2-2012

Pricing for Goodwill: A Threshold Quantile Regression Approach

Heng JU
Shanghai Jiaotong University

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Pai XU
University of Hong Kong

Pricing for Goodwill:
A Threshold Quantile Regression Approach*

Heng Ju,[a] Liangjun Su,[b] Pai Xu[c]
[a] Shanghai University of Finance and Economics, China
[b] Singapore Management University, Singapore
[c] University of Hong Kong, Hong Kong

February 2012

Abstract

In the absence of other effective trust systems, an agent's reputation status becomes a critical factor in online transactions. A higher reputation category may give sellers an advantage in competition on online trading platforms. It is also possible that such reputation benefits provide sufficient incentives for sellers to adjust their pricing behavior. We here propose a simple economic model in which an online seller maximizes the sum of the profit from current sales and the possible future gain from a targeted higher reputation level. We show that the model can predict a jump in optimal pricing behavior. We adopt a quantile regression threshold model (QRTM) to identify and explore such a pricing pattern as the "goodwill effect" in this paper. The use of a QRTM also allows us to model the heterogeneous behavior of different online sellers. We apply the proposed estimation and testing strategies to a data set obtained from Taobao.com, a leading online trading platform in China. We find both heterogeneities and jumps in a seller's goodwill pricing strategy in our application.

JEL Classifications: L10; C12; C13

Key Words: Heterogeneity; Pricing strategy; Reputation; Structural change; Threshold quantile regression.

*Contact: Heng Ju, School of International Business, Shanghai University of Finance and Economics, E-mail: hengju@gmail.com; Phone: +86 20 6590 3149; Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903, E-mail: ljsu@smu.edu.sg; Phone: +65 6828 0386; Pai Xu, School of Economics and Finance, the University of Hong Kong, 908 K.K.Leung Building, Pokfulam, Hong Kong, E-mail: paixu@hku.hk; Phone: +852 2857 8501.

1 Introduction

The global connectivity of the Internet offers potential buyers and sellers unprecedented opportunities to engage in arms-length transactions with distant partners. Our societies thus need to develop new trust mechanisms that are capable of ensuring cooperation and efficiency in a universe of strangers. Online reputation systems currently draw upon the Internet's bi-directional communication capabilities to engineer large-scale word-of-mouth networks in which individuals share opinions on and experiences of a wide range of topics, including, of particular interest here, products and services. Empirical studies have documented extensive evidence to show that sellers enjoy greater benefits from a better reputation. See, for example, Bolton et al. (2004) and Resnick et al. (2006).

Although it is heartening to know that reputation confers rewards, it is also of interest to economists to know whether a reputation system exerts any influence on a seller's market behavior. The reputation scoring mechanism has recently emerged as one of the most promising solutions to the problem of trust-building on the Internet. Typically, a transaction party can obtain positive/negative or neutral feedback from her trading partner, which, accordingly, adds/deducts one point or zero point to her total reputation score.[1] Buyers observe sellers' reputation scores when searching for a potential transaction partner. In this paper, we investigate whether and how a selling strategy may be affected by a reputation system of this kind.

Taobao.com is a Chinese-language online shopping website similar to eBay. It facilitates business-to-consumer and consumer-to-consumer retail by providing a platform for businesses and individual entrepreneurs to open online retail stores that cater mainly to consumers in mainland China, Hong Kong, Macau, and Taiwan. Sellers are able to post both new and used goods for sale on the Taobao marketplace, although the overwhelming majority of products is brand-new merchandise sold at a fixed price.[2] At the end of 2010, Taobao reported more than 370 million registered users and more than 800 million product listings. In 2010, the gross merchandise volume sold on Taobao.com was valued at 400 billion Chinese yuan (approximately US$61 billion), about ten times greater than the value sold in 2007. In January 2011, Taobao.com ranked 13th overall in Alexa's Internet traffic rankings.

Taobao.com exhibits an interesting phenomenon. When posting an item for sale, sellers sometimes explicitly indicate that it is "on sale" for the purpose of striving for the next category of reputation. It is this observation that motivates the study reported in this paper. We first provide an economic model to explain the pricing strategy that a seller may adopt when the benefit from a better reputation concerns the pricing decision. In our model, the seller maximizes the sum of the profit of current sales and the possible future gain from a better reputation. We demonstrate that, at a certain threshold on the reputation level, a seller may decide to undercut the current price in exchange for the future gain. Such a pricing pattern entails a "jump" or "structural change" in the pricing rule. We therefore refer to this pattern as "pricing for goodwill" in this paper.

In view of the presence of heterogeneous sellers in the market, we recognize that high-end sellers

[1] For example, eBay employs a version of this type of reputation mechanism in practice.

[2] Auctions are another selling mechanism on Taobao.com, although they account for a very small percentage of transactions.

may adopt completely different pricing strategies than middle- and low-end sellers. Indeed, Cabral and Hortascu (2010) acknowledged the existence of significant unobservable seller heterogeneity in the electronic market. Such heterogeneity motivates us to adopt a quantile regression model *a la* Koenker and Bassett (1978) to investigate the jump behavior of pricing. It is well known that quantile regressions are a flexible way to model the heterogeneous influences of explanatory variables on the response variable of interest, which is the selling price here. To allow for possible structural breaks, we extend the mean regression threshold model of Hansen (2000) and the median regression threshold model of Caner (2002) to our quantile regression model, which we term the quantile regression threshold model (QRTM hereafter). Based on the QRTM, we first test for the existence of a change (threshold) point in pricing behavior, and we then consider the estimation of both the change point and the jump size. When threshold points are detected at different quantile indices, it is also interesting to determine whether they are the same. Therefore, we also develop an inferential method that allows us to test for the existence of a common break and estimate the common change point and jump size.

We collected trading data for the iPod Nano from Taobao.com for the last four months of 2009, to which we apply the proposed methodology. Our empirical results indicate that sellers at different quantiles (of prices) exhibit quite different pricing behavior, although most of them employ a pricing for goodwill strategy predicted in the model.

It is worth mentioning that, from the application perspective, this paper also subtly enriches the empirical literature on the regression discontinuity design (RDD). In the typical RDD framework, researchers are interested in the causal effect of a binary intervention or treatment. This design arises frequently in studies of administrative decisions. The basic idea behind the RDD is that assignment to a treatment group is determined by whether the value of a predictor (a covariate) lies on one side of a fixed threshold. Then, any discontinuity in the conditional distribution of the outcome as a function of the covariate at the cutoff value can serve as evidence for the causal effect of the treatment.[3]

At the heart of identifying assumptions to validate the RDD framework, the covariate is connected with the potential outcomes in a continuous (smooth) way. However, it has gradually caught practitioners' attention that public knowledge of the treatment assignment rule may threaten such a continuity assumption. Calling this the "manipulation problem," McCrary (2008) pointed out that "when the individuals know of the selection rule for treatment, are interested in being treated, and have time to fully adjust their behavior accordingly," the validity of the identification arguments in the RDD approach may fail to hold. McCrary proposed a test for the discontinuity at the cutoff in the density function of the covariate.[4] This paper instead provides a complete picture of how agents adjust their behavior when approaching the treatment threshold (if we consider the "next reputation category" as a treatment). We contribute to the literature by documenting a scenario in which, at individuals' optimal behavior, another endogenous cutoff may occur in accordance with the incentive to achieve an exogenous threshold for the treatment.

The rest of the paper is organized as follows. Section 2 provides an economic model to illustrate the

[3]Empirical applications of such a framework are abundant: see, Imbens and Lemieux (2008) for a detailed survey.

[4]In a recent working paper, Bajari et al. (2010) investigated how a RDD estimation strategy can be modified to solve the manipulation problem in a health care contract application.

pricing pattern when goodwill is of concern. We then examine the testing and estimation strategies for this pricing pattern in a QRTM framework in Section 3. The empirical application is investigated in Section 4. Finally, Section 5 presents some concluding remarks. All mathematical proofs are collected in the appendix.

2 An Economic Model of Pricing for Goodwill

In this section, we present an economic model to investigate the pricing behavior of sellers when they take the goodwill effect into account when making decisions.[5]

Consider a monopolist with current reputation status (score) r who is selling a product with zero marginal cost.[6] A one-shot demand is $Q(p) = 1 - \alpha p$ (for $0 \leq p \leq 1/\alpha$, to guarantee non-negative sales). Among the sales made, the seller can receive a number of good reviews. When accumulating these good reviews to exceed a threshold $\bar{r}$, the seller can receive an extra (exogenous) profit β. The empirical literature has documented extensive evidence to show that sellers with a superior reputation generate significantly higher profits. This β can be thought of as the discounted future profit from operating with a better business reputation. Thus, the seller's expected profit function is given by

$$\begin{aligned}\Pi(p; r) &= \Pi_1(p) + \Pi_2(p; r)\\ &= \Pi_1(p) + \beta \cdot \Pr[R(p, e) \geq (\bar{r} - r)],\end{aligned}$$

where $R(p, e)$ denotes the accrued good reviews from sales by charging a price p, and e is a random factor that generates the randomness of Π_2 for any given (p, r). Therefore, Π_1 denotes the profit a seller obtains from the market without any concerns over goodwill benefits, and Π_2 is the expected gain in extra profit from goodwill.

We further specify $R(p, e) = 1 - \alpha p - e$. Note that in such a specification we implicitly assume that more sales (from charging lower prices) tend to generate more good reviews. Moreover, e can be understood as the part of the sales that incur bad reviews. Then, the probability of benefiting from goodwill is

$$\Pr[1 - \alpha p - e > \bar{r} - r] = F(1 - \alpha p - \bar{r} + r),$$

where F is the cumulative distribution function of e with density f, which is everywhere differentiable on its domain $[0, 1]$. The seller's profit function becomes

$$\Pi(p; r) = p(1 - \alpha p) + \beta F(1 - \alpha p - \bar{r} + r).$$

Taking the first-order derivative of Π with respect to p yields the first-order condition (FOC):

$$\frac{\partial \Pi}{\partial p}(p; r) = (1 - 2\alpha p) - \alpha\beta f(1 - \alpha p - \bar{r} + r) = 0. \tag{2.1}$$

[5] We are greatly indebted to Wing Suen who inspired us to establish this economic model.

[6] The zero marginal cost assumption is innocuous. The monopoly assumption can be thought of as a simplification of monopolistic competition, under which the firm's demand is residual demand.

It is worth noting that equation (2.1) implies, for any given r, that $\frac{\partial \Pi}{\partial p}(p;r) < 0$ if $p \geq \frac{1}{2\alpha} \equiv p^m$. Therefore, the optimal price in the model must entail a price cut from p^m if the concerns of goodwill matter.

Beyond the price cut from p^m at the optimum, we hope to generate a jump at a certain reputation level as the pricing strategy for goodwill in the model equilibrium. To this end, we make the following assumptions on the density function f. Let f' and f'' denote the first- and second-order derivatives of f.

Assumption M1. There exists $\hat{e} \in (0,1]$ such that $f(\hat{e}) < 1/(\alpha\beta)$ and $f'(\hat{e}) = 0$. Moreover, $f'(e) > 0 \;\; \forall\, e < \hat{e}$, and $f'(e) < 0 \;\; \forall\, e > \hat{e}$.

Assumption M2. There exists $\tilde{e} \in (0, \hat{e})$ such that $f'(\tilde{e}) > 2/(\alpha\beta)$. Moreover, $\lim_{e \to 0} f'(e) < 2/(\alpha\beta)$.

Assumption M1 implies that f is a unimodal density function. The height restriction on f in this assumption ensures that the first-order condition, equation (2.1), is equipped with a solution, thereby effectively ruling out an uninteresting case in which the goodwill effect would dominate over the current monopolistic pricing (and therefore the seller would charge zero price). Assumption M2 requires a special curvature on f to the left of its mode. This curvature induces increasing marginal returns on a segment of Π, which implies that the profit function Π is not globally concave.[7] Indeed, it is this particular curvature that delivers the pricing strategy for goodwill in the following proposition.

Proposition 2.1 *Suppose Assumptions M1 and M2 hold. Then the seller's optimal pricing strategy entails a regime change. That is, there exist a threshold of reputation γ_0 and two different pricing regimes $p_1^*(r)$ and $p_2^*(r)$ such that the seller's optimal pricing rule $p^*(r)$ is*

$$p^*(r) = \begin{cases} p_1^*(r) & \text{if } r \leq \gamma_0 \\ p_2^*(r) & \text{if } r > \gamma_0 \end{cases},$$

where $\frac{\partial p_1^*(r)}{\partial r} < 0$.

Proof. Assumptions M1 and M2 together imply that there must exist two points E_1, $E_2 \in (0, \hat{e})$ such that $f'(E_1) = f'(E_2) = 2/(\alpha\beta)$. Without loss of generality, we assume that $E_1 < E_2$, which in turn implies that $f(E_1) < f(E_2)$ by M1. Define v_1 and v_2 such that

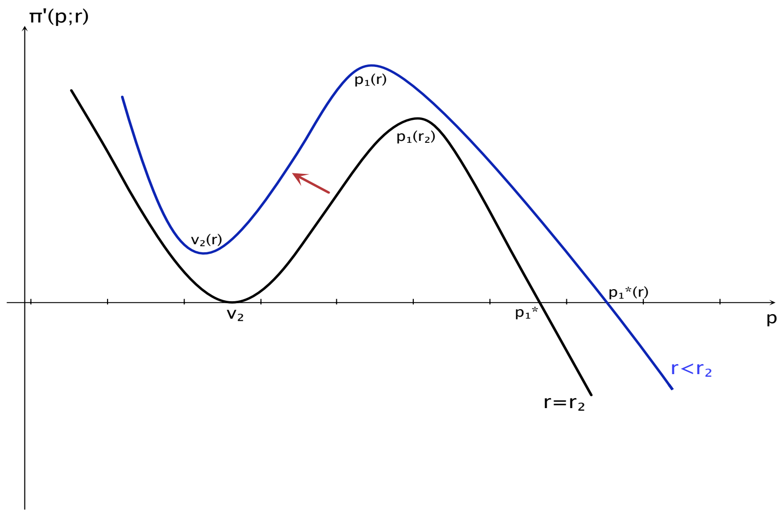
$$1 - 2\alpha v_1 - \beta\alpha f(E_1) = 0 \text{ and } 1 - 2\alpha v_2 - \beta\alpha f(E_2) = 0. \tag{2.2}$$

Then, we must have $v_1 > v_2$. Further, we define r_1 and r_2 such that

$$1 - \alpha v_1 - \bar{r} + r_1 = E_1 \text{ and } 1 - \alpha v_2 - \bar{r} + r_2 = E_2. \tag{2.3}$$

[7] We can extend the model by allowing for more general curvatures on the tails of f. Our major findings on pricing strategy in the model remain valid, but the extension unnecessarily complicates the analysis by introducing multiple optimal solutions. Therefore, we decide to retain the most simplifying assumption for ease of exposition.

Figure 1: Pricing Strategy When $r \leq r_2$



Consider r_1,

$$
\begin{aligned}
r_1 &= \bar{r} - 1 + \alpha v_1 + E_1 \\
&= \bar{r} - 1 + \alpha \frac{1 - \beta\alpha f(E_1)}{2\alpha} + E_1 \\
&= \bar{r} - \frac{1}{2} - \frac{\beta\alpha f(E_1)}{2} + E_1.
\end{aligned}
$$

Analogously, $r_2 = \bar{r} - \frac{1}{2} - \frac{\beta\alpha f(E_2)}{2} + E_2$. Therefore, by the mean value theorem there exists $\ddot{e} \in (E_1, E_2)$ such that

$$
\begin{aligned}
r_1 - r_2 &= -\frac{\beta\alpha}{2}[f(E_1) - f(E_2)] + (E_1 - E_2) \\
&= (E_1 - E_2)\left[1 - \frac{\beta\alpha}{2} f'(\ddot{e})\right] > 0,
\end{aligned}
$$

where the last inequality follows from the fact that $f'(e) > 2/(\alpha\beta)$ for any $e \in (E_1, E_2)$ by M2. Consequently we have shown that $r_1 > r_2$.

To understand the optimal pricing strategy in the model, we consider three cases: (1) $r \leq r_2$, (2) $r \geq r_1$, and (3) $r_2 < r < r_1$.

Case 1. When $r \leq r_2$.

At r_2, the point $p = v_2$ makes the FOC in (2.1) hold by construction. Further, $\frac{\partial^2 \Pi}{\partial p^2}(v_2; r_2) = -2\alpha + \alpha^2 \beta f'(E_2) = 0$ and $p = v_2$ is an inflexion point on the graph $\Pi(\cdot; r_2)$. Define $p_1(r_2) = v_2 + \frac{E_2 - E_1}{\alpha}$.

Using (2.2), (2.3), and the fact that $f'(E_1) = 2/(\alpha\beta)$, we can readily verify that $\frac{\partial \Pi}{\partial p}(p_1(r_2); r_2) = 1 - 2\alpha p_1(r_2) - \alpha\beta f(E_1) = 2(r_1 - r_2) > 0$ and $\frac{\partial^2 \Pi}{\partial p^2}(p_1(r_2); r_2) = -2\alpha + \alpha^2 \beta f'(E_1) = 0$. Therefore, $p_1(r_2)$ corresponds to a local maximum on the graph of $\frac{\partial \Pi}{\partial p}(\cdot; r_2)$ and an inflexion point on the $\Pi(\cdot; r_2)$ graph by M2. As $\frac{\partial \Pi}{\partial p}(p^m; r_2) < 0$, there must exist: $p_1^* \in (p_1(r_2), p^m)$ such that $\frac{\partial \Pi}{\partial p}(p_1^*; r_2) = 0$. Moreover, p_1^* is the unique maximum. (Refer to Figure 1.)

Extending to the case of $r < r_2$, define two local extremes, $v_2(r)$ and $p_1(r)$ on the function $\frac{\partial \Pi}{\partial p}(\cdot; r)$ with $v_2(r) < p_1(r)$. By definition, $\frac{\partial^2 \Pi}{\partial p^2}(v_2(r); r) = \frac{\partial^2 \Pi}{\partial p^2}(p_1(r); r) = 0$. It is easily verified that $\frac{\partial \Pi}{\partial p}(v_2; r) > 0$, $\frac{\partial \Pi}{\partial p}(p_1(r_2); r) > \frac{\partial \Pi}{\partial p}(p_1(r_2); r_2) > 0$, $\frac{\partial^2 \Pi}{\partial p^2}(p_1(r_2); r) < 0$, $\frac{\partial \Pi}{\partial p}(p_1^*; r) > 0$, and $\frac{\partial \Pi}{\partial p}(v_2(r); r) > \frac{\partial \Pi}{\partial p}(v_2(r); r_2) > 0 \;\forall\; r < r_2$.[8] The first three inequalities imply that $\forall\; r < r_2$, the graph of $\frac{\partial \Pi}{\partial p}(\cdot; r)$ can be obtained by shifting that of $\frac{\partial \Pi}{\partial p}(\cdot; r_2)$ to the upper left, and the last two, in conjunction with the fact that $\frac{\partial \Pi}{\partial p}(p^m; r) < 0$, imply the existence of a unique local maximum $p_1^*(r) \in (p_1^*, p^m) \;\; \forall\, r \le r_2$. By the FOC $\frac{\partial \Pi}{\partial p}(p_1^*(r); r) = 0$ and the implicit function theorem, we have

$$\frac{\partial p_1^*(r)}{\partial r} = -\frac{\frac{\partial^2 \Pi}{\partial p \partial r}(p_1^*(r); r)}{\frac{\partial^2 \Pi}{\partial p^2}(p_1^*(r); r)} = \frac{\alpha\beta f'(1 - \alpha p_1^*(r) - \bar{r} + r)}{\frac{\partial^2 \Pi}{\partial p^2}(p_1^*(r); r)} < 0$$

because $f'(e) > 0$ for any $e < \hat{e}$, $1 - \alpha p_1^*(r) - \bar{r} + r < 1 - \alpha p_1^* - \bar{r} + r_2 < \hat{e}$, and $\frac{\partial^2 \Pi}{\partial p^2}(p_1^*(r); r) < 0$. That is, $p_1^*(r)$ is **decreasing** in r.

Case 2. When $r \ge r_1$.

Note again that at r_1, the point $p = v_1$ is an inflexion point on the graph of $\Pi(\cdot; r_1)$. Similar to the arguments in Case (1), define $p_2(r_1) = v_1 - \frac{E_2 - E_1}{\alpha}$. Because $\frac{\partial \Pi}{\partial p}(p_2(r_1); r_1) < 0$ and $\lim_{p \to 0} \frac{\partial \Pi}{\partial p}(p; r_1) > 0$ by M1, there exists a $p_2^* \in (0, p_2(r_1))$ such that $\Pi(p; r_1)$ achieves a local maximum. (Refer to Figure 2.)

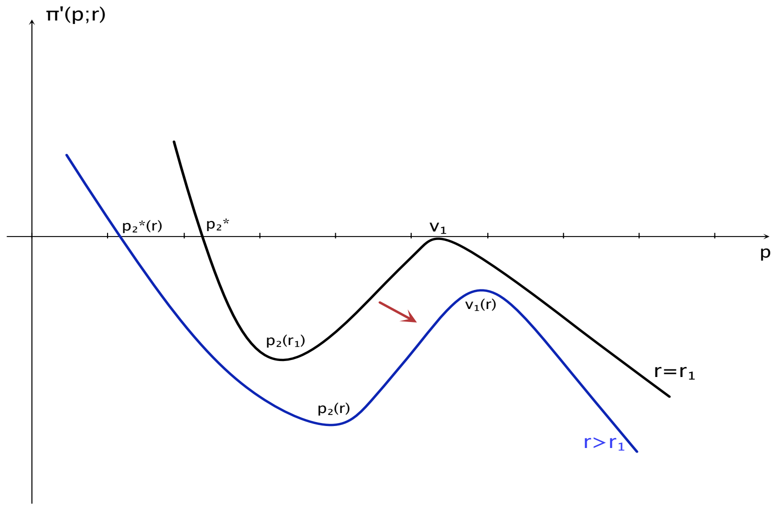
To extend to the case of $r > r_1$, let $p_2(r)$ denote the local minimum point on the function $\frac{\partial \Pi}{\partial p}(\cdot; r)$. We can apply arguments analogous to the case of $r < r_2$ to show that the graph of $\frac{\partial \Pi}{\partial p}(\cdot; r)$ can be obtained by shifting that of $\frac{\partial \Pi}{\partial p}(\cdot; r_2)$ down and right, and there exists a unique $p_2^*(r) \in (0, p_2(r)) \;\; \forall\, r > r_1$ that maximizes profits. However, noting that $1 - \alpha p_2^*(r) - \bar{r} + r$ can be either larger or smaller than $\hat{e}$, $f'(1 - \alpha p_2^*(r) - \bar{r} + r)$ can take positive, negative, or zero values, which implies that $p_2^*(r)$ may be **either increasing or decreasing** when $r > r_1$.[9]

---

[8] To obtain the last claim, we first note that $\frac{\partial^2 \Pi}{\partial p^2}(v_2(r); r) = -2\alpha + \alpha^2 \beta f'(1 - \alpha v_2(r) - \bar{r} + r) = 0$ implies that $f'(1 - \alpha v_2(r) - \bar{r} + r) = 2/(\alpha\beta)$. In fact, noting that $\frac{\partial \Pi}{\partial p}(\cdot; r)$ is convex around $v_2(r)$ under M2, we must have $1 - \alpha v_2(r) - \bar{r} + r = E_2 < \hat{e}$. It follows from M1 that $f(1 - \alpha v_2(r) - \bar{r} + r) < f(1 - \alpha v_2(r) - \bar{r} + r_2)\; \forall r < r_2$. Consequently, $\frac{\partial \Pi}{\partial p}(v_2(r); r) = 1 - \alpha v_2(r) - \alpha\beta f(1 - \alpha v_2(r) - \bar{r} + r) > 1 - \alpha v_2(r) - \alpha\beta f(1 - \alpha v_2(r) - \bar{r} + r_2) = \frac{\partial \Pi}{\partial p}(v_2(r); r_2) > 0$.

[9] Note that $1 - \alpha p_2^*(r) - \bar{r} + r > 1 - \alpha p^* - \bar{r} + r_1$, and nothing ensures that $1 - \alpha p_2^*(r) - \bar{r} + r < \hat{e}$ as $r > r_1$.

Figure 2: Pricing Strategy When $r \geq r_1$



Case 3. When $r_2 < r < r_1$**.**

There exist two local maxima, $p_1^*(r) \in (p_1(r), p^m)$ and $p_2^*(r) \in (0, p_2(r))$. Let $\triangle(r) = \Pi(p_1^*(r); r) - \Pi(p_2^*(r); r)$. By the envelope theorem and FOC,

$$\begin{aligned} \frac{\partial \triangle (r)}{\partial r} &= \beta f(1 - \alpha p_1^*(r) - \bar{r} + r) - \beta f(1 - \alpha p_2^*(r) - \bar{r} + r) \\ &= \frac{1 - 2\alpha p_1^*(r)}{\alpha} - \frac{1 - 2\alpha p_2^*(r)}{\alpha} \\ &= 2\left[p_2^*(r) - p_1^*(r)\right] < 0. \end{aligned}$$

Moreover, noting that $\triangle(r_2) > 0$ and $\triangle(r_1) < 0$, there must exist a unique point $\gamma_0 \in (r_2, r_1)$ such that $\triangle(\gamma_0) = 0$. It follows that the seller should adopt $p_1^*(r)$ if $r \leq \gamma_0$ and $p_2^*(r)$ otherwise, and the desired optimal pricing strategy holds. ■

Intuitively, the discontinuous pricing strategy occurs as follows. The restrictions in Assumptions M1 and M2 produce a peculiar shape of $\frac{\partial \Pi}{\partial p}(\cdot; r)$. Along with the increase in p, $\frac{\partial \Pi}{\partial p}(\cdot; r)$ is initially downward sloping and convex, then becomes positive sloping and concave, and then eventually slopes downward again. Thus, for any given $r \neq r_1$, r_2, there are three possible ways that $\frac{\partial \Pi}{\partial p}(\cdot; r)$ intersects with the horizontal axis:

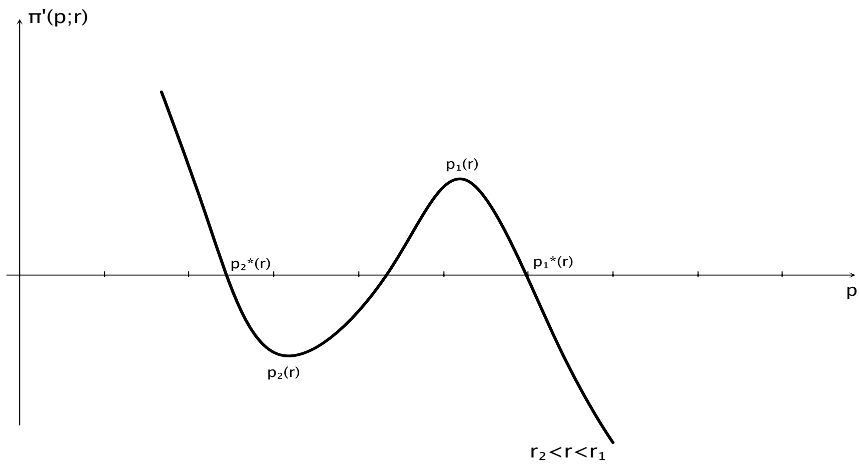
Case 1: the intersection occurs in the concave region alone (refer to Figure 1);

Case 2: the intersection occurs in the convex region alone (refer to Figure 2); and

Case 3: the intersection occurs in both regions (refer to Figure 3).

In the preceding proof, we show that the pricing scheme in Case 1 is associated with small values of

Figure 3: Pricing Strategy When $r \in (r_2, r_1)$



r, whereas that in Case 2 corresponds to large r's. Then, in Case 3, there exists a threshold value of r such that the seller will switch between the two pricing schemes. It is the presence of segment of a positively sloped $\frac{\partial \Pi}{\partial p}(\cdot; r)$ that makes the profit function $\Pi(;, r)$ exhibits a bimodal shape, which in turn induces discontinuity in the optimal pricing. If it were not the case, the profit function would be globally concave and pricing scheme change may not occur.

We first take a closer look at Case 1 by considering a slight increase in p when $p < v_2$. When r is small, the marginal profit in the current monopoly pricing always dominates the marginal cost of losing the potential benefit of goodwill. Therefore, the unique maximum of Π occurs in the concave region of $\frac{\partial \Pi}{\partial p}(\cdot; r)$ in this case. The decreasing pricing pattern in r simply reflects the fact that the potential benefit of goodwill becomes more significant as r increases.

Parallel to the first case, we next consider a decrease in p at $p > v_1$ in Case 2. When r is small, the loss of marginal profit in the current monopoly pricing may now be compensated for by the potential gain from future goodwill. Therefore, the unique maximum in this case occurs in the convex region of $\frac{\partial \Pi}{\partial p}(\cdot; r)$. In this case, as r is sufficiently close to $\bar{r}$, the pricing decision may face e on either side of $\hat{e}$. Therefore, the pricing pattern in r results in an ambiguous sign. On the graphs, this entails slope changes on $\frac{\partial \Pi}{\partial p}(\cdot; r)$ as it shifts down and right. Figure 2 shows only one of the possibilities, that is, that the curve gets a flatter slope when $\frac{\partial \Pi}{\partial p}(\cdot; r)$ shifts in connection with an increase in r.

In the pricing situation in Case 3, the seller needs to choose between two local maxima, $p_1^*(r)$ and $p_2^*(r)$, as illustrated in Figure 3. Switching from $p_1^*(r)$ to $p_2^*(r)$ induces a trade-off between the two areas in the region, $(p_2^*(r), p_1^*(r))$. The size of the gain is represented by the area below the horizontal axis, whereas the magnitude of loss is shown by the area above the horizontal axis. Consider a seller with an r close to r_2. As the gain from changing is not significant enough to compensate for the loss, the seller will continue to charge $p_1^*(r)$. However, along with the increase in r, there must exist a γ_0 that makes it worthwhile for the seller to switch to pricing regime $p_2^*(r)$.

In summary, the seller's optimal pricing schedule when $r \in [\gamma_0, \bar{r}]$ is different from that when $r < \gamma_0$. There is a discontinuity in the pricing function, which occurs at γ_0. In our model, such a "jump" reflects the local maxima switches at the seller's optimal pricing decision. It is this particular pricing pattern that is referred as the "goodwill effect" in this paper.

The mechanism for generating the discontinuity in the pricing strategy relies on the shape of underlying density function f. The exact location of γ_0, however, hinges on the model parameters of α and β, which capture the market demand situation and the seller's perceived gain from future goodwill, respectively. We may naturally expect that the sellers are heterogeneous and that different sellers face different sets of such model parameters upon making their decisions. As in any typical empirical work, our economic model can also accommodate other covariates to control the observed heterogeneity across sellers. Beyond this, we adopt quantile regression approach to our econometric analysis. Quantile regressions are a flexible way to model the heterogeneous influences of explanatory variables on the response variable, which is the selling price here. We thus develop a method to identify and estimate the heterogeneity in the pricing patterns for goodwill in the threshold quantile regression framework.

To proceed, it is worth mentioning that nothing in the economic model can ensure that the optimal pricing strategy before or after a certain threshold value (γ_0 above) is linear in reputation level r. Despite this, we still focus on a linear QRTM in the next section for two reasons. First, the linear QRTM can be regarded as a first-order approximation to the true optimal pricing strategy. Second, the vector of regressors in our model is allowed to include any transformation of the reputation level (say, r, r^2, r^3, etc.), and thus the quadratic or cubic terms can be included in the regression model to take into account the potential nonlinearity in the optimal pricing strategy.

3 A Quantile Regression Threshold Model

In this section, we first test for the existence of a change (threshold) point in pricing behavior. After confirming the change point, we then consider the estimation of both the change point and the jump size.

3.1 A simple quantile regression threshold model

Let $\{y_i, x_i, r_i\}_{i=1}^n$ be an independent sample, where y_i and r_i are real-valued and x_i is a $k \times 1$ random vector. The threshold variable r_i may be an element of x_i, and is assumed to have a continuous distribution. We assume that the τth conditional quantile of y_i, given x_i and r_i, is given by

$$Q_\tau (z_i) = \alpha_\tau' x_i 1 \{r_i \leq \gamma_\tau\} + \beta_\tau' x_i 1 \{r_i > \gamma_\tau\}, \tag{3.1}$$

where $z_i \equiv (x_i', r_i)'$ if $r_i \notin x_i$, and $z_i \equiv x_i$ otherwise; $1\{A\}$ is an indicator function that takes a value of one if A holds true, and zero otherwise; and $\delta_\tau \equiv \alpha_\tau - \beta_\tau$ may be nonzero for some unknown threshold point γ_τ. If δ_τ is 0 for all γ_τ on the support of r_i for all $\tau \in (0,1)$, then we can say there is no structural change in the quantile regression model (3.1). For technical simplicity, below we assume that γ_τ can only take values in a compact set Γ.

When applying the foregoing econometric model to study the optimal pricing behavior for goodwill, the threshold variable r_i is the reputation level, the response variable y_i is the observed price, and the regressor vector x_i can be specified in different ways. For example, one can set $x_i = (1, r_i)'$, which specifies linear pricing behavior before and after the potential break point γ_τ. Alternatively, one can set $x_i = \left(1, r_i, r_i^2\right)'$ to allow the optimal prices to depend on the reputation level in a nonlinear pattern. Of course, we can also include other control variables in x_i to account for the influence of other variables on prices.

Let $\theta_{1\tau} = (\alpha_\tau', \beta_\tau')'$ and $\theta_\tau = (\theta_{1\tau}', \gamma_\tau)'$. Define the "check function" $\rho_\tau(\cdot)$ by $\rho_\tau(u) = (\tau - 1\{u < 0\})\, u$. Following Koenker and Bassett (1978), we obtain the quantile estimate $\hat{\theta}_\tau$ of θ_τ as

$$\hat{\theta}_\tau = \underset{\theta_\tau}{\text{argmin}} S_{n\tau}(\theta_\tau) \text{ with } S_{n\tau}(\theta_\tau) = \sum_{i=1}^{n} \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right), \tag{3.2}$$

where

$$z_i(\gamma) = \left(x_i' 1\{r_i \le \gamma\},\ x_i' 1\{r_i > \gamma\}\right)'.$$

For this minimization, there is no closed form solution. In fact, the objective function is not convex in all of its parameters, and so it is difficult to obtain the global minimizer. Nevertheless, we can consider the profile quantile regression. For this, we first pretend that γ_τ is known, and obtain an estimate of $(\alpha_\tau, \beta_\tau)$ by

$$\left(\hat{\alpha}_\tau(\gamma_\tau), \hat{\beta}_\tau(\gamma_\tau)\right) \equiv \left(\hat{\alpha}(\tau, \gamma_\tau), \hat{\beta}(\tau, \gamma_\tau)\right) = \underset{\alpha_\tau, \beta_\tau}{\text{argmin}} S_{n\tau}(\alpha_\tau, \beta_\tau, \gamma_\tau). \tag{3.3}$$

Let $\hat{S}_{n\tau}(\gamma_\tau) \equiv S_{n\tau}(\hat{\alpha}_\tau(\gamma_\tau), \hat{\beta}_\tau(\gamma_\tau), \gamma_\tau)$. Then, we can estimate γ_τ by

$$\hat{\gamma}_\tau = \underset{\gamma \in \Gamma}{\text{argmin}} \hat{S}_{n\tau}(\gamma). \tag{3.4}$$

Noting that $\hat{S}_{n\tau}(\gamma)$ takes on fewer than n distinct values, we can define $\hat{\gamma}_\tau$ uniquely by choosing γ_τ over $\Gamma_n = \Gamma \cap \{q_1, q_2, \cdots, q_n\}$. Then, computing $\hat{\gamma}_\tau$ requires at most n function evaluations. As Hansen (2000) and Caner (2002) suggested, if n is large, then we can approximate Γ by a grid. After $\hat{\gamma}_\tau$ is obtained, we can compute the estimates of α_τ and β_τ as $\hat{\alpha}_\tau = \hat{\alpha}_\tau(\hat{\gamma}_\tau)$ and $\hat{\beta}_\tau = \hat{\beta}_\tau(\hat{\gamma}_\tau)$, respectively.

3.2 Test for the existence of a change point

The above computation is meaningful only if γ_τ is identified, in which case a structural change occurs for the τth conditional quantile regression. It is thus worthwhile to consider a test for the existence of a structural change before embarking on the estimation of γ_τ.

Let $\mathcal{T}=[\underline{\tau},\overline{\tau}]\subset(0,1)$. Let $\Theta_1\subset\mathbb{R}^{2k}$ denote the compact support for $\theta_{1\tau}$.[10] Let $\theta^0_{n1\tau}=\left(\alpha^{0\prime}_{n\tau},\beta^{0\prime}_{n\tau}\right)'$ denote the true value of $\theta_{1\tau}$. We allow $\theta^0_{n1\tau}$ to be n-dependent to facilitate the study of estimates of $\theta_{1\tau}$ in the case where we have a structural change but the jump size shrinks to zero as the sample size $n\to\infty$ (see Assumption A7 below). But, for notational simplicity, we suppress the dependence of $\theta^0_{n1\tau}$, $\alpha^0_{n\tau}$, and $\beta^0_{n\tau}$ on n and write them as $\theta^0_{1\tau}$, α^0_{τ}, and β^0_{τ}, respectively.

The null hypothesis is

$$\mathbb{H}_0: Q_\tau(z_i)=z_i(\gamma)'\theta^0_{1\tau} \text{ for some } \theta^0_{1\tau}\in\Theta_1 \text{ with } \alpha^0_\tau=\beta^0_\tau \text{ for all } \tau\in\mathcal{T}, \tag{3.5}$$

regardless of the value of $\gamma\in\Gamma$, and the alternative hypothesis is

$$\mathbb{H}_1: Q_\tau(z_i)=z_t\left(\gamma^0_\tau\right)'\theta^0_{1\tau} \text{ for some } \theta^0_{1\tau}\in\Theta_1 \text{ with } \alpha^0_\tau\neq\beta^0_\tau \text{ for some } \tau\in\mathcal{T}, \tag{3.6}$$

where $\gamma^0_\tau\in\Gamma$.

Clearly, both $\mathbb{H}_0$ and $\mathbb{H}_1$ are composite hypotheses, and they are designed to test the existence of a structural change at an arbitrary quantile point τ. For different τ's, the structural changes are allowed to occur at different threshold values γ^0_τ under $\mathbb{H}_1$. If we restrict our attention to a single quantile τ, i.e., $\mathcal{T}=\{\tau\}$, then we can consider the following null hypothesis.

$$\mathbb{H}_{0\tau}: Q_\tau(z_i)=z_i(\gamma)'\theta^0_{1\tau} \text{ for some } \theta^0_{1\tau}\in\Theta_1 \text{ with } \alpha^0_\tau=\beta^0_\tau, \tag{3.7}$$

regardless of the value of $\gamma\in\Gamma$, and the alternative hypothesis becomes

$$\mathbb{H}_{1\tau}: Q_\tau(z_i)=z_i\left(\gamma^0_\tau\right)'\theta^0_{1\tau} \text{ for some } \theta^0_{1\tau}\in\Theta_1 \text{ with } \alpha^0_\tau\neq\beta^0_\tau \text{ and } \gamma^0_\tau\in\Gamma. \tag{3.8}$$

The above formulation motivates us to consider the following τth quantile regression of y_i on $z_i(\gamma)$.

$$\hat\theta_1(\tau,\gamma)=\left(\hat\alpha(\tau,\gamma)',\ \hat\beta(\tau,\gamma)'\right)'=\underset{\theta_1}{\operatorname{argmin}}\sum_{i=1}^n\rho_\tau\left(y_i-\theta_1'z_i(\gamma)\right). \tag{3.9}$$

Noting that γ is not identified under $\mathbb{H}_0$ for any $\tau\in\mathcal{T}$, we cannot consider the estimation of γ_τ analogously to (3.4). Nevertheless, we can study the asymptotic property of $\hat\theta_1(\tau,\gamma)$ under $\mathbb{H}_0$, and propose a test for the null hypothesis of no structural change for all τ over $\mathcal{T}$ by considering the asymptotic behavior of $\hat\delta(\tau,\gamma)\equiv\hat\alpha(\tau,\gamma)-\hat\beta(\tau,\gamma)$ over the compact set $\mathcal{T}\times\Gamma$.

To proceed, we make the following assumptions.

Assumption A1. (y_i,z_i), $i=1,\cdots,n$, are independent and identically distributed (IID).

Assumption A2. $E\|x_i\|^2<\infty$.

Assumption A3. The conditional distribution function (CDF) $F(\cdot|z)$ of y_i given $z_i=z$ admits a Lebesgue probability density function (PDF) $f(\cdot|z)$ such that (i) $f(\cdot|z)$ is continuous for each z, and (ii) $f(\cdot|z)$ is uniformly bounded for each z.

Assumption A4. The threshold variable r_i is continuously distributed with continuous PDF $g(\cdot)$.

Assumption A5. Let $\Omega_1(\tau,\gamma)\equiv E[f\left(\theta^{0\prime}_{1\tau}z_i(\gamma)|z_i\right)z_i(\gamma)z_i(\gamma)']$. $\Omega_1(\tau,\gamma)$ is positive definite for each $(\tau,\gamma)\in\mathcal{T}\times\Gamma$.

[10] In principle, we can allow the support of $\theta_{1\tau}$ to be τ-dependent, and write Θ_1 as $\Theta_{1\tau}$.

Assumption A1 requires IID observations, but it can be weakened to allow for time series observations by using the concept of mixing processes, as in Bai (1995), Hansen (2000), Caner (2002), Su and Xiao (2008), and Galvao et al. (2010). Assumptions A2-A4 specify standard conditions on the threshold regressions; see, e.g., Galvao et al. (2011).

The following theorem is adapted from Galvao et al. (2011).

Theorem 3.1 *Suppose Assumptions A1-A5 hold. Then, under $\mathbb{H}_0$, $\hat{\theta}_1(\tau,\gamma)$ admits the following uniform Bahadur representation*

$$\sqrt{n}\left(\hat{\theta}_1(\tau,\gamma)-\theta_{1\tau}^0\right)=\Omega_1(\tau,\gamma)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\tau\left(y_i-\alpha_\tau^{0\prime}x_i\right)z_i(\gamma)+o_P(1),$$

where $\psi_\tau(u)=\tau-1\{u<0\}$, and $o_P(1)$ holds uniformly in $(\tau,\gamma)\in\mathcal{T}\times\Gamma$.

Let $\Omega_0(\gamma_1,\gamma_2)\equiv E[z_i(\gamma_1)z_i(\gamma_2)']$. Let $l^\infty(\mathcal{T}\times\Gamma)$ denote the space of all bounded functions on $\mathcal{T}\times\Gamma$ equipped with the uniform topology. From Theorem 3.1, we can easily obtain that under $\mathbb{H}_0$,

$$\sqrt{n}\left(\hat{\theta}_1(\tau,\gamma)-\theta_{1\tau}^0\right)\Rightarrow\Omega_1(\tau,\gamma)^{-1}W(\tau,\gamma)\ \text{ in }\ \left(l^\infty(\mathcal{T}\times\Gamma)\right)^{2k},$$

where $\Rightarrow$ denotes weak convergence, and $W(\tau,\gamma)$ is a zero-mean Gaussian process on $\mathcal{T}\times\Gamma$ with covariance kernel

$$E\left[W(\tau_1,\gamma_1)W(\tau_2,\gamma_2)'\right]=(\tau_1\wedge\tau_2-\tau_1\tau_2)\,\Omega_0(\gamma_1,\gamma_2).$$

Observe that

$$\Omega_0(\gamma,\gamma)=\begin{pmatrix}\Omega(\gamma) & \mathbf{0}_{k\times k}\\ \mathbf{0}_{k\times k} & \Omega^*(\gamma)\end{pmatrix},\ \Omega_1(\tau,\gamma)=\begin{pmatrix}\Omega(\tau,\gamma) & \mathbf{0}_{k\times k}\\ \mathbf{0}_{k\times k} & \Omega^*(\tau,\gamma)\end{pmatrix},$$

where $\Omega(\gamma)=E[x_ix_i'1\{r_i\le\gamma\}]$, $\Omega^*(\gamma)=E[x_ix_i'1\{r_i>\gamma\}]$, $\Omega(\tau,\gamma)=E\left[x_ix_i'1\{r_i\le\gamma\}f\left(\alpha_\tau^{0\prime}x_i|z_i\right)\right]$, and $\Omega^*(\tau,\gamma)=E\left[x_ix_i'1\{r_i>\gamma\}f\left(\beta_\tau^{0\prime}x_i|z_i\right)\right]$. Let $\hat{\delta}(\tau,\gamma)=\hat{\alpha}(\tau,\gamma)-\hat{\beta}(\tau,\gamma)=R\hat{\theta}_1(\tau,\gamma)$, where $R=[I_k,-I_k]$ and I_k is a $k\times k$ identity matrix. The block diagonality of Ω_0 and Ω_1 implies the asymptotic independence between $\hat{\alpha}(\tau,\gamma)$ and $\hat{\beta}(\tau,\gamma)$. It follows that for each (τ,γ)

$$\sqrt{n}\hat{\delta}(\tau,\gamma)\xrightarrow{d}N\left(\mathbf{0}_{k\times1},\tau(1-\tau)V(\tau,\gamma)\right),\tag{3.10}$$

where $\xrightarrow{d}$ denotes convergence in distribution, and

$$V(\tau,\gamma)\equiv\Omega(\tau,\gamma)^{-1}\Omega(\gamma)\Omega(\tau,\gamma)^{-1}+\Omega^*(\tau,\gamma)^{-1}\Omega^*(\gamma)\Omega^*(\tau,\gamma)^{-1}.\tag{3.11}$$

Let

$$\hat{\Omega}(\gamma)\equiv n^{-1}\sum_{i=1}^n x_ix_i'1\{r_i\le\gamma\},\ \hat{\Omega}^*(\gamma)\equiv n^{-1}\sum_{i=1}^n x_ix_i'-\hat{\Omega}(\gamma),$$

$$\hat{\Omega}(\tau,\gamma)\equiv(2nh)^{-1}\sum_{i=1}^n 1\{|y_i-x_i'\hat{\alpha}_\tau|\le h\}x_ix_i'1\{r_i\le\gamma\},\ \text{and}$$

$$\hat{\Omega}^*(\tau,\gamma)\equiv(2nh)^{-1}\sum_{i=1}^n 1\left\{\left|y_i-x_i'\hat{\beta}_\tau\right|\le h\right\}x_ix_i'1\{r_i>\gamma\},$$

where $h \equiv h(n)$ is a bandwidth parameter such that $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$ (see Koenker, 2005, pp. 80-81). It is easy to show that the above estimators are uniformly consistent with $\Omega(\gamma)$, $\Omega^*(\gamma)$, $\Omega(\tau,\gamma)$, and $\Omega^*(\tau,\gamma)$, respectively, over $\mathcal{T} \times \Gamma$. Thus, a uniformly consistent estimate of $V(\tau,\gamma)$ is given by

$$\hat{V}(\tau,\gamma) \equiv \hat{\Omega}(\tau,\gamma)^{-1} \hat{\Omega}(\gamma) \hat{\Omega}(\tau,\gamma)^{-1} + \hat{\Omega}^*(\tau,\gamma)^{-1} \hat{\Omega}^*(\gamma) \hat{\Omega}^*(\tau,\gamma)^{-1}. \tag{3.12}$$

Following Qu (2008), Su and Xiao (2008), and Galvao et al. (2010), we can consider the sup-Wald statistic for testing $\mathbb{H}_0$ given by

$$\sup W_n \equiv \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} W_n(\tau,\gamma), \tag{3.13}$$

where $W_n(\tau,\gamma) = n\hat{\delta}(\tau,\gamma)' [\tau(1-\tau)\hat{V}(\tau,\gamma)]^{-1}\hat{\delta}(\tau,\gamma)$, and we reject $\mathbb{H}_0$ for large values of $\sup W_n$. By the above discussions, the Slutsky theorem, and the continuous mapping theorem (CMT), we have

$$\begin{aligned} \sup W_n \xrightarrow{d} \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} & \frac{1}{\tau(1-\tau)} W(\tau,\gamma)' \Omega_1(\tau,\gamma)^{-1} R' \left[R\Omega_1(\tau,\gamma)^{-1} \Omega_0(\gamma,\gamma) \Omega_1(\tau,\gamma)^{-1} R'\right]^{-1} \\ & \times R\Omega_1(\tau,\gamma)^{-1} W(\tau,\gamma). \end{aligned} \tag{3.14}$$

Let A be a $p_1 \times q$ matrix and B a $q \times p_2$ matrix such that rank$(A) = q =$rank(B). Let A^+ denote the Moore-Penrose generalized inverse of A. Then,

$$(AB)^+ = B^+A^+ = B'(BB')^{-1}(A'A)^{-1}A' \tag{3.15}$$

by Fact 6.4.8 in Bernstein (2005). Using (3.15) repeatedly, we can readily show that the right-hand side of (3.14) simplifies to

$$\sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} \frac{1}{\tau(1-\tau)} W(\tau,\gamma)' \Omega_0(\gamma,\gamma)^{-1} W(\tau,\gamma). \tag{3.16}$$

The following corollary summarizes the asymptotic distribution of $\sup W_n$.

Corollary 3.2 *Suppose Assumptions A1-A5 hold. Then, under* $\mathbb{H}_0$,

$$\sup W_n \xrightarrow{d} \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} \frac{1}{\tau(1-\tau)} W(\tau,\gamma)' \Omega_0(\gamma,\gamma)^{-1} W(\tau,\gamma).$$

Like Su and Xiao (2008, Theorem 3.3), the limiting distribution of $\sup W_n$ depends only on the Gaussian process $W(\tau,\gamma)$. Unlike Su and Xiao (2008), Corollary 3.2 indicates that the limiting distribution of $\sup W_n$ is not pivotal because $\bar{W}(\tau,\gamma) \equiv \Omega_0(\gamma,\gamma)^{-1/2} W(\tau,\gamma)$ is a Gaussian process with covariance kernel

$$E\left[\bar{W}(\tau_1,\gamma_1)\bar{W}(\tau_2,\gamma_2)'\right] = (\tau_1 \wedge \tau_2 - \tau_1\tau_2)\Omega_0(\gamma_1,\gamma_1)^{-1/2}\Omega_0(\gamma_1,\gamma_2)\Omega_0(\gamma_2,\gamma_2)^{-1/2},$$

and $\Omega_0(\gamma_1,\gamma_2)$ is not parameter-free. This implies that the critical values for the $\sup W_n$ test cannot be tabulated in general.

Nevertheless, given the simple structure of $W(\tau,\gamma)$, we can readily simulate the critical values for the $\sup W_n$ test statistic. Given $\{z_i, i=1,...,n\}$, we can consistently estimate $\Omega_0(\gamma,\gamma)$ by

$$\hat{\Omega}_0(\gamma,\gamma) = \begin{pmatrix} \hat{\Omega}(\gamma) & \mathbf{0}_{k\times k} \\ \mathbf{0}_{k\times k} & \hat{\Omega}^*(\gamma) \end{pmatrix}$$

and then simulate the critical values of the $\sup W_n$ as follows.

1. Generate $\{u_i, i=1,...,n\}$ IID uniform on $[0,1]$.
2. Calculate $w_n(\tau,\gamma) = n^{-1/2}\sum_{i=1}^n [\tau - 1\{u_i \le \tau\}] z_i(\gamma)$.
3. Compute $\sup W_n^* \equiv \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} \frac{1}{\tau(1-\tau)} w_n(\tau,\gamma)' \hat{\Omega}_0(\gamma,\gamma)^{-1} w_n(\tau,\gamma)$. [11]
4. Repeat steps 1-3 B times and denote the resulting $\sup W_n^*$ test statistics as $\sup W_{n,j}^*$ for $j = 1,\cdots,B$.

Obviously, $w_n(\tau,\gamma) \Rightarrow W(\tau,\gamma)$ in $(l^\infty(\mathcal{T}\times\Gamma))^{2k}$. When B is sufficiently large, the asymptotic critical value of the level α test based on $\sup W_n$ is approximately given by the empirical upper α-quantile of $\{\sup W_{n,j}^*,\ j=1,...,B\}$. Therefore, we can reject the null hypothesis $\mathbb{H}_0$ if the simulated p-value

$$p^* \equiv B^{-1}\sum_{j=1}^{B} 1\left\{\sup W_{n,j}^* \ge \sup W_n\right\}$$

is smaller than the prescribed nominal level of significance α.

Note that by choosing $\mathcal{T}$ as a large compact subset of $(0,1)$, the above test can detect various violations of the null hypothesis. Alternatively, by specifying $\mathcal{T} = \{\tau\}$ we can consider the test of structural change at a single quantile τ. In the case in which we reject $\mathbb{H}_{0\tau}$ for the specified τ, we can further consider the estimation of the break point γ_τ under $\mathbb{H}_{1\tau}$.

3.3 Asymptotic properties of $\hat{\theta}_\tau$ under $\mathbb{H}_{1\tau}$

Let $\hat{\theta}_{1\tau} = (\hat{\alpha}_\tau', \hat{\beta}_\tau')'$. In this subsection, we first prove the consistency of $\hat{\theta}_\tau = (\hat{\theta}_{1\tau}', \hat{\gamma}_\tau)'$ under $\mathbb{H}_{1\tau}$, and then report the convergence rates and asymptotic distributions of $\hat{\theta}_{1\tau}$ and $\hat{\gamma}_\tau$.

To study the consistency of $\hat{\theta}_\tau$, we add the following identification condition.

Assumption A6. Let $\Delta(z_i,\theta_\tau) \equiv z_i(\gamma_\tau)'\theta_{1\tau} - z_i(\gamma_\tau^0)'\theta_{1\tau}^0$. There exists: $c_0 > 0$ such that $P(|\Delta(z_i,\theta_\tau)| > c_0) > 0$ for all $\theta_\tau \in \Theta$ such that $\theta_\tau \ne \theta_\tau^0$, where $\Theta = \Theta_1 \times \Gamma$.

In the special case of $x_i = (1, r_i)'$, we can write $\alpha_\tau^0 = (\alpha_{0\tau}^0, \alpha_{1\tau}^0)'$ and $\beta_\tau^0 = (\beta_{0\tau}^0, \beta_{1\tau}^0)'$, where $\alpha_{0\tau}^0$ and $\beta_{0\tau}^0$ are the true values of the intercept parameters before and after the break, and $\alpha_{1\tau}^0$ and $\beta_{1\tau}^0$ are the

[11] In practice, we compute the $\sup W_n$ by constructing a fine partition $\mathcal{T}_{m_1}\times\Gamma_{m_2} \subset \mathcal{T}\times\Gamma$ by a finite grid of $m_1 \times m_2$ points. In our applications, we set $m_1 = m_2 = 81$ and choose $\mathcal{T}_{81} = \{0.10, 0.11, \cdots, 0.90\}$ and Γ_{81} as the collection of the τth quantile of q_i for $\tau \in \mathcal{T}_{81}$. To obtain the simulated p-value, one can choose a finer partition because of the fast speed of computing $\sup W_{n,j}$.

true values of the slope parameters before and after the break. Let $d_\tau \equiv (\alpha_{0\tau} - \beta_{0\tau}) + (\alpha_{1\tau} - \beta_{1\tau})\,\gamma_\tau^0$. Then, a sufficient condition for Assumption A6 to hold is

$$d_\tau \neq 0.$$

The following theorem establishes the strong consistency of $\hat{\theta}_\tau$.

Theorem 3.3 *Suppose that Assumptions A1-A4 and A6 hold. Then,*

$$\hat{\theta}_\tau = \theta_\tau^0 + o_{a.s.}(1),$$

where $\theta_\tau^0 = (\theta_{1\tau}^{0\prime}, \gamma_\tau^0)'$.

To study the convergence rates of $\hat{\theta}_{1\tau}$ and $\hat{\gamma}_\tau$, we note that the convergence rate of $\hat{\gamma}_\tau$ depends on the size of the structural change and that $\hat{\theta}_{1\tau}$ and $\hat{\gamma}_\tau$ typically have different convergence rates.

The following assumption specifies the magnitude of structural change in the coefficients.

Assumption A7. Let $\delta_\tau^0 \equiv \delta_{n\tau}^0 \equiv \alpha_{n\tau}^0 - \beta_{n\tau}^0$. $\delta_\tau^0 = v_\tau n^{-a}$ with $v_\tau \neq 0$ and $a \in (0, \frac{1}{2})$.

Note that Hansen (2000) and Caner (2002) also allow $a \in (0, \frac{1}{2})$, in which case the break magnitude $\delta_\tau^0 = \delta_{n\tau}^0$ shrinks to 0 as $n \to \infty$. We also remark on the case of $a = 0$.

The following theorem establishes the convergence rates of $\hat{\theta}_{1\tau}$ and $\hat{\gamma}_\tau$ under $\mathbb{H}_{1\tau}$.

Theorem 3.4 *Suppose that Assumptions A1-A4 and A6-A7 hold. Then,*

$$n^{1/2}\left(\hat{\theta}_{1\tau} - \theta_{1\tau}^0\right) = O_P(1) \;\; \textit{and}\; n^{1-2a}\left(\hat{\gamma}_\tau - \gamma_\tau^0\right) = O_P(1).$$

In fact, the above convergence rates can also be obtained in the case of $a = 0$ in A7. To study the asymptotic distributions of $\hat{\theta}_{1\tau}$ and $\hat{\gamma}_\tau$, we add the following assumption.

Assumption A8. (i) Let $N_\tau(\gamma) \equiv E[x_i x_i' | r_i = \gamma]$, $D_\tau(\gamma) \equiv E\left[f\left(\alpha_\tau^{0\prime} x_i | z_i\right) x_i x_i' | r_i = \gamma\right]$, $N_\tau \equiv N_\tau\left(\gamma_\tau^0\right)$, and $D_\tau \equiv D_\tau\left(\gamma_\tau^0\right)$. $N_\tau(\gamma)$ and $D_\tau(\gamma)$ are continuous at γ_τ^0. $v_\tau' N_\tau v_\tau > 0$, $v_\tau' D_\tau v_\tau > 0$, and $g\left(\gamma_\tau^0\right) > 0$.

(ii) $E\|x_i\|^4 < \infty$.

Theorem 3.5 *Suppose that Assumptions A1-A4 and A6-A8 hold. Then,*

(i) $n^{1/2}\left(\hat{\theta}_{1\tau} - \theta_{1\tau}^0\right) \xrightarrow{d} N\left(\mathbf{0}_{2k\times 1}, \tau(1-\tau)\Sigma\left(\tau, \gamma_\tau^0\right)\right)$ *where* $\Sigma(\tau,\gamma) \equiv \Omega_1(\tau,\gamma)^{-1}\Omega_0(\gamma,\gamma)\Omega_1(\tau,\gamma)^{-1}$.

(ii) $n^{1-2a}\left(\hat{\gamma}_\tau - \gamma_\tau^0\right) \xrightarrow{d} \frac{\lambda_\tau}{4\mu_\tau^2} \underset{r\in(-\infty,\infty)}{\textit{argmax}} \left\{W(r) - \frac{1}{2}|r|\right\}$ *where* $\lambda_\tau \equiv v_\tau' N_\tau v_\tau g\left(\gamma_\tau^0\right)$, $\mu_\tau \equiv v_\tau' D_\tau v_\tau g\left(\gamma_\tau^0\right)$, *and* $W(\cdot)$ *is a two-sided Brownian motion.*[12]

(iii) $\hat{\theta}_{1\tau}$ *and* $\hat{\gamma}_\tau$ *are asymptotically independent.*

Based on the above theorem, we can conduct asymptotic tests for both the coefficient and threshold parameters. Because $\hat{\theta}_{1\tau}$ is asymptotically normally distributed, the statistical inferences for $\theta_{1\tau}$ are standard. It is worth mentioning that the result in Theorem 3.5(i) continues to hold even if one allows

[12] A two-sided Brownian motion on the real line is defined as $W(r) = W_1(-r)\,1\{r \leq 0\} + W_2(r)\,1\{r > 0\}$, where $W_1(r)$ and $W_2(r)$ are two independent standard Brownian motions on $[0,\infty)$.

$a = 0$ in Assumption A7. However, the asymptotic distribution of $\hat{\gamma}_\tau$ in Theorem 3.5(ii) remains valid only for the case of $a \in \left(0, \frac{1}{2}\right)$ in A7, analogously to the case of the least squares threshold regression in Hansen (2000). In the case of $a = 0$, if we assume the independence of $\varepsilon_{i\tau}$ and z_i, then we can apply the result of Koul et al. (2003) and demonstrate that $n\left(\hat{\gamma}_\tau - \gamma_\tau^0\right)$ converges in distribution to the argmin of a two-sided compound Poisson process. However, such an independence assumption seems too strong, and thus we focus only on the case of $a \in \left(0, \frac{1}{2}\right)$. In the following subsection, we study a likelihood ratio test for γ_τ.

3.4 A likelihood ratio test for threshold parameter γ_τ

To test the hypotheses of threshold parameter γ_τ, we follow Hansen (2000) and Caner (2002) and consider a likelihood ratio (LR) statistic. The null hypothesis is

$$H_{0\tau} : \gamma_\tau = \gamma_\tau^0, \tag{3.17}$$

and the LR statistic is

$$LR_{n\tau}\left(\gamma_\tau\right) = S_{n\tau}\left(\hat{\theta}_1\left(\tau, \gamma_\tau\right), \gamma_\tau\right) - S_{n\tau}\left(\hat{\theta}_{1\tau}, \hat{\gamma}_\tau\right).$$

We reject $H_{0\tau}$ for large values of $LR_{n\tau}\left(\gamma_\tau^0\right)$.

The following theorem establishes the asymptotic distribution of $LR_{n\tau}\left(\gamma_\tau^0\right)$ under $H_{0\tau}$ defined in (3.17).

Theorem 3.6 *Suppose that Assumptions A1-A4 and A6-A8 hold. Then, under $H_{0\tau}$,*

$$LR_{n\tau}\left(\gamma_\tau^0\right) \xrightarrow{d} \frac{\lambda_\tau}{4\mu_\tau}\Xi,$$

where $\Xi \equiv \sup_{r\in(-\infty,\infty)}\left\{2W\left(r\right) - |r|\right\}$.

Alternatively, we can consider the following test statistic.

$$LR_{n\tau}^a\left(\gamma_\tau^0\right) \equiv S_{n\tau}\left(\hat{\theta}_{1\tau}, \gamma_\tau^0\right) - S_{n\tau}\left(\hat{\theta}_{1\tau}, \hat{\gamma}_\tau\right).$$

Following the proof of Theorem 3.6, we can readily show that $LR_{n\tau}^a\left(\gamma_\tau^0\right) \xrightarrow{d} \frac{\lambda_\tau}{4\mu_\tau}\Xi$. That is, $LR_{n\tau}\left(\gamma_\tau^0\right)$ and $LR_{n\tau}^a\left(\gamma_\tau^0\right)$ share the same asymptotic distribution under $H_{0\tau}$.

Theorem 3.6 indicates that we can consider the following normalized LR test statistic.

$$NLR_{n\tau}\left(\gamma_\tau^0\right) = \frac{4\tilde{\delta}_\tau'\tilde{D}_\tau\tilde{\delta}_\tau}{\tilde{\delta}_\tau'\tilde{N}_\tau\tilde{\delta}_\tau}LR_{n\tau}\left(\gamma_\tau^0\right),$$

where $\tilde{\delta}_\tau \equiv \hat{\alpha}_\tau\left(\gamma_\tau^0\right) - \hat{\beta}_\tau\left(\gamma_\tau^0\right)$; $\tilde{N}_\tau$ is a consistent local linear (or constant) estimate of $N_\tau\left(\gamma_\tau^0\right) \equiv E\left[x_i x_i' | r_i = \gamma_\tau^0\right]$ by using the bandwidth h_1 and the kernel K;

$$\tilde{D}_\tau = \hat{E}_{h_1}\left[\hat{f}_{h_1}\left(\hat{\alpha}_\tau\left(\gamma_\tau^0\right)' x_i \; | z_i\right) x_i x_i' | r_i = \gamma_\tau^0\right];$$

$\hat{f}_{h_1}(\cdot|z_i)$ is a kernel estimate for the density of y_i given z_i by using the bandwidth h_1 and the kernel K; and $\hat{E}_{h_1}\left(\cdot|r_i = \gamma_\tau^0\right)$ is a kernel estimate of $E\left[f\left(\alpha_\tau^{0\prime} x_i|z_i\right) x_i x_i' | r_i = \gamma_\tau^0\right]$ by using the bandwidth h_1, the kernel K, and the observations on $\hat{f}_{h_1}\left(\hat{\alpha}_\tau\left(\gamma_\tau^0\right)' x_i\,|z_i\right) x_i x_i'$ and r_i. Under the assumption that z_i is compactly supported with bounded density that is bounded away from 0 on its support, we realize that one can obtain both estimates by the local linear method to avoid boundary bias and the asymptotic trimming issue. See Fan et al. (1996) for the local linear estimation of conditional density.

Under standard conditions, we can show that $\frac{\hat{\delta}_\tau' \hat{D}_\tau \hat{\delta}_\tau}{\hat{\delta}_\tau' \hat{N}_\tau \hat{\delta}_\tau} \to \frac{\mu_\tau}{\lambda_\tau}$ in probability. Then, by the Slutsky lemma, we have

$$NLR_{n\tau}\left(\gamma_\tau^0\right) \xrightarrow{d} \Xi.$$

That is, $NLR_{n\tau}\left(\gamma_\tau^0\right)$ is asymptotically pivotal. It is well known that $\sup_{r\leq 0}[2W(r) - |r|]$ and $\sup_{r\geq 0}[2W(r) - |r|]$ are independent exponential random variables with distribution function $1 - e^{-z}$ such that the CDF of Ξ is given by $P(\Xi \leq z) = \left(1 - e^{-z/2}\right)^2$. We can easily tabulate the asymptotic critical values for the normalized statistic $NLR_{n\tau}\left(\gamma_\tau^0\right)$. See Hansen (2000, p. 582) for more details.

3.5 Inference in the case of a common break

The above analysis can be extended to the case in which different conditional quantile functions share a common threshold value. In this case, joint analysis of multiple quantile regressions may improve the accuracy of the common threshold estimate.

Recall that $\mathcal{T} \equiv [\underline{\tau}, \overline{\tau}] \subset (0,1)$. If, for all $\tau \in \mathcal{T}$, the conditional quantile regressions share a common threshold γ, then we can estimate γ by

$$\hat{\gamma} \equiv \hat{\gamma}(\Pi) \equiv \underset{\gamma\in\Gamma}{\operatorname{argmin}}\ \hat{S}_{n,\Pi}(\gamma),$$

where $\hat{S}_{n,\Pi}(\gamma) = \int S_{n\tau}\left(\hat{\theta}_1(\tau,\gamma), \gamma\right) d\Pi(\tau)$, and Π is a user-specified probability distribution function defined on $\mathcal{T}$.

After we obtain the estimate $\hat{\gamma}$ of γ, we can estimate α_τ^0 and β_τ^0 by $\hat{\alpha}_\tau = \hat{\alpha}(\tau,\hat{\gamma})$ and $\hat{\beta}_\tau = \hat{\beta}(\tau,\hat{\gamma})$, respectively. As before, let $\hat{\theta}_1(\tau,\gamma) \equiv (\hat{\alpha}(\tau,\gamma)', \hat{\beta}(\tau,\gamma)')'$ and $\hat{\theta}_{1\tau} \equiv (\hat{\alpha}_\tau', \hat{\beta}_\tau')'$. To study the asymptotic properties of $\hat{\gamma}$, $\hat{\theta}_{1\tau}$, and $\hat{\theta}_1(\tau,\gamma)$, we add the following assumption.

Assumption A9. (i) Let $\theta_1^0(\tau,\gamma) \equiv \operatorname{argmin}_{\theta_1} S(\theta_1;\tau,\gamma)$, where $S(\theta_1;\tau,\gamma) \equiv E[\rho_\tau(y_i - \theta_1' z_i(\gamma))]$. There exists a $\gamma^0 \in \Gamma$ such that $\gamma^0 = \operatorname{argmin}_\gamma S\left(\theta_1^0(\tau,\gamma);\tau,\gamma\right)$ for all τ.

(ii) Let $\theta_{1\tau}^0 \equiv \theta_1^0\left(\tau,\gamma^0\right)$ and $\Delta(z_i,\tau,\gamma) \equiv \theta_1^0(\tau,\gamma)' z_i(\gamma) - \theta_{1\tau}^{0\prime} z_i\left(\gamma^0\right)$, and $\int P(|\Delta(z_i,\tau,\gamma)| > 0)\, d\Pi(\tau) > 0$ for all $\gamma \neq \gamma^0$.

(iii) Let $\bar{\Omega}_1(\tau,\gamma) \equiv E[f\left(\theta_1^0(\tau,\gamma) z_i(\gamma)|z_i\right) z_i(\gamma) z_i(\gamma)']$, and $\bar{\Omega}_1(\tau,\gamma)$ is positive definite for all $(\tau,\gamma) \in \mathcal{T}\times\Gamma$.

Observe that $E[\rho_\tau(y_i - \theta_1' z_i(\gamma))]$ is convex in θ_1 for all $(\tau,\gamma)\in\mathcal{T}\times\Gamma$, and $\theta_1^0(\tau,\gamma)$ in A9(i) exists and is uniquely defined. It is also continuous in (τ,γ) by an application of the maximum theorem. The first-order condition for the minimization of $S(\theta_1;\tau,\gamma)$ with respect to θ_1 implies that

$$E\left[\psi_\tau\left(y_i - \theta_1^0(\tau,\gamma)' z_i(\gamma)\right) z_i(\gamma)\right] = 0 \text{ for all } (\tau,\gamma)\in\mathcal{T}\times\Gamma, \tag{3.18}$$

which will be used in the proof of Theorem 3.7 below. The last part of A9(i) simply restricts the conditional quantile regression from sharing a common break γ^0, which does not depend on $\tau \in \mathcal{T}$. Like Assumption A6, A9(ii) is an identification condition and requires that γ^0 is the unique common break. A9(iii) extends A5. Note that $\bar{\Omega}_1\left(\tau,\gamma^0\right) = \Omega_1\left(\tau,\gamma^0\right)$ under A9(i)-(ii).

The following theorem summarizes the important properties of $\hat{\gamma}$, $\hat{\theta}_{1\tau}$, and $\hat{\theta}_1\left(\tau,\gamma\right)$.

Theorem 3.7 *Suppose that Assumptions A1-A4 and A6-A9 hold. Then,*

(i) $\hat{\gamma} = \gamma^0 + o_P(1)$ *and* $\hat{\theta}_{1\tau} = \theta^0_{1\tau} + o_P(1)$ *for each* $\tau \in \mathcal{T}$;

(ii) $\sqrt{n}\left(\hat{\theta}_1\left(\tau,\gamma\right) - \theta^0_1\left(\tau,\gamma\right)\right) = \bar{\Omega}_1\left(\tau,\gamma\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\tau\left(y_i - \theta^0_1\left(\tau,\gamma\right)' z_i\left(\gamma\right)\right) z_i\left(\gamma\right) + o_P(1)$, *where* $o_P(1)$ *holds uniformly over* $\mathcal{T}\times\Gamma$;

(iii) $n^{1/2}\left(\hat{\theta}_{1\tau} - \theta^0_{1\tau}\right) \xrightarrow{d} N\left(\mathbf{0}_{2k\times 1}, \tau\left(1-\tau\right)\Sigma\left(\tau,\gamma^0\right)\right)$;

(iv) $n^{1-2a}\left(\hat{\gamma} - \gamma^0\right) \xrightarrow{d} \frac{\lambda_0}{4\mu_0^2} \underset{r\in(-\infty,\infty)}{\operatorname{argmax}} \left\{W(r) - \frac{1}{2}|r|\right\}$, *where* $\lambda_0 \equiv \{\int\sqrt{v_\tau' N_\tau\left(\gamma^0\right)v_\tau}d\Pi\left(\tau\right)\}^2 g\left(\gamma^0\right)$ *and* $\mu_0 \equiv \int v_\tau' D_\tau\left(\gamma^0\right) v_\tau d\Pi\left(\tau\right) g\left(\gamma^0\right)$; *and*

(v) $\hat{\theta}_{1\tau}$ *and* $\hat{\gamma}$ *are asymptotically independent.*

Theorem 3.7(i) implies the consistency of the parameter estimates. Theorem 3.7(ii) extends the Bahadur uniform representation result in Theorem 3.1 to allow a single common break in the quantile processes. Theorem 3.7(iii) indicates that the first-order asymptotic distribution of $\hat{\theta}_{1\tau}$ is the same as that obtained before. This is largely due to the asymptotic orthogonality of the coefficient parameter estimate and the threshold parameter estimate. The last two parts of Theorem 3.7 are parallel to those in Theorem 3.5.

As in the previous subsection, we can consider testing the following composite hypothesis of common threshold parameter γ.

$$H_0 : \gamma_\tau = \gamma^0 \text{ for all } \tau \in \mathcal{T}. \tag{3.19}$$

We now investigate two LR-type statistics. The first is

$$LR_{n1}\left(\gamma\right) = \hat{S}_{n,\Pi}\left(\gamma\right) - \hat{S}_{n,\Pi}\left(\hat{\gamma}\right) = \int\left[S_{n\tau}\left(\hat{\theta}_1\left(\tau,\gamma\right),\gamma\right) - S_{n\tau}\left(\hat{\theta}_{1\tau},\hat{\gamma}\right)\right]d\Pi\left(\tau\right),$$

which is useful if we strongly believe that the quantile regressions for different τ's share a common threshold point but are not sure about the exact value of that threshold. The second statistic is

$$LR_{n2}\left(\gamma\right) = \int\left[S_{n\tau}\left(\hat{\theta}_1\left(\tau,\gamma\right),\gamma\right) - S_{n\tau}\left(\hat{\theta}_1\left(\tau,\hat{\gamma}_\tau\right),\hat{\gamma}_\tau\right)\right]d\Pi\left(\tau\right).$$

We reject H_0 for large values of $LR_{nj}\left(\gamma^0\right)$, $j = 1, 2$.

The following theorem establishes the asymptotic distributions of $LR_{n1}\left(\gamma^0\right)$ and $LR_{n2}\left(\gamma^0\right)$ under H_0 in (3.19).

Theorem 3.8 *Suppose that Assumptions A1-A4 and A6-A9 hold. Then, under* H_0, *we have*

(i) $LR_{n1}\left(\gamma^0\right) \xrightarrow{d} \frac{\lambda_0}{4\mu_0}\Xi$, *and*

(ii) $LR_{n2}\left(\gamma^0\right) \xrightarrow{d} \int\frac{\lambda_\tau}{4\mu_\tau}d\Pi\left(\tau\right)\Xi$,

where Ξ *is as defined in Theorem 3.6.*

Clearly, the asymptotic distribution of $LR_{n1}\left(\gamma^0\right)$ is analogous to that of $LR_{n\tau}\left(\gamma^0\right)$ in Theorem 3.6 under $H_{0\tau}$ for a specific quantile τ, and that of $LR_{n2}\left(\gamma^0\right)$ is a weighted version of that of $LR_{n\tau}\left(\gamma^0\right)$.

4 Empirical Application: Taobao.com

In this paper, we investigate the reputation and pricing patterns on Taobao.com, the dominant online trading platform in China.

4.1 Data

We collect trading data on the iPod Nano IV 8G on Taobao.com from September to December 2009. Two factors motivate us to choose the iPod Nano for this study. First, developed by Apple, the iPod Nano has become a popular choice among young consumers in China. This group of consumers is more familiar, and therefore more comfortable, with the trading rules and logistics of online transactions. Consequently, this group is more likely to become the target of online promotions. Second, the iPod Nano is designed to differentiate itself substantially from the other digital media players available on the market. Therefore, to a large extent, we sustain our analysis on a homogeneous product.

In our sample, we observe posting prices (price), the reputation score and category of the seller at the time of posting (reputation score), whether postage is included in the posted prices (postage), the total number of items sold by the seller (total items), the sales volume last recorded per posting item (sales volume), the rate of good reviews obtained by the seller (rate of good reviews), and the seller's location (area code).[13]

The reputation scoring system on Taobao.com works as follows. Once a transaction is completed, a buyer who is a member of Taobao.com is qualified to review the seller's service according to her experience of the transaction. In addition to any written comments, the review has to conclude with a rating of "good," "neutral," or "bad." In accordance with the buyer's review, the seller accrues one point for a "good" review, loses one point for a "bad" one, and gets nothing for a "neutral" review. Taobao.com also categorizes sellers' reputation status based on their reputation scores. Table 1 lists the 20 categories. For example, a seller with a reputation score between 4 and 10 falls into the "1-heart"category. The categories progress with numbers 1 to 5 and from heart to diamond, crown, and gold crown.

Figure 4 plots the histogram of each seller's reputation category based on the data. Most of the sellers are spread between Categories 1 and 9 (1-heart to 4-diamonds). Very few have accumulated more than 50,000 good reviews (Category 12). From the data, we spot a possible exogenous cutoff that might provide incentives for the sellers on Taobao.com to price for goodwill: that is, 5000, the point at which a seller moves from Category 9 (4-diamonds) to Category 10 (5-diamonds). Sellers near the cutoff of 5000 are strongly motivated to move up to 5-diamonds, where they can enjoy a higher

[13]We also observe the actual transaction prices. However, these prices have a great deal of noise, due to the options of an additional set menu at each seller's store. We therefore decide to focus on the posting price in our empirical analysis.

Table 1: Reputation Scoring System on Taobao.com

Category	Points needed	Category icon	Category	Points needed	Category icon
1	4-10	1-heart	11	10001-20000	1-crown
2	11-40	2-hearts	12	20001-50000	2-crowns
3	41-90	3-hearts	13	50001-100000	3-crowns
4	91-150	4-hearts	14	100001-200000	4-crowns
5	151-250	5-hearts	15	200001-500000	5-crowns
6	251-500	1-diamond	16	500001-1000000	1-gold crown
7	501-1000	2-diamonds	17	1000001-2000000	2-gold crowns
8	1001-2000	3-diamonds	18	2000001-5000000	3-gold crowns
9	2001-5000	4-diamonds	19	5000001-10000000	4-gold crowns
10	5001-10000	5-diamonds	20	10000000+	5-gold crowns

Figure 4: Histogram of Reputation Category

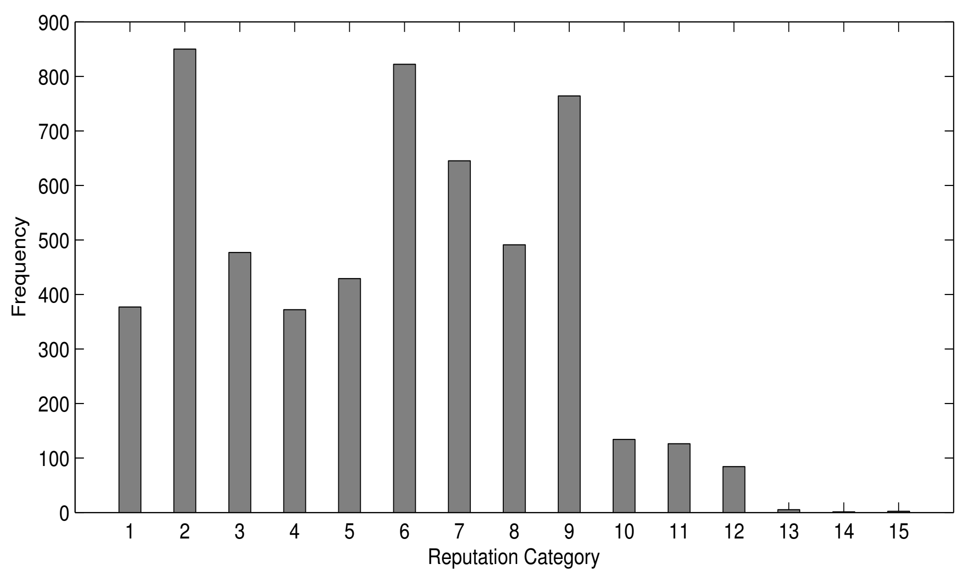


Table 2: Summary Statistics

Variables	Min	Max	Mean	Median	Std dev
price	508	1500	1010.34	1050	166.01
reputation score	251	4995	1412.34	824	1243.00
total items	0	36811	537.62	236	1518.73
postage	0	120	11.61	10	11.35
sales volume	0	300	3.27	0	12.15
rate of good reviews	0	100	89.16	99.83	30.65

NOTE: The sample includes only sellers that belong to Categories 6 to 9. The total number of observations is 2727.

pay-off from their better reputation, with a tremendous reduction in competition. We therefore define "goodwill" in this application by a seller's reputation classified as Category 10 or above.

In the application, we face the potential issue of sellers' maturity. For example, a new seller may have a greater chance of being "badly" behaved, as the reputation concern is of less significance to him. Taking this possibility into account, we regard sellers with a reputation score of less than 250 (Category 5) as rookies and exclude them from our data analysis.[14] Therefore, the sample for this study includes only sellers with reputation scores between 250 and 5000, and the total number of observations is n =2727.

Table 2 lists the basic summary statistics of the sample variables. We observe a substantial amount of variation in prices, which touches on the core of our study, that is, whether reputation contributes to providing a causal term for such rich variations in price. We observe only limited information on sellers in the dataset, among which "total items" is the most important. Recall that it represents the total number of items for sale in a particular online store (the seller). We view this variable as a proxy for a seller's scale and specialization. The significant variation observed in total items may reflect the fact that seller heterogeneity is at work. The sales volume variable exhibits much less variation. Lastly, the variation in the rate of good reviews indicates that it is less likely for sellers to get a bad or neutral review than a good one.[15] This is consistent with other empirical findings that only reviewers who provide good reviews tend to break the silence. See, for example, Dellarocas and Wood (2008). This pattern partially validates our theoretical model, in which a distribution that does not elicit good reviews plays a central role in equilibrium pricing. Our focus on the left tail of the distribution becomes more relevant.

4.2 Empirical analysis

First, we conduct testing for the existence of a change point in the data following the approach suggested in Section 3.2. Table 3 reports the results. We first consider the test of the null hypothesis of no change points for all quantiles between 0.1 and 0.9, i.e., $\mathcal{T} = [0.1, 0.9]$ in (3.5). The p-value for such a test is 0.000 to ensure that we reject the null of no breaks at all conventional significance levels

[14] We also experiment with defining rookies as sellers with reputation scores below Category 4 or 6, but our major empirical findings remain valid.

[15] The rate of good reviews is defined as the percentage of good reviews out of the total number of reviews.

Table 3: Test for the Existence of Breaks

$\mathcal{T}$	Test stat	p-value	Critical value		
			10%	5%	1%
0.1 - 0.9	807.01	0.00	27.63	29.54	33.76
0.1	190.86	0.00	21.02	22.93	27.54
0.2	278.62	0.00	20.18	21.89	25.79
0.3	361.25	0.00	19.90	21.55	24.93
0.4	387.43	0.00	19.86	21.44	24.88
0.5	807.01	0.00	19.71	21.22	24.87
0.6	103.69	0.00	19.68	21.16	24.62
0.7	56.19	0.00	19.88	21.39	24.81
0.8	15.48	0.55	20.20	22.01	25.87
0.9	14.11	0.80	21.05	22.85	27.79
0.71	39.18	0.00	19.90	21.40	25.05
0.72	36.08	0.00	19.91	21.50	25.04
0.73	25.06	0.01	19.96	21.55	25.22
0.74	25.20	0.01	20.00	21.53	24.86
0.75	24.71	0.01	20.07	21.68	25.10
0.76	22.54	0.03	20.06	21.68	25.12
0.77	20.43	0.08	20.06	21.72	25.28
0.78	14.07	0.76	20.07	21.69	25.22
0.79	12.97	0.90	20.12	21.78	25.59

(1%, 5%, and 10%). Then, we test for several typical quantiles, i.e., $\mathcal{T} = \{\tau\}$ in (3.5) for $\tau = 0.1, 0.2, ..., 0.9$. Using the 5% nominal levels, we find that change points exist for quantiles up to 0.7 and that the breaks do not occur for high quantiles (0.8 and 0.9). To take a closer look, we repeat the testing procedure for quantiles between 0.7 and 0.8, from which we find that the structural breaks occur only when $\tau \leq 0.76$ at the 5% nominal level.

Second, we estimate the model for the quantiles identified to have breaks. Table 4 reports the parameter estimates for the typical quantiles, that is, $\tau = 0.1, 0.2, ..., 0.7$. Figure 5 shows the plots of the quantile regression line before and after the breaks for a number of representative quantiles. Our estimates show that jumps occur among sellers at all quantiles under investigation. The size of these jumps can be as significant as -370.79, which is about 37% of the mean price in the sample. The estimated slope parameters before the jumps are small, and some are not statistically significant at the 5% level. Nevertheless, we tend to have more significant slope estimates after the change point, and the slope parameters after the jumps tend to be much larger in magnitude than those before the jumps. These estimates justify our quantile regression approach. The heterogeneous pricing behavior may reflect differences across sellers and market demand situations in online markets.

In making inferences for the estimates of γ_0, one may notice that the asymptotic sampling distributions depend on unknown parameters. This may in turn imply poor finite sample behavior if we follow Theorem 3.5. Instead, we resort to the normalized LR test proposed in Theorem 3.6. Table 5 reports the 95% confidence intervals for the jump location estimates ($\hat{\gamma}$). We find the upper bound

Figure 5: Estimates of Quantile Regression

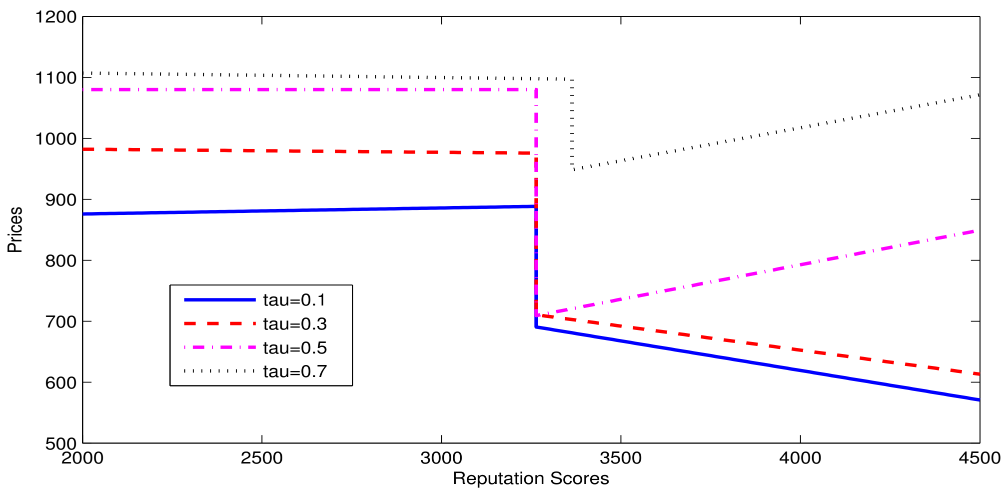


Table 4: Estimation Results

τ	Jump Size	γ	α_τ		β_τ	
			intercept	slope	intercept	slope
0.1	-197.88	3264	856.10	**0.010**	1006.98	**-0.097**
			(6.21)	(0.003)	(34.15)	(0.029)
0.2	-224.28	3264	947.22	**-0.007**	969.33	**-0.083**
			(5.73)	(0.003)	(30.59)	(0.025)
0.3	-265.17	3264	992.30	-0.005	967.87	**-0.079**
			(6.04)	(0.004)	(43.50)	(0.044)
0.4	-300.63	3264	1037.13	**-0.008**	735.55	-0.008
			(6.4)	(0.004)	(42.03)	(0.044)
0.5	-370.79	3264	1080.00	0.000	339.14	**0.113**
			(6.53)	(0.004)	(42.61)	(0.043)
0.6	-224.47	3240	1097.85	0.000	634.86	0.074
			(6.02)	(0.004)	(44.00)	(0.047)
0.7	-148.52	3364	1121.87	**-0.007**	584.56	**0.108**
			(6.31)	(0.004)	(52.40)	(0.050)

NOTE: Numbers in parentheses are standard errors, and numbers in bold indicate that the corresponding slope coefficients are statistically significant at the 10% level. All intercepts are statistically significant at the 1% level.

Table 5: Inference on the Estimates of Breaks

τ	γ	95% lower bound	95% upper bound
0.1	3264	3259	3271
0.2	3264	3259	3271
0.3	3264	3246	3271
0.4	3264	3246	3271
0.5	3264	3259	3271
0.6	3240	3232	3248
0.7	3364	3355	3367
0.1 - 0.5	3264	3259	3271
0.1 - 0.6	3252	3252	3252
mean regression	3240	3134	3272

for the 95% confidence intervals by $\inf\{\gamma : \gamma > \hat{\gamma} \ \& \ NLR(\gamma) \leq c_{0.95}\}$, where $c_{0.95}$ denotes the 0.95-level critical value for Ξ. Accordingly, the lower bound for the 95% confidence intervals is defined by $\sup\{\gamma : \gamma < \hat{\gamma} \ \& \ NLR(\gamma) \leq c_{0.95}\}$. Clearly, for $\tau = 0.1 - 0.5$, even though the estimates of the change points are the same, the 95% confidence intervals may be different.

Based on the estimates in Table 4, we strong believe that the quantiles between 0.1 and 0.5 may have a common structural break. Hence, we implement the testing methods discussed in Section 3.5. These results are also reported in Table 5. The jump is estimated to occur at 3264. We then follow the testing approach for LR_1 in Theorem 3.8 to build up a 95% confidence interval for this estimate, which coincides with the 95% confidence interval for $\tau = 0.1$, 0.2 and 0.3. Finally, we implement the LR_2 test in Theorem 3.8 for the quantiles between 0.1 and 0.6 because now we are not sure whether there exists a common structural break and if so what value it takes. The estimate of γ_0 is 3252. However, the null hypothesis of a common break in this case is rejected at this estimate.

Two remarks are in order for a thorough empirical analysis. First, given the fact that we found a common structural break among a few typical quantiles, one may wonder whether a least square threshold regression can also identify the break. To address this concern, we estimate Hansen's (2000) least squares threshold model. The results are reported in the last row of Table 5. It suggests that the estimated threshold in least squares mean regression differs from the most of the break estimates in the quantile regression framework. Moreover, the confidence interval in the least square estimation is much wider than those obtained in quantile regressions. The differences shed some lights on the necessity of using quantile regression models for the consideration of heterogeneities.

Our second concern arises in line of the manipulation problem raised by McCrary (2008). McCrary argued that some varieties of manipulation (e.g., complete manipulation) on the running variable in RDD may lead to identification problems while others may not. He developed a test of manipulation related to continuity of the running variable density function when the potential discontinuity point is known. Here we follow McCrary (2008) closely to test the discontinuity of the density function of the running variable (r) at the estimated cutoff point 3264 (for $0.1 \leq \tau \leq 0.5$). The estimated log difference of the left and right density limits at this point is 2.5968 with a standard error of 0.3932, which suggests a large t-ratio that rejects the null hypothesis of continuity at any conventional significance levels. Even so, because the sellers do not have any complete control on the reputation score and the latter also has idiosyncratic element which is determined by the buyers, the discontinuity at the density of reputation score does not lead to identification problems for the optimal pricing strategy. (c.f. Footnote 4 in McCrary, 2008.) On the contrary, we believe it offers partial support for our empirical analysis.

4.3 Robustness check

Although the goodwill pricing pattern is found in the data, there remain critical issues in the foregoing empirical analysis. First, we did not consider the possible impact of the observed covariates on the posting prices. As a robustness check, we repeat the previous empirical exercises by including all observed variables.[16]

Following the testing approach suggested in Section 3.2, we detect the existence of change points in the data for all of the quantiles between 0.1 and 0.9. We then estimate the model for quantiles $\tau = 0.1, ..., 0.9$. The estimation and inference results are reported in Table 6.

The jump sizes are evaluated at the mean values of each covariate in the quantile regressions. It is observed that the price-cut pattern occurs for all $\tau \in [0.2, 0.9]$, and the jump sizes are much smaller than those unconditional on the covariates. Moreover, roughly speaking, the larger the τ, the higher the reputation level at which the jump occurs (that is the closer to the exogenous cutoff of goodwill). Although jumps are identified in the quantiles of $\tau = 0.8, 0.9$, they are smaller in magnitude. What can be concluded is that for the sellers in most of the quantiles, a price-cut strategy may be useful when their reputation scores are in the range of 3200 to 3400.

A jump-up occurs at the quantile regression of $\tau = 0.1$. This inconsistent finding may indicate that sellers posting extremely low prices may possibly have objectives other than an enhanced reputation. If this is indeed the case, then our model cannot, in general, explain the pricing behavior of these sellers.

Our choice of the iPod Nano for this study stemmed from our concern with product homogeneity. An additional concern is whether a seller would choose this product to accrue good reviews to obtain the goodwill benefit. To address this issue, we repeat the testing and estimation procedure with a much more restrictive sample, that is, the sellers with fewer than 100 items in total to sell. These sellers are smaller in scale and possibly more specialized in selling electronic items. Our major findings on the pricing patterns remain valid with this restrictive sample. However, we also acknowledge that this issue may be significantly more complicated. In particular, consumers' willingness to provide positive reviews in exchange for lower prices may be dependent on product-specific characteristics. The issue of consumer responsiveness to this type of product is beyond the scope of this paper and is therefore left for future research.

[16] We also considered including r^2 in the QRTM but found its coefficient not to be significantly different from 0 at the 10% level for all quantiles under investigation. We thus decided to augment our QRTM in the previous subsection only by the covariates listed in Table 2.

Table 6: Estimation and Inference of Structural Breaks

τ	Jump Size	γ	95% lower bound	95% upper bound
0.1	277.68	1984	1975	2018
0.2	-77.83	3264	3231	3271
0.3	-125.60	2979	2975	3002
0.4	-101.01	3252	3232	3272
0.5	-52.38	3252	3247	3261
0.6	-87.07	3252	3232	3272
0.7	-64.23	3252	3247	3261
0.8	-28.00	3364	3355	3367
0.9	-27.64	3849	3750	3947

5 Conclusion

We investigate the pricing strategy that an online seller may employ when the concern for goodwill exerts an effect on the pricing decision. In theory, such pricing entails a price reduction at a reputation level threshold. A mechanical reason may be that sellers switch between the local extremes of the profit function, which, accordingly, are determined by the shape features on the underlying distributions. In connection with our model prediction, we herein propose a threshold quantile regression-based testing and estimation framework to identify and make inferences on this particular pricing strategy. We then apply the proposed methodology to Taobao.com and find both heterogeneities and jumps in sellers' pricing strategies for goodwill.

There are a few dimensions along which this work could be extended. First, we provide a simple but interesting theoretical model to explain the estimated pricing pattern in the data. Determining both whether other mechanisms can predict the same pricing strategy and which alternative caters best for the application is important to understanding the reputation effect on Internet markets. Doing so is beyond the scope of this paper, but is certainly worth exploring.

Second, in terms of heterogeneity, we consider and document only the across-quantile difference in posting prices. However, determining what drives these differences among sellers would certainly be an interesting issue to pursue. Unfortunately, due to the limited information on the seller side in the current dataset, we are restrained from looking into this issue in this paper.

Third, we adopt a parametric threshold quantile regression framework to investigate the functional relationship between prices and reputation levels in this paper. Even though we examine the validity of the pricing pattern with the specification of polynomial forms, a more flexible nonparametric approach (e.g., Oka, 2010) seems an appealing direction for future research. Product differentiation may be yet another interesting extension. We leave these for future research.

6 Appendix: Proof of the main results

To prove the main results in Section 3, we first define some notation. Recall $z_i \equiv (x_i', r_i)'$ if $r_i \notin x_i$ and $z_i \equiv x_i$ otherwise, $z_i(\gamma_\tau) \equiv [x_i' 1\{r_i \leq \gamma_\tau\} \;\; x_i' 1\{r_i > \gamma_\tau\}]'$, $\psi_\tau(u) \equiv \tau - 1(u < 0)$, $\theta_\tau \equiv (\theta_{1\tau}', \gamma_\tau)'$, $\theta_{1\tau} \equiv$

$(\alpha_\tau', \beta_\tau')'$, and the true value of $\theta_\tau, \theta_{1\tau}, \alpha_\tau, \beta_\tau$, and γ_τ are denoted as $\theta_\tau^0, \theta_{1\tau}^0, \alpha_\tau^0, \beta_\tau^0$, and γ_τ^0, respectively. Let

$$\varepsilon_{i\tau}(\theta_\tau) \equiv y_i - \theta_{1\tau}' z_i(\gamma_\tau) \text{ and } \varepsilon_{i\tau} = \varepsilon_{i\tau}\left(\theta_\tau^0\right). \tag{6.1}$$

Note that the τth conditional quantile of $\varepsilon_{i\tau} = \varepsilon_{i\tau}\left(\theta_\tau^0\right)$ given z_i is 0, i.e., $E\left[\psi_\tau(\varepsilon_{i\tau})|z_i\right] = 0$. Let

$$S_{n\tau}(\theta_\tau) = \sum_{i=1}^n \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right). \tag{6.2}$$

Let $\|\cdot\|$ denote the Euclidean norm. Note that for all $\theta_\tau \in \mathbb{R}^{2k+1}$, we have $\|z_i(\gamma_\tau)\| = \|x_i\|$, $|\theta_{1\tau}' z_i(\gamma_\tau)| \le \|\theta_{1\tau}\| \|x_i\|$, and $|\theta_{1\tau}' z_i(\gamma_\tau) - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau)| \le \|\theta_{1\tau} - \theta_{1\tau}^*\| \|x_i\|$, and

$$\|z_i(\gamma_\tau) - z_i(\gamma_\tau^*)\| \le \sqrt{2}\|x_i\| 1\{\gamma_\tau \wedge \gamma_\tau^* < r_i \le \gamma_\tau \vee \gamma_\tau^*\} \le \sqrt{2}\|x_i\| 1\{|r_i - \gamma_\tau| \le |\gamma_\tau^* - \gamma_\tau|\}. \tag{6.3}$$

We use C to denote a generic large positive constant whose exact value may change across lines.

We now state and prove some lemmata that are used in later proofs.

Lemma 6.1 $\lim_{\eta\to 0} E \sup_{\theta_\tau^* \in N_\eta(\theta_\tau)} \left|\rho_\tau\left(y_i - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right)\right| = 0$ *for any* $\theta_\tau \in \Theta$, *where* $N_\eta(\theta_\tau) \equiv \{\theta_\tau^* = (\theta_{1\tau}^{*\prime}, \gamma_\tau^*)' \in \Theta : \|\theta_{1\tau}^* - \theta_{1\tau}\| < \eta,\ |\gamma_\tau^* - \gamma_\tau| < \eta\}$ *denotes an* η*-neighborhood of* $\theta_\tau \in \Theta$ *and* $\eta > 0$.

Proof. Let $\Delta_{i\tau} \equiv \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*) - \theta_{1\tau}' z_i(\gamma_\tau)$. Then by the triangle inequality and (6.3),

$$\begin{aligned}
|\Delta_{i\tau}| &\le |\theta_{1\tau}'[z_i(\gamma_\tau) - z_i(\gamma_\tau^*)]| + \left|(\theta_{1\tau} - \theta_{1\tau}^*)' z_i(\gamma_\tau^*)\right| \\
&\le \left\{\sqrt{2} \times\ 1\{\gamma_\tau \wedge \gamma_\tau^* < r_i \le \gamma_\tau \vee \gamma_\tau^*\}\|\theta_{1\tau}\| + \|\theta_{1\tau} - \theta_{1\tau}^*\|\right\}\|x_i\| \\
&\le \left\{\sqrt{2} \times\ 1\{|r_i - \gamma_\tau| \le |\gamma_\tau - \gamma_\tau^*|\}\|\theta_{1\tau}\| + \|\theta_{1\tau} - \theta_{1\tau}^*\|\right\}\|x_i\| \\
&\le \left\{\sqrt{2} \times\ 1\{|r_i - \gamma_\tau| \le \eta\}\|\theta_{1\tau}\| + \eta\right\}\|x_i\|.
\end{aligned}$$

By Knight's (1998) identity (see also Koenker (2005, p. 121)),

$$\begin{aligned}
\rho_\tau\left(y_i - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right) &= \rho_\tau\left(\varepsilon_{i\tau}(\theta_\tau) - \Delta_{i\tau}\right) - \rho_\tau\left(\varepsilon_{i\tau}(\theta_\tau)\right) \\
&= -\Delta_{i\tau}\psi_\tau\left(\varepsilon_{i\tau}(\theta_\tau)\right) + \int_0^{\Delta_{i\tau}} \left[1\{u_{i\tau} \le s\} - 1\{u_{i\tau} \le 0\}\right] ds.
\end{aligned}$$

It follows that

$$\begin{aligned}
&E \sup_{\theta_\tau^* \in N_\eta(\theta_\tau)} \left|\rho_\tau\left(y_i - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right)\right| \\
&\le 2E|\Delta_{i\tau}| \le 2\sqrt{2}P\left(|r_i - \gamma_\tau| \le \eta\right)^{1/2}\|x_i\|_2\|\theta_{1\tau}\| + \eta\|x_i\| + 2\eta E\|x_i\| \to 0
\end{aligned}$$

as $\eta \to 0$, where $\|x_i\|_2 \equiv \{E\|x_i\|^2\}^{1/2}$. ■

Let $D_{n\tau}(\theta_{1\tau}, \gamma_\tau) \equiv S_n(\theta_{1\tau}, \gamma_\tau) - S_n\left(\theta_\tau^0, \gamma_\tau^0\right)$, $D_{1n\tau}(w_1) \equiv S_n\left(\theta_{1\tau}^0 + n^{-1/2}w_1, \gamma_\tau^0\right) - S_n\left(\theta_{1\tau}^0, \gamma_\tau^0\right)$, and $D_{n2}(\theta_{1\tau}, \gamma_\tau) \equiv S_n(\theta_{1\tau}, \gamma_\tau) - S_n\left(\theta_{1\tau}, \gamma_\tau^0\right)$. Then we have

$$D_{n\tau}(\theta_{1\tau}, \gamma_\tau) = D_{1n\tau}\left(n^{1/2}(\theta_{1\tau} - \theta_{1\tau}^0)\right) + D_{2n\tau}(\theta_{1\tau}, \gamma_\tau). \tag{6.4}$$

Let $\bar{D}_{2n\tau}(w_1, w_2) \equiv D_{2n\tau}\left(\theta_{1\tau}^0 + n^{-1/2}w_1, \gamma_\tau^0 + n^{-1+2a}w_2\right)$. The following two lemmas study the asymptotic properties of $D_{1n\tau}(w_1)$ and $\bar{D}_{2n\tau}(w_1, w_2)$.

Lemma 6.2 *For every $M \in (0,\infty)$, we have* $\sup_{\|w_1\|\leq M} |D_{1n\tau}(w_1) + n^{-1/2} w_1' \sum_{i=1}^n z_i(\gamma_\tau^0)\psi_\tau(\varepsilon_{i\tau})$ $-\frac{1}{2}w_1'\Omega_{1\tau}w_1| = o_P(1)$ *where* $\Omega_{1\tau} \equiv \Omega_1(\tau,\gamma_\tau^0) = E[z_i(\gamma_\tau^0) z_i(\gamma_\tau^0)' f(\theta_{1\tau}^{0\prime} z_i(\gamma_\tau^0)|z_i)]$.

Proof. Let $d_{1n\tau}(w_1) \equiv D_{1n\tau}(w_1) + n^{-1/2}w_1'\sum_{i=1}^n z_i(\gamma_\tau^0)\psi_\tau(\varepsilon_{i\tau}) - \frac{1}{2}w_1'\Omega_{1\tau}w_1$. By Knight's identity,

$$\begin{aligned}
D_{1n\tau}(w_1) &= \sum_{i=1}^n \left[\rho_\tau\left(\varepsilon_{i\tau} - n^{-1/2}w_1'z_i(\gamma_\tau^0)\right) - \rho_\tau(\varepsilon_{i\tau})\right] \\
&= -n^{-1/2}w_1'\sum_{i=1}^n z_i(\gamma_\tau^0)\psi_\tau(\varepsilon_{i\tau}) + n^{-1/2}w_1'\sum_{i=1}^n z_i(\gamma_\tau^0)\int_0^1 \pi_i(w_1,s)\,ds,
\end{aligned}$$

where $\pi_i(w_1,s) = 1\{\varepsilon_{i\tau} \leq sn^{-1/2}w_1'z_i(\gamma_\tau^0)\} - 1\{\varepsilon_{i\tau}\leq 0\}$. It follows that

$$\begin{aligned}
d_{1n\tau}(w_1) &= n^{-1/2}w_1'\sum_{i=1}^n z_i(\gamma_\tau^0)\int_0^1 \pi_i(w_1,s)\,ds - \frac{1}{2}w_1'\Omega_{1\tau}w_1 \\
&= \int_0^1 n^{-1/2}\sum_{i=1}^n \left[w_1'z_i(\gamma_\tau^0)\pi_i(w_1,s) - sn^{-1/2}w_1'\Omega_{1\tau}w_1\right] ds \\
&= \int_0^1 n^{-1/2}\sum_{i=1}^n w_1'z_i(\gamma_\tau^0)\{\pi_i(w_1,s) - E[\pi_i(w_1,s)|z_i]\}\,ds \\
&+ \int_0^1 n^{-1/2}\sum_{i=1}^n \left\{w_1'z_i(\gamma_\tau^0)E[\pi_i(w_1,s)|z_i] - sn^{-1/2}w_1'\Omega_{1\tau}w_1\right\} ds \\
&\equiv d_{1n\tau,1}(w_1) + d_{1n\tau,2}(w_1).
\end{aligned}$$

The pointwise convergence of $d_{1n\tau,1}(w_1)$ can be established by the Chebyshev inequality. Its uniform convergence results follows from similar arguments as used in the proof of Lemma 1 in Su and Xiao (2008). For $d_{1n\tau,2}(w_1)$ we apply the Taylor expansion with integral remainder and the triangle inequality to obtain

$$\begin{aligned}
&\sup_{\|w_1\|\leq b}|d_{1n\tau,2}(w_1)| \\
&= \sup_{\|w_1\|\leq b}\left|n^{-1/2}w_1'\sum_{i=1}^n z_i(\gamma_\tau^0)\int_0^1\left[F\left(\varsigma_{i\tau} + sn^{-1/2}w_1'z_i(\gamma_\tau^0)|z_i\right) - F(\varsigma_{i\tau}|z_i)\right]ds - \frac{1}{2}w_1'\Omega_{1\tau}w_1\right| \\
&\leq \sup_{\|w_1\|\leq b}\left|n^{-1}w_1'\sum_{i=1}^n z_i(\gamma_\tau^0)z_i(\gamma_\tau^0)'w_1\int_0^1\int_0^1\left[f\left(\varsigma_{i\tau} + stn^{-1/2}w_1'z_i(\gamma_\tau^0)|z_i\right) - f(\varsigma_{i\tau}|z_i)\right]dtds\right| \\
&+ \frac{1}{2}\sup_{\|w_1\|\leq M}\left|n^{-1}w_1'\sum_{i=1}^n\left[z_i(\gamma_\tau^0)z_i(\gamma_\tau^0)'f(\varsigma_{i\tau}|z_i) - \Omega_{1\tau}\right]w_1\right| \\
&\leq \left|M^2n^{-1}\sum_{i=1}^n\left\|z_i(\gamma_\tau^0)z_i(\gamma_\tau^0)'\right\|\int_0^1\int_0^1\sup_{\|w_1\|\leq M}\left|f\left(\varsigma_{i\tau} + stn^{-1/2}w_1'z_i(\gamma_\tau^0)|z_i\right) - f(\varsigma_{i\tau}|z_i)\right|dtds\right| \\
&+ \frac{M^2}{2}\left\|n^{-1}\sum_{i=1}^n\left[z_i(\gamma_\tau^0)z_i(\gamma_\tau^0)'f(\varsigma_{i\tau}|z_i) - \Omega_{1\tau}\right]\right\| \\
&= o_P(1) + o_P(1) = o_P(1),
\end{aligned}$$

where $\varsigma_{i\tau} \equiv \theta_{1\tau}^{0\prime} z_i\left(\gamma_\tau^0\right)$, and the last line follows from the dominated convergence theorem and the weak law of large numbers (LLN). ■

Lemma 6.3 *Let $w_2 \in \mathbb{R}$. For every $M \in (0,\infty)$, we have $\sup_{\|w_1\|\leq M}\left|\bar{D}_{2n\tau}(w_1,w_2) - \bar{D}_{2n\tau}(0,w_2)\right| = o_P(1)$.*

Proof. Without loss of generality, we consider the case $w_2 > 0$. Let $1_i(r) \equiv 1\left\{\gamma_\tau^0 < r_i \leq \gamma_\tau^0 + r\right\}$. Let $r \equiv \gamma_\tau - \gamma_\tau^0$ and $\Delta_{1\tau} \equiv \theta_{1\tau} - \theta_{1\tau}^0$, $a_1(x) \equiv (\mathbf{0}_{1\times p}, x')'$, $a_2(x) \equiv (x', \mathbf{0}_{1\times p})'$ and $a(x) \equiv a_1(x) - a_2(x)$, where $\mathbf{0}_{1\times p}$ denotes a $1\times p$ vector of zeros. Noting that $z_i(\gamma_\tau) - z_i\left(\gamma_\tau^0\right) = -a(x_i)\,1_i(r)$ when $r>0$, we observe that: (i) if $1_i(r) = 0$,

$$\begin{aligned} y_i - \theta_{1\tau}' z_i(\gamma_\tau) &= \varepsilon_{i\tau} - \theta_{1\tau}^{0\prime}\left[z_i(\gamma_\tau) - z_i\left(\gamma_\tau^0\right)\right] - \Delta_{1\tau}'\left[z_i(\gamma_\tau) - z_i\left(\gamma_\tau^0\right)\right] - \Delta_{1\tau}' z_i\left(\gamma_\tau^0\right) \\ &= \varepsilon_{i\tau} - \Delta_{1\tau}' z_i\left(\gamma_\tau^0\right) = y_i - \theta_{1\tau}' z_i\left(\gamma_\tau^0\right); \end{aligned}$$

and (ii) if $1_i(r) = 1$, $z_i\left(\gamma_\tau^0\right) = a_1(x_i)$ and $z_i(\gamma_\tau) = a_2(x_i)$. It follows that if $r>0$, then [17]

$$\begin{aligned} &\rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i\left(\gamma_\tau^0\right)\right) \\ &= \left[\rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i\left(\gamma_\tau^0\right)\right)\right] 1_i(r) \\ &= \left\{\rho_\tau\left(\varepsilon_{i\tau} - \theta_{1\tau}^{0\prime}\left[z_i(\gamma_\tau) - z_i\left(\gamma_\tau^0\right)\right] - \Delta_{1\tau}' z_i(\gamma_\tau)\right) - \rho_\tau\left(\varepsilon_{i\tau} - \Delta_{1\tau}' z_i\left(\gamma_\tau^0\right)\right)\right\} 1_i(r) \\ &= \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a(x_i) - \Delta_{1\tau}' a_2(x_i)\right) - \rho_\tau\left(\varepsilon_{i\tau} - \Delta_{1\tau}' a_1(x_i)\right)\right] 1_i(r). \end{aligned} \tag{6.5}$$

By Knight's identity, we have

$$\begin{aligned} &\bar{D}_{2n\tau}(w_1,w_2) - \bar{D}_{2n\tau}(0,w_2) \\ &= D_{2n\tau}\left(\theta_{1\tau}^0 + n^{-1/2}w_1, \gamma_\tau^0 + n^{-1+2a}w_2\right) - D_{2n\tau}\left(\theta_{1\tau}^0, \gamma_\tau^0 + n^{-1+2a}w_2\right) \\ &= \sum_{i=1}^n \Big\{\left[\rho_\tau\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a(x_i) - n^{-1/2} w_1' a_2(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})\right] \\ &- \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})\right] \\ &- \left[\rho_\tau\left(\varepsilon_{i\tau} - n^{-1/2} w_1' a_1(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})\right]\Big\} 1_i\left(n^{-1+2a}w_2\right) \\ &= n^{-1/2}\sum_{i=1}^n \psi_\tau(\varepsilon_{i\tau}) w_1' a(x_i)\, 1_i\left(n^{-1+2a}w_2\right) \\ &+ \sum_{i=1}^n \int_{-\theta_{1\tau}^{0\prime} a(x_i)}^{-\theta_{1\tau}^{0\prime} a(x_i) + n^{-1/2} w_1' a_2(x_i)} \left[1\{\varepsilon_{i\tau} \leq s\} - 1\{\varepsilon_{i\tau} \leq 0\}\right] ds\; 1_i\left(n^{-1+2a}w_2\right) \\ &- \sum_{i=1}^n \int_0^{n^{-1/2} w_1' a_1(x_i)} \left[1\{\varepsilon_{i\tau} \leq s\} - 1\{\varepsilon_{i\tau} \leq 0\}\right] ds\, 1_i\left(n^{-1+2a}w_2\right) \\ &\equiv d_{2n\tau,1}(w_1) + d_{2n\tau,2}(w_1) - d_{2n\tau,3}(w_1), \text{ say.} \end{aligned}$$

To study the uniform bound for $d_{2n\tau,1}(w_1)$, we consider the class of functions

$$\mathcal{F}_1 = \left\{m_1(y,z;\; w_1,\gamma) : w_1 \in \Theta_M, \gamma \in \Gamma\right\},$$

[17] Similarly, if $r<0$, then we can show that $\rho_\tau\left(y_i - \theta_{1\tau}' \dot{m}_{\gamma_\tau}(z_i)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' \dot{m}_{\gamma_\tau^0}(z_i)\right) = [\rho_\tau(\varepsilon_{i\tau} - \theta_{1\tau}^{0\prime} a(x_i) - \Delta_{1\tau}' a_1(x_i)) - \rho_\tau\left(\varepsilon_{i\tau} - \Delta_{1\tau}' a_2(x_i)\right)]\bar{1}_i(r)$, where $\bar{1}_i(r) = 1\left\{\gamma_\tau^0 + r < q_i \leq \gamma_\tau^0\right\}$.

where $m_1(y,z;\ w_1,\gamma) = \psi_\tau(y - \theta_{1\tau}^{0\prime} z(\gamma)) w_1' a(x) 1\{q \le \gamma\}$, $z(\gamma) = [x' 1\{q \le \gamma\}\ \ x' 1\{q > \gamma\}]'$, and $\Theta_M = \{\omega_1 : \|\omega_1\| \le M\}$. Let

$$\begin{aligned}
\mathcal{F}_{1,1} &= \{f_{1,1}(y,z;\gamma) = 1\{q \le \gamma\} : \gamma \in \Gamma\}, \\
\mathcal{F}_{1,2} &= \left\{f_{1,2}(y,z;\gamma) = \tau - 1\left\{y - \theta_{1\tau}^{0\prime} z(\gamma) < 0\right\} : \gamma \in \Gamma\right\}, \\
\mathcal{F}_{1,3} &= \{f_{1,3}(y,z;w_1) = w_1' a(x) : w_1 \in \Theta_M\}.
\end{aligned}$$

By Lemma 2.6.15 of Van der Vaart and Wellner (1996, hereafter VW), $\mathcal{F}_{1,1}$ is a VC-subgraph class. Noting that $\theta_{1\tau}^{0\prime} z(\gamma) = \beta_\tau^{0\prime} x + \delta_\tau^{0\prime} x 1\{q \le \gamma\}$ where $\delta_\tau^0 = \alpha_\tau^0 - \beta_\tau^0$, $\mathcal{F}_{1,2}$ is also a VC-subgraph class by Lemma 2.6.15(viii) of VW. $\mathcal{F}_{1,3}$ is Euclidean for the envelope $C\|x\|$ by Theorem 2.7.11 of VW or Lemma 2.13 of Pakes and Pollard (1989, PP hereafter). Noting that the VC-subgraph class is Euclidean for every envelope and the product of Euclidean classes of functions is also Euclidean (see Lemmas 2.13 and 2.14(iii) of PP), we conclude that $\mathcal{F}_1 = \mathcal{F}_{1,1} \cdot \mathcal{F}_{1,2} \cdot \mathcal{F}_{1,3}$ is Euclidean. Then by Assumption A2 and Lemma 2.17 of PP we have

$$\sup_{\|w_1\| \le M} |d_{2n\tau,1}(w_1)| = \sup_{\|w_1\| \le M} \left| n^{-1/2} \sum_{i=1}^n \left[m_1\left(y_i, z_i;\ w_1, \gamma_\tau^0 + n^{-1+2a} w_2\right) - m_1\left(y_i, z_i;\ w_1, \gamma_\tau^0\right)\right] \right| = o_P(1)$$

as $n^{-1+2a} \to 0$ as $n \to \infty$ by Assumption A7.

Next we study $d_{2n\tau,3}(w_1)$. Write $d_{2n\tau,3}(w_1) = n^{-1/2} \sum_{i=1}^n w_1' a_1(x_i) \int_0^1 [1\{\varepsilon_{i\tau} \le n^{-1/2} w_1' a_1(x_i) s\}$ $-1\{\varepsilon_{i\tau} \le 0\}] ds\ \ 1_i\left(n^{-1+2a} w_2\right)$. Let $m_2(y,z;\ w_1,\bar{w}_1) = w_1' a_1(x) \int_0^1 1\left\{y - \theta_{1\tau}^{0\prime} z(\gamma) \le \bar{w}_1' a_1(x) s\right\} \times$ $1\{\gamma_\tau^0 < q \le \gamma_\tau^0 + n^{-1+2a} w_2\}$. We consider the class of functions

$$\mathcal{F}_2 = \{m_2(y,z;\ w_1,\bar{w}_1) : w_1 \in \Theta_M,\ \bar{w}_1 \in \Theta_M\}.$$

Let

$$\mathcal{F}_{2,1} = \left\{ f_{2,2}(y,z;\ \bar{w}_1) = \int_0^1 1\left\{y - \theta_{1\tau}^{0\prime} z\left(\gamma_\tau^0\right) \le \bar{w}_1' a_1(x) s\right\} ds : \bar{w}_1 \in \Theta_M \right\}.$$

By Andrews (1994, p. 2270), both $\mathcal{F}_{1,3}$ and $\mathcal{F}_{2,1}$ belong to the type I class of functions and satisfy the Pollard's entropy condition. Noting that $\mathcal{F}_2$ can be written as the product of $\mathcal{F}_{1,3}$, $\mathcal{F}_{2,1}$, and a fixed indicator function $1\left\{\gamma_\tau^0 < q \le \gamma_\tau^0 + n^{-1+2a} w_2\right\}$, it also satisfies that Pollard's entropy condition and is stochastically equicontinuous with respect to the pseudometric defined by

$$\rho\left((w_1,\bar{w}_1),(w_1^*,\bar{w}_1^*)\right) = \left\{ E\left[\left| m_2(y_i,z_i;\ w_1,\bar{w}_1) - m_2(y_i,z_i;\ w_1^*,\bar{w}_1^*)\right|^2 \right] \right\}^{1/2}.$$

Consequently, letting $\overline{m}_2(y_i, z_i;\ w_1, \bar{w}_1) = m_2(y_i, z_i;\ w_1, \bar{w}_1) - E[m_2(y_i, z_i;\ w_1, \bar{w}_1)]$, we have

$$
\begin{aligned}
&\sup_{\|w_1\|\le M} |d_{2n\tau,3}(w_1)| \\
&\le \sup_{\|w_1\|\le M} \left| n^{-1/2} \sum_{i=1}^{n} E\left[m_2\left(y_i, z_i;\ w_1, n^{-1/2} w_1\right) - m_2(y_i, z_i;\ w_1, 0) \right] \right| \\
&+ \sup_{\|w_1\|\le M} \left| n^{-1/2} \sum_{i=1}^{n} \left[\bar{m}_2\left(y_i, z_i;\ w_1, n^{-1/2} w_1\right) - \bar{m}_2(y_i, z_i;\ w_1, 0) \right] \right| \\
&= \sup_{\|w_1\|\le M} \left| n^{-1/2} \sum_{i=1}^{n} E\left[m_2\left(y_i, z_i;\ w_1, n^{-1/2} w_1\right) - m_2(y_i, z_i;\ w_1, 0) \right] \right| + o_P(1) \\
&= \sup_{\|w_1\|\le M} \left| n^{-1/2} \sum_{i=1}^{n} E\left\{ w_1' a_1(x_i) \int_0^1 \left[F\left(n^{-1/2} w_1' a_1(x_i) s |\ z_i\right) - F(0|z_i) \right] ds\ 1_i\left(n^{-1+2a} w_2\right) \right\} \right| + o_P(1) \\
&\le M \left| n^{-1/2} \sum_{i=1}^{n} E\left\{ \|a_1(x_i)\| \int_0^1 \left[F\left(n^{-1/2} M \|a_1(x_i)\| s \mid z_i\right) - F(0|z_i) \right] ds\ 1_i\left(n^{-1+2a} w_2\right) \right\} \right| + o_P(1) \\
&\le M n^{1/2} \left\| \|a_1(x_i)\| \int_0^1 \left[F\left(n^{-1/2} M \|a_1(x_i)\| s \mid z_i\right) - F(0|z_i) \right] ds \right\|_2 \left\| 1_i\left(n^{-1+2a} w_2\right) \right\|_2 + o_P(1) \\
&= n^{1/2} O\left(n^{-1/2}\right) o\left(n^{-1+2a}\right) + o_P(1) = o_P(1),
\end{aligned}
$$

where $\|A\|_2 \equiv \{E\|A\|^2\}^{1/2}$.

By the same token, we can show that $\sup_{\|w_1\|\le M} |d_{2n\tau,2}(w_1)| = o_P(1)$. This completes the proof of the lemma. ■

Proof of Theorem 3.3

First, observe that $\hat{\theta}_\tau$ is also minimizing $\bar{S}_{n\tau}(\theta_\tau) \equiv n^{-1} \sum_{i=1}^{n} s_\tau(y_i, z_i; \theta_\tau)$, where $s_\tau(y_i, z_i; \theta_\tau) \equiv \rho_\tau(y_i - \theta_{1\tau}' z_i(\gamma_\tau)) - \rho_\tau\left(y_i - \theta_{1\tau}^{0\prime} z_i\left(\gamma_\tau^0\right)\right)$. Let $\Delta(z_i, \theta_\tau) \equiv \theta_{1\tau}' z_i(\gamma_\tau) - \theta_{1\tau}^{0\prime} z_i\left(\gamma_\tau^0\right)$ and $\varsigma_\tau(\theta_\tau) \equiv E[s_\tau(y_i, z_i; \theta_\tau)]$. Then by Knight's identity, the compactness of Θ_1 and Assumption A2, $E|s_\tau(y_i, x_i; \theta_\tau)| \le 2E|\Delta(z_i, \theta_\tau)| \le 2(\|\alpha_\tau\| + \|\beta_\tau\|) E\|x_i\| < \infty$. This ensures that $S_n(\theta_\tau) = \varsigma_\tau(\theta_\tau) + o_{a.s.}(1)$ for each $\theta_\tau \in \Theta$ by the strong LLN. By the proof of Lemma 2 in Galvao et al. (2011), the class of functions $\mathcal{F} = \{s_\tau(y, z; \theta) : \theta \in \Theta\}$is Glivenko-Cantelli. It follows that $\bar{S}_{n\tau}(\theta_\tau) = \varsigma_\tau(\theta_\tau) + o_{a.s.}(1)$ uniformly in $\theta_\tau \in \Theta$.

Let $l_\tau(c) \equiv E[\rho_\tau(\varepsilon_{i\tau} - c) - \rho_\tau(\varepsilon_{i\tau})]$. Knight's identity implies $l_\tau(c) > 0$ for any $c \ne 0$. Then by the law of iterated expectations and Assumption A6, we have

$$\varsigma_\tau(\theta_\tau) = E[E[\rho_\tau(\varepsilon_{i\tau} - \Delta(z_i, \theta_\tau)) - \rho_\tau(\varepsilon_{i\tau}) | z_i]] = E[l_\tau(\Delta(z_i, \theta_\tau))] > 0 \text{ for all } \theta_\tau \ne \theta_\tau^0.$$

By Lemma (6.1) $\varsigma_\tau(\theta_\tau)$ is continuous in θ_τ. It follows that θ_τ^0 is the unique minimizer of $\varsigma_\tau(\theta_\tau)$ and $\hat{\theta}_\tau \to \hat{\theta}_\tau$ a.s. ■

Proof of Theorem 3.4

The proof is similar to that of Theorem 3.2 in Koul et al. (2003), so we only sketch the main difference. Let $\Omega(\sigma) \equiv \{\theta_\tau \in \Theta : \left\|\theta_{1\tau} - \theta_{1\tau}^0\right\| < \sigma,\ \left|\gamma_\tau - \gamma_\tau^0\right| < \sigma\}$, where $\sigma \in (0, 1)$ can be chosen sufficiently small by Theorem 3.3. Let $b \in (0, \infty)$. Define

$$N_{1b} \equiv \{\theta_\tau \in \Omega(\sigma) :\ \left|\gamma_\tau - \gamma_\tau^0\right| > bn^{2a-1}\} \text{ and } N_{2b} \equiv \{\theta_\tau \in \Omega(\sigma) : \left\|\theta_{1\tau} - \theta_{1\tau}^0\right\| > bn^{-1/2}\}.$$

Noting that $\inf_{\theta_\tau \in N_{1b} \cup N_{2b}} D_{n\tau}(\theta_{1\tau}, \gamma_\tau) = \min\{\inf_{\theta_\tau \in N_{1b}} D_{n\tau}(\theta_{1\tau}, \gamma_\tau), \inf_{\theta_\tau \in N_{2b}} D_{n\tau}(\theta_{1\tau}, \gamma_\tau)\}$, one can prove the theorem by showing that for any $\kappa \in (0,1]$, $c_1 \in (0,\infty)$ and $c_2 \in (0,\infty)$, there exists $b \in (0,\infty)$ and n_0 such that

$$P\left(\inf_{\theta_\tau \in N_{jb}} D_{n\tau}(\theta_{1\tau}, \gamma_\tau) > c_j\right) > 1 - \kappa \text{ for } n > n_0,\ j = 1, 2, \tag{6.6}$$

because then $\inf_{\theta_\tau \in N_{1b} \cup N_{2b}} D_{n\tau}(\theta_{1\tau}, \gamma_\tau) > c_1 \wedge c_2 > 0$ with positive probability, implying that $\hat{\theta}_\tau \notin N_{1b} \cup N_{2b}$ as $D_{n\tau}(\hat{\theta}_{1\tau}, \hat{\gamma}_\tau) \equiv S_{n\tau}(\hat{\theta}_{1\tau}, \hat{\gamma}_\tau) - S_{n\tau}(\theta^0_{1\tau}, \gamma^0_\tau) < 0$ by the definition of $\hat{\theta}_\tau = (\hat{\theta}'_{1\tau}, \hat{\gamma}_\tau)'$.

Noting that for $j = 1, 2$,

$$\inf_{\theta_\tau \in N_{jb}} D_{n\tau}(\theta_{1\tau}, \gamma_\tau) \geq \inf_{\theta_\tau \in N_{jb}} D_{1n\tau}\left(n^{1/2}(\theta_{1\tau} - \theta^0_{1\tau})\right) + \inf_{\theta_\tau \in N_{jb}} D_{2n\tau}(\theta_{1\tau}, \gamma_\tau) \equiv D_{n\tau,j1} + D_{n\tau,j2}, \text{ say,}$$

it suffices to analyze $D_{n\tau,11}$, $D_{n\tau,12}$, $D_{n\tau,21}$, and $D_{n\tau,22}$.

We first analyze $D_{n\tau,12}$. By Koul et al. (2003) it suffices to show that for all $\kappa \in (0,1]$, $c_1 \in (0,\infty)$, there exists $c_0 < \infty$, $b_0 \in (0,\infty)$, $\sigma \in (0,1)$ and n_0 such that $c_0 b_0 g(\gamma^0_\tau)/2 > c_1$ and that

$$P\left(\inf_{\theta_\tau \in N_{jb}} \frac{D_{2n\tau}(\theta_{1\tau}, \gamma_\tau)}{nK(|\gamma_\tau - \gamma^0_\tau|)} > c_0\right) > 1 - \kappa/2 \text{ for all } n > n_0, \tag{6.7}$$

where $K(r) \equiv E[1_i(r)]$ and $1_i(r) \equiv 1\{\gamma^0_\tau < r_i \leq \gamma^0_\tau + r\}$. Let $r \equiv \gamma_\tau - \gamma^0_\tau$ and $\Delta_{1\tau} \equiv \theta_{1\tau} - \theta^0_{1\tau}$. Without loss of generality, assume that $r > 0$. Then by (6.5) we can decompose $n^{-1} D_{2n\tau}(\theta_{1\tau}, \gamma_\tau)$ as follows

$$\begin{aligned}
n^{-1} D_{2n\tau}(\theta_{1\tau}, \gamma_\tau) &= n^{-1} \sum_{i=1}^n \left[\rho_\tau(y_i - \theta'_{1\tau} z_i(\gamma_\tau)) - \rho_\tau(y_i - \theta'_{1\tau} z_i(\gamma^0_\tau))\right] \\
&= n^{-1} \sum_{i=1}^n \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau} a(x_i) - \Delta'_{1\tau} a_2(x_i)\right) - \rho_\tau(\varepsilon_{i\tau} - \Delta'_{1\tau} a_1(x_i))\right] 1_i(r) \\
&= n^{-1} \sum_{i=1}^n \left[\rho_\tau(\varepsilon_{i\tau}) - \rho_\tau(\varepsilon_{i\tau} - \Delta'_{1\tau} a_1(x_i))\right] 1_i(r) \\
&\quad + n^{-1} \sum_{i=1}^n \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau} a(x_i) - \Delta'_{1\tau} a_2(x_i)\right) - \rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau} a(x_i)\right)\right] 1_i(r) \\
&\quad + n^{-1} \sum_{i=1}^n \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau} a(x_i)\right) - \rho_\tau(\varepsilon_{i\tau}) - p(r_i)\right] 1_i(r) \\
&\quad + n^{-1} \sum_{i=1}^n \left[p(r_i) - p(\gamma^0_\tau)\right] 1_i(r) + p(\gamma^0_\tau)[K_n(r) - K(r)] + p(\gamma^0_\tau) K(r) \\
&\equiv d_{n\tau 1}(\theta_{1\tau}, r) + d_{n\tau 2}(\theta_{1\tau}, r) + d_{n\tau 3}(r) + d_{n\tau 4}(r) + d_{n\tau 5}(r) + d_{n\tau 6}(r), \text{ say,}
\end{aligned}$$

where $p(\gamma^0_\tau) = E\{\left[\rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau} a(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})\right] | r_i = \gamma^0_\tau\}$ and $K_n(r) = n^{-1} \sum_{i=1}^n 1_i(r)$. By Knight's identity, the law of iterated expectations, and Fubini's theorem, we can readily show that $p(\gamma)$ is strictly positive and continuous in γ under our assumptions. Following Koul et al. (2003) and Hansen (2000) we can show that the first five terms in the last decomposition are asymptotically negligible in comparison with $K(r)$ uniformly on N_{1b} by modifying the proof of their Lemma 3.2 to accommodate our definition of N_{1b}. For example, we need to prove Claim (i) below in order to analyze $d_{n\tau 4}(r)$ and $d_{n\tau 5}(r)$ because then $\sup_{B/n^{1-2a}\leq r\leq\sigma}|d_{n\tau 4}(r)/K(r)| \leq \sup_{0\leq r\leq\sigma}|p(\gamma_\tau^0+r)-p(\gamma_\tau^0)|$ $\sup_{B/n^{1-2a}\leq r\leq\sigma}|K_n(r)/K(r)| \to 0$ and $\sup_{B/n^{1-2a}\leq r\leq\sigma}|d_{n\tau 5}(r)/K(r)| \leq C\sup_{B/n^{1-2a}\leq r\leq\sigma}|K_n(r)$ $/K(r)-1| \to 0$ and by passing $n\to\infty$ and then $\sigma\to 0$. Similarly, Claim (ii) below is needed to show that $d_{n\tau 3}(\theta_{1\tau}, r)/K(r)$ is asymptotically negligible on the set N_{1b}.

Claim. *For each $\kappa>0$, $c>0$, there exists a constant $B\in(0,\infty)$ such that for all $\sigma\in(0,1)$ and $n\geq[B/\sigma]+1$, we have*

(i) $P\left(\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|\frac{K_n(r)}{K(r)}-1\right|<c\right)>1-\kappa,$

(ii) $P\left(\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|\frac{R_n(r)}{K(r)}\right|<c\right)>1-\kappa,$

where $R_n(r)=n^{-1}\sum_{i=1}^n\{J(x_i,\varepsilon_{i\tau})-E[J(x_i,\varepsilon_{i\tau})|r_i=r]\}1_i(r)$ *and* $J(x_i,\varepsilon_{i\tau})\equiv\rho_\tau\left(\varepsilon_{i\tau}+\theta_{1\tau}^{0\prime}a(x_i)\right)-\rho_\tau(\varepsilon_{i\tau})$.

It follows that $n^{-1}D_{2n\tau}(\theta_{1\tau},\gamma_\tau)/K(r)=p\left(\gamma_\tau^0\right)+o_P(1)>c>0$ with probability approaching 1 as $n\to\infty$ uniformly in $\theta_\tau\in N_{1b}$, and thus (6.7) follows.

The analyses of $D_{n\tau,11}$, $D_{n\tau,12}$, and $D_{n\tau,22}$ are analogous to those of the corresponding terms in the proof of Theorem 3.2 in Koul et al. (2003) and thus omitted. ■

Proof of Theorem 3.5

(i) By Theorem 3.4, it suffices to study the asymptotic behavior of $D_{n\tau}(\theta_{1\tau},\gamma_\tau)$ by restricting our attention to the case where $n^{1/2}\left\|\theta_{1\tau}-\theta_{1\tau}^0\right\|\leq M$ and $n^{1-2a}\left|\gamma_\tau-\gamma_\tau^0\right|\leq M$ for some large but fixed positive number M. Let $w_{1n\tau}\equiv n^{1/2}(\hat\theta_{1\tau}-\theta_{1\tau}^0)$ and $w_{2n\tau}\equiv n^{1-2a}\left(\hat\gamma_\tau-\gamma_\tau^0\right)$. Then by (6.4),

$$D_{n\tau}\left(\theta_{1\tau}^0+n^{-1/2}w_1,\gamma_\tau^0+n^{-1+2a}w_2\right)=D_{1n\tau}(w_1)+\bar D_{2n\tau}(w_1,w_2)$$

where recall $\bar D_{2n\tau}(w_1,w_2,)=D_{2n\tau}\left(\theta_{1\tau}^0+n^{-1/2}w_1,\gamma_\tau^0+n^{-1+2a}w_2\right)$. By Lemmata 6.2 and 6.3, we have

$$D_{n\tau}\left(\theta_{1\tau}^0+n^{-1/2}w_1,\gamma_\tau^0+n^{-1+2a}w_2\right)=\bar D_{1n\tau}(w_1)+\bar D_{2n\tau}(0,w_2)+o_P(1),$$

where $o_P(1)$ holds uniformly over the set $\|w_1\|\leq M$ and $|w_2|\leq M$, $\bar D_{1n\tau}(w_1)=-n^{-1/2}w_1'\sum_{i=1}^n z_i\left(\gamma_\tau^0\right)$ $\times\psi_\tau(\varepsilon_{i\tau})+\frac{1}{2}w_1'\Omega_{1\tau}w_1$, and

$$\bar D_{2n\tau}(0,w_2)=\begin{cases}\sum_{i=1}^n\left[\rho_\tau\left(\varepsilon_{i\tau}+\theta_{1\tau}^{0\prime}a(x_i)\right)-\rho_\tau(\varepsilon_{i\tau})\right]1_i\left(n^{-1+2a}w_2\right) & \text{if } w_2>0\\ \sum_{i=1}^n\left[\rho_\tau\left(\varepsilon_{i\tau}-\theta_{1\tau}^{0\prime}a(x_i)\right)-\rho_\tau(\varepsilon_{i\tau})\right]\bar 1_i\left(n^{-1+2a}w_2\right) & \text{if } w_2\leq 0\end{cases}. \tag{6.8}$$

Thus $D_{n\tau}(\hat\theta_{1\tau},\hat\gamma_\tau)=D_{1n\tau}(w_{1n})+\bar D_{2n\tau}(w_{1n},w_{2n})=\bar D_{1n\tau}(w_{1n})+\bar D_{2n\tau}(0,w_{2n})+o_P(1)$. Noting that $\bar D_{1n\tau}(w_1)$ and $\bar D_{2n\tau}(0,w_2)$ are free of w_2 and w_1, respectively, and (w_{1n},w_{2n}) is a minimizer of $D_{n\tau}\left(\theta_{1\tau}^0+n^{-1/2}w_1,\gamma_\tau^0+n^{-1+2a}w_2\right)$ with respect to (w_1,w_2), the asymptotic distribution of w_{1n} is determined by that of the minimizer of $\bar D_{1n\tau}(w_1)$ with respect to w_1, and similarly the asymptotic distribution of w_{2n} is determined by that of the minimizer of $\bar D_{2n\tau}(0,w_2)$ with respect to w_2.

Noting that $\bar D_{1n\tau}(w_1)$ is convex in w_1, we can readily apply the convexity lemma to obtain

$$w_{1n\tau}=n^{-1/2}\Omega_{1\tau}^{-1}\sum_{i=1}^n z_i\left(\gamma_\tau^0\right)\psi_\tau(\varepsilon_{i\tau})+o_P(1)\xrightarrow{d}N\left(\mathbf{0}_{2k\times 1},\tau(1-\tau)\Sigma\left(\tau,\gamma_\tau^0\right)\right).$$

where recall $\Sigma(\tau,\gamma)=\Omega_1(\tau,\gamma)^{-1}\Omega_0(\gamma,\gamma)\Omega_1(\tau,\gamma)^{-1}$. This proves (i).

We now prove (ii). By reversing the argument used to obtain (6.5) we find that it is convenient to rewrite $\bar{D}_{2n\tau}(0,w_2)$ as

$$\begin{aligned}\bar{D}_{2n\tau}(0,w_2)&=\sum_{i=1}^{n}\left[\rho_\tau\left(\varepsilon_{i\tau}+\theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0\right)-\theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0+n^{-1+2a}w_2\right)\right)-\rho_\tau\left(\varepsilon_{i\tau}\right)\right]\\&=\sum_{i=1}^{n}\left[\rho_\tau\left(\varepsilon_{i\tau}+\delta_\tau^{0\prime}\Delta x_i\left(w_2\right)\right)-\rho_\tau\left(\varepsilon_{i\tau}\right)\right],\end{aligned}$$

where $\Delta x_i(w_2)=x_i\left[1\left\{r_i\le\gamma_\tau^0+n^{-1+2a}w_2\right\}-1\left\{r_i\le\gamma_\tau^0\right\}\right]$ and (recall) $\delta_\tau^0=\delta_{n\tau}^0=\alpha_{n\tau}^0-\beta_{n\tau}^0=v_\tau n^{-a}$. Using Knight's identity,

$$\begin{aligned}\bar{D}_{2n\tau}(0,w_2)&=\sum_{i=1}^{n}\psi_\tau\left(\varepsilon_{i\tau}\right)\delta_\tau^{0\prime}\Delta x_i\left(w_2\right)+\sum_{i=1}^{n}\int_0^{-\delta_\tau^{0\prime}\Delta x_i(w_2)}\left[1\left\{\varepsilon_{i\tau}\le s\right\}-1\left\{\varepsilon_{i\tau}\le 0\right\}\right]ds\\&\equiv\bar{D}_{2n\tau,1}\left(w_2\right)+\bar{D}_{2n\tau,2}\left(w_2\right),\text{ say.}\end{aligned}$$

Assume that $w_2>0$. Using arguments similar to those used in the proof of Lemma A.11 in Hansen (2000), we can readily show that

$$\bar{D}_{2n\tau,1}\left(w_2\right)\Rightarrow B_\tau\left(w_2\right),$$

where $B_\tau(w_2)$ is a Brownian motion with variance $E[B_\tau(1)^2]=v_\tau' E[x_ix_i'|r_i=\gamma_\tau^0]v_\tau\ g\left(\gamma_\tau^0\right)\equiv\lambda_\tau$. Analogously to the proof of Lemma 6.3 and using arguments as used in the proof of Lemma A.10 in Hansen (2000), we can show that uniformly in w_2 on a compact set

$$\begin{aligned}\bar{D}_{2n\tau,2}\left(w_2\right)&=-\sum_{i=1}^{n}\delta_\tau^{0\prime}\Delta x_i\left(w_2\right)\int_0^1\left[F\left(\theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0\right)-s\delta_\tau^{0\prime}\Delta x_i\left(w_2\right)|z_i\right)-F\left(\theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0\right)|z_i\right)\right]ds+o_P(1)\\&=\delta_\tau^{0\prime}\sum_{i=1}^{n}f\left(\alpha_\tau^{0\prime}x_i|z_i\right)\Delta x_i\left(w_2\right)\Delta x_i\left(w_2\right)'\delta_\tau^0+o_P(1)=\mu_\tau\left|w_2\right|+o_P(1),\end{aligned}$$

where $\mu_\tau\equiv v_\tau' E[f\left(\alpha_\tau^{0\prime}x_i|z_i\right)x_ix_i'|r_i=\gamma_\tau^0]v_\tau g\left(\gamma_\tau^0\right)$. Noting that $B_\tau(w_2)$ is a Brownian motion with variance λ_τ and thus can be written as $-\sqrt{\lambda_\tau}W_1(w_2)$ with $W_1(w_2)$ being a standard Brownian motion, we have,

$$\bar{D}_{2n\tau}(0,w_2)\Rightarrow\mu_\tau\left|w_2\right|-\sqrt{\lambda_\tau}W_1\left(w_2\right)\ \text{ if } w_2>0. \tag{6.9}$$

Similarly, we can show that

$$\bar{D}_{2n\tau}(0,w_2)\Rightarrow\mu_\tau\left|w_2\right|-\sqrt{\lambda_\tau}W_2\left(-w_2\right)\ \text{ if } w_2\le 0, \tag{6.10}$$

where $W_2(w_2)$ is a standard Brownian motion that is independent of $W_1(w_2)$. Then by the continuous mapping theorem (CMT) and following the proof of Theorem 1 in Hansen (2000), we have

$$\begin{aligned}w_{2n\tau}&\xrightarrow{d}\underset{-\infty<w_2<\infty}{\arg\max}-\left\{\mu_\tau\left|w_2\right|+\sqrt{\lambda_\tau}W\left(w_2\right)\right\}\\&=\frac{\lambda_\tau}{4\mu_\tau^2}\underset{-\infty<r<\infty}{\arg\max}-\left\{\mu_\tau\left|\frac{\lambda_\tau}{4\mu_\tau^2}r\right|+\sqrt{\lambda_\tau}W\left(\frac{\lambda_\tau}{4\mu_\tau^2}r\right)\right\}\\&=\frac{\lambda_\tau}{4\mu_\tau^2}\underset{-\infty<r<\infty}{\arg\max}\left\{-\frac{\lambda_\tau}{4\mu_\tau}|r|+\frac{\lambda}{2\mu_\tau}W(r)\right\}=\frac{\lambda_\tau}{4\mu_\tau^2}\underset{-\infty<r<\infty}{\arg\max}\left\{-\frac{1}{2}|r|+W(r)\right\},\end{aligned}\tag{6.11}$$

by the change of variables $w_2 = \left(\lambda_\tau/(4\mu_\tau^2)\right) r$ and the distributional equality $W\left(c^2 r\right) \equiv cW(r)$. [18]

(iii) follows from the direct asymptotic covariance calculations. ■

Proof of Theorem 3.6

Recall that $w_{1n\tau} \equiv n^{1/2}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0)$ and $w_{2n\tau} \equiv n^{1-2a}\left(\hat{\gamma}_\tau - \gamma_\tau^0\right)$. Let $\bar{w}_{1n\tau} \equiv n^{1/2}\left(\hat{\theta}_1\left(\tau, \gamma_\tau^0\right) - \theta_{1\tau}^0\right)$. By (6.4), the relationship between $D_{2n\tau}$ and $\bar{D}_{2n\tau}$, and Lemma 6.3, we have

$$\begin{aligned} LR_{n\tau}\left(\gamma_\tau^0\right) &= \left[D_{1n\tau}\left(\bar{w}_{1n\tau}\right) - D_{1n\tau}\left(w_{1n\tau}\right)\right] + \left[\bar{D}_{2n\tau}\left(\bar{w}_{1n\tau}, 0\right)\right) - \bar{D}_{2n\tau}\left(w_{1n\tau}, w_{2n\tau}\right)\right)\right] \\ &= \left[D_{1n\tau}\left(\bar{w}_{1n\tau}\right) - D_{1n\tau}\left(w_{1n\tau}\right)\right] - \bar{D}_{2n\tau}\left(0, w_{2n\tau}\right) + o_P(1). \end{aligned}$$

By the proofs of Theorems 3.5(i) and 3.7(ii) below (assuming $\mathcal{T} = \{\tau\}$), $w_{1n\tau}$ and $\bar{w}_{1n\tau}$ are asymptotically equivalent, implying that $D_{1n\tau}\left(\bar{w}_{1n\tau}\right) - D_{1n\tau}\left(w_{1n\tau}\right) = o_P(1)$ by the CMT. This, in conjunction with the analysis of $\bar{D}_{2n\tau}(0, w_2)$ in the the proof of Theorem 3.5(ii) and the CMT, implies that

$$LR_{n\tau}\left(\gamma_\tau^0\right) = -\bar{D}_{2n\tau}\left(0, w_{2n\tau}\right) + o_P(1) \xrightarrow{d} \sup_{w_2}\left\{-\mu_\tau\left|w_2\right| + \sqrt{\lambda_\tau}W\left(w_2\right)\right\} \tag{6.12}$$

By the change of variables $w_2 = \left(\lambda_\tau/(4\mu_\tau^2)\right) r$ and the distributional equality $W\left(c^2 r\right) \equiv cW(r)$, we have

$$\begin{aligned} \sup_{w_2} - \left\{\mu_\tau\left|w_2\right| + \sqrt{\lambda_\tau}W\left(w_2\right)\right\} &= \sup_r\left\{-\mu_\tau\left|\frac{\lambda_\tau}{4\mu_\tau^2}r\right| + \sqrt{\lambda_\tau}W\left(\frac{\lambda_\tau}{4\mu_\tau^2}r\right)\right\} \\ &= \frac{\lambda_\tau}{4\mu_\tau}\sup_r\left\{-|r| + 2W(r)\right\}. \end{aligned} \tag{6.13}$$

Consequently, we have $LR_{n\tau}\left(\gamma_\tau^0\right) \xrightarrow{d} \frac{\lambda_\tau}{4\mu_\tau}\sup_r\left\{-|r| + 2W(r)\right\}$. ■

Proof of Theorem 3.7

(i) First, observe that $\hat{\theta}_1(\tau, \gamma)$ defined in (3.9) is also minimizing $\bar{\mathcal{S}}_n(\theta_1, \tau, \gamma) \equiv n^{-1}\sum_{i=1}^n s\left(y_i, z_i; \theta_1, \tau, \gamma\right)$ with respect to θ_1, where $s\left(y_i, z_i; \theta_1, \tau, \gamma\right) \equiv \rho_\tau\left(y_i - \theta_1' z_i(\gamma)\right) - \rho_\tau\left(y_i - \theta_{1\tau}^{0\prime} z_i\left(\gamma^0\right)\right)$. Let $\varsigma(\theta_1, \gamma, \tau) \equiv E[s\left(y_i, z_i; \theta_1, \tau, \gamma\right)]$. By Lemma 6.1, $\varsigma(\theta_1, \tau, \gamma)$ is continuous in (θ_1, γ). It is straightfoward to show that it is also continuous in τ. Thus $\varsigma(\theta_1, \tau, \gamma)$ is continuous over $\Theta_1 \times \mathcal{T} \times \Gamma$. By Lemma 2 in Galvao et al. (2011), $\{s(y, z; \theta_1, \gamma, \tau) : (\theta_1, \tau, \gamma) \in \Theta_1 \times \mathcal{T} \times \Gamma\}$ is Glivenko-Cantelli. In conjunction with the pointwise convergence, this implies that

$$\sup_{(\theta_1, \tau, \gamma) \in \Theta_1 \times \mathcal{T} \times \Gamma}\left|\bar{\mathcal{S}}_n(\theta_1, \tau, \gamma) - \varsigma(\theta_1, \tau, \gamma)\right| \xrightarrow{p} 0.$$

As remarked after Assumption A9, $\theta_1^0(\tau, \gamma) = \arg\min_{\theta_1 \in \Theta_1}\varsigma(\theta_1, \tau, \gamma)$ is uniquely defined. It follows from Lemma B.1 of Chernozhukov and Hansen (2006) that

$$\sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma}\left\|\hat{\theta}_1(\tau, \gamma) - \theta_1^0(\tau, \gamma)\right\| = o_P(1). \tag{6.14}$$

[18] It is easy to see that Hansen's proofs of his Lemmas A.10-A.11 break down in the case $a = 0$.

Let $D_n(\gamma) \equiv \hat{S}_{n\pi}(\gamma) - \sum_{i=1}^{n} \int \rho_\tau \left(y_i - \theta_{1\tau}^{0\prime} z_i \left(\gamma^0 \right) \right) d\Pi(\tau)$. (6.14) implies that uniformly over Γ

$$
\begin{aligned}
n^{-1} D_n(\gamma) &= n^{-1} \sum_{i=1}^{n} \int s\left(y_i, z_i; \hat{\theta}_1(\tau, \gamma), \tau, \gamma \right) d\Pi(\tau) \\
&= n^{-1} \sum_{i=1}^{n} \int \left[\rho_\tau \left(y_i - \hat{\theta}_1(\tau, \gamma)' z_i(\gamma) \right) - \rho_\tau \left(y_i - \theta_{1\tau}^{0\prime} z_i \left(\gamma^0 \right) \right) \right] d\Pi(\tau) \\
&= n^{-1} \sum_{i=1}^{n} \int \left[\rho_\tau \left(y_i - \theta_1^0(\tau, \gamma)' z_i(\gamma) \right) - \rho_\tau \left(y_i - \theta_{1\tau}^{0\prime} z_i \left(\gamma^0 \right) \right) \right] d\Pi(\tau) + o_P(1) \\
&= D(\gamma) + o_P(1)
\end{aligned}
$$

where $D(\gamma) = \int E\left[\rho_\tau \left(\varepsilon_{i\tau} - \left(\theta_1^0(\tau, \gamma) z_i(\gamma) - \theta_{1\tau}^{0\prime} z_i \left(\gamma^0 \right) \right) \right) - \rho_\tau(\varepsilon_{i\tau}) \right] d\Pi(\tau)$ and $\varepsilon_{i\tau} = y_i - \theta_{1\tau}^{0\prime} z_i \left(\gamma^0 \right)$. Again, by the fact that $E\left[\rho_\tau(\varepsilon_{i\tau} - c) - \rho_\tau(\varepsilon_{i\tau}) \right] > 0$ for all $c \neq 0$, $D(\gamma)$ is minimized iff $\theta_1^0(\tau, \gamma) z_i(\gamma) = \theta_{1\tau}^{0\prime} z_i \left(\gamma^0 \right)$ a.s., i.e., iff $\gamma = \gamma^0$ by Assumption A9(ii). By invoking Lemma B.1 of Chernozhukov and Hansen (2006) again, we have $\hat{\gamma} = \gamma^0 + o_P(1)$. This, in conjunction with (6.14) and the continuity of $\theta_1^0(\tau, \cdot)$, implies that $\hat{\theta}_1(\tau) = \hat{\theta}_1(\tau, \hat{\gamma}) = \theta_1^0(\tau, \hat{\gamma}) + o_P(1) = \theta_{1\tau}^0 + o_P(1)$.

(ii) By the computational property of quantile regression (e.g., Lemma A2 in Ruppert and Carroll (1980)), uniformly in (τ, γ) $o_P(1) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau \left(y_i - \hat{\theta}_1(\tau, \gamma)' z_i(\gamma) \right) z_i(\gamma)$. For $u = (y, z')'$, define the map $f \longmapsto \mathbb{G}_n f(u) \equiv n^{-1/2} \sum_{i=1}^{n} \{ f(u_i) - E[f(u_i)] \}$ for any measurable function f. Let $f_1(u; \theta_1, \tau, \gamma) = \psi_\tau(y_i - \theta_1{}' z(\gamma)) z(\gamma)$. Noting that $\{ f_1(u; \theta_1, \tau, \gamma) : (\theta_1, \tau, \gamma) \in \Theta_1 \times \mathcal{T} \times \Gamma \}$ is Euclidean and

$$
\begin{aligned}
E\left[f_1\left(u_i;\ \theta_1(\tau, \gamma), z_i(\gamma), \gamma \right) \right] &= E\left[\psi_\tau \left(y_i - \theta_1(\tau, \gamma)' z_i(\gamma) \right) z_i(\gamma) \right] \\
&= E\left\{ \left[F\left(\theta_1^0(\tau, \gamma)' z_i(\gamma) | z_i \right) - F\left(\theta_1(\tau, \gamma)' z_i(\gamma) | z_i \right) \right] z_i(\gamma) \right\} \\
&\to -E\left[\bar{\Omega}_1(\tau, \gamma) \right] \left(\theta_1(\tau, \gamma) - \theta_1^0(\tau, \gamma) \right)
\end{aligned}
$$

when $\theta_1(\tau, \gamma) \to \theta_1^0(\tau, \gamma)$ uniformly in (τ, γ), we have

$$
\begin{aligned}
o_P(1) &= n^{-1/2} \sum_{i=1}^{n} \psi_\tau \left(y_i - \hat{\theta}_1(\tau, \gamma)' z_i(\gamma) \right) z_i(\gamma) \\
&= \mathbb{G}_n f_1 \left(u_i;\ \hat{\theta}_1(\tau, \gamma), \tau, \gamma \right) + \sqrt{n} E\left[f_1(u_i;\ \theta_1, \tau, \gamma) \right] |_{\theta_1 = \hat{\theta}_1(\tau, \gamma)} \\
&= \mathbb{G}_n f_1 \left(u_i;\ \theta_1^0(\tau, \gamma), \tau, \gamma \right) + o_P(1) + \sqrt{n} E\left[f_1(u_i;\ \theta_1, \tau, \gamma) \right] |_{\theta_1 = \hat{\theta}_1(\tau, \gamma)} \\
&= \mathbb{G}_n f_1 \left(u_i;\ \theta_1^0(\tau, \gamma), \tau, \gamma \right) + o_P(1) - \Omega_1(\tau, \gamma) \sqrt{n} \left\{ \hat{\theta}_1(\tau, \gamma) - \theta_1^0(\tau, \gamma) \right\}.
\end{aligned}
$$

Thus we have the following uniform Bahadur representation

$$
\sqrt{n} \left(\hat{\theta}_1(\tau, \gamma) - \theta_1^0(\tau, \gamma) \right) = \bar{\Omega}_1(\tau, \gamma)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau \left(y_i - \theta_1^0(\tau, \gamma)' z_i(\gamma) \right) z_i(\gamma) + o_P(1), \tag{6.15}
$$

where we have used the fact $E[\psi_\tau(y_i - \theta_1^0(\tau, \gamma)' z_i(\gamma)) | z_i] = 0$ by (3.18) and the last $o_P(1)$ holds uniformly over $\mathcal{T} \times \Gamma$.

(iii) Let $w_1(\tau) \equiv \sqrt{n}\left(\theta_{1\tau} - \theta_{1\tau}^0\right)$, $w_2 \equiv n^{1-2a}\left(\gamma - \gamma^0\right)$, $w_{1n}(\tau) \equiv \sqrt{n}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0)$ and $w_{2n} \equiv n^{1-2a}\left(\hat{\gamma} - \gamma^0\right)$. Let $\{\theta_{1\tau}\}$ denote $\{\theta_{1\tau}\}_{\tau\in\mathcal{T}}$, and similarly for $\{\hat{\theta}_{1\tau}\}$. As in the proof of Theorems 3.4 and 3.5, we continue to use the decomposition for $D_{n\tau}(\theta_{1\tau}, \gamma)$ in (6.4), where the only difference is that γ and γ^0 are now τ-invariant. Noting that $\left(\{\hat{\theta}_{1\tau}\}, \hat{\gamma}\right) = \text{argmin}_{\{\theta_{1\tau}\},\gamma} \int D_{n\tau}(\theta_{1\tau}, \gamma)\, d\Pi(\tau)$, we have

$$(w_{1n}(\cdot), w_{2n}) = \underset{w_1(\cdot),\ w_2}{\text{argmin}} \left[\mathcal{D}_{1n}(w_1) + \bar{\mathcal{D}}_{2n}(w_1, w_2)\right].$$

where $\mathcal{D}_{1n}(w_1) = \int D_{1n\tau}(w_1(\tau))\, d\Pi(\tau)$ and $\bar{\mathcal{D}}_{2n}(w_1, w_2) = \int \bar{D}_{2n\tau}(w_1(\tau), w_2)\, d\Pi(\tau)$. It is easy to see that the results in Lemmata 6.2 and 6.3 can be strengthened to hold uniformly in (τ, w_1) over any compact set. It follows that

$$\mathcal{D}_{1n}(w_1) = -n^{-1/2} \int w_1(\tau)' \sum_{i=1}^{n} z_i\left(\gamma^0\right) \psi_\tau(\varepsilon_{i\tau})\, d\Pi(\tau) + \frac{1}{2} \int w_1(\tau)' \Omega_{1\tau} w_1(\tau)\, d\Pi(\tau) + o_P(1), \quad (6.16)$$

and

$$\bar{\mathcal{D}}_{2n}(w_1, w_2) = \bar{\mathcal{D}}_{2n}(0, w_2) + o_P(1) = \int \sum_{i=1}^{n} \left[\rho_\tau\left(\varepsilon_{i\tau} + \delta_\tau^{0\prime} \Delta x_i(w_2)\right) - \rho_\tau(\varepsilon_{i\tau})\right] d\Pi(\tau) + o_P(1). \quad (6.17)$$

Then the rest of the proof follows directly from the proof of Theorem 3.5 with obvious modification. (iv) follows by the straightforward covariance calculations. ■

Proof of Theorem 3.8.

(i) Let $w_{2n} \equiv n^{1-2a}\left(\hat{\gamma} - \gamma^0\right)$. Following closely the proof of Theorem 3.6, we obtain

$$LR_{n1}\left(\gamma^0\right) = -\int \bar{D}_{2n\tau}(0, w_{2n})\, d\Pi(\tau) + o_P(1) = -\bar{\mathcal{D}}_{2n}(0, w_{2n}) + o_P(1).$$

The result then follows by arguments analogous to those used in the proof of Theorem 3.6.

(ii) Let $\bar{w}_{1n}(\tau) = n^{1/2}\left(\hat{\theta}_1\left(\tau, \gamma^0\right) - \theta_1^0\left(\tau, \gamma^0\right)\right)$, $\tilde{w}_{1n}(\tau) = n^{1/2}\left(\hat{\theta}_1(\tau, \hat{\gamma}_\tau) - \theta_1^0\left(\tau, \gamma^0\right)\right)$, and $w_{2n}(\tau) = n^{1-2a}\left(\hat{\gamma}_\tau - \gamma^0\right)$. By Theorems 3.7 and 3.5, $\bar{w}_{1n}(\tau) = O_P(1)$, $\tilde{w}_{1n}(\tau) = O_p(1)$, and $w_{2n}(\tau) = O_p(1)$. Then by (6.4), (6.16), and (6.17), we have

$$\begin{aligned}
LR_{n2}\left(\gamma^0\right) &= \int \left[S_{n\tau}\left(\hat{\theta}_1\left(\tau, \gamma^0\right), \gamma^0\right) - S_{n\tau}\left(\hat{\theta}_1\left(\tau, \gamma^0\right), \hat{\gamma}_\tau\right)\right] d\Pi(\tau) \\
&\quad + \int \left[S_{n\tau}\left(\hat{\theta}_1\left(\tau, \gamma^0\right), \hat{\gamma}_\tau\right) - S_{n\tau}\left(\hat{\theta}_1(\tau, \hat{\gamma}_\tau), \hat{\gamma}_\tau\right)\right] d\Pi(\tau) \\
&= \int \left[\bar{D}_{2n\tau}(\bar{w}_{1n}(\tau), 0) - \bar{D}_{2n\tau}(\bar{w}_{1n}(\tau), w_{2n}(\tau))\right] d\Pi(\tau) \\
&\quad + \int [\bar{D}_{1n\tau}(\bar{w}_{1n}(\tau)) - \bar{D}_{1n\tau}(\tilde{w}_{1n}(\tau)) + \bar{D}_{2n\tau}(\bar{w}_{1n}(\tau), w_{2n}(\tau)) \\
&\qquad\qquad - \bar{D}_{2n\tau}(\tilde{w}_{1n}(\tau), w_{2n}(\tau))] d\Pi(\tau) \\
&= -\int \bar{D}_{2n\tau}(0, w_{2n}(\tau))\, d\Pi(\tau) + \int \left[\bar{D}_{1n\tau}(\bar{w}_{1n}(\tau)) - \bar{D}_{1n\tau}(\tilde{w}_{1n}(\tau))\right] d\Pi(\tau) + o_P(1) \\
&\equiv LR_{n2,1} + LR_{n2,2} + o_P(1), \text{ say.}
\end{aligned}$$

By (6.12), (6.13), and the CMT, under H_0

$$LR_{n2,1} = \int LR_{n\tau}\left(\gamma^0\right) d\Pi(\tau) + o_P(1) \xrightarrow{d} \int \frac{\lambda_\tau}{2\mu_\tau} d\Pi(\tau) \sup_r \left\{-\frac{1}{2}|r| + W(r)\right\}.$$

By Theorem 3.7(ii), $\bar{w}_{1n}(\tau) \Rightarrow \Omega_{1\tau}^{-1} W\left(\tau, \gamma^0\right)$. Then by the fact that $n^{-1/2}\sum_{i=1}^n z_i\left(\gamma^0\right)\psi_\tau\left(\varepsilon_{i\tau}\right) \Rightarrow W\left(\tau,\gamma^0\right)$, (6.16) and the CMT, we have

$$\begin{aligned}
&\int \bar{D}_{1n\tau}\left(\bar{w}_{1n}(\tau)\right) d\Pi(\tau) \\
&= -n^{-1/2}\int \bar{w}_{1n}(\tau)' \sum_{i=1}^n z_i\left(\gamma^0\right)\psi_\tau\left(\varepsilon_{i\tau}\right) d\Pi(\tau) + \frac{1}{2}\int \bar{w}_{1n}(\tau)'\,\Omega_{1\tau}\bar{w}_{1n}(\tau)\, d\Pi(\tau) + o_P(1) \\
&\xrightarrow{d} -\frac{1}{2}\int W\left(\tau,\gamma^0\right)'\,\Omega_{1\tau}^{-1} W\left(\tau,\gamma^0\right) d\Pi(\tau).
\end{aligned}$$

The same result holds for $\int \bar{D}_{1n\tau}\left(\tilde{w}_{1n}(\tau)\right) d\Pi(\tau)$. As a consequence $LR_{n2,2} \xrightarrow{d} 0$, implying that $LR_{n2,2} = o_P(1)$. It follows that $LR_{n2}\left(\gamma^0\right) \xrightarrow{d} \int \frac{\lambda_\tau}{4\mu_\tau} d\Pi(\tau) \sup_r \{-|r| + 2W(r)\}$. ■

Acknowledgement. We sincerely thank Jacques Crémer, Han Hong, Thomas Lemieux, Michael Peters, and Wing Suen for their many useful comments. We are also grateful to the participants in the seminars at Peking University, the Shanghai Workshop on Industrial Organization and Competition Policy, and the 45th annual conference of the Canadian Economic Association. Pai Xu gratefully acknowledges the financial support from the University of Hong Kong under the seed funding scheme.

REFERENCES

Andrews, D. W. K. (1994), Empirical Process Methods in Econometrics, in R. F. Engle and D. L. McFadden (eds), Handbook of Econometrics IV, 2248-2294.

Bai, J. (1995), Least Absolute Deviation Estimation of a Shift. Econometric Theory 11, 403-436.

Bajari, P., H. Hong, M. Park, and R. Town (2010), Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health Care. Working paper, Department of Economics, University of Minnesota.

Bernstein, D. S. (2005), Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory. Princeton University Press, Princeton, NJ.

Bolton, G. E., E. Katok, and A. Ockenfels (2004), How Effective are Online Reputation Mechanisms? An Experimental Study. Management Science 50(11), 1587-1602.

Cabral, L. and A. Hortacsu (2010), The Dynamics of Seller Reputation: Evidence from eBay. Journal of Industrial Economics LVIII, 54-78.

Caner, M. (2002), A Note on Least Absolute Deviation Estimation of a Threshold Model. Econometric Theory 18, 800-814.

Chernozhukov, V. and C. Hansen (2006), Instrumental Quantile Regression Inference for Structural and Treatment Effect Models. Journal of Econometrics 127, 491-525.

Dellarocas, C. and C. A. Wood (2008), The Sound of Silence in Online Feedback: Estimating Trading Risks in the Presence of Reporting Bias. Management Science 54(3), 460-476.

Fan J., Q. Yao, and H. Tong (1996), Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems. Biometrika 83, 189-206.

Galvao, A. F., K. Kato, G. Montes-Rojas, and J. Olmo (2010), A Uniform Test for Linearity in Quantile Regression. Working paper, Department of Economics, University of Iowa.

Galvao, A. F., G. Montes-Rojas, and J. Olmo (2011), Threshold Quantile Autoregression Models. Journal of Time Series Analysis, forthcoming.

Hansen, B. E. (2000), Sample Splitting and Threshold Estimation. Econometrica 68, 575-605.

Imbens, G. W. and T. Lemieux (2008), Regression Discontinuity Designs: A Guide to Practice. Journal of Econometrics 142(2), 615-635.

Knight, K. (1998), Limiting Distributions for L_1 Regression Estimators under General Conditions. Annals of Statistics 26, 755-770.

Koenker, R. (2005), Quantile Regression. Cambridge University Press, Cambridge.

Koenker, R. and W. G. Bassett (1978), Regression Quantiles. Econometrica 46, 33-50.

Koul, H. L., L. Qian, and D. Surgailis (2003), Asymptotics of M-estimators in Two-phase Linear Regression Models. Stochastic Processes and their Applications 103, 123-154.

McCrary J. (2008), Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. Journal of Econometrics 142(2), 698-714.

Oka, T. (2010), Nonparametric Quantile Regression with Discontinuities. Working paper, Dept. of Economics, Boston University.

Pakes, A. and D. Pollard (1989), Simulation and the Asymptotics of Optimization Estimators. Econometrica 57, 1027-1057.

Qu, Z. (2008), Testing for Structural Change in Regression Quantiles. Journal of Econometrics 146, 170-184.

Resnick, P., R. Zeckhauser, J. Swanson, and K. Lockwood (2006), The Value of Reputation on eBay: A Controlled Experiment. Experimental Economics 9(2), 79-101.

Ruppert, D. and R. J. Carroll (1980), Trimmed Least Squares Estimation in the Linear Model. Journal of the American Statistical Association 75, 828-838.

Su, L. and Z. Xiao (2008), Testing for Parameter Stability in Quantile Regression Models. Statistics & Probability Letters 78, 2768-2775.

Tirole J. (1988), The Theory of Industrial Organization. MIT Press, Cambridge, MA.

Van der Vaart, W. and J. A. Wellner (1996), Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York.