Common Threshold in Quantile Regressions
with an Application to Pricing for Reputation

Liangjun Su\textsuperscript{a}, Pai Xu\textsuperscript{b}
\textsuperscript{a} Singapore Management University, Singapore
\textsuperscript{b} University of Hong Kong, Hong Kong

December 20, 2016

Abstract

The paper develops a systematic estimation and inference procedure for quantile regression models where there may exist a common threshold effect across different quantile indices. We first propose a sup-Wald test for the existence of a threshold effect, and then study the asymptotic properties of the estimators in a threshold quantile regression model under the shrinking-threshold-effect framework. We consider several tests for the presence of a common threshold value across different quantile indices and obtain their limiting distributions. We apply our methodology to study the pricing strategy for reputation via the use of a dataset from Taobao.com. In our economic model, 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, which is considered as “reputation effect” in this paper. The use of threshold quantile regression model allows us to identify and explore the reputation effect and its heterogeneity in data. We find both reputation effects and common thresholds for a range of quantile indices in seller’s pricing strategy in our application.

\textbf{JEL Classifications:} L10; C12; C13

\textbf{Key Words:} Common threshold effect; Pricing strategy; Regime change; Specification test; Threshold quantile regression
1 Introduction

Since Tong (1978, 1983) threshold models have become very popular in econometrics and statistics. Early literature focuses on the modeling of the conditional mean in time series context. See, for example, Chan (1993) and Hansen (2000) on the asymptotic distribution theory for the threshold estimator in the fixed-threshold-effect and shrinking-threshold-effect frameworks, respectively, and Tong (2011) and Hansen (2011) for reviews on the development and applications of the threshold regression models in statistics and economics. Robust estimation of the threshold parameter has not been analyzed in the literature until Caner (2002) who derives the asymptotic distribution of the LAD estimator of the threshold parameter. Kato (2009) extends the convexity arguments of Pollard (1991) to the case where estimators are obtained as stochastic processes and applies this technique to study the inferential problems for the LAD estimator in threshold models. Cai and Stander (2008), Cai (2010), and Galvao et al. (2011) study the asymptotic properties of the parameter estimators in threshold quantile autoregressive models. Yu (2012) studies likelihood-based estimation and inference in threshold regression models.

This paper studies the estimation and inference in threshold quantile regression (TQR) models when it is uncertain whether there is a threshold effect at any quantile index, and if there is any effect, whether or not the threshold point depends on the quantile index. Early literature on TQR models assumes the existence of a threshold first and studies the asymptotic properties for the estimators of both regression coefficients and threshold parameters. Examples include but are not limited to Caner (2002), Kato (2009), Cai and Stander (2008), Cai (2010), and Galvao et al. (2011). However, it is only until Lee et al. (2011) that the test for the existence of threshold effect has been developed. They propose a general likelihood-ratio-based method for testing threshold effects in regression models that include the quantile regression as a special case. But they can only test whether the threshold effect exists for a single quantile index. More recently, Yu (2014) considers the estimation and testing in TQR by assuming that the threshold parameter is invariant to the quantile index \( \tau \). As expected, his estimator is more efficient than some of the existing estimators (e.g., LAD or least squares estimators) and his test is more powerful than the conventional tests based solely on the LAD or least squares estimators if the underlying assumption of common threshold parameter across different quantile indices holds true. On the other hand, an estimation and testing procedure of this kind may be invalid if the above underlying assumption is violated. Therefore, it is important to consider an inferential procedure which does not rely on either one of the following two assumptions: (i) there exists a threshold effect in the quantile regression, and (ii) the threshold parameter is invariant to the quantile index.

This paper thus develops a systematic procedure on estimation and inference of quantile regression models in the absence of either assumptions mentioned above. First, similar to Galvao et al. (2014) who develop a test of linearity against threshold effects in the quantile regression framework, we propose a sup-Wald test for the absence of a threshold effect. The supremum is taken over a compact subset \( T \) in \((0,1)\) where the quantile index lies. In principle, our test has power as long as the threshold effect exists for some quantile index \( \tau \in T \). It is possible that the threshold effect does not exist for \( \tau \) in a subset \( T_1 \) of \( T \), but it is present for \( \tau \) that lies in the complement of \( T_1 \) relative to \( T \). In this case, the single quantile-index-based test of Lee et al. (2011) loses power if the quantile index \( \tau \) happens to lie in \( T_1 \).
When we motivate the TQR model from the random coefficient perspective, we can easily construct a data generating process (DGP) where the threshold effect exists only for quantile indices in a subset of $(0, 1)$. This phenomenon is also confirmed in our empirical application.

Second, we study the asymptotic properties of the estimators of both the threshold parameter and regression coefficients. In the TQR framework, it is typically assumed that the regression coefficients are dependent on the quantile index and composite quantile regression should be called upon otherwise. Then a question arises - whether we should allow the threshold parameter to depend on the quantile index. Previous works, such as Caner (2002) and Kato (2009) consider quantile regressions at a fixed quantile index and their estimators may be inefficient if the threshold parameter is invariant to the quantile index. In contrast, Yu (2014) assumes that the threshold parameter is invariant to the quantile index and considers efficient estimation under such an assumption. Nevertheless, his estimation and inference can be misleading if the common threshold specification is incorrect. In fact, such an assumption can be easily violated in a class of random coefficient models that motivates the TQR; see Case 3 in our simulation example. Therefore we take a precautionary step and propose to estimate both the regression coefficients and the threshold parameters separately for each quantile index of interest. We establish the asymptotic distribution theory for both estimators by following the shrinking-threshold-effect framework of Hansen (2000). In addition, we also propose a likelihood-ratio (LR) test for the threshold parameter. Admittedly, our procedure avoids the potential issue of misspecification but sacrifices on efficiency in the case of a common threshold value.

Thirdly and most importantly, we study the estimation and inference theory associated with a common threshold value. We first study the asymptotic properties of the estimators of the regression coefficients and the threshold parameter in the presence of a common threshold value. Then we consider tests for the presence of a common threshold value. Here, we are interested in testing whether the quantile-index-dependent threshold parameter $\gamma_{\tau}$ is the same for all $\tau \in \mathcal{T}$. We propose two LR-type tests, one is for the case where the common threshold value $\gamma^0$ is unobserved and estimated from data, and the other is the case where $\gamma^0$ is observed. The first case works when it is unclear whether a common threshold parameter value should be shared across different quantile indices, and when one has no idea about the potential value of the common threshold parameter. The rejection of the null in this case implies the absence of a common threshold value and therefore misspecification of the model as in Yu (2014). On the other hand, when we fail to reject the null, one can follow Yu (2014) and consider efficient inference under the assumption of a common threshold value. The second case works when it is strongly believed that the quantile regressions for different quantile indices share the common threshold value (say based on the result of the first LR test here) and one wants to test whether the true common threshold value is given by a hypothesized value $\gamma^0$. The RDD setup may be one example of the second case. The second test statistic can be inverted to obtain the confidence interval for the common threshold parameter too.

For the empirical application, we apply our proposed estimation and inference procedure to study the existence of a particular pricing pattern arising from reputation concerns in online sales. Empirical studies have documented extensive evidence to show that sellers enjoy greater benefits from a better reputation. See, e.g., 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. Our empirical application therefore focuses on investigating whether and
how a selling strategy may be affected by a reputation system in use.

We collected trading data from Taobao.com, the leading online shopping website in China. There 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. To explain such an interesting pattern in pricing, we construct an 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 demonstrate that there exists a threshold in reputation, starting from which the sellers engage in price-cuts in exchange for the rewards of better reputation. Thus our model predicts pricing-regime changes and accordingly discontinuity (or, more specifically a “jump”) in optimal pricing behavior. In turn, examining the relevance of theory prediction arising from reputation concerns amounts to testing for existence of threshold. In view of the presence of heterogeneous sellers in the market, we recognize that high-end sellers may adopt different pricing strategies than middle- and low-end sellers. Cabral and Hortacsu (2010) acknowledge the existence of significant unobservable seller heterogeneity in the electronic market. Such heterogeneity, indeed, motivates us to adopt a TQR model 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.

Beyond the difference, however, it is also fair to ask whether at all, or to what extent, there exists homogeneous pricing behavior across sellers. Particularly, in our application, it is certainly critical to know how common the sellers may share the same cut-off strategy in terms of these regime changes. This question instead motivates us to consider studying the case of common threshold. Our empirical results indicate that sellers at different quantiles (of prices) exhibit quite different pricing behavior, yet some of them do employ the same pricing for reputation strategy predicted in the model. Thus, we believe our methodology enriches the empirical analysis on dealing with heterogeneity by identifying the existence of a subset of homogeneous agents.

It is worth mentioning that, from the application perspective, this paper also subtly enriches the empirical literature on the RDD. In a typical RDD framework, researchers are interested in the causal effect of a binary intervention or treatment. This design arises frequently in the study 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/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. See Imbens and Lemieux (2008) for a detailed survey on the empirical applications. At the heart of identifying assumptions to validate the RDD framework, the covariate is connected with the potential outcomes in a continuous 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) points 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 proposes a test for the discontinuity at the cutoff in the density function of the covariate. 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).
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. In Section 2 we introduce the estimation and inferences in a TQR model where we propose three types of tests: one for testing the existence of a threshold effect, the second for testing the presence of a common threshold value, and the third for the inference of a common or non-common threshold parameter. The asymptotic properties of both the estimators and test statistics are reported. We conduct a sequence of Monte Carlo simulations in Section 3 to investigate the finite sample performance of our estimators and tests. We apply our methodology to study the pricing for reputation via the use of a dataset from Taobao.com in Section 4. Section 5 concludes. All technical assumptions and proofs of the main results are collected in the appendix.

2 Estimation and Inferences in Quantile Threshold Regression Models

In this section, we first introduce the TQR model and propose a sup-Wald test for the existence of a threshold effect. Then we study the asymptotic properties of the estimators of both the regression coefficients and the threshold parameter for a fixed quantile index $\tau$ and consider the inference issue associated with the threshold parameter. Lastly and most importantly, we study the estimation and inference issues associated with the presence of a common threshold effect.

2.1 A 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 be exogenous with a continuous probability density function (PDF) $g(\cdot)$. Let $z_i \equiv (x'_i, r_i)'$ if $r_i/ \in x_i$, and $z_i \equiv x_i$ otherwise. We assume that the $\tau$th conditional quantile of $y_i$, given $z_i$, is given by

$$Q_\tau(z_i) = \alpha'_\tau x_i 1\{r_i \leq \gamma_\tau\} + \beta'_\tau x_i 1\{r_i > \gamma_\tau\}, \quad (2.1)$$

where $1\{A\}$ is an indicator function that takes value one if $A$ holds true, and zero otherwise; and $\delta_\tau \equiv \alpha_\tau - \beta_\tau$ may be nonzero for some unknown threshold point $\gamma_\tau$. If $\delta_\tau$ is 0 for all $\gamma_\tau$ on the support of $r_i$ and for all $\tau \in (0, 1)$, then we can say there is no regime change in the quantile regression model (2.1). For technical simplicity, below we assume that $\gamma_\tau$ can only take values in a compact set $\Gamma \equiv [\underline{\gamma}, \bar{\gamma}]$.

Let $\theta_{1,\tau} \equiv (\alpha'_\tau, \beta'_\tau)'$ and $\theta_{\tau} \equiv (\theta'_{1,\tau}, \gamma_\tau)'$. Define the “check function” $\rho_\tau(\cdot)$ by $\rho_\tau(u) \equiv (\tau - 1\{u < 0\}) u$. Following Koenker and Bassett (1978), we obtain the quantile estimate $\hat{\theta}_\tau$ of $\theta_\tau$ as

$$\hat{\theta}_\tau = \operatorname{argmin}_{\theta_\tau} S_{n,\tau}(\theta_\tau) \quad \text{with} \quad S_{n,\tau}(\theta_\tau) = \sum_{i=1}^n \rho_\tau(y_i - \theta'_{1,\tau} z_i(\gamma_\tau)), \quad (2.2)$$

where $z_i(\gamma) \equiv (x'_i 1\{r_i \leq \gamma\}, x'_i 1\{r_i > \gamma\})'$.

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 \( \gamma_\tau \) is known, and obtain an estimate of \((\alpha_\tau, \beta_\tau)\) by

\[
\left( \hat{\alpha}_\tau (\gamma_\tau), \hat{\beta}_\tau (\gamma_\tau) \right) \equiv \left( \hat{\alpha} (\tau, \gamma_\tau), \hat{\beta} (\tau, \gamma_\tau) \right) = \arg \min_{\alpha_\tau, \beta_\tau} S_{n\tau} (\alpha_\tau, \beta_\tau, \gamma_\tau),
\]

(2.3)

where \( S_{n\tau} \) is convex in its first two arguments. Let \( \hat{S}_{n\tau} (\gamma) \equiv S_{n\tau} (\hat{\alpha}_\tau (\gamma), \hat{\beta}_\tau (\gamma), \gamma) \). Then we can estimate \( \gamma_\tau \) by

\[
\hat{\gamma}_\tau = \arg \min_{\gamma} \hat{S}_{n\tau} (\gamma).
\]

(2.4)

In view of the fact that \( \hat{S}_{n\tau} (\gamma) \) takes on less than \( n \) distinct values, we follow Hansen (2000, p.578) and define \( \hat{\gamma}_\tau \) by choosing \( \hat{\gamma}_\tau \) over \( \Gamma_n = \Gamma \cap \{ r_1, r_2, \cdots, r_n \} \). Then, computing \( \hat{\gamma}_\tau \) requires at most \( n \) function evaluations.\(^1\) However, if \( n \) is large, then we can approximate \( \Gamma \) by a grid as suggested by Hansen (2000). After \( \hat{\gamma}_\tau \) is obtained, we can compute the estimates of \( \alpha_\tau \) and \( \beta_\tau \) as \( \hat{\alpha}_\tau = \hat{\alpha}_\tau (\hat{\gamma}_\tau) \) and \( \hat{\beta}_\tau = \hat{\beta}_\tau (\hat{\gamma}_\tau) \), respectively.

### 2.2 Test the existence of a change point

The preceding computation procedure is meaningful only if \( \gamma_\tau \) is identified, in which case a regime change occurs for the \( \tau \)th conditional quantile regression. It is thus worthwhile to consider a test for the existence of a regime change before embarking on the estimation of \( \gamma_\tau \).

Let \( T \equiv [\tau, \tau] \subset (0, 1) \) and \( \Theta_1 \subset \mathbb{R}^k \) denote the compact support for \( \theta_{1\tau} \). In principle, we allow the support of \( \theta_{1\tau} \) to be \( \tau \)-dependent, and write \( \Theta_1 \) as \( \Theta_{1\tau} \). We use \( \Theta_1 \) instead of \( \Theta_{1\tau} \) mainly for notational simplicity. Let \( \theta_{01\tau} \equiv (\alpha_{0\tau}, \beta_{0\tau}) \) denote the true value of \( \theta_{1\tau} \). It is fair to comment here that we allow \( \theta_{01\tau} \) to be \( n \)-dependent in our framework to facilitate the study of the estimate of \( \theta_{1\tau} \), even in the case where we have a regime change but the jump size shrinks to zero as the sample size \( n \to \infty \). But, for notational simplicity, we suppress the dependence of \( \theta_{01\tau} \), \( \alpha_{0\tau} \), and \( \beta_{0\tau} \) on \( n \) and write them as \( \theta_{1\tau}^0 \), \( \alpha_{0\tau}^0 \), and \( \beta_{0\tau}^0 \), respectively.

The null hypothesis of no regime change is

\[
H_0: Q_\tau (z_i) = z_i (\gamma)^\prime \theta_{1\tau}^0 \quad \forall \gamma \in \Gamma \text{ for some } \theta_{1\tau}^0 \in \Theta_1 \text{ with } \alpha_{0\tau}^0 = \beta_{0\tau}^0 \text{ for all } \tau \in T.
\]

(2.5)

The alternative hypothesis is

\[
H_1: Q_\tau (z_i) = z_i (\gamma)^\prime \theta_{1\tau}^0 \quad \text{for some } (\gamma_{0\tau}^0, \theta_{1\tau}^0) \in \Gamma \times \Theta_1 \text{ with } \alpha_{0\tau}^0 \neq \beta_{0\tau}^0 \text{ for some } \tau \in T.
\]

(2.6)

Clearly, both \( H_0 \) and \( H_1 \) are composite hypotheses, which are designed to test the existence of a regime change at an arbitrary quantile point \( \tau \). For different \( \tau \)’s, the regime changes are allowed to occur at different threshold values \( \gamma_{0\tau}^0 \) under \( H_1 \). If we restrict our attention to a single quantile \( \tau \), i.e., \( T = \{ \tau \} \), then we can consider the following null hypothesis.

\[
H_{0\tau}: Q_\tau (z_i) = z_i (\gamma)^\prime \theta_{1\tau}^0 \text{ for some } \theta_{1\tau}^0 \in \Theta_1 \text{ with } \alpha_{0\tau}^0 = \beta_{0\tau}^0,
\]

(2.7)

\(^1\)Let \( r_{(1)} \leq r_{(2)} \leq \cdots \leq r_{(n-1)} \leq r_{(n)} \) denote the order statistics for \( \{ r_i, i = 1, \ldots, n \} \). It is well known that the solution to the threshold quantile regression is not unique in the sense if \( \hat{\gamma}_\tau \in [r_{(q)}, r_{(q+1)}) \) for some \( q \) is the solution, then any value on \( [r_{(q)}, r_{(q+1)}) \) is also the solution. For this reason, Hansen (2000) recommended searching the solutions over \( \Gamma \cap \{ r_1, \ldots, r_n \} \). But some researchers (e.g., Yu 2012) suggested to search the solutions over \( \Gamma \cap \{ \frac{1}{2}(r_{(1)} + r_{(2)}), \ldots, \frac{1}{2}(r_{(n-1)} + r_{(n)}) \} \) as the solution.
regardless of the value of $\gamma \in \Gamma$, and the alternative hypothesis becomes
\[ H_{1T} : Q_T(z_i) = z_i (\gamma_0 \tau^0) \theta_1^0, \] for some $\theta_1^0 \in \Theta_1$ with $\alpha^0_\gamma \neq \beta^0_\gamma$ and $\gamma_0 \in \Gamma$. \hfill (2.8)

The above formulation motivates us to consider the following $\tau$th quantile regression of $y_i$ on $z_i(\gamma)$.
\[ \hat{\theta}_1(\tau,\gamma) = \left( \hat{\alpha}(\tau,\gamma)', \hat{\beta}(\tau,\gamma)' \right)' = \arg\min_{\theta_1} \sum_{i=1}^n \rho_\tau(y_i - \theta_1'z_i(\gamma)). \hfill (2.9) \]

Even though $\gamma$ is not identified under $H_0$ for any $\tau \in \mathcal{T}$, we can study the asymptotic property of $\hat{\theta}_1(\tau,\gamma)$ under $H_0$ and propose a test for the null hypothesis of no regime change for all $\tau \in \mathcal{T}$ by considering the asymptotic behavior of $\hat{\delta}(\tau,\gamma) \equiv \hat{\alpha}(\tau,\gamma) - \hat{\beta}(\tau,\gamma)$ over the compact set $\mathcal{T} \times \Gamma$.

Here we consider a sup-Wald statistic for testing $H_0$. Let $\Omega(\gamma) \equiv E[x_i'x_1 \{ r_i \leq \gamma \}]$, $\Omega^*(\gamma) \equiv E[x_i'x_1 \{ r_i > \gamma \}]$, $\Omega(\tau,\gamma) \equiv E[x_i'x_{1\tau} \{ r_i \leq \gamma \} f(\alpha^0_\gamma x_i z_i)]$, and $\Omega^*(\tau,\gamma) \equiv E[x_i'x_{1\tau} \{ r_i > \gamma \} f(\beta^0_\gamma x_i z_i)]$, where $f(\cdot | z)$ denotes the conditional PDF of $y_i$ given $z_i = z$. By Lemma B.1 in Appendix B, we can readily show that for each $(\tau,\gamma) \in \mathcal{T} \times \Gamma$,
\[ \sqrt{n} \hat{\delta}(\tau,\gamma) \xrightarrow{d} N(0_{k \times 1}, (1 - \tau)V(\tau,\gamma)), \hfill (2.10) \]
where $d$ denotes convergence in distribution, and
\[ V(\tau,\gamma) \equiv \Omega(\tau,\gamma)^{-1} \Omega(\gamma) \Omega(\tau,\gamma)^{-1} + \sum_{i=1}^n \rho_\tau(y_i - \theta_1'z_i(\gamma)). \hfill (2.11) \]

Let $\Omega(\gamma) \equiv n^{-1} \sum_{i=1}^n x_i'x_i \{ r_i \leq \gamma \}$, $\Omega^*(\gamma) \equiv n^{-1} \sum_{i=1}^n x_i'x_i - \hat{\Omega}(\gamma)$, $\hat{\Omega}(\tau,\gamma) \equiv n^{-1} \sum_{i=1}^n \{ y_i - x_i'\hat{\delta} \} h_i x_i'x_i \{ r_i \leq \gamma \}$, and $\hat{\Omega}^*(\tau,\gamma) \equiv n^{-1} \sum_{i=1}^n \{ y_i - x_i'\hat{\delta} \} h_i x_i'x_i \{ r_i > \gamma \}$, where $h \equiv h(n)$ is a bandwidth parameter such that $h \rightarrow 0$ and $nh^2 \rightarrow \infty$ as $n \rightarrow \infty$ (see Koenker, 2005, pp. 80-81). Using Lemma B.1 and following the proof of Theorem 3 in Powell (1991), one can readily 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
\[ \hat{V}(\tau,\gamma) \equiv \hat{\Omega}(\tau,\gamma)^{-1} \hat{\Omega}(\gamma) \hat{\Omega}(\tau,\gamma)^{-1} + \hat{\Omega}^*(\tau,\gamma)^{-1} \hat{\Omega}^*(\gamma) \hat{\Omega}^*(\tau,\gamma)^{-1}. \hfill (2.12) \]
Following Qu (2008) and Su and Xiao (2008) who consider testing for a permanent structural change in time series quantile regression models and Galvao et al. (2014) who consider testing linearity against threshold effects in quantile regression models, we propose a sup-Wald statistic for testing $H_0$ given by
\[ \sup_{\tau,\gamma} W_n \equiv \sup_{\tau,\gamma} W_n(\tau,\gamma), \hfill (2.13) \]
where $W_n(\tau,\gamma) = n \hat{\delta}(\tau,\gamma)' \{ 1 - \tau \} \hat{V}(\tau,\gamma)^{-1} \hat{\delta}(\tau,\gamma)$.

The following theorem provides the asymptotic distribution of $\sup W_n$.

\[ \textbf{Theorem 2.1} \] Let $R \equiv [I_k, -I_k]$ with $I_k$ being a $k \times k$ identity matrix, $\Omega_0(\gamma_1, \gamma_2) \equiv E[z_i(\gamma_1) z_i(\gamma_2)' \mid z_i(\gamma_1) z_i(\gamma_2)' \mid]$, and $\Omega_1(\tau,\gamma) \equiv E[f(\theta_1^0 x_i z_i)] z_i(\gamma_1) z_i(\gamma_2)'$. Suppose that Assumptions A1-A5 in Appendix A hold. Suppose that $h \rightarrow 0$ and $nh^2 \rightarrow \infty$ as $n \rightarrow \infty$. Then, under $H_0$,
\[ \sup W_n \xrightarrow{d} \sup_{\tau,\gamma} \frac{1}{\tau (1 - \tau)} W(\tau,\gamma) \Omega_1(\tau,\gamma)^{-1} R^\prime V(\tau,\gamma)^{-1} R \Omega_1(\tau,\gamma)^{-1} W(\tau,\gamma) \]
where $W(\tau,\gamma)$ is a zero-mean Gaussian process on $\mathcal{T} \times \Gamma$ with covariance kernel $E[W(\tau_1,\gamma_1) W(\tau_2,\gamma_2)']$
\[ = (\tau_1 \wedge \tau_2 - \tau_1 \tau_2) \Omega_0(\gamma_1, \gamma_2). \]
The preceding theorem shows that the limiting distribution of sup $W_n$ depends on the bi-parameter Gaussian process $W(\tau, \gamma)$. It is not pivotal and one cannot tabulate the critical values for the sup $W_n$ test. Nevertheless, given the simple structure of $W(\tau, \gamma)$, we can readily simulate the critical values for the sup $W_n$ test statistic. Observing that $\Omega_1(\tau, \gamma) = \left( \Omega(\tau, \gamma) \begin{pmatrix} 0_{k \times k} \\ 0_{k \times k} \end{pmatrix} \right)$, we can consistently estimate it by $\hat{\Omega}_1(\tau, \gamma) = \left( \hat{\Omega}(\tau, \gamma) \begin{pmatrix} 0_{k \times k} \\ 0_{k \times k} \end{pmatrix} \hat{\Omega}^*(\tau, \gamma) \right)$. Following Hansen (1996), we propose to simulate the critical values for the sup $W_n$ statistic with the following procedure:

1. Generate $\{u_i, i = 1, \ldots, n\}$ independently from the uniform distribution on $[0, 1]$;
2. Calculate $Z_n(\tau, \gamma) = n^{-1/2} \sum_{i=1}^n \left[ \tau - 1 \{ u_i \leq \tau \} \right] z_i(\gamma)$;
3. Compute $\sup W_n^* = \sup_{(\tau, \gamma) \in T \times \Gamma} \frac{1}{(1-\tau)\tau} Z_n(\tau, \gamma)' \hat{\Omega}_1(\tau, \gamma)^{-1} R' \hat{V}(\tau, \gamma)^{-1} \hat{R} \hat{\Omega}_1(\tau, \gamma)^{-1} Z_n(\tau, \gamma)$;
4. Repeat steps 1-3 $B$ times and denote the resulting sup $W_n^*$ test statistics as $sup_{n,j} W_n^*$ for $j = 1, \ldots, B$.
5. Calculate the simulated $p$-value for the sup $W_n$ test as $p^*_W(n) = \frac{1}{B} \sum_{j=1}^B 1\{ sup_{n,j} W_n^* \geq sup W_n \}$.

In practice, we compute the sup $W_n$ by constructing a finite partition $T_{m_1} \times T_{m_2} \subset T \times \Gamma$ by a finite grid of $m_1 \times m_2$ points. In our applications, we set $m_1 = m_2 = 81$ and choose $T_{81} = \{0.10, 0.11, \ldots, 0.90\}$ and $\Gamma_{81}$ as the collection of the $\tau$th quantile of $q_1$ for $\tau \in T_{81}$. To obtain the simulated $p$-value, one can choose a finer partition because of the fast speed of computing sup $W_n^*$.

One can readily show that $Z_n(\tau, \gamma) \Rightarrow W(\tau, \gamma)$ in $(C^\infty(T \times \Gamma))^2$, where $\Rightarrow$ denotes weak convergence and $C^\infty(T \times \Gamma)$ the space of all bounded functions on $T \times \Gamma$ equipped with the uniform topology. When $B$ is sufficiently large, the asymptotic critical value of the level $\alpha$ test based on sup $W_n$ is approximately given by the empirical upper $\alpha$-quantile of $\{ sup_{n,j} W_n^* \}$. Therefore, we can reject the null hypothesis $H_0$ if the simulated $p$-value $p^*_W$ is smaller than the prescribed nominal level of significance $\alpha$.

Note that by choosing $T$ as a large compact subset of $(0, 1)$, the above test can detect various violations of the null hypothesis. Alternatively, specifying $T = \{\tau\}$ allows us to consider the test of structural or regime change at a single quantile $\tau$. In the case where we reject $H_{0\tau}$ for the specified $\tau$, we can further consider estimating the location of the change point $\gamma_\tau$ under $H_{1\tau}$.

**Remark 1.** Using $\delta_\tau = \alpha_\tau - \beta_\tau$, we can rewrite (2.1) as

$$Q_\tau(z_i) = \alpha_\tau' x_i + \delta_\tau' x_i 1 \{ r_i > \gamma_\tau \}.$$  \hspace{1cm} (2.14)

If $\gamma_\tau$ is known, one can follow Koenker and Machado (1999) and consider several tests for the null hypothesis $H_0 : \delta_\tau^0 = 0 \forall \tau \in T$, which include a sup-Wald test. Our sup-Wald test can be regarded as an extension of theirs to allow for the presence of unknown threshold effect. It is also motivated from the literature on threshold regression; see, e.g., Hansen (1996) and Galvao et al. (2014) who only consider sup-Wald tests too. In principle, any other smooth functional can be used in place of our sup functional and our Wald statistic can be replaced by the LM or LR statistic. Let $T_n(\tau, \gamma)$ denote any one of these three statistics indexed by $(\tau, \gamma)$. A general test statistic can be written as a function $g(T_n)$, where $g(\cdot)$ maps functionals on $T \times \Gamma$ to the real line $\mathbb{R}$ and we treat $T_n = \{ T_n(\tau, \gamma) : (\tau, \gamma) \in T \times \Gamma \}$ as a random function on $T \times \Gamma$. 

8
A basic requirement on \( g \) is that it is monotonic and continuous. The continuity is required for us to apply the continuous mapping theorem. For brevity and clarity, we limit our attention to the sup functional and focus on the Wald statistic only in this paper.

**Remark 2.** Our test for the absence of threshold effect can be put in the general framework of hypothesis tests when a nuisance parameter (\( \gamma \) here) is present only under the alternative. The statistics literature goes back to at least Davies (1977, 1987). In econometrics, Andrews and Ploberger (1994, AP hereafter) and Hansen (1996) have made some fundamental contributions. AP focus on the asymptotic optimality issue. Hansen (1996) proposes a heteroskedasticity-robust sup-Wald test and suggests a useful simulation method to obtain the simulating critical values. But he does not discuss optimal choice of tests. Given AP, it is natural to ask consider an asymptotically optimal test for the absence of threshold effect, say, by proposing an average exponential LM, LR, or Wald test that maximizes the weighted average local power among a class of tests. We find that the extension of AP’s optimal test to the current framework does not appear appealing mainly for the following two reasons. First, as emphasized by AP, their optimality results only apply in correctly specified maximum likelihood contexts (see Remark 3 on p.1394 in AP). The correct specification of the likelihood function is crucial for their optimality results but is generally not needed for the usual sup-Wald-type test. In our TQR model, the loss function is associated with the asymmetric Laplacean density:

\[
f_\tau(u) = \tau (1 - \tau) \exp(-\rho_\tau(u)), \text{ where } \rho_\tau(\cdot) \text{ is the check function.}
\]

Despite this connection, it is rather implausible to assume the quantile error term follows the Laplacean distribution; see Koenker and Machado (1999, p.1298). Without assuming the correct specification of the error distribution (which is what a robust method like quantile regression tries to avoid exactly), we cannot address the optimality issue of a test for the absence of quantile threshold effect. Second, even if we make the above stringent distributional assumption on the quantile error term, the extension is challenging due to the nonsmoothness of the check function \( \rho_\tau(\cdot) \). AP requires that the density function be twice continuously differentiable with respect to the parameter of interest (\( \theta_1 = (\alpha'_\tau, \beta'_\tau)' \) in our notation), which is violated in our setup. So their proof strategy which relies on the second order Taylor expansion of the log-likelihood function repeatedly breaks down and some alternative method must be called upon.

In addition, Andrews and Ploberger (1995) demonstrate that the sup-Wald test and their asymptotically equivalent tests, namely, the sup-LR and sup-LM tests, are asymptotically admissible in the sense that they are the best tests against alternatives that are sufficiently distant from the null hypothesis in large samples. It is natural to ask whether our sup-Wald test is also asymptotic admissible. Again, Andrews and Ploberger (1995) work in a correctly specified maximum likelihood framework where the density function is twice continuously differentiable. Our quantile threshold regression does not fit into this framework and for this reason, we do not verify the potential asymptotic admissibility of our sup-Wald test here.

### 2.3 Asymptotic properties and inference under \( \mathbb{H}_{1\tau} \)

In this subsection, we first investigate the asymptotic properties of \( \hat{\theta}_{1\tau} \) and \( \hat{\gamma}_\tau \) under \( \mathbb{H}_{1\tau} \), and then focus on the inference issue associated with the threshold parameter \( \gamma_\tau \). The analysis here will pave a way to the test and inference issue associated with the common threshold parameter in the next subsection.

Let \( \theta_1 \equiv (\alpha'_\tau, \beta'_\tau)' \) and \( \hat{\theta}_\tau \equiv (\hat{\alpha}'_\tau, \hat{\beta}'_\tau)' \). Let \( N_\tau(\gamma) \equiv E \left[ x_i x_i' | r_i = \gamma \right], D_{\tau}(\gamma) \equiv E \left[ f(\alpha'_\tau x_i | z_i) x_i x_i' | r_i = \gamma \right], N_r \equiv N_{\tau}(\gamma_0^0), \text{ and } D_r \equiv D_{\tau}(\gamma_0^0). \) We study the strong consistency of \( \hat{\theta}_\tau \) and the asymptotic distributions...
of \( \hat{\theta}_{1\tau} \) and \( \hat{\gamma}_{\tau} \) in the following two theorems, respectively.

**Theorem 2.2** Suppose that Assumptions A1-A6 in Appendix A hold. Then \( \hat{\theta}_{\tau} = \theta^0_{\tau} + o_{a.s.}(1) \) where \( \theta^0_{\tau} = (\theta^0_{1\tau}, \gamma^0_{\tau})' \).

**Theorem 2.3** Suppose that Assumptions A1-A8 in Appendix A hold. Then

\[
\begin{align*}
(i) & \quad n^{1/2}(\hat{\theta}_{1\tau} - \theta^0_{1\tau}) \xrightarrow{d} N(0_{2k \times 1}, \tau(1-\tau) \Sigma(\tau, \gamma^0_{\tau})), \\
(ii) & \quad n^{-2a}(\hat{\gamma}_{\tau} - \gamma^0_{\tau}) \xrightarrow{d} \frac{\lambda_{\tau}}{\mu_{\tau}} \arg\max \{ W(r) - \frac{1}{2} |r| \},
\end{align*}
\]

where the parameter \( a \) is defined in Assumption A7, \( W(\cdot) \) is a two-sided Brownian motion, \( \Sigma(\tau, \gamma) \equiv \Omega_1(\tau, \gamma)^{-1} \Omega_0(\gamma, \gamma) \Omega_1(\tau, \gamma)^{-1} \), \( \lambda_{\tau} \equiv (1-\tau) v^T_\tau \Sigma(\tau, \gamma^0_{\tau}) v_\tau \), \( \mu_{\tau} \equiv v^T_\tau D_\tau v_\tau g(\gamma^0_{\tau}) \), \( g(\cdot) \) denotes the PDF of \( r_i \), and \( v_\tau \) are defined in Assumption A7.

Recall that 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 \( W_1(\cdot) \) and \( W_2(\cdot) \) are two independent standard Brownian motions on \([0, \infty)\). It is worth mentioning that the result in Theorem 2.3(i) continues to hold even if one allows \( a = 0 \) in Assumption A7. However, the asymptotic distribution of \( \hat{\gamma}_{\tau} \) in Theorem 2.3(ii) remains valid only for the case of \( a \in (0, \frac{1}{2}) \) in Assumption A7, analogously to the case of the least squares threshold regression in Hansen (2000) and the LAD threshold regression in Caner (2002). In particular, the limiting distribution of our quantile threshold effect estimator is invariant to the quantile index up to the scalar effect \( \frac{\lambda_{\tau}}{\mu_{\tau}} \) and the random component in the limiting distribution, i.e., \( \arg\max \{ W(r) - \frac{1}{2} |r| \} \) also appears in Hansen (2000) and Caner (2002). 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 \( n(\hat{\gamma}_{\tau} - \gamma^0_{\tau}) \) 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 \( a \in (0, \frac{1}{2}) \). One may also be interested in the asymptotic behavior of \( \hat{\gamma}(\tau) \equiv \hat{\gamma}_{\tau} \) when the estimators are treated as a stochastic process \( \{ \hat{\gamma}(\cdot) \} \) indexed by the quantile index \( \tau \) on the set \( T \). But because the limiting process is non-Gaussian and cannot be simply characterized by its mean and covariance kernel, we are unable to formally study this challenging issue.

Based on Theorem 2.3, 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. We will focus on the study of statistical inferences for \( \gamma_{\tau} \) below.

To make inferences about \( \gamma_{\tau} \), one may be tempted to apply the asymptotic distribution result in Theorem 2.3(ii). But since the asymptotic distribution of \( \hat{\gamma}_{\tau} \) depends on some nuisance parameters, inferences based on it tend to be poor in finite samples. Below, we follow the spirit of Hansen (2000) and consider an LR statistic to test the hypotheses about the threshold parameter \( \gamma_{\tau} \). Specifically, we are interested in testing the null hypothesis

\[
H_{0\tau} : \gamma_{\tau} = \gamma^0_{\tau}. \tag{2.15}
\]

We consider the LR statistic: \( LR_{n\tau}(\gamma^0_{\tau}) = S_{n\tau}(\hat{\theta}_{1\tau}, \gamma^0_{\tau}) - S_{n\tau}(\hat{\theta}_{1\tau}, \hat{\gamma}_{\tau}) \). We reject \( H_{0\tau} \) for large values of \( LR_{n\tau}(\gamma^0_{\tau}) \). The following theorem establishes the asymptotic distribution of \( LR_{n\tau}(\gamma^0_{\tau}) \) under \( H_{0\tau} \).

**Theorem 2.4** Suppose that Assumptions A1-A8 in Appendix A hold. Then under \( H_{0\tau} \), \( LR_{n\tau}(\gamma^0_{\tau}) \xrightarrow{d} \frac{\lambda_{\tau}}{\mu_{\tau}} \Xi \), where \( \Xi \equiv \sup_{r \in (-\infty, \infty)} \{ 2W(r) - |r| \} \), and \( \lambda_{\tau}, \mu_{\tau}, \) and \( W(\cdot) \) are as defined in Theorem 2.3.
Theorem 2.4 indicates that \( LR_{n\tau}(\gamma^0_\tau) \) is not asymptotically pivotal under the null hypothesis. To obtain an asymptotically pivotal test statistic, we need to estimate \( \lambda_f \) and \( \mu_f \) and consider the following normalized LR test statistic:

\[
NLR_{n\tau}(\gamma^0_\tau) = \frac{2\delta'_\tau \hat{D}_\tau \delta_\tau}{(1 - \tau) \hat{G}_\tau \hat{N}_\tau \delta_\tau} LR_{n\tau}(\gamma^0_\tau),
\]

where \( \delta_\tau \equiv \hat{a}_\tau(\gamma^0_\tau) - \hat{\beta}_\tau(\gamma^0_\tau) \), \( \hat{N}_\tau \equiv \hat{N}_\tau(\gamma^0_\tau) \) is a consistent local linear (or constant) estimate of \( N_{\tau_\tau}(\gamma^0_\tau) \) by using the bandwidth \( h_1 \) and the kernel \( K \); \( \hat{D}_\tau = \hat{D}(\gamma^0_\tau) = \hat{E}_{h_1}(|\hat{f}_\tau(\hat{\gamma}^0_\tau x_i | z_i) x_i x'_i | \gamma^0_\tau|, \hat{f}_h (\cdot | z_i) \) is a kernel estimate for the density of \( y_i \) given \( z_i \) by using the bandwidth \( h_1 \) and the kernel \( K \), and \( \hat{E}_{h_1} (\cdot | \gamma^0_\tau) \) is a kernel estimate of \( E \left[ f \left( (\alpha^0_x x_i | z_i) x_i x'_i | r_i = \gamma^0_\tau \right) \right] \) by using the bandwidth \( h_1 \), the kernel \( K \), and the observations on \( \hat{f}_h_1(\hat{a}_\tau(\gamma^0_\tau) x_i | z_i) x_i x'_i \) and \( r_i \). Under the assumption that \( z_i \) is compactly supported with bounded density that is bounded away from 0 on its support, one can obtain both estimates by the local linear method (e.g., Fan et al. (1996)) to avoid boundary bias and the asymptotic trimming issue. In this case, we can regress \( \phi(\gamma^0_\tau y_i / h_y) \) on \( z_i \) by using the local linear method with kernel \( K \) and bandwidth \( h_1 \) in order to obtain \( f(y_i | z_i) \), where \( h_y \) is another bandwidth and \( \phi(\cdot) \) is the standard normal PDF.

Under standard conditions, we can readily show that \( \frac{\hat{D}_\tau \hat{N}_\tau \delta_\tau}{(1 - \tau) \hat{G}_\tau \hat{N}_\tau \delta_\tau} \to \frac{\delta_\tau}{\delta_\tau} \) in probability. Then, by the Slutsky lemma, we have \( NLR_{n\tau}(\gamma^0_\tau) \to d \Xi \). That is, \( NLR_{n\tau}(\gamma^0_\tau) \) is asymptotically pivotal. It is well known that \( \sup_{r \in [0, \infty]} 2W(r) - |r| \) and \( \sup_{r \geq 0} 2W(r) - |r| \) are independent exponential random variables with distribution function \( 1 - e^{-z} \) such that the CDF of \( \Xi \) 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 \( NLR_{n\tau}(\gamma^0_\tau) \). See Hansen (2000, p. 582) for more details. In addition, we can invert the \( NLR_{n\tau}(\gamma^0_\tau) \) statistic to obtain the asymptotic 1 - \( \alpha \) confidence interval for the normalized statistic \( NLR_{n\tau}(\gamma^0_\tau) \). See Hansen (2000, p. 582) for more details. In addition, we can invert the \( NLR_{n\tau}(\gamma^0_\tau) \) statistic to obtain the asymptotic 1 - \( \alpha \) confidence interval for the common threshold parameter \( \gamma^0_\tau \) is given by \( C_{1 - \alpha} = \{ \gamma : NLR_{n\tau}(\gamma) \leq \Xi_{1 - \alpha}, \gamma \in \Gamma_n \} \), where \( \Xi_{1 - \alpha} \) is the 1 - \( \alpha \) upper percentile of \( \Xi \) (e.g., \( \Xi_{1 - 0.1} = 5.94, 7.35 \) and 10.59 for \( \alpha = 0.1, 0.05, \) and 0.01, respectively).

### 2.4 Test and Inference for the common threshold parameter across quantiles

In data analysis, we may suspect that different conditional quantile functions share a common threshold value. If it were the case, joint analysis of multiple quantile regressions would improve the accuracy of the common threshold estimate. In this subsection, we show that our previous analysis can be naturally extended to the case of a common break, which turns out to be relevant in our empirical application later.

Recall that \( \mathcal{T} \equiv [\underline{\tau}^\tau, \overline{\tau}^\tau] \subset (0, 1) \). We first propose a test for the hypothesis of common threshold value on the set \( \mathcal{T} \). That is, we consider testing the null hypothesis

\[
H_0 : \gamma_\tau = \gamma^0 \quad \text{for all } \tau \in \mathcal{T} \quad \text{and some } \gamma^0 \in \Gamma,
\]

versus the alternative hypothesis

\[
H_1 : \text{There is no } \gamma \in \Gamma \text{ such that } \gamma_\tau = \gamma \text{ for all } \tau \in \mathcal{T}.
\]

The test of the null hypothesis in (2.17) serves as a specification test for the key assumption of common threshold value in Yu (2014). If the null is rejected, inferences in Yu (2014) would be invalid. Otherwise,
one can follow Yu (2014) and conduct inferences that tend to be more efficient than those based on a single quantile regression.

In general, $\gamma^0$ is not observed. Under $H_0 : \gamma_\tau = \gamma^0$ for all $\tau \in \mathcal{T}$, we can estimate $\gamma^0$ by

$$
\hat{\gamma} \equiv \hat{\gamma}(\Pi) \equiv \arg\min_{\hat{\gamma} \in \mathcal{T}} \hat{S}_{n,\Pi}(\gamma),
$$

(2.19)

where $\hat{S}_{n,\Pi}(\gamma) = \int S_{n\tau}(\hat{\theta}_1(\tau, \gamma), \gamma)d\Pi(\tau)$, and $\Pi$ is a user-specified probability distribution function defined on $\mathcal{T}$. After we obtain the estimate $\hat{\gamma}$ of $\gamma$, we can estimate $\alpha^0_\tau$ and $\beta^0_\tau$ 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 $\hat{\theta}_{1\tau} \equiv (\hat{\alpha}_\tau', \hat{\beta}_\tau')'$. The following theorem summarizes the important properties of $\hat{\gamma}, \hat{\theta}_1(\tau, \gamma)$, and $\hat{\theta}_{1\tau}$.

**Theorem 2.5** Suppose that Assumptions A1-A4 and A6-A9 in Appendix A 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) $\sqrt{n}(\hat{\theta}_1(\tau, \gamma) - \theta^0_1(\tau, \gamma)) = \Omega(\tau, \gamma)^{-1}\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(y_i - \theta^0_1(\tau, \gamma)' z_i(\gamma)) z_i(\gamma) + o_p(1)$ uniformly in $(\tau, \gamma) \in \mathcal{T} \times \Gamma$;

(iii) $n^{1/2}(\hat{\theta}_1(\tau, \gamma) - \theta^0_1(\tau, \gamma)) \stackrel{d}{\rightarrow} N(0_{2k \times 1}, \tau(1 - \tau) \sigma^2(\tau, \gamma^0))$;

(iv) $n^{1-2a}(\hat{\gamma} - \gamma^0) \stackrel{d}{\rightarrow} \frac{\lambda_0}{\mu_0} \arg\max_{\tau \in (-\infty, \infty)} \{W(\tau) - \frac{1}{2}|\tau|\}$;

where $\psi_\tau(u) \equiv \tau - 1\{u < 0\}$, the pseudo-true values $\gamma^0, \theta^0_{1\tau}$ and $\theta^0_1(\tau, \gamma)$ together with $\Omega(\tau, \gamma)$ are defined in Assumption A9, $\lambda_0 \equiv \{\int \sqrt{\tau(1 - \tau)} v_{1\tau}^\tau N_{1\tau}(\gamma^0) v_{1\tau}d\Pi(\tau)\}^2 g(\gamma^0)$, and $\mu_0 \equiv \int v_{1\tau}^2 D_{1\tau}(\gamma^0) v_{1\tau}d\Pi(\tau) g(\gamma^0)$ in (iv), and $W(\cdot)$ is as defined in Theorem 2.3.

Theorem 2.5(i) implies the consistency of the parameter estimates. Theorem 2.5(ii) extends the uniform Bahadur representation result in Lemma B.1 to allow a single common break in the quantile processes. The last two parts of Theorem 2.5 are parallel to those in Theorem 2.3. In particular, Theorem 2.5(iii) indicates that the first-order asymptotic distribution of $\hat{\theta}_{1\tau}$ is the same as that of $\hat{\theta}_1(\tau, \gamma)$ obtained before. This is as expected due to the asymptotic independence between the estimator of the regression coefficients and that of the threshold parameter.

Given the above estimator $\hat{\gamma}$ of $\gamma^0$ under $H_0$ and motivated from the early LR test for a threshold parameter, we consider the following LR test statistic:

$$
LR_n = \int \left[ S_{n\tau}(\hat{\theta}_{1\tau}, \hat{\gamma}) - S_{n\tau}(\hat{\theta}_{1\tau}, \gamma^0) \right] d\Pi(\tau).
$$

(2.20)

The following theorem reports the asymptotic distribution of the $LR_n$ statistic.

**Theorem 2.6** Suppose that Assumptions A1-A4 and A6-A9 in Appendix A hold. Then, under $H_0 : \gamma_\tau = \gamma^0$ for all $\tau \in \mathcal{T}$ and some $\gamma^0 \in \Gamma$, we have $LR_n \stackrel{d}{\rightarrow} c_{LR} \Xi$, where $c_{LR} = \int \frac{\lambda_0}{2\mu_0} d\Pi(\tau) - \frac{\lambda_0}{2\mu_0} \lambda_\tau, \mu_\tau$ and $\Xi$ are defined in Theorem 2.4, and $\lambda_0$ and $\mu_0$ are defined in Theorem 2.5.

To implement the $LR_n$ test, we need to estimate $c_{LR}$. Following the discussion after Theorem 2.4, we propose to estimate it by

$$
c_{LR} = \frac{1}{2} \int \frac{\tau(1 - \tau)}{\hat{\delta}_1^2 D_{1\tau} \hat{\delta}_\tau} d\Pi(\tau) - \left\{ \int \frac{\sqrt{\tau(1 - \tau)} \hat{\delta}_1^2 N_{1\tau} \hat{\delta}_\tau d\Pi(\tau)}{2 \int \hat{\delta}_1^2 D_{1\tau} \hat{\delta}_\tau d\Pi(\tau)} \right\}^2
$$

(2.21)
where \( \delta_\tau, \hat{N}_\tau, \) and \( \hat{D}_\tau \) are defined as before. Using the Cauchy-Schwarz inequality, we easily show that \( \tilde{c}_{LR} \geq 0 \) and the equality holds if and only if \( \delta_\tau^T \hat{N}_\tau \delta_\tau = (\delta_\tau^T \hat{D}_\tau \delta_\tau)^2 \) a.e.-\( \Pi \). It is easy to show that \( \tilde{c}_{LR} \) converges in probability to \( c_{LR} \). Then under under \( H_0 \), the normalized version of \( LR_n \) satisfies

\[
NLR_n \equiv \tilde{c}_{LR}^{-1}LR_n \overset{d}{\to} \Xi. \quad (2.22)
\]

In some applications, \( \gamma^0 \) is known under the null hypothesis of common threshold. This can be the case of the RDD framework where the potential discontinuity/threshold point is commonly observable. It can also be the case when one fails to reject the null hypothesis by using the above \( NLR_n \) test statistic, concludes that the quantile regression shares a common threshold value for \( \tau \in \mathcal{T} \), and then tries to test whether the common threshold value is given by particular value \( \gamma^0 \). In this case, we write the null hypothesis as

\[
H^*_0 : \gamma_\tau = \gamma^0 \text{ for all } \tau \in \mathcal{T} \quad (2.23)
\]

and the alternative hypothesis as

\[
H^*_1 : \gamma_\tau \neq \gamma^0 \text{ for some } \tau \in \mathcal{T}. \quad (2.24)
\]

In this case, we investigate the following LR statistic:

\[
LR_n (\gamma^0) = \int \left[ S_{n\tau} \left( \hat{\theta}_{1\tau}, \gamma^0 \right) - S_{n\tau} \left( \hat{\theta}_{1\tau}, \hat{\gamma} \right) \right] d\Pi (\tau). \quad (2.25)
\]

We reject \( H^*_0 \) for large values of \( LR_n (\gamma^0) \). The following theorem establishes the asymptotic distribution of \( LR_n (\gamma^0) \) under \( H^*_0 \) in (2.23).

**Theorem 2.7** Suppose that Assumptions A1-A4 and A6-A9 in Appendix A hold. Then, under \( H^*_0 \) we have \( LR_n (\gamma^0) \overset{d}{\to} \frac{\lambda_0}{\mu_0} \Xi \), where \( \lambda_0 \) and \( \mu_0 \) are as defined in Theorem 2.5.

Clearly, the asymptotic distribution of \( LR_n (\gamma^0) \) is analogous to that of \( LR_{n\tau} (\gamma^0) \) in Theorem 2.4 under \( H_{0\tau} \) for a specific quantile index \( \tau \). To implement the test, we consider the following normalized version of \( LR_n (\gamma^0) \):

\[
NLR_n (\gamma^0) = \frac{2 \int \delta_\tau^T \hat{D} (\gamma^0) \delta_\tau d\Pi (\tau)}{\left\{ \int \sqrt{\tau (1 - \tau)} \delta_\tau^T \hat{N}_\tau (\gamma^0) \delta_\tau d\Pi (\tau) \right\}^2} LR_n (\gamma^0), \quad (2.26)
\]

where \( \delta_\tau = \hat{\alpha}_\tau - \hat{\beta}_\tau \), and \( \hat{D} (\gamma^0) \) and \( \hat{N}_\tau (\gamma^0) \) are defined analogously to \( \hat{D} (\gamma^0) \) and \( \hat{N}_\tau (\gamma^0) \). One can readily show that \( NLR_n (\gamma^0) \overset{d}{\to} \Xi \) under \( H^*_0 \). The \( 1 - \alpha \) confidence interval for the common threshold parameter \( \gamma^0 \) is given by \( C_{1-\alpha} = \{ \gamma : NLR_n (\gamma) \leq \Xi_{1-\alpha}, \gamma \in \Gamma_n \} \).

By the definitions of the two LR statistics, namely \( LR_n \) in (2.20) and \( LR_n (\gamma^0) \) in (2.25), we use the unrestricted estimators \( \hat{\theta}_{1\tau} \) for the likelihood under both the null and alternative hypotheses. We can do this because the value \( \theta^0_{1\tau} \) is not affected by whether the threshold parameter \( \gamma_\tau \) is common or not, making \( \hat{\theta}_{1\tau} \) consistent under both the null and alternative. A close examination of the proofs of Theorems 2.6 and 2.7 suggests that such a use of the same estimators \( \hat{\theta}_{1\tau} \) generates some cancelling mechanism to greatly facilitate the proof. Conceptually, one may consider using the restricted estimators of \( \theta_{1\tau} \) under the null but that will greatly complicate the derivation of the asymptotic distributions of our LR test statistics.
3 Monte Carlo Simulations

In this section we conduct a set of Monte Carlo experiments to evaluate the finite sample performance of our tests and estimates.

3.1 Data generating process

We consider the following data generating process (DGP):

\[
y_i = \begin{cases} 
1 + \Phi^{-1}(v_i) + (a_0 + a_1 v_i)x_i, & \text{if } r_i \leq 0.5 + c|v_i - 0.5| \\
1 + \Phi^{-1}(v_i) + c_0 n^{-1/8} + (a_0 + a_1 v_i + c_0 n^{-1/8})x_i, & \text{if } r_i > 0.5 + c|v_i - 0.5|
\end{cases},
\]

(3.1)

where \( \Phi(\cdot) \) is the standard normal cumulative distribution function (CDF), \( x_i \)'s are independent and identically distributed (IID) from the Beta(2,3) distribution, \( r_i \)'s and \( v_i \)'s are independently generated from the uniform distribution on \([0,1]\). In our simulation, we chose \( a_0 = 1 \) and \( a_1 = 0.5 \) and consider various values for \( c_0 \) and \( c \). We consider both sample sizes of \( n = 200 \) and \( n = 400 \). The number of repetitions is set as 500.

Apparently, the \( \tau \)th conditional quantile of \( y_i \) given \( (x_i, r_i) \) is

\[
Q_{\tau}(x_i, r_i) = \{1 + \Phi^{-1}(\tau) + (1 + 0.5\tau)x_i\} 1\{r_i \leq 0.5 + c|\tau - 0.5|\} + \{1 + \Phi^{-1}(\tau) + c_0 n^{-1/8} + (1 + 0.5\tau + c_0 n^{-1/8})x_i\} 1\{r_i > 0.5 + c|\tau - 0.5|\}.
\]

Here, \( c_0 n^{-1/8} \) signifies the jump size for both the intercept and slope coefficients. (3.1) can accommodate various scenarios of interest by taking different values on \( c_0 \) and \( c \). In particular, we shall consider the following cases:

1. \( c_0 = 0 \). In this case, there is no quantile threshold effect irrespective of the value of \( c \).

2. \( c = 0 \) and any nonzero \( c_0 \). In this case, we have the quantile threshold effect and the threshold effect is common across all quantile indices \( \tau \) and given by \( \gamma^0 = 0.5 \).

3. \( c = 0.5 \) and any nonzero \( c_0 \). In this case, we have the quantile threshold effect \( \gamma^0_\tau = 0.5 + 0.5 \cdot |\tau - 0.5| \), which is varying over \( \tau \).

3.2 Test for the presence of quantile threshold effect

We first consider the sup-Wald test for the presence of quantile threshold effect at the three conventional significance levels, namely, 1%, 5%, and 10%. To obtain the simulated \( p \)-value for the test statistic, we need to choose the bandwidth \( h = h(\tau) \) to obtain the estimates \( \hat{\Omega}(\tau, \gamma) \) and \( \hat{\Omega}^*(\tau, \gamma) \). Following Koenker (2005), we set \( h(\tau) = \kappa \left( \Phi^{-1}(\tau + e_n) - \Phi^{-1}(\tau - e_n) \right) \), where \( \kappa \) is a robust estimate of the scale of the quantile residual, \( e_n = \frac{2}{3} n^{-1/3} \), and \( \Phi^{-1}(\cdot) \) is the inverse function of \( \Phi(\cdot) \). To implement the sup-Wald test, we first consider the test for a fixed quantile index where the set \( T \) is a singleton \( \{\tau\} \) for \( \tau = 0.1, 0.2, ..., \) and 0.9. Then we consider the test by setting \( T = [0.1, 0.9] \).
Table 1 reports the rejection frequency for testing the presence of quantile threshold effect. We consider 500 repetitions and generate 200 simulated samples of \{u_i\} in each repetition to obtain the simulated p-values. We made a few observations from Table 1. First, the top panel in Table 1 indicates that the sup-Wald test tends to be undersized where \(T = \{\tau\}\) is a singleton and \(\tau\) lies close to either 0 or 1, and it is moderately oversized for the 5% test when \(\tau\) takes value around 0.5. This is especially true when \(n = 200\). When \(T = [0.1, 0.9]\), both the 5% and 10% tests are undersized when \(n = 200\), but the sizes improve quickly when \(n\) is increased to 400.

Second, in terms of power, we compare three cases with different values of \(c_0\), i.e., \(c_0 = 0.5, 1\) and 2 in Table 1. It shows that for any given sample size, the larger value of \(c_0\) the higher power of our test. It is indeed intuitive. The power of our test to detect the presence of threshold effect depends on the underlying jump size at the thresholds in the DGP. Of course, it also depends on the choice of \(T\).

Next, we focus on the DGP with moderate jump size at threshold (i.e., \(c_0 = 1\)). We observe that our test has stable power, no matter whether the threshold effect is common or not. When \(n = 200\), the test based on \(T = [0, 1, 0.9]\) is not necessarily more powerful than those based on individual quantile index for \(\tau\) to take values not far away from 0.5. Nevertheless, for \(n = 400\), we observe that the test based on \(T = [0.1, 0.9]\) is more powerful than those based on individual quantile index \(\tau\). This suggest the benefit of taking supremum over \(\tau \in T\) in large samples.

3.3 Estimation Results

We consider the estimation of the TQR model in both cases of common threshold value (\(c = 0\)) and non-common threshold value (\(c = 0.5\)). We also specify \(c_0 = 0.5\) in DGP (3.1). Table 2 provides the mean squared error (MSE) of the estimates of quantile regression coefficients when \(\tau = 0.1, 0.2, ..., 0.9\) across 500 repetitions. The results in Table 2 are as expected. First, the MSEs are generally larger when \(\tau\) is close to 0 or 1 than when \(\tau\) is close to 0.5. Second, as the sample size \(n\) doubles, we observe that the MSEs are roughly halved, which is consistent with the \(\sqrt{n}\)-consistency of the quantile regression coefficient estimators.

3.4 Test for the presence of a common threshold value

We implement the LR test for the presence of a common threshold value. The test statistic used is \(NLR_n\) defined in (2.22). To construct the test statistic, we need to specify the probability distribution function \(\Pi(\cdot)\). Here, we specify \(\Pi(\cdot)\) through its PDF \(\pi(\cdot)\): \[\pi(\tau) = 2 - 4(\tau - 0.5)\text{sgn}(\tau - 0.5),\] where \(\text{sgn}(u) = -1\) if \(u \leq 0\) and 1 otherwise. One can check that \(\pi(\tau)\) is nonnegative and integrated into one on \([0,1]\). Apparently, this way of specification of \(\pi(\cdot)\) allows us to put more weight on \(\tau\) when it is around 0.5 than when \(\tau\) is close to either end because of the low estimation accuracy of the quantile regression coefficients when \(\tau\) is close to 0 or 1. To test from the presence of common threshold on
\( T = [\tau, \bar{\tau}] \), we approximate the integral as follows:

\[
\int_{\tau}^{\bar{\tau}} A(\tau) d\Pi(\tau) \approx \frac{1}{N} \sum_{j=1}^{N} A(\tau_j) \pi(\tau_j),
\]

where \( A(\cdot) \) is an arbitrary smooth function and \( \tau_1, \tau_2, \ldots, \tau_N \) are equally spaced on \( T \). The above approximation can be justified from the law of large numbers in statistics or the Riemann sum formula for approximating a definite integral, and the approximation error can be made arbitrarily well by choosing sufficiently large \( N \). We find through simulations that our simulations results are not sensitive to the choice of \( N \) provided \( N \geq 40 \). To save on computation, we fix \( N = 41 \) in our simulations.

We then let \( c \) vary from 0 to 1 in DGP. When \( c = 0 \), we examine the level behavior of our \( NLR_n \) test; when \( c > 0 \), we check whether \( NLR_n \) has any power to detect the deviation from the null hypothesis of a common threshold value. In particular, our concern is how the power function depends on the value of \( c \).

To construct the test statistic \( NLR_n \) in (2.22), we obtain the estimates \( \hat{D}_\tau \) and \( \hat{N}_\tau \) by following the remarks after Theorem 2.4. We consider the local linear estimates of the conditional density \( f(\cdot | z_i) \) by following Fan et al. (1996) closely. We choose the kernel function \( K \) as the standard normal PDF and specify the two bandwidth sequences by following Silverman’s normal reference rule: \( h_y = 1.06s_y n^{-1/5} \) and \( h_1 = 1.06s_z n^{-1/6} \) where \( s_z = (s_x, s_r) \), and \( s_y, s_x \) and \( s_r \) denotes the sample standard deviations of \( \{y_i\}, \{x_i\} \) and \( \{r_i\} \), respectively.

Figure 1 plots the rejection frequency of our \( NLR_n \) test for the presence of a common threshold value when the nominal level is fixed at 0.05 and the number of replications is 500. When \( c = 0 \), the rejection frequency is about 0.032 and 0.040 for \( n = 200 \) and 400, respectively. This suggests that the level of the test is well controlled. As \( c \) increases, we observe the stable increase of the power function. As \( n \) increases, the power also increases for the fixed positive value of \( c \). This indicates that our test has reasonably good power in detecting deviations from the null hypothesis of a common threshold value.

### 3.5 Inference on the threshold parameter

It is well known that the inference on the threshold parameter in a threshold regression is a hard problem despite the availability of the asymptotic distribution result. Here we consider the inference on the threshold parameter based on individual TQR or integrated QR in the case of a common threshold value.

Following Hansen (2000), we first consider the empirical coverage ratio and average length for the 95% confidence interval for \( \gamma_\tau^0 \) by inverting the \( NLR_{n, \tau} (\gamma_\tau^0) \) given in (2.16). To construct the test statistic, we obtain the estimates \( \hat{D}_\tau \) and \( \hat{N}_\tau \) as described above. The confidence interval is asymptotically valid no matter whether the threshold parameters are common across the quantile index \( \tau \) or not. In the presence of a common threshold value, we also consider the empirical coverage ratio and average length for the 95% confidence interval for the common threshold parameter \( \gamma^0 \) by inverting the \( \overline{NLR}_n (\gamma^0) \) given in (2.26).

Table 3 reports the results the empirical coverage ratio and average lengths for constructing the 95% confidence intervals of \( \gamma_\tau^0 \) or \( \gamma^0 \). It shows that, for both empirical coverage ratio and average lengths of the estimated confidence intervals, tail index (i.e., when \( \tau \) is close to 0 or 1) performs less satisfactory than the middle range of values for \( \tau \). As the sample size \( n \) increases, both measures get to improve, that
is, the average lengths decrease and coverage ratios approach to the nominal 95% confidence level. Most favorably, however, the inference on $\gamma_0$ seems to perform better in general than that of $\gamma_0^\tau$.

4 Empirical Application: Pricing for Reputation

In this section, we apply our methodology to investigate the reputation and pricing patterns on the dominant online trading platform in China, Taobao.com. Similar to eBay in the United States, Taobao provides a platform for businesses and individual entrepreneurs to open online retail stores that cater mainly to consumers in mainland China.

The reputation scoring system in 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 his/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 4 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. At the time of posting prices, the information on seller’s reputation scores and category will both be revealed to potential trading partners.

4.1 Data

We collect the trading data on the iPod Nano IV 8G on Taobao.com from September to December 2009. Two factors motivate the choice of 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.

We first plot the raw data of prices and reputation scores and the histogram of the sellers’ reputation category in Figures 2 and 3, respectively. Two observations can be made from these plots, which motivate our empirical study. First, Figure 2 indicates that the prices posted by the sellers with reputation scores 5000 or above are much less scattered with a significantly higher mean than other sellers in the market. It echoes the established results in the literature that the reputation rewards. Furthermore, Figure 3 suggests that most of the sellers are spread between Categories 1 and 9 (1-heart to 4-diamonds), and very few have accumulated more than 50,000 good reviews (Category 12). We therefore spot a possible exogenous cutoff that might provide incentives for the sellers on Taobao.com to price for a better reputation: 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 pay-off from their better reputation, with a tremendous reduction in competition. We therefore define “reputation” in this application by a seller’s reputation classified as Category 10 or above. Second, the
price distributions appear different for the reputation scores in the range of 2500-3500. We then suspect that there may involve a regime change in underlying pricing strategies occurred in this range, given the possible motives from the “reputation”.

4.2 An economic model of pricing for reputation

Next, we construct 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, 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 “regime change” in the pricing rule. We therefore refer to this pattern as “pricing for reputation”.

Consider a monopolist with current reputation status (score) $r$ who is selling a product with zero marginal cost. 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 $\beta$. The empirical literature has documented extensive evidence to show that sellers with a superior reputation generate significantly higher profits. This $\beta$ 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

$$\Pi_0(p; r) = \Pi_1(p; r) + \Pi_2(p; r) = \Pi_1(p; r) + \beta \Pr[R(p, e) \geq \bar{r} - r],$$

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 $\Pi_2$ for any given $(p, r)$. Therefore, $\Pi_1$ denotes the profit a seller obtains from the market without any concerns over reputation benefits, and $\Pi_2$ is the expected gain in extra profit from good reputation.

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 reputation 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

$$\Pi_0(p; r) = p(1 - \alpha p) + \beta F(1 - \alpha p - \bar{r} + r).$$

Let $f'$ and $f''$ denote the first- and second-order derivatives of $f$. We make the following assumptions on the density function $f$.

**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, $\lim_{e \to 0} f'(e) < 2/(\alpha \beta)$.

These assumptions require a special curvature on $f$ to the left of its mode. This curvature induces increasing marginal returns on a segment of $\Pi_0$, which implies that the profit function $\Pi_0$ is not globally
concave. Indeed, it is this particular curvature that delivers the pricing strategy for reputation in the following proposition.

**Proposition 4.1** 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 $\gamma_0$ 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 $\partial p^*_1(r)/\partial r < 0$.

We leave the technical proof of Proposition 4.1 and detailed discussion on the model intuition to the supplementary Appendix C. It is worth mentioning that Assumption M1 implies that $f$ is a unimodal density function. This is plausible in application, if one believes that, for example, the random process for a consumer writing a good review follows a binomial distribution. Then, the uncertainty a seller may face for not getting good reviews from sales may be well approximated by a normal distribution. The height restriction on $f$ in Assumption M1 ensures that the first-order condition is equipped with a solution, thereby effectively ruling out an uninteresting case in which the reputation effect would dominate over the current monopolistic pricing (and therefore the seller would charge zero price). 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.

Proposition 4.1 suggests that the seller’s optimal pricing schedule when $r \in [\gamma_0, \bar{r}]$ is different from that when $r < \gamma_0$. Such a regime change predicts two observable patterns in the seller’s optimal pricing. First, there exists a discontinuity in the pricing function, which occurs at $\gamma_0$. In our model, the “jump” reflects the local maxima switches at the seller’s optimal pricing decision. Second, the pricing function is always negatively sloped in $r$ before the regime change (that is, $r < \gamma_0$). However, the dependence of pricing function with respect to $r$ after $\gamma_0$ is indeterminate. In our model, such an ambiguity is induced by the unimodal shape of probability distribution on not incurring good reviews from sales. It is these particular pricing patterns that are referred to as the “reputation effect” in this paper.

Apparently, the exact value of $\gamma_0$ hinges on the model parameters of $\alpha$ and $\beta$, 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. Quantile regression is a flexible way to model the heterogeneous influences of explanatory variables on the response variable, which is the selling price here.

Our application of pricing strategy at Taobao.com connects our work to the literature of reputation. Empirical studies are abundant for the past decade, ensuing the emerging digital market of Internet. Among them, the most related branch should be the one which 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). Moreover, Cabral and Hortacsu (2010) explicitly acknowledge the existence of significant unobservable seller heterogeneity in the electronic markets. Our model and methodology developed above respect and include these features of internet market in the literature. It has also been noticed that only reviewers who provide good reviews tend to break the silence. See, e.g., 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. To the best knowledge of ours, we are not aware of any other theory model that is designed for reputation system in internet market of this kind.

4.3 Estimation and testing results

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 500 (Category 6 and below) as rookies and exclude them from our data analysis. Therefore, the sample for this study includes only sellers with reputation scores between 500 and 5000, and the total number of observations is $n = 1903$.

4.3.1 Testing for the existence of a quantile threshold effect

First, we conduct the sup-Wald test for the existence of a change point in the data following the approach suggested in Section 2.2. As in the simulations, we consider testing the null hypothesis of no threshold effect for all quantile indices between 0.1 and 0.9 (i.e., $\mathcal{T} = [0.1, 0.9]$ in (2.5)) and for nine individual quantile indices (i.e., $\mathcal{T} = \{\tau\}$ in (2.5) for $\tau = 0.1, 0.2, \ldots, 0.9$). The implementation is done as in the simulation section. Table 5 reports the test statistic, the simulated $p$-value, and the simulated critical values at the three conventional significance levels (1%, 5%, and 10%). The $p$-value for the sup-Wald test based on $\mathcal{T} = [0.1, 0.9]$ is 0.000, which offers strong evidence for the existence of a jump behavior in the pricing behavior. For the sup-Wald test based on individual quantile index $\tau$, we find that at the 5% nominal level, jump points exist for quantiles up to 0.7 and that the breaks do not occur for such high quantiles as 0.8 and 0.9.

4.3.2 Estimation

Given the above findings, we can estimate the quantile regression coefficients for $\tau = 0.1, 0.2, \ldots, 0.7$ when the quantile threshold effect is detected. Table 6 reports the parameter estimates for these typical quantiles. Figure 4 shows the plots of the quantile regression lines before and after the changes 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 slope parameters before the jumps are mostly negative among the statistically significant estimates, consistent with our model predictions. An exception occurs for quantile 0.1 where the slope estimate is positive and statistically significant. We will comment on this case in the supplementary material of this paper.
We tend to have more significant slope estimates after the change point, and they are much larger in magnitude than those before the jumps. Furthermore, we also observe that, for the quantiles below the median, the slope estimates are negative. They turn positive for the median and upper quantiles. Recall that our model predicts such differences on slope parameters, which is induced by the unimodal shape of the probability distribution on not incurring good reviews from sales. Particularly, the change in signs of slope parameters implies the shift from tails to the right of modal on the probability distribution of “not incurring good review from sales”. In other words, it is rather likely for the sellers posting prices at lower quantiles (up to the median) to receive good reviews from sales. In stark contrast, the sellers who post prices at upper quantiles may face certain probabilities of not incurring good feedback. We believe this finding concurs with general intuition on Internet market – a lower price is more likely to generate a positive feedback from customers. In general, our estimates justify not only the relevance of our economic model but also our quantile regression approach.

4.3.3 Testing and inference on the threshold parameter

We then test for the presence of a common threshold value for the quantile index \( \tau \in T \). We implement the \( NLR_n \) test statistic in (2.22) by specifying three choices for \( T \), namely, \( T = [0.1, 0.5] \), \( [0.1, 0.7] \), and \( [0.1, 0.9] \). The test statistic takes values 0.0161, 7.7198, and 53.6874, respectively, with the corresponding \( p \)-values 0.9999, 0.0417, and 0.000. That is, we have a common threshold value for \( T = [0.1, 0.5] \).

Table 7 reports the 95% confidence intervals for the jump location estimates (\( \hat{\gamma} \)). We find the upper bound for the 95% confidence intervals by \( \inf \{ \gamma : \gamma > \hat{\gamma} \text{ and } NLR(\gamma) \leq c_{0.95} \} \), where \( c_{0.95} \) denotes the 0.95-level critical value for \( \Xi \). Accordingly, the lower bound for the 95% confidence intervals is defined by \( \sup \{ \gamma : \gamma < \hat{\gamma} \text{ and } NLR(\gamma) \leq c_{0.95} \} \). 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.

The previous testing results of a common threshold suggest that the quantiles between 0.1 and 0.5 have a common threshold at which regime changes occur. Hence, we implement the estimating and inference procedure on these quantiles for the case of common threshold. These results are also reported in Table 7. The jump is estimated to occur at 3264, and its 95% confidence interval appears tighter than those of single quantile estimates.

Next, we address several concerns on the empirical exercise. First, we considered only linear specification in the previous analysis to ease the exhibition. But, non-linearity may occur in applications for various reasons. To capture this possible pattern in data, we augment our TQR model by including \( r^2 \) and interaction terms of \( r \) with other observables. We found the coefficients of these non-linear terms insignificantly different from 0 at the 10% level for all quantiles under investigation.

Second, in comparison with our TQR model, one may wonder whether a least squares threshold regression can identify as much about the pricing scheme changes as our TQR model. To address this concern, we estimate Hansen’s (2000) least squares threshold model. The results are reported in the last rows of Tables 6 and 7. It suggests several differences between the mean and quantile estimates. Among them, two are worth highlighting. One is that the slope estimates before and after the change in the mean regression are both statistically insignificant, indicating less support of dependence between prices and reputation scores. Moreover, the confidence interval for the estimated threshold in the least square estimation is much
wider than those obtained in quantile regressions. Generally speaking, these differences shed some lights on the necessity of using quantile regression models for the consideration of unobserved heterogeneities. In our application, the heterogeneous pricing behavior across quantiles may reflect differences across sellers and market demand situations in online markets.

Our last 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 develops a test of manipulation related to the 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.

\section{Conclusion}

We consider quantile regression models where there may exist a threshold effect and the threshold effect, if exists, may be common for all quantile indices in a proper subset of \((0, 1)\) or not. Following the literature, a sup-Wald test is proposed to determine the existence of a threshold effect in the quantile regression across quantiles. We propose two tests for the presence of a common threshold value across different quantile indices and obtain their limiting distributions. We apply our methodology to study the pricing for reputation via the use of a dataset from Taobao.com and find both the existence of a threshold effect across many quantiles and the presence of a common threshold value for across quantile indices in the set \([0.1, 0.5]\).

Several extensions are possible. First, we only consider a parametric TQR model. It is feasible to extend our analysis to the nonparametric TQR model as considered by Oka (2010). Second, once common threshold effect is detected, it is natural to consider more efficient inference on the threshold parameter. Yu (2014) has started this line of research and more can be done. Third, we only consider IID observations and the presence of one threshold. It is possible to extend to time series observations and multiple-threshold scenario (see, e.g., Lanne and Saikkonen 2002; Galvao et al. 2014). Fourth, the economic model in our application is a simple static model. It is interesting to extend it to the dynamic case. We leave these for future research.
Appendix

In this appendix we provide assumptions and prove all main results but Proposition 4.1 in the paper.

A Assumptions

We make the following assumptions.

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

Assumption A2. \(E \|x_i\|^2 < \infty\).

Assumption A3. The conditional CDF \(F(\cdot | z)\) of \(y_i\) given \(z_i = z\) admits a 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. \(\Omega_1 (\tau, \gamma)\) is positive definite for each \((\tau, \gamma) \in T \times \Gamma\).

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

Assumption A7. Let \(\delta_{\tau}^\prime \equiv \delta_{\tau n}^\prime \equiv (a_{\tau n 0}^0 - \beta_{\tau n 0}^0)\). \(\delta_0^0 = v_\tau n^{-a}\) with \(v_\tau \neq 0\) and \(a \in (0, \frac{1}{2})\).

Assumption A8. (i) \(N_\tau (\gamma)\) and \(D_\tau (\gamma)\) are continuous at \(\gamma_0^0\). \(v_\tau^0 N_\tau v_\tau > 0\), \(v_\tau^0 D_\tau v_\tau > 0\), and \(g(\gamma_0^0) > 0\). (ii) \(E \|x_i\|^4 < \infty\).

Assumption A9. (i) Let \(\theta_\tau^0(\tau, \gamma) \equiv \text{argmin}_{\theta_\tau} S (\theta_1; \tau, \gamma)\), where \(S (\theta_1; \tau, \gamma) \equiv E[\rho_\tau(y_i - \theta_1^0 z_i (\gamma))]\). There exists a \(\gamma_0^0 \in \Gamma\) such that \(\gamma_0^0 = \text{argmin}_{\gamma} S (\theta_1^0(\tau, \gamma); \tau, \gamma)\) for all \(\tau\). (ii) Let \(\theta_{1\tau}^0 \equiv \theta_1^0(\tau, \gamma_0^0)\) and \(\Delta(z_i, \tau, \gamma) \equiv \theta_{1\tau}^0(z_i, \gamma_0^0) - \theta_{0\tau}^0 z_i (\gamma_0^0)\), and \(\int P(\|\Delta(z_i, \tau, \gamma)\| > 0) d\Pi (\tau) > 0\) for all \(\gamma \neq \gamma_0^0\). (iii) Let \(\Omega_1 (\tau, \gamma) \equiv E[f(\theta_{1\tau}^0(\tau, \gamma)' z_i (\gamma)|z_i z_i (\gamma) z_i (\gamma)'|]\), and \(\Omega_1 (\tau, \gamma)\) is positive definite for all \((\tau, \gamma) \in T \times \Gamma\).

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), Galvao et al. (2011), and Galvao et al. (2014). Assumptions A2-A5 specify standard conditions on threshold quantile regressions; see, e.g., Galvao et al. (2011). Assumption A6 specifies the identification condition which is needed to establish the consistency of \(\hat{\theta}_\tau\) under \(\mathbb{H}_{1\tau}\). In the special case of \(x_i = (1, \tau_i)'\), we can write \(\alpha_\tau = (\alpha_{0\tau}, \alpha_{1\tau})'\) and \(\beta_\tau = (\beta_{0\tau}, \beta_{1\tau})'\), where \(\alpha_{0\tau}\) and \(\beta_{0\tau}\) are the true values of the intercept parameters before and after the break, and \(\alpha_{1\tau}\) and \(\beta_{1\tau}\) are the true values of the slope parameters before and after the break. Let \(d_\tau \equiv (\alpha_{0\tau} - \beta_{0\tau}) + (\alpha_{1\tau} - \beta_{1\tau}) \gamma_0^0\). Then, a sufficient condition for Assumption A6 to hold is \(d_\tau \neq 0\). Assumption A7 specifies the magnitude of change in the coefficients, which is also made in Hansen (2000) and Caner (2002).

Assumption A8 is needed to study the asymptotic distributions of \(\hat{\theta}_{1\tau}\) and \(\hat{\gamma}_\tau\) under \(\mathbb{H}_{1\tau}\). It is also used to establish the asymptotic distribution of the likelihood ratio test statistic in Sections 2.4 and 2.5. Assumption A9 is needed to establish the asymptotic distribution of the likelihood ratio test statistic in Section 2.5. Observe that \(E[\rho_\tau(y_i - \theta_1^0 z_i (\gamma))]\) is convex in \(\theta_1\) for all \((\tau, \gamma) \in T \times \Gamma\), and \(\theta_{1\tau}^0(\tau, \gamma)\) in A9(i) exists and is uniquely defined. It is also continuous in \((\tau, \gamma)\) by an application of the maximum theorem. The first-order condition for the minimization of \(S(\theta_1; \tau, \gamma)\) with respect to \(\theta_1\) implies that

\[
E [\psi_\tau (y_i - \theta_{1\tau}^0(\tau, \gamma)' z_i (\gamma))] z_i (\gamma) = 0 \quad \text{for all} \quad (\tau, \gamma) \in T \times \Gamma,
\]  

(A.1)
where $\psi_r(u) = \tau - 1 \{u < 0\}$. (A.1) will be used in the proof of Theorem 2.5 below. The last part of A9(i) simply restricts the conditional quantile regression from sharing a common break $\gamma^0$, which does not depend on $\tau \in \mathcal{T}$. Like Assumption A6, A9(ii) is an identification condition and requires that $\gamma^0$ be the unique common break. A9(iii) extends A5. Note that $\Omega_1(\tau, \gamma^0) = \Omega_1(\tau, \gamma^0)$ under A9(i)-(ii).

### B Proofs of the Main Results

We first prove some technical lemmas that are used in the proof of the main results in Section 2.

**Lemma B.1** Suppose that Assumptions A1-A5 hold. Then, under $H_0$, $\hat{\theta}_1(\tau, \gamma)$ admits the following uniform Bahadur representation: $\sqrt{n}(\hat{\theta}_1(\tau, \gamma) - \theta^0_{1,\tau}) = \Omega_1(\tau, \gamma)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_r(y_i - \alpha_{0,\tau}^0 x_i) z_i(\gamma) + o_p(1)$, where $\psi_r(u) = \tau - 1 \{u < 0\}$, and $o_p(1)$ holds uniformly in $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.

**Proof.** See the proof of Theorem 1 in Galvao et al. (2014). ■

To state and prove the next three lemmas, we first define some notation. Recall that $\theta_r \equiv (\theta_{1,\tau}^r, \gamma_r)^r$, $\theta_{1,\tau} \equiv (\alpha_r^r, \beta_r^r)^r$, and the true value of $\theta_r$, $\theta_{1,\tau}$, $\alpha_r$, $\beta_r$, and $\gamma_r$ are denoted as $\theta^0_{r}$, $\theta^0_{1,\tau}$, $\alpha^0$, $\beta^0$, and $\gamma^0$, respectively. Let $\varepsilon_{1,\tau}(\theta_r) \equiv y_i - \theta_{1,\tau}^1 z_i(\gamma_r)$ and $\varepsilon_{i,\tau} = \varepsilon_{i,\tau}(\theta^0_{r})$. Note that the $\tau$th conditional quantile of $\varepsilon_{i,\tau}$ given $z_i$ is 0, i.e., $E[\psi_r(\varepsilon_{i,\tau}) | z_i] = 0$. Let

$$S_{n\tau}(\theta_r) = \sum_{i=1}^{n} \rho_r(y_i - \theta_{1,\tau}^1 z_i(\gamma_r)). \tag{B.1}$$

Let $\|\cdot\|$ denote the Euclidean norm. Note that for all $\theta_r \in \mathbb{R}^{2k+1}$, we have $\|z_i(\gamma_r)\| = \|x_i\|$, $\|\theta^1_{1,\tau} z_i(\gamma_r)\| \leq \|\theta^0_{1,\tau} z_i(\gamma_r)\| \leq \|\theta^0_{1,\tau} - \theta^0_{1,\tau}\| \|x_i\|$, and

$$\|z_i(\gamma_r) - z_i(\gamma^0_r)\| \leq \sqrt{2} \|x_i\| \{r_i \leq |\gamma^0_r - \gamma_r|\}. \tag{B.2}$$

Let $D_{n\tau}(\theta_{1,\tau}, \gamma_{r}) = S_{n\tau}(\theta_{1,\tau}, \gamma_{r}) - S_{n\tau}(\theta^0_{1,\tau}, \gamma^0_r)$, $D_{1n\tau}(w_1) = S_{n\tau}(\theta^0_{1,\tau} + n^{-1/2} w_1, \gamma^0_r) - S_{n\tau}(\theta^0_{1,\tau}, \gamma^0_r)$, and $D_{2n\tau}(\theta_{1,\tau}, \gamma_{r}) = S_{n\tau}(\theta_{1,\tau} + \gamma_{r}) - S_{n\tau}(\theta^0_{1,\tau}, \gamma^0_r)$. Then we have

$$D_{n\tau}(\theta_{1,\tau}, \gamma_{r}) = D_{1n\tau}(\theta_{1,\tau} + \gamma_{r}) + D_{2n\tau}(\theta_{1,\tau}, \gamma_{r}). \tag{B.3}$$

Let $D_{2n\tau}(w_1, w_2) = D_{2n\tau}(\theta^0_{1,\tau} + n^{-1/2} w_1, \gamma^0_r + n^{-1/2} w_2)$. We will study the asymptotic properties of $D_{1n\tau}(w_1)$ and $D_{2n\tau}(w_1, w_2)$ in Lemmas B.3 and B.4, respectively.

**Lemma B.2** Suppose that Assumptions A1-A2 and A4 hold. Then $\lim_{\eta \to 0} E \sup_{\theta^* \in N_\eta(\theta_r)} |\rho_r(y_i - \theta^*_{1,\tau} z_i(\gamma^*_r)) - \rho_r(y_i - \theta^*_{1,\tau} z_i(\gamma^*_r))| = 0$ for any $\theta_r \in \Theta$, where $N_\eta(\theta_r) \equiv \{\theta^*_{r} = (\theta^*_{1,\tau}, \gamma^*_r)^r \in \Theta : \|\theta^*_{1,\tau} - \theta_{1,\tau}\| < \eta, |\gamma^*_r - \gamma_r| < \eta\}$ denotes an $\eta$-neighborhood of $\theta_r \in \Theta$ and $\eta > 0$.

**Proof.** Let $\Delta_{1,\tau} = \theta^*_{1,\tau} z_i(\gamma^*_r) - \theta_{1,\tau}^1 z_i(\gamma_r)$. Then by the triangle inequality and (B.2),

$$|\Delta_{1,\tau}| \leq \left|\theta^*_{1,\tau} z_i(\gamma^*_r) - z_i(\gamma^*_r)\right| + \left|\theta_{1,\tau}^1 - \theta^*_{1,\tau}\right|^1 z_i(\gamma^*_r) \right| \leq \left\{\sqrt{2} \times 1 \{r_i \leq \gamma^*_r \wedge \gamma^*_r < r_i \leq \gamma^*_r \vee \gamma^*_r\} \|\theta^0_{1,\tau} - \theta^0_{1,\tau}\| \|x_i\| \right\} \|x_i\| \leq \left\{\sqrt{2} \times 1 \{|r_i - \gamma^*_r| \leq \eta\} \|\theta^0_{1,\tau} - \theta^0_{1,\tau}\| \|x_i\| \right\} \|x_i\|.$$
By Knight’s (1998) identity (see also Koenker (2005, p. 121)),

\[
\rho_\tau(y_i - \theta_1^i z_i (\gamma^*) - y_i - \theta_1^i z_i (\gamma_\tau)) = \rho_\tau(\varepsilon_i | \theta_\tau - \Delta_\tau) - \rho(\varepsilon_i | \theta_\tau)
\]

\[
= -\Delta_\tau \psi_\tau(\varepsilon_i | \theta_\tau) + \int_0^{\Delta_\tau} [1 \{u_{i\tau} \leq s\} - 1 \{u_{i\tau} \leq 0\}] \, ds.
\]

It follows that \(E \sup_{\theta_\tau \in \Omega} |\rho_\tau(y_i - \theta_1^i z_i (\gamma^*)) - \rho_\tau(y_i - \theta_1^i z_i (\gamma_\tau))| \leq 2E |\Delta_\tau| \leq 2\sqrt{2} \times |P(r_i - \gamma_\tau| \leq \eta)|^{1/2} \|x_i\|_2 \|\theta_1\| + 2\sqrt{2} \eta||x_i|| \rightarrow 0 \) as \( \eta \rightarrow 0 \), where \( ||x_i||_2 = \{E||x_i||^2\}^{1/2} \).

**Lemma B.3** Suppose Assumptions A1-A5 hold. Then \( \sup_{\|w_i\| \leq M} |D_{1\tau r}(w_1) + n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) \psi_\tau(\varepsilon_i) - \frac{1}{2} w_1^2 \Omega_{1\tau r} w_1| = o_p(1) \) for every \( M \in (0, \infty) \), where \( \Omega_{1\tau} = \Omega ((\tau, \gamma^0_\tau) = E(z_i (\gamma_\tau^0) z_i (\gamma_\tau^0) f(\theta_1^0 z_i (\gamma_\tau^0) | z_i)).

**Proof.** Let \( d_{1\tau r}(w_1) = D_{1\tau r}(w_1) + n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) \psi_\tau(\varepsilon_i) - \frac{1}{2} w_1^2 \Omega_{1\tau r} w_1 \). By Knight’s identity, \( D_{1\tau r}(w_1) = -n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) \psi_\tau(\varepsilon_i) + n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) \int_0^1 \pi_i(w_1, s) \, ds \), where \( \pi_i(w_1, s) = 1 \{\varepsilon_i \leq sn^{-1/2}w_1^2 z_i (\gamma_\tau^0) \} - 1 \{\varepsilon_i \leq 0\} \). It follows that \( d_{1\tau r}(w_1) = n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) \times \int_0^1 \pi_i(w_1, s) \, ds - \frac{1}{2} w_1^2 \Omega_{1\tau} w_1 = d_{1\tau r,1}(w_1) + d_{1\tau r,2}(w_1) \), where

\[
d_{1\tau r,1}(w_1) = \int_0^1 n^{-1/2} \sum_{i=1}^n w_1 z_i (\gamma_\tau^0) \{\pi_i(w_1, s) - E[\pi_i(w_1, s) | z_i]\} \, ds,
\]

\[
d_{1\tau r,2}(w_1) = \int_0^1 n^{-1/2} \sum_{i=1}^n \left\{ w_1 z_i (\gamma_\tau^0) E[\pi_i(w_1, s) | z_i] - sn^{-1/2}w_1^2 \Omega_{1\tau} w_1 \right\} \, ds.
\]

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

\[
\sup_{\|w_i\| \leq b} |d_{1\tau r,2}(w_1) |
\]

\[
= \sup_{\|w_i\| \leq b} n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) \int_0^1 \left[ F(\zeta_\tau + sn^{-1/2}w_1^2 z_i (\gamma_\tau^0) | z_i) - F(\zeta_\tau | z_i) \right] \, ds - \frac{1}{2} w_1^2 \Omega_{1\tau} w_1
\]

\[
\leq \sup_{\|w_i\| \leq b} n^{-1/2}w_1 \sum_{i=1}^n z_i (\gamma_\tau^0) z_i (\gamma_\tau^0) \int_0^1 \int_0^1 \left[ f (\zeta_\tau + sn^{-1/2}w_1^2 z_i (\gamma_\tau^0) | z_i) - f(\zeta_\tau | z_i) \right] \, dtds
\]

\[
+ \frac{1}{2} \sup_{\|w_i\| \leq M} n^{-1/2}w_1 \sum_{i=1}^n \left[ z_i (\gamma_\tau^0) z_i (\gamma_\tau^0) f(\zeta_\tau | z_i) - \Omega_{1\tau} \right] \, w_1
\]

\[
\leq M^2 n^{-1/2} \sum_{i=1}^n \left[ z_i (\gamma_\tau^0) z_i (\gamma_\tau^0) \right] \int_0^1 \int_0^1 \sup_{\|w_i\| \leq M} \left[ f (\zeta_\tau + sn^{-1/2}w_1^2 z_i (\gamma_\tau^0) | z_i) - f(\zeta_\tau | z_i) \right] \, dtds
\]

\[
+ \frac{M^2}{2} \sum_{i=1}^n \left[ z_i (\gamma_\tau^0) z_i (\gamma_\tau^0) f(\zeta_\tau | z_i) - \Omega_{1\tau} \right]
\]

\[
= o_p(1) + o_p(1) = o_p(1),
\]

where \( \zeta_\tau \equiv \theta_1^0 z_i (\gamma_\tau^0) \), and the last line follows from the Lebesgue dominated convergence theorem and the weak LLN for IID observations.
Lemma B.4  Suppose that Assumptions A1-A5 hold. Then \( \sup_{\|w_i\| \leq M} |\bar{D}_{2n\tau} (w_1, w_2) - \tilde{D}_{2n\tau} (0, w_2)| = o_p \left( 1 \right) \) for every \( M \in (0, \infty) \) and \( w_2 \in \mathbb{R} \).

Proof. Without loss of generality, we consider the case \( w_2 > 0 \). Let \( \Delta_i \equiv \theta_{1, \tau} - \theta_{2, \tau} \), \( a_1 (x) \equiv (0_{1 \times p}, x)' \), \( a_2 (x) \equiv (x', 0_{1 \times p})' \) and \( a (x) \equiv a_1 (x) - a_2 (x) \), where \( 0_{1 \times p} \) denotes a \( 1 \times p \) vector of zeros. Noting that \( z_i (\gamma_\tau) - z_i (\gamma_\tau^0) = -a (x_i) I_i (r) \) when \( r > 0 \), we observe that: (i) if \( I_i (r) = 0 \),

\[
y_i - \theta_{1, \tau} z_i (\gamma_\tau) = \epsilon_{i \tau} - \theta_{1, \tau}^{\prime} \left[ z_i (\gamma_\tau) - z_i (\gamma_\tau^0) \right] - \Delta_{1, \tau}^{\prime} \left[ z_i (\gamma_\tau) - z_i (\gamma_\tau^0) \right] - \Delta_{1, \tau}^{\prime} z_i (\gamma_\tau^0) = \epsilon_{i \tau} - \Delta_{1, \tau}^{\prime} z_i (\gamma_\tau^0) = 0,
\]

and (ii) if \( I_i (r) = 1 \), \( z_i (\gamma_\tau^0) = a_1 (x_i) \) and \( z_i (\gamma_\tau) = a_2 (x_i) \). It follows that if \( r > 0 \), then

\[
\rho_r \left( y_i - \theta_{1, \tau}^{\prime} z_i (\gamma_\tau) \right) - \rho_r \left( y_i - \theta_{1, \tau}^{\prime} z_i (\gamma_\tau^0) \right) = \left[ \rho_r \left( y_i - \theta_{1, \tau}^{\prime} z_i (\gamma_\tau) \right) - \rho_r \left( y_i - \theta_{1, \tau}^{\prime} z_i (\gamma_\tau^0) \right) \right] \mathbf{1}_i (r)
\]

\[
= \left[ \rho_r \left( \epsilon_{i \tau} - \theta_{1, \tau}^{\prime} \left[ z_i (\gamma_\tau) - z_i (\gamma_\tau^0) \right] - \Delta_{1, \tau}^{\prime} \left[ z_i (\gamma_\tau) - z_i (\gamma_\tau^0) \right] - \Delta_{1, \tau}^{\prime} z_i (\gamma_\tau^0) \right) - \rho_r \left( \epsilon_{i \tau} - \Delta_{1, \tau}^{\prime} z_i (\gamma_\tau^0) \right) \right] \mathbf{1}_i (r)
\]

\[
= \left[ \rho_r \left( \epsilon_{i \tau} + \theta_{1, \tau}^{\prime} a_1 (x_i) - \Delta_{1, \tau}^{\prime} a_2 (x_i) - \rho_r \left( \epsilon_{i \tau} - \Delta_{1, \tau}^{\prime} a_1 (x_i) \right) \right) \right] \mathbf{1}_i (r).
\]

(B.4)

[Similarly, if \( r < 0 \), then \( \rho_r \left( y_i - \theta_{1, \tau}^{\prime} m_{\gamma_\tau} (z_i) \right) - \rho_r \left( y_i - \theta_{1, \tau}^{\prime} m_{\gamma_\tau^0} (z_i) \right) = \rho_r \left( \epsilon_{i \tau} - \theta_{1, \tau}^{\prime} a_1 (x_i) - \Delta_{1, \tau}^{\prime} a_1 (x_i) \right) - \rho_r \left( \epsilon_{i \tau} - \Delta_{1, \tau}^{\prime} a_2 (x_i) \right) \mathbf{1}_i (r) \). By Knight’s identity, we have

\]

\[
\tilde{D}_{2n\tau} (w_1, w_2) - \bar{D}_{2n\tau} (0, w_2)
\]

\[
= \sum_{i=1}^{n} \left[ \rho_r \left( \epsilon_{i \tau} + \theta_{1, \tau}^{\prime} a_1 (x_i) - \Delta_{1, \tau}^{\prime} a_2 (x_i) \right) \right] \mathbf{1}_i \left( n^{-1/2} w_2 \right)
\]

\[
= n^{-1/2} \sum_{i=1}^{n} \psi_r (\epsilon_{i \tau}) w_i a_1 (x_i) \mathbf{1}_i \left( n^{-1/2} w_2 \right)
\]

\[
+ \sum_{i=1}^{n} \int_{0}^{n^{-1/2} w_2 a_1 (x_i)} \left[ 1 \{ \epsilon_{i \tau} \leq s \} - 1 \{ \epsilon_{i \tau} \leq 0 \} \right] d s \mathbf{1}_i \left( n^{-1/2} w_2 \right)
\]

\[
- \sum_{i=1}^{n} \int_{0}^{n^{-1/2} w_2 a_1 (x_i)} \left[ 1 \{ \epsilon_{i \tau} \leq s \} - 1 \{ \epsilon_{i \tau} \leq 0 \} \right] d s \mathbf{1}_i \left( n^{-1/2} w_2 \right)
\]

\[
\equiv d_{2n\tau, 1} (w_1) + d_{2n\tau, 2} (w_1) - d_{2n\tau, 3} (w_1), \text{ say.}
\]

To study the uniform bound for \( d_{2n\tau, 1} (w_1) \), we consider the class of functions \( \mathcal{F}_1 = \{ m_1 (y, z; w_1, \gamma) : w_1 \in \Theta_M, \gamma \in \Gamma \} \), where \( m_1 (y, z; w_1, \gamma) = \psi_r (y - \theta_{1, \tau}^{\prime} z (\gamma)) w_i a_1 (x_i) \) \( 1 \{ q \leq \gamma \} \), \( z (\gamma) = [x' 1 \{ q \leq \gamma \} x' 1 \{ q > \gamma \}]' \), and \( \Theta_M = \{ \omega_1 : \| \omega_1 \| \leq M \} \). Let

\[
\mathcal{F}_{1, 1} = \{ f_{1, 1} (y, z; \gamma) = 1 \{ q \leq \gamma \} : \gamma \in \Gamma \},
\]

\[
\mathcal{F}_{1, 2} = \{ f_{1, 2} (y, z; \gamma) = \tau - 1 \{ y - \theta_{1, \tau}^{\prime} z (\gamma) < 0 \} : \gamma \in \Gamma \}, \text{ and }
\]

\[
\mathcal{F}_{1, 3} = \{ f_{1, 3} (y, z; w_1) = w_i a_1 (x) : w_1 \in \Theta_M \}.
\]
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 $\theta^p_{1,r}z (\gamma) = \beta^p_{1,r}x + \delta^p_{1,r}x \{ q \leq \gamma \}$ where $\delta^p_{1,r} = \alpha^p_{1,r} - \beta^p_{1,r}$, $\mathcal{F}_{1,2}$ 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), where $C$ is a large constant. 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

\[
\sup_{\| w_1 \| \leq M} |d_{2n,1} (w_1)| = \sup_{\| w_1 \| \leq M} n^{-1/2} \sum_{i=1}^{n} \left[ m_1 (y_i, z_i; w_1, \gamma_0 + n^{-1+2a}w_2) - m_1 (y_i, z_i; w_1, \gamma_0) \right] = o_P (1) \tag{1}
\]

as $n^{-1+2a} \to 0$ as $n \to \infty$ by Assumption A7. Next we study $d_{2n,3} (w_1)$. Write $d_{2n,3} (w_1) = n^{-1/2} \sum_{i=1}^{n} w_i \times a_1 (x_i) \int_0^1 \left\{ \varepsilon_{ir} \leq n^{-1/2}w_i^r a_1 (x_i) s \right\} -1 \{ \varepsilon_{ir} \leq 0 \} ds \{ n^{-1+2a}w_2 \}$. Let $m_2 (y, z; w_1, \bar{w}_{1}) = w_i^r a_1 (x) \int_0^1 \left\{ y \leq \theta_{1,r} z (\gamma) \leq \bar{w}_i a_1 (x) s \right\} \{ \gamma < \gamma_0 + n^{-1+2a}w_2 \}$. 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 $\mathcal{F}_{2,1} = \{ f_2, z (y; \bar{w}_1) = \int_0^1 \left\{ y - \theta_{1,r} z (\gamma) \leq \bar{w}_i a_1 (x) s \right\} ds : \bar{w}_1 \in \Theta_M \}$.

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 $1 \left\{ \gamma < \gamma_0 + n^{-1+2a}w_2 \right\}$, it also satisfies that Pollard’s entropy condition and is stochastically equicontinuous with respect to the pseudometric defined by

\[
\rho ((w_1, \bar{w}_1), (w_1^*, \bar{w}_1^*)) = \left\{ E \left[ \frac{m_2 (y, z; w_1, \bar{w}_1) - m_2 (y, z; w_1^*, \bar{w}_1^*)}{\gamma_0} \right]^2 \right\}^{1/2}.
\]

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

\[
\sup_{\| w_1 \| \leq M} |d_{2n,3} (w_1)| \leq \sup_{\| w_1 \| \leq M} n^{-1/2} \sum_{i=1}^{n} E \left[ m_2 (y_i, z_i; w_1, n^{-1/2}w_1) - m_2 (y_i, z_i; w_1, 0) \right] \tag{1}
\]

\[
+ \sup_{\| w_1 \| \leq M} n^{-1/2} \sum_{i=1}^{n} \left[ \bar{m}_2 (y_i, z_i; w_1, n^{-1/2}w_1) - \bar{m}_2 (y_i, z_i; w_1, 0) \right] \tag{1}
\]

\[
= \sup_{\| w_1 \| \leq M} n^{-1/2} \sum_{i=1}^{n} E \left\{ w_i a_1 (x_i) \int_0^1 \left[ F \left( n^{-1/2}w_i^r a_1 (x_i) s \right) | z_i \right] - F (0 | z_i) \right\} ds \{ n^{-1+2a}w_2 \} + o_P (1) \tag{1}
\]

\[
\leq M n^{1/2} \left\| \| a_1 (x_i) \| \right\| \int_0^1 \left[ F \left( n^{-1/2}M \| a_1 (x_i) \| s | z_i \right] - F (0 | z_i) \right\} ds \left\| n^{-1+2a}w_2 \right\| + o_P (1) \tag{1}
\]

\[
= n^{1/2} O \left( n^{-1/2} \right) o \left( n^{-1+2a} \right) + o_P (1) = o_P (1),
\]

where $\left\| A \right\|_2 \equiv \{ E \| A \|^2 \}^{1/2}$. By the same token, we can show that $\sup_{\| w_1 \| \leq M} |d_{2n,2} (w_1)| = o_P (1)$ . This completes the proof of the lemma. ■
Proof of Theorem 2.1. Recall that $\mathcal{F}(T \times \Gamma)$ denotes the space of all bounded functions on $T \times \Gamma$ equipped with the uniform topology. From Lemma B.1, we can easily obtain that under $\mathbb{H}_0$,
\[ \sqrt{n} \left( \hat{\theta}_1(\tau, \gamma) - \theta_{1,\tau}^0 \right) \Rightarrow \Omega_1(\tau, \gamma)^{-1} W(\tau, \gamma) \text{ in } (\mathcal{F}(T \times \Gamma))^{2k}. \]
Note that $\delta(\tau, \gamma) = \hat{\beta}(\tau, \gamma) \beta(\tau, \gamma) = R \hat{\theta}_1(\tau, \gamma)$ and $\hat{V}(\tau, \gamma)$ is a uniformly consistent estimator of $V(\tau, \gamma)$ under the stated conditions. By the Slutsky theorem and the continuous mapping theorem (CMT), we have the desired result. 

Proof of Theorem 2.2. First, observe that $\hat{\varsigma}$ equipped with the uniform topology. From Lemma B.1, we can easily obtain that under $\mathbb{H}_0$, we have $\mathbb{H}_0$ stability for $\hat{\varsigma}$. Let

$$(\tau_0) \equiv \left\{ \begin{array}{l} n_{\tau,j} \equiv n_{\tau,j} \left( \begin{array}{l} \tau \left( \gamma_j \right) - \gamma_j^0 \left( \gamma_j \right) \end{array} \right) \end{array} \right\} \text{ and } \mathcal{F}(\tau, \gamma)) \Rightarrow \mathbb{H}_0, \quad \mathbb{H}_0 \text{ stability for } \hat{\varsigma}.$$ 

By Lemma B.2 $\mathbb{H}_0$ stability for $\hat{\varsigma}$. It follows that $\theta_{0}^j$ is the unique minimizer of $\varsigma_\tau(\theta_j)$ and $\hat{\theta}_\tau \to \hat{\theta}_\tau$ a.s.

Proof of Theorem 2.3. First, we follow the proof of Theorem 3.2 in Koul et al. (2003) and prove that

$$n^{1/2}(\hat{\theta}_\tau - \theta_{1,\tau}^0) = O_P(1) \quad \text{and} \quad n^{1-2a}(\hat{\varsigma}_\tau - \gamma_j^0) = O_P(1).$$

Let $\Omega(\sigma) \equiv \{ \theta_j \in \Theta : \| \theta_j - \theta_{1,\tau}^0 \| < \kappa, \| \gamma_j - \gamma_j^0 \| < \kappa \}$, where $\kappa \in (0, 1)$ can be chosen sufficiently small by Theorem 2.2. Let $b \in (0, \infty)$. Define

$$N_{1b} = \{ \theta_j \in \Omega(\sigma) : \| \gamma_j - \gamma_j^0 \| > bn^{2a-1} \} \quad \text{and} \quad N_{2b} = \{ \theta_j \in \Omega(\sigma) : \| \theta_j - \theta_{1,\tau}^0 \| > bn^{-1/2} \}.$$ 

Noting that $\inf_{\theta_j \in N_{1b} \cup N_{2b}} D_{n} (\theta_j, \gamma_j) = \min \{ \inf_{\theta_j \in N_{1b}} D_{n} (\theta_j, \gamma_j), \inf_{\theta_j \in N_{2b}} D_{n} (\theta_j, \gamma_j) \}$, 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

$$P \left( \inf_{\theta_j \in N_{j}} D_{n} (\theta_j, \gamma_j) > c_j \right) > 1 - \kappa \quad \text{for } n > n_0, \quad j = 1, 2,$$

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

$$\inf_{\theta_j \in N_{j}} D_{n} (\theta_j, \gamma_j) \geq \inf_{\theta_j \in N_{j}} D_{n} (\theta_j, \gamma_j) \left( n^{1/2} (\theta_j - \theta_{1,\tau}^0) \right) + \inf_{\theta_j \in N_{j}} D_{n} (\theta_j, \gamma_j) \equiv D_{n \tau,j,1} + D_{n \tau,j,2}, \quad \text{say},$$ 

it suffices to analyze $D_{n \tau,j,1}$, $D_{n \tau,j,2}$, $D_{n \tau,j,21}$, and $D_{n \tau,j,22}$. We first analyze $D_{n \tau,j,2}$. 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 $c_0 b_0 (\gamma_j^0) / 2 > c_1$ and that

$$P \left( \inf_{\theta_j \in N_{j}} D_{n} (\theta_j, \gamma_j) > c_0 \right) > 1 - \kappa / 2 \quad \text{for all } n > n_0.$$ 

(B.7)
where $K (r) \equiv E [1_i (r)]$ and $1_i (r) \equiv 1 \{ \gamma^0_i < r_i \leq \gamma^0_i + r \}$. Let $r \equiv \gamma_r - \gamma^0_r$ and $\Delta_{1r} \equiv \theta_{1r} - \theta^0_{1r}$. Without loss of generality, assume that $r > 0$. Then by (B.4) we can decompose $n^{-1} D_{2n\tau} (\theta_{1r}, \gamma_r)$ as follows

$$n^{-1} D_{2n\tau} (\theta_{1r}, \gamma_r) = n^{-1} \sum_{i=1}^{n} [\rho_r ( y_i - \theta^0_{1r} z_i (\gamma_r)) - \rho_r ( y_i - \theta^0_{1r} z_i (\gamma^0_r))]$$

$$= n^{-1} \sum_{i=1}^{n} [\rho_r ( \varepsilon_{ir} + \theta^0_{1r} a (x_i) - \Delta^0_{1r} a_1 (x_i)) - \rho_r ( \varepsilon_{ir} - \Delta^0_{1r} a_1 (x_i))] \, 1_i (r)$$

$$= d_{n\tau 1} (\theta_{1r}, r) + d_{n\tau 2} (\theta_{1r}, r) + d_{n\tau 3} (r) + d_{n\tau 4} (r) + d_{n\tau 5} (r) + d_{n\tau 6} (r),$$

where

$$d_{n\tau 1} (\theta_{1r}, r) = n^{-1} \sum_{i=1}^{n} [\rho_r ( \varepsilon_{ir}) - \rho_r ( \varepsilon_{ir} - \Delta^0_{1r} a_1 (x_i))] \, 1_i (r),$$

$$d_{n\tau 2} (\theta_{1r}, r) = n^{-1} \sum_{i=1}^{n} [\rho_r ( \varepsilon_{ir} + \theta^0_{1r} a (x_i) - \Delta^0_{1r} a_2 (x_i)) - \rho_r ( \varepsilon_{ir} - \theta^0_{1r} a (x_i))] \, 1_i (r),$$

$$d_{n\tau 3} (r) = n^{-1} \sum_{i=1}^{n} [\rho_r ( \varepsilon_{ir} + \theta^0_{1r} a (x_i)) - \rho_r ( \varepsilon_{ir}) - p (r_i)] \, 1_i (r),$$

$$d_{n\tau 4} (r) = n^{-1} \sum_{i=1}^{n} [p (r_i) - p (\gamma^0_r)] \, 1_i (r),$$

$$d_{n\tau 5} (r) = p (\gamma^0_r) \left[ K_n (r) - K (r) \right],$$

$$d_{n\tau 6} (r) = p (\gamma^0_r) K (r),$$

$p (\gamma^0_r) = E \{ \rho_r ( \varepsilon_{ir} + \theta^0_{1r} a (x_i)) - \rho_r ( \varepsilon_{ir}) \} | r_i = \gamma^0_r \}$ 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 $p (\gamma^0_r)$ is strictly positive and continuous in $\gamma$ under our assumptions. Following Koul et al. (2003) (see also 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 (1) below in order to analyze $d_{n\tau 4} (r)$ and $d_{n\tau 5} (r)$ because then

$$\sup_{B/n^{1-2a} \leq r \leq \sigma} | d_{n\tau 4} (r) / K (r) | \leq \sup_{0 \leq r \leq \sigma} \left| p (\gamma^0_r + r) - p (\gamma^0_r) \right| \sup_{B/n^{1-2a} \leq r \leq \sigma} | K_n (r) / K (r) | \to 0$$

and

$$\sup_{B/n^{1-2a} \leq r \leq \sigma} | d_{n\tau 5} (r) / K (r) | \leq C \sup_{B/n^{1-2a} \leq r \leq \sigma} | K_n (r) / K (r) | - 1 \to 0$$

by passing $n \to \infty$ and then $\sigma \to 0$, where $C$ is a large constant. Similarly, Claim (2) below is needed to show that $d_{n\tau 3} (r) / K (r)$ 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

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

where $R_n (r) = n^{-1} \sum_{i=1}^{n} |J (x_i, \varepsilon_{ir}) - E [J (x_i, \varepsilon_{ir}) | r_i = r]| 1_i (r)$ and $J (x_i, \varepsilon_{ir}) \equiv \rho_r ( \varepsilon_{ir} + \theta^0_{1r} a (x_i)) - \rho_r ( \varepsilon_{ir})$.

It follows that $n^{-1} D_{2n\tau} (\theta_{1r}, \gamma_r) / K (r) = p (\gamma^0_r) + o_P (1) > c > 0$ with probability approaching 1 as $n \to \infty$ uniformly in $\theta_r \in N_{1b}$, and thus (B.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.
To prove (i), by (B.5) it suffices to study the asymptotic behavior of $D_{nt}(\theta_{1\tau}, \gamma_{\tau})$ by restricting our attention to the case where $n^{1/2} ||\theta_{1\tau} - \theta_0^\tau|| \leq M$ and $n^{1-2a} |\gamma_{\tau} - \gamma_0^\tau| \leq M$ for some large but fixed positive number $M$. Let $w_{1n\tau} \equiv n^{1/2}(\theta_{1\tau} - \theta_0^\tau)$ and $w_{2n\tau} \equiv n^{1-2a} (\gamma_{\tau} - \gamma_0^\tau)$. Then by (B.3), $D_{nt}(\theta_{1\tau} + n^{-1/2}w_1, \gamma_{\tau}^0 + n^{-1+2a}w_2) = D_{1n\tau}(w_1) + D_{2n\tau}(w_1, w_2)$, where recall that $D_{2n\tau}(w_1, w_2) = D_{2n\tau}\left( {\theta_1^\tau + n^{-1/2}w_1, \gamma_0^\tau + n^{-1+2a}w_2} \right)$. By Lemmas B.3 and B.4, we have

$$D_{nt}\left( {\theta_1^\tau + n^{-1/2}w_1, \gamma_0^\tau + n^{-1+2a}w_2} \right) = D_{1n\tau}(w_1) + D_{2n\tau}(0, w_2) + o_P(1),$$

where $o_P(1)$ holds uniformly over the set $||w_1|| \leq M$ and $|w_2| \leq M$, $D_{1n\tau}(w_1) = -n^{-1/2}w_1' \sum^n_{i=1} z_i (\gamma_0^\tau) \psi_{\tau}(\varepsilon_{i\tau}) + \frac{1}{2} w_1' \Omega_{1\tau} w_1$, and

$$D_{2n\tau}(0, w_2) = \left\{ \begin{array}{ll}
\sum^n_{i=1} \left[ \rho_{\tau} (\varepsilon_{i\tau} + \theta_0^\tau a(x_i)) - \rho_{\tau}(\varepsilon_{i\tau}) \right] & \text{if } w_2 > 0 \\
\sum^n_{i=1} \left[ \rho_{\tau} (\varepsilon_{i\tau} - \theta_0^\tau a(x_i)) - \rho_{\tau}(\varepsilon_{i\tau}) \right] & \text{if } w_2 \leq 0.
\end{array} \right.$$  \hspace{1cm} (B.8)

Thus $D_{nt}(\theta_{1\tau}, \gamma_{\tau}) = D_{1n\tau}(w_1) + D_{2n\tau}(w_1, w_2) = D_{1n\tau}(w_1) + D_{2n\tau}(0, w_2) + o_P(1)$. Noting that $D_{1n\tau}(w_1)$ and $D_{2n\tau}(0, w_2)$ are free of $w_2$ and $w_1$, respectively, and $(w_1, w_2)$ is a minimizer of $D_{nt}(\theta_1^\tau + n^{-1/2}w_1, \gamma_0^\tau + n^{-1+2a}w_2)$ with respect to $(w_1, w_2)$, the asymptotic distribution of $w_{1n\tau}$ is determined by that of the minimizer of $D_{1n\tau}(w_1)$ with respect to $w_1$, and similarly the asymptotic distribution of $w_{2n\tau}$ is determined by that of the minimizer of $D_{2n\tau}(0, w_2)$ with respect to $w_2$. Noting that $D_{1n\tau}(w_1)$ is convex in $w_1$, we can readily apply the convexity lemma to obtain

$$w_{1n\tau} = n^{-1/2} \Omega_{1\tau}^{-1} \sum^n_{i=1} z_i (\gamma_0^\tau) \psi_{\tau}(\varepsilon_{i\tau}) + o_P(1) \overset{d}{\rightarrow} N(0, \Sigma, \tau) \cdot N(\theta_{0k	imes 1}, \tau (1-\tau) \Sigma (\tau, \gamma_0^\tau)), \hspace{1cm} (B.9)$$

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

We now prove (ii). By reversing the argument used to obtain (B.4) in Appendix C we find that it is convenient to rewrite $D_{2n\tau}(0, w_2)$ as

$$D_{2n\tau}(0, w_2) = \sum^n_{i=1} \left[ \rho_{\tau} (\varepsilon_{i\tau} + \theta_0^\tau z_i (\gamma_0^\tau) - \theta_0^\tau \varepsilon_{i\tau}) (\gamma_0^\tau + n^{-1+2a}w_2) \right] - \rho_{\tau}(\varepsilon_{i\tau}),$$

where

$$\Delta x_i (w_2) = x_i \left[ 1 \{ r_i \leq \gamma_0^\tau + n^{-1+2a}w_2 \} - 1 \{ r_i \leq \gamma_0^\tau \} \right] \text{ and } \delta_0^\tau = \delta_0^{nr} - \delta_0^{nr} = v_\tau n^{-a}. \text{ By Knight’s identity, } D_{2n\tau}(0, w_2) = \sum^n_{i=1} \psi_{\tau} (\varepsilon_{i\tau}) \delta_0^\tau \Delta x_i (w_2) + \sum^n_{i=1} \int_0^s -\delta_0^\tau \Delta x_i (w_2) \left[ 1 \{ \varepsilon_{i\tau} \leq s \} - 1 \{ \varepsilon_{i\tau} \leq 0 \} \right] ds \equiv D_{2n\tau,1}(w_2) + D_{2n\tau,2}(w_2) \text{, say. Assume that } w_2 > 0. \text{ Using arguments similar to those used in the proof of Lemma A.11 in Hansen (2000), we can readily show that }$$

$$D_{2n\tau,1}(w_2) \Rightarrow B_{\tau} (w_2),$$

where $B_{\tau} (w_2)$ is a Brownian motion with variance $E[B_{\tau} (1)^2] = \tau (1-\tau) v_\tau' E[x_i x'_i | r_i = \gamma_0^\tau v_\tau g (\gamma_0^\tau) = \lambda_\tau$. Analogously to the proof of Lemma B.4 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

\[
D_{2n_r,2} (w_2) = - \sum_{i=1}^{n} \delta_{\tau}^0 \Delta x_i (w_2) \int_{0}^{1} \left[ F (\theta_{1i}^0 z_i (\gamma_r^0) - s \delta_{\tau}^0 \Delta x_i (w_2) | z_i) - F (\theta_{1i}^0 z_i (\gamma_r^0) | z_i) \right] ds + o_p (1)
\]

\[
= \frac{1}{2} \sum_{i=1}^{n} \Delta x_i (w_2) \Delta x_i (w_2)' \delta_{\tau}^0 f (\theta_{1i}^0 z_i (\gamma_r^0) | z_i) + o_p (1)
\]

\[
= \frac{1}{2} \sum_{i=1}^{n} x_i x_i' f (\theta_{1i}^0 z_i (\gamma_r^0) | z_i) \left[ 1 \{ r_i \leq \gamma_r^0 + n^{-1+2a} w_2 \} - 1 \{ r_i \leq \gamma_r^0 \} \right] + o_p (1)
\]

\[
= \frac{1}{2} \mu_r |w_2| + o_p (1),
\]

where \( \mu_r = v_t E[f (\alpha_{r,x_i} | z_i) x_i x_i' | r_i = \gamma_r^0] \). Noting that \( B_r (\cdot) \) is a Brownian motion with variance \( \lambda_r \) and thus can be written as \( -\sqrt{\lambda_r} W_1 (\cdot) \) with \( W_1 (\cdot) \) being a standard Brownian motion, we have,

\[
D_{2n_r} (0, w_2) \Rightarrow \frac{1}{2} \mu_r |w_2| - \sqrt{\lambda_r} W_1 (w_2) \text{ if } w_2 > 0.
\]

(B.10)

Similarly, we can show that

\[
D_{2n_r} (0, w_2) \Rightarrow \frac{1}{2} \mu_r |w_2| - \sqrt{\lambda_r} W_2 (-w_2) \text{ if } w_2 \leq 0,
\]

(B.11)

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

\[
w_{2n_r} \overset{d}{\to} \arg \max_{-\infty < w_2 < \infty} \left\{ - \frac{1}{2} \mu_r |w_2| + \sqrt{\lambda_r} W (w_2) \right\}
\]

\[
= \frac{\lambda_r}{\mu_r^2} \arg \max_{-\infty < r < \infty} \left\{ \frac{\lambda_r}{\mu_r^2} |r| + \sqrt{\lambda_r} W (r) \right\}.
\]

(B.12)

by the change of variables \( w_2 = (\lambda_r / \mu_r^2) r \) and the distributional equality \( W (c^2 r) \equiv c W (r) \).

**Proof of Theorem 2.4.** Recall that \( w_{1n_r} \equiv n^{-1/2} (\hat{\theta}_{1r} - \theta_{1r}^0) \) and \( w_{2n_r} \equiv n^{-1-2a} (\gamma_r - \gamma_r^0) \). By (B.3), the relationship between \( D_{2n_r} \) and \( D_{2n_r} \), and Lemma B.4, we have \( LR_{n_r} (\gamma_r^0) = D_{2n_r} (w_{1n_r}, 0) - D_{2n_r} (w_{1n_r}, w_{2n_r}) \) \( = -D_{2n_r} (0, w_{2n_r}) + o_p (1) \). By the analysis of \( D_{2n_r} (0, w_2) \) in the the proof of Theorem 2.3(ii) and the CMT, implies that

\[
LR_{n_r} (\gamma_r^0) = -D_{2n_r} (0, w_{2n_r}) + o_p (1) \overset{d}{\to} \sup_{w_2} \left\{ - \frac{1}{2} \mu_r |w_2| + \sqrt{\lambda_r} W (w_2) \right\}
\]

(B.13)

By the change of variables \( w_2 = (\lambda_r / \mu_r^2) r \) and the distributional equality \( W (c^2 r) \equiv c W (r) \), we have

\[
\sup_{w_2} \left\{ - \frac{1}{2} \mu_r |w_2| + \sqrt{\lambda_r} W (w_2) \right\} = \sup_r \left\{ - \frac{1}{2} \mu_r \left| \frac{\lambda_r}{\mu_r^2} r \right| + \sqrt{\lambda_r} W \left( \frac{\lambda_r}{\mu_r^2} r \right) \right\}
\]

\[
= \frac{\lambda_r}{\mu_r} \sup_r \left\{ - \frac{1}{2} |r| + W (r) \right\}.
\]

(B.14)

Consequently, we have \( LR_{n_r} (\gamma_r^0) \overset{d}{\to} \frac{\lambda_r}{2 \mu_r} \sup_r \{- |r| + 2 W (r)\} \).

31
Proof of Theorem 2.5. (i) First, observe that $\hat{\theta}_1(\tau, \gamma)$ defined in (2.9) is also minimizing $\bar{S}_n(\theta_1, \tau, \gamma) \equiv n^{-1}\sum_{i=1}^{n} s(y_i, z_i; \theta_1, \tau, \gamma)$ with respect to $\theta_1$, where $s(y_i, z_i; \theta_1, \tau, \gamma) \equiv \rho_\tau(y_i - \theta_1^0(\tau, \gamma) z_i)(\gamma)$. Let $\zeta(\theta_1, \gamma, \tau) \equiv E[s(y_i, z_i; \theta_1, \tau, \gamma)]$. By Lemma B.2, $\zeta(\theta_1, \gamma, \tau)$ is continuous in $(\theta_1, \gamma)$. It is straightforward to show that it is also continuous in $\tau$. Thus $\zeta(\theta_1, \tau, \gamma)$ is continuous over $\Theta_1 \times T \times \Gamma$. By Lemma 2 in Galvao et al. (2011), $\{s(y_i, z_i; \theta_1, \gamma, \tau) : (\theta_1, \gamma, \tau) \in \Theta_1 \times T \times \Gamma\}$ is Glivenko-Cantelli. In conjunction with the pointwise convergence, this implies that $\sup_{(\theta_1, \gamma, \tau) \in \Theta_1 \times T \times \Gamma} |\bar{S}_n(\theta_1, \tau, \gamma) - \zeta(\theta_1, \tau, \gamma)| \overset{P}{\to} 0$. As remarked after Assumption A9, $\theta_1^0(\tau, \gamma) = \arg \min_{\theta_1 \in \Theta_1} \zeta(\theta_1, \tau, \gamma)$ is uniquely defined. It follows from Lemma B.1 of Chernozhukov and Hansen (2006) that

$$\sup_{(\tau, \gamma) \in T \times \Gamma} \|\hat{\theta}_1(\tau, \gamma) - \theta_1^0(\tau, \gamma)\| = o_P(1). \quad (B.15)$$

Let $D_n(\gamma) \equiv \bar{S}_{n, \tau}(\gamma) - \sum_{i=1}^{n} \int \rho_\tau(y_i - \theta_1^0(\tau, \gamma) z_i)(\gamma) d\Pi(\tau)$. (B.15) implies that uniformly over $\Gamma$:

$$n^{-1}D_n(\gamma) = n^{-1} \sum_{i=1}^{n} \int s(y_i, z_i; \hat{\theta}_1(\tau, \gamma), \tau, \gamma) d\Pi(\tau)$$

$$= n^{-1} \sum_{i=1}^{n} \int \left[ \rho_\tau(y_i - \hat{\theta}_1(\tau, \gamma) z_i(\gamma)) - \rho_\tau(y_i - \theta_1^0(\tau, \gamma) z_i(\gamma)) \right] d\Pi(\tau)$$

$$= n^{-1} \sum_{i=1}^{n} \int \left[ \rho_\tau(y_i - \theta_1^0(\tau, \gamma) z_i(\gamma)) - \rho_\tau(y_i - \theta_1^0(\tau, \gamma) z_i(\gamma)) \right] d\Pi(\tau) + o_P(1)$$

$$= D(\gamma) + o_P(1)$$

where $D(\gamma) = \int E[\rho_\tau(\xi_{ir} - (\theta_1^0(\tau, \gamma) z_i(\gamma) - \theta_1(\tau, \gamma) z_i(\gamma))) - \rho_\tau(\xi_{ir})] d\Pi(\tau)$ and $\xi_{ir} = y_i - \theta_1^0(\tau, \gamma) z_i(\gamma)$. Again, by the fact that $E[\rho_\tau(\xi_{ir} - c) - \rho_\tau(\xi_{ir})] > 0$ for all $c \neq 0$, $D(\gamma)$ is minimized iff $\theta_1^0(\tau, \gamma) z_i(\gamma) = \theta_1^0(\tau, \gamma) z_i(\gamma)$ a.s., i.e., $\tau, \gamma \in \gamma$ 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 (B.15) and the continuity of $\theta_1^0(\tau, \gamma)$, implies that $\hat{\theta}_1(\tau, \gamma) = \theta_1^0(\tau, \gamma) + o_P(1) = \theta_1^0(\tau, \gamma) + o_P(1)$.

(ii) By the computational property of quantile regression (e.g., Lemma A2 in Ruppert and Carroll (1980)), uniformly in $(\tau, \gamma)$ $o_P(1) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau(y_i - \hat{\theta}_1(\tau, \gamma)' z_i(\gamma)) z_i(\gamma)$. For $u = (y, z)'$, define the map $f \mapsto G_n(f(u)) \equiv n^{-1/2} \sum_{i=1}^{n} \{f(u_i) - E[f(u_i)]\}$ for any measurable function $f$. Let $f_1(u; \theta_1, \gamma) = \psi_\tau(y_i - \theta_1^0(\gamma) z_i(\gamma)) z_i(\gamma)$. Noting that $\{f_1(u; \theta_1, \gamma) : (\theta_1, \gamma) \in \Theta_1 \times T \times \Gamma\}$ is Euclidean and $E[f_1(u; \theta_1, \gamma), \gamma] = E\{f(\theta_1^0(\tau, \gamma)' z_i(\gamma)) z_i(\gamma)\} - E[f(\theta_1^0(\tau, \gamma)' z_i(\gamma)) z_i(\gamma)]$ → $-E[f(\theta_1^0(\tau, \gamma))]' (\theta_1(\tau, \gamma) - \theta_1^0(\tau, \gamma))$ when $\theta_1(\tau, \gamma) \rightarrow \theta_1^0(\tau, \gamma)$ uniformly in $(\tau, \gamma)$, we have

$$o_P(1) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau(y_i - \hat{\theta}_1(\tau, \gamma)' z_i(\gamma)) z_i(\gamma)$$

$$= G_n f_1(u; \hat{\theta}_1(\tau, \gamma), \tau, \gamma) + \sqrt{n} E[f_1(u; \theta_1, \tau, \gamma)]|_{\theta_1=\hat{\theta}_1(\tau, \gamma)}$$

$$= G_n f_1(u; \theta_1^0(\tau, \gamma), \tau, \gamma) + o_P(1) + \sqrt{n} E[f_1(u; \theta_1, \tau, \gamma)]|_{\theta_1=\hat{\theta}_1(\tau, \gamma)}$$

$$= G_n f_1(u; \theta_1^0(\tau, \gamma), \tau, \gamma) + o_P(1) - \Omega_1(\tau, \gamma) \sqrt{n} \left\{\hat{\theta}_1(\tau, \gamma) - \theta_1^0(\tau, \gamma)\right\}.$$
where we have used the fact $E[\psi_\tau(y_1-\theta_1^0 (\tau, \gamma)^T z_i (\gamma))|z_i] = 0$ by (A.1) and the last $o_P (1)$ holds uniformly over $T \times \Gamma$.

(iii) and (iv) Let $w_1 (\tau) = \sqrt{n} (\theta_1 - \theta_1^0)$, $w_2 = n^{-1/2} (\gamma_1 - \gamma_0)$, $w_{1n} (\tau) = \sqrt{n} (\hat{\theta}_1 - \theta_1^0)$ and $w_{2n} = n^{-1/2} (\hat{\gamma}_1 - \gamma_0)$. Let $\{\theta_{1r}\}$ denote $\{\theta_{1r}\}_{r \in T}$, and similarly for $\{\hat{\theta}_{1r}\}$. As in the proof of Theorem 2.3, we continue to use the decomposition for $D_{nr} (\theta_{1r}, \gamma)$ in (B.3), where the only difference is that $\gamma$ and $\hat{\gamma}$ are now $\tau$-invariant. Noting that $(\{\hat{\theta}_{1r}\}, \hat{\gamma}) = \text{argmin}_{(\theta_{1r}, \gamma)} \int D_{nr} (\theta_{1r}, \gamma) d\Pi (\tau)$, we have

$$(w_{1n} (\cdot), w_{2n}) = \text{argmin}_{w_1 (\cdot), w_2} \left[ D_{1n} (w_1) + D_{2n} (w_1, w_2) \right],$$

where $D_{1n} (w_1) = \int D_{1nr} (w_1 (\tau)) d\Pi (\tau)$ and $D_{2n} (w_1, w_2) = \int D_{2nr} (w_1 (\tau), w_2) d\Pi (\tau)$. It is easy to see that the results in Lemmata B.3 and B.4 can be strengthened to hold uniformly in $(\tau, w_1)$ over any compact set. It follows that

$$D_{1n} (w_1) = -n^{-1/2} \int w_1 (\tau)^T \sum_{i=1}^n z_i (\gamma) \psi_\tau (\varepsilon_i) d\Pi (\tau) + \frac{1}{2} \int w_1 (\tau)^T \Omega_{1n} w_1 (\tau) d\Pi (\tau) + o_P (1), \quad (B.17)$$

and

$$D_{2n} (w_1, w_2) = D_{2n} (0, w_2) + o_P (1) = \int \sum_{i=1}^n \left[ \rho_\tau (\varepsilon_i + \delta_\gamma \Delta x_i (w_2)) - \rho_\tau (\varepsilon_i) \right] d\Pi (\tau) + o_P (1). \quad (B.18)$$

Then the rest of the proof follows directly from the proof of Theorem 2.3 with obvious modification. ■

**Proof of Theorem 2.6.** Recall that $w_{1n} = n^{-1/2} (\hat{\theta}_{1r} - \theta_1^0)$ and $w_{2n} = n^{-1/2} (\hat{\gamma}_1 - \gamma_0)$. Then $LR_n = \int [\tilde{D}_{2n} (w_{1n}, \tilde{w}_{2n}) - D_{2n} (w_{1n}, w_{2n})] d\Pi (\tau)$. The result in Lemma B.4 continues to hold when $\sup_{\|w_1\| \leq M}$ is replaced by $\sup_{\|w_1\| \leq M} \sup_{\tau \in T}$. It follows that

$$LR_n = \int \left[ \tilde{D}_{2n} (0, \tilde{w}_{2n}) - D_{2n} (0, w_{2n}) \right] d\Pi (\tau) + o_P (1). \quad (B.19)$$

Following the proof of Theorem 2.3(ii), we can readily show that $\tilde{D}_{2n} (0, w_2) \Rightarrow \mu \int w_2 + B (\tau, w_2)$, where $B (\tau, w_2)$ is a two-dimensional Gaussian process with the covariance kernel

$$Y (\tau, \bar{\tau}; w_1, \bar{w}_2) \equiv E \left[ B (\tau, w_2) B (\bar{\tau}, \bar{w}_2) \right] = \lambda (\tau, \bar{\tau}) \int b (w_2, w_2),$$

where $\lambda (\tau, \bar{\tau}) = \psi_\tau (x) E[x|x_i |r_i = \gamma \theta_1^0] v_{\gamma_0} (\gamma^0)$, and $b (w_2, \bar{w}_2) = (|w_2| \wedge |\bar{w}_2|) 1 (w_2 \bar{w}_2 \geq 0)$. For fixed $\tau$, $B (\tau, \cdot)$ reduces to the one-dimensional Gaussian process $B_\tau (\cdot)$ defined in the proof of Theorem 2.3(ii). As before, we can write $B_\tau (w_2) = -\sqrt{\lambda_\tau} W (w_2)$ where $W (\cdot)$ is a two-sided Brownian motion. Following the proof of Theorem 2.4, we can readily show that

$$- \int D_{2n} (0, \tilde{w}_{2n}) d\Pi (\tau) \Rightarrow \frac{\lambda_\tau}{2 \mu_0} \sup_{\tau} \left\{ - |r| + 2 W (r) \right\}. \quad (B.20)$$

This, in conjunction with (B.13)-(B.14), implies that $LR_n = c_{LR} \Xi + o_P (1)$, where $c_{LR} = \int \frac{\lambda_\tau}{2 \mu_0} d\Pi (\tau) - \frac{\lambda_\tau}{2 \mu_0}$ and $\Xi = \sup_{-\infty < r < \infty} \left\{ - |r| + 2 W (r) \right\}$. By the Cauchy-Schwarz inequality, one can easily verify that $c_{LR} \geq 0$ and the equality holds if and only if $\lambda_\tau = \mu_0^2$ a.s.-$\Pi$. ■
Proof of Theorem 2.7. Recall that $\tilde{w}_2^n \equiv n^{1-2a} (\gamma - \gamma^0)$. Following closely the proof of Theorem 2.4, we obtain $L \tilde{R}_n (\gamma^0) = - \int \tilde{D}_{2n^2} (0, w_{2n}) d\Pi (\tau) + o_P (1)$. The result then follows by arguments analogous to those used in the proof of Theorem 2.4. ■

Acknowledgements

The authors are grateful to an associate editor and two anonymous referees for their useful comments and suggestions. They sincerely thank Jacques Crémer, Han Hong, Thomas Lemieux, Thierry Magnac, Michael Peters, and Wing Suen for their many useful comments and Heng Ju for motivation and discussion on the subject. They are also thankful to the participants in the seminars at Peking University, University of Macau, the 2011 Summer International Econometrics Symposium in Chengdu, the Shanghai Workshop on Industrial Organization and Competition Policy, the 2012 Asian meeting of Econometric Society, and the 45th annual conference of the Canadian Economic Association.

Funding

Su gratefully acknowledges support from the Singapore Ministry of Education for Tier-2 Academic Research Fund (ARcF) under grant number MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. Xu gratefully acknowledges the generous financial support by the University of Hong Kong. The work was partially accomplished when Su was visiting HKU under the visiting scholar scheme.

References


Davies, R. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64: 247-254.
Davies, R. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74: 33-43.


Table 1: Rejection frequency for the test of existence of a threshold effect

<table>
<thead>
<tr>
<th>( T ) \ Level</th>
<th>( n = 200 )</th>
<th>( 1% )</th>
<th>( 5% )</th>
<th>( 10% )</th>
<th>( n = 400 )</th>
<th>( 1% )</th>
<th>( 5% )</th>
<th>( 10% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>No threshold effect (( c_0 = 0, c = 0 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.004</td>
<td>0.028</td>
<td>0.042</td>
<td>0.018</td>
<td>0.040</td>
<td>0.072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.010</td>
<td>0.034</td>
<td>0.050</td>
<td>0.026</td>
<td>0.050</td>
<td>0.072</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.018</td>
<td>0.054</td>
<td>0.068</td>
<td>0.032</td>
<td>0.050</td>
<td>0.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.024</td>
<td>0.052</td>
<td>0.080</td>
<td>0.036</td>
<td>0.068</td>
<td>0.100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.020</td>
<td>0.064</td>
<td>0.092</td>
<td>0.020</td>
<td>0.064</td>
<td>0.088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.040</td>
<td>0.068</td>
<td>0.098</td>
<td>0.026</td>
<td>0.054</td>
<td>0.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.030</td>
<td>0.062</td>
<td>0.084</td>
<td>0.022</td>
<td>0.048</td>
<td>0.078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.018</td>
<td>0.044</td>
<td>0.068</td>
<td>0.016</td>
<td>0.044</td>
<td>0.062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.008</td>
<td>0.026</td>
<td>0.054</td>
<td>0.012</td>
<td>0.056</td>
<td>0.094</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[0.1,0.9]</td>
<td>0.010</td>
<td>0.030</td>
<td>0.056</td>
<td>0.030</td>
<td>0.062</td>
<td>0.086</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common threshold value (( c_0 = 0.5, c = 0 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.050</td>
<td>0.102</td>
<td>0.148</td>
<td>0.084</td>
<td>0.198</td>
<td>0.272</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.040</td>
<td>0.104</td>
<td>0.146</td>
<td>0.114</td>
<td>0.204</td>
<td>0.286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.108</td>
<td>0.182</td>
<td>0.250</td>
<td>0.140</td>
<td>0.254</td>
<td>0.338</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.110</td>
<td>0.200</td>
<td>0.268</td>
<td>0.184</td>
<td>0.310</td>
<td>0.394</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.090</td>
<td>0.194</td>
<td>0.270</td>
<td>0.196</td>
<td>0.328</td>
<td>0.380</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.104</td>
<td>0.190</td>
<td>0.248</td>
<td>0.160</td>
<td>0.292</td>
<td>0.344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.082</td>
<td>0.158</td>
<td>0.220</td>
<td>0.120</td>
<td>0.288</td>
<td>0.348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.042</td>
<td>0.102</td>
<td>0.148</td>
<td>0.110</td>
<td>0.204</td>
<td>0.278</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.032</td>
<td>0.082</td>
<td>0.122</td>
<td>0.106</td>
<td>0.184</td>
<td>0.260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[0.1,0.9]</td>
<td>0.064</td>
<td>0.118</td>
<td>0.150</td>
<td>0.144</td>
<td>0.280</td>
<td>0.346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-common threshold value (( c_0 = 0.5, c = 0.5 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.058</td>
<td>0.126</td>
<td>0.172</td>
<td>0.124</td>
<td>0.226</td>
<td>0.312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.058</td>
<td>0.128</td>
<td>0.182</td>
<td>0.136</td>
<td>0.216</td>
<td>0.282</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.116</td>
<td>0.208</td>
<td>0.270</td>
<td>0.156</td>
<td>0.254</td>
<td>0.336</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.116</td>
<td>0.212</td>
<td>0.272</td>
<td>0.212</td>
<td>0.354</td>
<td>0.416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.094</td>
<td>0.210</td>
<td>0.274</td>
<td>0.188</td>
<td>0.330</td>
<td>0.402</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.094</td>
<td>0.172</td>
<td>0.242</td>
<td>0.156</td>
<td>0.252</td>
<td>0.332</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.070</td>
<td>0.128</td>
<td>0.182</td>
<td>0.096</td>
<td>0.232</td>
<td>0.312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.032</td>
<td>0.076</td>
<td>0.112</td>
<td>0.072</td>
<td>0.160</td>
<td>0.232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.022</td>
<td>0.056</td>
<td>0.088</td>
<td>0.056</td>
<td>0.122</td>
<td>0.178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[0.1,0.9]</td>
<td>0.052</td>
<td>0.136</td>
<td>0.174</td>
<td>0.148</td>
<td>0.268</td>
<td>0.340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$ \ Level</td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Common threshold value ($c_0 = 1$, $c = 0$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.204</td>
<td>0.390</td>
<td>0.480</td>
<td>0.542</td>
<td>0.740</td>
<td>0.812</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.258</td>
<td>0.460</td>
<td>0.546</td>
<td>0.720</td>
<td>0.856</td>
<td>0.894</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.422</td>
<td>0.596</td>
<td>0.680</td>
<td>0.788</td>
<td>0.902</td>
<td>0.944</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.474</td>
<td>0.652</td>
<td>0.734</td>
<td>0.854</td>
<td>0.930</td>
<td>0.954</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.508</td>
<td>0.666</td>
<td>0.744</td>
<td>0.852</td>
<td>0.926</td>
<td>0.942</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.522</td>
<td>0.692</td>
<td>0.756</td>
<td>0.830</td>
<td>0.924</td>
<td>0.946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.452</td>
<td>0.610</td>
<td>0.680</td>
<td>0.770</td>
<td>0.892</td>
<td>0.926</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.334</td>
<td>0.496</td>
<td>0.580</td>
<td>0.666</td>
<td>0.810</td>
<td>0.864</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.248</td>
<td>0.416</td>
<td>0.524</td>
<td>0.550</td>
<td>0.740</td>
<td>0.802</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$[0.1,0.9]$</td>
<td>0.374</td>
<td>0.550</td>
<td>0.648</td>
<td>0.848</td>
<td>0.940</td>
<td>0.960</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-common threshold value ($c_0 = 1$, $c = 0.5$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.250</td>
<td>0.440</td>
<td>0.520</td>
<td>0.576</td>
<td>0.706</td>
<td>0.780</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.358</td>
<td>0.532</td>
<td>0.612</td>
<td>0.686</td>
<td>0.844</td>
<td>0.886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.526</td>
<td>0.668</td>
<td>0.740</td>
<td>0.836</td>
<td>0.916</td>
<td>0.942</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>0.588</td>
<td>0.738</td>
<td>0.806</td>
<td>0.858</td>
<td>0.934</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.562</td>
<td>0.690</td>
<td>0.774</td>
<td>0.852</td>
<td>0.924</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.476</td>
<td>0.638</td>
<td>0.696</td>
<td>0.798</td>
<td>0.896</td>
<td>0.926</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.350</td>
<td>0.524</td>
<td>0.622</td>
<td>0.726</td>
<td>0.856</td>
<td>0.900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.224</td>
<td>0.412</td>
<td>0.504</td>
<td>0.552</td>
<td>0.766</td>
<td>0.836</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.114</td>
<td>0.248</td>
<td>0.344</td>
<td>0.366</td>
<td>0.564</td>
<td>0.692</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$[0.1,0.9]$</td>
<td>0.410</td>
<td>0.570</td>
<td>0.626</td>
<td>0.852</td>
<td>0.938</td>
<td>0.972</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Common threshold value ($c_0 = 2$, $c = 0$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.956</td>
<td>0.994</td>
<td>0.998</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.986</td>
<td>0.996</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.992</td>
<td>0.998</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.96</td>
<td>0.99</td>
<td>0.994</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$[0.1,0.9]$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Non-common threshold value ($c_0 = 2$, $c = 0.5$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.938</td>
<td>0.972</td>
<td>0.980</td>
<td>0.994</td>
<td>0.998</td>
<td>0.998</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>0.984</td>
<td>0.994</td>
<td>0.996</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>0.996</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.988</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.960</td>
<td>0.982</td>
<td>0.994</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>0.880</td>
<td>0.950</td>
<td>0.970</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$[0.1,0.9]$</td>
<td>0.994</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: MSE of estimates of the quantile regression coefficients in the presence of threshold effect

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\alpha_\tau$</th>
<th>$\beta_\tau$</th>
<th>$\alpha_\tau$</th>
<th>$\beta_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>intercept</td>
<td>slope</td>
<td>intercept</td>
<td>slope</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>$n = 400$</td>
<td>$n = 200$</td>
<td>$n = 400$</td>
</tr>
</tbody>
</table>

Common threshold value ($c_0 = 0.5, c = 0$)

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\alpha_\tau$</th>
<th>$\beta_\tau$</th>
<th>$\alpha_\tau$</th>
<th>$\beta_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.322</td>
<td>1.764</td>
<td>0.542</td>
<td>2.767</td>
</tr>
<tr>
<td>0.2</td>
<td>0.211</td>
<td>1.208</td>
<td>0.228</td>
<td>1.242</td>
</tr>
<tr>
<td>0.3</td>
<td>0.176</td>
<td>0.979</td>
<td>0.207</td>
<td>1.057</td>
</tr>
<tr>
<td>0.4</td>
<td>0.169</td>
<td>0.902</td>
<td>0.205</td>
<td>1.053</td>
</tr>
<tr>
<td>0.5</td>
<td>0.166</td>
<td>0.890</td>
<td>0.214</td>
<td>1.070</td>
</tr>
<tr>
<td>0.6</td>
<td>0.170</td>
<td>0.908</td>
<td>0.199</td>
<td>0.999</td>
</tr>
<tr>
<td>0.7</td>
<td>0.194</td>
<td>0.952</td>
<td>0.221</td>
<td>1.207</td>
</tr>
<tr>
<td>0.8</td>
<td>0.235</td>
<td>1.251</td>
<td>0.314</td>
<td>1.719</td>
</tr>
<tr>
<td>0.9</td>
<td>0.300</td>
<td>1.559</td>
<td>0.797</td>
<td>4.898</td>
</tr>
</tbody>
</table>

Non-common threshold value ($c_0 = 0.5, c = 0.5$)

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\alpha_\tau$</th>
<th>$\beta_\tau$</th>
<th>$\alpha_\tau$</th>
<th>$\beta_\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.301</td>
<td>1.662</td>
<td>0.343</td>
<td>1.735</td>
</tr>
<tr>
<td>0.2</td>
<td>0.194</td>
<td>1.025</td>
<td>0.301</td>
<td>1.556</td>
</tr>
<tr>
<td>0.3</td>
<td>0.155</td>
<td>0.870</td>
<td>0.252</td>
<td>1.205</td>
</tr>
<tr>
<td>0.4</td>
<td>0.166</td>
<td>0.904</td>
<td>0.231</td>
<td>1.146</td>
</tr>
<tr>
<td>0.5</td>
<td>0.160</td>
<td>0.904</td>
<td>0.210</td>
<td>1.104</td>
</tr>
<tr>
<td>0.6</td>
<td>0.162</td>
<td>0.875</td>
<td>0.204</td>
<td>0.987</td>
</tr>
<tr>
<td>0.7</td>
<td>0.203</td>
<td>1.047</td>
<td>0.231</td>
<td>1.211</td>
</tr>
<tr>
<td>0.8</td>
<td>0.215</td>
<td>1.148</td>
<td>0.401</td>
<td>1.966</td>
</tr>
<tr>
<td>0.9</td>
<td>0.295</td>
<td>1.502</td>
<td>0.440</td>
<td>2.175</td>
</tr>
</tbody>
</table>

Figure 1: The power function of the $NLR_n$ test for a common threshold value
Table 3: Coverage ratio for the 95% confidence interval of $\gamma_\tau$ and its average length

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$\gamma_\tau$</th>
<th>coverage ratio</th>
<th>average length</th>
<th>coverage ratio</th>
<th>average length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 200</td>
<td></td>
<td></td>
<td>n = 400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Common threshold value ($c_0 = 0.5$, $c = 0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.5</td>
<td>0.996</td>
<td>0.681</td>
<td>0.976</td>
<td>0.674</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5</td>
<td>0.982</td>
<td>0.659</td>
<td>0.980</td>
<td>0.656</td>
</tr>
<tr>
<td>0.3</td>
<td>0.5</td>
<td>0.970</td>
<td>0.637</td>
<td>0.960</td>
<td>0.609</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5</td>
<td>0.974</td>
<td>0.610</td>
<td>0.946</td>
<td>0.567</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.948</td>
<td>0.594</td>
<td>0.946</td>
<td>0.527</td>
</tr>
<tr>
<td>0.6</td>
<td>0.5</td>
<td>0.950</td>
<td>0.588</td>
<td>0.946</td>
<td>0.526</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>0.962</td>
<td>0.600</td>
<td>0.934</td>
<td>0.549</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5</td>
<td>0.976</td>
<td>0.628</td>
<td>0.946</td>
<td>0.600</td>
</tr>
<tr>
<td>0.9</td>
<td>0.5</td>
<td>0.982</td>
<td>0.660</td>
<td>0.978</td>
<td>0.612</td>
</tr>
<tr>
<td>0.1-0.9</td>
<td>0.5</td>
<td>0.918</td>
<td>0.572</td>
<td>0.946</td>
<td>0.513</td>
</tr>
<tr>
<td></td>
<td>Non-common threshold value ($c_0 = 0.5$, $c = 0.5$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.7</td>
<td>0.986</td>
<td>0.670</td>
<td>0.978</td>
<td>0.665</td>
</tr>
<tr>
<td>0.2</td>
<td>0.65</td>
<td>0.982</td>
<td>0.639</td>
<td>0.978</td>
<td>0.611</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6</td>
<td>0.980</td>
<td>0.622</td>
<td>0.968</td>
<td>0.581</td>
</tr>
<tr>
<td>0.4</td>
<td>0.55</td>
<td>0.964</td>
<td>0.606</td>
<td>0.956</td>
<td>0.558</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.916</td>
<td>0.599</td>
<td>0.948</td>
<td>0.537</td>
</tr>
<tr>
<td>0.6</td>
<td>0.55</td>
<td>0.918</td>
<td>0.584</td>
<td>0.936</td>
<td>0.527</td>
</tr>
<tr>
<td>0.7</td>
<td>0.6</td>
<td>0.940</td>
<td>0.599</td>
<td>0.942</td>
<td>0.551</td>
</tr>
<tr>
<td>0.8</td>
<td>0.65</td>
<td>0.934</td>
<td>0.626</td>
<td>0.946</td>
<td>0.603</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7</td>
<td>0.950</td>
<td>0.671</td>
<td>0.964</td>
<td>0.674</td>
</tr>
</tbody>
</table>

Table 4: The reputation scoring system on Taobao.com

<table>
<thead>
<tr>
<th>Category</th>
<th>Points needed</th>
<th>Category icon</th>
<th>Category</th>
<th>Points needed</th>
<th>Category icon</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4-10</td>
<td>1-heart</td>
<td>11</td>
<td>10001-20000</td>
<td>1-crown</td>
</tr>
<tr>
<td>2</td>
<td>11-40</td>
<td>2-hearts</td>
<td>12</td>
<td>20001-50000</td>
<td>2-crowns</td>
</tr>
<tr>
<td>3</td>
<td>41-90</td>
<td>3-hearts</td>
<td>13</td>
<td>50001-100000</td>
<td>3-crowns</td>
</tr>
<tr>
<td>4</td>
<td>91-150</td>
<td>4-hearts</td>
<td>14</td>
<td>100001-200000</td>
<td>4-crowns</td>
</tr>
<tr>
<td>5</td>
<td>151-250</td>
<td>5-hearts</td>
<td>15</td>
<td>200001-500000</td>
<td>5-crowns</td>
</tr>
<tr>
<td>6</td>
<td>251-500</td>
<td>1-diamond</td>
<td>16</td>
<td>500001-1000000</td>
<td>1-gold crown</td>
</tr>
<tr>
<td>7</td>
<td>501-1000</td>
<td>2-diamonds</td>
<td>17</td>
<td>1000001-200000</td>
<td>2-gold crowns</td>
</tr>
<tr>
<td>8</td>
<td>1001-2000</td>
<td>3-diamonds</td>
<td>18</td>
<td>2000001-500000</td>
<td>3-gold crowns</td>
</tr>
<tr>
<td>9</td>
<td>2001-5000</td>
<td>4-diamonds</td>
<td>19</td>
<td>5000001-1000000</td>
<td>4-gold crowns</td>
</tr>
<tr>
<td>10</td>
<td>5001-10000</td>
<td>5-diamonds</td>
<td>20</td>
<td>10000000+</td>
<td>5-gold crowns</td>
</tr>
</tbody>
</table>
Table 5: The sup-Wald test for the existence of quantile threshold effect

<table>
<thead>
<tr>
<th>( \bar{\tau} )</th>
<th>Test stat</th>
<th>( p )-value</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0.1,0.9]</td>
<td>596.11</td>
<td>0.00</td>
<td>19.88</td>
<td>21.54</td>
<td>25.09</td>
</tr>
<tr>
<td>0.1</td>
<td>151.43</td>
<td>0.00</td>
<td>15.68</td>
<td>17.34</td>
<td>20.63</td>
</tr>
<tr>
<td>0.2</td>
<td>190.86</td>
<td>0.00</td>
<td>15.31</td>
<td>16.80</td>
<td>20.13</td>
</tr>
<tr>
<td>0.3</td>
<td>279.62</td>
<td>0.00</td>
<td>15.28</td>
<td>16.71</td>
<td>20.28</td>
</tr>
<tr>
<td>0.4</td>
<td>359.55</td>
<td>0.00</td>
<td>15.28</td>
<td>16.72</td>
<td>19.95</td>
</tr>
<tr>
<td>0.5</td>
<td>596.11</td>
<td>0.00</td>
<td>15.19</td>
<td>16.60</td>
<td>19.77</td>
</tr>
<tr>
<td>0.6</td>
<td>103.87</td>
<td>0.00</td>
<td>15.18</td>
<td>16.55</td>
<td>20.32</td>
</tr>
<tr>
<td>0.7</td>
<td>61.00</td>
<td>0.00</td>
<td>15.24</td>
<td>16.68</td>
<td>20.35</td>
</tr>
<tr>
<td>0.8</td>
<td>12.47</td>
<td>0.36</td>
<td>15.46</td>
<td>16.98</td>
<td>20.19</td>
</tr>
<tr>
<td>0.9</td>
<td>9.25</td>
<td>0.89</td>
<td>15.94</td>
<td>17.67</td>
<td>21.41</td>
</tr>
</tbody>
</table>

Table 6: Estimation results

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Jump Size</th>
<th>( \gamma )</th>
<th>( \alpha_\tau )</th>
<th>( \beta_\tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>intercept</td>
<td>slope</td>
</tr>
<tr>
<td>0.1</td>
<td>-197.88</td>
<td>3264</td>
<td>856.10</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(6.21)</td>
<td>(0.003)</td>
<td>(34.15)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>0.2</td>
<td>-224.28</td>
<td>3264</td>
<td>947.22</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(5.73)</td>
<td>(0.003)</td>
<td>(30.59)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>0.3</td>
<td>-265.17</td>
<td>3264</td>
<td>992.30</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(6.04)</td>
<td>(0.004)</td>
<td>(43.50)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>0.4</td>
<td>-300.63</td>
<td>3264</td>
<td>1037.13</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(6.4)</td>
<td>(0.004)</td>
<td>(42.03)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>0.5</td>
<td>-370.79</td>
<td>3264</td>
<td>1080.00</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(6.53)</td>
<td>(0.004)</td>
<td>(42.61)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>0.6</td>
<td>-224.47</td>
<td>3240</td>
<td>1097.85</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(6.02)</td>
<td>(0.004)</td>
<td>(44.00)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>0.7</td>
<td>-148.52</td>
<td>3364</td>
<td>1121.87</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(6.31)</td>
<td>(0.004)</td>
<td>(52.40)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>mean regression</td>
<td>3240</td>
<td>1043.00</td>
<td>0.006</td>
<td>804.39</td>
</tr>
<tr>
<td></td>
<td>(4.78)</td>
<td>(0.004)</td>
<td>(89.36)</td>
<td>(0.023)</td>
</tr>
</tbody>
</table>

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 7: Inference on the threshold parameter

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Jump Size</th>
<th>( \gamma_\tau ) or ( \gamma )</th>
<th>95% lower bound</th>
<th>95% upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>3264</td>
<td>3168</td>
<td>3271</td>
<td>3271</td>
</tr>
<tr>
<td>0.2</td>
<td>3264</td>
<td>3182</td>
<td>3271</td>
<td>3271</td>
</tr>
<tr>
<td>0.3</td>
<td>3264</td>
<td>3168</td>
<td>3271</td>
<td>3271</td>
</tr>
<tr>
<td>0.4</td>
<td>3264</td>
<td>3166</td>
<td>3271</td>
<td>3271</td>
</tr>
<tr>
<td>0.5</td>
<td>3264</td>
<td>3231</td>
<td>3271</td>
<td>3271</td>
</tr>
<tr>
<td>0.6</td>
<td>3240</td>
<td>3232</td>
<td>3248</td>
<td>3248</td>
</tr>
<tr>
<td>0.7</td>
<td>3364</td>
<td>3337</td>
<td>3367</td>
<td>3367</td>
</tr>
<tr>
<td>0.1 - 0.5</td>
<td>3264</td>
<td>3197</td>
<td>3271</td>
<td>3271</td>
</tr>
<tr>
<td>mean regression</td>
<td>3240</td>
<td>3134</td>
<td>3272</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Scatter plot of the raw data

Figure 3: The histogram of reputation category
Figure 4: The estimated quantile regression lines before and after the jump point
In this supplement we first prove proposition 4.1 and then conduct robustness check for the empirical application in the above paper.

C Proof of Proposition 4.1

First, taking the first-order derivative of $\Pi$ with respect to $p$ yields the first-order condition (FOC):

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

It is worth noting that equation (C.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.

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

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

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

$$1 - \alpha v_1 - \bar{r} + r_1 = E_1 \text{ and } 1 - \alpha v_2 - \bar{r} + r_2 = E_2. \quad (C.3)$$

Note that

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

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 $\hat{e} \in (E_1, E_2)$ such that

$$r_1 - r_2 = -\frac{\beta \alpha}{2} [f(E_1) - f(E_2)] + (E_1 - E_2) = (E_1 - E_2) \left[1 - \frac{\beta \alpha}{2} f'(\hat{e})\right] > 0,$$
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:** \( r \leq r_2 \).

At \( r_2 \), the point \( p = v_2 \) makes the FOC in (C.1) hold by construction. Further, \( \frac{\partial \Pi}{\partial p}(v_2; r_2) = -2\alpha + \alpha^2 \beta f'(E_2) = 0 \) 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 (C.2), (C.3), and the fact that \( f'(E_2) = 2/(\alpha \beta) \), we can readily verify that \( \frac{\partial \Pi}{\partial p}(p_1(r_2); r_2) = 1 - 2\alpha p_1(r_2) - \alpha \beta f(E_2) = 2(r_1 - r_2) > 0 \) and \( \frac{\partial^2 \Pi}{\partial p^2}(p_1(r_2); r_2) = -2\alpha + \alpha^2 \beta f'(E_2) = 0 \). 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 6.) 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 \( \frac{\partial \Pi}{\partial p}(v_2; r_2) > 0 \), \( \frac{\partial \Pi}{\partial p}(p_1(r_2); r_2) > \frac{\partial \Pi}{\partial p}(p_1(r_2); r_2) > 0 \), \( \frac{\partial \Pi}{\partial p}(p_1^*(r); r_2) > 0 \), and \( \frac{\partial \Pi}{\partial p}(v_2(r); r_2) > 0 \forall r < r_2 \). 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 \( p_1^*(r) \in (p_1^*, p^m) \forall r \leq r_2 \).

By the FOC \( \frac{\partial \Pi}{\partial p}(p_1^*(r); r) = 0 \) and the implicit function theorem, we have

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

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 \( \frac{\partial \Pi}{\partial p}(p_1^*(r); r) < 0 \). That is, \( p_1^*(r) \) is decreasing in \( r \).

**Case 2:** \( r \geq r_1 \).

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 5.)

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 \( p_2^* \in (0, p_2(r)) \) \( \forall r > r_1 \) 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 \( r > r_1 \). [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 \( r > r_1 \).]

**Case 3:** \( r_2 < r < r_1 \).
There exist two local maxima, \( p^*_1(r) \in (p_1(r), \bar{p}) \) and \( p^*_2(r) \in (0, p_2(r)) \). (Refer to Figure 7.) Let \( \Delta(r) = \Pi(p^*_1(r); r) - \Pi(p^*_2(r); r) \). By the envelope theorem and FOC,
\[
\frac{\partial \Delta(r)}{\partial r} = \beta f(1 - \alpha p_1^*(r) - \bar{r} + r) - \beta f(1 - \alpha p_2^*(r) - \bar{r} + r)
= \frac{1 - 2\alpha p_1^*(r)}{\alpha} - \frac{1 - 2\alpha p_2^*(r)}{\alpha}
= 2[p_2^*(r) - p_1^*(r)] < 0.
\]
Moreover, noting that \( \Delta(r_2) > 0 \) and \( \Delta(r_1) < 0 \), there must exist a unique point \( \gamma_0 \in (r_2, r_1) \) such that \( \Delta(\gamma_0) = 0 \). 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.

**Remark on the intuition.**

Intuitively, the discontinuous pricing strategy occurs as follows. The restrictions in Assumptions M1 and M2 produce a peculiar shape of \( \Pi'(\cdot; r) \). Along with the increase in \( p \), \( \Pi'(\cdot; r) \) is initially downward sloping and convex, then becomes positive sloping and concave, and then eventually slopes downwards again. Thus, in order for the FOC in (C.1) to hold, there are three possible ways that \( \Pi'(\cdot; r) \) intersects the horizontal \( p \)-axis:

- Case 1: the intersection occurs in the concave region alone (refer to Figure 6);
- Case 2: the intersection occurs in the convex region alone (refer to Figure 5); and
- Case 3: the intersection occurs in both regions (refer to Figure 7).

In the proof, we show that the pricing scheme in Cases 1 and 2 are associated with small and large values of \( r \), respectively, in Case 3 there exists a threshold value \( \gamma_0 \) such that the seller will switch between the two pricing schemes when \( r \) increases from a number below \( \gamma_0 \) to one above \( \gamma_0 \). It is the presence of a positively sloped segment of \( \Pi'(\cdot; r) \) that makes the profit function \( \Pi(\cdot; r) \) exhibit 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 a change in pricing scheme may not occur.

We first take a close look at Case 1 by considering a slight change in \( p \). 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 \( \Pi \) occurs in the concave region of \( \Pi'(\cdot; r) \) in this case. Parallel to the first case, we next consider a change in \( p \) in Case 2. 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 \( \Pi'(\cdot; r) \).

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 7. 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...
Figure 5: Pricing strategy when $r \geq r_1$

Figure 6: Pricing strategy when $r \leq r_2$
to charge \( p_1^*(r) \). However, along with the increase in \( r \), there must exist a \( \gamma_0 \) that makes it worthwhile for the seller to switch to the pricing regime \( p_2^*(r) \).

In Case 1 where the value of \( r \) is small, the tail of \( f \), the distribution of not recruiting good reviews, is relevant. As a matter of fact, being to the left of the mode implies that the optimal prices will decrease with \( r \). Such a decreasing pricing pattern simply reflects the fact that the potential benefit of goodwill becomes more significant as \( r \) increases. However, in Case 2, 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.

D Robustness check

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). 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 A.1 lists the basic summary statistics for our data. 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 variation in price. We observe only limited information on sellers in the dataset, among which “total items” is the most important. It represents the total number of items for sale in a particular online store. 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 sellers’ 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.

Although the reputation pricing pattern is found in the data, there remain critical issues in the foregoing 
empirical analysis. For example, 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. We thus augment our TQR model in Section 4 only by the covariates listed in Table A.1.

Following the testing approach suggested in Section 2.2, 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 \( \tau = 0.1, \ldots, 0.9 \). 
The estimation and inference results for the threshold parameters are reported in Table A.2. From these 
estimates, we suspect quantiles 0.4 – 0.7 may share a common threshold. Therefore, we did the test for 
common threshold among these quantiles, whose results strongly supports the null of common threshold. 
Then, the estimated common threshold parameter is 3252, with a confidence interval of [3246,3260].

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 \( \tau \), the higher 
the reputation level at which the jump occurs (that is the closer to the exogenous cutoff of reputation). 
Although jumps are identified in the quantiles of \( \tau = 0.8,0.9 \), they are smaller in magnitude, relative to 
other quantiles. 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 \( \tau = 0.1 \). Recall that the slope estimate before the jump 
for \( \tau = 0.1 \) is statistically significant and positive in Table 6. These inconsistent findings 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 
flow 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 project and is therefore left for future research.
Table A.1: Summary statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>508</td>
<td>1400</td>
<td>995.10</td>
<td>1050</td>
<td>171.43</td>
</tr>
<tr>
<td>reputation score</td>
<td>501</td>
<td>4995</td>
<td>1875.48</td>
<td>1392</td>
<td>1225.58</td>
</tr>
<tr>
<td>total items</td>
<td>7</td>
<td>36811</td>
<td>645.30</td>
<td>288</td>
<td>1787.90</td>
</tr>
<tr>
<td>postage</td>
<td>0</td>
<td>100</td>
<td>11.40</td>
<td>10</td>
<td>10.42</td>
</tr>
<tr>
<td>sales volume</td>
<td>0</td>
<td>300</td>
<td>3.84</td>
<td>1</td>
<td>12.86</td>
</tr>
<tr>
<td>rate of good reviews</td>
<td>0</td>
<td>100</td>
<td>89.06</td>
<td>99.83</td>
<td>30.79</td>
</tr>
</tbody>
</table>

NOTE: The sample includes only sellers that belong to Categories 7 to 9. The total number of observations is 1903.

Table A.2: Estimation and inference on the threshold parameter

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Jump Size</th>
<th>$\gamma$</th>
<th>95% lower bound</th>
<th>95% upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>277.68</td>
<td>1984</td>
<td>1975</td>
<td>2018</td>
</tr>
<tr>
<td>0.2</td>
<td>-77.83</td>
<td>3264</td>
<td>3231</td>
<td>3271</td>
</tr>
<tr>
<td>0.3</td>
<td>-125.60</td>
<td>2979</td>
<td>2975</td>
<td>3002</td>
</tr>
<tr>
<td>0.4</td>
<td>-101.01</td>
<td>3252</td>
<td>3232</td>
<td>3272</td>
</tr>
<tr>
<td>0.5</td>
<td>-52.38</td>
<td>3252</td>
<td>3247</td>
<td>3261</td>
</tr>
<tr>
<td>0.6</td>
<td>-87.07</td>
<td>3252</td>
<td>3232</td>
<td>3272</td>
</tr>
<tr>
<td>0.7</td>
<td>-64.23</td>
<td>3252</td>
<td>3247</td>
<td>3261</td>
</tr>
<tr>
<td>0.8</td>
<td>-28.00</td>
<td>3364</td>
<td>3355</td>
<td>3367</td>
</tr>
<tr>
<td>0.9</td>
<td>-27.64</td>
<td>3849</td>
<td>3750</td>
<td>3947</td>
</tr>
<tr>
<td>0.4-0.7</td>
<td></td>
<td>3252</td>
<td>3246</td>
<td>3260</td>
</tr>
</tbody>
</table>