RedNIE
Red Nacional de Investigadores en Economía

Location-Scale and Compensated Effects in Unconditional Quantile Regressions

Julián Martinez-Iriarte (UC Santa Cruz)

Gabriel Montes-Rojas (UBA/CONICET)

Yixiao Sun (UC San Diego)

DOCUMENTO DE TRABAJO N° 127

Marzo de 2022

Citar como:
Martinez-Iriarte, Julián, Gabriel Montes-Rojas y Yixiao Sun (2022). Location-Scale and Compensated Effects in Unconditional Quantile Regressions. *Documento de trabajo RedNIE N°127*.

Location-Scale and Compensated Effects in Unconditional Quantile Regressions*

Julian Martinez-Iriarte† Gabriel Montes-Rojas‡ Yixiao Sun§

January 10, 2022

Abstract

This paper proposes an extension of the unconditional quantile regression analysis to (i) location-scale shifts, and (ii) compensated shifts. The first case is intended to study a counterfactual policy analysis aimed at increasing not only the mean or location of a covariate but also its dispersion or scale. The compensated shift refers to a situation where a shift in a covariate is compensated at a certain rate by another covariate. Not accounting for these possible scale or compensated effects will result in an incorrect assessment of the potential policy effects on the quantiles of an outcome variable. More general interventions and compensated shifts are also considered. The unconditional policy parameters are estimated with simple semi-parametric estimators, for which asymptotic properties are studied. Monte Carlo simulations are implemented to study their finite sample performances, and the proposed approach is applied to a Mincer equation to study the effects of a location-scale shift in education on the unconditional quantiles of wages.

Keywords: Quantile regression, unconditional policy effect, unconditional regression.

JEL: J01, J31.

*For helpful conversations, we thank Javier Alejo, George Bulman and Augusto Nieto-Barthaburu. All errors remain our own.

†Department of Economics, UC Santa Cruz. E-mail: jmart425@ucsc.edu
‡CONICET and Universidad de Buenos Aires. E-mail: gabriel.montes@fce.uba.ar
§Department of Economics, UC San Diego. E-mail: yisun@ucsd.edu

1 Introduction

In many research areas, it is important to assess the distributional effects of covariates on an outcome variable. Several methods have been implemented in the literature to study this. A prolific line of research is a combination of conditional mean and quantile regression models together with micro simulation exercises, as in Autor, Katz, and Kearney (2005), Machado and Mata (1995), and Melly (2005) (see Fortin, Lemieux, and Firpo (2011) for a review). A more recent and popular method is the recentered influence function (RIF) regression of Firpo, Fortin, and Lemieux (2009), which directly estimates the effect of a covariate change on a functional of the unconditional distribution of the outcome variable. The functional of interest can be the mean, quantile, or any other aspect of the unconditional distribution.

Consider, as an example, the unconditional quantile of the outcome variable Y. Let F_Y be the unconditional distribution function of Y, then the τ-quantile of F_Y is defined by

$$Q_\tau[Y] := \arg\min\{q : \tau \leq F_Y(q)\} \quad \text{for} \quad \tau \in (0,1).$$

We seek to study how $Q_\tau[Y]$ changes when we induce an infinitesimal change in a covariate $X \in \mathbb{R}$, allowing the presence of other observable covariates W and unobservable covariates collected in U. These covariates and the outcome variable are related via a structural or causal function h so that $Y = h(X,W,U)$. We consider a sequence of policy experiments that change X into $X_\delta = \mathcal{G}(X;\delta)$ for a smooth function $\mathcal{G}(\cdot;\cdot)$. The policy experiments are indexed by δ satisfying $\mathcal{G}(X;0) = X$. That is, $\delta = 0$ corresponds to the *status quo* policy. With this induced change in X, the outcome variable becomes $Y_\delta = h(X_\delta,W,U) = h(\mathcal{G}(X;\delta),W,U)$ where the distribution of (X,W,U) is held constant. Our policy experiment has a *ceteris paribus* interpretation at the population level: we change X into X_δ while holding the stochastic dependence among X,W, and U constant. Such a policy experiment is implementable if the covariate X is not a causal factor for either W or U. In this case, when we intervene X and change it into X_δ, W and U will not change. The main parameter of interest is the marginal effect of the change on the unconditional quantile of the outcome variable:

$$\Pi_\tau := \lim_{\delta\to 0} \frac{Q_\tau[Y_\delta] - Q_\tau[Y]}{\delta}.$$

Firpo, Fortin, and Lemieux (2009) consider a pure location shift $X_\delta = X + \delta$. This shift affects the entire unconditional distribution of $Y = h(X,W,U)$, moving it towards a counterfactual distribution of $Y_\delta = h(X_\delta,W,U)$. One of the main results in Firpo, Fortin, and Lemieux (2009, p.958, eq. (6)) is that Π_τ can be represented as an average derivative:

$$\Pi_\tau = E\left[\dot{\psi}_x\left(X,W\right)\right],$$

where

$$\dot{\psi}_x(x,w) = \frac{\partial E[\psi(Y,\tau,F_Y)|X=x,W=w]}{\partial x},$$

$\psi(y,\tau,F_Y) = [\tau - 1\{y \leq Q_\tau[Y]\}] / f_Y(Q_\tau[Y])$ is the influence function of the quantile functional, and $f_Y(Q_\tau[Y])$ is the unconditional density of Y evaluated at the τ-quantile $Q_\tau[Y]$. The unconditional quantile effect Π_τ can then be estimated by first running an unconditional quantile regression (henceforth, UQR), which involves regressing the influence function $\psi(Y_i,\tau,F_Y)$ on the covariates (X_i, W_i) and then taking an average of the partial derivatives of the regression function with respect to X.

The same method is applicable to other functionals of interest — we only need to replace $\psi(Y_i,\tau,F_Y)$ by the influence function underlying the functional we care about. This leads to the general RIF regression of Firpo, Fortin, and Lemieux (2009). The potential simplicity and flexibility that the methodology offers motivate subsequent research to expand the use of RIF regressions. On the empirical side, after its introduction, RIF regressions became a popular method for analyzing and identifying the distributional effects on outcomes in terms of changes in observed characteristics in areas such as labor economics, income and inequality, health economics, and public policy. On the theoretical side, the RIF type of regression has been used to study the effect of a change in a discrete covariate.[1] More recent research on UQR and RIF regressions includes the high-dimensional setting of Sasaki, Ura, and Zhang (2020) and the two-sample problem of Inoue, Li, and Xu (2021).

This paper extends the UQR and RIF regression in several ways. First, we allow simultaneous location and scale shifts in a continuous covariate. The main goal is to study a case where a counterfactual policy analysis aiming at increasing the location or the mean of a covariate might also affect its dispersion. For example, we may consider $X_\delta = X(1+\delta)^{-1} + \delta$. We find that in this case, the marginal effect has a closed-form expression. In order to interpret the scale effect, we introduce the quantile-standard deviation elasticity: the percentage change in the unconditional quantiles of the outcome associated with a 1% change in the standard deviation of the target covariate.

Second, we consider the case of compensated location changes in two covariates. This happens when a location shift in one covariate induces a location shift in another covariate. For example, $Y = h(X_1, X_2, W, U)$ for two scalar target covariates X_1 and X_2, and the policy induces $X_{1\delta} = X_1 + \delta$ and $X_{2\delta} = X_2 - \delta$. We show that the compensated effect can be obtained as a linear combination of individual effects obtained by considering one change at a time.

Third, while we focus mainly on location-scale and compensated location shifts, we consider a general framework that includes these two types of shifts as special cases. In fact, our framework allows for any smooth and invertible intervention of the target covariates.

Fourth, we allow the target covariates to be endogenous, and we characterize the asymptotic

[1] In such a case, we may consider a shift in the probability mass function. The discrete case was initially studied by Firpo, Fortin, and Lemieux (2009). See Rothe (2012), Martinez-Iriarte (2020), and Martinez-Iriarte and Sun (2021b) for further studies.

bias of the unconditional effect estimator when the endogeneity is not appropriately accounted for. We eliminate the endogeneity bias using a control function approach.

Fifth, as a complement to the existing literature that focuses on changing the *marginal distribution* of the target covariates, we consider changing the *values* of the target covariates directly. An advantage of our approach is that the changes under consideration are directly implementable. We note that it may not be easy to induce a desired shift in the marginal distribution, and when possible, such a shift is often achieved via transforming the target covariates, which is what we consider here.

Finally, we propose consistent and asymptotically normal semiparametric estimators of the location-scale effect and the compensated effect. The estimators can be easily implemented in empirical work using either a probit or logit specification of the conditional distribution function. We conduct an extensive Monte Carlo study evaluating the finite sample performances of the location-scale effect estimator and the accuracy of the normal approximation. Simulation results show that the estimator works reasonably well under different specifications and that the standard normal distribution provides a good approximation to the finite sample distribution of a studentized test statistic introduced in this paper.

As potential applications of our proposed approach, consider the following empirical examples to motivate its use.

Example 1. ***Effect of increasing education on wage inequality.*** *In a Mincer equation, log wages Y are modeled as a function of certain observable covariates such as education. A study of the effect of a shift in education on wage inequality could be implemented using our proposed framework. We can accommodate a counterfactual policy experiment where there may be not only a general increase in education but also a change in its dispersion.*

Example 2. ***Trade integration and skill distribution*** *Gu, Malik, Pozzoli, and Rocha (2019) document the impact of trade integration on both the mean and the standard deviation of the skill distribution across municipalities in Denmark. Moreover, as argued by Hanushek and Woessmann (2008), skills are related to the income distribution. Thus, a quantification of the impact of a scale effect in the skills distribution on the quantiles of the income distribution appears to be relevant.*

Example 3. ***Financial return and risk.*** *Consider the study of two assets X and W in a portfolio investment framework with stochastic returns Y. We are interested in how changes in the returns of asset X affect the distribution of Y through its unconditional quantiles. A typical exercise involves analyzing changes in the returns (location) and risk (scale) of X. Ignoring the structural interpretations if identification fails, we can still use the proposed framework to decompose the relative contribution of each effect. This could be applied to Value-at-Risk models; see, for instance, Engle and Manganelli (2004).*

We illustrate the proposed method with an empirical application related to Example 1: the effect of changing education on wage inequality, decomposing it into location and scale effects. Empirical results reveal the contrasting nature of the two effects. The location effects are seen to be positive and relatively similar across quantiles. On the other hand, the scale effects are highly heterogeneous and monotonically decreasing across quantiles. Hence, the scale effects can more than offset the location effects. This shows that not accounting for both shifts may result in a biased assessment of the policy effects on the quantiles of the outcome variable.

The paper is organized as follows. Section 2 defines and studies the location-scale marginal effects in one covariate. Section 3 proposes and studies a compensated change in two covariates. Section 4 describes the estimators of the location-scale effect and the compensated effect and studies their asymptotic properties. Section 5 reports the finite sample performance of the location-scale effect estimator and the associated tests, and Section 6 presents the empirical application. Section 7 concludes. The proofs are in the Appendix. Calculation details for two theoretical examples are given in the Supplementary Appendix.

A word on notation: we use $F_{Y|X}(y|x)$ and $f_{Y|X}(y|x)$ to denote the cumulative distribution function and the probability density function of Y, respectively, conditional on $X = x$. For a random variable Z, the unconditional τ-quantile is denoted by $Q_\tau[Z]$, *i.e.*, $\Pr(Z \leq Q_\tau[Z]) = \tau$. For a pair of random variables Z_1 and Z_2, the conditional quantile is denoted by $Q_\tau[Z_1|z_2]$, *i.e.*, $\Pr(Z_1 \leq Q_\tau[Z_1|z_2]|Z_2 = z_2) = \tau$. We adopt the following notational conventions:

$$\frac{\partial E(Z|X)}{\partial X} = \left.\frac{\partial E\left(Z|X=x\right)}{\partial x}\right|_{x=X}, \quad \frac{\partial F_{Z|X}(z|X)}{\partial X} = \left.\frac{\partial F_{Z|X}(z|X=x)}{\partial x}\right|_{x=X}.$$

2 Location-scale marginal effects

2.1 Basic setting and main results

We start with a general structural model $Y = h(X, W, U)$, where the function h is unknown, and we only observe (X, W) and Y. Here X is univariate and is our *target* variable. The dimension of W is left unrestricted, and U collects all unobserved causal factors of Y. Consider the following location-scale shift of X,

$$X_\delta = \frac{X-\mu}{s(\delta)} + \mu + \ell(\delta). \tag{1}$$

Here, μ is a *known* parameter, $\ell(\delta)$ is the location shift, and $s(\delta) > 0$ is the scale shift.[2] Under (1) with $\mu = \mu_X$ for $\mu_X := E\left(X\right)$, we have $E[X_\delta] = \mu + \ell(\delta)$ and the variance is $V[X_\delta] = s(\delta)^{-2}V[X]$. In this case, $\ell(\delta)$ affects only the location, and $s(\delta)$ affects only the scale. When $\mu \neq \mu_X$, then $E[X_\delta] = \mu + \ell(\delta) + s(\delta)^{-1}\left[E\left(X\right) - \mu\right]$ and $V[X_\delta] = s(\delta)^{-2}V[X]$. In this case, $s\left(\delta\right)$ affects both the location and the scale. We allow for a general μ that includes, for example, $\mu = 0$ and $\mu = \mu_X$ as special cases.

We view $s(\delta)$ and $\ell(\delta)$ as functions of the scalar δ, and assume that they are continuously differentiable. We further assume that $s(0) = 1$ and $\ell(0) = 0$ so that $X_0 = X$. The case studied

[2] μ is given by the policy maker or calibrated. Note that if $Q_\tau[X]$ is the τ-quantile of X, then

$$Q_\tau[X_\delta] = \frac{Q_\tau[X]-\mu}{s(\delta)} + \mu + \ell(\delta).$$

by Firpo, Fortin, and Lemieux (2009) amounts to setting $s(\delta) \equiv 1$ and $\ell(\delta) = \delta$, and thus, does not account for the scale effect and is independent of the choice of μ. To include the scale effect, we could set $s(\delta) = 1 + \delta$ and $\ell(\delta) = \delta$. A special case of this model is the case with only a scale shift (i.e., $\ell(\delta) = 0$) so that $X_\delta = (X - \mu)/s(\delta) + \mu$.

To allow for a more general policy function that includes the location-scale shift in (1) as a special case, we consider the intervention:

$$X_\delta = \mathcal{G}(X; \delta)$$

for some smooth function $\mathcal{G}(\cdot;\cdot)$ that is invertible in its first argument. We will refer to $\mathcal{G}(\cdot;\cdot)$ as the policy function. We want to compare the quantiles of

$$Y = h(X, W, U) \tag{2}$$

to the quantiles of

$$Y_\delta = h(X_\delta, W, U) = h(\mathcal{G}(X;\delta), W, U), \tag{3}$$

where the distribution of (X, W, U) in (3) is held the same as that in (2). To understand the latter condition, we can consider two parallel worlds: the worlds before and after the intervention. For each given δ, let $\mathcal{G}^{-1}(x;\delta)$ be the inverse function of $\mathcal{G}(x;\delta)$ such that $\mathcal{G}(\mathcal{G}^{-1}(x;\delta);\delta) = x$. After applying the inverse transform to the target covariate in the post-intervention world, the distribution of $\left(\mathcal{G}^{-1}(X^\delta;\delta), W^\delta, U^\delta\right)$ in the post-intervention world is assumed to be the same as that of (X, W, U) in the pre-intervention world. Here, no change is induced on W and U and so $\left(W^\delta, U^\delta\right)$ is actually the same as (W, U) for every individual in the population.

Formally, our parameter of interest, the *marginal effect for the τ-quantile*, is defined as

$$\Pi_\tau := \lim_{\delta \to 0} \frac{Q_\tau[Y_\delta] - Q_\tau[Y]}{\delta},$$

whenever this limit exists. For the location-scale shift that depends on μ, we write Π_τ as Π_τ^μ.

For notational economy, we write $x^\delta = \mathcal{G}^{-1}(x;\delta)$. Then $X_\delta = x$ if and only if $X = x^\delta$. Define the Jacobian of the inverse transform $x \mapsto x^\delta := \mathcal{G}^{-1}(x;\delta)$ as

$$J(x^\delta;\delta) := \frac{\partial x^\delta}{\partial x} = \left[\frac{\partial \mathcal{G}(x;\delta)}{\partial x}\right]^{-1}\Bigg|_{x = x^\delta}.$$

Then, the joint probability density functions of the covariate vector before and after the intervention satisfy

$$f_{X_\delta, W}(x, w) = J(x^\delta;\delta) \cdot f_{X,W}(x^\delta, w).$$

For $\varepsilon > 0$, define $\mathcal{N}_\varepsilon := \{\delta : |\delta| \le \varepsilon\}$. We maintain the following assumption.

Assumption 1. *(i.a) For some $\varepsilon > 0$, $\mathcal{G}(x;\delta)$ is continuously differentiable on $\mathcal{X} \otimes \mathcal{N}_\varepsilon$, where $\mathcal{X}$ is the support of X.*

(i.b) $\mathcal{G}(x;\delta)$ *is strictly increasing in* x *for each* $\delta \in \mathcal{N}_\varepsilon$.

(i.c) $\mathcal{G}(x;0) = x$ *for all* $x \in \mathcal{X}$.

(ii) for $\delta \in \mathcal{N}_\varepsilon$, *the conditional density of* U *satisfies* $f_{U|X_\delta,W}(u|x,w) = f_{U|X,W}(u|x^\delta,w)$, *and the support* $\mathcal{U}$ *of* U *given* X *and* W *does not depend on* (X,W).

(iii.a) $x \mapsto f_{X,W}(x,w)$ *is continuously differentiable for all* $w \in \mathcal{W}$ *and*

$$\int_{\mathcal{W}} \int_{\mathcal{X}} \sup_{\delta \in \mathcal{N}_\varepsilon} \left| \frac{\partial \left[J\left(x^\delta;\delta\right) f_{X,W}(x^\delta,w) \right]}{\partial \delta} \right| dxdw < \infty$$

where $\mathcal{W}$ *is the support of* W.

(iii.b) $x \mapsto f_{U|X,W}(u|x,w)$ *is continuously differentiable for all* (u,w) *and*

$$\int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \sup_{\delta \in \mathcal{N}_\varepsilon} \left| \frac{\partial}{\partial \delta} \left[f_{U|X,W}(u|x^\delta,w) f_{X,W}(x^\delta,w) \right] \right| dudxdw < \infty,$$

$$\int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \sup_{\delta \in \mathcal{N}_\varepsilon} \left| \frac{\partial f_{X,W}(x^\delta,w)}{\partial \delta} \right| f_{U|X,W}(u|x,w) dudxdw < \infty.$$

(iv) $f_{X,W}(x,w)$ *is equal to* 0 *on the boundary of the support of* X *given* $W = w$ *for all* $w \in \mathcal{W}$.

(v) $f_Y(Q_\tau[Y]) > 0$.

Remark 1. *Assumption 1(i) imposes some restrictions on the policy function* $\mathcal{G}(x;\delta)$. *It is reasonable that* $\mathcal{G}(x;\delta)$ *is strictly increasing in* x, *as non-monotonic and non-invertible functions do not seem to be practically relevant. The strictly increasing property implies that* $J(x;\delta) > 0$ *for all* $x \in \mathcal{X}$ *and* $\delta \in \mathcal{N}_\varepsilon$. *The condition that* $\mathcal{G}(x;0) = x$ *says that there is no intervention when* $\delta = 0$, *and it implies that* $J(x;0) = 1$ *for all* $x \in \mathcal{X}$. *Assumption 1(ii) assumes that how* U *depends on the covariate vector is maintained when we induce a change in the covariate vector. Note that Assumption 1(ii) is different from* $f_{U|X_\delta,W}(u|x,w) = f_{U|X,W}(u|x,w)$, *which in general can not hold when* U *depends on* X *and* W. *The counterfactual model in (3) says that we maintain the structure of the causal system. Assumption 1(ii) says that we also maintain how the unobservable depends on the observables. As discussed above, we also implicitly assume that* $\left(\mathcal{G}^{-1}(X^\delta;\delta), W^\delta\right)$ *has the same distribution as* (X,W). *The rest of Assumption 1 consists of regularity conditions.*

Remark 2. *Assumption 1 does not assume that* U *is independent of* (X,W). *It does not assume that* U *is conditionally independent of* X *given* W *either. Assumption 2 below will impose identification assumptions.*

The following theorem characterizes the effects of the policy change on the distribution of Y_δ and its quantiles.

Theorem 1. *Let Assumption 1 hold.*

(i) For each $(x,w) \in \mathcal{X} \otimes \mathcal{W}$,

$$\lim_{\delta \to 0} \frac{f_{X_\delta,W}(x,w) - f_{X,W}(x,w)}{\delta} = \frac{\partial}{\partial x} \left[\kappa(x) f_{X,W}(x,w) \right],$$

where

$$\kappa(x) := \left.\frac{\partial x^{\delta}}{\partial \delta}\right|_{\delta=0} = -\left.\frac{\partial \mathcal{G}(x;\delta)}{\partial \delta}\right|_{\delta=0}.$$

(ii) As $\delta \to 0$, we have

$$\begin{aligned}&\frac{F_{Y_\delta}(y) - F_Y(y)}{\delta} \\ &\to E\left[\left(-\frac{\partial F_{Y|X,W}(y|X,W)}{\partial X} + \mathbb{1}\{h(X,W,U) \leq y\}\frac{\partial \ln f_{U|X,W}(U|X,W)}{\partial X}\right)\kappa(X)\right]\end{aligned}$$

uniformly in $y \in \mathcal{Y}$, the support of Y.

(iii) The marginal effect of the intervention $X_\delta = \mathcal{G}(X;\delta)$ on the τ-quantile of the outcome variable Y can be represented by

$$\Pi_\tau = A_\tau - B_\tau \tag{4}$$

where

$$\begin{aligned}A_\tau &= -E\left[\frac{\partial E[\psi(Y,\tau,F_Y)|X,W]}{\partial X}\kappa(X)\right], \\ B_\tau &= -E\left[\psi(Y,\tau,F_Y)\frac{\partial \ln f_{U|X,W}(U|X,W)}{\partial X}\kappa(X)\right],\end{aligned}$$

and

$$\psi(y,\tau,F_Y) = \frac{\tau - 1(y < Q_\tau[Y])}{f_Y(Q_\tau[Y])}.$$

Remark 3. *To understand Theorem 1(i), we can write*

$$f_{X_\delta,W}(x,w) - f_{X,W}(x,w) = f_{X_\delta,W}(x,w) - f_{X,W}(x^\delta,w) + f_{X,W}(x^\delta,w) - f_{X,W}(x,w).$$

It is quite intuitive that the second term is approximately $\delta \cdot \kappa(x) \cdot \frac{\partial f_{X,W}(x,w)}{\partial x}$ when δ is small. For the first term, we note that $X_\delta = x$ if and only if $X = x^\delta$, and so this term reflects the effect from the Jacobian of the transformation. Indeed, $f_{X_\delta,W}(x,w) - f_{X,W}(x^\delta,w) = \left[J(x^\delta;\delta) - J(x^\delta;0)\right] f_{X,W}(x^\delta,w)$ as $J(x^\delta;0) = 1$. The first term is then approximately equal to $\delta \cdot f_{X,W}(x,w) \cdot \frac{\partial J(x,\delta)}{\partial \delta}\big|_{\delta=0}$. But

$$\left.\frac{\partial J(x,\delta)}{\partial \delta}\right|_{\delta=0} = \left.\frac{\partial}{\partial \delta}\frac{\partial x^\delta}{\partial x}\right|_{\delta=0} = \left.\frac{\partial}{\partial x}\frac{\partial x^\delta}{\partial \delta}\right|_{\delta=0} = \frac{\partial \kappa(x)}{\partial x}$$

and hence the first term is approximately $\delta \cdot f_{X,W}(x,w) \cdot \frac{\partial \kappa(x)}{\partial x}$. Combining these two approximations yields Theorem 1(i).

Remark 4. *By definition, $\kappa(x)$ measures the marginal change of the inverse function $\mathcal{G}^{-1}(x;\delta)$ as we increase δ from zero infinitesimally. Since $\mathcal{G}(x;0) = x$, Theorem 1(i) shows that $\kappa(x)$ is equal to the negative of the marginal change of $\mathcal{G}(x;\delta)$ at $\delta = 0$. Theorems 1 (ii) and (iii) show that only $\kappa(x)$ appears in the marginal effect and the Jacobian does not. This is not surprising, as what matters for the marginal* *effect is the marginal change in the policy function.*

Remark 5. *Theorem 1(iii) represents the structural parameter Π_τ in terms of statistical objects. While the first term A_τ is identifiable, the second term B_τ, which involves the conditional density of U given X and W, is not. If we use $\hat{A}_\tau$, a consistent estimator of A_τ as an estimator of Π_τ, then the second term B_τ is the asymptotic bias of $\hat{A}_\tau$. Similar results have been established in Martinez-Iriarte and Sun (2021a) but only for location changes. If we do not have the identification condition such as what is given in Assumption 2 below, Theorem 1(iii) allows us to use a bound approach to bound B_τ and infer the range of the policy effect or conduct a sensitivity analysis similar to that in Martinez-Iriarte (2020).*

Remark 6. *While the paper focuses on the quantile functional, Theorem 1(iii) is formulated in a general way. The result holds for any Hadamard differentiable functional and for the mean functional. We only need to replace $\psi(y, \tau, F_Y)$ by the influence function of the functional that we are interested in. For example, for the mean functional, we can replace $\psi(y, \tau, F_Y)$ by $y - E(Y)$, and Theorem 1(iii) remains valid.*

To identify Π_τ, we make the following independence or conditional independence assumption.

Assumption 2. *For $\delta \in \mathcal{N}_\varepsilon$, the unobservable U satisfies either $f_{U|X,W}(u|x,w) = f_{U|X,W}(u|x^\delta, w) = f_U(u)$ or $f_{U|X,W}(u|x,w) = f_{U|X,W}(u|x^\delta, w) = f_{U|W}(u|w)$.*

Under the above assumption, $\partial \ln f_{U|X,W}(u|x,w)/\partial x = 0$ and the second term B_τ in (4) vanishes. In this case, $\Pi_\tau = A_\tau$ and hence is identified.

For the location-scale shift given in (1), we have

$$\kappa(x) = \dot{s}(0)(x - \mu) - \dot{\ell}(0),$$

where $\dot{s}(\delta) = ds(\delta)/d\delta$ and $\dot{\ell}(\delta) = d\ell(\delta)/d\delta$. The corollary below then follows directly from Theorem 1(iii).

Corollary 1. *Let Assumption 1 hold with Assumption 1 (ii) strengthened to Assumption 2. Then*

$$\Pi_\tau = -E\left[\frac{\partial E[\psi(Y,\tau,F_Y)|X,W]}{\partial X}\kappa(X)\right] = \frac{1}{f_Y(Q_\tau[Y])}E\left[\frac{\partial F_{Y|X,W}(Q_\tau[Y]|X,W)}{\partial X}\kappa(X)\right].$$

For the location and scale shift in (1) with $\ell(0) = 0$, $s(0) = 1$, and $s(\delta) > 0$, we have

$$\Pi_\tau^\mu = \Pi_{\tau,L} + \Pi_{\tau,S}^\mu, \tag{5}$$

where

$$\Pi_{\tau,L} = -\frac{\dot{\ell}(0)}{f_Y(Q_\tau[Y])}\int_{\mathcal{W}}\int_{\mathcal{X}}\frac{\partial F_{Y|X,W}(Q_\tau[Y]|x,w)}{\partial x}f_{X,W}(x,w)dxdw,$$
$$\Pi_{\tau,S}^\mu = \frac{\dot{s}(0)}{f_Y(Q_\tau[Y])}\int_{\mathcal{W}}\int_{\mathcal{X}}\frac{\partial F_{Y|X,W}(Q_\tau[Y]|x,w)}{\partial x}(x-\mu)f_{X,W}(x,w)dxdw.$$

Remark 7. *Both conditions in Assumption 2 require that* $f_{U|X,W}(u|x,w) = f_{U|X,W}(u|x^{\delta},w)$. *This is related to the assumption in Firpo, Fortin, and Lemieux (2009, pp.955-957), framed as "maintaining the conditional distribution of Y given X unaffected." In essence, Firpo, Fortin, and Lemieux (2009) requires* $f_{U|X}(u|x) = f_{U|X}(u|x^{\delta})$. *When this condition fails, we may still have* $f_{U|X,W}(u|x,w) = f_{U|X,W}(u|x^{\delta},w)$. *Such a condition has also been used in Hsu, Lai, and Lieli (2020) and Spini (2021) in a context of extrapolation to populations with different distributions of the covariates.*

Remark 8. *The first condition in Assumption 2 is satisfied if U is independent of* (X,W). *The second condition in Assumption 2 is a conditional independence assumption, which is commonly used to achieve identification. When W consists of only causal variables entering the causal function* $h(X,W,U)$, *Assumption 2 may not hold. In this case, we can find control variables* W_c *so that*

$$f_{U|X,W,W_c}(u|x,w,w_c) = f_{U|X,W,W_c}(u|x^{\delta},w,w_c) = f_{U|W,W_c}(u|w,w_c).$$

After replacing W by $W^* = (W,W_c)$, *Corollary 1 continues to hold. To see this, we can write the structural function as* $h^*(X,W^*,U)$, *but* $h^*(X,W^*,U) = h(X,W,U)$. *That is, we include the control variables in the structural function and restrict the structural function to be a constant function of the control variables. With such a conceptual change, our proof goes through without any change.*

Remark 9. *The second part of Corollary 1 is specific to the location-scale change. The overall effect* Π_{τ} *can be decomposed into the sum of* $\Pi_{\tau,L}$ *and* $\Pi^{\mu}_{\tau,S}$. *Here* $\Pi_{\tau,L}$ *is the location effect, and is the estimand studied by Firpo, Fortin, and Lemieux (2009) when we set* $\dot{\ell}(0) = 1$ *and* $s(\delta) \equiv 1$. $\Pi^{\mu}_{\tau,S}$ *is the scale effect, and it is one of the main objects of interest in this study.*

Remark 10. *Corollary 1 shows that the scale effect under a general* μ *is linearly related to the location effect and the scale effect under the specific* $\mu = u_X$:

$$\Pi^{\mu}_{\tau,S} = \tilde{\mu}\Pi_{\tau,L} + \Pi^{\mu_X}_{\tau,S}, \tag{6}$$

where

$$\tilde{\mu} = (\mu - \mu_X)\,\frac{\dot{s}\,(0)}{\dot{\ell}\,(0)}.$$

The slope $\tilde{\mu}$ *is proportional to* $\mu - \mu_X$ *and independent of* τ. *We will refer to* $\Pi^{\mu_X}_{\tau,S}$ *as the pure scale effect, as it is not related to the location effect.*

In the rest of this section, we focus on the location-scale shift. To better understand the location and scale effects in Corollary 1, consider the case that X and U are independent and there is no W. Then

$$\begin{aligned}\Pi_{\tau,L} &= -\frac{\dot{\ell}\,(0)}{f_Y(Q_{\tau}[Y])}\int_{\mathcal{X}}\frac{\partial F_{Y|X}(Q_{\tau}[Y]|x)}{\partial x}f_X(x)dx,\\ \Pi^{\mu}_{\tau,S} &= \frac{\dot{s}\,(0)}{f_Y(Q_{\tau}[Y])}\int_{\mathcal{X}}\frac{\partial F_{Y|X}(Q_{\tau}[Y]|x)}{\partial x}\,(x-\mu)\,f_X(x)dx.\end{aligned} \tag{7}$$

Define

$$X_{\tau,F}(x) = \frac{\partial F_{Y|X}(Q_\tau[Y]|x)}{\partial x} = \frac{\partial \Pr(Y \le Q_\tau[Y]|X = x)}{\partial x},$$

which measures how $\Pr(Y \le Q_\tau[Y]|X = x)$ will change when we induce a small change in x. By definition,

$$X_{\tau,F}(x) = \lim_{\Delta \to 0} \frac{\Pr(Y \le Q_\tau[Y]|X = x + \Delta) - \Pr(Y \le Q_\tau[Y]|X = x)}{\Delta}. \tag{8}$$

Intuitively, when x is changed into $x + \Delta$, the value of Y will cross $Q_\tau[Y]$ from above for a subset of individuals, and the value of Y will cross $Q_\tau[Y]$ from below for another subset of individuals. The difference in the fractions of individuals in these two subsets is the numerator of (8). $X_{\tau,F}(x)$ is then the limit value of the difference rescaled by the induced change in x.

Note that $X_{\tau,F}(x)$ is possibly a nonlinear function of x. For notational simplicity, let $X_{\tau,F} = X_{\tau,F}(X)$. To sign the location effect and the pure scale effect, consider the best linear prediction of $X_{\tau,F}$ using $X - \mu_X$ as the predictor:

$$X_{\tau,F} = c_{0\tau}^* + (X - \mu_X) c_{1\tau}^* + e_\tau,$$

where $E(e_\tau) = 0$ and $cov(X, e_\tau) = 0$. By definition,

$$c_{0\tau}^* = E[X_{\tau,F}] = \int_{\mathcal{X}} \frac{\partial F_{Y|X}(Q_\tau[Y]|x)}{\partial x} f_X(x) dx,$$
$$c_{1\tau}^* = \frac{cov(X_{\tau,F}, X)}{var(X)} = \frac{1}{var(X)} \int_{\mathcal{X}} \frac{\partial F_{Y|X}(Q_\tau[Y]|x)}{\partial x} (x - \mu_X) f_X(x) dx.$$

Therefore,

$$\Pi_{\tau,L} = -\dot{\ell}(0) \frac{1}{f_Y(Q_\tau[Y])} c_{0\tau}^* \text{ and } \Pi_{\tau,S}^{\mu_X} = \dot{s}(0) \frac{var(X)}{f_Y(Q_\tau[Y])} c_{1\tau}^*.$$

The signs of $\Pi_{\tau,L}$ and $\Pi_{\tau,S}^{\mu_X}$ can then be determined from the signs of the best predictive intercept and slope coefficient.

To sign the location effect $\Pi_{\tau,L}$, we can assess whether $\Pr(Y \le Q_\tau[Y]|X = x)$ is increasing in x or not. If $\dot{\ell}(0) > 0$ and $\Pr(Y \le Q_\tau[Y]|X = x)$ is increasing in x on average, more precisely, $E[X_{\tau,F}] \ge 0$, then $\Pi_{\tau,L} \le 0$. As an example, consider the case that $h(x, u)$ is decreasing in x for each u. In this case, $\Pr(Y \le Q_\tau[Y]|X = x)$ is increasing in x for all $x \in \mathcal{X}$, and so $\Pi_{\tau,L} \le 0$ if $\dot{\ell}(0) > 0$.

It is a bit more challenging to sign the pure scale effect $\Pi_{\tau,S}^{\mu_X}$. The best linear predictive coefficient $c_{1\tau}^*$ depends not only on the function form of $X_{\tau,F}(x)$ but also on the distribution of X. We consider two examples below.

Example 4. ***Normal Location Model.*** *Consider a typical linear model $Y = \alpha + X\beta + U$, where X and U are independent $N(0,1)$. We have: $\Pi_{\tau,L} = \dot{\ell}(0)\beta$ and*

$$\Pi_{\tau,S}^{\mu_X} = -\dot{s}(0) \frac{\beta^2}{\beta^2 + 1} (Q_\tau[Y] - \alpha) = -\dot{s}(0) \frac{\beta^2}{\sqrt{\beta^2 + 1}} Q_\tau[U].$$

See Section S.1 in the Supplementary Appendix for details. While the location effect is constant across τ, the pure scale effect varies across quantiles and does not depend on the sign of β. The coefficient on the scale effect (i.e., $\beta^2/(\beta^2+1)$) has a "signal-to-noise-ratio" interpretation. Indeed, $\Pi^{\mu_X}_{\tau,S} = -\dot{s}(0)E[X\beta|Y = Q_\tau[Y]]$. See Theorem 2 below. This can be regarded as an inverse prediction problem. Given $Y = Q_\tau[Y]$, we want to predict or extract the signal $X\beta$. The predictive coefficient, given by $var(X\beta)/var(Y)$, is precisely $\beta^2/(\beta^2+1)$.

The next example represents the pure scale effect under increasingly restrictive assumptions, culminating with a generalization of Example 4. Details are given in Section S.2 in the Supplementary Appendix.

Example 5. ***Normal Covariates.*** *Consider the linear model $Y = \alpha + X\beta + U$ where X and U are independent. Suppose we only assume that $X \sim N(\mu_X, \sigma_X^2)$. We can use Stein's lemma (see, for example, Casella and Berger (2001, pp.124-125) and references therein) to gain some insight on the pure scale effect. Stein's lemma states that for a differentiable function m such that $E[|m'(X)|] < \infty$, $E[m(X)(X-\mu_X)] = \sigma_X^2 E[m'(X)]$ whenever $X \sim N(\mu_X, \sigma_X^2)$. Taking $m(x) = X_{\tau,F}(x)$ and using Stein's lemma, we can express the pure scale effect as*

$$\begin{aligned}\Pi^{\mu_X}_{\tau,S} &= \frac{\dot{s}(0)}{f_Y(Q_\tau[Y])}E\left[X_{\tau,F}(X)(X-\mu_X)\right] = \frac{\dot{s}(0)}{f_Y(Q_\tau[Y])}\sigma_X^2 E\left[\frac{\partial X_{\tau,F}(X)}{\partial X}\right] \\ &= \frac{\dot{s}(0)}{f_Y(Q_\tau[Y])}\sigma_X^2 E\left[\frac{\partial^2 F_{Y|X}(Q_\tau[Y]|X)}{\partial X^2}\right].\end{aligned}$$

Therefore, when X is normal and $\dot{s}(0) > 0$, the pure scale effect is non-negative (non-positive) if $F_{Y|X}(Q_\tau[Y]|x)$ is a convex (concave) function of x. It is interesting to see that the location effect depends on the first order derivative of $F_{Y|X}(Q_\tau[Y]|x)$ while the pure scale effect depends on its second-order derivative.

If $F_{Y|X}(Q_\tau[Y]|x) = G_\tau(a_\tau + b_\tau x)$ for some link function G_τ and parameters a_τ and b_τ, then

$$X_{\tau,F}(x) = b_\tau \dot{G}_\tau(a_\tau + b_\tau x)$$

and

$$\frac{\partial X_{\tau,F}(x)}{\partial x} = b_\tau^2 \ddot{G}_\tau(a_\tau + b_\tau x)$$

where $\dot{G}_\tau$ and $\ddot{G}_\tau$ are the first order and second order derivatives of G_τ. Hence

$$\Pi^{\mu_X}_{\tau,S} = \dot{s}(0)\frac{\sigma_X^2 b_\tau^2}{f_Y(Q_\tau[Y])}E\left[\ddot{G}_\tau(a_\tau + b_\tau X)\right].$$

For example, when U is also normal so that G_τ is the standard normal cumulative distribution function (cdf), we obtain a generalization of Example 4:

$$\Pi^{\mu_X}_{\tau,S} = -\dot{s}(0)\frac{\sigma_X^2\beta^2}{\sqrt{\sigma_X^2\beta^2+1}}Q_\tau[U].$$

The above result reduces to that of Example 4 when we set $\sigma_X^2 = 1$. Regardless of whether $\sigma_X^2 = 1$, $\Pi_{\tau,S}^{\mu_X}$ does not depend on μ_X. Thus, the mean of X does not play a role in the pure scale effect.

To understand why, in Example 5, the pure scale effect does not depend on μ_X and $\text{sign}(\beta)$, we write

$$
\begin{aligned}
cov\left(\frac{\partial F_{Y|X}(Q_\tau[Y]|X)}{\partial x}, X\right) &= -\beta cov\left(f_U(Q_\tau[Y] - \alpha - X\beta), X\right) \\
&= cov\left(f_U(Q_\tau[U^\circ] + X^\circ), X^\circ\right),
\end{aligned}
$$

where $X^\circ := -\beta\left(X - \mu_X\right)$ and $U^\circ = U - X^\circ$. Now, if $X - \mu_X$ is symmetrically distributed around zero, then the above covariance does not depend on sign(β) as the distributions of X° and U° remain the same if we flip the sign of β. Also, since the distributions of X° and U° do not depend on μ_X, the above covariance does not depend on μ_X. On the other hand, for the denominator of the scale effect, we have

$$
\begin{aligned}
f_Y(Q_\tau[Y]) &= f_Y(Q_\tau[\alpha + X\beta + U]) = f_Y(Q_\tau[\alpha + (X - \mu_X)\beta + U + \mu_X\beta]) \\
&= f_Y(\alpha + Q_\tau[U^\circ + \mu_X\beta]) = f_Y(\alpha + Q_\tau[U^\circ] + \mu_X\beta) = f_{U^\circ}\left(Q_\tau[U^\circ]\right).
\end{aligned}
$$

If $X - \mu_X$ is symmetrically distributed around zero, then the distribution of U° does not depend on μ_X or sign(β). Hence, $f_Y(Q_\tau[Y])$ does not depend on μ_X or sign(β).

Since both the numerator and the denominator of $\Pi_{\tau,S}^{\mu_X}$ are invariant to μ_X and sign(β), we obtain the following proposition immediately.

Proposition 1. *Consider the linear model $Y = \alpha + X\beta + U$ where X and U are independent. If $X - E[X]$ is symmetrically distributed around zero, then the pure scale effect does not depend either on $E[X]$ or on sign(β).*

2.2 Interpretation of the scale effects

Consider a situation where we only care about the scale effect, that is, we set $\ell(\delta) \equiv 0$. Then, we have $x^\delta = \mu + (X - \mu)/s(\delta)$ and $\sigma_{X_\delta} = \sigma_X / s(\delta)$. To interpret $\Pi_{\tau,S}^{\mu}$, we assume $Q_\tau[Y_\delta] \neq 0$ and consider the following *Y-quantile-X-standard-deviation elasticity*

$$
\mathcal{E}_{\tau,\delta} := \frac{dQ_\tau[Y_\delta]}{Q_\tau[Y_\delta]}\left(\frac{d\sigma_{X_\delta}}{\sigma_{X_\delta}}\right)^{-1}.
$$

By straightforward calculations, we have

$$
\mathcal{E}_{\tau,\delta} = -\frac{1}{Q_\tau[Y]}\frac{dQ_\tau[Y_\delta]}{d\delta}\left(\frac{1}{s(\delta)}\frac{ds(\delta)}{d\delta}\right)^{-1}.
$$

When $s(0)=1$ and $\dot{s}(0)\neq 0$, the elasticity at $\delta=0$ is

$$\mathcal{E}_{\tau,\delta=0}=-\frac{\Pi^{\mu}_{\tau,S}}{\dot{s}(0)\,Q_{\tau}[Y]}. \tag{9}$$

Therefore, a 1% *decrease* in the standard deviation of X results in a $\Pi^{\mu}_{\tau,S}/\left(\dot{s}(0)\,Q_{\tau}[Y]\right)$ % change in the τ-quantile of Y.

Example 5 (Continued)**.** *In this case, the elasticity of $\Pi^{\mu_X}_{\tau,S}$ at $\delta=0$ is:*

$$\mathcal{E}_{\tau,\delta=0}=\frac{\sigma_X^2\beta^2}{\sqrt{\sigma_X^2\beta^2+1}}\frac{Q_{\tau}[U]}{Q_{\tau}[Y]}=\frac{\sigma_X^2\beta^2}{\sqrt{\sigma_X^2\beta^2+1}}\frac{Q_{\tau}[U]}{\alpha+\mu_X\beta+\sqrt{\sigma_X^2\beta^2+1}Q_{\tau}(U)}.$$

So, $\mathcal{E}_{\tau,\delta=0}$ is positive if $Q_{\tau}[U]$ and $Q_{\tau}[Y]$ have the same sign. When $\alpha=0$ and $\mu_X=0$, $\mathcal{E}_{\tau,\delta=0}=\sigma_X^2\beta^2/(\sigma_X^2\beta^2+1)$, which is positive for all quantile levels.

To find the value of $s(\delta)$ corresponding to a Δ% change in the standard deviation of X, we let

$$\sigma_{X_{\delta}}=\frac{\sigma_X}{s(\delta)}=\left(1+\frac{\Delta}{100}\right)\sigma_X.$$

We then obtain

$$s(\delta)=\left(1+\frac{\Delta}{100}\right)^{-1}.$$

For $\Delta=-1$, which corresponds to 1% decrease in σ_X, we have $s(\delta)=(1-1/100)^{-1}=1.0101$.

Often times, when the outcome is strictly positive as in prices and wages, we are interested in $\log Y$. In such a case we denote the marginal scale effect by $\tilde{\Pi}^{\mu}_{\tau,S}$, and, since we set $\ell(\delta)\equiv 0$ and there is no location effect, it is given by

$$\tilde{\Pi}^{\mu}_{\tau,S}:=\lim_{\delta\to 0}\frac{Q_{\tau}[\log Y_{\delta}]-Q_{\tau}[\log Y]}{\delta}.$$

Since $\log(\cdot)$ is a strictly increasing transformation, we have

$$\tilde{\Pi}^{\mu}_{\tau,S}=\lim_{\delta\to 0}\frac{\log Q_{\tau}[Y_{\delta}]-\log Q_{\tau}[Y]}{\delta},$$

and we can relate $\tilde{\Pi}^{\mu}_{\tau,S}$ to $\Pi^{\mu}_{\tau,S}$ by

$$\tilde{\Pi}^{\mu}_{\tau,S}=\frac{1}{Q_{\tau}[Y]}\Pi^{\mu}_{\tau,S}.$$

Comparing this last expression to (9), we obtain that the elasticity at $\delta=0$ is

$$\mathcal{E}_{\tau,\delta=0}=-\frac{\tilde{\Pi}^{\mu}_{\tau,S}}{\dot{s}(0)}.$$

This says that a 1% *increase* in the standard deviation of X results in a $-\tilde{\Pi}^{\mu}_{\tau,S}/\dot{s}(0)$ % change in the τ-quantile of Y. When $\dot{s}(0) = -1$, the scale effect $\tilde{\Pi}^{\mu}_{\tau,S}$ (based on $\log(Y)$) can be interpreted directly as the Y-quantile-X-standard-deviation elasticity.

2.3 Relation to Conditional Effects

In order to explore the relationship between conditional quantile regression coefficients and unconditional effects, we introduce a "conditional" version of the unconditional effect given in Corollary 1. For a given $(x, w) \in \mathcal{X} \otimes \mathcal{W}$, this is defined as

$$\Pi_\tau(x,w) := \lim_{\delta \to 0} \frac{Q_\tau[Y_\delta|x,w] - Q_\tau[Y|x,w]}{\delta},$$

whenever this limit exists. In the above, $Q_\tau[Y|x,w]$ is the conditional quantile of Y given $X = x$ and $W = w$, and $Q_\tau[Y_\delta|x,w]$ is the conditional quantile of $Y_\delta = h(\mathcal{G}(X,\delta), W, U)$ given $X = x$ and $W = w$. Under essentially the same assumptions as in Theorem 1 and Corollary 1, we can obtain

$$\Pi_\tau(x,w) = \frac{1}{f_{Y|X,W}(Q_\tau[Y|x,w]|x,w)} \frac{\partial F_{Y|X,W}(Q_\tau[Y|x,w]|z,w)}{\partial z}\bigg|_{z=x} \kappa(x).$$

This is a "non-integrated" version of Π_τ. There are two differences between $\Pi_\tau(x,w)$ and Π_τ. First, instead of the unconditional quantile $Q_\tau[Y]$, the conditional quantile $Q_\tau[Y|x,w]$ is used in $\Pi_\tau(x,w)$. Second, instead of the unconditional density f_Y, the conditional density $f_{Y|X,W}$ is used in $\Pi_\tau(x,w)$. Note that "iterating the expectation" is unlikely to work, *i.e.*, $E[\Pi_\tau(X,W)]$ is, in general, not equal to Π_τ. Due to this fact, we follow Firpo, Fortin, and Lemieux (2009) and use the matching function:

$$\xi_\tau(x,w) = \left\{\eta : Q_\eta[Y|x,w]) = Q_\tau[Y]\right\}. \tag{10}$$

This function matches the unconditional quantile at quantile level τ with the conditional quantile (conditioning on $X = x, W = w$) at the quantile level $\xi_\tau(x,w)$.

Theorem 2. *If $\Pi_\tau(x,w)$ exists for all τ in the support of $\xi_\tau(X,W)$, then the unconditional marginal effect can be represented as*

$$\Pi_\tau = E\left[\Pi_{\xi_\tau(X,W)}(X,W)\frac{f_{Y|X,W}(Q_\tau[Y]|X,W)}{f_Y(Q_\tau[Y])}\right],$$

and as a (reverse) projection:

$$\Pi_\tau = E\left[\Pi_{\xi_\tau(X,W)}(X,W)\middle|Y = Q_\tau[Y]\right].$$

The first representation is the counterpart of Proposition 1(ii) of Firpo, Fortin, and Lemieux (2009). The second representation appears to be new. It does not rely on any shape or dimension restriction on the structural model $Y = h(X,W,U)$.

Because $F_{Y|X,W}(Q_\tau[Y|x,w]|x,w) = \tau$, by implicitly differentiating, we have

$$\frac{\partial Q_\tau[Y|x,w]}{\partial x} = -\frac{1}{f_{Y|X,W}(Q_\tau[Y|x,w]|x,w)} \frac{\partial F_{Y|X,W}(Q_\tau[Y|x,w]|z,w)}{\partial z}\bigg|_{z=x}.$$

We can then write $\Pi_\tau(x,w)$ in terms of the conditional quantile effect:

$$\Pi_\tau(x,w) = -\frac{\partial Q_\tau[Y|x,w]}{\partial x}\kappa(x) = \frac{\partial Q_\tau[Y|x,w]}{\partial x}\frac{\partial \mathcal{G}(x;\delta)}{\partial \delta}\bigg|_{\delta=0}.$$

Using Theorem 2, we then have

$$\Pi_\tau = -E\left\{\left[\frac{\partial Q_{\xi_\tau(X,W)}[Y|z,W])}{\partial z}|_{z=X}\right]\kappa(X)\frac{f_{Y|X,W}(Q_\tau[Y]|X,W)}{f_Y(Q_\tau[Y])}\right\} \tag{11}$$

and

$$\Pi_\tau = -E\left\{\left[\frac{\partial Q_{\xi_\tau(X,W)}[Y|z,W])}{\partial z}|_{z=X}\right]\kappa(X)\bigg|Y = Q_\tau[Y]\right\}. \tag{12}$$

Hence, Π_τ is a weighted average of quantile derivatives. This suggests an alternative way to estimate Π_τ. The problem of estimating a weighted average of quantile derivatives has been recently studied in Lee (2021). However, our problem is different. In both (11) and (12), the average is taken over a random quantile level $\xi_\tau(X,W)$, instead of a fixed quantile level. The random quantile level arises from the matching function given in (10). We leave this alternative way to estimate Π_τ for future research.

Example 6. *Consider the location-scale shift with no covariate W. Suppose that the conditional quantiles are linear: $Q_\tau[Y|X=x] = a_\tau + xb_\tau$. Then*

$$\frac{\partial Q_{\xi_\tau(X)}[Y|z])}{\partial z}\bigg|_{z=X} = b_{\xi_\tau(X)}$$

and so

$$\Pi_\tau^{\mu_X} = \underbrace{\dot{\ell}(0)E\left[b_{\xi_\tau(X)}|Y = Q_\tau[Y]\right]}_{=\Pi_{\tau,L}}\underbrace{-\dot{s}(0)E\left[b_{\xi_\tau(X)}(X-\mu_X)|Y = Q_\tau[Y]\right]}_{=\Pi_{\tau,S}^{\mu_X}}.$$

In Example 4, we have $b_\tau = \beta$ for every τ, and $\dot{s}(0) = 1$. The pure scale effect is then

$$\Pi_{\tau,S}^{\mu_X} = -\beta E\left[(X-\mu_X)|Y = Q_\tau[Y]\right] = -\beta\frac{Cov(X,Y)}{Var(Y)}(Q_\tau[Y]-\alpha) = -\frac{\beta^2}{\beta^2+1}(Q_\tau[Y]-\alpha),$$

which is what we obtained before.

3 Compensated Marginal Effects

In this section, we consider the case where a location shift in one covariate is compensated by a location shift in another covariate. In a model $Y = h(X_1, X_2, W, U)$ where both X_1 and X_2 are univariate, we consider the limiting effect of the simultaneous location shift $X_{1\delta} = X_1 + \ell_1(\delta)$ and $X_{2\delta} = X_2 + \ell_2(\delta)$ for some smooth functions $\ell_1(\delta)$ and $\ell_2(\delta)$ satisfying $\ell_1(0) = \ell_2(0) = 0$. In the simplest case, we have $\ell_1(\delta) = \delta$ and $\ell_2(\delta) = -p\delta$ for some $p \geq 0$. Here, p can be interpreted as the "relative price" of X_1 in terms of X_2. An example is the following: a policy targeted towards increasing the level of education can, at the same time, reduce the experience of workers. As with the case of the scale shift, neglecting this possible side effect of the policy might lead to an inconsistent estimator of its effect.

With the above motivation, we now consider a more general setting that allows for a general change in X_1 and X_2. We induce a change in $X = (X_1, X_2)'$ so that it becomes $X_\delta = (X_{1\delta}, X_{2\delta})'$. We do not specify the exact form of the change, but we use the simultaneous location shift as a working example. We assume that

$$X_\delta = \mathcal{G}(x;\delta) = (\mathcal{G}_1(X;\delta), \mathcal{G}_2(X;\delta))'$$

for a smooth and invertible bivariate function $\mathcal{G} = (\mathcal{G}_1, \mathcal{G}_2)'$. We allow $X_{1\delta}$ and $X_{2\delta}$ to depend on both X_1 and X_2. A special case is that $\mathcal{G}_1(X;\delta)$ is a function of X_1 only and $\mathcal{G}_2(X;\delta)$ is a function of X_2 only.

In this general setting, the original outcome is given by

$$Y = h(X_1, X_2, W, U) = h(X, W, U),$$

and the counterfactual outcome is given by

$$Y_\delta = h(X_{1\delta}, X_{2\delta}, W, U) = h(\mathcal{G}_1(X;\delta), \mathcal{G}_2(X;\delta), W, U). \tag{13}$$

The distribution of (X, W, U) is kept the same in the above two equations. We want to identify the following quantity

$$\Pi_{\tau,C} := \lim_{\delta \to 0} \frac{Q_\tau[Y_\delta] - Q_\tau[Y]}{\delta}, \tag{14}$$

whenever this limit exists. We refer to $\Pi_{\tau,C}$ as the *compensated marginal effect for the τ-quantile.*

Let $x = (x_1, x_2)'$. As before, we define $x^\delta = \left(x_1^\delta, x_2^\delta\right)'$ such that $\mathcal{G}\left(x^\delta;\delta\right) = x$. By construction, $X_\delta = x$ if and only if $X = x^\delta$. Define the Jacobian matrix as

$$J\left(x^\delta;\delta\right) := \frac{\partial x^\delta}{\partial x'} = \begin{pmatrix} \frac{\partial x_1^\delta}{\partial x_1} & \frac{\partial x_1^\delta}{\partial x_2} \\ \frac{\partial x_2^\delta}{\partial x_1} & \frac{\partial x_2^\delta}{\partial x_2} \end{pmatrix} = \left(\frac{\partial \mathcal{G}(x;\delta)}{\partial x'}\right)^{-1}\Bigg|_{x=x^\delta},$$

where the second equality follows from differentiating $\mathcal{G}\left(x^\delta, \delta\right) = x$ with respect to x and then solving for $\partial x^{\delta}/\partial x'$.

Assumption 3. *(i) For some $\varepsilon > 0$, each component function of $\mathcal{G}(x;\delta)$ is continuously differentiable on $\mathcal{X} \otimes \mathcal{N}_{\varepsilon}$.*

(i.b) $\mathcal{G}(x;\delta)$ is an invertible function of x each $\delta \in \mathcal{N}_{\varepsilon}$.

(i.c) $\mathcal{G}(x;0) = x$ for all $x \in \mathcal{X}$.

(ii) for $\delta \in \mathcal{N}_{\varepsilon}$, the conditional density of U satisfies $f_{U|X_{\delta},W}(u|x,w) = f_{U|X,W}(u|x^{\delta},w)$ and the support $\mathcal{U}$ of U conditional on X and W does not depend on (X,W).

(iii) Assumption 1 (iii.a) holds with $J(x^{\delta};\delta)$ replaced by $\det\left[J(x^{\delta};\delta)\right]$ and Assumption 1 (iii.b) holds.

(iv) $f_{X,W}(x,w)$ is equal to 0 on the boundary of the support of X_1 given $W = w$ and $X_2 = x_2$ for all $w \in \mathcal{W}$ and $x_2 \in \mathcal{X}_2$, the support of X_2, and symmetrically, $f_{X,W}(x,w)$ is equal to 0 on the boundary of the support of X_2 given $W = w$ and $X_1 = x_1$ for all $w \in \mathcal{W}$ and $x_1 \in \mathcal{X}_1$, the support of X_1.

(v) $f_Y(Q_\tau[Y]) > 0$.

Assumption 3 is a modified version of Assumption 1 adapted to the case with two target covariates. Under Assumption 3(i.c), we have $J(x;0) = I_2$, the 2×2 identity matrix. Since $\det[J(x,0)] = 1$, by continuity, $\det\left[J(x^{\delta};\delta)\right] > 0$ when δ is small enough. Hence, there is no need to take the absolute value of $\det\left[J(x^{\delta};\delta)\right]$ when converting the pdf of (X,W) into that of (X_{δ},W).

Define the local change function as

$$\kappa(x) = (\kappa_1(x), \kappa_2(x))' := \left.\frac{\partial x^{\delta}}{\partial \delta}\right|_{\delta=0}.$$

Under Assumption 3, we have

$$\kappa(x) = -\left.\frac{\partial \mathcal{G}(x;\delta)}{\partial \delta}\right|_{\delta=0}.$$

Theorem 3. *Let Assumption 3 hold. Then*

$$\Pi_{\tau,C} = -E\left[\frac{\partial E[\psi(Y,\tau,F_Y)|X,W]}{\partial X'}\kappa(X)\right] + E\left[\psi(Y,\tau,F_Y)\frac{\partial \ln f_{U|X,W}(U|X,W)}{\partial X'}\kappa(X)\right], \quad (15)$$

where, as before,

$$\psi(y,\tau,F_Y) = \frac{\tau - 1(y < Q_\tau[Y])}{f_Y(Q_\tau[Y])}.$$

The theorem takes the same form as Theorem 1. Under the assumption that $X_{j\delta}$ is a function of X_j only for $j = 1$ and 2, $\kappa_j(x)$ depends on x_j only, and the effect from changing X_1 into $X_{1\delta}$ and that from changing X_2 into $X_{2\delta}$ are additively separable.

Corollary 2. *Let Assumptions 2 and 3 hold. Then*

$$\Pi_{\tau,C} = -E\left[\frac{\partial E[\psi(Y,\tau,F_Y)|X,W]}{\partial X'}\kappa(X)\right].$$

For the case of a simultaneous location shift $X_{1\delta} = X_1 + \ell_1(\delta)$ *and* $X_{2\delta} = X_2 + \ell_2(\delta)$*, we have*

$$\kappa(x) = -\left(\dot{\ell}_1(0), \dot{\ell}_2(0)\right)',$$

and so

$$\begin{aligned}\Pi_{\tau,C} = &-\frac{\dot{\ell}_1(0)}{f_Y(Q_\tau[Y])}\int_{\mathcal{W}}\int_{\mathcal{X}}\frac{\partial F_{Y|X,W}(Q_\tau[Y]|x,w)}{\partial x_1}f_{X,W}(x,w)dxdw \\ &-\frac{\dot{\ell}_2(0)}{f_Y(Q_\tau[Y])}\int_{\mathcal{W}}\int_{\mathcal{X}}\frac{\partial F_{Y|X,W}(Q_\tau[Y]|x,w)}{\partial x_2}f_{X,W}(x,w)dxdw.\end{aligned} \tag{16}$$

Corollary 2 shows that the compensated effect from the simultaneous location shift is a linear combination of two location effects: one where the target variable is X_1 and the other where the target variable is X_2. Thus, we can write: $\Pi_{\tau,C} = \Pi_{\tau,L,1} + \Pi_{\tau,L,2}$. This additive result follows because we have two unrelated location changes whose effects are, in essence, captured by the sum of two partial derivatives. This is convenient since it immediately allows us to obtain the bias if we omit the possible simultaneous change in a covariate different from the target variable.

Corollary 1 in Firpo, Fortin, and Lemieux (2009) considers the case of a simultaneous location shift in k covariates, and delivers a $k \times 1$ vector of marginal effects. Theorem 3 and Corollary 2 complement such a result by showing how to interpret a linear combination of the entries of the vector of marginal effects. Furthermore, Theorem 3 and Corollary 2 allow for the intervention of a target covariate to depend on another target covariate. Here we consider only two target covariates for ease of exposition. Our results can be easily extended to the case with more than two target covariates.

Our framework can accommodate more complicated policy interventions, such as simultaneous location-scale shifts in two target variables. In a potential application, a compensated change may substitute the mean of one target variable with the variance of another target variable. Given the generality of $\mathcal{G}(x;\delta)$, Corollary 2 is general enough to accommodate various compensating policies.

4 Estimation of location-scale effects

4.1 Location-scale effects

In this section, we focus on the estimation of Π_τ^μ given in (5). The estimator involves several preliminary steps. Firstly, for a given quantile, we need to estimate $Q_\tau[Y]$. This is given by

$$\hat{q}_\tau = \arg\min_q \sum_{i=1}^n \left(\tau - \mathbb{1}\{Y_i \le q\}\right)(Y_i - q). \tag{17}$$

Next, we need to estimate the density of Y evaluated at $Q_\tau[Y]$. This can be estimated by

$$\hat{f}_Y(\hat{q}_\tau) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_h(Y_i - \hat{q}_\tau) \tag{18}$$

where $\mathcal{K}_h(u) = h^{-1}\mathcal{K}(h^{-1}u)$ for a given kernel $\mathcal{K}$ and a bandwidth h. For the average derivative of the conditional cdf, we propose either a logit model as in Firpo, Fortin, and Lemieux (2009) or a probit model. We model:

$$F_{Y|X,W}(Q_\tau[Y]|x,w) = G(x\alpha_\tau + w'\beta_\tau) \tag{19}$$

where $G(\cdot)$ is either the cdf of a logistic random variable (logit) or a standard normal random variable (probit). Let $Z_i = (X_i', W_i')'$ and $\theta_\tau = (\alpha_\tau', \beta_\tau')'$. We estimate θ_τ by the maximum likelihood estimator:

$$\begin{aligned}\hat{\theta}_\tau &:= (\hat{\alpha}_\tau, \hat{\beta}_\tau')' = \arg\max_{\theta\in\Theta}\sum_{i=1}^{n} l_i(\theta;\hat{q}_\tau)\\ &= \arg\max_{\theta\in\Theta}\sum_{i=1}^{n}\left\{\mathbb{1}\{Y_i \le \hat{q}_\tau\}\log\left[G(Z_i'\theta)\right] + \mathbb{1}\{Y_i > \hat{q}_\tau\}\log\left[1 - G(Z_i'\theta)\right]\right\},\end{aligned} \tag{20}$$

where Θ is a compact parameter space that contains θ_τ as an interior point. The estimator of Π_τ^μ is then

$$\hat{\Pi}_\tau^\mu = \hat{\Pi}_{\tau,L} + \hat{\Pi}_{\tau,S}^\mu$$

where

$$\hat{\Pi}_{\tau,L} = -\frac{\dot{\ell}(0)}{\hat{f}_Y(\hat{q}_\tau)}\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau, \tag{21}$$

$$\hat{\Pi}_{\tau,S}^\mu = \frac{\dot{s}(0)}{\hat{f}_Y(\hat{q}_\tau)}\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau(X_i - \mu). \tag{22}$$

In the above, g is the derivative of G, that is, the logistic density or the standard normal density. In order to establish the asymptotic distribution of $\hat{\Pi}_\tau^\mu$, we need the following three sets of assumptions, one for each preliminary estimation step.

Assumption 4. ***Quantile.*** *The density of Y is positive, continuous, and differentiable at $Q_\tau[Y]$.*

Assumption 5. ***Logit/Probit.*** *For G either the cdf of a logistic or a standard normal random variable, we have*

(i) $F_{Y|Z}(Q_\tau[Y]|z) = G(z'\theta_\tau)$ *for an interior point* $\theta_\tau \in \Theta$ *and* $\hat{\theta}_\tau = \theta_\tau + o_p(1)$.

(ii) For

$$H_i(\theta;q) = \frac{\partial^2 l_i(\theta;q)}{\partial\theta\partial\theta'},$$

the Hessian of observation i, the following holds

$$\sup_{(\theta,q)\in\mathcal{N}} \left\| \frac{1}{n}\sum_{i=1}^{n} H_i(\theta;q) - E[H_i(\theta;q)] \right\| \xrightarrow{p} 0,$$

where $\mathcal{N}$ is a neighborhood of $(\theta_\tau', Q_\tau[Y]')'$, and $H := E[H_i(\theta_\tau; Q_\tau[Y])]$ is negative definite.

(iii) For the score s_i defined by

$$s_i(\theta, q) = \frac{\partial l_i(\theta; q)}{\partial \theta},$$

the following stochastic equicontinuity assumption holds:

$$\frac{1}{n}\sum_{i=1}^{n} \left\{ s_i(\theta_\tau; \hat{q}_\tau) - E\left[s_i(\theta_\tau; q)\right] |_{q=\hat{q}_\tau} \right\} = \frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; Q_\tau[Y]) + o_p(n^{-1/2}),$$

and the map $q \mapsto E\left[s_i(\theta_\tau; q)\right]$ is continuously differentiable at $Q_\tau[Y]$ with

$$\left.\frac{\partial E\left[s_i(\theta_\tau; q)\right]}{\partial q}\right|_{q=Q_\tau[Y]} =: H_Q.$$

(iv) For $\tilde{X}_i = (1, X_i)'$ and $\tilde{Z}_i = (1, Z_i')'$, the following uniform law of large numbers holds:

$$\sup_{\theta\in\mathcal{N}_\theta} \left\| \frac{1}{n}\sum_{i=1}^{n} \dot{g}(Z_i'\theta)\tilde{X}_i\tilde{Z}_i' - E\left[\dot{g}(Z_i'\theta)\tilde{X}_i\tilde{Z}_i'\right] \right\| \xrightarrow{p} 0,$$

where $\mathcal{N}_\theta$ is a neighborhood of θ_τ, $\dot{g}$ is the derivative of g, and

$$M := \begin{pmatrix} E\left[\dot{g}(Z_i'\theta_\tau)\alpha_\tau X_i' + g(Z_i'\theta_\tau)\right] & E\left[\dot{g}(Z_i'\theta_\tau)\alpha_\tau W_i'\right] \\ E\left[\dot{g}(Z_i'\theta_\tau)\alpha_\tau X_i X_i' + g(Z_i'\theta_\tau)X_i'\right] & E\left[\dot{g}(Z_i'\theta_\tau)\alpha_\tau X_i W_i'\right] \end{pmatrix}$$

exists.

Assumption 6. ***Density.***

(i) The kernel function $K(\cdot)$ satisfies (i) $\int_{-\infty}^{\infty} K(u)du = 1$, (ii) $\int_{-\infty}^{\infty} u^2 K(u)du < \infty$, and (iii) $K(u) = K(-u)$, and it is twice differentiable with Lipschitz continuous second-order derivative $K''(u)$ satisfying (i) $\int_{-\infty}^{\infty} K''(u)udu < \infty$ and (ii) there exist positive constants C_1 and C_2 such that $|K''(u_1) - K''(u_2)| \leq C_2 |u_1 - u_2|^2$ for $|u_1 - u_2| \geq C_1$.

(ii) As $n \uparrow \infty$, the bandwidth satisfies: $h \downarrow 0$, $nh^3 \uparrow \infty$, and $nh^5 = O(1)$.

Under Assumption 4, $\hat{q}_\tau$ given in (17) is asymptotically linear with

$$\hat{q}_\tau - Q_\tau[Y] = \frac{1}{n}\sum_{i=1}^{n} \frac{\tau - \mathbb{1}\left\{Y_i \leq Q_\tau[Y]\right\}}{f_Y(Q_\tau[Y])} + o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, \tau, F_Y) + o_p(n^{-1/2}).$$

See, for example, Serfling (1980). Assumption 5 is mostly necessary to deal with the preliminary estimator $\hat{q}_\tau$ that enters the likelihood in (20). Assumption 6 is taken from Martinez-Iriarte and Sun (2021b).

The following lemma contains the influence function for the maximum likelihood estimator $\hat{\theta}_\tau$.

Lemma 1. *Under Assumptions 4 and 5, we have*

$$\hat{\theta}_\tau - \theta_\tau = -H^{-1}\frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; Q_\tau[Y]) - H^{-1}H_Q\frac{1}{n}\sum_{i=1}^{n}\psi(Y_i, \tau, F_Y) + o_p(n^{-1/2}).$$

Theorem 4. *Under Assumptions 4, 5, and 6, the estimators given in* (21) *and* (22) *satisfy*

$$\begin{pmatrix} \hat{\Pi}_{\tau,L} \\ \hat{\Pi}^{\mu}_{\tau,S} \end{pmatrix} - \begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}\Phi_{i,\tau} + O\left(h^2\right) + o_p(n^{-1/2}) + o_p(n^{-1/2}h^{-1/2}),$$

where

$$\begin{aligned} \Phi_{i,\tau} = & \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_\mu \left[g(Z_i'\theta_\tau)\alpha_\tau \tilde{X}_i - Eg(Z_i'\theta_\tau)\alpha_\tau \tilde{X}_i \right] \\ & - \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_\mu M H^{-1} s_i(\theta_\tau; Q_\tau[Y]) \\ & - \left[\begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix} \frac{\dot{f}_Y(Q_\tau[Y])}{f_Y\left(Q_\tau[Y]\right)} + \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_\mu M H^{-1} H_Q \right] \psi(Y_i, \tau, F_Y) \\ & - \begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix} \frac{1}{f_Y\left(Q_\tau[Y]\right)} \left\{ \mathcal{K}_h\left(Y_i - Q_\tau[Y]\right) - E\mathcal{K}_h\left(Y_i - Q_\tau[Y]\right) \right\}, \end{aligned}$$

$\dot{f}_Y(\cdot)$ *is the derivative of* $f_Y(\cdot)$, *and*

$$D_\mu = \begin{pmatrix} D_L' \\ D_{\mu,S}' \end{pmatrix} = \begin{pmatrix} -\dot{\ell}(0) & 0 \\ -\mu\dot{s}(0) & \dot{s}(0) \end{pmatrix}.$$

Theorem 4 establishes the contribution from each estimation step. In particular, the last term in $n^{-1}\sum_{i=1}^{n}\Phi_{i,\tau}$ is the contribution from estimating the density of Y non-parametrically. This term converges at a non-parametric rate, which is slower than other terms. As a result, the asymptotic distribution of the location-scale effect estimator is determined by the last term in $n^{-1}\sum_{i=1}^{n}\Phi_{i,\tau}$. However, we do not recommend dropping all other terms. Instead, we write the asymptotic normality result in the form

$$\left[\frac{1}{n^2}\sum_{i=1}^{n}\hat{\Phi}_{i,\tau}\hat{\Phi}_{i,\tau}'\right]^{-1/2}\left[\begin{pmatrix} \hat{\Pi}_{\tau,L} \\ \hat{\Pi}^{\mu}_{\tau,S} \end{pmatrix} - \begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix}\right] \xrightarrow{d} N(0, I_2) \tag{23}$$

as $n \uparrow \infty$, $nh^3 \uparrow \infty$, and $nh^5 \downarrow 0$ where $\hat{\Phi}_{i,\tau}$ is a plug-in estimator of $\Phi_{i,\tau}$. In particular,

$$\left[n^{-2}\sum_{i=1}^{n}(l_1'\hat{\Phi}_{i,\tau})^2\right]^{-1/2}\left(\hat{\Pi}_{\tau,L}-\Pi_{\tau,L}\right)\xrightarrow{d}N(0,1),$$
$$\left[n^{-2}\sum_{i=1}^{n}(l_2'\hat{\Phi}_{i,\tau})^2\right]^{-1/2}\left(\hat{\Pi}^{\mu}_{\tau,S}-\Pi^{\mu}_{\tau,S}\right)\xrightarrow{d}N(0,1), \tag{24}$$

where $l_1=(1,0)'$ and $l_2=(0,1)'$. The above results hold under some additional but standard regularity conditions such as the nonsingularity of the probability limit of $n^{-2}\sum_{i=1}^{n}\hat{\Phi}_{i,\tau}\hat{\Phi}_{i,\tau}'$. Inferences based on these results account for the estimation errors from all estimation steps and are more reliable in finite samples. This is supported by simulation evidence not reported here. On the other hand, if we parametrize the density of Y and estimate it at the parametric $\sqrt{n}$-rate, then the last term in $n^{-1}\sum_{i=1}^{n}\Phi_{i,\tau}$ will take a different form and will be of the same order as the other terms. In this case, the location-scale effect estimator is $\sqrt{n}$-asymptotically normal, and all the terms in Theorem 4 will contribute to the asymptotic variance. With an obvious modification of the last term in $\Phi_{i,\tau}$, the asymptotic normality can be presented in the same way as in (23).

Let

$$\Gamma^{\mu}_{\tau,S}=D'_{\mu,S}E\left[\frac{\partial F_{Y|X,W}(Q_\tau[Y]|X,W)}{\partial X}(X-\mu)\right]$$

be the numerator of $\Pi^{\mu}_{\tau,S}$. Then the scale effect $\Pi^{\mu}_{\tau,S}$ is zero if and only if $\Gamma^{\mu}_{\tau,S}=0$. To test the null hypothesis $H_0:\Pi^{\mu}_{\tau,S}=0$, we can equivalently test the null hypothesis $H_0:\Gamma^{\mu}_{\tau,S}=0$. Unlike $\Pi^{\mu}_{\tau,S}$, $\Gamma^{\mu}_{\tau,S}$ can be estimated at the parametric rate even if $f_Y(\cdot)$ is not parametrically specified. More specifically, under Assumption 5, we can estimate $\Gamma^{\mu}_{\tau,S}$ by

$$\hat{\Gamma}^{\mu}_{\tau,S}:=D'_{\mu,S}\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i,$$

where $D'_{\mu,S}=(-\mu,1)$ upon setting $\dot{s}(0)=1$ without loss of generality.

Under the assumptions of Theorem 4, we can show that

$$\hat{\Gamma}^{\mu}_{\tau,S}-\Gamma^{\mu}_{\tau,S}=D'_{\mu,S}\frac{1}{n}\sum_{i=1}^{n}\Phi^{\Gamma}_{i,\tau}+o_p\left(\frac{1}{\sqrt{n}}\right),$$

where

$$\Phi^{\Gamma}_{i,\tau}=g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i-E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]$$
$$-MH^{-1}s_i(\theta_\tau;Q_\tau[Y])-MH^{-1}H_Q\psi(Y_i,\tau,F_Y).$$

Define

$$V_\tau=\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}E(D'_{\mu,S}\Phi^{\Gamma}_{i,\tau})^2.$$

Under some regularity conditions such as $V_\tau > 0$, we have $V_\tau^{-1/2}(\hat{\Gamma}^{\mu}_{\tau,S} - \Gamma^{\mu}_{\tau,S}) \xrightarrow{d} N(0,1)$. To test $H_0 : \Gamma^{\mu}_{\tau,S} = 0$, we construct the test statistic

$$t^{\mu}_{\tau,S} := \frac{\hat{\Gamma}^{\mu}_{\tau,S}}{\sqrt{\hat{V}_\tau}} \text{ for } \hat{V}_\tau = \frac{1}{n^2}\sum_{i=1}^{n}(D'_{\mu,S}\hat{\Phi}^{\Gamma}_{i,\tau})^2,$$

where

$$\hat{\Phi}^{\Gamma}_{i,\tau} = g(Z_i\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i - \frac{1}{n}\sum_{i=1}^{n} g(Z_i\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i - \hat{M}\hat{H}^{-1}s_i(\hat{\theta}_\tau;\hat{q}_\tau) - \hat{M}\hat{H}^{-1}\hat{H}_Q\hat{\psi}(Y_i,\tau,F_Y). \tag{25}$$

In the above, $\hat{\psi}(Y_i,\tau,F_Y) = [\tau - 1\{Y_i \le \hat{q}_\tau\}]/\hat{f}_Y(\hat{q}_\tau)$ and the score $s_i(\hat{\theta}_\tau;\hat{q}_\tau)$ is obtained by evaluating the expression given in (A.7) at $\theta = \hat{\theta}_\tau$ and $q = \hat{q}_\tau$. $\hat{M}$, $\hat{H}$, and $\hat{H}_Q$ are the sample versions of M, H, and H_Q, respectively. Details are given in the proof of the corollary below.

Corollary 3. *Let the assumptions of Theorem 4 hold. Assume that* $V_\tau^{-1/2}(\hat{\Gamma}^{\mu}_{\tau,S} - \Gamma^{\mu}_{\tau,S}) \xrightarrow{d} N(0,1)$ *for some* $V_\tau > 0$ *and* $\hat{V}_\tau/V_\tau \xrightarrow{p} 1$. *Then, under the null hypothesis* $H_0 : \Pi^{\mu}_{\tau,S} = 0$,

$$t^{\mu}_{\tau,S} \xrightarrow{d} N(0,1).$$

4.2 Compensated Effects

In this section, we focus on the estimation of $\Pi_{\tau,C}$ given in (16). We use the same estimators of the quantile, the density of Y, and the parameters in the probit/logit model. We only need to make some minor notational changes. As before $\theta_\tau = (\alpha'_\tau, \beta'_\tau)'$, $\hat{\theta}_\tau = (\hat{\alpha}'_\tau, \hat{\beta}'_\tau)'$ and $Z_i = (X'_i, W'_i)'$ but now $\alpha_\tau = (\alpha_{\tau,1}, \alpha_{\tau,2})'$, $\hat{\alpha}_\tau = (\hat{\alpha}_{\tau,1}, \hat{\alpha}_{\tau,2})'$ and $X_i = (X'_{1i}, X'_{2i})'$. As in the case with the location-scale effect, we estimate $\Pi_{\tau,C}$ by

$$\hat{\Pi}_{\tau,C} = \hat{\Pi}_{\tau,L,1} + \hat{\Pi}_{\tau,L,2}$$

where

$$\hat{\Pi}_{\tau,L,1} = -\frac{\dot{\ell}_1(0)}{\hat{f}_Y(\hat{q}_\tau)}\frac{1}{n}\sum_{i=1}^{n} g(Z'_i\hat{\theta}_\tau)\hat{\alpha}_{\tau,1}, \tag{26}$$

$$\hat{\Pi}_{\tau,L,2} = -\frac{\dot{\ell}_2(0)}{\hat{f}_Y(\hat{q}_\tau)}\frac{1}{n}\sum_{i=1}^{n} g(Z'_i\hat{\theta}_\tau)\hat{\alpha}_{\tau,2}. \tag{27}$$

For the next theorem, we need the following modification of Assumption 5.

Assumption 7. ***Logit/Probit II.*** *Assumption 5 holds with (iv) replaced by the following:*

$$\sup_{\theta\in\mathcal{N}_\theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\dot{g}(Z'_i\theta)Z'_i - E\left[\dot{g}(Z'_i\theta)Z'_i\right]\right\| \xrightarrow{p} 0,$$

$$\sup_{\theta\in\mathcal{N}_\theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\dot{g}(Z'_i\theta) - E\left[\dot{g}(Z'_i\theta)\right]\right\| \xrightarrow{p} 0,$$

where $\mathcal{N}_\theta$ is a neighborhood of θ_τ and

$$M_L = \left(E\left[\dot{g}(Z_i'\theta_\tau)\alpha_\tau X_i' + g(Z_i'\theta_\tau)\right], \quad E\left[\dot{g}(Z_i'\theta)\alpha_\tau W_i'\right] \right)$$

exists.

Theorem 5. *Under Assumptions 4, 6, and 7, the estimators given in (26) and (27) satisfy*

$$\begin{pmatrix} \hat{\Pi}_{\tau,L,1} \\ \hat{\Pi}_{\tau,L,2} \end{pmatrix} - \begin{pmatrix} \Pi_{\tau,L,1} \\ \Pi_{\tau,L,2} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \Phi_{i,\tau}^{L} + O\left(h^2\right) + o_p(n^{-1/2}) + o_p(n^{-1/2}h^{-1/2}),$$

where

$$\begin{aligned} \Phi_{i,\tau}^{L} = & \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_L g(Z_i'\theta_\tau)\alpha_\tau - E\left[g(Z_i'\theta_\tau)\alpha_\tau\right] \\ & - \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_L M_L H^{-1} s_i(\theta_\tau; Q_\tau[Y]) \\ & - \left[\begin{pmatrix} \Pi_{\tau,L,1} \\ \Pi_{\tau,L,2} \end{pmatrix} \frac{\dot{f}_Y(Q_\tau[Y])}{f_Y\left(Q_\tau[Y]\right)} + \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_L M_L H^{-1} H_Q \right] \psi(Y_i, \tau, F_Y) \\ & - \begin{pmatrix} \Pi_{\tau,L,1} \\ \Pi_{\tau,L,2} \end{pmatrix} \frac{1}{f_Y\left(Q_\tau[Y]\right)} \left\{ \mathcal{K}_h\left(Y_i - Q_\tau[Y]\right) - E\mathcal{K}_h\left(Y_i - Q_\tau[Y]\right) \right\} \end{aligned}$$

and

$$D_L = \begin{pmatrix} -\dot{\ell}_1(0) & 0 \\ 0 & -\dot{\ell}_2(0) \end{pmatrix}.$$

For the asymptotic normality, the discussions after Theorem 4 are still applicable.

In the special case that $\ell_1(\delta) = \delta$ and $\ell_2(\delta) = -p\delta$, it suffices to change D_L to $\mathrm{diag}(1, -p)$. It is possible that p, the relative price X_1 in terms of X_2, has to be estimated by $\hat{p}$ based on an independent sample. In that case, the estimator of the compensated effect would be

$$\hat{\Pi}_{\tau,L} = \hat{\Pi}_{\tau,L,1} - \hat{p}\hat{\Pi}_{\tau,L,2}.$$

If the sample size $\tilde{n}$ of the independent sample for estimating p is much larger than n (i.e., $\tilde{n}/n \to \infty$), then the expansion in Theorem 5 still holds.

5 Monte Carlo experiments

In this section, we use Monte Carlo simulations to evaluate the finite sample performances of the proposed estimators and tests of location and scale effects. We employ the same data generating process as in Examples 4 and 5 for which we have derived the closed-form expressions for the location and scale effects. In particular, we let

$$Y = \alpha + X\beta + U,$$

where $X \sim N(\mu_X, \sigma_X^2)$ and $U \sim N(0,1)$. We set $\alpha = 0$ and $\dot{s}(0) = \dot{\ell}(0) = 1$. Then, from the results in Examples 4 and 5, the true location effect is $\Pi_{\tau,L} = \beta$, and the true scale effect is

$$\Pi_{\tau,S}^{\mu_X} = -\frac{\sigma_X^2 \beta^2}{\sqrt{\sigma_X^2 \beta^2 + 1}} Q_\tau[U].$$

We consider quantiles $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and sample sizes $n = 500$ and $n = 1000$. The number of simulations is set to $10,000$ for each experiment.

We implement our estimators in `Matlab`. The unconditional quantile estimator in equation (17) is easily computed as an order statistic. The density function is estimated as a kernel density estimator as in equation (18) using a standard normal kernel. For the bandwidth choice in the kernel density estimation, we use a modified version of Silverman's rule of thumb. More specifically, since we require $nh^3 \uparrow \infty$ and $nh^5 \downarrow 0$ as $n \uparrow \infty$, we take $h = 1.06\hat{\sigma}_Y n^{-1/4}$, where $\hat{\sigma}_Y$ is the sample standard deviation of Y.

5.1 Bias, variance and mean squared error

In this subsection, we consider the biases, variances, and mean-squared errors of the proposed location and scale effects estimators. For each effect estimator, we consider either a probit or a logit specification for the conditional cdf $F_{Y|X}(Q_\tau[Y]|X)$. Under our data generating process, the probit for $F_{Y|X}(Q_\tau[Y]|X)$ is correctly specified while the logit is misspecified.

The bias, variance, and mean-squared error are reported in Table 1 when $\mu_X = 0$, $\beta = 1$ and $\sigma_X^2 = 1$ so that the true location effect is 1 for any τ and the true scale effect is $-0.707 Q_\tau[U]$. To save space, simulation results for other values of β and σ_X^2 are omitted.

Table 1 shows that the effect estimator based on the probit specification outperforms that based on the logit one. This is consistent with the correct specification of probit. For each estimator, the bias decreases as the sample size n increases. The variance also decreases as the sample size n increase, and as a result, the MSE also becomes smaller when the sample size grows. For our purposes, the scale-effect estimator performs well. For non-central quantiles, the difference in the scale-effect estimates under the probit and logit specifications is in general larger than the difference in the location-effect estimates. For central quantiles, the probit and logit specifications lead to more or less the same estimates for both the scale effect and the location effect.

5.2 Accuracy of the normal approximation

In this subsection, we investigate the finite sample accuracy of the normal approximation given in (24). Using the same data generating process as in the previous subsection and employing the

Table 1: The biases, variances, and mean-squared errors of the location and scale effects estimators with $\beta = 1$ and $\sigma_X^2 = 1$.

		$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
				$n = 500$		
Bias	Π_L (probit)	-0.015	0.013	0.023	0.012	-0.016
	Π_L (logit)	-0.016	0.012	0.023	0.012	-0.016
	Π_S (probit)	-0.008	0.008	0.000	-0.007	0.008
	Π_S (logit)	0.039	0.034	0.000	-0.034	-0.039
Variance	Π_L (probit)	0.019	0.010	0.008	0.010	0.019
	Π_L (logit)	0.019	0.010	0.008	0.010	0.020
	Π_S (probit)	0.032	0.007	0.003	0.008	0.033
	Π_S (logit)	0.033	0.007	0.003	0.008	0.034
MSE	Π_L (probit)	0.019	0.010	0.009	0.010	0.019
	Π_L (logit)	0.020	0.011	0.009	0.010	0.020
	Π_S (probit)	0.033	0.007	0.003	0.008	0.033
	Π_S (logit)	0.035	0.009	0.003	0.009	0.035
				$n = 1000$		
Bias	Π_L (probit)	-0.011	0.009	0.017	0.008	-0.013
	Π_L (logit)	-0.011	0.009	0.017	0.008	-0.013
	Π_S (probit)	-0.007	0.005	-0.000	-0.004	0.010
	Π_S (logit)	0.041	0.032	-0.000	-0.031	-0.038
Variance	Π_L (probit)	0.011	0.006	0.005	0.006	0.011
	Π_L (logit)	0.011	0.006	0.005	0.006	0.011
	Π_S (probit)	0.018	0.004	0.001	0.004	0.017
	Π_S (logit)	0.018	0.004	0.001	0.004	0.018
MSE	Π_L (probit)	0.011	0.006	0.005	0.006	0.011
	Π_L (logit)	0.011	0.006	0.005	0.006	0.011
	Π_S (probit)	0.018	0.004	0.001	0.004	0.017
	Π_S (logit)	0.020	0.005	0.001	0.005	0.019

probit specification, we simulate the distributions of the studentized statistics

$$\left[n^{-2} \sum_{i=1}^{n} (l_1' \hat{\Phi}_{i,\tau})^2 \right]^{-1/2} (\hat{\Pi}_{\tau,L} - \Pi_{\tau,L})$$

and

$$\left[n^{-2} \sum_{i=1}^{n} (l_2' \hat{\Phi}_{i,\tau})^2 \right]^{-1/2} (\hat{\Pi}^{\mu}_{\tau,S} - \Pi^{\mu}_{\tau,S}).$$

We plot each distribution and compare it with the standard normal distribution. We consider $\beta \in \{0.25, 0.50, 0.75, 1\}$ and the same values τ as in the previous subsection. Simulation results for the two sample sizes $n = 500$, and $n = 1000$ are qualitatively similar, and we report only the case when $n = 1000$ here. Figures 1–4 report the (simulated) finite sample distributions when $\sigma_X^2 = 1$ and $n = 1000$ for some selected values of β and τ together with a standard normal density that is superimposed on each figure. It is clear from these figures that the standard normal distribution provides an accurate approximation to the distribution of the studentized test statistic for both the location and scale effects.

Table 2 reports the empirical coverage of 95% confidence intervals for the location and scale effects. The empirical coverage is close to the nominal coverage in all cases. This is consistent with Figures 1–4. We may then conclude that the normal approximation can be reliably used for

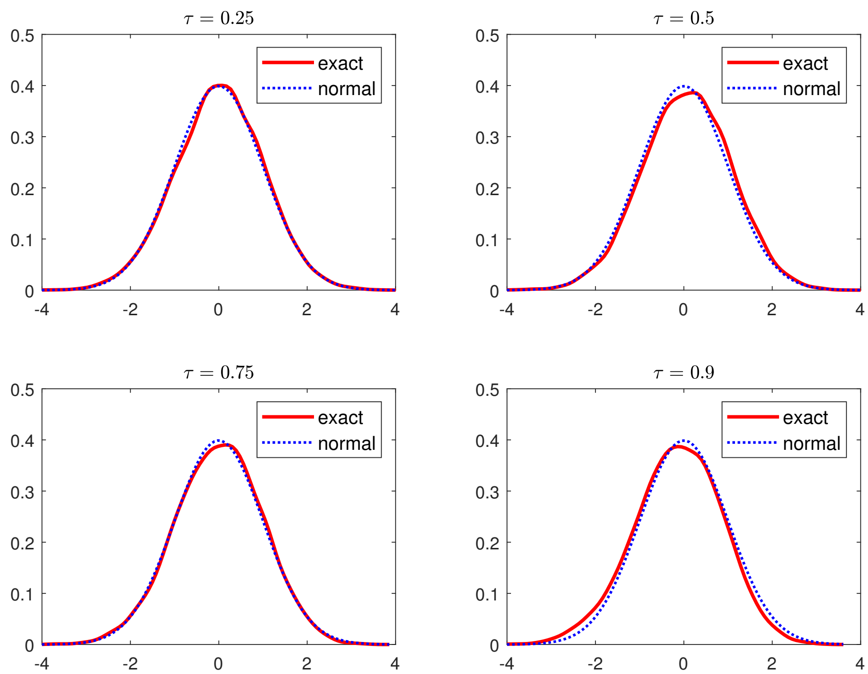

Figure 1: Finite sample exact distribution of the studentized location effect statistic when $\beta = 0.25$, $\sigma_X^2 = 1$, and $n = 1000$.

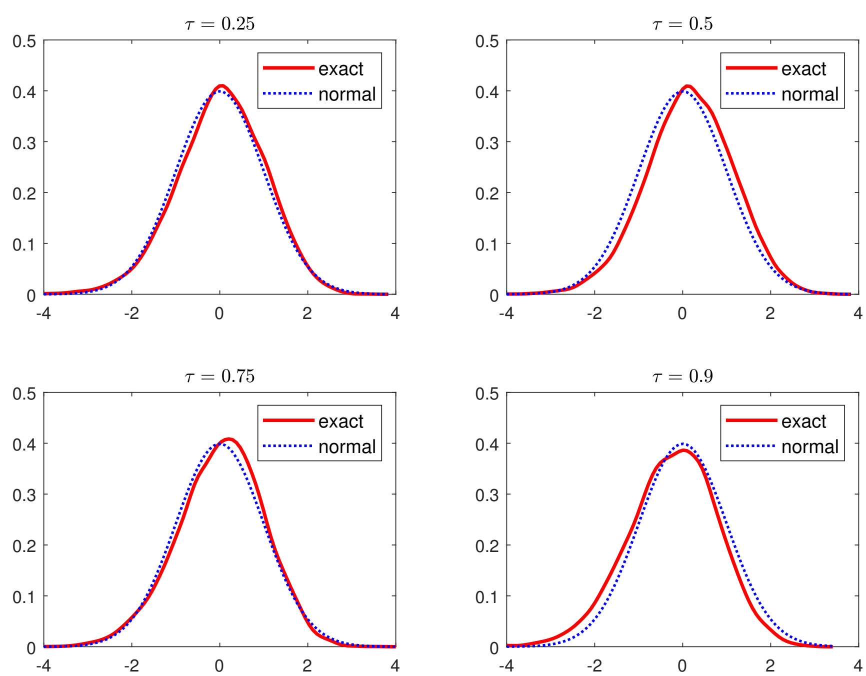

Figure 2: Finite sample exact distribution of the studentized location effect statistic when $\beta = 0.75$, $\sigma_X^2 = 1$, and $n = 1000$.

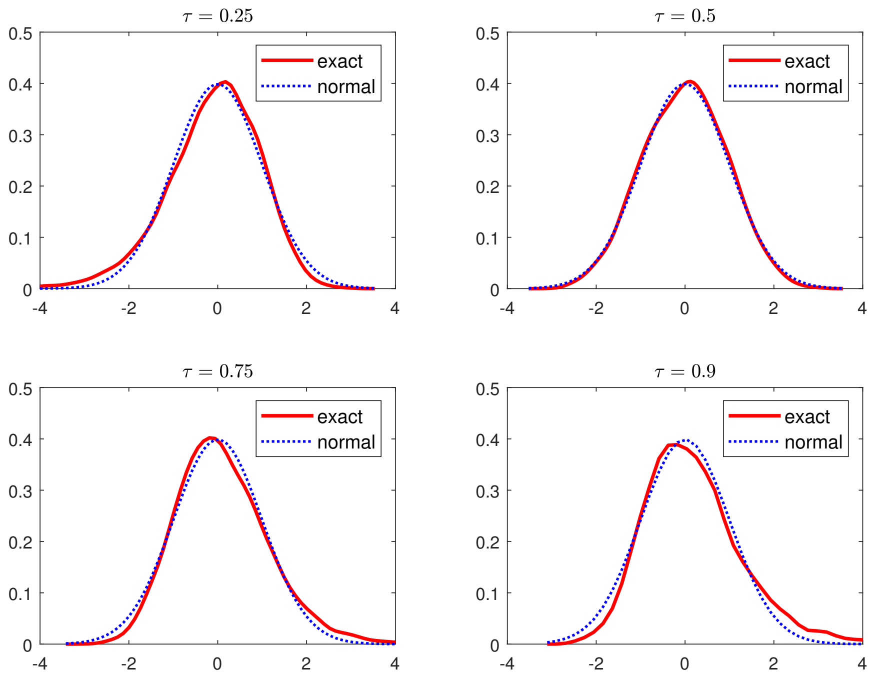

Figure 3: Finite sample exact distribution of the studentized scale effect statistic when $\beta = 0.25$, $\sigma_X^2 = 1$, and $n = 1000$.

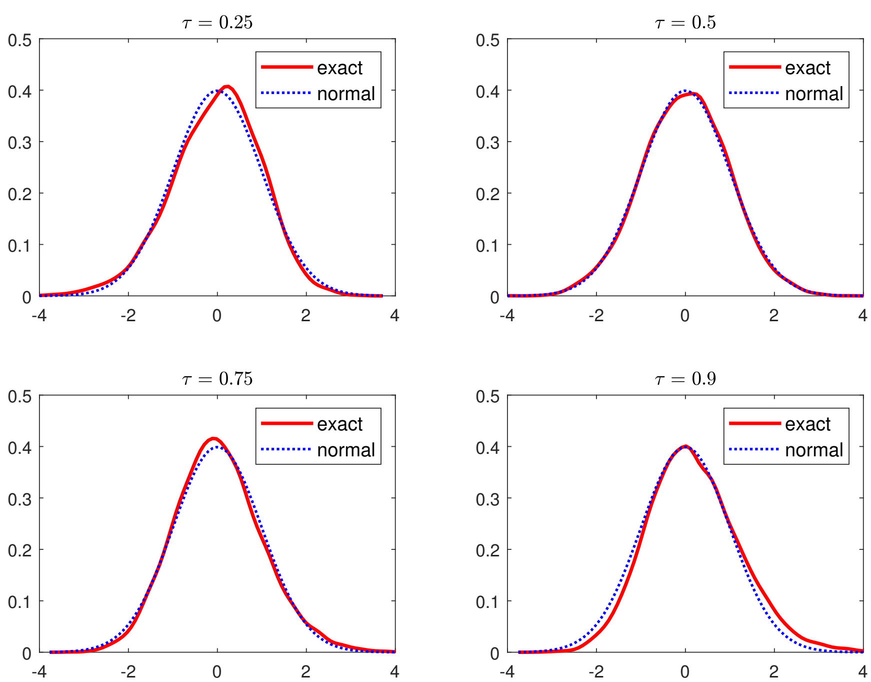

Figure 4: Finite sample exact distribution of the studentized scale effect statistic when $\beta = 0.75$, $\sigma_X^2 = 1$, and $n = 1000$.

Table 2: Empirical coverage of 95% confidence intervals for the location and scale effects when $\sigma_X^2 = 1$.

	β	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
	$n = 500$					
Location	0.25	0.946	0.950	0.951	0.950	0.947
	0.5	0.942	0.952	0.950	0.953	0.938
	0.75	0.940	0.954	0.952	0.956	0.937
	1	0.937	0.957	0.950	0.957	0.935
Scale	0.25	0.900	0.921	0.973	0.916	0.902
	0.5	0.930	0.943	0.957	0.939	0.928
	0.75	0.937	0.950	0.954	0.946	0.933
	1	0.939	0.952	0.951	0.945	0.933
	$n = 1000$					
Location	0.25	0.948	0.951	0.951	0.954	0.945
	0.5	0.946	0.950	0.952	0.957	0.943
	0.75	0.945	0.952	0.953	0.957	0.940
	1	0.941	0.952	0.952	0.958	0.942
Scale	0.25	0.922	0.939	0.965	0.940	0.921
	0.5	0.938	0.949	0.955	0.950	0.933
	0.75	0.942	0.951	0.952	0.952	0.938
	1	0.939	0.952	0.950	0.953	0.940

making inference on the location and scale effects.

5.3 Power of the t-test of a zero scale effect

To investigate the power of the t-test proposed in Corollary 3, we simulate the following model:

$$Y = \alpha + X\beta + U,$$

where

$$\begin{pmatrix} X \\ U \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

Here we set $\alpha = 0$, $\mu_X = 1$ and $\dot{s}(0) = 1$. When $\beta = 0$, X is excluded from the outcome equation and thus the scale effect is 0. The null hypothesis of a zero scale effect corresponds to the case that $\beta = 0$. The power of the test is obtained by varying β around 0 in a grid from -0.4 to 0.4 with an increment of 0.01.

Figure 5 graphs the size-adjusted power of the t-test for different quantile levels when $n = 500$ and when $n = 1000$. The power is calculated using the probit specification of $F_{Y|X}(Q_\tau[Y]|X)$. The size adjustment is based on the empirical critical value such that the test rejects the null 5% of the time. Figure 5 shows that the power increases as β deviates more from its null value of zero, and that for a given nonzero value of β, the power increases with the sample size. Results not reported here show that the test has a quite accurate size in that the empirical rejection probability under the null is close to 5%, the nominal level of the test.

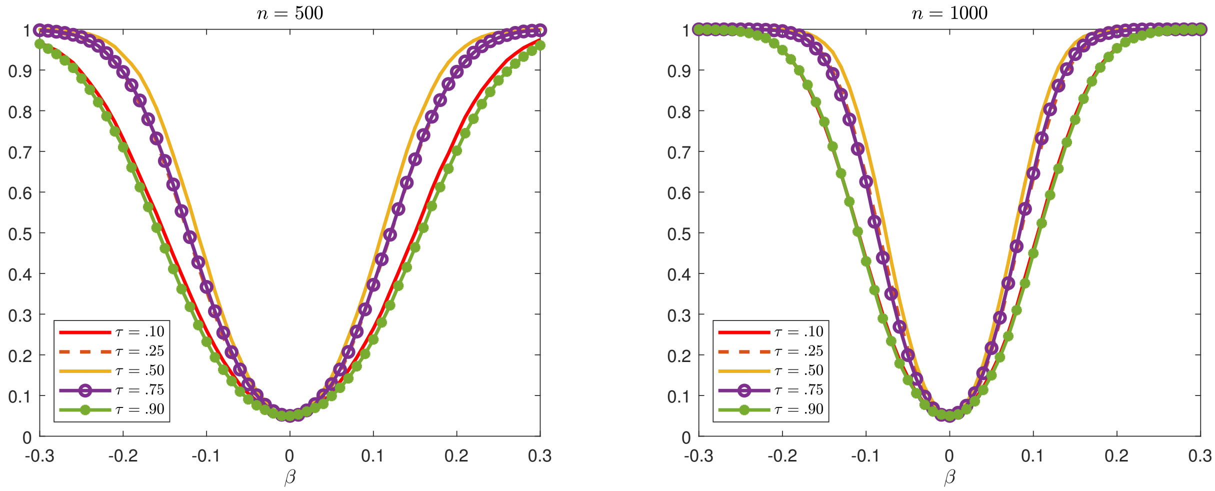

Figure 5: Size-adjusted power of the t-test for a zero scale effect.

6 Empirical application

In order to illustrate the proposed approach, we use a household labor survey from Wooldridge (2002) that can be accessed online for replication.[3] The idea is to evaluate the effects of *education* on the quantile of the unconditional distribution of *log wage*. In this application, $Y = lwage$, which is log hourly wage, and $X = educ$, which is years of education. The controls are: $W = [exper\ tenure\ nonwhite\ female]$, where *exper* is years of working experience, *tenure* is years with current employer, *nonwhite* is a dummy that equals 1 if the individual is non-white, and *female* is a dummy that equals 1 if the individual is female. We assume that Assumption 2 holds for this choice of W.

While the main goal is to study the scale effect, we also present results for the location effect. For the mean of years of education μ_X, we let $\mu_X = 12.29$ based on the Barro-Lee Data on Educational Attainment.[4] We set $\mu = \mu_X = 12.29$ to study the location and scale effects. In a similar fashion to the Monte Carlo analysis, we consider $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. The sample size for the household labor survey is $n = 526$, which is comparable to $n = 500$ in the simulation exercises. We compute the standard errors using the approximation in (24).

The most interesting results in Table 3 appear in the unconditional scale effects. As discussed in Section 2.2, the scale effects can be interpreted as percentage changes of the unconditional quantiles. Consider the scale effect for $\tau = 0.10$. Both the probit and logit specifications suggest an effect of about .045. Then, using the quantile-standard deviation elasticity, a 1% decrease in the standard deviation of education would produce a positive effect of .045% on the unconditional quantile at the quantile level $\tau = 0.10$. Given that the sample standard deviation of *educ* is 2.77,

[3]See http://fmwww.bc.edu/ec-p/data/wooldridge/wage1.des and http://fmwww.bc.edu/ec-p/data/wooldridge/wage1.dta for the data in the Stata data file format.

[4]The dataset is available from https://databank.worldbank.org/reports.aspx?source=EducationStatistics We use the series "Barro-Lee: Average years of total schooling, age 25+, total" for the US between 1970-2010 and find that the average years of schooling is 12.29.

Table 3: Effects of location-scale shifts in education on the unconditional quantiles of log-wage.

		$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = .90$
Location (probit)	Estimate	0.039	0.062	0.101	0.101	0.118
		(0.008)	(0.011)	(0.015)	(0.016)	(0.021)
	95% CI_L	0.025	0.041	0.072	0.069	0.076
	95% CI_U	0.054	0.083	0.129	0.132	0.160
Location (logit)	Estimate	0.038	0.065	0.103	0.100	0.120
		(0.007)	(0.010)	(0.015)	(0.016)	(0.021)
	95% CI_L	0.024	0.044	0.074	0.069	0.080
	95% CI_U	0.053	0.085	0.131	0.132	0.160
Scale (probit)	Estimate	0.045	0.029	-0.025	-0.103	-0.203
		(0.014)	(0.011)	(0.013)	(0.028)	(0.065)
	95% CI_L	0.018	0.007	-0.051	-0.158	-0.330
	95% CI_U	0.071	0.052	0.001	-0.049	-0.077
Scale (logit)	Estimate	0.045	0.034	-0.024	-0.110	-0.227
		(0.014)	(0.012)	(0.014)	(0.029)	(0.066)
	95% CI_L	0.017	0.011	-0.051	-0.167	-0.356
	95% CI_U	0.072	0.058	0.002	-0.053	-0.099

Notes: Standard errors are in parenthesis.

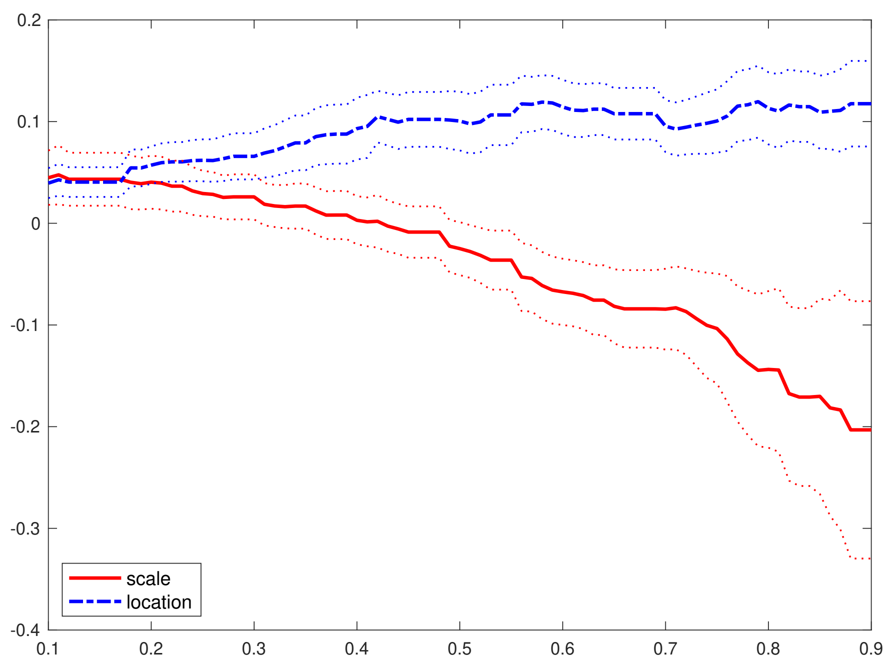

Figure 6: Point and interval estimates of location and scale effects of education on the unconditional quantiles of log-wage based on the probit specification: $\Pi_{\tau,S}$ (solid red) and $\Pi_{\tau,L}$ (dashed blue).

the 1% decrease is approximately a change in the standard deviation from 2.77 to 2.74. Consider now the scale effect for $\tau = 0.50$. In this case, both probit and logit specifications provide a statistically insignificant effect (at the 5% level). Confront this with the results of Examples 4 and 5 where in the linear model $Y = \alpha + X\beta + U$, the scale effect is proportional to $Q_\tau[U]$. Thus, $\hat{\Pi}_{0.50,S} \approx 0$ is consistent with a linear model and U symmetric around 0. Finally, consider the scale effect for $\tau = 0.90$, again using both probit and logit specifications. In this case, the effects are negative, suggesting a 1% decrease in the standard deviation would reduce the upper $\tau = 0.90$ quantile by .20% (probit) and .23% (logit). Overall this analysis shows that the scale effects are monotonically decreasing in τ. This can be seen in Figure 6 that plots, for a finer grid of τ,[5] the probit estimates for both the location (dashed blue) and scale (solid red) effects.

How can this be interpreted? The location effects suggest that the marginal contribution of one more year of education benefits more the upper parts of the unconditional distribution of wages. The scale effects suggest the contrary. Reducing the overall dispersion of education would increase the lower quantile wages, but reduce the upper ones.

7 Conclusion

This paper has provided a general procedure to analyze the distributional impact of changes in covariates on an outcome variable. The standard unconditional quantile regression analysis focuses on a particular impact coming from a pure location shift. We study a more general location-scale model and show how to additively decompose the total effect into a location effect and a scale effect. They can be separately analyzed and estimated. To complement the existing results, we focus on how to define and estimate a change in the scale of a covariate. Additionally, we consider the case of compensated location changes in different covariates. We show how this can be obtained from the usual vector-valued unconditional quantile regressions. More generally, we have provided a framework to study the unconditional policy effects generated by a smooth and invertible intervention of one or more target variables.

References

Autor, D. H., L. S. Katz, and M. S. Kearney (2005): "Rising wage inequality: The role of composition and prices," NBER Working Paper 11628.

Casella, G., and R. L. Berger (2001): *Statistical Inference, 2nd. edition*. Duxbury, Pacific Grove, CA.

Engle, R., and S. Manganelli (2004): "Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economics Statistics*, 22(4), 367–381.

[5] For Figure 6 we use $\tau = 0.10, 0.11, ..., 0.89, 0.90$.

Firpo, S., N. Fortin, and T. Lemieux (2009): "Unconditional quantile regression," *Econometrica*, 77(3), 953–973.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition methods in economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4, pp. 1–12. Amsterdam: Elsevier.

Gu, G. W., S. Malik, D. Pozzoli, and V. Rocha (2019): "Trade-induced Skill Polarization," *Economic Inquiry*, 58(1), 241–259.

Hanushek, E. A., and L. Woessmann (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 3(46), 607–668.

Hsu, Y.-C., T.-C. Lai, and R. P. Lieli (2020): "Counterfactual Treatment Effects: Estimation and Inference," *Journal of Business and Economic Statistics*, Forthcoming.

Inoue, A., T. Li, and Q. Xu (2021): "Two Sample Unconditional Quantile Effect," ARXIV: https://arxiv.org/pdf/2105.09445.pdf.

Lee, Y.-Y. (2021): "Nonparametric Weighted Average Quantile Derivative," *Econometric Theory*, pp. 1–39.

Machado, J. A. F., and J. Mata (1995): "Counterfactual decomposition of changes in wage distributions using quantile," *Journal of Applied Econometrics*, 20, 445–465.

Martinez-Iriarte, J. (2020): "Sensitivity Analysis in Unconditional Quantile Effects," Working Paper.

Martinez-Iriarte, J., and Y. Sun (2021a): "Characterizing Asymptotic Biases of Unconditional Regression Estimators of Policy Effects Under Endogeneity," Working Paper.

——— (2021b): "Identification and Estimation of Unconditional Policy Effects of an Endogenous Binary Treatment," Working Paper.

Melly, B. (2005): "Decomposition of differences in distribution using quantile regressions," *Labour Economics*, 12, 577–590.

Rothe, C. (2012): "Partial Distributional Policy Effects," *Econometrica*, 80(5), 2269–2301.

Sasaki, Y., T. Ura, and Y. Zhang (2020): "Unconditional Quantile Regression with High Dimensional Data," Working Paper.

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Spini, P. (2021): "Robustness, Heterogeneous Treatment Effects and Covariate Shifts," Working Paper.

van der Vaart, A. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Appendix

A.1 Proof of Theorem 1

Part (i). To obtain the joint density of (X_δ, W), we note that

$$F_{X_\delta,W}(x,w) = \Pr(X_\delta \le x, W \le w) = \Pr(X \le x^\delta, W \le w) = F_{X,W}(x^\delta, w),$$

and so

$$f_{X_\delta,W}(x,w) = \frac{\partial x^\delta}{\partial x} \cdot f_{X,W}(x^\delta, w) = J(x^\delta;\delta) f_{X,W}(x^\delta, w).$$

Evaluated at $\delta = 0$, $J\left(x^\delta;\delta\right)$ is 1 and $f_{X_\delta,W}(x,w)$ is $f_{X,W}(x,w)$. Given this, we expand $f_{X_\delta,W}(x,w) - f_{X,W}(x,w)$ around $\delta = 0$, which is possible under Assumptions 1(i) and (iii.a). Observing that

$$\begin{aligned}
\frac{\partial J(x^\delta;\delta)}{\partial x^\delta} &= \frac{\partial}{\partial x}\frac{\partial x^\delta}{\partial x^\delta} = 0, \\
\left.\frac{\partial J(x^\delta;\delta)}{\partial \delta}\right|_{\delta=0} &= \left.\frac{\partial}{\partial \delta}\frac{\partial x^\delta}{\partial x}\right|_{\delta=0} = \left.\frac{\partial}{\partial x}\frac{\partial x^\delta}{\partial \delta}\right|_{\delta=0} = \frac{\partial \kappa\left(x\right)}{\partial x},
\end{aligned}$$

we have

$$\begin{aligned}
& f_{X_\delta,W}(x,w) - f_{X,W}(x,w) \\
&= J(x^\delta;\delta) f_{X,W}(x^\delta,w) - f_{X,W}(x,w) \\
&= \delta \left[\frac{\partial J\left(x^\delta;\delta\right)}{\partial x^\delta}\frac{\partial x^\delta}{\partial \delta} + \frac{\partial J\left(x^\delta;\delta\right)}{\partial \delta}\right]\Bigg|_{\delta=0} f_{X,W}(x,w) \\
&+ \delta J\left(x;0\right) \frac{\partial f_{X,W}(x,w)}{\partial x}\left.\frac{\partial x^\delta}{\partial \delta}\right|_{\delta=0} + \delta R_1(x,w,\delta) \\
&= \delta \left.\frac{\partial J\left(x^\delta;\delta\right)}{\partial \delta}\right|_{\delta=0} f_{X,W}(x,w) + \delta J\left(x;0\right) \frac{\partial f_{X,W}(x,w)}{\partial x}\left.\frac{\partial x^\delta}{\partial \delta}\right|_{\delta=0} + \delta R_1(x,w,\delta) \\
&= \delta \left[\frac{\partial \kappa\left(x\right)}{\partial x} f_{X,W}(x,w) + \frac{\partial f_{X,W}(x,w)}{\partial x}\kappa\left(x\right)\right] + \delta R_1(x,w,\delta) \\
&= \delta \frac{\partial}{\partial x}\left[\kappa\left(x\right) f_{X,W}(x,w)\right] + \delta R_1(x,w,\delta),
\end{aligned}$$

where, for $\tilde{\delta}(x,w)$ between 0 and δ,

$$R_1(x,w,\delta) = \left\{ \frac{\partial \left[J\left(x^\delta;\delta\right) f_{X,W}(x^\delta,w) \right]}{\partial \delta} \bigg|_{\delta=\tilde{\delta}(x,w)} - \frac{\partial \left[J\left(x^\delta;\delta\right) f_{X,W}(x^\delta,w) \right]}{\partial \delta} \bigg|_{\delta=0} \right\}. \tag{A.1}$$

By the continuity of the derivative of $J\left(x^\delta;\delta\right) f_{X,W}(x^\delta,w)$ with respect to δ, we have $R_1(x,w,\delta) = o(1)$ for each (x,w) as $\delta \to 0$.

It remains to show that $\kappa(x) = -\frac{\partial \mathcal{G}(x;\delta)}{\partial \delta}\Big|_{\delta=0}$. Differentiating both sides of $\mathcal{G}\left(x^\delta;\delta\right) = x$ with respect to δ, we obtain that

$$\frac{\partial x^\delta}{\partial \delta} = -\left(\frac{\partial \mathcal{G}\left(x^\delta;\delta\right)}{\partial x^\delta} \right)^{-1} \frac{\partial \mathcal{G}\left(x^\delta;\delta\right)}{\partial \delta}$$

and so

$$\kappa(x) := \frac{\partial x^\delta}{\partial \delta}\bigg|_{\delta=0} = -\left(\frac{\partial \mathcal{G}(x;0)}{\partial x} \right)^{-1} \frac{\partial \mathcal{G}(x;\delta)}{\partial \delta}\bigg|_{\delta=0} = -\frac{\partial \mathcal{G}(x;\delta)}{\partial \delta}\bigg|_{\delta=0},$$

where we have used $\mathcal{G}(x;0) = x$.

Part (ii). Consider first the counterfactual distribution F_{Y_δ}:

$$F_{Y_\delta}(y) = \int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\left\{ h(x,w,u) \le y \right\} f_{U|X_\delta,W}(u|x,w) f_{X_\delta,W}(x,w) du dx dw,$$

where for simplicity we have assumed that the support of X conditional on any $W = w$ does not depend on w and we have denoted the support by $\mathcal{X}$. By Assumption 1(ii), $f_{U|X_\delta,W}(u|x,w) = f_{U|X,W}(u|x^\delta,w)$. So we can write

$$\begin{aligned}
&F_{Y_\delta}(y) \\
&= \int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\left\{ h(x,w,u) \le y \right\} f_{U|X_\delta,W}(u|x,w) f_{X_\delta,W}(x,w) du dx dw \\
&= \int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\left\{ h(x,w,u) \le y \right\} f_{U|X,W}(u|x^\delta,w) f_{X_\delta,W}(x,w) du dx dw \\
&= \underbrace{\int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\left\{ h(x,w,u) \le y \right\} f_{U|X,W}(u|x,w) f_{X,W}(x,w) du dx dw}_{=F_Y(y)} \\
&+ \int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\left\{ h(x,w,u) \le y \right\} f_{U|X,W}(u|x,w) \left[f_{X_\delta,W}(x,w) - f_{X,W}(x,w) \right] du dx dw \\
&+ \int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\left\{ h(x,w,u) \le y \right\} \left[f_{U|X,W}(u|x^\delta,w) - f_{U|X,W}(u|x,w) \right] f_{X_\delta,W}(x,w) du dx dw.
\end{aligned}$$

Hence, we have

$$\frac{F_{Y_\delta}(y) - F_Y(y)}{\delta} := G_{1,\delta}(y) + G_{2,\delta}(y),$$

where

$$G_{1,\delta}(y) = \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u) \le y\} f_{U|X,W}(u|x,w)\frac{1}{\delta}\left[f_{X_\delta,W}(x,w) - f_{X,W}(x,w)\right] dudxdw$$
$$= \int_{\mathcal{W}}\int_{\mathcal{X}} F_{Y|X,W}(y|x,w)\frac{1}{\delta}\left[f_{X_\delta,W}(x,w) - f_{X,W}(x,w)\right] dxdw, \tag{A.2}$$

and

$$G_{2,\delta}(y) = \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u) \le y\}$$
$$\times \frac{1}{\delta}\left[f_{U|X,W}(u|x^\delta,w) - f_{U|X,W}(u|x,w)\right] f_{X_\delta,W}(x,w)dudxdw. \tag{A.3}$$

We first consider the term $G_{1,\delta}(y)$. Using Part (i) and Assumption 1(iv), we have

$$\begin{aligned}
G_{1,\delta}(y) &= \int_{\mathcal{W}}\int_{\mathcal{X}} F_{Y|X,W}(y|x,w)\frac{\partial\left[\kappa(x) f_{X,W}(x,w)\right]}{\partial x} \\
&\quad + \int_{\mathcal{W}}\int_{\mathcal{X}} F_{Y|X,W}(y|x,w)R_1(x,w,\delta)dxdw \\
&= -\int_{\mathcal{W}}\int_{\mathcal{X}} \frac{\partial F_{Y|X,W}(y|x,w)}{\partial x}\kappa(x) f_{X,W}(x,w)dxdw \\
&\quad + \int_{\mathcal{W}}\int_{\mathcal{X}} F_{Y|X,W}(y|x,w)R_1(x,w,\delta)dxdw,
\end{aligned}$$

where the second equality follows from integration by parts. Under Assumption 1(iii.a), we can use the dominated convergence theorem to obtain

$$\lim_{\delta\to 0}\sup_{y\in\mathcal{Y}}\left|\int_{\mathcal{W}}\int_{\mathcal{X}} F_{Y|X,W}(y|x,w)R_1(x,w,\delta)dxdw\right| = 0.$$

Thus, we have that $G_{1,\delta}(y)$ converges to $G_{1,0}(y)$, given by

$$G_{1,0}(y) := -\int_{\mathcal{W}}\int_{\mathcal{X}} \frac{\partial F_{Y|X,W}(y|x,w)}{\partial x'}\kappa(x) f_{X,W}(x,w)dxdw$$

uniformly in $y \in \mathcal{Y}$, as $\delta \to 0$.

Next, we consider $G_{2,\delta}(y)$. Using Assumption 1(iii.b), we have

$$\begin{aligned}
&\left[f_{U|X,W}(u|x^\delta,w) - f_{U|X,W}(u|x,w)\right] f_{X,W}(x^\delta,w) \\
&= \frac{\partial f_{U|X,W}(u|x^\delta,w)}{\partial x^{\delta\prime}} f_{X,W}(x^\delta,w)\frac{\partial x^\delta}{\partial \delta}\bigg|_{\delta=0}\cdot\delta + \delta R_2(u,x,w,\delta) \\
&= \frac{\partial f_{U|X,W}(u|x,w)}{\partial x'} f_{X,W}(x,w)\kappa(x)\,\delta + \delta R_2(u,x,w,\delta),
\end{aligned}$$

where

$$
\begin{aligned}
&R_2(u,x,w,\delta) \\
&= \left.\frac{\partial\left[f_{U|X,W}(u|x^{\delta},w)f_{X,W}(x^{\delta},w)\right]}{\partial\delta}\right|_{\delta=\tilde{\delta}(u,x,w)} - \left.\frac{\partial\left[f_{U|X,W}(u|x^{\delta},w)f_{X,W}(x^{\delta},w)\right]}{\partial\delta}\right|_{\delta=0} \\
&- \left[f_{U|X,W}(u|x,w)\left.\frac{\partial f_{X,W}(x^{\delta},w)}{\partial\delta}\right|_{\delta=\tilde{\delta}(u,x,w)} - f_{U|X,W}(u|x,w)\left.\frac{\partial f_{X,W}(x^{\delta},w)}{\partial\delta}\right|_{\delta=0} \right].
\end{aligned}
$$

Note that in the above, the transpose on x is not relevant but we keep it so that the same lines of arguments can be used for proving Theorem 3. Hence

$$
\begin{aligned}
G_{2,\delta}(y) &= \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u)\le y\}\frac{\partial f_{U|X,W}(u|x,w)}{\partial x'} f_{X,W}(x,w)\kappa(x)\,dudxdw \\
&+ \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u)\le y\} R_2(u,x,w,\delta)dudxdw.
\end{aligned}
$$

Under Assumption 1(iii.b), we can invoke the dominated convergence theorem to get

$$
\lim_{\delta\to 0}\sup_{y\in\mathcal{Y}}\left|\int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u)\le y\} R_2(u,x,w,\delta)dudxdw\right| = 0.
$$

Hence, uniformly in $y\in\mathcal{Y}$, as $\delta\to 0$, $G_{2,\delta}(y)$ converges to

$$
\begin{aligned}
G_{2,0}(y) &= \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u)\le y\}\frac{\partial f_{U|X,W}(u|x,w)}{\partial x'}\kappa(x) f_{X,W}(x,w)dudxdw \\
&= \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \mathbb{1}\{h(x,w,u)\le y\}\frac{\partial \ln f_{U|X,W}(u|x,w)}{\partial x'}\kappa(x) f_{U|X,W}(u|x,w) f_{X,W}(x,w)dudxdw \\
&= E\left[\mathbb{1}\{h(X,W,U)\le y\}\frac{\partial \ln f_{U|X,W}(U|X,W)}{\partial X'}\kappa(X)\right].
\end{aligned}
$$

Combining the above results yields

$$
\begin{aligned}
&\frac{F_{Y_\delta}(y) - F_Y(y)}{\delta} \\
&\to G_{10}(y) + G_{20}(y) \\
&= E\left[\left(-\frac{\partial F_{Y|X,W}(y|X,W)}{\partial X'} + \mathbb{1}\{h(X,W,U)\le y\}\frac{\partial \ln f_{U|X,W}(U|X,W)}{\partial X'}\right)\kappa(X)\right] \\
&:= G(y)
\end{aligned}
$$

uniformly over $y\in\mathcal{Y}$ as $\delta\to 0$.

Part (iii). Note that $\psi(y,\tau,F_Y)$ is the influence function of the quantile functional. Using Part

(ii) and Assumption 1(v), we have

$$\Pi_\tau = \int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) dG\left(y\right) = \int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) dG_{1,0}\left(y\right) + \int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) dG_{2,0}\left(y\right)$$

by Lemma 21.3 in van der Vaart (1998). Now

$$\begin{aligned}
\int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) dG_{1,0}\left(y\right) &= \int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) dG_{1,0}\left(y\right) \\
&= -\int_{\mathcal{W}}\int_{\mathcal{X}} \left[\int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) \frac{\partial f_{Y|X,W}(y|x,w)}{\partial x'} dy\right] \kappa\left(x\right) f_{X,W}(x,w) dx dw \\
&= -\int_{\mathcal{W}}\int_{\mathcal{X}} \frac{\partial}{\partial x'} \left[\int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) f_{Y|X,W}(y|x,w) dy\right] \kappa\left(x\right) f_{X,W}(x,w) dx dw \\
&= -\int_{\mathcal{W}}\int_{\mathcal{X}} \frac{\partial E\left[\psi\left(Y,\tau,F_Y\right)|X=x,W=w\right]}{\partial x'} \kappa\left(x\right) f_{X,W}(x,w) dx dw
\end{aligned}$$

and

$$\begin{aligned}
&\int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) dG_{2,0}\left(y\right) \\
&= \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \left[\int_{\mathcal{Y}} \psi\left(y,\tau,F_Y\right) d\mathbb{1}\left\{h(x,w,u) \le y\right\}\right] \frac{\partial \ln f_{U|X,W}(u|x,w)}{\partial x'} \kappa\left(x\right) \\
&\times f_{U|X,W}(u|x,w) f_{X,W}(x,w) du dx dw \\
&= \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \psi\left(h(x,w,u),\tau,F_Y\right) \frac{\partial \ln f_{U|X,W}(u|x,w)}{\partial x'} \kappa\left(x\right) \\
&\times f_{U|X,W}(u|x,w) f_{X,W}(x,w) du dx dw.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pi_\tau = &-\int_{\mathcal{W}}\int_{\mathcal{X}} \frac{\partial E\left[\psi\left(y,\tau,F_Y\right)|X=x,W=w\right]}{\partial x'} \kappa\left(x\right) f_{X,W}(x,w) dx dw dy \\
&+ \int_{\mathcal{W}}\int_{\mathcal{X}}\int_{\mathcal{U}} \psi\left(h(x,w,u),\tau,F_Y\right) \frac{\partial \ln f_{U|X,W}(u|x,w)}{\partial x'} \kappa\left(x\right) \\
&\times f_{U|X,W}(u|x,w) f_{X,W}(x,w) du dx dw.
\end{aligned}$$

A.2 Proof of Theorem 2

The conditional version of Π_τ is

$$\Pi_\tau(x,w) = \frac{1}{f_{Y|X,W}(Q_\tau[Y|x,w]|x,w)} \left.\frac{\partial F_{Y|X,W}(Q_\tau[Y|x,w]|z,w)}{\partial z}\right|_{z=x} \kappa\left(x\right).$$

Let $\xi_\tau(x,w)$ be the quantile implied by the matching function in (10). The conditional effect at this particular quantile is then

$$\Pi_{\xi_\tau(x,w)}(x,w) = \frac{1}{f_{Y|X,W}(Q_\tau[Y]|x,w)} \frac{\partial F_{Y|X,W}(Q_\tau[Y]|x,w)}{\partial x} \kappa(x).$$

But Corollary 1 says that

$$\Pi_\tau = \frac{1}{f_Y(Q_\tau[Y])} \int_{\mathcal{W}} \int_{\mathcal{X}} \frac{\partial F_{Y|X,W}(Q_\tau[Y]|x,w)}{\partial x} \kappa(x) f_{X,W}(x,w)\, dxdw.$$

It follows that we can reweigh $\Pi_\tau(x,w)$ to obtain Π_τ:

$$\begin{aligned}
\Pi_\tau &= \int_{\mathcal{W}} \int_{\mathcal{X}} \Pi_{\xi_\tau(x,w)}(x,w) \frac{f_{Y|X,W}(Q_\tau[Y]|x,w)}{f_Y(Q_\tau[Y])} f_{X,W}(x,w) dxdw \\
&= E\left[\Pi_{\xi_\tau(X,W)}(X,W) \frac{f_{Y|X,W}(Q_\tau[Y]|X,W)}{f_Y(Q_\tau[Y])}\right].
\end{aligned}$$

To obtain the second representation, we note that

$$\frac{f_{Y|X,W}(Q_\tau[Y]|x,w)}{f_Y(Q_\tau[Y])} f_{X,W}(x,w) = f_{X,W|Y}(x,w|Q_\tau[Y]),$$

and so we obtain:

$$\begin{aligned}
\Pi_\tau &= \int_{\mathcal{W}} \int_{\mathcal{X}} \Pi_{\xi_\tau(x,w)}(x,w) f_{X,W|Y}(x,w|Q_\tau[Y]) dxdw \\
&= E\left[\Pi_{\xi_\tau(X,W)}(X,W) \middle| Y = Q_\tau[Y]\right].
\end{aligned}$$

A.3 Proof of Theorem 3

The proof of this Theorem is very similar to the proof of Theorem 1. The following decomposition still holds

$$\frac{F_{Y_\delta}(y) - F_Y(y)}{\delta} := G_{1,\delta}(y) + G_{2,\delta}(y),$$

where

$$\begin{aligned}
G_{1,\delta}(y) &= \int_{\mathcal{W}} \int_{\mathcal{X}} F_{Y|X,W}(y|x,w) \frac{[f_{X_\delta,W}(x,w) - f_{X,W}(x,w)]}{\delta} dxdw, \\
G_{2,\delta}(y) &= \int_{\mathcal{W}} \int_{\mathcal{X}} \int_{\mathcal{U}} \mathbb{1}\{h(x,w,u) \le y\} \frac{\left[f_{U|X,W}(u|x^\delta,w) - f_{U|X,W}(u|x,w)\right]}{\delta} f_{X_\delta,W}(x,w) dudxdw.
\end{aligned}$$

We first consider the term $G_{1,\delta}(y)$. Under the assumptions given, we have

$$f_{X_\delta,W}(x,w) = \det\left[J(x^\delta;\delta)\right] f_{X,W}(x^\delta,w).$$

Evaluated at $\delta = 0$, $f_{X_\delta,W}(x,w)$ is $f_{X,W}(x,w)$. Given this, we expand $f_{X_\delta,W}(x,w) - f_{X,W}(x,w)$ around $\delta = 0$, which is possible under Assumptions 3(i) and (iii). We have

$$\begin{aligned}
& f_{X_\delta,W}(x,w) - f_{X,W}(x,w) \\
&= \delta \frac{\partial \left[\det\left[J\left(x^\delta;\delta\right)\right] f_{X,W}(x^\delta,w)\right]}{\partial \delta}\bigg|_{\delta=0} + \delta R_1(x,w,\delta) \\
&= \delta \frac{\partial \det\left[J\left(x^\delta;\delta\right)\right]}{\partial \delta}\bigg|_{\delta=0} f_{X,W}(x,w) + \delta \det\left[J(x^\delta;\delta)\right]\left(\frac{\partial x^\delta}{\partial \delta}\right)' \frac{\partial f_{X,W}(x^\delta,w)}{\partial x^\delta}\bigg|_{\delta=0} + \delta R_1(x,w,\delta) \\
&= \delta \frac{\partial \det\left[J\left(x^\delta;\delta\right)\right]}{\partial \delta}\bigg|_{\delta=0} f_{X,W}(x,w) + \delta \kappa\left(x\right)' \frac{\partial f_{X,W}(x,w)}{\partial x} + \delta R_1(x,w,\delta),
\end{aligned} \tag{A.4}$$

where, for $\tilde{\delta}(x,w)$ between 0 and δ,

$$R_1(x,w,\delta) = \left\{ \frac{\partial \left[\det\left[J\left(x^\delta;\delta\right)\right] f_{X,W}(x^\delta,w)\right]}{\partial \delta}\bigg|_{\delta=\tilde{\delta}(x,w)} - \frac{\partial \left[\det\left[J\left(x^\delta;\delta\right)\right] f_{X,W}(x^\delta,w)\right]}{\partial \delta}\bigg|_{\delta=0} \right\}.$$

Using the arguments similar to those in the proof of Theorem 1, we can show that $G_{1,\delta}\left(y\right)$ converges to

$$\begin{aligned}
G_{1,0}\left(y\right) &:= \int_{\mathcal{W}}\int_{\mathcal{X}} \frac{\partial \det\left[J\left(x^\delta;\delta\right)\right]}{\partial \delta}\bigg|_{\delta=0} F_{Y|X,W}(y|x,w) f_{X,W}(x,w)dxdw \\
&+ \int_{\mathcal{W}}\int_{\mathcal{X}} F_{Y|X,W}(y|x,w)\left(\frac{\partial x^\delta}{\partial \delta}\right)'\bigg|_{\delta=0} \frac{\partial f_{X,W}(x,w)}{\partial x} dxdw \\
&:= G_{1,0}^{(1)}\left(y\right) + G_{1,0}^{(2)}\left(y\right)
\end{aligned}$$

uniformly in $y \in \mathcal{Y}$, as $\delta \to 0$.

Using Assumption 3 and the fact that $\frac{\partial x_i^\delta}{\partial x_j}|_{\delta=0} = \mathbb{1}\left\{i=j\right\}$, we have

$$\begin{aligned}
\frac{\partial \det\left[J\left(x^\delta;\delta\right)\right]}{\partial \delta}\bigg|_{\delta=0} &= \frac{\partial}{\partial \delta}\left(\frac{\partial x_1^\delta}{\partial x_1}\frac{\partial x_2^\delta}{\partial x_2} - \frac{\partial x_1^\delta}{\partial x_2}\frac{\partial x_2^\delta}{\partial x_1}\right)\bigg|_{\delta=0} \\
&= \left(\frac{\partial \kappa_1\left(x\right)}{\partial x_1}\frac{\partial x_2^\delta}{\partial x_2} + \frac{\partial x_1^\delta}{\partial x_1}\frac{\partial \kappa_2\left(x\right)}{\partial x_2} - \frac{\partial \kappa_1\left(x\right)}{\partial x_2}\frac{\partial x_2^\delta}{\partial x_1} - \frac{\partial x_1^\delta}{\partial x_2}\frac{\partial \kappa_2\left(x\right)}{\partial x_1}\right)\bigg|_{\delta=0} \\
&= \frac{\partial \kappa_1\left(x\right)}{\partial x_1} + \frac{\partial \kappa_2\left(x\right)}{\partial x_2}.
\end{aligned}$$

So

$$G_{1,0}^{(1)}\left(y\right) = \int_{\mathcal{W}}\int_{\mathcal{X}}\left(\frac{\partial \kappa_1\left(x\right)}{\partial x_1} + \frac{\partial \kappa_2\left(x\right)}{\partial x_2}\right) F_{Y|X,W}(y|x,w) f_{X,W}(x,w)dxdw.$$

Next, note that

$$\left(\frac{\partial x^\delta}{\partial \delta}\right)'\bigg|_{\delta=0} \frac{\partial f_{X,W}(x,w)}{\partial x} = \kappa\left(x\right)' \frac{\partial f_{X,W}(x,w)}{\partial x}.$$

Using integration by parts, we can show that for $j = 1$ and 2,

$$\int_{\mathcal{W}} \int_{\mathcal{X}} F_{Y|X,W}(y|x,w) \left[\kappa_j(x) \frac{\partial f_{X,W}(x,w)}{\partial x_j} \right] dxdw$$
$$= -\int_{\mathcal{W}} \int_{\mathcal{X}} f_{X,W}(x,w) \frac{\partial \left[F_{Y|X,W}(y|x,w) \kappa_j(x) \right]}{\partial x_j} dxdw.$$

So

$$\begin{aligned}
G_{1,0}^{(2)}(y) &= \int_{\mathcal{W}} \int_{\mathcal{X}} F_{Y|X,W}(y|x,w) \left[\kappa(x)' \frac{\partial f_{X,W}(x,w)}{\partial x} \right] dxdw \\
&= -\int_{\mathcal{W}} \int_{\mathcal{X}} f_{X,W}(x,w) \left(\frac{\partial \left[F_{Y|X,W}(y|x,w) \kappa_1(x) \right]}{\partial x_1} + \frac{\partial \left[F_{Y|X,W}(y|x,w) \kappa_2(x) \right]}{\partial x_2} \right) dxdw \\
&= -\int_{\mathcal{W}} \int_{\mathcal{X}} f_{X,W}(x,w) \left(\frac{\partial \left[F_{Y|X,W}(y|x,w) \right]}{\partial x_1} \kappa_1(x) + \frac{\partial \left[F_{Y|X,W}(y|x,w) \right]}{\partial x_2} \kappa_2(x) \right) dxdw \\
&\quad - \int_{\mathcal{W}} \int_{\mathcal{X}} f_{X,W}(x,w) F_{Y|X,W}(y|x,w) \left(\frac{\partial \kappa_1(x)}{\partial x_1} + \frac{\partial \kappa_2(x)}{\partial x_2} \right) dxdw.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& G_{1,0}(y) \\
&= -\int_{\mathcal{W}} \int_{\mathcal{X}} \left[\frac{\partial F_{Y|X,W}(y|x,w)}{\partial x_1} \kappa_1(x) + \frac{\partial F_{Y|X,W}(y|x,w)}{\partial x_2} \kappa_2(x) \right] f_{X,W}(x,w) dxdw \\
&= -\int_{\mathcal{W}} \int_{\mathcal{X}} \left[\frac{\partial \left[F_{Y|X,W}(y|x,w) \right]}{\partial x'} \kappa(x) \right] f_{X,W}(x,w) dxdw \\
&= -E \left[\frac{\partial F_{Y|X,W}(y|X,W)}{\partial X'} \kappa(X) \right].
\end{aligned}$$

For $G_{2,\delta}(y)$, the proof of Theorem 1 remains valid, and we have that $G_{2,\delta}(y)$ converges to

$$G_{2,0}(y) := E \left[\mathbb{1} \left\{ h(X,W,U) \leq y \right\} \frac{\partial \ln f_{U|X,W}(U|X,W)}{\partial X'} \kappa(X) \right]$$

uniformly in $y \in \mathcal{Y}$, as $\delta \to 0$.

Invoking the same argument as that in the proof of Theorem 1, we obtain the desired result.

A.4 Proof of Lemma 1

The main complication in this lemma is that the dependent variable is $\mathbb{1} \{Y_i \leq \hat{q}_\tau\}$. This means that the preliminary estimator $\hat{q}_\tau$ might affect the asymptotic distribution of $\hat{\alpha}_\tau$ and $\hat{\beta}_\tau$.

As mentioned in the main text, under Assumption 4,

$$\hat{q}_\tau - Q_\tau[Y] = \frac{1}{n}\sum_{i=1}^{n} \frac{\tau - \mathbb{1}\{Y_i \le Q_\tau[Y]\}}{f_Y(Q_\tau[Y])} + o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, \tau, F_Y) + o_p(n^{-1/2}).$$

Recall that

$$\hat{\theta}_\tau = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \left\{ \mathbb{1}\{Y_i \le \hat{q}_\tau\} \log\left[G(Z_i'\theta)\right] + \mathbb{1}\{Y_i > \hat{q}_\tau\} \log\left[1 - G(Z_i'\theta)\right] \right\}.$$

Let $s_i(\theta; \hat{q}_\tau)$ denote the score for observation i. Then, under Assumption 5(i), we have

$$\frac{1}{n}\sum_{i=1}^{n} s_i(\hat{\theta}_\tau; \hat{q}_\tau) = 0.$$

Taking a mean-value expansion (element-by-element), we obtain

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} s_i(\hat{\theta}_\tau; \hat{q}_\tau)}_{=0} = \frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; \hat{q}_\tau) + \frac{1}{n}\sum_{i=1}^{n} H_i(\tilde{\theta}_\tau; \hat{q}_\tau)\left(\hat{\theta}_\tau - \theta_\tau\right),$$

where $\tilde{\theta}_\tau$ is between θ_τ and $\hat{\theta}_\tau$ and can be different for different rows of H_i. Under the assumption of the uniform law of large numbers for the Hessian (i.e., Assumption 5(ii)), we obtain

$$\frac{1}{n}\sum_{i=1}^{n} H_i(\tilde{\theta}_\tau, \hat{q}_\tau) \xrightarrow{p} E[H_i(\theta_\tau; Q_\tau[Y])] =: H.$$

We have then

$$0 = \frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; \hat{q}_\tau) + H\left(\hat{\theta}_\tau - \theta_\tau\right) + o_p\left(\left\|\hat{\theta}_\tau - \theta_\tau\right\|\right). \tag{A.5}$$

Now, we use the stochastic equicontinuity in Assumption 5(iii):

$$\frac{1}{n}\sum_{i=1}^{n}\left(s_i(\theta_\tau; \hat{q}_\tau) - E\left[s_i(\theta_\tau; q)\right]|_{q=\hat{q}_\tau}\right) = \frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; Q_\tau[Y]) + o_p(n^{-1/2}).$$

Here we have used that $E[s_i(\theta_\tau; Q_\tau[Y])] = 0$: the score evaluated at the true quantile has expected value 0. Plugging this back into (A.5), we obtain

$$0 = E\left[s_i(\theta_\tau; q)\right]|_{q=\hat{q}_\tau} + \frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; Q_\tau[Y]) + H\left(\hat{\theta}_\tau - \theta_\tau\right) + o_p\left(\left\|\hat{\theta}_\tau - \theta_\tau\right\|\right). \tag{A.6}$$

Here $E\left[s_i(\theta_\tau; q)\right]|_{q=\hat{q}_\tau}$ is random because we first compute the expectation $E\left[s_i(\theta_\tau; q)\right]$ for a fixed q and then replace q by $\hat{q}_\tau$, which is random. To show that $E\left[s_i(\theta_\tau; q)\right]|_{q=\hat{q}_\tau}$ is $O_p(n^{-1/2})$, we observe that (see equation 15.18 in Wooldridge (2002))

$$s_i(\theta;q) = \frac{g(Z_i'\theta)Z_i\left[\mathbb{1}\left\{Y_i \leq q\right\} - G\left(Z_i'\theta\right)\right]}{G\left(Z_i'\theta\right)\left[1 - G\left(Z_i'\theta\right)\right]}. \tag{A.7}$$

Therefore, using the law of iterated expectations, we obtain

$$E\left[s_i(\theta;q)\right] = E\left[\frac{g(Z_i'\theta)Z_i\left[F_{Y|Z}(q|Z_i) - G\left(Z_i'\theta\right)\right]}{G\left(Z_i'\theta\right)\left[1 - G\left(Z_i'\theta\right)\right]}\right].$$

So

$$H_Q = \left.\frac{\partial E\left[s_i(\theta_\tau;q)\right]}{\partial q}\right|_{q=Q_\tau[Y]} = E\left[\frac{g(Z_i'\theta_\tau)Z_i\left[f_{Y|Z}(Q_\tau[Y]|Z_i)\right]}{G\left(Z_i'\theta_\tau\right)\left[1 - G\left(Z_i'\theta_\tau\right)\right]}\right]. \tag{A.8}$$

We have

$$\begin{aligned}
E\left[s_i(\theta_\tau;q)\right]|_{q=\hat{q}_\tau} &= \underbrace{E\left[s_i(\theta_\tau;Q_\tau[Y])\right]}_{=0} + \underbrace{\left.\frac{\partial E\left[s_i(\theta_\tau;q)\right]}{\partial q}\right|_{q=Q_\tau[Y]}}_{=H_Q}\left(\hat{q}_\tau - Q_\tau[Y]\right) + o_p(n^{-1/2}) \\
&= H_Q\left(\hat{q}_\tau - Q_\tau[Y]\right) + o_p(n^{-1/2}),
\end{aligned}$$

which implies that $E\left[s_i(\theta_\tau;q)\right]|_{q=\hat{q}_\tau} = O_p(n^{-1/2})$. Going back to (A.6), we obtain

$$\left\|H\left(\hat{\theta}_\tau - \theta_\tau\right) + o_p\left(\left\|\hat{\theta}_\tau - \theta_\tau\right\|\right)\right\| \leq \left\|E\left[s_i(\theta_\tau;q)\right]|_{q=\hat{q}_\tau}\right\| + \left\|\frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau;Q_\tau[Y])\right\|,$$

which implies that

$$\hat{\theta}_\tau - \theta_\tau = O_p(n^{-1/2}).$$

Furthermore, since H is negative definite, then we have

$$\begin{aligned}
\hat{\theta}_\tau - \theta_\tau &= \underbrace{-H^{-1}\frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau;Q_\tau[Y])}_{\text{Usual influence function}} - \underbrace{H^{-1}E\left[s_i(\theta_\tau;q)\right]|_{q=\hat{q}_\tau}}_{\text{Contribution of } \hat{q}_\tau} + o_p(n^{-1/2}) \\
&= -H^{-1}\frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau;Q_\tau[Y]) - H^{-1}H_Q\left(\hat{q}_\tau - Q_\tau[Y]\right) + o_p(n^{-1/2}) \\
&= -H^{-1}\frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau;Q_\tau[Y]) - H^{-1}H_Q\frac{1}{n}\sum_{i=1}^{n}\psi(Y_i,\tau,F_Y) + o_p(n^{-1/2}).
\end{aligned} \tag{A.9}$$

A.5 Proof of Theorem 4

To establish the joint asymptotic distribution of the estimators of the location and scale effect, we need to obtain the asymptotic distribution of $\hat{f}_Y\left(\hat{q}_\tau\right)$. By Lemma 6 in Martinez-Iriarte and Sun

(2021b), we have that

$$\hat{f}_Y(y) - f_Y(y) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h\left(Y_i - y\right) - E\left[K_h\left(Y - y\right)\right] + B_f(y) + o_p(h^2), \tag{A.10}$$

where the bias is

$$B_{f_Y}(y) = \frac{1}{2}h^2 f_Y''^2(y)\int_{-\infty}^{\infty} u^2\mathcal{K}(u)du.$$

Moreover, we can write

$$\hat{f}_Y(\hat{q}_\tau) - \hat{f}_Y(Q_\tau[Y]) = \dot{f}_Y(Q_\tau[Y])\left(\hat{q}_\tau - Q_\tau[Y]\right) + o_p(n^{-1/2}h^{-1/2}),$$

where $\dot{f}_Y$ is the derivative of the density. Thus, we have that

$$\begin{aligned}
&\hat{f}_Y\left(\hat{q}_\tau\right) - f_Y\left(Q_\tau[Y]\right) \\
&= \hat{f}_Y\left(\hat{q}_\tau\right) - \hat{f}_Y\left(Q_\tau[Y]\right) + \hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right) \\
&= \dot{f}_Y(Q_\tau[Y])\left(\hat{q}_\tau - Q_\tau[Y]\right) + \hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right) + o_p(n^{-1/2}h^{-1/2}).
\end{aligned} \tag{A.11}$$

The first term captures the uncertainty associated with estimating the quantile, and the second term captures the uncertainty associated with estimating the density.

Next, we can write the location and scale effects as

$$\begin{pmatrix}\hat{\Pi}_{\tau,L}\\ \hat{\Pi}^{\mu}_{\tau,S}\end{pmatrix} - \begin{pmatrix}\Pi_{\tau,L}\\ \Pi^{\mu}_{\tau,S}\end{pmatrix} = D_\mu\left[\frac{n^{-1}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i}{\hat{f}_Y\left(\hat{q}_\tau\right)} - \frac{E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]}{f_Y\left(Q_\tau[Y]\right)}\right].$$

Now

$$\begin{aligned}
&\frac{n^{-1}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i}{\hat{f}_Y\left(\hat{q}_\tau\right)} - \frac{E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]}{f_Y\left(Q_\tau[Y]\right)} \\
&= \frac{n^{-1}\sum_{i=1}^{n}\left[g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i\right] - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]}{\hat{f}_Y\left(\hat{q}_\tau\right)} - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]\frac{\hat{f}_Y\left(\hat{q}_\tau\right) - f_Y\left(Q_\tau[Y]\right)}{\hat{f}_Y\left(\hat{q}_\tau\right)f_Y\left(Q_\tau[Y]\right)} \\
&= \frac{n^{-1}\sum_{i=1}^{n}\left[g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i\right] - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]}{\hat{f}_Y\left(\hat{q}_\tau\right)} \\
&- \frac{E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]}{f_Y\left(Q_\tau[Y]\right)^2}\left\{\dot{f}_Y(Q_\tau[Y])\left(\hat{q}_\tau - Q_\tau[Y]\right) + \hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right)\right\} + o_p(n^{-1/2}h^{-1/2}).
\end{aligned}$$

Taking a mean-value expansion (element-by-element), we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i &= \frac{1}{n}\sum_{i=1}^{n} g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i \\
&+ \left(\frac{1}{n}\sum_{i=1}^{n}\dot{g}(Z_i'\tilde{\theta}_\tau)\tilde{\alpha}_\tau\tilde{X}_i Z_i'\right)\left(\hat{\theta}_\tau - \theta_\tau\right) + \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\tilde{\theta}_\tau)\tilde{X}_i\right)\left(\hat{\alpha}_\tau - \alpha_\tau\right).
\end{aligned}$$

Using the uniform law of large numbers in Assumption 5(iv), we have

$$\frac{1}{n}\sum_{i=1}^{n}\dot{g}(Z_i'\tilde{\theta}_\tau)\tilde{\alpha}_\tau\tilde{X}_iZ_i' \xrightarrow{p} M_1 := \underbrace{E\left[\dot{g}(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_iZ_i'\right]}_{2\times\dim(Z)}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\tilde{\theta}_\tau)\tilde{X}_i \xrightarrow{p} M_2 := \underbrace{E\left[g(Z_i'\theta_\tau)\tilde{X}_i\right]}_{2\times 1}.$$

Therefore,

$$\begin{aligned}&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]\right)\\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]\right) + M_1\sqrt{n}(\hat{\theta}_\tau-\theta_\tau) + M_2\sqrt{n}(\hat{\alpha}_\tau-\alpha_\tau) + o_p(1).\end{aligned}$$

The first term captures the uncertainty in estimating the expected value, and the second and third terms capture the uncertainty in estimating the logit/probit model, and it has already incorporated the contribution of the preliminary estimator $\hat{q}_\tau$ of $Q_\tau[Y]$. To ease notation, define $M := M_1 + (M_2, O)$ where O is a $2\times\dim(W)$ matrix of zeros. An explicit expression of M is given in Assumption 5(iv). Thus, we can write:

$$\begin{aligned}&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\tilde{X}_i - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]\right)\\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]\right) + M\sqrt{n}\left(\hat{\theta}_\tau-\theta_\tau\right) + o_p(1).\end{aligned} \tag{A.12}$$

It then follows that

$$\begin{aligned}\begin{pmatrix}\hat{\Pi}_{\tau,L}\\ \hat{\Pi}^{\mu}_{\tau,S}\end{pmatrix} - \begin{pmatrix}\Pi_{\tau,L}\\ \Pi^{\mu}_{\tau,S}\end{pmatrix} &= \frac{1}{f_Y\left(Q_\tau[Y]\right)}D_\mu\left[\frac{1}{n}\sum_{i=1}^{n}g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i - E\left[g(Z_i'\theta_\tau)\alpha_\tau\tilde{X}_i\right]\right]\\ &\quad + \frac{1}{f_Y\left(Q_\tau[Y]\right)}D_\mu M\left(\hat{\theta}_\tau-\theta_\tau\right) - \begin{pmatrix}\Pi_{\tau,L}\\ \Pi^{\mu}_{\tau,S}\end{pmatrix}\frac{\dot{f}_Y(Q_\tau[Y])}{f_Y\left(Q_\tau[Y]\right)}\left(\hat{q}_\tau - Q_\tau[Y]\right)\\ &\quad - \begin{pmatrix}\Pi_{\tau,L}\\ \Pi^{\mu}_{\tau,S}\end{pmatrix}\frac{\left(\hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right)\right)}{f_Y\left(Q_\tau[Y]\right)} + o_p(n^{-1/2}) + o_p(n^{-1/2}h^{-1/2}).\end{aligned}$$

Plugging the asymptotic representation of $\sqrt{n}\left(\hat{\theta}_\tau - \theta_\tau\right)$ in (A.9), we obtain:

$$\begin{aligned}
\begin{pmatrix} \hat{\Pi}_{\tau,L} \\ \hat{\Pi}^{\mu}_{\tau,S} \end{pmatrix} - \begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix} &= \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_\mu \left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\theta_\tau)\alpha_\tau \tilde{X}_i - E\left[g(Z_i'\theta_\tau)\alpha_\tau \tilde{X}_i\right]\right] \\
&- \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_\mu M H^{-1} \frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; Q_\tau[Y]) \\
&- \left[\begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix} \frac{\dot{f}_Y(Q_\tau[Y])}{f_Y\left(Q_\tau[Y]\right)} + \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_\mu M H^{-1} H_Q \right] \frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, \tau, F_Y) \\
&- \begin{pmatrix} \Pi_{\tau,L} \\ \Pi^{\mu}_{\tau,S} \end{pmatrix} \frac{\hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right)}{f_Y\left(Q_\tau[Y]\right)} + o_p(n^{-1/2}) + o_p(n^{-1/2}h^{-1/2}).
\end{aligned}$$

Plugging the representation of $\hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right)$ in (A.10) completes the proof.

A.6 Proof of Corollary 3

The result has been proved in the main text. Here we give the expressions for $\hat{M}$, $\hat{H}$, and $\hat{H}_Q$. For $\hat{M}$ and $\hat{H}$, we have

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} \dot{g}(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau X_i' + g(Z_i'\hat{\theta}_\tau), & \dot{g}(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau W_i' \\ \dot{g}(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau X_i X_i' + g(Z_i'\hat{\theta}_\tau) X_i, & \dot{g}(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau X_i W_i' \end{pmatrix}$$

and

$$\hat{H} = \frac{1}{n}\sum_{i=1}^{n} \frac{g(Z_i'\hat{\theta}_\tau)^2}{G(Z_i'\hat{\theta}_\tau)(1 - G(Z_i'\hat{\theta}_\tau))} \begin{pmatrix} X_i^2 & X_i W_i' \\ X_i W_i & W_i W_i' \end{pmatrix}.$$

For $\hat{H}_Q$, we note that

$$H_Q = \left.\frac{\partial E\left[s_i(\theta_\tau; q)\right]}{\partial q}\right|_{q=Q_\tau[Y]} = E\left[\frac{g(Z_i'\theta_\tau) Z_i f_{Y|Z}(Q_\tau[Y]|Z_i)}{G(Z_i'\theta_\tau)\left[1 - G(Z_i'\theta_\tau)\right]}\right].$$

Let

$$\Lambda(Z_i, \theta_\tau) := \frac{g(Z_i'\theta_\tau) Z_i}{G(Z_i'\theta_\tau)\left[1 - G(Z_i'\theta_\tau)\right]}.$$

Then

$$
\begin{aligned}
H_Q &= E[\Lambda(Z_i,\theta_\tau) f_{Y|Z}(Q_\tau[Y]|Z_i)] \\
&= \int_{\mathcal{Z}} \Lambda(z,\theta_\tau) f_{Y|Z}(Q_\tau[Y]|z) f_Z(z) dz \\
&= f_Y(Q_\tau[Y]) \int_{\mathcal{Z}} \Lambda(z,\theta_\tau) \frac{f_{Y,Z}(Q_\tau[Y],z)}{f_Y(Q_\tau[Y]) f_Z(z)} f_Z(z) dz \\
&= f_Y(Q_\tau[Y]) \int_{\mathcal{Z}} \Lambda(z,\theta_\tau) f_{Z|Y}(z|Q_\tau[Y]) dz \\
&= f_Y(Q_\tau[Y]) E[\Lambda(Z,\theta_\tau)|Y = Q_\tau[Y])].
\end{aligned}
$$

To estimate the conditional expectation, we may use a vector version of the Nadaraya-Watson estimator:

$$
\hat{E}[\Lambda(Z,\hat{\theta}_\tau)|Y = \hat{q}_\tau] = \frac{\sum_{i=1}^n K_h(Y_i - \hat{q}_\tau)\Lambda(Z_i,\hat{\theta}_\tau)}{\sum_{i=1}^n K_h(Y_i - \hat{q}_\tau)},
$$

where K_h is the rescaled kernel $K_h(Y_i - y) = h^{-1} K\left((Y_i - y)/h\right)$ for a kernel function $K(\cdot)$. We can then estimate H_Q by

$$
\begin{aligned}
\hat{H}_Q &= \hat{f}_Y(\hat{q}_\tau)\hat{E}[\Lambda(Z,\hat{\theta}_\tau)|Y = \hat{q}_\tau] \\
&= \left[\frac{1}{n}\sum_{i=1}^n K_h(Y_i - \hat{q}_\tau)\right] \cdot \frac{\sum_{i=1}^n K_h(Y_i - \hat{q}_\tau)\Lambda(Z_i,\hat{\theta}_\tau)}{\sum_{i=1}^n K_h(Y_i - \hat{q}_\tau)} \\
&= \frac{1}{n}\sum_{i=1}^n K_h(Y_i - \hat{q}_\tau)\Lambda(Z_i,\hat{\theta}_\tau). \qquad \text{(A.13)}
\end{aligned}
$$

It is worth pointing out that, in the logistic case, $G(z) = (1 + \exp(-z))^{-1}$, we have the convenient identity $g(z) = G(z)(1 - G(z))$. Thus, $\Lambda(Z_i,\hat{\theta}_\tau) = Z_i$ and the estimation of H and H_Q becomes simpler.

A.7 Proof of Theorem 5

The proof of this theorem is similar to that of Theorem 4. We outline the main steps and omit the details here. We have

$$
\begin{pmatrix} \hat{\Pi}_{\tau,L,1} \\ \hat{\Pi}_{\tau,L,2} \end{pmatrix} - \begin{pmatrix} \Pi_{\tau,L,1} \\ \Pi_{\tau,L,2} \end{pmatrix} = D_L \left[\frac{n^{-1}\sum_{i=1}^n g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau}{\hat{f}_Y(\hat{q}_\tau)} - \frac{E\left[g(Z_i'\theta_\tau)\alpha_\tau\right]}{f_Y(Q_\tau[Y])} \right].
$$

But

$$
\begin{aligned}
&\frac{n^{-1}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau}{\hat{f}_Y\left(\hat{q}_\tau\right)} - \frac{E\left[g(Z_i'\theta_\tau)\alpha_\tau\right]}{f_Y\left(Q_\tau[Y]\right)} \\
&= \frac{n^{-1}\sum_{i=1}^{n}\left[g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau\right] - E\left[g(Z_i'\theta_\tau)\alpha_\tau\right]}{f_Y\left(Q_\tau[Y]\right)} \\
&- \frac{E\left[g(Z_i'\theta_\tau)\alpha_\tau\right]}{f_Y\left(Q_\tau[Y]\right)^2}\left\{\dot{f}_Y(Q_\tau[Y])\left(\hat{q}_\tau - Q_\tau[Y]\right) + \hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right)\right\} + o_p(n^{-1/2}h^{-1/2}).
\end{aligned}
$$

Now

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\hat{\alpha}_\tau &= \frac{1}{n}\sum_{i=1}^{n} g(Z_i'\theta_\tau)\alpha_\tau \\
&+ \left(\frac{1}{n}\sum_{i=1}^{n}\dot{g}(Z_i'\tilde{\theta}_\tau)\tilde{\alpha}_\tau Z_i'\right)(\hat{\theta}_\tau - \theta_\tau) + \left(\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\tilde{\theta}_\tau)\right)(\hat{\alpha}_\tau - \alpha_\tau) \\
&= \frac{1}{n}\sum_{i=1}^{n} g(Z_i'\hat{\theta}_\tau)\alpha_\tau + M_L(\hat{\theta}_\tau - \theta_\tau) + o_p\left(n^{-1/2}\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\begin{pmatrix}\hat{\Pi}_{\tau,L,1}\\ \hat{\Pi}_{\tau,L,2}\end{pmatrix} - \begin{pmatrix}\Pi_{\tau,L,1}\\ \Pi_{\tau,L,2}\end{pmatrix} \\
&= \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_L\left[\frac{1}{n}\sum_{i=1}^{n} g(Z_i'\theta_\tau)\alpha_\tau - E\left[g(Z_i'\theta_\tau)\alpha_\tau\right]\right] \\
&- \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_L M_L H^{-1}\frac{1}{n}\sum_{i=1}^{n} s_i(\theta_\tau; Q_\tau[Y]) \\
&- \left[\begin{pmatrix}\Pi_{\tau,L,1}\\ \Pi_{\tau,L,2}\end{pmatrix}\frac{\dot{f}_Y(Q_\tau[Y])}{f_Y\left(Q_\tau[Y]\right)} + \frac{1}{f_Y\left(Q_\tau[Y]\right)} D_L M_L H^{-1} H_Q\right]\frac{1}{n}\sum_{i=1}^{n}\psi(Y_i,\tau,F_Y) \\
&- \begin{pmatrix}\Pi_{\tau,L,1}\\ \Pi_{\tau,L,2}\end{pmatrix}\frac{\hat{f}_Y\left(Q_\tau[Y]\right) - f_Y\left(Q_\tau[Y]\right)}{f_Y\left(Q_\tau[Y]\right)} + o_p(n^{-1/2}) + o_p(n^{-1/2}h^{-1/2}).
\end{aligned}
$$

Combining this with (A.10) leads to the desired result.

Supplementary Appendix

S.1 Details of Example 4

Let $\varepsilon \sim N(0,1)$. Before the location-scale shift,

$$Y = \alpha + X\beta + U \sim N(\alpha, 1+\beta^2) := \alpha + \sqrt{1+\beta^2}\varepsilon,$$

and the τ-quantile $Q_\tau[Y]$ of Y is $\alpha + \sqrt{1+\beta^2}\epsilon_\tau$ where ϵ_τ is the τ-quantile of ε. After the location-scale shift with

$$X_\delta = X/s(\delta) + \ell(\delta) \sim N(\ell(\delta), s^{-2}(\delta)),$$

we have

$$Y_\delta = \alpha + X_\delta\beta + U \sim N\left[\alpha + \beta\ell(\delta), 1+\beta^2 s^{-2}(\delta)\right] := \alpha + \beta\ell(\delta) + \sqrt{1+\beta^2 s^{-2}(\delta)}\varepsilon,$$

and the τ-quantile $Q_\tau[Y_\delta]$ of Y_δ is $\alpha + \beta\ell(\delta) + \sqrt{1+\beta^2 s^{-2}(\delta)}\epsilon_\tau$. Hence

$$\begin{aligned}
\Pi_\tau &= \lim_{\delta\to 0} \frac{\beta\ell(\delta) + \sqrt{1+\beta^2 s^{-2}(\delta)}\epsilon_\tau - \sqrt{1+\beta^2}\epsilon_\tau}{\delta} \\
&= \dot{\ell}(0)\beta - \dot{s}(0)\frac{\beta^2}{\sqrt{\beta^2+1}} Q_\tau[U] \\
&= \dot{\ell}(0)\beta - \dot{s}(0)\frac{\beta^2}{\sqrt{\beta^2+1}} \frac{Q_\tau[Y]-\alpha}{\sqrt{\beta^2+1}} \\
&= \dot{\ell}(0)\beta - \dot{s}(0)\frac{\beta^2}{\beta^2+1}\left(Q_\tau[Y]-\alpha\right) \\
&:= \Pi_{\tau,L} + \Pi_{\tau,S},
\end{aligned}$$

where $\Pi_{\tau,L} = \beta\dot{\ell}(0)$ is the location effect and $\Pi_{\tau,S} = -\dot{s}(0)\frac{\beta^2}{\beta^2+1}\left(Q_\tau[Y]-\alpha\right)$ is the scale effect.

Next, we have

$$E[X|Y=y] = \frac{Cov(X,Y)}{Var(Y)}(y-\alpha) = \frac{\beta}{\beta^2+1}(y-\alpha).$$

Taking $y = Q_\tau[Y]$ yields

$$E[X|Y=Q_\tau[Y]] = \frac{\beta}{\beta^2+1}\left(Q_\tau[Y]-\alpha\right).$$

Therefore, we obtain the alternative expression $\Pi_{\tau,S} = -\dot{s}(0)E[X\beta|Y=Q_\tau[Y]]$.

S.2 Details of Example 5

When $F_{Y|X}(Q_\tau[Y]|x) = G(a_\tau + b_\tau x)$ for a standard normal cdf G. We have

$$\Pi^{\mu_X}_{\tau,S} = \frac{\dot{s}(0)}{f_Y(Q_\tau[Y])}\sigma_X^2 b_\tau^2 E\left[\dot{g}(a_\tau + b_\tau X)\right], \tag{S.1}$$

where

$$g(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \text{ and } \dot{g}(y) = -g(y)y.$$

Therefore,

$$E\left[\dot{g}(a_\tau + b_\tau X)\right] = -\frac{1}{2\pi\sigma_X} \int_{-\infty}^{\infty} (a_\tau + b_\tau x) \exp\left(-\frac{1}{2}\left[(a_\tau + b_\tau x)^2 + \left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]\right) dx.$$

First, we complete the squares to recover a Gaussian pdf. We have

$$(a_\tau + b_\tau x)^2 + \left(\frac{x-\mu_X}{\sigma_X}\right)^2 = \left(b_\tau^2 + \frac{1}{\sigma_X^2}\right) x^2 + 2\left(a_\tau b_\tau - \frac{\mu_X}{\sigma_X^2}\right) x + b_\tau^2 + \frac{\mu_X^2}{\sigma_X^2}.$$

Define

$$\begin{aligned}
K_{1,\tau} &:= \left(b_\tau^2 + \frac{1}{\sigma_X^2}\right)^{-1}, \\
K_{2,\tau} &:= -K_{1,\tau}\left(a_\tau b_\tau - \frac{\mu_X}{\sigma_X^2}\right), \\
K_{3,\tau} &:= K_{1,\tau}\left(b_\tau^2 + \frac{\mu_X^2}{\sigma_X^2}\right).
\end{aligned}$$

Then, we have

$$\begin{aligned}
(a_\tau + b_\tau x)^2 + \left(\frac{x-\mu_X}{\sigma_X}\right)^2 &= K_{1,\tau}^{-1}\left(x^2 - 2K_{2,\tau}x + K_{3,\tau}\right) \\
&= K_{1,\tau}^{-1}\left(x^2 - 2K_{2,\tau}x + K_{2,\tau}^2 - K_{2,\tau}^2 + K_{3,\tau}\right) \\
&= K_{1,\tau}^{-1}\left(x - K_{2,\tau}\right)^2 + K_{1,\tau}^{-1}\left(K_{3,\tau} - K_{2,\tau}^2\right).
\end{aligned}$$

It then follows that

$$\begin{aligned}
&\exp\left(-\frac{1}{2}\left[(a_\tau + b_\tau x)^2 + \left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[K_{1,\tau}^{-1}\left(x - K_{2,\tau}\right)^2 + K_{1,\tau}^{-1}\left(K_{3,\tau} - K_{2,\tau}^2\right)\right]\right) \\
&= \sqrt{2\pi K_{1,\tau}} \exp\left[K_{1,\tau}^{-1}\left(K_{3,\tau} - K_{2,\tau}^2\right)\right] \cdot \frac{1}{\sqrt{2\pi K_{1,\tau}}} \exp\left(-\frac{1}{2}K_{1,\tau}^{-1}\left(x - K_{2,\tau}\right)^2\right).
\end{aligned}$$

Next, we go back to the integral that we are interested in. For $\mathbb{X} \sim N(K_{2,\tau}, K_{1,\tau})$, we have

$$
\begin{aligned}
&E\left[\dot{g}(a_\tau + b_\tau X)\right] \\
&= -\frac{1}{\sqrt{2\pi}\sigma_X}\sqrt{K_{1,\tau}}\exp\left[K_{1,\tau}^{-1}\left(K_{3,\tau} - K_{2,\tau}^2\right)\right] \times E\left(a_\tau + b_\tau \mathbb{X}\right) \\
&= -\frac{\sqrt{K_{1,\tau}}\exp\left[K_{1,\tau}^{-1}\left(K_{3,\tau} - K_{2,\tau}^2\right)\right]}{\sqrt{2\pi}\sigma_X}\left(a_\tau + b_\tau K_{2,\tau}\right).
\end{aligned}
$$

Now, consider the case where $Y = \alpha + X\beta + U$ and $X \perp U$, and U is a standard normal. Note that

$$
\begin{aligned}
F_{Y|X}(Q_\tau[Y]|x) &= \Pr\left(\alpha + X\beta + U < Q_\tau[Y]|X = x\right) \\
&= \Pr\left(U < Q_\tau[Y] - \alpha - x\beta|X = x\right) = G(Q_\tau[Y] - \alpha - x\beta).
\end{aligned}
$$

So, in this case, $a_\tau = Q_\tau[Y] - \alpha$, $b_\tau = -\beta$. Therefore,

$$
\begin{aligned}
K_{1,\tau} &:= \left(b_\tau^2 + \frac{1}{\sigma_X^2}\right)^{-1} = \left(\beta^2 + \frac{1}{\sigma_X^2}\right)^{-1} \\
K_{2,\tau} &:= -K_{1,\tau}\left(a_\tau b_\tau - \frac{\mu_X}{\sigma_X^2}\right) = K_{1,\tau}\left(\left(Q_\tau[Y] - \alpha\right)\beta + \frac{\mu_X}{\sigma_X^2}\right) \\
K_{3,\tau} &:= K_{1,\tau}\left(a_\tau^2 + \frac{\mu_X^2}{\sigma_X^2}\right) = K_{1,\tau}\left(\left(Q_\tau[Y] - \alpha\right)^2 + \frac{\mu_X^2}{\sigma_X^2}\right).
\end{aligned}
$$

Now, by some simple algebra, we have

$$
\begin{aligned}
K_{1,\tau}^{-1}\left(K_{3,\tau} - K_{2,\tau}^2\right) &= \left(Q_\tau[Y] - \alpha\right)^2 + \frac{\mu_X^2}{\sigma_X^2} - \frac{\left(\left(Q_\tau[Y] - \alpha\right)\beta + \frac{\mu_X}{\sigma_X^2}\right)^2}{\beta^2 + \frac{1}{\sigma_X^2}} \\
&= \frac{\left(Q_\tau[Y] - \alpha - \mu_X\beta\right)^2}{\sigma_X^2\beta^2 + 1}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&E\left[\dot{g}(a_\tau + b_\tau X)\right] \\
&= -\left(Q_\tau[Y] - \alpha\right)\frac{\sqrt{K_{1,\tau}}\exp\left[-\frac{1}{2}\frac{\left(Q_\tau[Y]-\alpha-\mu_X\beta\right)^2}{\sigma_X^2\beta^2+1}\right]}{\sqrt{2\pi}\sigma_X} + \beta\frac{\sqrt{K_{1,\tau}}\exp\left[-\frac{1}{2}\frac{\left(Q_\tau[Y]-\alpha-\mu_X\beta\right)^2}{\sigma_X^2\beta^2+1}\right]}{\sqrt{2\pi}\sigma_X}K_{2,\tau} \\
&= f_Y(Q_\tau[Y])\left[-\left(Q_\tau[Y] - \alpha\right) + \beta\frac{\left(Q_\tau[Y] - \alpha\right)\beta + \frac{\mu_X}{\sigma_X^2}}{\beta^2 + \frac{1}{\sigma_X^2}}\right] \\
&= f_Y(Q_\tau[Y])\frac{\alpha + \mu_X\beta - Q_\tau[Y]}{\sigma_X^2\beta^2 + 1},
\end{aligned}
$$

where we have used

$$f_Y(Q_\tau[Y]) = \frac{\exp\left[-\frac{1}{2}\frac{(Q_\tau[Y]-\alpha-\mu_X\beta)^2}{\sigma_X^2\beta^2+1}\right]}{\sqrt{2\pi\left(\sigma_X^2\beta^2+1\right)}} = \frac{\sqrt{K_{1,\tau}}\exp\left[-\frac{1}{2}\frac{(Q_\tau[Y]-\alpha-\mu_X\beta)^2}{\sigma_X^2\beta^2+1}\right]}{\sqrt{2\pi}\sigma_X}.$$

Going back to (S.1), we obtain

$$\begin{aligned}
\Pi_{\tau,S}^{\mu_X} &= \frac{\dot{s}(0)}{f_Y(Q_\tau[Y])}\sigma_X^2 b_\tau^2 E\left[\dot{g}(a_\tau + b_\tau X)\right] \\
&= \dot{s}(0)\sigma_X^2\beta^2\frac{\alpha+\mu_X\beta - Q_\tau[Y]}{\sigma_X^2\beta^2+1} \\
&= \dot{s}(0)\sigma_X^2\beta^2\frac{\alpha+\mu_X\beta - \left[\alpha+\mu_X\beta+\sqrt{\sigma_X^2\beta^2+1}Q_\tau\left(U\right)\right]}{\sigma_X^2\beta^2+1} \\
&= -\dot{s}(0)\frac{\sigma_X^2\beta^2}{\sqrt{\sigma_X^2\beta^2+1}}Q_\tau[U],
\end{aligned}$$

where we have used $Q_\tau[Y] = \alpha + \mu_X\beta + \sqrt{\sigma_X^2\beta^2+1}Q_\tau\left(U\right)$.