cdot,\cdot)$ and $\widetilde{\kappa}_n(\cdot,\cdot)$, we estimate $\widetilde{B}^0$ by plug-in, giving

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

$\widetilde{B}_n^\sharp$ is a $c\times c$ matrix with (j,t) block submatrices that estimate $\widetilde{B}^0_{\tau_j\tau_t}$, for j and $t=1,2,\ldots,p$. It immediately follows that $\widetilde{B}_n^\sharp \overset{\mathbb{P}}{\to} \widetilde{B}^0$.

(b) If $\mathcal{M}$ is misspecified and the nuisance effect is absent, we let $J_{ni} := [J_{\tau_1ni}', J_{\tau_2ni}',\ldots,J_{\tau_pni}']'$ and $\widehat{J}_{ni} := [\widehat{J}_{\tau_1ni}', \widehat{J}_{\tau_2ni}',\ldots,\widehat{J}_{\tau_pni}']'$ to define the estimates

$$\widehat{B}_n := \frac{1}{n}\sum_{i=1}^n J_{ni}J_{ni}'.$$

It immediately follows that $\widehat{B}_n \overset{\mathbb{P}}{\to} B^*$ by applying Theorem 2.

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

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

to estimate B^0 by plug-in, giving

$$\widehat{B}_n^\sharp := \int_\gamma\int_{\bar{\gamma}} \nabla_\theta\rho_\tau(\gamma,\widehat{\theta}_n)\widehat{\kappa}_n(\gamma,\bar{\gamma})\nabla_\theta'\rho_\tau(\bar{\gamma},\widehat{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\bar{\gamma}).$$

$\widehat{B}_n^\sharp$ is $c\times c$ matrix with (j,t) block submatrices that estimate $B^0_{\tau_j\tau_t}$ for j and $t=1,2,\ldots,p$. It immediately follows that $\widehat{B}_n^\sharp \overset{\mathbb{P}}{\to} B^0$ by applying Theorem 2.

(e) In the Online Supplement, we formally state the limit consequences of the asymptotic variance estimates without proofs, as they are established as corollaries of the univariate quantile regression estimators. □

5 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

$$\mathbb{H}_o : R(\theta^*) = 0 \qquad \text{versus} \qquad \mathbb{H}_a : R(\theta^*) \neq 0, \tag{5}$$

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 $r = \text{rank}[\nabla'_{\theta} R(\theta^*)]$. For example, it may be of interest to test the hypothesis $x_\tau(\cdot) \equiv c$ for some $c \in \mathbb{R}$, which can be be reformulated under the above framework. Thus, if the model is specified as $x_\tau(\gamma) = \theta_{\tau 1} + \theta_{\tau 2}\gamma$, testing the invariant quantile function hypothesis is equivalent to testing whether $\theta^*_{\tau 2} = 0$ or not. Then, depending on the model assumptions, the invariant quantile function hypothesis can be easily rewritten in terms of the unknown parameter.

Inference concerning hypotheses such as (5) on the unknown parameters of multiple quantile functions can be made in parallel to the standard testing procedures based on maximum likelihood (ML) estimation, viz., Wald, LM, and LR test principles. Although Koenker and Bassett (1982) applied the LR test principle to the median estimator by forming the likelihood function based on the Laplace distribution, past research has only occasionally used LM and LR test principles in the quantile regression context. Furthermore, to our knowledge there is no prior application in the functional data context. It is therefore useful to formally define these test principles and demonstrate their asymptotic behaviors in the present framework. Another goal of applying the three test principles is to provide a straightforward testing methodology for empirical functional data. As it turns out, the asymptotic behavior of our tests parallel those of the standard Wald, LM, and QLR tests applied to random variables. The test statistics asymptotically follow chi-squared distributions under the null and diverge at the rate of n under the alternative, which is the same as in standard tests. This parallel with standard inferential methodology helps to enhance its practical appeal for empirical analysis.

Since the objective functions $q_n(\cdot)$ and $\widehat{q}_n(\cdot)$ are formed by integral transformations, the three tests can be constructed analogous to those under mean-regression. First, define the following constrained estimators of the parameters

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

We call $\ddot{\theta}_n$ the *constrained multiple two-stage quasi-functional quantile regression (constrained quasi-2MFQR)* estimator. This estimator involves constrained Lagrangian estimation, analogous to constrained ML estimation. Although the function $\widehat{q}_n(\cdot)$ is not continuously differentiable, it is 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 distribution of $\ddot{\theta}_n$.

Assumption 11. *(i)* $\lambda_{\min}(A^*) > 0$*; and (ii)* $\lambda_{\min}(B^*) > 0$. □

Assumption 12. *(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 $\rho_\tau(\gamma, \theta_\tau^0) = \theta_{1\tau}^0 + \theta_{2\tau}^0\gamma$ 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. In finite sample sizes the probability of crossing is always greater than zero, implying that it is extremely difficult to specify a correct model for empirical data with a finite sample size. In prior literature, specification tests are developed for the conventional quantile regression model (see Bierens and Ginther, 2001; Escanciano and Velasco, 2010; Escanciano and Goh, 2014; Dong et al., 2019, among others), and they are useful in specifying quantile functions without crossings. But no such specification test is available in the prior literature for function space dependent variable cases and this topic is left for future research.

The limit distribution of the constrained quasi-2MFQR estimator is given in the following lemma.

Lemma 2. *Given Assumptions 2, 3, 8, 9, 10, and 12,* $\sqrt{n}(\ddot{\theta}_n - \theta^*) + \sqrt{n}(\Omega A^*)^{-1}D^{*\prime}[D^*(\Omega A^*)^{-1}D^{*\prime}]^{-1}$ $R(\theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \{I + (\Omega A^*)^{-1}D^{*\prime}[D^*(\Omega A^*)^{-1}D^{*\prime}]^{-1}D^*\}\widetilde{C}^*\{I + D^{*\prime}[D^*(\Omega A^*)^{-1}D^{*\prime}]^{-1}D^*(\Omega A^*)^{-1}\})$, *where* $\Omega := \mathrm{diag}[\omega_1 I_{c_1}, \omega_2 I_{c_2}, \ldots, \omega_p I_{c_p}]$ *such that for each* $j = 1, 2, \ldots, p$, $\omega_j > 0$, *and* $D^* := \nabla_\theta' R(\theta^*)$. □

Under $\mathbb{H}_o$, $\sqrt{n}(\ddot{\theta}_n - \theta^*)$ is asymptotically distributed and centred at zero, whereas it is not bounded under $\mathbb{H}_a$. It is therefore useful in forming the tests discussed below. Note that the constrained quasi-2MFQR limit distribution is 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 just as CPS (2022) define them in the mean estimation context using functional data. The Wald test uses the unconstrained quasi-2MFQR estimator giving

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

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

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

Under Assumption 7 and given consistency of $\widetilde{\theta}_n$ for θ^*, $\widetilde{A}_n$ is consistent for A^*. The Wald test assesses the magnitude of $R(\widetilde{\theta}_n)$ in suitable metrics and unless $R(\theta^*) = 0$, the test is not bounded in probability.

The LM test is constructed as

$$\mathcal{L}\ddot{\mathcal{M}}_n := n\ddot{Q}_n'\ddot{A}_n^{-1}\ddot{D}_n'\{\ddot{D}_n\ddot{C}_n\ddot{D}_n'\}^{-1}\ddot{D}_n\ddot{A}_n^{-1}\ddot{Q}_n,$$

where we let $\ddot{Q}_n := n^{-1}\sum_{i=1}^n \ddot{J}_{ni}$ with $\ddot{J}_{ni} := [\ddot{J}_{\tau_1 ni}', \ddot{J}_{\tau_2 ni}', \ldots, \ddot{J}_{\tau_p ni}']'$, $\ddot{D}_n := \nabla_\theta' R(\ddot{\theta}_n)$, $\ddot{C}_n := \ddot{A}_n^{-1}\ddot{B}_n\ddot{A}_n^{-1}$, and $\ddot{B}_n := n^{-1}\sum_{i=1}^n \ddot{J}_{ni}\ddot{J}_{ni}' - \ddot{V}_n\widehat{P}_n^{-1}\ddot{K}_n' - \ddot{K}_n\widehat{P}_n^{-1}\ddot{V}_n' + \ddot{K}_n\widehat{P}_n^{-1}\widehat{H}_n\widehat{P}_n^{-1}\ddot{K}_n'$ with $\ddot{V}_n := n^{-1}\sum_{i=1}^n \ddot{J}_{ni}S_i'$ and $\ddot{K}_n := [\ddot{K}_{\tau_1 n}', \ddot{K}_{\tau_2 n}', \ldots, \ddot{K}_{\tau_p n}']'$ such that for each $j = 1, 2, \ldots, p$,

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

$\ddot{K}_{\tau n} := n^{-1}\sum_{i=1}^{n}\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\ddot{\theta}_{\tau n})\widehat{f}_{\gamma n}(\rho_\tau(\gamma,\ddot{\theta}_{\tau n}))\nabla'_\pi \widetilde{G}_i(\gamma,\widehat{\pi}_n)d\mathbb{Q}(\gamma)$. Further, define $\ddot{A}_n := \mathrm{diag}[\ddot{A}_{\tau_1 n}, \ddot{A}_{\tau_2 n}, \ldots, \ddot{A}_{\tau_p n}]$, where for each $j = 1, 2, \ldots, p$,

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

Given Assumption 7 and the consistency of $\ddot{\theta}_n$ for θ^* under the null, it follows that $\ddot{A}_n$ is consistent for A^* under the null. Note that $\ddot{J}_{ni}$ is defined in parallel to $\widetilde{J}_{ni}$. The only difference is that $\widetilde{\theta}_n$ in $\widetilde{J}_{ni}$ is replaced by $\ddot{\theta}_n$. Although Lemma 2 implies that the limit distribution of $\sqrt{n}(\ddot{\theta}_n - \theta^*)$ is influenced by the selection of Ω, the LM test statistic is defined without direct association with Ω. If $\mathbb{H}_o$ holds, quasi-2MFQR estimator converges to θ^*, so that $\ddot{Q}_n$ converges to zero. Otherwise, $\ddot{Q}_n$ does not converge 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 $\mathbb{H}_o$, $\ddot{\theta}_n$ and $\widetilde{\theta}_n$ both converge to θ^*, so that the distance between $\widehat{q}_n(\ddot{\theta}_n)$ and $\widehat{q}_n(\widetilde{\theta}_n)$ converges to zero. But $\ddot{\theta}_n$ does not converge to θ^* under $\mathbb{H}_a$, so that 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

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

The QLR statistic is nonnegative because $\ddot{\theta}_n$ minimizes the objective function subject to the restrictions $R(\theta) = 0$, whereas $\widetilde{\theta}_n$ minimizes the same objective function 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 4. *For any sequence c_n such that $c_n = o(n)$, if Assumptions 2, 3, 7, 8, 9, 10, and 12 hold, (i) $\ddot{\mathcal{W}}_n \overset{\mathrm{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty}\mathbb{P}(\ddot{\mathcal{W}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; (ii) $\mathcal{L}\ddot{\mathcal{M}}_n \overset{\mathrm{A}}{\sim} \mathcal{X}_r^2$ 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) $\mathcal{Q}\ddot{\mathcal{L}}\mathcal{R}_n \overset{\mathrm{A}}{\sim} \widetilde{W}'(D^*(\Omega A^*)^{-1}D^{*\prime})^{-1}\widetilde{W}$ 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 $\widetilde{W} \sim \mathcal{N}(0, D^*\widetilde{C}^*D^{*\prime})$.* □

Remarks 5. (a) According to Theorem 4, 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 statistics 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 $\mathbb{H}_a$.

(b) Although we do not pursue the approach here as an alternative inferential method, it is straightforward to perform resampling on the functional observations to apply a bootstrap procedure and construct a confidence interval. This approach is a topic for future research.

(c) In the Online Supplement, we discuss the Wald, LM, and QLR tests defined by (constrained) quasi-MFQR estimator. □

6 Simulations

6.1 Simple Linear Specification

Simulations are 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 are generated according to the regularity conditions in Section 5.

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 $\tau \in (0, 1)$,

$$\mathcal{M}_\tau := \{\rho_\tau(\gamma, \theta_\tau) := \theta_\tau + \theta_\tau\gamma, \theta_\tau \in \Theta := [-1/2, 1/2]\}. \tag{6}$$

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

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

and the quasi-MFQR 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. The quasi-MFQR 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. Note that the grid distance is extremely narrow, allowing for the minimization of $q_n(\cdot)$ with a high degree of precision. Alternatively, if n is large enough, $q_n(\cdot)$ can be minimized through convex optimization with a moderate tolerance level because $q_n(\cdot)$ is stochastically differentiable at the limit (see Pollard, 1985).

Wald, LM, and QLR tests are constructed for the following hypotheses: $\mathbb{H}_o : [\theta^*_{\tau_1}, \theta^*_{\tau_2}]' = r$ and $\mathbb{H}_a : [\theta^*_{\tau_1}, \theta^*_{\tau_2}]' \ne r$, 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 $A^* := \text{diag}[A^*_{\tau_1}, A^*_{\tau_2}] = \text{diag}[7/8, 7/8]$

with

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

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

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 $f_\gamma(\rho_{\tau_j}(\gamma, \theta^*_{\tau_j})) = 1/(1+\gamma)$. For the simulation, we estimate the density function $f_\gamma(\rho_{\tau_j}(\gamma, \theta^*_{\tau_j}))$ for given τ_j by kernel density estimation. We use a Gaussian kernel with Silverman'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

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

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 4 (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 4 (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.

6.2 Nonlinear Specification

In this section, we extend the baseline setup from the previous session to a more flexible and realistic data-generating structure. This allows us to examine how the proposed procedures perform when the underlying functional observations exhibit richer curvature and variability that changes with γ.

Let $\{G_i : \Gamma \to \mathbb{R} : i = 1, 2, \ldots, n\}$ denote a collection of i.i.d. functional observations, where each individual contributes a continuous realization $G_i(\cdot)$ on $\gamma \in \Gamma = [0, 1]$. The observable functions are generated according to

$$G_i(\gamma) = \sigma_0 X_i + \sigma_1 X_i \gamma + \sigma_2 X_i \gamma^2 + s(\gamma)\, \varepsilon_i. \tag{7}$$

The individual functions $G_i(\cdot)$ are driven jointly by the random scale component X_i and the idiosyncratic shock process ε_i, with $X_i \sim_{\text{iid}} U[0, 1]$ and $\varepsilon_i \sim_{\text{iid}} N(0, 1)$ independently. The stochastic dispersion of $G_i(\gamma)$ is governed by the smooth scale function $s(\gamma) = b_0 + b_1\gamma + b_2\gamma^2$, where b_0, b_1, and b_2 are fixed constants. Note that this specification preserves the quadratic structure of the conditional quantile function in γ.

For each quantile $\tau \in (0, 1)$, we consider

$$\mathcal{M}_\tau := \Big\{\rho_\tau(\gamma, \theta_\tau) := \theta_{\tau,0} + \theta_{\tau,1}\gamma + \theta_{\tau,2}\gamma^2,\ \theta_\tau = (\theta_{\tau,0}, \theta_{\tau,1}, \theta_{\tau,2})' \in \Theta\Big\}.$$

Under the data-generating mechanism introduced above, $\mathcal{M}_\tau$ is correctly specified for the τ-th conditional quantile function of $G_i(\gamma)$. Consequently, the true coefficients are given by

$$\theta^0_{\tau,0} = \sigma_0\, \tau + b_0\, \Phi^{-1}(\tau), \qquad \theta^0_{\tau,1} = \sigma_1\, \tau + b_1\, \Phi^{-1}(\tau), \qquad \theta^0_{\tau,2} = \sigma_2\, \tau + b_2\, \Phi^{-1}(\tau),$$

where $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function.

Firstly, we consider a set of standard Monte Carlo diagnostics to assess the finite sample accuracy of the estimator. For each coefficient $j \in \{0, 1, 2\}$, the simulation analogues of Bias, RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error) are computed as

$$\text{Bias}_{\tau,j} = \frac{1}{R}\sum_{r=1}^{R}\big(\hat{\theta}^{(r)}_{\tau n,j} - \theta^0_{\tau,j}\big), \quad \text{RMSE}_{\tau,j} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\big(\hat{\theta}^{(r)}_{\tau n,j} - \theta^0_{\tau,j}\big)^2}, \quad \text{MAE}_{\tau,j} = \frac{1}{R}\sum_{r=1}^{R}\big|\hat{\theta}^{(r)}_{\tau n,j} - \theta^0_{\tau,j}\big|.$$

where we examine the three quantiles $\tau \in \{0.25, 0.50, 0.75\}$ using $R = 5000$ replications. These criteria summarize how closely the estimator recovers the underlying coefficients across repeated samples, capturing both systematic deviation and overall dispersion. For simulations, the scale function is parameterized by $(b_0, b_1, b_2) = (0.01, -0.1, 0.1)$, while $(\sigma_0, \sigma_1, \sigma_2)$ is set to $(0.1, 0.2, 0.3)$. The adjunct probability measure $\mathbb{Q}(\cdot)$ is taken to be the uniform distribution on Γ and other implementation details follow those in Section 6.1.

Further, we explore a misspecified setting by augmenting the quadratic structure with an additional nonlinear component. In particular, we consider the same data-generating mechanism as before, except that the functional observation includes an extra term $h(\gamma)$, so that

$$G_i(\gamma) = (\sigma_0 X_i) + (\sigma_1 X_i)\gamma + (\sigma_2 X_i)\gamma^2 + h(\gamma) + s(\gamma)\,\varepsilon_i.$$

The misspecification enters solely through the additive function

$$h(\gamma) = \delta \sum_{k=1}^{5} c_k \phi_k(\gamma).$$

The nonlinear deviation $h(\gamma)$ is constructed from five sieve basis elements: $\phi_1(\gamma) = \sin(2\pi\gamma)$, $\phi_2(\gamma) = \cos(2\pi\gamma)$, and three interior cubic B-spline functions $\phi_3(\gamma)$, $\phi_4(\gamma)$, $\phi_5(\gamma)$ generated from clamped knots at $\{0, 0, 0, 0, 0.25, 0.5, 0.75, 1, 1, 1, 1\}$. To prevent the spline components from being captured by the quadratic polynomial space, the B-spline functions are orthogonalized with respect to that space and then normalized. The Bias, RMSE, and MAE are then evaluated with respect to the pseudo–true values $\theta^*_{\tau,j}$ for $j \in \{0, 1, 2\}$. The sieve coefficients are set to $c_k = (0.5, 0.4, 0.3, 0.2, 0.1)$ with $\delta = 0.15$.

In addition, we examine the impact of nuisance effects that occur when latent functional curves are not directly observed but reconstructed from data observed only at a finite collection of points. We generate discrete measurements at $T = 250$ equally spaced points $\{\gamma_t\}_{t=1}^{T}$, according to

$$Y_{it} = G_i(\gamma_t) + \eta_{it}, \qquad \eta_{it} \sim_{\text{iid}} N(0, \delta_\eta^2).$$

where the construction of G_i follows equation (7). The process $\{Y_{it}\}$ can be viewed as a discretized and noise-contaminated version of the underlying functional observation $G_i(\cdot)$. The noise level δ_η is set sufficiently small so that it introduces only minimal distortion and is designed to have almost no effect on the true quantile structure. To obtain $\widehat{G}_i(\cdot)$, we follow Zhang and Chen (2007) who show that local polynomial kernel smoothing applied to discrete noisy measurements yields consistent functional estimators whose influence on subsequent inference becomes asymptotically negligible under mild regularity conditions.

Table 2 reports the resulting accuracy measures for the three simulation designs described above. Under correct specification, the estimator recovers the target coefficients with minimal finite sample deviation and exhibits the expected improvement as the sample size increases. Under misspecification, the estimator closely tracks the pseudo–true slope and curvature parameters, and the corresponding error measures decline with the sample size in accordance with the quasi-consistency theory in Section 3. In the nuisance-parameter design, the finite sample patterns remain broadly comparable to those under correct specification, with only minor distortions arising from the reconstruction step.

Beyond estimation accuracy we assess the finite sample performance of inference procedures. The experiments are based on the quadratic model in equation (7). For each coefficient $j \in \{0, 1, 2\}$, inference is conducted using a joint hypothesis across the three quantile levels, namely $\mathbb{H}_o : [\,\theta^*_{\tau_1,j},\, \theta^*_{\tau_2,j},\, \theta^*_{\tau_3,j}\,]' = r_j$ and $\mathbb{H}_a : [\,\theta^*_{\tau_1,j},\, \theta^*_{\tau_2,j},\, \theta^*_{\tau_3,j}\,]' \neq r_j$, where $r_j := (\sigma_j\tau_1 + b_j\Phi^{-1}(\tau_1),\, \sigma_j\tau_2 + b_j\Phi^{-1}(\tau_2),\, \sigma_j\tau_3 + b_j\Phi^{-1}(\tau_3))'$. We examine the Wald, LM, and QLR tests developed in Section 5 and evaluate how closely their empirical rejection probabilities track their corresponding asymptotic limits. The simulation structure parallels the inference design in Section 6.1 but is adapted to accommodate the quadratic specification employed in our main analysis.

The first panel of Table 3 reports the empirical size of the three tests. Since the remaining coefficient cases yield patterns that are nearly identical, we focus on reporting the inference results for $\theta_{\tau,2}$ only. Overall, the Wald, LM, and QLR tests exhibit rejection probabilities that lie close to their nominal levels, with accuracy improving steadily as the sample size n increases. The Wald test shows mild oversizing in smaller samples, but this discrepancy diminishes for $n \geq 170$. The LM test displays the most stable size performance across all nominal levels and sample sizes. The QLR test tends to be slightly liberal at the 1% level, although this deviation also lessens as n grows. Taken together, the results align well with our asymptotic theory and indicate that all three procedures achieve near-nominal size when the sample size reaches 250.

For the analysis of power, we retain the same functional structure for $G_i(\gamma)$ in (7) but modify the distribution of X_i to introduce controlled deviations from the null model. Specifically, we generate $X_i := \mathbf{1}\{W_i > \beta\}Z_i + \mathbf{1}\{W_i \leq \beta\}Z_i^{1/2}$, where $W_i \sim_{\text{iid}} U[0,1]$, $Z_i \sim_{\text{iid}} U[0,1]$, and the two sequences are mutually independent. The parameter β governs the strength of the departure from the null: when $\beta = 0$, the distribution of X_i coincides with that in the correctly specified model, whereas increasing β introduces progressively stronger deviations through a mixture of linear and nonlinear transformations of Z_i.

The empirical power can be found in Table 3 for the Wald, LM, and QLR tests under progressively stronger alternatives generated by increasing values of β. Even for a moderate level of deviation, the tests exhibit substantial power, and for larger departures the rejection probabilities approach one. For both sample sizes $n = 100$ and $n = 250$, the rejection rates rise monotonically with β, confirming that all three statistics respond appropriately to departures from the null model. When $n = 250$, the resulting power values are close to one across all nominal levels. Overall, the patterns align closely with the theoretical predictions in Section 5, which establish that the statistics achieve increasing power under a broad range of alternatives and attain near-perfect discrimination for sufficiently large n.

7 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 CPS (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 (quasi-)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 dataset was used in CPS (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. By this condition, individuals with zero income at some time points are removed from the analysis. 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.

7.1 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. As the right side is defined using the dummy variable d_i, the model delivered by $\rho_\tau(\gamma, \theta_\tau^M, \theta_\tau^F)$ may look different from that in (1). However, the model $\rho_\tau(\gamma, \theta_\tau^M, \theta_\tau^F)$ distinguishes the two separate models by using the dummy variable d_i. Hence, estimation of the unknown parameters θ_τ^F and θ_τ^M through the above model is equivalent to estimating the unknown parameters in the models for males and females separately. 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 4 and 5 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. Fourth, although Tables 4 and 5 do not contain the estimation results for the unknown parameters, all coefficients are not equal to zero. In fact, the hypothesis that all of $\theta_{\tau_2}^M$, $\theta_{\tau_3}^M$, $\theta_{\tau_4}^M$, and $\theta^M_{\tau_5}$ are equal to zero was rejected, meaning that the quantile function is not invariant with respect to γ. This result is also valid for the female data, and Figure S.2 in the Supplement shows the estimated paths of $x_\tau(\cdot)$ for $\tau = 0.25$, 0.50, and 0.75 that are obtained by specifying the quadratic, cubic, and quartic models. All estimated paths are apparently different from a constant function.

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 6 and 7 contain the test outcomes for $\tau = 0.25, 0.5$, and 0.75. Table 6 reports the test results using the samples of the original LIPs, while Table 7 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 7 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.

7.2 Inference on income paths for later work years

We repeat the exercises conducted in Section 7.1 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 8 and 9 report inferential findings on gender effects and are summarized as follows. Table 8 gives test results from the Wald, LM, and QLR statistics. Similar to Table 4, 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 7.1. 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 9, 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 $\tau = 0.75$. Third, as in Section 7.1, we could reject the hypothesis that all coefficients are equal to zero for both male and female data, meaning that the quantile function is not invariant with respect to γ even for later work year data. Figure S.3 in the Supplement shows the estimated paths of $x_\tau(\cdot)$ for $\tau = 0.25$, 0.50, and 0.75 that are obtained by specifying the quadratic, cubic, and quartic models. All estimated paths again differ from a constant function.

Finally, Tables 10 and 11 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 6 and 7. Table 10 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 8, 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 CPS (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.

8 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. 6

——— (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. 16, 3, 18

ANGRIST, J., V. CHERNOZHUKOV, AND I. FERNÁNDEZ-VAL (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563. 3, 8, 2

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. 34

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. 34

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. 34

BIERENS, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458. 10

BIERENS, H. J. AND D. K. GINTHER (2001): "Integrated Conditional Moment Testing of Quantile Regression Models," *Empirical Economics*, 26, 307–324. 27

BILLINGSLEY, P. (1968, 1999): *Convergence of Probability Measures*, New York: John Wiley & Sons. 16, 3, 4

BOSQ, D. (2000): *Linear Processes in Function Spaces: Theory and Applications*, New York: Springer. 1

CARDOT, H., C. CRAMBES, AND P. SARDA (2005): "Quantile Regression when the Covariates are Functions," *Nonparametric Statistics*, 17, 841–856. 1

CHANG, Y., B. HU, AND J. Y. PARK (2019): "Econometric Analysis of Functional Dynamics in the Presence of Persistence," Tech. rep., Indiana University. 1

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. 1

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Armsterdam: North Holland. Chapter 76, pp. 5549–5632, vol. 6. 5

CHILÈS, J.-P. AND P. DELFINER (1999): *Geostatistics: Modelling Spatial Uncertainty*, New York: John Wiley & Sons. 5, 10

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. 17, 34

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. 1, 2, 5, 7, 8, 10, 27, 33, 34, 37

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. 1

DAVIES, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247–254. 6, 10

——— (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternatives," *Biometrika*, 74, 33–43. 6, 10

DONG, C., G. LI, AND X. FENG (2019): "Lack-of-Fit Tests for Quantile Regression Models," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81, 629–648. 18, 27

ESCANCIANO, J. C. AND S.-C. GOH (2014): "Specification Analysis of Linear Quantile Models," *Journal of Econometrics*, 178, 495–507. 18, 27

ESCANCIANO, J. C. AND C. VELASCO (2010): "Specification Tests of Parametric Dynamic Conditional Quantiles," *Journal of Econometrics*, 159, 209–221. 27

FAN, J. AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*, Monographs on Statistics and Applied Probability 66, New York: Chapamand & Hall/CRC. 11, 13

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. 1

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. 34

GARDEAZABAL, J. AND A. UGIDOS (2005): "Gender Wage Discrimination at Quantiles," *Journal of Population Economics*, 18, 165–179. 34

HORVATH, L. AND P. KOKOSZKA (2012): *Inference for Functional Data with Applications*, vol. 200, New York: Springer-Verlag. 1

HUIZINGA, F. (1990): "An Overlapping Generations Model of Wage Determination," *The Scandinavian Journal of Economics*, 92, 81–98. 34

KATO, K. (2012): "Estimation in Functional Linear Quantile Regression," *The Annals of Statistics*, 40, 3108–3136. 1, 8

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. 34

KIM, T.-H. AND H. WHITE (2003): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," in *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*, Emerald Group Publishing Limited, 107–132. 3, 8, 2

KNIGHT, K. (1998): "Limiting Distributions for L_1 Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755–770. 14, 2

KOENKER, R. (2005): *Quantile Regression*, Econometric Society Monographs, Cambridge University Press. 3

KOENKER, R. AND G. BASSETT (1982): "Tests of Linear Hypotheses and l"1 Estimation," *Econometrica*, 50, 1577–1583. 26

KOENKER, R. AND G. J. BASSETT (1978): "Regrssion Quantiles," *Econometrica*, 46, 33–50. 7

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. 34

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. 1

LI, M., K. WANG, A. MAITY, AND A.-M. STAICU (2022): "Inference in Functional Linear Quantile Regression," *Journal of Multivariate Analysis*, 190, 104985. 1

LIGHT, A. AND M. URETA (1995): "Early-Career Work Experience and Gender Wage Differentials," *Journal of Labor Economics*, 13, 121–154. 34

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. 34

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302. 5, 34

——— (1974): Schooling, Experience and Earnings, New York: National Bureau of Economic Research. 5, 34

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. 5, 17, 34

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167. 23

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. 14

NICODEMO, C. (2009): "Gender Pay Gap and Quantile Regression in European Families," IZA Discussion Papers 3978, Institute of Labor Economics (IZA). 34

OBERHOFER, W. AND H. HAUPT (2016): "Asymptotic Theory for Nonlinear Quantile Regression under Weak Dependence," *Econometric Theory*, 32, 686–713. 8, 13, 14, 1, 2, 16

PHILLIPS, P. C. B. (2015): "Halbert White Jr. Memorial JFEC Lecture: Pitfalls and Possibilities in Predictive Regression," *Journal of Financial Econometrics*, 13, 521–555. 3, 27

PHILLIPS, P. C. B. AND L. JIANG (2016): "Parametric Autoregression in Function Space," Working Paper, Singapore Management University. 1

——— (2025): "Cross section curve data autoregression," *Cowles Foundation Discussion Paper No. CFDP 2439*, 48. 1, 4

POLLARD, D. (1984): *Convergence of Stochastic Processes*, New York: Springer-Verlag. 16, 3

——— (1985): "New Ways to Prove Central Limit Theorem," *Econometric Theory*, 1, 295–314. 29, 14, 15

RAMSAY, J. AND B. SILVERMAN (1997): *The Analysis of Functional Data*, New York: Springer. 1

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. 1

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. 1

SAKELLARIOU, C. (2004a): "Gender-earnings Differentials using Quantile Regressions," *Journal of Labor Research*, 25, 457–468. 34

——— (2004b): "The Use of Quantile Regressions in Estimating Gender Wage Differentials: A Case Study of the Philippines," *null*, 36, 1001–1007. 34

WANG, Y. (2011): *Smoothing Splines: Methods and Applications*, New York: Chapman & Hall/CRC. 5, 10

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25. 4, 9

WISWALL, M. AND B. ZAFAR (2017): "Preference for the Workplace, Investment in Human Capital, and Gender," *The Quarterly Journal of Economics*, 133, 457–507. 35

WOOLDRIDGE, J. AND H. WHITE (1988): "Some Invariance Principles and Central Limit Theorems for Dependence Heterogeneous Processess," *Econometric Theory*, 4, 210–230. 23, 8, 26

ZHANG, J.-T. AND J. CHEN (2007): "Statistical Inferences for Functional Data," *The Annals of Statistics*, 35, 1052 – 1079. 3, 5, 6, 10, 32

ZOU, H. AND M. YUAN (2008): "Composite Quantile Regression and the Oracle Model Selection Theory," *The Annals of Statistics*, 36, 1108–1126. 23

Statistics	Levels\n						
Size of the Tests							
Statistics	Levels\n	50	100	200	300	400	500
$\bar{\mathcal{W}}_n$	10%	13.84	13.20	11.22	11.14	10.98	10.88
	5%	8.18	7.88	6.26	5.50	5.80	5.74
	1%	2.60	2.46	1.88	1.22	1.62	1.34
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	11.34	9.82	10.18	9.70	10.44	9.66
	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
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	10.62	10.42	9.86	10.42	10.08	10.22
	5%	5.46	5.38	4.74	4.98	5.40	4.90
	1%	1.22	1.24	1.12	0.90	1.16	0.98
Power of the Tests							
Statistics	Levels\n	20	40	60	80	100	120
$\bar{\mathcal{W}}_n$	10%	81.36	97.60	99.76	100.0	100.0	100.0
	5%	73.14	95.14	99.30	99.94	99.96	100.0
	1%	55.98	87.66	96.90	99.46	99.82	99.98
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	74.10	97.00	99.84	100.0	100.0	100.0
	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
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	86.56	98.66	99.94	100.0	100.0	100.0
	5%	78.38	97.36	99.86	100.0	100.0	100.0
	1%	57.28	90.28	98.56	99.84	99.96	100.0
Local Power of the Tests							
Statistics	Levels\n	50	100	200	300	400	500
$\bar{\mathcal{W}}_n$	10%	84.80	82.04	81.76	79.28	80.12	79.14
	5%	77.86	73.68	72.00	69.92	70.28	69.40
	1%	60.90	55.08	50.60	48.30	47.90	47.32
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	83.16	80.60	80.12	78.00	78.96	77.74
	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
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	89.88	87.22	87.30	85.20	86.32	85.56
	5%	83.10	79.92	79.42	76.96	78.00	77.06
	1%	64.30	60.48	57.60	56.24	56.86	55.30

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 $Z_i \sim_{\text{iid}} U[0,1]$ and $W_i \sim_{\text{iid}} U[0,1]$. For $\tau_1 = 1/3$ and $\tau_2 = 2/3$, we let $\rho_{\tau_j}(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ 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 $\theta^*_{\tau_2} \neq 1/6$.

τ	Est.	Errors	Correctly Specified		Misspecified		with Nuisance	
			$n = 100$	$n = 250$	$n = 100$	$n = 250$	$n = 100$	$n = 250$
0.25	$\hat{\theta}_0$	Bias	0.0067	0.0060	0.0193	0.0192	0.0072	0.0069
		RMSE	0.0081	0.0072	0.0200	0.0194	0.0086	0.0076
		MAE	0.0071	0.0067	0.0193	0.0192	0.0071	0.0069
	$\hat{\theta}_1$	Bias	-0.0642	-0.0640	0.1805	0.1781	-0.0688	-0.0675
		RMSE	0.0700	0.0666	0.1834	0.1797	0.0734	0.0701
		MAE	0.0643	0.0640	0.1805	0.1781	0.0677	0.0675
	$\hat{\theta}_2$	Bias	0.0646	0.0639	-0.1981	-0.1931	0.0678	0.0683
		RMSE	0.0702	0.0675	0.2019	0.1952	0.0760	0.0718
		MAE	0.0649	0.0642	0.1981	0.1931	0.0683	0.0683
0.50	$\hat{\theta}_0$	Bias	0.0002	0.0000	-0.0495	-0.0493	-0.0000	0.0001
		RMSE	0.0055	0.0036	0.0498	0.0498	0.0298	0.0188
		MAE	0.0044	0.0028	0.0495	0.0497	0.0045	0.0029
	$\hat{\theta}_1$	Bias	0.0002	0.0000	0.2824	0.2787	0.0056	0.0043
		RMSE	0.0289	0.0199	0.2837	0.2796	0.0343	0.0224
		MAE	0.0230	0.0158	0.2824	0.2787	0.0236	0.0150
	$\hat{\theta}_2$	Bias	0.0010	0.0001	-0.2693	-0.2620	-0.0054	-0.0041
		RMSE	0.0345	0.0210	0.2722	0.2642	0.0343	0.0224
		MAE	0.0270	0.0178	0.2693	0.2620	0.0272	0.0179
0.75	$\hat{\theta}_0$	Bias	-0.0071	-0.0067	-0.1248	-0.1198	-0.0070	-0.0070
		RMSE	0.0085	0.0076	0.1260	0.1205	0.0085	0.0077
		MAE	0.0073	0.0069	0.1248	0.1198	0.0073	0.0070
	$\hat{\theta}_1$	Bias	0.0679	0.0661	0.4758	0.4570	0.0753	0.0750
		RMSE	0.0738	0.0695	0.4815	0.4601	0.0808	0.0774
		MAE	0.0681	0.0667	0.4758	0.4570	0.0754	0.0750
	$\hat{\theta}_2$	Bias	-0.0691	-0.0671	-0.4544	-0.4387	-0.0764	-0.0759
		RMSE	0.0766	0.0711	0.4605	0.4417	0.0831	0.0790
		MAE	0.0696	0.0675	0.4544	0.4387	0.0766	0.0759

Table 2: FINITE-SAMPLE ACCURACY OF THE ESTIMATOR UNDER CORRECT, MISSPECIFIED, AND NUISANCE DESIGNS. The table reports the finite sample accuracy of the estimator under three data-generating designs: correct specification, misspecification, and nuisance effect. The baseline model is $G_i(\gamma) := \sigma_0 X_i + \sigma_1 X_i \gamma + \sigma_2 X_i \gamma^2 + s(\gamma)\varepsilon_i$ for $\gamma \in \Gamma := [0, 1]$, with $s(\gamma) := b_0 + b_1\gamma + b_2\gamma^2$, where $(b_0, b_1, b_2) = (0.01, -0.1, 0.1)$ and $(\sigma_0, \sigma_1, \sigma_2) = (0.1, 0.2, 0.3)$. The covariates satisfy $X_i \sim_{\text{iid}} U[0, 1]$ and the errors satisfy $\varepsilon_i \sim_{\text{iid}} N(0, 1)$. The misspecified design perturbs the distribution of X_i through a non-linear mixture mechanism based on sieve components, while the nuisance design reconstructs $G_i(\cdot)$ from noisy discrete measurements.

		Size of the Tests				
Statistics	Levels$\backslash n$	90	130	170	210	250
$\bar{\mathcal{W}}_n$	10%	0.1082	0.1057	0.1043	0.1034	0.1002
	5%	0.0529	0.0518	0.0513	0.0512	0.0504
	1%	0.0121	0.0111	0.0102	0.0093	0.0096
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	0.1032	0.1034	0.1014	0.0999	0.0994
	5%	0.0523	0.0514	0.0499	0.0495	0.0506
	1%	0.0120	0.0100	0.0104	0.0098	0.0098
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	0.1171	0.1168	0.1125	0.1052	0.1006
	5%	0.0637	0.0630	0.0590	0.0530	0.0504
	1%	0.0248	0.0208	0.0156	0.0108	0.0108
		Power of the Tests ($n = 100$)				
Statistics	Levels$\backslash\beta$	0.2	0.4	0.6	0.8	1.0
$\bar{\mathcal{W}}_n$	10%	0.9896	0.9932	0.9972	0.9990	0.9992
	5%	0.9514	0.9696	0.9856	0.9932	0.9984
	1%	0.9230	0.9328	0.9686	0.9894	0.9970
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	0.9780	0.9918	0.9972	0.9988	0.9991
	5%	0.9546	0.9698	0.9850	0.9958	0.9978
	1%	0.9178	0.9442	0.9714	0.9922	0.9970
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	0.9448	0.9570	0.9692	0.9722	0.9864
	5%	0.8944	0.9124	0.9358	0.9426	0.9664
	1%	0.8512	0.8748	0.9042	0.9246	0.9496
		Power of the Tests ($n = 250$)				
Statistics	Levels$\backslash\beta$	0.2	0.4	0.6	0.8	1.0
$\bar{\mathcal{W}}_n$	10%	0.9914	0.9970	1.0000	1.0000	1.0000
	5%	0.9614	0.9872	0.9996	0.9998	1.0000
	1%	0.9250	0.9742	0.9964	0.9996	1.0000
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	0.9912	0.9976	1.0000	1.0000	1.0000
	5%	0.9588	0.9868	0.9992	1.0000	1.0000
	1%	0.9226	0.9744	0.9964	0.9996	1.0000
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	0.9882	0.9962	0.9984	0.9990	1.0000
	5%	0.9552	0.9828	0.9958	0.9974	0.9986
	1%	0.9224	0.9688	0.9946	0.9956	0.9984

Table 3: EMPIRICAL REJECTION RATES OF THE TEST STATISTICS UNDER THE NULL AND ALTERNATIVE HYPOTHESES. The table reports the empirical rejection rates of the Wald, LM, and QLR test statistics under the quadratic model. We let $G_i(\gamma) := \sigma_0 X_i + \sigma_1 X_i \gamma + \sigma_2 X_i \gamma^2 + s(\gamma)\varepsilon_i$ for $\gamma \in \Gamma := [0,1]$, with $s(\gamma) := b_0 + b_1\gamma + b_2\gamma^2$, where $(b_0, b_1, b_2) = (0.01, -0.1, 0.1)$, $(\sigma_0, \sigma_1, \sigma_2) = (0.1, 0.2, 0.3)$, and $\varepsilon_i \sim_{\text{iid}} N(0,1)$. Inferences are conducted for the joint hypotheses $\mathbb{H}_0^{(2)} : (\theta^*_{\tau_1,2}, \theta^*_{\tau_2,2}, \theta^*_{\tau_3,2})' = r_2$ and $\mathbb{H}_a^{(2)} : (\theta^*_{\tau_1,2}, \theta^*_{\tau_2,2}, \theta^*_{\tau_3,2})' \neq r_2$, where $r_2 := (\sigma_2\tau_1 + b_2\Phi^{-1}(\tau_1),\ \sigma_2\tau_2 + b_2\Phi^{-1}(\tau_2),\ \sigma_2\tau_3 + b_2\Phi^{-1}(\tau_3))'$ for $(\tau_1, \tau_2, \tau_3) = (0.25, 0.50, 0.75)$. For size calculations, the covariates follow $X_i := \mathbb{1}\{W_i > \beta\}Z_i + \mathbb{1}\{W_i \leq \beta\}Z_i^{1/2}$ with $\beta = 0$, where $W_i \sim_{\text{iid}} U[0,1]$ and $Z_i \sim_{\text{iid}} U[0,1]$.

		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**

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

Table 4: 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.

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	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

Table 5: 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
Inference results on the quantiles of the log income path across different education levels								
$\tau = 0.25$	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**

Table 6: 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*

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

Table 7: 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.

		Education Level	Wald	LM	QLR
Inference results on the quantiles of the log income path across different genders					
$\tau = 0.25$	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 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 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.

		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

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

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 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**

Table 10: 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.

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

Table 11: 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 applied to Lifetime Income Curves'*

by

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

[a]Yonsei University

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

[e]National University of Singapore

This Supplement is an Appendix that provides proofs of the results in the paper, including the lemmas, as well as some additional empirical findings. Proofs are given in Section D, and we discuss the quasi-2FQR estimator when the model is driven by covariates in Section E. Additional simulations and empirical applications are in Sections F and G.

A Estimation of Single Quantile Function

In this section, we discuss the large sample distributions of the quasi-FQR, FQR, and 2FQR estimators by supposing that $\mathcal{M}_\tau$ is correctly specified or misspecified.

A.1 Quasi-FQR Estimation

To analyze the quasi-FQR estimator, we first provide the regularity condition for the functional data without the nuisance effects that is obtained from Assumption 5 by supposing π_* is known:

Assumption S.1. *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 $t = 1, 2, \ldots, c_\tau$, $\sup_{\theta_\tau} \left|(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau t})\rho_\tau(\,\cdot\,,\theta_\tau)\right| \leq M$.* □

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 we apply the approximation approach in Oberhofer and Haupt (2016, pp. 710–711):

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

using the continuity conditions in Assumption 4, where $\nabla_{\theta_\tau} := \partial/\partial\theta_\tau$ and

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

*The co-editor, Degui Li, and three anonymous referees provided very helpful comments for which we are most grateful. Cho acknowledges research support from the Yonsei University Research Grant of 2023; 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; and Seo acknowledges research support from the Tier 1 grant FY2023-FRC1-003.

From this approximation we obtain (3). 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, and further enables analysis of the FQR estimator as a special case of the quasi-FQR estimator, as detailed below.

Although it is not necessary to represent the quasi-FQR estimator in the form of (3) for empirical analysis, 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), which is useful in deriving its limit distribution. We note the following three key features from (3). First, the approximation on the right side is analogous to that of standard quasi-maximum likelihood estimation, which is essentially the result of defining the objective function through an integral transformation. Second, 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 required in Assumption S.2 below. Third, the limit distribution of the quasi-FQR estimator is determined mainly by the other components on the right side of (3). From Assumptions 4 and S.1 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 $J_{\tau i} := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)\left(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta_\tau^*)\} - \tau\right)d\mathbb{Q}(\gamma)$ 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

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

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. In addition, the covariance matrix B_τ^* can be related to previous research. Although our interest lies in the marginal distribution of $G_i(\gamma)$, the covariance matrix of the QR estimator examined by Kim and White (2003) and Angrist et al. (2006) can be obtained from B_τ^* if we let $\mathbb{Q}(\cdot)$ be a distribution function with a point mass and if we let $\nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)$ be their conditioning variable.

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

Assumption S.2. *(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 S.1. *Given Assumptions 1, 4, S.1, and S.2, if* $\mathcal{M}_\tau$ *is misspecified,* $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta_\tau^*) \overset{\text{A}}{\sim} \mathcal{N}(0, C_\tau^*)$, *where* $C_\tau^* := A_\tau^{*-1}B_\tau^*A_\tau^{*-1}$. □

The distribution of the quasi-FQR estimator is therefore asymptotically normal with a variance matrix that has a sandwich-form. Even though we are dealing with functional data, Theorem S.1 implies that we can develop a testing methodology for the unknown parameter using the Wald principle in a manner similar to the random variable 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 3.2.

A.2 FQR Estimation

When $\mathcal{M}_\tau$ is correctly specified the asymptotic theory given in Theorem S.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 $F_\gamma(\rho_\tau(\gamma, \theta_\tau^0)) = \tau$. It follows that $(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^0)\} - \tau) = (\mathbb{1}\{U_i(\gamma) \leq \tau\} - \tau)$, 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 $\mathbb{E}[\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^0)\}] = \tau$. Therefore,

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

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 S.1. *Given Assumptions 1 and 4, 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 $\bar{\gamma} \in \Gamma$, $\mathbb{E}[\mathcal{G}_\tau(\gamma)\mathcal{G}_\tau(\bar{\gamma})]$ $= \kappa_\tau(\gamma, \bar{\gamma}) := \mathbb{E}\left[\mathbb{1}\{U_i(\gamma) \leq \tau\}\mathbb{1}\{U_i(\bar{\gamma}) \leq \tau\}\right] - \tau^2$.* □

Note that $\kappa_\tau(\gamma, \bar{\gamma})$ generalizes the core covariance matrix component of the quantile regression estimator provided in Koenker (2005) by letting the kernel be indexed by the index of the functional data. Note that $\kappa_\tau(\gamma, \gamma) = \tau(1-\tau)$, as in Koenker (2005, p. 73), and therefore from Lemma S.1, 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 S.1. Lemma S.1 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, 1968, 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 S.1 satisfies Ossiander's L^2 entropy condition.

A remark is warranted on Lemma S.1. Although Assumptions 1 and 4 assume that $G_i(\cdot)$ is continuous almost surely, the stochastic equicontinuity in Lemma S.1 can be established for a wider class of functions.

For example, Billingsley (1968, 1999) demonstrates that stochastic equicontinuity is effective even when continuous functional observations are extended to càdlàg functions (right-continuous functions with left limits). As another example, Phillips and Jiang (2025, lemma A) establish stochastic equicontinuity in L_2 Hilbert space by employing the maximal martingale inequality in proving an FCLT. Exploiting these stochastic equicontinuity conditions we can obtain the weak limit of $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{G_i(\cdot) \leq \rho_\tau(\cdot, \theta_\tau^0)\} - \tau)$ similarly to Lemma S.2.

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

Assumption S.3. *(i) $\lambda_{\min}(A_\tau^0) > 0$, where $A_\tau^0 := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0) f_\gamma(\rho_\tau(\gamma, \theta_\tau^0))\nabla'_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0) d\mathbb{Q}(\gamma)$ 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 $B_\tau^0 := \int_\gamma \int_{\bar\gamma} \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0)\kappa_\tau(\gamma, \bar\gamma)\nabla'_{\theta_\tau}\rho_\tau(\bar\gamma, \theta_\tau^0) d\mathbb{Q}(\gamma) d\mathbb{Q}(\bar\gamma)$ and $\kappa_\tau(\gamma, \bar\gamma) := \mathbb{E}[\mathbb{1}\{F_\gamma(G_i(\gamma)) \leq \tau\}\mathbb{1}\{F_\gamma(G_i(\bar\gamma)) \leq \tau\}] - \tau^2$.* □

Theorem S.2. *Given Assumptions 1, 4, S.1, and S.3, 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 useful for developing testing methodologies and is obtained in a different way from that of quasi-FQR, 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 3.2.

A.3 2FQR Estimation

As in Section A.2, the limit theory of the quasi-2FQR estimator continues to apply for 2FQR 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 $\bar\pi_{\gamma n i}$ between π^* and $\widehat\pi_n$, it follows that

$$\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0) \sum_{i=1}^n \left(\mathbb{1}\{\widehat G_i(\gamma) \leq \rho_\tau(\gamma, \theta_\tau^0)\} - \tau\right) d\mathbb{Q}(\gamma) \tag{S.2}$$

$$= \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma, \theta_\tau^0) \sum_{i=1}^n \left(\mathbb{1}\{\widetilde G_i(\gamma, \pi^*) + \nabla'_\pi \widetilde G_i(\gamma, \bar\pi_{\gamma n i})(\widehat\pi_n - \pi^*) \leq \rho_\tau(\gamma, \theta_\tau^0)\} - \tau\right) d\mathbb{Q}(\gamma).$$

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

Lemma S.2. *Given Assumptions 4, 2, 3, and 5, if $\mathcal{M}_\tau$ is correctly specified, $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{\widetilde G_i(\cdot, \pi^*) + \nabla'_\pi \widetilde G_i(\cdot, \bar\pi_{(\cdot)ni})(\widehat\pi_n - \pi^*) \leq \rho_\tau(\cdot, \theta_\tau^0)\} - \tau) \Rightarrow \widetilde{\mathcal{G}}_\tau(\cdot)$, where $\nabla'_\pi := \partial/\partial\pi'$ and $\widetilde{\mathcal{G}}_\tau(\cdot)$ is a zero-mean*

Gaussian process such that for each γ *and* $\bar\gamma$, $\mathbb{E}[\widetilde{\mathcal{G}}_\tau(\gamma)\widetilde{\mathcal{G}}_\tau(\bar\gamma)] = \widetilde{\kappa}_\tau(\gamma,\bar\gamma)$ *with*

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

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 S.1, 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 $\int_\gamma\int_{\bar\gamma} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*)\widetilde{\kappa}_\tau(\gamma,\bar\gamma)\nabla'_{\theta_\tau}\rho_\tau\ (\gamma,\theta_\tau^*)d\mathbb{Q}(\gamma)d\mathbb{Q}(\bar\gamma)$, so that $\widetilde{B}_\tau^*$ can be estimated by first estimating the covariance kernel $\widetilde{\kappa}_\tau(\cdot,\cdot)$ as detailed in Section A.4.

Lemma S.2 is established by showing stochastic equicontinuity, as in Lemma S.1. 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 $\widehat{\mu}_{ni}(\gamma) := \nabla'_\pi G_i(\gamma,\bar\pi_{\gamma ni})(\widehat{\pi}_n - \pi^*)$, we have

$$\begin{aligned}&\mathbb{1}\{G_i(\gamma,\pi^*) + \nabla'_\pi G_i(\gamma,\bar\pi_{\gamma ni})(\widehat{\pi}_n - \pi^*) \le \rho_\tau(\gamma,\theta_\tau^0)\} - \tau = \mathbb{1}\{G_i(\gamma,\pi^*) \le \rho_\tau(\gamma,\theta_\tau^0)\} - \tau\\ &\quad + \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^0) < G_i(\gamma) \le \rho_\tau(\gamma,\theta_\tau^0) - \widehat{\mu}_{ni}(\gamma)\} - \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^0) - \widehat{\mu}_{ni}(\gamma) < G_i(\gamma) \le \rho_\tau(\gamma,\theta_\tau^0)\}. \quad \text{(S.3)}\end{aligned}$$

Lemma S.1 applies to the first term on the right side, converging weakly to $\mathcal{G}_\tau(\cdot)$. But the second and third terms of (S.3) 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 2FQR estimator when $\mathcal{M}_\tau$ is correctly specified the following regularity conditions are employed. These conditions match those in Assumption 6 and are employed for the same reason.

Assumption S.4. *(i)* $\lambda_{\min}(A_\tau^0) > 0$*; (ii)* $\lambda_{\min}(L_\tau^0) > 0$*; and (iii)* $\lambda_{\min}(\widetilde{B}_\tau^0) > 0$*, where*

$$L_\tau^0 := \begin{bmatrix} H^* & V_\tau^{0\prime} \\ V_\tau^0 & B_\tau^0 \end{bmatrix},$$

$V_\tau^0 := \int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le \rho_\tau(\gamma,\theta_\tau^0)\} - \tau)\cdot S_i']d\mathbb{Q}(\gamma)$, $\widetilde{B}_\tau^0 := \int_\gamma\int_{\bar\gamma} \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)\widetilde{\kappa}_\tau(\gamma,\bar\gamma)$ $\nabla'_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)d\mathbb{Q}(\gamma)d\mathbb{Q}(\bar\gamma)$*, and* B_τ^0 *is defined in Assumption S.3.* □

Assumption S.4 implies that S_i, $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\theta_\tau^0)\} - \tau)d\mathbb{Q}(\gamma)$ and $\int_\gamma \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^0)$ $(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma,\theta_\tau^0)\} - \tau)d\mathbb{Q}(\gamma)$ all have positive definite variance matrices.

Theorem S.3. *Given Assumptions 2, 3, 4, 5, and S.4, if* $\mathcal{M}_\tau$ *is correctly specified,* $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_\tau^0) \overset{\text{A}}{\sim} \mathcal{N}(0,\widetilde{C}_\tau^0)$*, where* $\widetilde{C}_\tau^0 := A_\tau^{0\,-1}\widetilde{B}_\tau^0 A_\tau^{0\,-1}$*.* □

Even when $\mathcal{M}_\tau$ is correctly specified, the limit distribution of the 2FQR estimator differs from that of the FQR estimator in Theorem S.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 $\widetilde{\theta}_{\tau n}$.

A.4 Asymptotic Variance Matrix Estimation

In this section, we provide consistent asymptotic variance matrix estimators of the FQR, quasi-FQR, and 2FQR estimators.

The consistent asymptotic variance matrix estimators are as follows. First, we define the asymptotic variance matrix estimator of the quasi-FQR estimator as $\widehat{B}_{\tau n} := n^{-1}\sum_{i=1}^n J_{\tau ni}J'_{\tau ni}$, where

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

Second, we define the asymptotic variance matrix estimator of the 2FQR estimator as

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

where

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

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

Finally, the asymptotic variance matrix estimator of the FQR estimator is defined as

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

where

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

These estimators are consistent for the asymptotic variance matrices as the following theorem states:

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

(ii) Given Assumption 2, 3, 4, 5, 7, if $\mathcal{M}_\tau$ is correctly specified and Assumption S.4 holds, $\widetilde{B}^\sharp_{\tau n} \overset{\mathbb{P}}{\to} \widetilde{B}^0_\tau$.

□

B Joint Estimation of Multiple Quantile Functions

We redefine the quasi-2MFQR estimator by modifying the underlying the DGP and model conditions. If $\mathcal{M}$ is misspecified and the nuisance effects do not exist, we call the estimator the quasi-multiple functional quantile regression (quasi-MFQR) estimator and denote it as $\widehat{\theta}_n := (\widehat{\theta}'_{\tau_1 n}, \widehat{\theta}'_{\tau_2 n}, \ldots, \widehat{\theta}'_{\tau_p n})'$, where for each j, $\widehat{\theta}_{\tau_j n}$ denotes the quasi-FQR estimator obtained from $\mathcal{M}_{\tau_j}$. In addition, if $\mathcal{M}$ is correctly specified, we call $\widetilde{\theta}_n$ and $\widehat{\theta}_n$ the 2MFQR and MFQR estimators, respectively.

B.1 Quasi-MFQR Estimations

In this section, we discuss the large sample distributions of the quasi-MFQR estimator by supposing that $\mathcal{M}$ is misspecified but the nuisance effects do not exist.

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 $\{w_j\}$. For this estimator, we provide the tailored conditions as follows:

Assumption S.5. *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 $j, t = 1, \ldots, c_{\tau_\ell}$, $\sup_{(\gamma, \theta_{\tau_\ell})} |(\partial^2 / \partial\theta_{\tau_\ell j} \partial\theta_{\tau_\ell t}) \rho_{\tau_\ell}(\gamma, \theta_{\tau_\ell})| \leq M$.* □

Assumption S.6. *(i) $\lambda_{\min}(A^*) > 0$; and (ii) $\lambda_{\min}(B^*) > 0$, where $A^* := \mathrm{diag}[A^*_{\tau_1}, A^*_{\tau_2}, \ldots, A^*_{\tau_p}]$ and $B^* := \mathbb{E}[J_i J_i']$.* □

Assumptions S.5 and S.6 extend Assumptions S.1 and S.2, ensuring that the regular bound conditions in Assumptions S.1 and S.2 apply to the case of multiple quantile function estimation. In addition, they are obtained by removing the effects of the nuisance parameters in Assumptions 9 and 10.

Given these conditions, we obtain the large sample distribution of the quasi-MFQR estimator as a corollary of Theorem 3.

Corollary S.1. *Given Assumptions 2, 3, 8, S.5, and S.6, if $\mathcal{M}$ is misspecified, $\sqrt{n}(\widehat{\theta}_n - \theta^*) \overset{\mathrm{A}}{\sim} \mathcal{N}(0, C^*)$, where $C^* := A^{*-1} B^* A^{*-1}$.* □

B.2 2MFQR Estimation

We next examine the large sample distribution of the 2MFQR estimator by supposing that $\mathcal{M}$ is correctly specified and the nuisance effects are present. For this purpose we define some notation. Let $\nabla_\theta \rho_\tau(\gamma, \theta^0) := \mathrm{diag}[\nabla_{\theta_{\tau_1}} \rho_{\tau_1}(\gamma, \theta^0_{\tau_1}), \nabla_{\theta_{\tau_2}} \rho_{\tau_2}(\gamma, \theta^0_{\tau_2}), \ldots, \nabla_{\theta_{\tau_p}} \rho_{\tau_p}(\gamma, \theta^0_{\tau_p})]$ and

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

These quantities are defined with some abuse of notation: $\nabla_\theta \rho_\tau(\gamma, \theta^0)$ 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 $\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma, \theta^0)\}$ is a $p \times 1$ column vector instead of a scalar.

The joint limit distribution of the 2MFQR estimator is efficiently obtained by collecting the asymptotic approximations in (3) into a vector. Specifically, (3) can be rewritten as

$$\sqrt{n}(\widetilde{\theta}_n - \theta^0) = -A^{0\,-1}\left(\int_\gamma \nabla_\theta \rho_\tau(\gamma, \theta^0)\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma, \theta^0)\} - \tau\Big)\, d\mathbb{Q}(\gamma)\right) + o_{\mathbb{P}}(1), \quad \text{(S.4)}$$

where $A^0 := \text{diag}[A^0_{\tau_1}, A^0_{\tau_2}, \ldots, A^0_{\tau_p}]$ and $\tau := (\tau_1, \tau_2, \ldots, \tau_p)'$. So the limit distribution of the 2MFQR estimator is determined by the two factors in the leading term on the right side of (S.4). 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 S.2 shows that each component in $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{\widehat{G}_i(\cdot) \leq \rho_\tau(\cdot, \theta^0)\} - \tau)$ weakly converges to a Gaussian stochastic process and the following lemma proves that the full vector converges weakly to a vector Gaussian process.

Lemma S.3. *Given Assumptions 2, 3, 8, and 9, if $\mathcal{M}$ is correctly specified, $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{\widehat{G}_i(\cdot) \leq \rho_\tau(\cdot, \theta^0)\} - \tau) \Rightarrow \widetilde{\mathcal{G}}(\cdot) := [\widetilde{\mathcal{G}}_{\tau_1}(\cdot), \widetilde{\mathcal{G}}_{\tau_2}(\cdot), \cdots, \widetilde{\mathcal{G}}_{\tau_p}(\cdot)]'$, where $\widetilde{\mathcal{G}}(\cdot)$ is a mean-zero Gaussian process such that for j and $t = 1, 2, \ldots, p$, and γ and $\bar{\gamma} \in \Gamma$, the covariance kernel is*

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

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

Lemma S.3 extends Lemma S.2 and specializes to it when $p = 1$. Further, when $j = t$, $\kappa_{\tau_j\tau_t}(\cdot, \cdot)$ and $\widetilde{\kappa}_{\tau_j\tau_t}(\cdot, \cdot)$ are identical to $\kappa_{\tau_j}(\cdot, \cdot)$ and $\widetilde{\kappa}_{\tau_j}(\cdot, \cdot)$. Lemma S.3 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

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

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 t-th block column matrix

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

The limit distribution of the 2MFQR estimator is obtained under the following regularity conditions.

Assumption S.7. *(i) $\lambda_{\min}(A^0) > 0$; (ii) $\lambda_{\min}(L^0) > 0$; and (iii) $\lambda_{\min}(\widetilde{B}^0) > 0$, where*

$$L^0 := \left[\begin{array}{cc} H^* & V^{0\prime} \\ V^0 & B^0 \end{array} \right],$$

$V^0 := [V^{0\prime}_{\tau_1}, V^{0\prime}_{\tau_2}, \ldots, V^{0\prime}_{\tau_p}]'$, B^0 is a $c \times c$ matrix such that for each j and $t = 1, 2, \ldots p$, its j-th block row and t-th block column matrix is $B^0_{\tau_j \tau_t} := \int_\gamma \int_{\bar{\gamma}} \nabla_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^0_{\tau_j}) \kappa_{\tau_j \tau_t}(\gamma, \bar{\gamma}) \nabla'_{\theta_{\tau_j}} \rho_{\tau_t}(\bar{\gamma}, \theta^0_{\tau_t}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\bar{\gamma})$, and $\widetilde{B}^0$ is defined as just above this assumption. □

By Assumption S.7, S_i, $\int_\gamma \nabla_\theta \rho_\tau(\gamma, \theta^0)(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^*) \leq \rho_\tau(\gamma, \theta^0)\} - \tau) d\mathbb{Q}(\gamma)$ and $\int_\gamma \nabla_\theta \rho_\tau(\gamma, \theta^0)(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_\tau(\gamma, \theta^0)\} - \tau) d\mathbb{Q}(\gamma)$ have positive definite variance matrices. In the diagonal block with $j = t$ the matrices $B^0_{\tau_j \tau_t}$ and $\widetilde{B}^0_{\tau_j \tau_t}$ are identical to $B^0_{\tau_j}$ and $\widetilde{B}^0_{\tau_j}$ as defined earlier in Section A.3, Assumptions S.3 and S.4. Under Assumption S.7 and earlier conditions, the limit distribution of the 2MFQR estimator is non-degenerate and given in the following theorem.

Theorem S.5. *Under Assumptions 2, 3, 8, 9, and S.7, 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 S.5 relates closely to the misspecifed case. Specifically, when $\theta^* = \theta^0$, we have $\widetilde{B}^* = \widetilde{B}^0$ and $A^* = A^0$, so that $\widetilde{C}^* = \widetilde{C}^0$.

B.3 MFQR Estimation

The large sample distribution of the MFQR estimator, obtained by supposing that $\mathcal{M}$ is correct and that there are no nuisance effects, is obtained by applying Theorem S.5. In particular, by setting $S_i \equiv 0$ in the definition of $\widetilde{\kappa}_{\tau_j \tau_t}(\gamma, \bar{\gamma})$, which leads to $\widetilde{\kappa}_{\tau_j \tau_t}(\cdot, \cdot) = \kappa_{\tau_j \tau_t}(\cdot, \cdot)$ and $\widetilde{B}^0 = B^0$. The following assumption tailors Assumption S.6 in the context of correct model assumption:

Assumption S.8. *(i) $\lambda_{\min}(A^0) > 0$; and (ii) $\lambda_{\min}(B^0) > 0$, where B^0 is a $c \times c$ matrix such that for each j and $t = 1, 2, \ldots p$, its j-th block row and t-th block column matrix is $B^0_{\tau_j \tau_t} := \int_\gamma \int_{\bar{\gamma}} \nabla_{\theta_{\tau_j}} \rho_{\tau_j}(\gamma, \theta^0_{\tau_j}) \kappa_{\tau_j \tau_t}(\gamma, \bar{\gamma}) \nabla'_{\theta_{\tau_j}} \rho_{\tau_t}(\bar{\gamma}, \theta^0_{\tau_t}) d\mathbb{Q}(\gamma) d\mathbb{Q}(\bar{\gamma})$.* □

As a consequence, we have the following corollary:

Corollary S.2. *Under Assumptions 2, 3, 8, S.5, and S.8, if $\mathcal{M}$ is correctly specified, $\sqrt{n}(\widehat{\theta}_n - \theta^0) \overset{\text{A}}{\sim} \mathcal{N}(0, C^0)$, where $C^0 := A^{0\,-1} B^0 A^{0\,-1}$.* □

B.4 Variance matrix estimation

The limit theory of Sections 4.1, B.1, B.2, and B.3 enables hypothesis testing on the unknown model parameters once the relevant asymptotic variance matrices are estimated. The approach follows Section A.4 closely and is only briefly detailed here.

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

$$\widehat{B}_n := \frac{1}{n}\sum_{i=1}^{n} J_{ni}J'_{ni}.$$

It immediately follows that $\widehat{B}_n \overset{\mathbb{P}}{\to} B^*$ under Assumptions 1, 8, S.5, and S.6 by applying Theorem S.4.

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

$$\widehat{\kappa}_n(\gamma,\bar{\gamma}) := \frac{1}{n}\sum_{i=1}^{n}\Big(\mathbb{1}\{G_i(\gamma) \le \rho_\tau(\gamma,\widehat{\theta}_n)\} - \tau\Big)\Big(\mathbb{1}\{G_i(\bar{\gamma}) \le \rho_\tau(\bar{\gamma},\widehat{\theta}_n)\} - \tau\Big)',$$

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

and $\widetilde{\zeta}_n(\gamma) := n^{-1}\sum_{i=1}^{n}(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_\tau(\gamma,\widetilde{\theta}_n)\} - \tau)S'_i$. 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 t-th column blocks being $\kappa_{\tau_j\tau_t}(\cdot,\cdot)$. Likewise, $\widetilde{\zeta}_n(\cdot)$ turns out to be consistent for $\zeta(\cdot) := \mathbb{E}[(\mathbb{1}\{\widetilde{G}_i(\cdot,\pi^*) \le \rho_\tau(\cdot,\theta^0)\} - \tau)S'_i]$ 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

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

$$\widetilde{B}^\sharp_n := \int_\gamma\int_{\bar{\gamma}} \nabla_\theta\rho_\tau(\gamma,\widetilde{\theta}_n)\widetilde{\kappa}_n(\gamma,\bar{\gamma})\nabla'_\theta\rho_\tau(\bar{\gamma},\widetilde{\theta}_n)d\mathbb{Q}(\gamma)d\mathbb{Q}(\bar{\gamma}) - \int_\gamma \nabla_\theta\rho_\tau(\gamma,\widetilde{\theta}_n)\widetilde{\zeta}_n(\gamma)d\mathbb{Q}(\gamma)\widehat{P}_n^{-1}\widehat{K}'_n$$
$$-\widehat{K}_n\widehat{P}_n^{-1}\int_\gamma \widetilde{\zeta}_n(\gamma)'\nabla'_\theta\rho_\tau(\gamma,\widetilde{\theta}_n)d\mathbb{Q}(\gamma) + \widehat{K}_n\widehat{P}_n^{-1}\widehat{H}_n\widehat{P}_n^{-1}\widehat{K}'_n.$$

$\widehat{B}^\sharp_n$ and $\widetilde{B}^\sharp_n$ are $c\times c$ matrices with (j,t) block submatrices that estimate $B^0_{\tau_j\tau_t}$ and $\widetilde{B}^0_{\tau_j\tau_t}$, for j and $t = 1,2,\ldots,p$. It immediately follows that $\widehat{B}^\sharp_n \overset{\mathbb{P}}{\to} B^0$ under Assumptions 1, 8, S.5, and S.6 by applying Theorem 2. Similarly $\widetilde{B}^\sharp_n \overset{\mathbb{P}}{\to} \widetilde{B}^0$ under Assumptions 1, 8, S.5, and S.8. The result is given in the following Corollary whose proof is almost identical to that of Theorem S.4 and is omitted.

Corollary S.3. *(i) Given Assumption 1, 8, and S.5, (i.a) if $\mathcal{M}$ is misspecified and Assumption S.6 holds, $\widehat{B}_n \overset{\mathbb{P}}{\to} B^*$; (i.b) if $\mathcal{M}$ is correctly specified and Assumption S.8 holds, $\widehat{B}^\sharp_n \overset{\mathbb{P}}{\to} B^0$;*

(ii) Given Assumption 2, 3, 5, 7, 8, and 9, if $\mathcal{M}_\tau$ is correctly specified and Assumption S.7 holds, $\widetilde{B}^\sharp_n \overset{\mathbb{P}}{\to} \widetilde{B}^0$. □

C Joint Inference of Multiple Quantile Functions

In this section, we focus on the quasi-MFQR estimator and discuss testing the hypotheses given in (5) in parallel to Section 5.

The constrained quasi-2MFQR estimator can be adjusted in the context of the absent nuisance effects. Given that we form the objective function $q_n(\cdot)$ by integral transformation, we first define the following constrained estimator of the parameters

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

We call $\bar{\theta}_n$ the *constrained quasi-multiple functional quantile regression (constrained quasi-MFQR)* estimator. This estimator involves constrained Lagrangian estimation, analogous to constrained ML estimation.

The limit distribution of the constrained quasi-MFQR estimator is given in the following lemma.

Lemma S.4. *Given Assumptions 1, 8, 11, 12, and S.5,* $\sqrt{n}(\bar{\theta}_n - \theta^*) + \sqrt{n}(\Omega A^*)^{-1} D^{*\prime}[D^*(\Omega A^*)^{-1} D^{*\prime}]^{-1} R(\theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, \{I + (\Omega A^*)^{-1} D^{*\prime}[D^*(\Omega A^*)^{-1} D^{*\prime}]^{-1} D^*\}\, C^* \{I + D^{*\prime}[D^*(\Omega A^*)^{-1} D^{*\prime}]^{-1} D^*(\Omega A^*)^{-1}\})$, *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* $C^* := A^{*-1} B^* A^{*-1}$. □

Under $\mathbb{H}_o$, $\sqrt{n}(\bar{\theta}_n - \theta^*)$ is asymptotically distributed and centred at zero, whereas it is not bounded under $\mathbb{H}_a$. It is therefore useful in forming the tests discussed below. Note that the constrained quasi-MFQR limit distribution is 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 test uses the unconstrained quasi-MFQR estimator giving

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

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 $\widehat{A}_n := \text{diag}[\widehat{A}_{\tau_1 n}, \widehat{A}_{\tau_2 n}, \ldots, \widehat{A}_{\tau_p n}]$ and for each $j = 1, 2, \ldots, p$,

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

Under Assumption 7 and given consistency of $\widehat{\theta}_n$ for θ^*, $\widehat{A}_n$ is consistent for A^*. The Wald tests assess the magnitudes of $R(\widehat{\theta}_n)$ in a suitable metric and unless $R(\theta^*) = 0$, the test is not bounded in probability.

The LM test is constructed as

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

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

$\bar{A}_n^{-1}\bar{B}_n\bar{A}_n^{-1}$, $\bar{B}_n := n^{-1}\sum_{i=1}^n \bar{J}_{ni}\bar{J}_{ni}'$ such that for each $j = 1, 2, \ldots, p$,

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

Further, define $\bar{A}_n := \text{diag}[\bar{A}_{\tau_1 n}, \bar{A}_{\tau_2 n}, \ldots, \bar{A}_{\tau_p n}]$, where for each $j = 1, 2, \ldots, p$,

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

We construct the LR test by estimating the quantile functions under both hypotheses. Under the null $\mathbb{H}_o$, $\bar{\theta}_n$ and $\widehat{\theta}_n$ both converge to θ^*, so that the distance between $q_n(\bar{\theta}_n)$ and $q_n(\widehat{\theta}_n)$ converges to zero. But $\bar{\theta}_n$ does not converge to θ^* under $\mathbb{H}_a$, so that the distance between $q_n(\bar{\theta}_n)$ and $q_n(\widehat{\theta}_n)$ is non zero asymptotically. These distances then form the basis of the following QLR test

$$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n := 2n\{q_n(\bar{\theta}_n) - q_n(\widehat{\theta}_n)\}.$$

The QLR statistic is nonnegative because $\bar{\theta}_n$ minimizes the objective function subject to the restrictions $R(\theta) = 0$, whereas both $\widehat{\theta}_n$ minimizes 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.

Corollary S.4. *For any sequence c_n such that $c_n = o(n)$, if Assumptions 1, 7, 8, 11, 12, and S.5 hold, (i) $\bar{\mathcal{W}}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty}\mathbb{P}(\bar{\mathcal{W}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; (ii) $\mathcal{L}\bar{\mathcal{M}}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2$ under $\mathbb{H}_o$ and $\lim_{n\to\infty}\mathbb{P}(\mathcal{L}\bar{\mathcal{M}}_n \geq c_n) = 1$ under $\mathbb{H}_a$; and (iii) $\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n \overset{\text{A}}{\sim} W'(D^*(\Omega A^*)^{-1}D^{*\prime})^{-1}W$ 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 $W \sim \mathcal{N}(0, D^*C^*D^{*\prime})$.* □

According to Corollary S.4, the Wald, LM, and QLR statistics are each bounded in probability under $\mathbb{H}_o$ but unbounded under $\mathbb{H}_a$. As Corollary S.4 follows from Theorem 4, we do not separately prove Corollary S.4.

D Proofs

In this section, we provide the proofs of the main and supplementary claims.

D.1 Proofs of the Main Claims

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: The derivation of (3) is almost identical to that of (S.1) and for brevity is not repeated. Instead, we focus on deriving the limit distribution from (3).

If we apply (3) 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 (S.5), as given in the proof of Lemma S.2, to the misspecified model giving, for each $\gamma \in \Gamma$,

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

Here, we applied the ULLN to obtain $n^{-1} \sum_{i=1}^{n} \nabla_{\pi} \widetilde{G}_i(\cdot, \bar{\pi}_{\gamma n i}) \stackrel{\mathbb{P}}{\to} \mathbb{E}[\nabla_{\pi} \widetilde{G}_i(\cdot, \pi^{*})]$ by using Assumption 5. It now follows that

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

Here, Assumptions 2 and 5 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 $\int_{\gamma} \nabla_{\theta_{\tau}} \rho_{\tau}(\gamma, \theta_{\tau}^{*})(\mathbb{1}\{\widetilde{G}_i(\gamma, \pi^{*}) \leq \rho_{\tau}(\gamma, \theta_{\tau}^{*})\} - \tau)$ 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 3 and 4 imply that $\mathbb{E}[J_{\tau i}] = 0$ and $\mathbb{E}[S_i] = 0$. Furthermore, Assumption 6 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 $j = 1, 2, \ldots, c_{\tau}$, $\mathbb{E}[J_{\tau i j}^2] < \infty$ and $\mathbb{E}[S_{ij}^2] < \infty$ by Assumptions 3 and 5. It now follows by the multivariate CLT that $\sqrt{n}(\widetilde{\theta}_{\tau n} - \theta_{\tau}^{*}) \stackrel{\text{A}}{\sim} \mathcal{N}(0, \widetilde{C}_{\tau}^{*})$, as required. ■

Proof of Corollary 1: The desired result follows by applying Theorem 1 to $\rho_{\tau}(\cdot, \widetilde{\theta}_{\tau n})$ using Delta method. ■

Proof of Theorem 2: The consistency of $\widetilde{B}_{\tau n}$ is not detailed because follows in a similar fashion to the consistency of $\widehat{B}_{\tau n}$ given in Theorem S.4 (i). In particular, given the moment conditions in Assumption 5 and the condition for the other consistent estimators for P^{*}, H^{*}, and K_{τ}^{*} as given in Assumption 7, we have $\widehat{J}_{\tau n i} = \widehat{J}_{\tau i} + o_{\mathbb{P}}(1)$, from which the result $\widetilde{B}_{\tau n} \stackrel{\mathbb{P}}{\to} \widetilde{B}_{\tau}^{*}$ follows. ■

Proof of Theorem 3: Theorem 1 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, $\widetilde{B}^*$ is positive definite by Assumption 10. 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 Corollary 2: The desired result follows by applying Theorem 3 to $\rho_\tau(\cdot, \widetilde{\theta}_n)$ using Delta method. ■

Proof of Corollary 3: This follows as a corollary of Theorem 2. ■

Proof of Lemma 2: 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 constrained 2FQR estimator (see also Newey and McFadden, 1994, section 7). The asymptotic first-order conditions are

$$\Omega \ddot{Q}_n + \ddot{D}_n' \ddot{\lambda}_n = o_{\mathbb{P}}(1) \quad \text{and} \quad R(\ddot{\theta}_n) \equiv 0, \tag{S.5}$$

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

$$\Omega \ddot{Q}_n = \Omega \widehat{Q}_n + \Omega A^*(\ddot{\theta}_n - \theta^*) + o_{\mathbb{P}}(1) \quad \text{and} \quad R(\ddot{\theta}_n) = R(\theta^*) + D^*(\widetilde{\theta}_n)(\ddot{\theta}_n - \theta^*) + o_{\mathbb{P}}(1), \tag{S.6}$$

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

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

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 3. Next,

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

Here, we note that Ω and A^* are block diagonal matrices, so that the 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 4: (i) 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 3 (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) Solving for $\ddot{\lambda}_n$ from (S.5) and (S.6), $\sqrt{n}\ddot{\lambda}_n = -(E^{*-1}D^*(\Omega A^*)^{-1})\sqrt{n}\Omega\widehat{Q}_n - E^{*-1}\sqrt{n}R(\theta^*)$ $+o_{\mathbb{P}}(\sqrt{n})$. Given that $\sqrt{n}\Omega\widehat{Q}_n \overset{\text{A}}{\sim} \mathcal{N}(0, \Omega\widetilde{B}^*\Omega)$, it follows that

$$\sqrt{n}\ddot{\lambda}_n + E^{*-1}\sqrt{n}R(\theta^*) \overset{\text{A}}{\sim} \mathcal{N}(0, E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}), \tag{S.7}$$

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

$$n\ddot{\lambda}_n'\{E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}\}^{-1}\ddot{\lambda}_n \overset{\text{A}}{\sim} \mathcal{X}_r^2. \tag{S.8}$$

Using the fact that $E^* := -D^*(\Omega A^*)^{-1}D^{*\prime}$ we have $\{E^{*-1}D^*\widetilde{C}^*D^{*\prime}E^{*-1}\}^{-1} = E^*(D^*\widetilde{C}^*D^{*\prime})^{-1}E^* =$ $D^*(\Omega A^*)^{-1}D^{*\prime}(D^*\widetilde{C}^*D^{*\prime})^{-1}D^*\ (\Omega A^*)^{-1}\ D^{*\prime}$. 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 2 and the consistency of $\widetilde{A}_n$ for A^*. The last equality follows from (S.5) 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 (S.8).

Under $\mathbb{H}_a$, note that $\Omega\ddot{Q}_n + \ddot{D}_n'\ddot{\lambda}_n = o_{\mathbb{P}}(1)$ from (S.5) and $\sqrt{n}\ddot{\lambda}_n = O_{\mathbb{P}}(\sqrt{n})$ from (S.7), 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 12, 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 2 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 3. 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,

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

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. ■

D.2 Proofs of the Supplementary Claims

Proof of Theorem S.1: The proof follows reasoning similar to that of Oberhofer and Haupt (2016). Applying Lemma 2N of Oberhofer and Haupt (2016), we first obtain that for w such that $\|w\| = 1$,

$$\widehat{R}_{ln}(w) \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w'\int_{\gamma}\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})(\mathbb{1}\{G_i(\gamma) \leq \rho_\tau(\gamma,\widehat{\theta}_{\tau n})\} - \tau)d\mathbb{Q}(\gamma) \leq \widehat{R}_{un}(w),$$

where

$$\widehat{R}_{ln}(w) := -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\gamma}\mathbb{1}\{G_i(\gamma) = \rho_\tau(\gamma,\widehat{\theta}_{\tau n})\}|w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})|\mathbb{1}\{w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n}) < 0\}d\mathbb{Q}(\gamma),$$

$$\widehat{R}_{un}(w) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\gamma}\mathbb{1}\{G_i(\gamma) = \rho_\tau(\gamma,\widehat{\theta}_{\tau n})\}|w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})|\mathbb{1}\{w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n}) \geq 0\}d\mathbb{Q}(\gamma).$$

Furthermore, applying Lemma 9N of Oberhofer and Haupt (2016) demonstrates that

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

so that if we show that $\widehat{R}_{ln}(w) = o_{\mathbb{P}}(1)$ and $\widehat{R}_{un}(w) = o_{\mathbb{P}}(1)$, then (3) follows. For this derivation, we let

$$\widehat{R}_n(w) := \widehat{R}_{ln}(w) - \widehat{R}_{un}(w) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\gamma}\mathbb{1}\{G_i(\gamma) = \rho_\tau(\gamma,\widehat{\theta}_{\tau n})\}|w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})|d\mathbb{Q}(\gamma)$$

and show that $\widehat{R}_n(w) = o_{\mathbb{P}}(1)$ by noting that $\widehat{R}_n(w) = o_{\mathbb{P}}(1)$ if and only if $\widehat{R}_{ln}(w) = o_{\mathbb{P}}(1)$ and $\widehat{R}_{un}(w) = o_{\mathbb{P}}(1)$. If we let $B(\theta_{\tau 0}, d) := \{\theta_\tau : \|\theta_\tau - \theta_{\tau 0}\| \leq d\}$, then for a sufficiently large $m <$

∞, there are finite numbers of open balls covering $\Theta_\tau(n,m) := \{\theta_\tau : \sqrt{n}\|\theta_\tau - \theta_\tau^*\| \leq m\}$, viz., $\Theta_\tau(n,m) \subset \cup_{j=1}^{n(d)} B(\theta_\tau(j,d),d)$ such that for any $d > 0$, $n(d) < \infty$, where $\theta_\tau(j,d)$ is the center of the j-th open ball. Given this, Assumption 4 implies that for any $\theta_\tau \in B(\theta_\tau(j,d),d)$, there exists $\bar{\theta}_\tau$, θ_τ^a, and $\theta_\tau^b \in B(\theta_\tau(j,d),d)$ such that

$$\int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau)|^2 d\mathbb{Q}(\gamma) \leq \int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma)$$

and

$$\int_\gamma \mathbb{1}\{G_i(\gamma) = \rho_\tau(\gamma,\theta_\tau)\} d\mathbb{Q}(\gamma) \leq \int_\gamma \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^a) \leq G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^b)\} d\mathbb{Q}(\gamma).$$

If we further let

$$\bar{R}_n(j,r) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_\gamma \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^a) \leq G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^b)\} d\mathbb{Q}(\gamma) \int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma),$$

then it follows that $0 \leq \widehat{R}_n(w) \leq \bar{R}_n(j,r)$ by noting that for any $\theta_\tau \in B(\theta_\tau(j,d),d)$,

$$\begin{aligned}&\int_\gamma \mathbb{1}\{G_i(\gamma) = \rho_\tau(\gamma,\theta_\tau)\}|w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau)| d\mathbb{Q}(\gamma)\\ &\quad\leq \int_\gamma \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^a) \leq G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^b)\} d\mathbb{Q}(\gamma) \int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma).\end{aligned}$$

We here note that for each $j = 1, 2, \ldots, n(d)$,

$$\begin{aligned}&\mathbb{E}[\int_\gamma \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^a) \leq G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^b)\} d\mathbb{Q}(\gamma) \int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma)]\\ &\quad= \int_\gamma \{F_\gamma(\rho_\tau(\gamma,\theta_\tau^b)) - F_\gamma(\rho_\tau(\gamma,\theta_\tau^a))\} d\mathbb{Q}(\gamma) \int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma),\end{aligned}$$

and

$$\text{var}[\bar{R}_n(j,r)] \leq \int_\gamma \{F_\gamma(\rho_\tau(\gamma,\theta_\tau^b)) - F_\gamma(\rho_\tau(\gamma,\theta_\tau^a))\} d\mathbb{Q}(\gamma) \left(\int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma)\right)^2.$$

Further, for some $\widetilde{\theta}_\tau$ between θ_τ^a and θ_τ^b, we have

$$F_\gamma(\rho_\tau(\gamma,\theta_\tau^b)) - F_\gamma(\rho_\tau(\gamma,\theta_\tau^a)) = f_\gamma(\rho_\tau(\gamma,\widetilde{\theta}_\tau))\sqrt{n}(\theta_\tau^b - \theta_\tau^a)\frac{1}{\sqrt{n}}\nabla_{\theta_\tau}\rho_\tau(\gamma,\widetilde{\theta}_\tau) = o(1)$$

because $f_{(\cdot)}(\cdot)$ and $\nabla_{\theta_\tau}\rho_\tau(\cdot,\cdot)$ are uniformly bounded on $\Gamma \times \Theta_\tau$ by Assumption 4 and $\sqrt{n}(\theta_\tau^b - \theta_\tau^a) \leq 2d$. It therefore follows that $\bar{R}_n(j,r) = o_{\mathbb{P}}(1)$, so that

$$\mathbb{E}[\int_\gamma \mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^a) \leq G_i(\gamma) \leq \rho_\tau(\gamma,\theta_\tau^b)\} d\mathbb{Q}(\gamma) \int_\gamma |w'\nabla_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_\tau)|^2 d\mathbb{Q}(\gamma)] = o(1),$$

and $\text{var}[\bar{R}_n(j,r)] = o(1)$, leading to $\widehat{R}_n(w) = o_{\mathbb{P}}(1)$, and this again leads to (3).

Second, note that (3) implies that $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta^*_\tau) = -A^{*-1}_\tau n^{-1/2}\sum_{i=1}^n J_{\tau i} + o_{\mathbb{P}}(1)$. Given that θ^*_τ is identified as given in Assumption 4, the first-order condition holds, so that $\mathbb{E}[J_{\tau i}] = 0$. Assumption S.2 also implies that $B^*_\tau := \mathbb{E}[J_{\tau i}J'_{\tau i}]$ is positive definite. Furthermore, Assumption S.1 implies that for each $j = 1, 2, \ldots, c_\tau$, $\mathbb{E}[J^2_{\tau ij}] < \infty$, 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 $\sqrt{n}(\widehat{\theta}_{\tau n} - \theta^*_\tau) \overset{\text{A}}{\sim} \mathcal{N}(0, A^{*-1}_\tau B^*_\tau A^{*-1}_\tau)$. This completes the proof. ■

Proof of Lemma S.1: We show stochastic equicontinuity of $n^{-1/2}\sum_{i=1}^n(\mathbb{1}\{G_i(\cdot) \le \rho_\tau(\cdot, \theta^0_\tau)\} - \tau)$ using Ossiander's L^2 entropy condition: for some $\nu > 0$ and $C > 0$,

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

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

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

by noting that $F_\gamma(\rho_\tau(\gamma, \theta^0_\tau)) = \tau$ and $F_\gamma(\rho_\tau(\bar{\gamma}, \theta^0_\tau)) = \tau$. 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 Assumptions 1 and 4 imply that $U_i(\cdot)$ is Lipschitz continuous almost surely: for some $C > 0$, $|U_i(\gamma) - U_i(\bar{\gamma})| \le C\|\gamma - \bar{\gamma}\|$. 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 $\bar{\gamma}$,

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

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

Proof of Theorem S.2: Given Lemma S.1, we note by continuous mapping that

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

which follows a normal distribution since $\mathcal{G}_\tau(\cdot)$ is a Gaussian stochastic process. Further note that applying dominated convergence using Assumption S.1 gives

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

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

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 S.3. 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 Lemma S.2: 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 $+\nabla'_\pi\widetilde{G}_i(\gamma,\bar{\pi}_{\gamma ni})(\widehat{\pi}_n-\pi^*)$, respectively. If we further let x, c, and a be $\widehat{G}_i(\gamma)$, $x_\tau(\gamma)$, and $\widehat{\mu}_{ni}(\gamma)$, respectively, it now follows that $\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)+\nabla'_\pi\widetilde{G}_i(\gamma,\bar{\pi}_{\gamma ni})(\widehat{\pi}_n-\pi^*) \le \rho_\tau(\gamma,\theta_\tau^0)\} = \mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\} + \mathbb{1}\{x_\tau(\gamma) < \widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)-\widehat{\mu}_{ni}(\gamma)\} - \mathbb{1}\{x_\tau(\gamma)-\widehat{\mu}_{ni}(\gamma) < \widetilde{G}_i(\gamma,\pi^*) \le x_\tau(\gamma)\}$. Note that Assumption 5 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,

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

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

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)+\nabla_\pi'\widetilde{G}_i(\gamma,\bar{\pi}_{\gamma ni})(\widehat{\pi}_n-\pi^*)\le\rho_\tau(\gamma,\theta_\tau^0)\}-\tau\Big)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)\le x_\tau(\gamma)\}-\tau\Big)-f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla_\pi'\widetilde{G}_i(\gamma,\pi^*)]\sqrt{n}(\widehat{\pi}_n-\pi^*)+o_{\mathbb{P}}(1)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\Big(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)\le x_\tau(\gamma)\}-\tau\Big)-f_\gamma(x_\tau(\gamma))\mathbb{E}[\nabla_\pi'\widetilde{G}_i(\gamma,\pi^*)]P^{*-1}S_i\Big]+o_{\mathbb{P}}(1). \qquad \text{(S.9)}
\end{aligned}$$

Given this expression, we compute the covariance kernel using the summand on the right side of (S.9), viz.,

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

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

We next prove that the left side of (S.9) 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 (S.9) satisfies the bound condition to apply Ossiander's L^2 entropy condition: for some C and $\nu>0$,

$$\mathbb{E}\left(\sup_{\|\gamma-\bar{\gamma}\|<\delta}|(\mathbb{1}\{\widetilde{G}_i(\gamma,\pi^*)\le x_\tau(\gamma)\}-\varsigma(\gamma)S_i)-(\mathbb{1}\{\widetilde{G}_i(\bar{\gamma},\pi^*)\le x_\tau(\bar{\gamma})\}-\varsigma(\bar{\gamma})S_i)|^2\right)^{1/2}\le C\delta^\nu. \qquad \text{(S.10)}$$

We here note that

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

In the proof of Lemma S.1, we already saw that there are C_1 and $\nu_1 > 0$ such that

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

Next, Assumptions 4, 2, and 5 imply that $\varsigma(\cdot)$ is Lipschitz continuous, because the product of two Lipschitz continuous functions is Lipschitz continuous: for some $m > 0$, $\|\varsigma(\gamma) - \varsigma(\bar{\gamma})\| \leq m\|\gamma - \bar{\gamma}\|$, so that for some C_2 and $\nu_2 > 0$,

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

by letting $C_2 := ms \cdot \max_{j=1,\ldots,s} \mathbb{E}[|S_{ij}|^2]$ and $\nu_2 = 1$. Note that $\max_{j=1,\ldots,s} \mathbb{E}[|S_{ij}|^2] < \infty$ from Assumption 3. We note that

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

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

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

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 (S.10) follows. This shows that

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

is stochastically equicontinuous, completing the proof. ■

Proof of Theorem S.3: Given Lemma S.2, we note that

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

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. Then, by dominated convergence using Assumption 5,

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

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

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

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

giving the desired result. ■

Proof of Theorem S.4: (i.a) If we apply the mean-value theorem to $\rho_\tau(\gamma,\widehat{\theta}_{\tau n})$ and $\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n})$ around the unknown parameter θ_τ^*, for each γ, there are $\bar{\theta}_{\tau\gamma}^*$ and $\acute{\theta}_{\tau\gamma}^*$ such that

$$\rho_\tau(\gamma,\widehat{\theta}_{\tau n}) = \rho_\tau(\gamma,\theta_\tau^*) + \nabla'_{\theta_\tau}\rho_\tau(\gamma,\bar{\theta}_{\tau\gamma}^*)(\widehat{\theta}_{\tau n}-\theta_\tau^*), \quad \text{and}$$
$$\nabla_{\theta_\tau}\rho_\tau(\gamma,\widehat{\theta}_{\tau n}) = \nabla_{\theta_\tau}\rho_\tau(\gamma,\theta_\tau^*) + \nabla_{\theta_\tau}^2\rho_\tau(\gamma,\acute{\theta}_{\tau\gamma}^*)(\widehat{\theta}_{\tau n}-\theta_\tau^*).$$

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

$$\begin{aligned}\mathbb{1}\{G_i(\gamma)\le\rho_\tau(\gamma,\widehat{\theta}_{\tau n})\}-\tau &= \mathbb{1}\{G_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau\\ &+\mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^*)<G_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^*)-\widehat{\nu}_n(\gamma)\}-\mathbb{1}\{\rho_\tau(\gamma,\theta_\tau^*)-\widehat{\nu}_n(\gamma)<G_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^*)\}\\ &= \mathbb{1}\{G_i(\gamma)\le\rho_\tau(\gamma,\theta_\tau^*)\}-\tau+o_{\mathbb{P}}(1), \end{aligned} \tag{S.11}$$

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

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

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

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

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

(i.b) Given the second-order differentiability of $\rho_\tau(\gamma,\cdot)$ in Assumption 4 and Theorem 1 (i), we apply (S.11) to obtain

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

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

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

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

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

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

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

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

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

$$\mathbb{E}\left[\sup_{\|(\gamma,\gamma')-(\gamma'',\gamma''')\|<\delta}\left|\mathbb{1}\{U_i(\gamma) \le \tau\}\mathbb{1}\{U_i(\gamma') \le \tau\} - \mathbb{1}\{U_i(\gamma'') \le \tau\}\mathbb{1}\{U_i(\gamma''') \le \tau\}\right|^2\right] \le C^0\delta^\nu. \tag{S.12}$$

We first note that

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

Therefore,

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

Next observe that

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

by applying Cauchy-Schwarz. Note that

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

It follows that

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

which implies

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

We have already seen in the proof of Lemma S.1 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 (S.12) 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.

(ii) Proof of consistency of $\widetilde{B}_{\tau n}^{\sharp}$ is not detailed because it follows in a similar fashion to the consistency of $\widehat{B}_{\tau n}^{\sharp}$. In Given the moment conditions in Assumption 5 and the condition for the other consistent estimators for P^*, H^*, and K_τ^* in Assumption 7, we have $\widehat{J}_{\tau ni}=\widehat{J}_{\tau i}+o_{\mathbb{P}}(1)$, from which the result $\widetilde{B}_{\tau n}^{\sharp}\overset{\mathbb{P}}{\to}\widetilde{B}_\tau^0$ follows. ■

Proof of Corollary S.1: The desired proof follows from the proof of Theorem 3 by letting $S_i\equiv 0$. ■

Proof of Lemma S.3: To show the claim, for each $j=1,2,\ldots,p$, for notational simplicity we first set

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

Second, note that Lemma S.2 implies that for any $\epsilon_j>0$ and $\eta_j>0$, there exist n_{0j} and $\delta_j>0$ such that if $n>n_{0j}$,

$$\mathbb{P}\left(\sup_{\|\gamma-\bar{\gamma}\|<\delta_j}|\widehat{\omega}_{nj}(\gamma)-\widehat{\omega}_{nj}(\bar{\gamma})|>\epsilon_j\right)<\eta_j. \tag{S.13}$$

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 $n > n_0$,

$$\mathbb{P}\left(\sup_{\|\gamma-\bar{\gamma}\|<\delta}\left|\sum_{j=1}^{p}\lambda_j(\widehat{\omega}_{nj}(\gamma)-\tau_j)-\sum_{j=1}^{p}\lambda_j(\widehat{\omega}_{nj}(\bar{\gamma})-\tau_j)\right|>\epsilon\right)<\eta, \tag{S.14}$$

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 (S.14), we let $\epsilon > 0$ and $\eta > 0$ and show 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 $n > n_0 := \max[n_{01}, n_{02}, \ldots, n_{0p}]$,

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

from (S.13). On the other hand, if $\lambda_j = 0$,

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

Therefore,

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

and we also note that

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

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

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

This completes the proof. ■

Proof of Theorem S.5: Given Lemma S.3, note that

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

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 9,

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

defining $\widetilde{\kappa}(\cdot,\cdot) : \Gamma \times \Gamma \mapsto \mathbb{R}^{p\times p}$ such that its j-th row and t-th column element is $\widetilde{\kappa}_{\tau_j,\tau_t}(\cdot,\cdot)$ given in Lemma S.3,

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

which yields $\widetilde{B}^0$. Therefore, $\int_\gamma \nabla_\theta \rho_\tau(\gamma, \theta^0)\widetilde{\mathcal{G}}(\gamma) d\mathbb{Q}(\gamma) \sim \mathcal{N}(0, \widetilde{B}^0)$. Given that $\widetilde{B}^0$ is positive definite by Assumption S.7, it follows that

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

as desired, completing the proof. ■

Proof of Corollary S.2: The desired proof follows from the proof of Theorem S.5 by letting $S_i \equiv 0$. ■

Proof of Corollary S.3: The proof is parallel to that of Theorem S.4. ■

Proof of Lemma S.3: The proof is parallel to that of Lemma 2. ■

Proof of Corollary S.4: The proof is parallel to that of Theorem 4. ■

E Quasi-2FQR Estimation with Covariates

In this section, we examine the quasi-2FQR estimation when the model is driven by covariates. We suppose that the covariates exist as a set of random variables $X_i \in \mathbb{R}^k$ ($k \in \mathbb{R}^k$). Our goal is to estimate the conditional quantile function of $\widetilde{G}_i(\cdot)$ on $X_i = x$ consistently. For this goal, we suppose that the following parametric model is specified:

$$\mathcal{S}_\tau := \{\rho_\tau(\cdot, \theta_\tau, x) : \Gamma \mapsto \mathbb{R} | \theta_\tau \in \Theta_\tau \in \mathbb{R}^{c_\tau}\},$$

where $c_\tau \in \mathbb{N}$. Note that $\mathcal{S}_\tau$ corresponds to $\mathcal{M}_\tau$. The only difference is in the fact that $\mathcal{S}_\tau$ is defined through $\rho_\tau(\cdot, \theta_\tau, x)$, which is a function of the conditioning variable X_i. From now, for notation simplicity, we let

$\rho_{\tau i}(\cdot, \theta_\tau) := \rho_\tau(\cdot, \theta_\tau, X_i)$, and we estimate the unknown parameter by minimizing the following function: for each $\theta_\tau \in \Theta_\tau$,

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

producing $\dot{\theta}_{\tau n} := \arg\min_{\theta_\tau \in \Theta_\tau} \dot{q}_{\tau n}(\theta_\tau)$.

Assumption S.9. *(i) Let $(\mathbf{\Omega}, \mathcal{F}, \mathbb{P})$ 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) $\{(\widetilde{G}_i(\cdot), X_i')' : \widetilde{G}_i(\cdot) : \mathbf{\Omega} \times \Gamma \times \Pi \mapsto \mathbb{R} \text{ and } X_i \in \mathbb{R}^k\}_{i=1}^n$ $(k \in \mathbb{N})$ is a set of iid observations such that (ii.a) for each $(\gamma, \pi) \in \Gamma \times \Pi$, $(\widetilde{G}_i(\gamma, \pi), X_i')'$ 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 $(\mathbf{\Omega} \times \Gamma, \mathcal{F} \otimes \mathcal{G}, \mathbb{P} \cdot \mathbb{Q})$ 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 $f_\gamma(\cdot)$ is uniformly bounded, where for some π^*, $F_\gamma(\cdot, X_i)$ and $f_\gamma(\cdot, X_i)$ are the conditional CDF and PDF of $\widetilde{G}_i(\gamma, \pi^*)$ on X_i, respectively.* □

Assumption S.10. *(i) For each $\theta_\tau \in \Theta_\tau$, $\rho_{\tau i}(\,\cdot\,, \theta_\tau)$ is $\mathcal{F} \otimes \mathcal{G}$-measurable, where Θ_τ is a compact and convex set in $\mathbb{R}^{c_\tau}$ $(c_\tau \in \mathbb{N})$; (ii) for each $\gamma \in \Gamma$, $\rho_{\tau i}(\gamma, \,\cdot\,) \in \mathcal{C}^{(2)}(\Theta_\tau)$ with probability 1 and for $j, t = 1, 2, \ldots, c_\tau$, $(\partial/\partial\theta_{\tau j})\rho_{\tau i}(\cdot, \cdot) \in \mathcal{C}(\Gamma \times \Theta_\tau)$ and $(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau t})\rho_{\tau i}(\cdot, \cdot) \in \mathcal{C}(\Gamma \times \Theta_\tau)$ with probability 1; (iii) for each $\theta_\tau \in \Theta_\tau$, $\rho_{\tau i}(\,\cdot\,, \theta_\tau) \in \mathcal{L}_{ip}(\Gamma)$ with probability 1; (iv) if we let $\dot{q}_\tau(\theta_\tau) := \int_\gamma \int \xi_\tau\{g(\gamma) - \rho_\tau(\gamma, \theta_\tau, x)\} d\mathbb{P}(g(\gamma), x) d\mathbb{Q}(\gamma)$, $\theta_\tau^\dagger := \arg\min_{\theta_\tau} q_\tau(\theta_\tau)$ is unique and interior to Θ_τ, where $\mathbb{P}(g(\gamma), x)$ is the joint distribution of $\widetilde{G}_i(\gamma, \pi^*)$ and X_i; and (v) $f_{(\cdot)}(\rho_\tau(\cdot, \cdot), X_i) \in \mathcal{C}(\Gamma \times \Theta_\tau)$ with probability 1.* □

Assumption S.11. *For some $M_i \in L^2(\mathbb{P})$, (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) $\sup_{(\gamma,\theta_\tau)} |\rho_{\tau i}(\gamma, \theta_\tau)| \leq M_i$; (iv) for each $j = 1, 2, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau)} |(\partial/\partial\theta_{\tau j})\rho_{\tau i}(\gamma, \theta_\tau)| \leq M_i$; (v) for each $j, t = 1, \ldots, c_\tau$, $\sup_{(\gamma,\theta_\tau)} |(\partial^2/\partial\theta_{\tau j}\partial\theta_{\tau t})\rho_{\tau i}(\gamma, \theta_\tau)| \leq M_i$; and (vi) for each $j = 1, 2, \ldots, s$, $\mathbb{E}[(\partial/\partial\pi_j)\widetilde{G}_i(\cdot, \pi^*)] \in \mathcal{L}_{ip}(\Gamma)$ and for each θ_τ, $f_{(\cdot)}(\rho_\tau(\cdot, \theta_\tau), X_i) \in \mathcal{L}_{ip}(\Gamma)$ with probability 1.* □

For the next regularity condition, we define the following:

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

We further let $J_{\tau i}^\dagger := \int_\gamma \nabla_{\theta_\tau} \rho_{\tau i}(\gamma, \theta_\tau^\dagger) \left(\mathbb{1}\{G_i(\gamma) \leq \rho_{\tau i}(\gamma, \theta_\tau^\dagger)\} - \tau \right) d\mathbb{Q}(\gamma)$ and with $B_\tau^\dagger := \mathbb{E}[J_{\tau i}^\dagger J_{\tau i}^{\dagger\prime}]$. Next, let $\dot{J}_{\tau i}^\dagger := \int_\gamma \nabla_{\theta_\tau} \rho_{\tau i}(\gamma, \theta_\tau^\dagger)(\mathbb{1}\{\widehat{G}_i(\gamma) \leq \rho_{\tau i}(\gamma, \theta_\tau^\dagger)\} - \tau) d\mathbb{Q}(\gamma)$ with its covariance matrix, denoted by $\dot{B}_\tau^\dagger$ such that

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

where $K_\tau^\dagger := \int_\gamma \int_x \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^\dagger, x) f_\gamma(\rho_\tau(\gamma, \theta_\tau^\dagger, x), x) \nabla'_\pi g(\gamma, \pi^*) d\mathbb{P}(g(\gamma), x) d\mathbb{Q}(\gamma)$. Given this, we further impose the following condition.

Assumption S.12. *(i)* $\lambda_{\min}(A_\tau^\dagger) > 0$*; (ii)* $\lambda_{\min}(L_\tau^\dagger) > 0$*; and (iii)* $\lambda_{\min}(\dot{B}_\tau^\dagger) > 0$*, where*

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

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

These assumption extends those in Section 2. Assumptions S.9, S.10, S.11, and S.12 correspond to Assumptions 2, 4, 5, and 6, respectively, and they are obtained by supposing the model condition $\mathcal{S}_\tau$ instead of $\mathcal{M}_\tau$. Using the assumptions we also prove the following corollary corresponding to Theorem 1 as follows.

Corollary S.5. *Given Assumptions 3, S.9, S.10, S.11, and S.12, if* $\mathcal{S}_\tau$ *is misspecified,* $\sqrt{n}(\dot{\theta}_{\tau n} - \theta_\tau^\dagger) \overset{\text{A}}{\sim} \mathcal{N}(0, \dot{C}_\tau^\dagger)$*, where* $\dot{C}_\tau^\dagger := A_\tau^{\dagger-1} \dot{B}_\tau^\dagger A_\tau^{\dagger-1}$. □

Remarks S.1. (a) As the proof of Corollary S.5 is parallel to Theorem 1, we only provide an outline of the proof. We first obtain the following, analogous to (3):

$$\sqrt{n}(\dot{\theta}_{\tau n} - \theta_\tau^\dagger) = -A_{\tau n}^{\dagger-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_\gamma \nabla_{\theta_\tau} \rho_{\tau i}(\gamma, \theta_\tau^\dagger) \left(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_{\tau i}(\gamma, \theta_\tau^\dagger)\} - \tau \right) d\mathbb{Q}(\gamma) + o_{\mathbb{P}}(1),$$

where $\dot{J}_{\tau i}^\dagger := \int_\gamma \nabla_{\theta_\tau} \rho_{\tau i}(\gamma, \theta_\tau^\dagger) \left(\mathbb{1}\{\widehat{G}_i(\gamma) \le \rho_{\tau i}(\gamma, \theta_\tau^\dagger)\} - \tau \right) d\mathbb{Q}(\gamma)$ and

$$A_{\tau n}^\dagger := \int_\gamma \frac{1}{n} \sum_{i=1}^n \nabla_{\theta_\tau} \rho_{\tau i}(\gamma, \theta_\tau^\dagger) f_\gamma(\rho_{\tau i}(\gamma, \theta_\tau^\dagger), X_i) \nabla_{\theta_\tau}' \rho_{\tau i}(\gamma, \theta_\tau^\dagger) d\mathbb{Q}(\gamma).$$

Next, we show that $A_{\tau n}^\dagger \overset{\mathbb{P}}{\to} A_\tau^\dagger$ and $n^{-1/2} \sum_{i=1}^n \dot{J}_{\tau i}^\dagger \overset{\text{A}}{\sim} \mathcal{N}(0, \dot{B}_\tau^\dagger)$, to produce Corollary S.5.

(b) If the nuisance effects do not exist, so that $\widehat{G}_i(\cdot)$ is replaced by $G_i(\cdot)$, the asymptotic distribution of the quasi-FQR estimator is obtained as another corollary of Corollary S.5. That is, $\sqrt{n}(\dot{\theta}_{\tau n} - \theta_\tau^\dagger) \overset{\text{A}}{\sim} \mathcal{N}(0, C_\tau^\dagger)$, where $C_\tau^\dagger := A_\tau^{\dagger-1} B_\tau^\dagger A_\tau^{\dagger-1}$.

(c) If $\mathcal{S}_\tau$ is correctly specified, so that for some $\theta_\tau^\ddagger$, $\rho_\tau(\cdot, \theta_\tau^\ddagger, X_i) = x_\tau(\cdot, X_i)$ with probability 1, where for each $\gamma \in \Gamma$ and $x \in \mathbb{R}^k$, $x_\tau(\gamma, x) := \inf\{z \in \mathbb{R} : F_\gamma(z, x) \ge \tau\}$, where $F_\gamma(\cdot, x)$ is the conditional CDF of $G(\gamma)$ on $X_i = x$, then we can use Corollary S.5 to obtain the limit distribution of the 2FQR estimator: $\sqrt{n}(\dot{\theta}_{\tau n} - \theta_\tau^\ddagger) \overset{\text{A}}{\sim} \mathcal{N}(0, \dot{C}_\tau^\ddagger)$, where $\dot{C}_\tau^\ddagger := A_\tau^{\ddagger-1} \dot{B}_\tau^\ddagger A_\tau^{\ddagger-1}$,

$$A_\tau^\ddagger := \int_\gamma \int_x \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^\ddagger, x) f_\gamma(\rho_\tau(\gamma, \theta_\tau^\ddagger, x), x) \nabla_{\theta_\tau}' \rho_\tau(\gamma, \theta_\tau^\ddagger, x) d\mathbb{P}(x) d\mathbb{Q}(\gamma)$$

and

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

such that $B_\tau^\ddagger := \mathbb{E}[J_{\tau i}^\ddagger J_{\tau i}^{\ddagger\prime}]$, $J_{\tau i}^\ddagger := \int_\gamma \nabla_{\theta_\tau} \rho_{\tau i}(\gamma, \theta_\tau^\ddagger) \left(\mathbb{1}\{G_i(\gamma) \le \rho_{\tau i}(\gamma, \theta_\tau^\ddagger)\} - \tau \right) d\mathbb{Q}(\gamma)$, and $K_\tau^\ddagger := \int_\gamma \int_x \nabla_{\theta_\tau} \rho_\tau(\gamma, \theta_\tau^\ddagger, x) f_\gamma(\rho_\tau(\gamma, \theta_\tau^\ddagger, x), x) \nabla_\pi' g(\gamma, \pi^*) d\mathbb{P}(g(\gamma), x) d\mathbb{Q}(\gamma)$.

(d) In case the nuisance effects do not exist, the limit distribution of FQR estimator is obtained as follows: $\sqrt{n}(\dot{\theta}_{\tau n} - \theta^{\ddagger}_{\tau}) \overset{\text{A}}{\sim} \mathcal{N}(0, C^{\ddagger}_{\tau})$, where $C^{\ddagger}_{\tau} := A^{\ddagger-1}_{\tau} B^{\ddagger}_{\tau} A^{\ddagger-1}_{\tau}$. □

F Additional Simulations

Using the linear specification in Section 6.1 we provide further simulation findings under alternative choices of the adjunct probability measure. In particular, the following five adjunct probability measures were used. First, let $\mathbb{Q}$ be another uniform distribution with a non-zero measure on $[3/4, 1]$ for the same $\Gamma = [1/2, 1]$. Compared to the above adjunct probability measure, this concentrates the probability measure to a narrower interval. Second, let $\mathbb{Q}$ be the uniform distribution on $\Gamma = [0, 1]$, so that the adjunct probability measure is equally distributed over a wider interval than those considered before. Third, let $\Gamma = [0, 1]$ and use the following beta distribution as the adjunct probability measure

$$f(\gamma; \alpha, \beta) = \frac{\gamma^{\alpha}(1-\gamma)^{\beta}}{B(\alpha, \beta)},$$

where $\gamma \in [0, 1]$. We let $(\alpha, \beta) = (2, 2)$. With this selection, the PDF weighs the central point of the interval more than any other points. Fourth, use the same beta distribution but let $\alpha = 1$ and $\beta = 3$. With this setting, the left-end point has the highest PDF level, that monotonically decreases as γ increases. Finally, we let $\alpha = 5$ and $\beta = 1$ for the beta distribution, and this PDF shows a reverse shape from the fourth setting. Figure S.1 illustrates the shapes of the last three beta distributions.

The simulation results are reported in Tables S.1 – S.5 in the format of Table 1. Rejection rates of the Wald, LM, and QLR tests are computed under the null, fixed alternative, and local alternative hypotheses. The findings from this simulation reveal that the rejection rates are insensitive to the adjunct probability measures. The figures in Tables S.1 – S.5 are more or less similar. This similarity may not hold for any DGP or model. It applies because functional observations are linear functions of γ, as stated in Section 6.

Further simulations were conducted to corroborate the limit theory findings of the paper. For example, when the functional observation is nonlinear, the asymptotic distribution of the FQR estimator depends on the selection of the adjunct probability measure, reflecting Theorem S.2. This fact also implies that estimating the unknown parameter using a parameterized adjunct probability measure can deliver a more efficient FQR estimator. For brevity these results are not reported here.

G Supplementary Empirical Applications

This section provides additional empirical material for Section 7. 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 S.2 plots the estimated LIPs using work experiences over 0–40 years, and Figure S.3 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 S.2 and S.3. In particular, the curves in Figure S.3 generally have less curvature and are closer to linearity than those of Figure S.2 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 S.6 and S.7 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.

Size of the Tests

Statistics	Levels\n	50	100	200	300	400	500
	10%	13.88	13.02	11.18	11.16	10.84	10.78
$\bar{\mathcal{W}}_n$	5%	8.16	7.70	6.30	5.52	5.66	5.70
	1%	2.58	2.40	1.86	1.24	1.58	1.36
	10%	11.48	9.90	10.14	9.68	10.34	9.70
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	5.22	5.02	5.20	4.94	5.28	4.62
	1%	1.20	1.08	1.14	0.90	1.10	0.88
	10%	10.72	10.32	9.84	10.40	9.88	10.18
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	5.52	5.42	4.76	4.94	5.56	4.82
	1%	1.30	1.20	1.20	0.92	1.16	1.02

Power of the Tests

Statistics	Levels\n	20	40	60	80	100	120
	10%	81.36	97.60	99.76	100.0	100.0	100.0
$\bar{\mathcal{W}}_n$	5%	73.06	95.08	99.32	99.96	99.96	100.0
	1%	55.84	87.54	96.86	99.46	99.82	99.98
	10%	74.06	96.94	99.82	100.0	100.0	100.0
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	66.02	94.10	99.52	99.90	99.98	100.0
	1%	35.20	80.54	96.18	99.38	99.92	99.98
	10%	86.60	98.60	99.94	100.0	100.0	100.0
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	78.30	97.38	99.82	100.0	100.0	100.0
	1%	57.18	90.30	98.52	99.86	99.96	100.0

Local Power of the Tests

Statistics	Levels\n	50	100	200	300	400	500
	10%	84.72	81.80	81.56	79.10	80.06	78.92
$\bar{\mathcal{W}}_n$	5%	77.44	73.56	71.60	69.90	70.34	68.90
	1%	60.50	54.86	50.34	48.50	47.82	46.86
	10%	83.08	80.52	80.04	78.06	79.00	77.38
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	72.34	69.24	69.20	68.66	69.22	67.40
	1%	48.66	45.84	44.66	44.72	45.98	43.30
	10%	89.70	87.18	87.20	85.12	86.38	85.44
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	82.82	79.60	79.34	76.90	78.02	76.74
	1%	64.10	60.34	57.56	56.06	56.66	55.28

Table S.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 $\gamma \in \Gamma := [1/2, 1]$ 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 $\rho_{\tau_j}(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ 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 $\theta^*_{\tau_2} \neq 1/6$. The adjunct probability measure we employ here is the uniform distribution on $[3/4, 1]$.

Size of the Tests							
Statistics	Levels\n	50	100	200	300	400	500
$\bar{\mathcal{W}}_n$	10%	13.62	13.18	11.30	11.14	10.84	10.82
	5%	8.10	7.86	6.38	5.56	5.76	5.72
	1%	2.56	2.44	1.86	1.24	1.62	1.32
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	9.78	9.82	9.94	9.70	10.32	9.58
	5%	4.94	4.94	5.04	4.86	4.92	4.52
	1%	0.78	0.90	1.04	0.78	1.00	0.94
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	10.52	10.46	9.70	10.42	9.80	10.18
	5%	5.36	5.42	4.78	4.98	5.38	4.78
	1%	1.20	1.22	1.16	0.94	1.20	0.96

Power of the Tests							
Statistics	Levels\n	20	40	60	80	100	120
$\bar{\mathcal{W}}_n$	10%	81.32	97.60	99.74	100.0	100.0	100.0
	5%	73.02	95.10	99.32	99.94	99.96	100.0
	1%	55.74	87.52	96.70	99.48	99.82	99.98
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	75.76	96.68	99.84	100.0	100.0	100.0
	5%	60.06	92.82	99.24	99.92	99.96	100.0
	1%	31.84	76.36	94.80	99.08	99.84	99.96
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	86.52	98.66	99.94	100.0	100.0	100.0
	5%	78.22	97.34	99.84	100.0	100.0	100.0
	1%	57.06	90.20	98.56	99.86	99.94	100.0

Local Power of the Tests							
Statistics	Levels\n	50	100	200	300	400	500
$\bar{\mathcal{W}}_n$	10%	84.66	81.70	81.64	79.66	79.94	78.90
	5%	77.48	73.36	71.90	69.92	70.28	69.14
	1%	60.44	54.68	50.36	48.10	47.86	46.72
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	80.58	79.96	79.48	77.78	79.28	77.82
	5%	71.22	69.56	68.90	68.52	69.26	66.96
	1%	44.38	45.18	43.92	44.84	45.50	43.16
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	89.76	87.20	87.32	85.32	86.34	85.44
	5%	82.88	79.58	79.32	77.22	77.94	76.84
	1%	64.06	60.04	57.58	56.20	56.84	55.42

Table S.2: 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 $\gamma \in \Gamma := [0, 1]$ 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 $\rho_{\tau_j}(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ 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 $\theta^*_{\tau_2} \neq 1/6$. The adjunct probability measure we employ here is the uniform distribution on Γ.

			Size of the Tests				
Statistics	Levels$\backslash n$	50	100	200	300	400	500
$\bar{\mathcal{W}}_n$	10%	13.60	13.08	11.00	10.98	10.78	10.60
	5%	8.08	7.70	6.22	5.50	5.66	5.62
	1%	2.60	2.44	1.84	1.24	1.62	1.32
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	9.88	9.84	9.82	9.64	10.26	9.54
	5%	4.94	4.92	5.04	4.82	5.02	4.58
	1%	0.84	0.88	1.04	0.78	1.06	0.92
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	10.64	10.44	9.74	10.36	9.84	10.02
	5%	5.62	5.38	4.72	5.06	5.46	4.62
	1%	1.26	1.24	1.10	0.92	1.22	0.98
			Power of the Tests				
Statistics	Levels$\backslash n$	20	40	60	80	100	120
$\bar{\mathcal{W}}_n$	10%	81.40	97.48	99.74	100.0	100.0	100.0
	5%	73.02	95.06	99.28	99.94	99.96	100.0
	1%	55.70	87.36	96.76	99.50	99.82	99.98
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	75.94	96.68	99.84	100.0	100.0	100.0
	5%	60.34	92.84	99.24	99.90	99.96	100.0
	1%	32.14	76.24	94.84	99.08	99.84	99.96
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	86.56	98.60	99.94	100.0	100.0	100.0
	5%	78.22	97.34	99.84	100.0	100.0	100.0
	1%	57.16	90.08	98.52	99.86	99.98	100.0
			Local Power of the Tests				
Statistics	Levels$\backslash n$	50	100	200	300	400	500
$\bar{\mathcal{W}}_n$	10%	84.88	81.64	81.76	79.16	79.76	78.80
	5%	77.80	73.40	71.64	69.44	70.08	69.00
	1%	60.38	54.52	50.18	47.46	47.66	46.92
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	80.42	79.96	79.48	77.38	79.04	77.92
	5%	71.18	69.58	69.06	68.16	69.10	67.22
	1%	44.46	45.22	44.14	44.66	45.70	43.24
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	89.78	87.06	87.36	85.10	86.26	85.46
	5%	83.10	79.78	79.36	76.78	78.00	76.94
	1%	64.08	60.26	57.52	55.90	56.62	55.42

Table S.3: 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 $\gamma \in \Gamma := [0, 1]$ 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 $\rho_{\tau_j}(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ 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 $\theta^*_{\tau_2} \neq 1/6$. The adjunct probability measure we employ here is the beta distribution on Γ with $\alpha = 2$ and $\beta = 2$.

Statistics	Levels\n	50	100	200	300	400	500
Size of the Tests							
$\bar{\mathcal{W}}_n$	10%	13.70	13.14	11.26	11.20	10.90	10.82
	5%	8.02	7.84	6.34	5.74	5.66	5.70
	1%	2.58	2.44	1.84	1.28	1.58	1.34
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	9.70	9.84	9.88	9.70	10.32	9.60
	5%	4.94	4.92	5.14	4.86	5.00	4.60
	1%	0.84	0.92	1.02	0.78	0.98	0.92
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	10.50	10.34	9.92	10.54	10.04	10.14
	5%	5.32	5.40	4.82	5.26	5.42	4.70
	1%	1.16	1.28	1.16	0.88	1.18	0.96

Statistics	Levels\n	20	40	60	80	100	120
Power of the Tests							
$\bar{\mathcal{W}}_n$	10%	81.42	97.60	99.76	100.0	100.0	100.0
	5%	73.16	95.12	99.26	99.94	99.96	100.0
	1%	55.98	87.48	96.86	99.46	99.82	99.98
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	75.92	96.68	99.82	100.0	100.0	100.0
	5%	59.98	92.78	99.22	99.90	99.96	100.0
	1%	31.96	76.28	94.74	99.04	99.84	99.96
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	86.62	98.64	99.94	100.0	100.0	100.0
	5%	78.44	97.44	99.82	100.0	100.0	100.0
	1%	57.38	90.12	98.46	99.86	99.96	100.0

Statistics	Levels\n	50	100	200	300	400	500
Local Power of the Tests							
$\bar{\mathcal{W}}_n$	10%	84.92	81.82	81.66	79.12	79.74	79.08
	5%	77.90	73.64	71.90	69.64	70.10	69.50
	1%	60.64	55.30	50.40	48.14	47.80	47.12
$\mathcal{L}\bar{\mathcal{M}}_n$	10%	80.64	80.02	79.38	77.44	79.06	77.98
	5%	71.38	69.72	68.98	68.40	69.20	67.44
	1%	44.70	45.52	44.06	44.48	45.54	43.22
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	10%	89.88	87.32	87.28	85.22	86.28	85.48
	5%	83.20	79.82	79.34	77.10	77.84	77.18
	1%	64.08	60.68	57.90	55.74	56.64	55.32

Table S.4: 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 $\gamma \in \Gamma := [0, 1]$ 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 $\rho_{\tau_j}(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ 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 $\theta^*_{\tau_2} \neq 1/6$. The adjunct probability measure we employ here is the beta distribution on Γ with $\alpha = 1$ and $\beta = 3$.

Size of the Tests

Statistics	Levels\\n	50	100	200	300	400	500
	10%	13.98	13.36	11.24	11.38	11.02	10.92
$\bar{\mathcal{W}}_n$	5%	8.28	8.02	6.36	5.72	5.90	5.80
	1%	2.60	2.58	1.88	1.30	1.62	1.36
	10%	9.88	9.90	9.86	9.72	10.38	9.60
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	4.96	5.02	5.08	4.82	4.98	4.64
	1%	0.84	0.94	1.02	0.82	1.00	0.94
	10%	10.66	10.50	9.78	10.48	10.00	10.28
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	5.42	5.50	4.76	5.18	5.44	4.74
	1%	1.24	1.34	1.20	0.94	1.18	0.98

Power of the Tests

Statistics	Levels\\n	20	40	60	80	100	120
	10%	81.38	97.60	99.70	100.0	100.0	100.0
$\bar{\mathcal{W}}_n$	5%	73.14	95.14	99.24	99.94	99.96	100.0
	1%	56.04	87.70	96.66	99.48	99.82	99.98
	10%	75.78	96.66	99.84	100.0	100.0	100.0
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	60.08	92.82	99.24	99.90	99.96	100.0
	1%	31.90	76.38	94.84	99.04	99.84	99.96
	10%	86.52	98.62	99.94	100.0	100.0	100.0
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	78.26	97.42	99.84	100.0	100.0	100.0
	1%	57.10	90.26	98.48	99.90	99.96	100.0

Local Power of the Tests

Statistics	Levels\\n	50	100	200	300	400	500
	10%	84.76	82.08	81.84	79.40	80.24	79.14
$\bar{\mathcal{W}}_n$	5%	77.94	73.70	72.36	69.96	70.46	69.44
	1%	60.88	54.92	50.86	48.38	48.08	47.40
	10%	80.50	80.08	79.68	77.36	79.08	77.78
$\mathcal{L}\bar{\mathcal{M}}_n$	5%	71.24	69.58	69.16	68.22	69.04	67.14
	1%	44.40	45.16	44.14	44.76	45.68	43.14
	10%	89.82	87.14	87.48	85.30	86.42	85.56
$\mathcal{Q}\bar{\mathcal{L}}\mathcal{R}_n$	5%	83.02	79.88	79.62	77.02	78.12	77.22
	1%	64.20	60.46	58.24	56.08	57.06	55.62

Table S.5: 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 $\gamma \in \Gamma := [0,1]$ 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 $\rho_{\tau_j}(\gamma, \theta_{\tau_j}) := \theta_{\tau_j} + \theta_{\tau_j}\gamma$ 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 $\theta^*_{\tau_2} \neq 1/6$. The adjunct probability measure we employ here is the beta distribution on Γ with $\alpha = 5$ and $\beta = 1$.

		Male			Female		
Estimated errors of the quantiles of the original log income path							
		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 S.6: 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 S.7: 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.

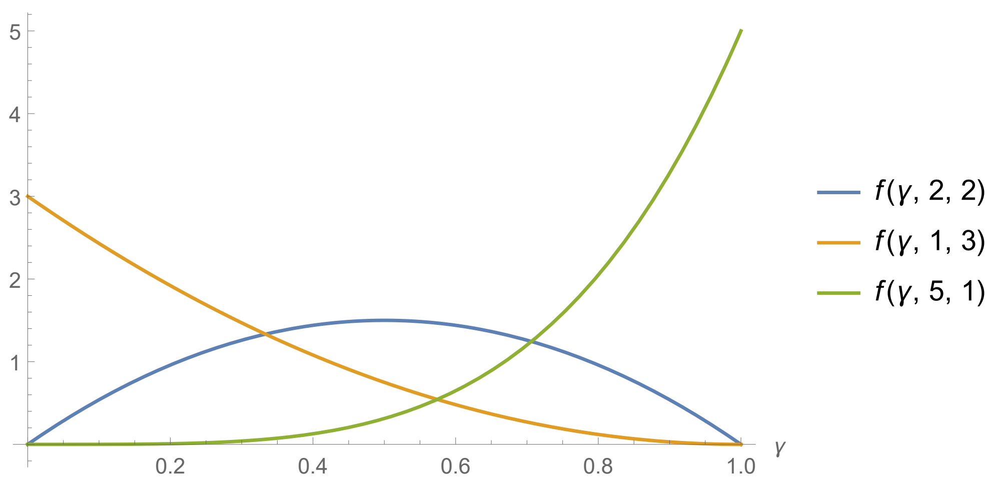

Figure S.1: SHAPES OF BETA PROBABILITY DENSITY FUNCTIONS FOR DIFFERENT PARAMETERS.

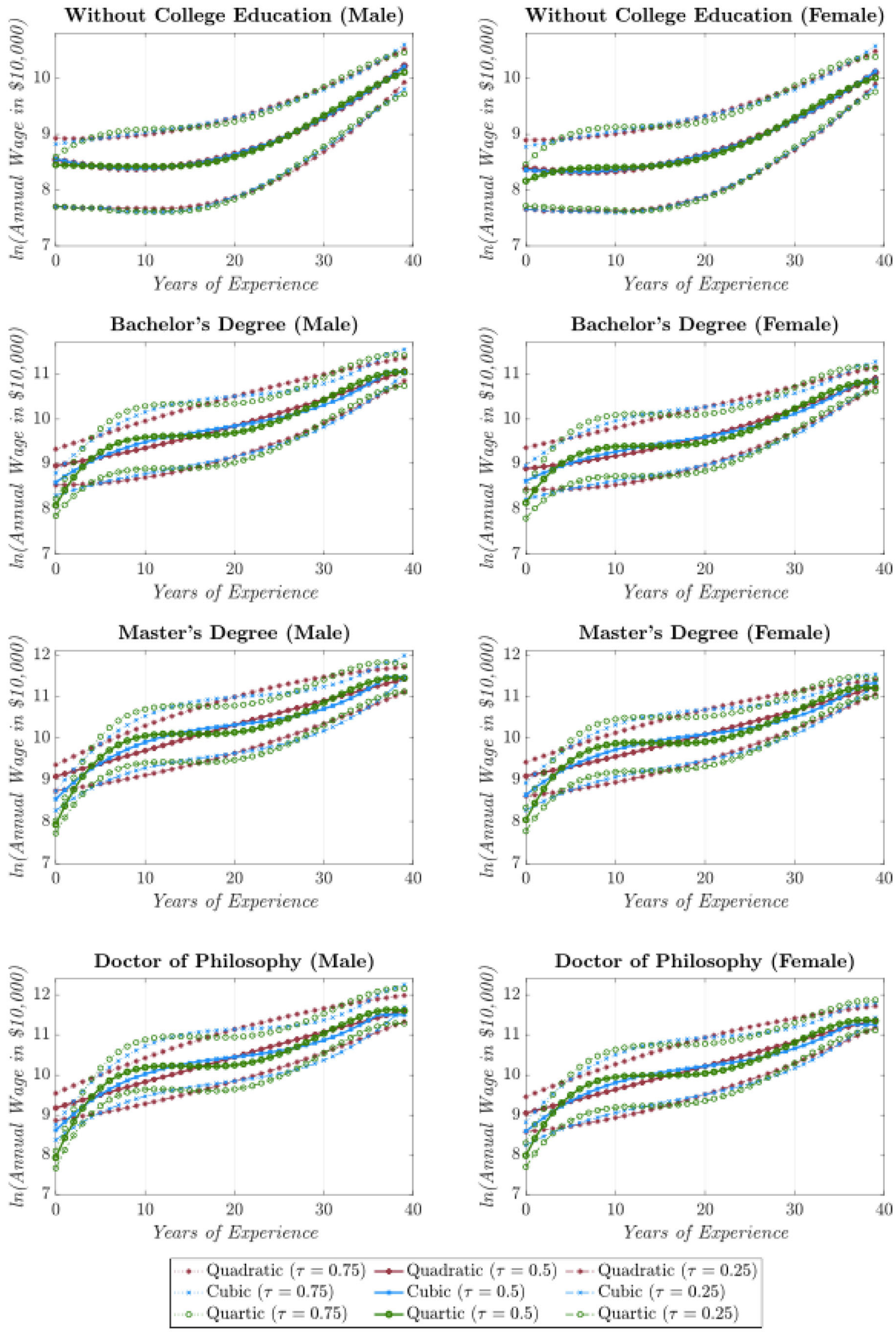

Figure S.2: ESTIMATED QUANTILE FUNCTIONS OVER 0 TO 40 WORK EXPERIENCE YEARS USING THE ORIGINAL LIPS.

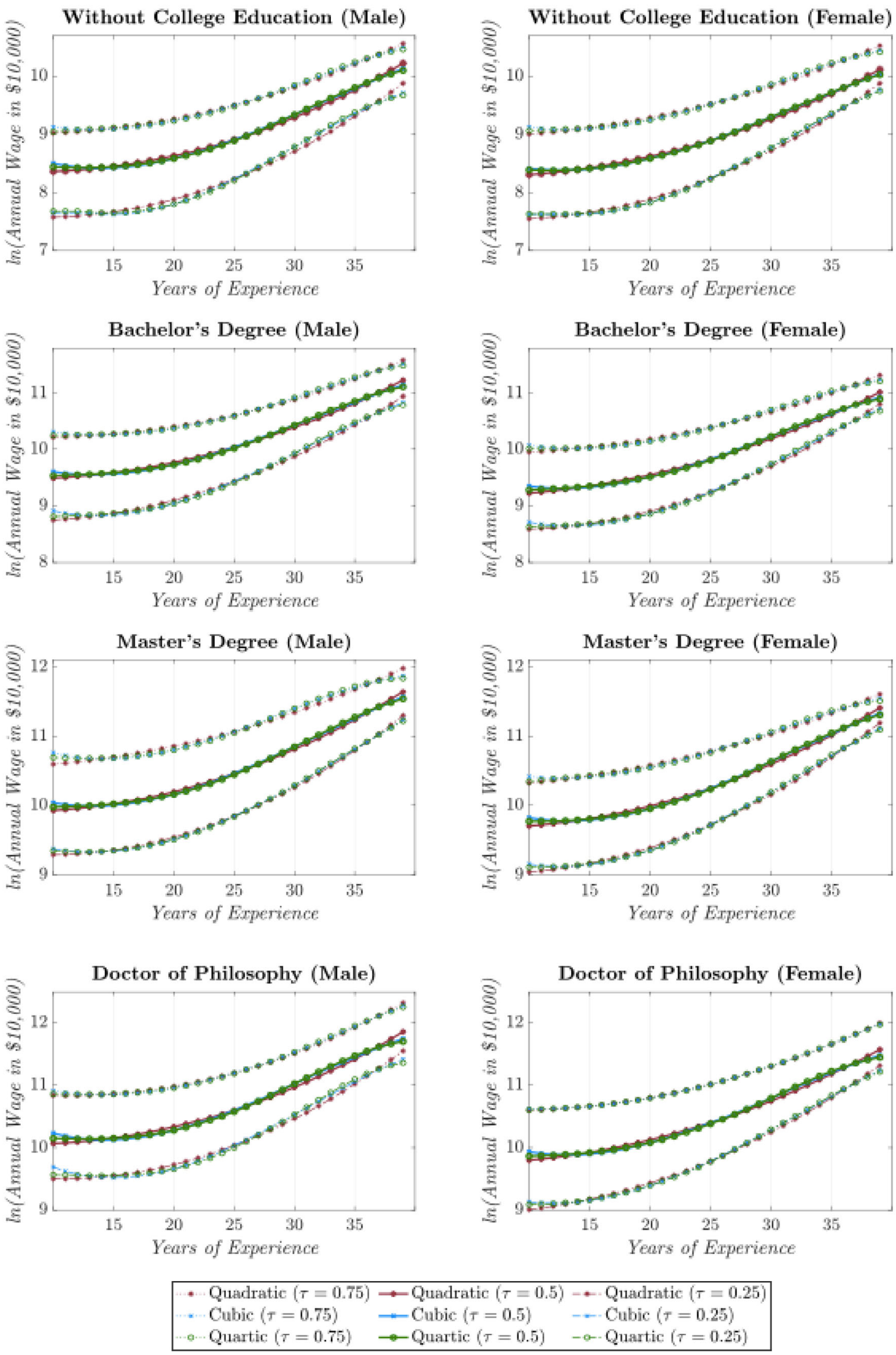

Figure S.3: ESTIMATED QUANTILE FUNCTIONS OVER 10 TO 40 WORK EXPERIENCE YEARS USING THE ORIGINAL LIPS.