Research Collection School Of Economics   School of Economics

Common threshold in quantile regressions with an application to pricing for reputation

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

Pai XU
University of Hong Kong

Common threshold in quantile regressions with an application to pricing for reputation

Liangjun Su[a] and Pai Xu[b]

[a]School of Economics, Singapore Management University, Singapore; [b]School of Economics and Finance, University of Hong Kong, Hong Kong

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 through the use of a data set 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.

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 τ. 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 $\mathcal{T}$ 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 $\tau \in \mathcal{T}$. It is possible that the threshold effect does not exist for τ in a subset $\mathcal{T}_1$ of $\mathcal{T}$, but it is present for τ that lies in the complement of $\mathcal{T}_1$ relative to $\mathcal{T}$. In this case, the single quantile index-based test of Lee et al. (2011) loses power if the quantile index τ happens to lie in $\mathcal{T}_1$. 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 done by 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 γ_τ 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 γ^0 is unobserved and estimated from data, and the other is the case where γ^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 γ^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 use the same pricing for reputation strategy predicted in the model. Thus, we believe that 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). We contribute to the literature by documenting a scenario in which, at individuals' optimal behavior, another endogenous cutoff may occur in accordance with the incentive to achieve an exogenous threshold for the treatment.

The rest of the paper is organized as follows. 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 noncommon 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 through the use of a data set 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 τ and consider the inference issue associated with the threshold parameter. Finally 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 $\left\{y_i, x_i, r_i\right\}_{i=1}^{n}$ 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 $r_i \notin x_i$, and $z_i \equiv x_i$ otherwise. We assume that the τth conditional quantile of y_i, given z_i, is given by:

$$Q_\tau(z_i) = \alpha_\tau' x_i 1\{r_i \leq \gamma_\tau\} + \beta_\tau' x_i 1\{r_i > \gamma_\tau\}, \tag{2.1}$$

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 γ_τ. If δ_τ is 0 for all γ_τ 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 γ_τ 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 θ_τ as:

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

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 γ_τ is known and obtain an estimate of $(\alpha_\tau, \beta_\tau)$ by:

$$\left(\hat{\alpha}_\tau(\gamma_\tau), \hat{\beta}_\tau(\gamma_\tau)\right) \equiv \left(\hat{\alpha}(\tau, \gamma_\tau), \hat{\beta}(\tau, \gamma_\tau)\right) = \underset{\alpha_\tau, \beta_\tau}{\operatorname{argmin}} S_{n\tau}(\alpha_\tau, \beta_\tau, \gamma_\tau), \tag{2.3}$$

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 γ_τ by:

$$\hat{\gamma}_\tau = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \hat{S}_{n\tau}(\gamma). \tag{2.4}$$

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 γ_τ over $\Gamma_n = \Gamma \cap \{r_1, r_2, \ldots, r_n\}$. Then, computing $\hat{\gamma}_\tau$ requires at most n function evaluations.[1] However, if n is large, then we can approximate Γ by a grid as suggested by Hansen (2000). After $\hat{\gamma}_\tau$ is obtained, we can compute the estimates of α_τ and β_τ as $\hat{\alpha}_\tau = \hat{\alpha}_\tau(\hat{\gamma}_\tau)$ and $\hat{\beta}_\tau = \hat{\beta}_\tau(\hat{\gamma}_\tau)$, respectively.

2.2. Test the existence of a change point

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

Let $\mathcal{T} \equiv [\underline{\tau}, \overline{\tau}] \subset (0, 1)$ and $\Theta_1 \subset \mathbb{R}^{2k}$ denotes the compact support for $\theta_{1\tau}$. In principle, we allow the support of $\theta_{1\tau}$ to be τ-dependent, and write Θ_1 as $\Theta_{1\tau}$. We use Θ_1 instead of $\Theta_{1\tau}$ mainly for notational simplicity. Let $\theta_{n1\tau}^0 \equiv \left(\alpha_{n\tau}^{0\prime}, \beta_{n\tau}^{0\prime}\right)'$ denote the true value of $\theta_{1\tau}$. It is fair to comment here that we allow $\theta_{n1\tau}^0$ 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 $\theta_{n1\tau}^0$, $\alpha_{n\tau}^0$, and $\beta_{n\tau}^0$ on n and write them as $\theta_{1\tau}^0$, α_τ^0, and β_τ^0, respectively.

The null hypothesis of no regime change is

$$\mathbb{H}_0 : Q_\tau(z_i) = z_i(\gamma)' \theta_{1\tau}^0 \ \forall \gamma \in \Gamma \text{ for some } \theta_{1\tau}^0 \in \Theta_1 \text{ with } \alpha_\tau^0 = \beta_\tau^0 \text{ for all } \tau \in \mathcal{T}. \tag{2.5}$$

The alternative hypothesis is

$$\mathbb{H}_1 : Q_\tau(z_i) = z_i\left(\gamma_\tau^0\right)' \theta_{1\tau}^0 \text{ for some } \left(\gamma_\tau^0, \theta_{1\tau}^0\right) \in \Gamma \times \Theta_1 \text{ with } \alpha_\tau^0 \neq \beta_\tau^0 \text{ for some } \tau \in \mathcal{T}. \tag{2.6}$$

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

$$\mathbb{H}_{0\tau} : Q_\tau(z_i) = z_i(\gamma)' \theta_{1\tau}^0 \text{ for some } \theta_{1\tau}^0 \in \Theta_1 \text{ with } \alpha_\tau^0 = \beta_\tau^0, \tag{2.7}$$

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

$$\mathbb{H}_{1\tau} : Q_\tau(z_i) = z_i\left(\gamma_\tau^0\right)' \theta_{1\tau}^0 \text{ for some } \theta_{1\tau}^0 \in \Theta_1 \text{ with } \alpha_\tau^0 \neq \beta_\tau^0 \text{ and } \gamma_\tau^0 \in \Gamma. \tag{2.8}$$

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

$$\hat{\theta}_1(\tau, \gamma) = \left(\hat{\alpha}(\tau, \gamma)', \hat{\beta}(\tau, \gamma)'\right)' = \underset{\theta_1}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau\left(y_i - \theta_1' z_i(\gamma)\right). \tag{2.9}$$

Even though γ is not identified under $\mathbb{H}_0$ for any $\tau \in \mathcal{T}$, we can study the asymptotic property of $\hat{\theta}_1(\tau, \gamma)$ under $\mathbb{H}_0$ and propose a test for the null hypothesis of no regime change for all τ over $\mathcal{T}$ by considering the asymptotic behavior of $\hat{\delta}(\tau, \gamma) \equiv \hat{\alpha}(\tau, \gamma) - \hat{\beta}(\tau, \gamma)$ over the compact set $\mathcal{T} \times \Gamma$.

Here we consider a sup-Wald statistic for testing $\mathbb{H}_0$. Let $\Omega(\gamma) \equiv E\left[x_i x_i' 1\{r_i \leq \gamma\}\right]$, $\Omega^*(\gamma) \equiv E\left[x_i x_i' 1\{r_i > \gamma\}\right]$, $\Omega(\tau, \gamma) \equiv E\left[x_i x_i' 1\{r_i \leq \gamma\} f\left(\alpha_\tau^{0\prime} x_i | z_i\right)\right]$, and $\Omega^*(\tau, \gamma) \equiv E[x_i x_i' 1\{r_i > \gamma\} \times f\left(\beta_\tau^{0\prime} x_i | z_i\right)]$, where $f(\cdot|z)$ denotes the conditional PDF of y_i given $z_i = z$. By Lemma B.1 in Appendix B,

---

[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 unqiue 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.

we can readily show that for each $(\tau, \gamma) \in \mathcal{T} \times \Gamma$,

$$\sqrt{n}\hat{\delta}(\tau, \gamma) \overset{d}{\to} N\left(\mathbf{0}_{k \times 1}, \tau(1 - \tau) V(\tau, \gamma)\right), \tag{2.10}$$

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

$$V(\tau, \gamma) \equiv \Omega(\tau, \gamma)^{-1} \Omega(\gamma) \Omega(\tau, \gamma)^{-1} + \Omega^*(\tau, \gamma)^{-1} \Omega^*(\gamma) \Omega^*(\tau, \gamma)^{-1}. \tag{2.11}$$

Let $\hat{\Omega}(\gamma) \equiv n^{-1} \sum_{i=1}^{n} x_i x_i' 1\{r_i \leq \gamma\}$, $\hat{\Omega}^*(\gamma) \equiv n^{-1} \sum_{i=1}^{n} x_i x_i' - \hat{\Omega}(\gamma)$, $\hat{\Omega}(\tau, \gamma) \equiv (2nh)^{-1} \sum_{i=1}^{n} 1$ $\{|y_i - x_i'\hat{\alpha}_\tau| \leq h\} x_i x_i' 1\{r_i \leq \gamma\}$, and $\hat{\Omega}^*(\tau, \gamma) \equiv (2nh)^{-1} \sum_{i=1}^{n} 1\{|y_i - x_i'\hat{\beta}_\tau| \leq h\} x_i x_i' 1\{r_i > \gamma\}$, where $h \equiv h(n)$ is a bandwidth parameter such that $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$ (see Koenker (2005, pp. 80–81)). 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:

$$\hat{V}(\tau, \gamma) \equiv \hat{\Omega}(\tau, \gamma)^{-1} \hat{\Omega}(\gamma) \hat{\Omega}(\tau, \gamma)^{-1} + \hat{\Omega}^*(\tau, \gamma)^{-1} \hat{\Omega}^*(\gamma) \hat{\Omega}^*(\tau, \gamma)^{-1}. \tag{2.12}$$

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 $\mathbb{H}_0$ given by:

$$\sup W_n \equiv \sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} W_n(\tau, \gamma), \tag{2.13}$$

where $W_n(\tau, \gamma) = n\hat{\delta}(\tau, \gamma)' [\tau(1 - \tau) \hat{V}(\tau, \gamma)]^{-1} \hat{\delta}(\tau, \gamma)$.

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

Theorem 2.1. *Let $R \equiv [I_k, -I_k]$ with I_k being a $k \times k$ identity matrix, $\Omega_0(\gamma_1, \gamma_2) \equiv E[z_i(\gamma_1) z_i(\gamma_2)']$, and $\Omega_1(\tau, \gamma) \equiv E[f(\theta_{1\tau}^{0\prime} z_i(\gamma) | z_i) z_i(\gamma) z_i(\gamma)']$. Suppose that Assumptions A1–A5 in Appendix A hold. Suppose that $h \to 0$ and $nh^2 \to \infty$ as $n \to \infty$. Then, under $\mathbb{H}_0$,*

$$\sup W_n \overset{d}{\to} \sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} \frac{1}{\tau(1 - \tau)} W(\tau, \gamma)' \Omega_1(\tau, \gamma)^{-1} R' V(\tau, \gamma)^{-1} R\Omega_1(\tau, \gamma)^{-1} W(\tau, \gamma)$$

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 $\Omega_1(\tau, \gamma) = \begin{pmatrix} \Omega(\tau, \gamma) & \mathbf{0}_{k \times k} \\ \mathbf{0}_{k \times k} & \Omega^*(\tau, \gamma) \end{pmatrix}$, we can consistently estimate it by $\hat{\Omega}_1(\tau, \gamma) \equiv \begin{pmatrix} \hat{\Omega}(\tau, \gamma) & \mathbf{0}_{k \times k} \\ \mathbf{0}_{k \times k} & \hat{\Omega}^*(\tau, \gamma) \end{pmatrix}$. 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 $Z_n(\tau, \gamma) = n^{-1/2} \sum_{i=1}^{n} [\tau - 1\{u_i \leq \tau\}] z_i(\gamma)$;
3. Compute $\sup W_n^* \equiv \sup_{(\tau, \gamma) \in \mathcal{T} \times \Gamma} \frac{1}{\tau(1 - \tau)} Z_n(\tau, \gamma)' \hat{\Omega}_1(\tau, \gamma)^{-1} R' \hat{V}(\tau, \gamma)^{-1} R\hat{\Omega}_1(\tau, \gamma)^{-1} Z_n(\tau, \gamma)$;
4. Repeat Steps 1–3 B times and denote the resulting $\sup W_n^*$ test statistics as $\sup W_{n,j}^*$ for $j = 1, \ldots, B$.
5. Calculate the simulated p-value for the $\sup W_n$ test as $p_W^* = \frac{1}{B} \sum_{j=1}^{B} 1\{\sup W_{n,j}^* \geq \sup W_n\}$.

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

One can readily show that $Z_n(\cdot, \cdot) \Rightarrow W(\cdot, \cdot)$ in $(\ell^\infty(\mathcal{T} \times \Gamma))^{2k}$, where $\Rightarrow$ denotes weak convergence and $\ell^\infty(\mathcal{T} \times \Gamma)$ the space of all bounded functions on $\mathcal{T} \times \Gamma$ equipped with the uniform topology. When B is sufficiently large, the asymptotic critical value of the level α test based on sup W_n is approximately given by the empirical upper α-quantile of $\{\sup W_{n,j}^*, j = 1, \ldots, B\}$. Therefore, we can reject the null hypothesis $\mathbb{H}_0$ if the simulated p-value p_W^* is smaller than the prescribed nominal level of significance α.

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

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

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

If γ_τ is known, one can follow Koenker and Machado (1999) and consider several tests for the null hypothesis $\mathbb{H}_0 : \delta_\tau^0 = 0 \, \forall \, \tau \in \mathcal{T}$, 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 (τ, γ). A general test statistic can be written as a function $g(T_n)$, where $g(\cdot)$ maps functionals on $\mathcal{T} \times \Gamma$ to the real line $\mathbb{R}$ and we treat $T_n = \{T_n(\tau, \gamma) : (\tau, \gamma) \in \mathcal{T} \times \Gamma\}$ as a random function on $\mathcal{T} \times \Gamma$. 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 (γ_τ 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 considering 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 asymptotic Laplacean density: $f_\tau(u) = \tau(1 - \tau) \exp(-\rho_\tau(u))$, where $\rho_\tau(\cdot)$ 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 differential* with respect to the parameter of interest ($\theta_{1\tau} = (\alpha_\tau', \beta_\tau')'$ 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 section, 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 γ_τ. The analysis here will pave a way to the test and inference issue associated with the common threshold parameter in the next section.

Let $\hat{\theta}_{1\tau} \equiv (\hat{\alpha}'_\tau, \hat{\beta}'_\tau)'$ and $\hat{\theta}_\tau \equiv (\hat{\theta}'_{1\tau}, \hat{\gamma}_\tau)'$. Let $N(\gamma) \equiv E\left[x_i x'_i | r_i = \gamma\right]$, $D_\tau(\gamma) \equiv E\left[f\left(\alpha^{0\prime}_\tau x_i | z_i\right) x_i x'_i | r_i = \gamma\right]$, $N_\tau \equiv N_\tau\left(\gamma^0_\tau\right)$, and $D_\tau \equiv D_\tau\left(\gamma^0_\tau\right)$. 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 $\theta^0_\tau = (\theta^{0\prime}_{1\tau}, \gamma^0_\tau)'$.*

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

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

(ii) $n^{1-2a}\left(\hat{\gamma}_\tau - \gamma^0_\tau\right) \xrightarrow{d} \dfrac{\lambda_\tau}{\mu^2_\tau} \underset{r \in (-\infty, \infty)}{argmax} \{W(r) - \tfrac{1}{2}|r|\}$,

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

Recall that a two-sided Brownian motion on the real line is defined as $W(r) = W_1(-r)\mathbf{1}\{r \leq 0\} + W_2(r)\mathbf{1}\{r > 0\}$, 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 $(\frac{\lambda_\tau}{\mu^2_\tau})$ and the random component in the limiting distribution, i.e., $\underset{r \in (-\infty, \infty)}{argmax} \{W(r) - \tfrac{1}{2}|r|\}$ 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 $n\left(\hat{\gamma}_\tau - \gamma^0_\tau\right)$ converges in distribution to the argmin of a two-sided compound Poisson process. However, such an independence assumption seems too strong, and thus we focus only on the case of $a \in (0, \frac{1}{2})$. 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 τ on the set $\mathcal{T}$. 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 γ_τ below.

To make inferences about γ_τ, 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 γ_τ. Specifically, we are interested in testing the null hypothesis

$$H_{0\tau} : \gamma_\tau = \gamma_\tau^0. \tag{2.15}$$

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

Theorem 2.4. *Suppose that Assumptions A1–A8 in Appendix A hold. Then under $H_{0\tau}$, $LR_{n\tau}\left(\gamma_\tau^0\right) \xrightarrow{d} \frac{\lambda_\tau}{2\mu_\tau} \Xi$, where $\Xi \equiv \sup_{r\in(-\infty,\infty)} \{2W(r) - |r|\}$, and λ_τ, μ_τ, and $W(\cdot)$ are as defined in Theorem 2.3.*

Theorem 2.4 indicates that $LR_{n\tau}\left(\gamma_\tau^0\right)$ is not asymptotically pivotal under the null hypothesis. To obtain an asymptotically pivotal test statistic, we need to estimate λ_τ and μ_τ and consider the following normalized LR test statistic:

$$NLR_{n\tau}\left(\gamma_\tau^0\right) = \frac{2\hat{\delta}_\tau' \hat{D}_\tau \hat{\delta}_\tau}{\tau(1-\tau)\hat{\delta}_\tau' \hat{N}_\tau \hat{\delta}_\tau} LR_{n\tau}\left(\gamma_\tau^0\right), \tag{2.16}$$

where $\hat{\delta}_\tau \equiv \hat{\alpha}_\tau\left(\gamma_\tau^0\right) - \hat{\beta}_\tau\left(\gamma_\tau^0\right)$, $\hat{N}_\tau = \hat{N}_\tau\left(\gamma_\tau^0\right)$ is a consistent local linear (or constant) estimate of $N_\tau\left(\gamma_\tau^0\right)$ using the bandwidth h_1 and the kernel K; $\hat{D}_\tau = \hat{D}\left(\gamma_\tau^0\right) = \hat{E}_{h_1}[\hat{f}_{h_1}(\hat{\alpha}_\tau\left(\gamma_\tau^0\right)' x_i | z_i) x_i x_i' | \gamma_\tau^0]$, $\hat{f}_{h_1}(\cdot|z_i)$ is a kernel estimate for the density of y_i given z_i using the bandwidth h_1 and the kernel K, and $\hat{E}_{h_1}\left(\cdot | \gamma_\tau^0\right)$ is a kernel estimate of $E\left[f\left(\alpha_\tau^{0\prime} x_i | z_i\right) x_i x_i' | r_i = \gamma_\tau^0\right]$ using the bandwidth h_1, the kernel K, and the observations on $\hat{f}_{h_1}(\hat{\alpha}_\tau\left(\gamma_\tau^0\right)' x_i | z_i) x_i x_i'$ 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 $\phi\left((y_i - y)/h_y\right)$ on z_i using the local linear method with kernel K and bandwidth h_1 to obtain $f\left(y|z_i\right)$, where h_y is another bandwidth and $\phi(\cdot)$ is the standard normal PDF.

Under standard conditions, we can readily show that $\frac{\hat{\delta}_\tau' \hat{D}_\tau \hat{\delta}_\tau}{\tau(1-\tau)\hat{\delta}_\tau' \hat{N}_\tau \hat{\delta}_\tau} \to \frac{\mu_\tau}{\lambda_\tau}$ in probability. Then, by the Slutsky lemma, we have $NLR_{n\tau}\left(\gamma_\tau^0\right) \xrightarrow{d} \Xi$. That is, $NLR_{n\tau}\left(\gamma_\tau^0\right)$ is asymptotically pivotal. It is well known that $\sup_{r\le0}[2W(r) - |r|]$ and $\sup_{r\ge0}[2W(r) - |r|]$ are independent exponential random variables with distribution function $1 - e^{-z}$ such that the CDF of Ξ is given by $P(\Xi \le z) = \left(1 - e^{-z/2}\right)^2$. We can easily tabulate the asymptotic critical values for the normalized statistic $NLR_{n\tau}\left(\gamma_\tau^0\right)$. See Hansen (2000, p. 582) for more details. In addition, we can invert the $NLR_{n\tau}\left(\gamma_\tau^0\right)$ statistic to obtain the asymptotic $1 - \alpha$ confidence interval for $\gamma_\tau^0 : 1 - \alpha$ confidence interval for the common threshold parameter γ^0 is given by $C_{1-\alpha} = \{\gamma : NLR_{n\tau}(\gamma) \le \Xi_{1-\alpha}, \gamma \in \Gamma_n\}$, where $\Xi_{1-\alpha}$ is the $1 - \alpha$ upper percentile of Ξ (e.g., $\Xi_{1-\alpha} = 5.94, 7.35$ 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 section, 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 $\mathcal{T} \equiv [\underline{\tau}, \overline{\tau}] \subset (0, 1)$. 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

$$H_0 : \gamma_\tau = \gamma^0 \text{ for all } \tau \in \mathcal{T} \text{ and some } \gamma^0 \in \Gamma, \tag{2.17}$$

versus the alternative hypothesis

$$H_1 : \text{There is no } \gamma \in \Gamma \text{ such that } \gamma_\tau = \gamma \text{ for all } \tau \in \mathcal{T}. \tag{2.18}$$

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

$$\tilde{\gamma} \equiv \tilde{\gamma}(\Pi) \equiv \underset{\gamma \in \Gamma}{\text{argmin }} \hat{S}_{n,\Pi}(\gamma), \tag{2.19}$$

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

Theorem 2.5. *Suppose that Assumptions A1–A4 and A6–A9 in Appendix A hold. Then,*
(i) $\tilde{\gamma} = \gamma^0 + o_P(1)$ *and* $\tilde{\theta}_{1\tau} = \theta_{1\tau}^0 + o_P(1)$ *for each* $\tau \in \mathcal{T}$;
(ii) $\sqrt{n}\left(\hat{\theta}_1(\tau, \gamma) - \theta_1^0(\tau, \gamma)\right) = \bar{\Omega}_1(\tau, \gamma)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau\left(y_i - \theta_1^0(\tau, \gamma)' z_i(\gamma)\right) z_i(\gamma) + o_P(1)$ *uniformly in* $(\tau, \gamma) \in \mathcal{T} \times \Gamma$;
(iii) $n^{1/2}\left(\tilde{\theta}_{1\tau} - \theta_{1\tau}^0\right) \overset{d}{\to} N\left(\mathbf{0}_{2k \times 1}, \tau(1 - \tau)\Sigma(\tau, \gamma^0)\right)$;
(iv) $n^{1-2a}\left(\tilde{\gamma} - \gamma^0\right) \overset{d}{\to} \frac{\lambda_0}{\mu_0^2} \underset{r \in (-\infty, \infty)}{\text{argmax}} \{W(r) - \frac{1}{2}|r|\}$;
where $\psi_\tau(u) \equiv \tau - 1\{u < 0\}$, *the pseudo-true values* γ^0, $\theta_{1\tau}^0$ *and* $\theta_1^0(\tau, \gamma)$ *together with* $\bar{\Omega}_1(\tau, \gamma)$ *are defined in Assumption A9,* $\lambda_0 \equiv \{\int \sqrt{\tau(1 - \tau)} v_\tau' N_\tau(\gamma^0) v_\tau d\Pi(\tau)\}^2 g(\gamma^0)$, *and* $\mu_0 \equiv \int v_\tau' D_\tau(\gamma^0) v_\tau d\Pi(\tau) g(\gamma^0)$ *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 $\tilde{\theta}_{1\tau}$ is the same as that of $\hat{\theta}_{1\tau}$ 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 $\tilde{\gamma}$ of γ^0 under H_0 and motivated from the early LR test for a threshold parameter, we consider the following LR test statistic:

$$LR_n = \int \left[S_{n\tau}(\hat{\theta}_{1\tau}, \tilde{\gamma}) - S_{n\tau}(\hat{\theta}_{1\tau}, \hat{\gamma}_\tau)\right] d\Pi(\tau). \tag{2.20}$$

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* $LR_n \overset{d}{\to} c_{LR}\Xi$, *where* $c_{LR} = \int \frac{\lambda_\tau}{2\mu_\tau} d\Pi(\tau) - \frac{\lambda_0}{2\mu_0}$, λ_τ, μ_τ *and* Ξ *are defined in Theorem 2.4, and* λ_0 *and* μ_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

$$\tilde{c}_{LR} = \frac{1}{2} \int \frac{\tau\,(1-\tau)\,\hat{\delta}'_\tau \hat{N}_\tau \hat{\delta}_\tau}{\hat{\delta}'_\tau \hat{D}_\tau \hat{\delta}_\tau} d\Pi\,(\tau) - \frac{\left\{ \int \sqrt{\tau\,(1-\tau)\,\hat{\delta}'_\tau \hat{N}_\tau \hat{\delta}_\tau}\, d\Pi\,(\tau) \right\}^2}{2 \int \hat{\delta}'_\tau \hat{D}_\tau \hat{\delta}_\tau\, d\Pi\,(\tau)} \tag{2.21}$$

where $\hat{\delta}_\tau$, $\hat{N}_\tau$, and $\hat{D}_\tau$ are defined as before. Using the Cauchy–Schwarz inequality, we can easily show that $\tilde{c}_{LR} \geq 0$ and the equality holds if and only if $\hat{\delta}'_\tau \hat{N}_\tau \hat{\delta}_\tau = (\hat{\delta}'_\tau \hat{D}_\tau \hat{\delta}_\tau)^2$ a.e.-Π. 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

$$NLR_n \equiv \tilde{c}_{LR}^{-1} LR_n \xrightarrow{d} \Xi. \tag{2.22}$$

In some applications, γ^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 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 γ^0. In this case, we write the null hypothesis as

$$H_0^* : \gamma_\tau = \gamma^0 \text{ for all } \tau \in \mathcal{T} \tag{2.23}$$

and the alternative hypothesis as

$$H_1^* : \gamma_\tau \neq \gamma^0 \text{ for some } \tau \in \mathcal{T}. \tag{2.24}$$

In this case, we investigate the following LR statistic:

$$\overline{LR}_n\left(\gamma^0\right) = \int \left[S_{n\tau}\left(\hat{\theta}_{1\tau}, \gamma^0\right) - S_{n\tau}\left(\hat{\theta}_{1\tau}, \tilde{\gamma}\right) \right] d\Pi\,(\tau). \tag{2.25}$$

We reject H_0^* for large values of $\overline{LR}_n\left(\gamma^0\right)$. The following theorem establishes the asymptotic distribution of $\overline{LR}_n\left(\gamma^0\right)$ 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* $\overline{LR}_n\left(\gamma^0\right) \xrightarrow{d} \frac{\lambda_0}{2\mu_0} \Xi$, *where λ_0 and μ_0 are as defined in Theorem 2.5.*

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

$$\overline{NLR}_n\left(\gamma^0\right) = \frac{2 \int \tilde{\delta}'_\tau \tilde{D}\left(\gamma^0\right) \tilde{\delta}_\tau d\Pi\,(\tau)}{\left\{ \int \sqrt{\tau\,(1-\tau)\,\tilde{\delta}'_\tau \tilde{N}_\tau\left(\gamma^0\right) \tilde{\delta}_\tau}\, d\Pi\,(\tau) \right\}^2} \overline{LR}_n\left(\gamma^0\right), \tag{2.26}$$

where $\tilde{\delta}_\tau = \tilde{\alpha}_\tau - \tilde{\beta}_\tau$, and $\tilde{D}\left(\gamma^0\right)$ and $\tilde{N}_\tau\left(\gamma^0\right)$ are defined analogously to $\hat{D}\left(\gamma_\tau^0\right)$ and $\hat{N}_\tau\left(\gamma_\tau^0\right)$. One can readily show that $\overline{NLR}_n\left(\gamma^0\right) \xrightarrow{d} \Xi$ under H_0^*. The $1 - \alpha$ confidence interval for the common threshold parameter γ^0 is given by $C_{1-\alpha} = \left\{ \gamma : \overline{NLR}_n\left(\gamma\right) \leq \Xi_{1-\alpha},\, \gamma \in \Gamma_n \right\}$.

By the definitions of the two LR statistics, namely, LR_n in (2.20) and $\overline{LR}_n\left(\gamma^0\right)$ 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 $\theta_{1\tau}^0$ is not affected by whether the threshold parameter γ_τ 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):

$$y_i = \begin{cases} [1 + \Phi^{-1}(v_i)] + (a_0 + a_1 v_i)x_i, & \text{if } r_i \leq 0.5 + c|v_i - 0.5| \\ [1 + \Phi^{-1}(v_i) + c_0 n^{-1/8}] + (a_0 + a_1 v_i + c_0 n^{-1/8})x_i, & \text{if } r_i > 0.5 + c|v_i - 0.5| \end{cases}, \quad (3.1)$$

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 τth conditional quantile of y_i given (x_i, r_i) is

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

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 τ 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 $\gamma_\tau^0 = 0.5 + 0.5 \cdot |\tau - 0.5|$, which is varying over τ.

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 κ 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 $\mathcal{T}$ is a singleton $\{\tau\}$ for $\tau = 0.1, 0.2, \ldots,$ and 0.9. Then we consider the test by setting $\mathcal{T} = [0.1, 0.9]$.

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 $\mathcal{T} = \{\tau\}$ is a singleton and τ lies close to either 0 or 1, and it is moderately oversized for the 5% test when τ takes value around 0.5. This is especially true

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

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

(*Continued*)

Table 1. Continued.

$\mathcal{T}$ \ Level	$n = 200$			$n = 400$		
	1%	5%	10%	1%	5%	10%
Common threshold value ($c_0 = 2, c = 0$)						
0.1	0.956	0.994	0.998	0.998	1	1
0.2	0.986	0.996	1	1	1	1
0.3	1	1	1	1	1	1
0.4	1	1	1	1	1	1
0.5	1	1	1	1	1	1
0.6	1	1	1	1	1	1
0.7	0.998	1	1	1	1	1
0.8	0.992	0.998	0.998	1	1	1
0.9	0.96	0.99	0.994	1	1	1
[0.1, 0.9]	1	1	1	1	1	1
Noncommon threshold value ($c_0 = 2, c = 0.5$)						
0.1	0.938	0.972	0.980	0.994	0.998	0.998
0.2	0.984	0.994	0.996	1	1	1
0.3	0.996	1	1	1	1	1
0.4	1	1	1	1	1	1
0.5	1	1	1	1	1	1
0.6	0.998	1	1	1	1	1
0.7	0.988	0.998	1	1	1	1
0.8	0.960	0.982	0.994	1	1	1
0.9	0.880	0.950	0.970	1	1	1
[0.1, 0.9]	0.994	1	1	1	1	1

when $n = 200$. When $\mathcal{T} = [0.1, 0.9]$, 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 $\mathcal{T}$.

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 $\mathcal{T} = [0.1, 0.9]$ is not necessarily more powerful than those based on individual quantile index for τ to take values not far away from 0.5. Nevertheless, for $n = 400$, we observe that the test based on $\mathcal{T} = [0.1, 0.9]$ is more powerful than those based on individual quantile index τ. This suggests the benefit of taking supremum over $\tau \in \mathcal{T}$ 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 $\tau = 0.1, 0.2, \ldots, 0.9$ across 500 repetitions. The results in Table 2 are as expected. First, the MSEs are generally larger when τ is close to 0 or 1 than when τ 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

Table 2. MSE of estimates of the quantile regression coefficients in the presence of threshold effect.

| | $n = 200$ | | | | $n = 400$ | | | |
| | α_τ | | β_τ | | α_τ | | β_τ | |
τ	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
Common threshold value ($c_0 = 0.5, c = 0$)								
0.1	0.322	1.764	0.542	2.767	0.170	0.819	0.153	0.848
0.2	0.211	1.208	0.228	1.242	0.117	0.553	0.119	0.682
0.3	0.176	0.979	0.207	1.057	0.110	0.566	0.099	0.535
0.4	0.169	0.902	0.205	1.053	0.100	0.535	0.091	0.500
0.5	0.166	0.890	0.214	1.070	0.098	0.542	0.080	0.472
0.6	0.170	0.908	0.199	0.999	0.114	0.572	0.085	0.470
0.7	0.194	0.952	0.221	1.207	0.109	0.543	0.100	0.529
0.8	0.235	1.251	0.314	1.719	0.124	0.634	0.098	0.517
0.9	0.300	1.559	0.797	4.898	0.150	0.704	0.156	0.836
Noncommon threshold value ($c_0 = 0.5, c = 0.5$)								
0.1	0.301	1.662	0.343	1.735	0.157	0.754	0.220	1.140
0.2	0.194	1.025	0.301	1.556	0.103	0.508	0.139	0.851
0.3	0.155	0.870	0.252	1.205	0.096	0.507	0.112	0.680
0.4	0.166	0.901	0.231	1.146	0.096	0.499	0.098	0.550
0.5	0.160	0.904	0.210	1.104	0.099	0.551	0.082	0.482
0.6	0.162	0.875	0.204	0.987	0.109	0.527	0.086	0.499
0.7	0.203	1.047	0.231	1.211	0.089	0.457	0.106	0.562
0.8	0.215	1.148	0.401	1.966	0.115	0.588	0.117	0.646
0.9	0.295	1.502	0.440	2.175	0.150	0.705	0.205	1.273

$\Pi(\cdot)$. Here, we specify $\Pi(\cdot)$ through its PDF $\pi(\cdot)$:

$$\pi(\tau) = 2 - 4(\tau - 0.5)\,\text{sgn}(\tau - 0.5), \qquad (3.2)$$

where $\text{sgn}(u) = -1$ if $u \le 0$ and 1 otherwise. One can check that $\pi(\tau)$ is non-negative and integrated into one on $[0, 1]$. Apparently, this way of specification of $\pi(\cdot)$ allows us to put more weight on τ when it is around 0.5 than when τ is close to either end because of the low estimation accuracy of the quantile regression coefficients when τ is close to 0 or 1. To test from the presence of common threshold on $\mathcal{T} = [\underline{\tau}, \overline{\tau}]$, we approximate the integral as follows:

$$\int_{\underline{\tau}}^{\overline{\tau}} A(\tau)\, d\Pi(\tau) \approx \frac{1}{N} \sum_{j=1}^{N} A(\tau_j)\, \pi(\tau_j),$$

where $A(\cdot)$ is an arbitrary smooth function and $\tau_1, \tau_2, \ldots, \tau_N$ are equally spaced on $\mathcal{T}$. 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 simulation results are not sensitive to the choice of N provided $N \ge 40$. To save on computation, we fix $N = 41$ in our simulations.

We then let c varies 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: $h_y = 1.06 s_y n^{-1/5}$ and $h_1 = 1.06 s_z n^{-1/6}$ where $s_z = (s_x, s_r)$, and s_y, s_x, and s_r denote 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
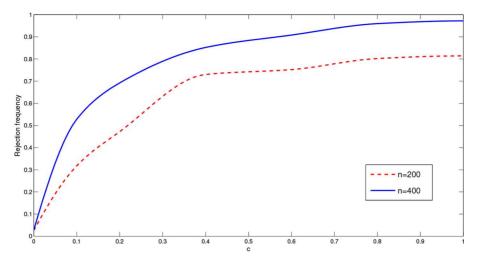
Figure 1. The power function of the NLR_n test for a common threshold value.

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 in 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 γ_τ^0 by inverting the $NLR_{n\tau}\left(\gamma_\tau^0\right)$ 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 τ 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 γ^0 by inverting the $\overline{NLR}_n\left(\gamma^0\right)$ given in (2.26).

Table 3 reports the results the empirical coverage ratio and average lengths for constructing the 95% confidence intervals of γ_τ^0 or γ^0. It shows that, for both empirical coverage ratio and average lengths of the estimated confidence intervals, tail index (i.e., when τ is close to 0 or 1) performs less satisfactory than the middle range of values for τ. 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 γ^0 seems to perform better in general than that of γ_τ^0.

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

Table 3. Coverage ratio for the 95% confidence interval of γ_τ and its average length.

		$n = 200$		$n=400$	
τ	γ_τ	Coverage ratio	Average length	Coverage ratio	Average length
Common threshold value ($c_0 = 0.5, c = 0$)					
0.1	0.5	0.996	0.681	0.976	0.674
0.2	0.5	0.982	0.659	0.980	0.656
0.3	0.5	0.970	0.637	0.960	0.609
0.4	0.5	0.974	0.610	0.946	0.567
0.5	0.5	0.948	0.594	0.946	0.527
0.6	0.5	0.950	0.588	0.946	0.526
0.7	0.5	0.962	0.600	0.934	0.549
0.8	0.5	0.976	0.628	0.946	0.600
0.9	0.5	0.982	0.660	0.978	0.612
0.1–0.9	0.5	0.918	0.572	0.946	0.513
Noncommon threshold value ($c_0 = 0.5, c = 0.5$)					
0.1	0.7	0.986	0.670	0.978	0.665
0.2	0.65	0.982	0.639	0.978	0.611
0.3	0.6	0.980	0.622	0.968	0.581
0.4	0.55	0.964	0.606	0.956	0.558
0.5	0.5	0.916	0.599	0.948	0.537
0.6	0.55	0.918	0.584	0.936	0.527
0.7	0.6	0.940	0.599	0.942	0.551
0.8	0.65	0.934	0.626	0.946	0.603
0.9	0.7	0.950	0.671	0.964	0.674

Table 4. The reputation scoring system on Taobao.com.

Category	Points needed	Category icon	Category	Points needed	Category icon
1	4–10	1-heart	11	10,001–20,000	1-crown
2	11–40	2-hearts	12	20,001–50,000	2-crowns
3	41–90	3-hearts	13	50,001–100,000	3-crowns
4	91–150	4-hearts	14	100,001–200,000	4-crowns
5	151–250	5-hearts	15	200,001–500,000	5-crowns
6	251–500	1-diamond	16	500,001–1,000,000	1-gold crown
7	501–1,000	2-diamonds	17	1,000,001–2,000,000	2-gold crowns
8	1,001–2,000	3-diamonds	18	2,000,001–5,000,000	3-gold crowns
9	2,001–5,000	4-diamonds	19	5,000,001–10,000,000	4-gold crowns
10	5,001–10,000	5-diamonds	20	10,000,000+	5-gold crowns

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."
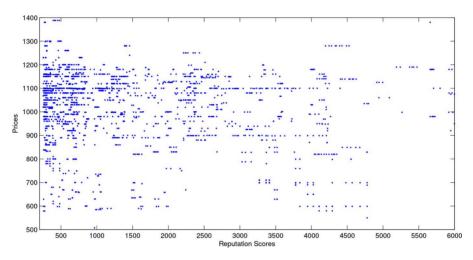
Figure 2. Scatter plot of the raw data.
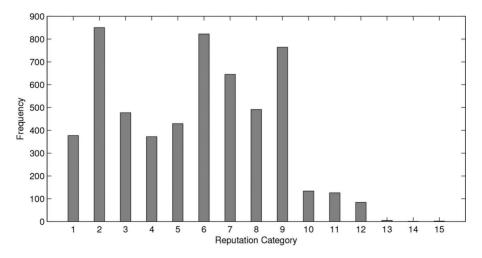
Figure 3. The histogram of reputation category.

4.2. 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 β. The empirical literature has documented extensive evidence to show that sellers with a superior reputation generate significantly higher profits. This β can be thought of as the discounted future profit from operating with a better business reputation. Thus, the seller's expected profit function is given by

$$\Pi_0(p; r) = \Pi_1(p; r) + \Pi_2(p; r)$$

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

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 Π_2 for any given (p, r). Therefore, Π_1 denotes the profit a seller obtains from the market without any concerns over reputation benefits, and Π_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 Π_0, which implies that the profit function Π_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 γ_0 and two different pricing regimes $p_1^*(r)$ and $p_2^*(r)$ such that the seller's optimal pricing rule $p^*(r)$ is*

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

where $\partial p_1^*(r)/\partial r < 0$.

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 γ_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 (i.e., $r < \gamma_0$). However, the dependence of pricing function with respect to r after γ_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 γ_0 hinges on the model parameters of α and β, which capture the market demand situation and the seller's perceived gain from future goodwill, respectively. We may naturally expect that the sellers are heterogeneous and that different sellers face different sets of such model parameters upon making their decisions. As in any typical empirical work, our economic model can also accommodate other covariates to control the observed heterogeneity across sellers. 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 5,000, 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 τ, 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

Table 5. The sup-Wald test for the existence of quantile threshold effect.

$\mathcal{T}$	Test stat	p-value	Critical value		
			10%	5%	1%
[0.1, 0.9]	596.11	0.00	19.88	21.54	25.09
0.1	151.43	0.00	15.68	17.34	20.63
0.2	190.86	0.00	15.31	16.80	20.13
0.3	279.62	0.00	15.28	16.71	20.28
0.4	359.55	0.00	15.28	16.72	19.95
0.5	596.11	0.00	15.19	16.60	19.77
0.6	103.87	0.00	15.18	16.55	20.32
0.7	61.00	0.00	15.24	16.68	20.35
0.8	12.47	0.36	15.46	16.98	20.19
0.9	9.25	0.89	15.94	17.67	21.41

Table 6. Estimation results.

τ	Jump size	γ	α_τ		β_τ	
			Intercept	Slope	Intercept	Slope
0.1	−197.88	3,264	856.10	**0.010**	1006.98	**−0.097**
			(6.21)	(0.003)	(34.15)	(0.029)
0.2	−224.28	3,264	947.22	**−0.007**	969.33	**−0.083**
			(5.73)	(0.003)	(30.59)	(0.025)
0.3	−265.17	3,264	992.30	−0.005	967.87	**−0.079**
			(6.04)	(0.004)	(43.50)	(0.044)
0.4	−300.63	3,264	1037.13	**−0.008**	735.55	−0.008
			(6.4)	(0.004)	(42.03)	(0.044)
0.5	−370.79	3,264	1080.00	0.000	339.14	**0.113**
			(6.53)	(0.004)	(42.61)	(0.043)
0.6	−224.47	3,240	1097.85	0.000	634.86	0.074
			(6.02)	(0.004)	(44.00)	(0.047)
0.7	−148.52	3,364	1121.87	**−0.007**	584.56	**0.108**
			(6.31)	(0.004)	(52.40)	(0.050)
Mean regression	−178.67	3,240	1043.00	−0.006	804.39	0.0125
			(4.78)	(0.004)	(89.36)	(0.023)

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.
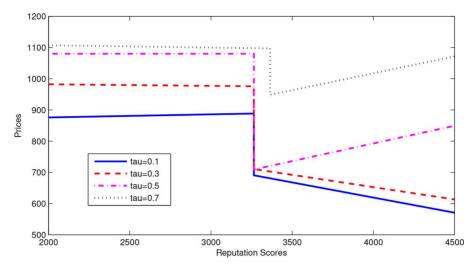
Figure 4. The estimated quantile regression lines before and after the jump point.

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 $\tau \in \mathcal{T}$. We implement the NLR_n test statistic in (2.22) by specifying three choices for $\mathcal{T}$, namely, $\mathcal{T} = [0.1, 0.5]$, $[0.1, 0.7]$, 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 $\mathcal{T} = [0.1, 0.5]$.

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 Ξ. 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 3,264, and its 95% confidence interval appears tighter than those of single quantile estimates.

Table 7. Inference on the threshold parameter.

τ	γ_τ or γ	95% lower bound	95% upper bound
0.1	3,264	3,168	3,271
0.2	3,264	3,182	3,271
0.3	3,264	3,168	3,271
0.4	3,264	3,166	3,271
0.5	3,264	3,231	3,271
0.6	3,240	3,232	3,248
0.7	3,364	3,337	3,367
0.1–0.5	3,264	3,197	3,271
Mean regression	3,240	3,134	3,272

Next, we address several concerns on the empirical exercise. First, we considered only linear specification in the previous analysis to ease the exhibition. But, nonlinearity 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 nonlinear 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 3,264 (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.

5. 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 through the use of a data set 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 $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.

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

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

Assumption A8. (i) $N_\tau(\gamma)$ and $D_\tau(\gamma)$ are continuous at γ_τ^0. $\nu_\tau' N_\tau \nu_\tau > 0$, $\nu_\tau' D_\tau \nu_\tau > 0$, and $g(\gamma_\tau^0) > 0$. (ii) $E \|x_i\|^4 < \infty$.

Assumption A9. (i) Let $\theta_1^0(\tau, \gamma) \equiv \operatorname{argmin}_{\theta_1} S(\theta_1; \tau, \gamma)$, where $S(\theta_1; \tau, \gamma) \equiv E[\rho_\tau(y_i - \theta_1' z_i(\gamma))]$. There exists a $\gamma^0 \in \Gamma$ such that $\gamma^0 = \operatorname{argmin}_\gamma S(\theta_1^0(\tau, \gamma); \tau, \gamma)$ for all τ. (ii) Let $\theta_{1\tau}^0 \equiv \theta_1^0(\tau, \gamma^0)$ and $\Delta(z_i, \tau, \gamma) \equiv \theta_1^0(\tau, \gamma)' z_i(\gamma) - \theta_{1\tau}^{0'} z_i(\gamma^0)$, and $\int P(|\Delta(z_i, \tau, \gamma)| > 0) d\Pi(\tau) > 0$ for all $\gamma \neq \gamma^0$. (iii) Let $\bar{\Omega}_1(\tau, \gamma) \equiv E[f(\theta_1^0(\tau, \gamma)' z_i(\gamma) | z_i) z_i(\gamma) z_i(\gamma)']$, and $\bar{\Omega}_1(\tau, \gamma)$ is positive definite for all $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.

Assumption A1 requires IID observations, but it can be weakened to allow for time series observations 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 $x_i = (1, r_i)'$, we can write $\alpha_\tau^0 = (\alpha_{0\tau}^0, \alpha_{1\tau}^0)'$ and $\beta_\tau^0 = (\beta_{0\tau}^0, \beta_{1\tau}^0)'$, where $\alpha_{0\tau}^0$ and $\beta_{0\tau}^0$ are the true values of the intercept parameters before and after the break, and $\alpha_{1\tau}^0$ and $\beta_{1\tau}^0$ are the true values of the slope parameters

before and after the break. Let $d_\tau \equiv (\alpha_{0\tau} - \beta_{0\tau}) + (\alpha_{1\tau} - \beta_{1\tau}) \gamma_\tau^0$. Then, a sufficient condition for Assumption A6 to hold is $d_\tau \neq 0$. 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 $E\left[\rho_\tau\left(y_i - \theta_1' z_i(\gamma)\right)\right]$ is convex in θ_1 for all $(\tau, \gamma) \in \mathcal{T} \times \Gamma$, and $\theta_1^0(\tau, \gamma)$ in A9(i) exists and is uniquely defined. It is also continuous in (τ, γ) by an application of the maximum theorem. The first-order condition for the minimization of $S(\theta_1; \tau, \gamma)$ with respect to θ_1 implies that

$$E\left[\psi_\tau\left(y_i - \theta_1^0(\tau, \gamma)' z_i(\gamma)\right) z_i(\gamma)\right] = 0 \text{ for all } (\tau, \gamma) \in \mathcal{T} \times \Gamma, \tag{A.1}$$

where $\psi_\tau(u) = \tau - 1\{u < 0\}$. (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 γ^0, which does not depend on $\tau \in \mathcal{T}$. Like Assumption A6, A9(ii) is an identification condition and requires that γ^0 be the unique common break. A9(iii) extends A5. Note that $\bar{\Omega}_1(\tau, \gamma^0) = \Omega_1(\tau, \gamma^0)$ 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 $\mathbb{H}_0$, $\hat{\theta}_1(\tau, \gamma)$ admits the following uniform Bahadur representation: $\sqrt{n}(\hat{\theta}_1(\tau, \gamma) - \theta_{1\tau}^0) = \Omega_1(\tau, \gamma)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(y_i - \alpha_\tau^{0'} x_i) z_i(\gamma) + o_P(1)$, where $\psi_\tau(u) = \tau - 1\{u < 0\}$, and $o_P(1)$ holds uniformly in $(\tau, \gamma) \in \mathcal{T} \times \Gamma$.*

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 $\theta_\tau \equiv (\theta_{1\tau}', \gamma_\tau)'$, $\theta_{1\tau} \equiv (\alpha_\tau', \beta_\tau')'$, and the true value of θ_τ, $\theta_{1\tau}$, α_τ, β_τ, and γ_τ is denoted as θ_τ^0, $\theta_{1\tau}^0$, α_τ^0, β_τ^0, and γ_τ^0, respectively. Let $\varepsilon_{i\tau}(\theta_\tau) \equiv y_i - \theta_{1\tau}' z_i(\gamma_\tau)$ and $\varepsilon_{i\tau} = \varepsilon_{i\tau}(\theta_\tau^0)$. Note that the τth conditional quantile of $\varepsilon_{i\tau}$ given z_i is 0, i.e., $E[\psi_\tau(\varepsilon_{i\tau})|z_i] = 0$. Let

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

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

$$\|z_i(\gamma_\tau) - z_i(\gamma_\tau^*)\| \leq \sqrt{2} \|x_i\| 1\left\{|r_i - \gamma_\tau| \leq |\gamma_\tau^* - \gamma_\tau|\right\}. \tag{B.2}$$

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

$$D_{n\tau}(\theta_{1\tau}, \gamma_\tau) = D_{1n\tau}\left(n^{1/2}(\theta_{1\tau} - \theta_{1\tau}^0)\right) + D_{2n\tau}(\theta_{1\tau}, \gamma_\tau). \tag{B.3}$$

Let $\bar{D}_{2n\tau}(w_1, w_2) \equiv D_{2n\tau}\left(\theta_{1\tau}^0 + n^{-1/2} w_1, \gamma_\tau^0 + n^{-1+2a} w_2\right)$. We will study the asymptotic properties of $D_{1n\tau}(w_1)$ and $\bar{D}_{2n\tau}(w_1, w_2)$ in Lemmas B.3 and B.4, respectively.

Lemma B.2. *Suppose that Assumptions A1–A2 and A4 hold. Then* $\lim_{\eta \to 0} E \sup_{\theta_\tau^* \in N_\eta(\theta_\tau)} |\rho_\tau(y_i - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*)) - \rho_\tau(y_i - \theta_{1\tau}' z_i(\gamma_\tau))| = 0$ *for any* $\theta_\tau \in \Theta$, *where* $N_\eta(\theta_\tau) \equiv \{\theta_\tau^* = (\theta_{1\tau}^{*\prime}, \gamma_\tau^*)' \in \Theta : \|\theta_{1\tau}^* - \theta_{1\tau}\| < \eta, |\gamma_\tau^* - \gamma_\tau| < \eta\}$ *denotes an* η-*neighborhood of* $\theta_\tau \in \Theta$ *and* $\eta > 0$.

Proof. Let $\Delta_{i\tau} \equiv \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*) - \theta_{1\tau}' z_i(\gamma_\tau)$. Then by the triangle inequality and (B.2),

$$|\Delta_{i\tau}| \le |\theta_{1\tau}'[z_i(\gamma_\tau) - z_i(\gamma_\tau^*)]| + \left|(\theta_{1\tau} - \theta_{1\tau}^*)' z_i(\gamma_\tau^*)\right|$$

$$\le \left\{ \sqrt{2} \times 1\{\gamma_\tau \wedge \gamma_\tau^* < r_i \le \gamma_\tau \vee \gamma_\tau^*\} \|\theta_{1\tau}\| + \|\theta_{1\tau} - \theta_{1\tau}^*\| \right\} \|x_i\|$$

$$\le \left\{ \sqrt{2} \times 1\{|r_i - \gamma_\tau| \le |\gamma_\tau - \gamma_\tau^*|\} \|\theta_{1\tau}\| + \|\theta_{1\tau} - \theta_{1\tau}^*\| \right\} \|x_i\|$$

$$\le \left\{ \sqrt{2} \times 1\{|r_i - \gamma_\tau| \le \eta\} \|\theta_{1\tau}\| + \eta \right\} \|x_i\|.$$

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

$$\rho_\tau\left(y_i - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right)$$

$$= \rho_\tau\left(\varepsilon_{i\tau}(\theta_\tau) - \Delta_{i\tau}\right) - \rho_\tau\left(\varepsilon_{i\tau}(\theta_\tau)\right)$$

$$= -\Delta_{i\tau} \psi_\tau\left(\varepsilon_{i\tau}(\theta_\tau)\right) + \int_0^{\Delta_{i\tau}} [1\{u_{i\tau} \le s\} - 1\{u_{i\tau} \le 0\}]\, ds.$$

It follows that $E \sup_{\theta_\tau^* \in N_\eta(\theta_\tau)} \left|\rho_\tau\left(y_i - \theta_{1\tau}^{*\prime} z_i(\gamma_\tau^*)\right) - \rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right)\right| \le 2E|\Delta_{i\tau}| \le 2\sqrt{2} \times [P(|r_i - \gamma_\tau| \le \eta)]^{1/2} \|x_i\|_2 \|\theta_{1\tau}\| + 2\sqrt{2}\eta E \|x_i\| \to 0$ as $\eta \to 0$, where $\|x_i\|_2 \equiv \{E \|x_i\|^2\}^{1/2}$.

Lemma B.3. *Suppose Assumptions A1–A5 hold. Then* $\sup_{\|w_1\| \le M} |D_{1n\tau}(w_1) + n^{-1/2} w_1' \sum_{i=1}^n z_i(\gamma_\tau^0) \psi_\tau(\varepsilon_{i\tau}) - \frac{1}{2} w_1' \Omega_{1\tau} w_1| = o_P(1)$ *for every* $M \in (0, \infty)$, *where* $\Omega_{1\tau} \equiv \Omega_1(\tau, \gamma_\tau^0) = E[z_i(\gamma_\tau^0) z_i(\gamma_\tau^0)' f(\theta_{1\tau}^{0\prime} z_i(\gamma_\tau^0) |z_i)]$.

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

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

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

The pointwise convergence of $d_{1n\tau,1}(w_1)$ 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 $d_{1n\tau,2}(w_1)$, we apply the Taylor expansion with integral remainder and the triangle inequality to

obtain

$$\sup_{\|w_1\| \le b} \left| d_{1n\tau,2}(w_1) \right|$$

$$= \sup_{\|w_1\| \le b} \left| n^{-1/2} w_1' \sum_{i=1}^{n} z_i\left(\gamma_\tau^0\right) \int_0^1 \left[F\left(\varsigma_{i\tau} + s n^{-1/2} w_1' z_i\left(\gamma_\tau^0\right) | z_i\right) - F\left(\varsigma_{i\tau} | z_i\right) \right] ds - \frac{1}{2} w_1' \Omega_{1\tau} w_1 \right|$$

$$\le \sup_{\|w_1\| \le b} \left| n^{-1} w_1' \sum_{i=1}^{n} z_i\left(\gamma_\tau^0\right) z_i\left(\gamma_\tau^0\right)' w_1 \int_0^1 \int_0^1 \left[f\left(\varsigma_{i\tau} + stn^{-1/2} w_1' z_i\left(\gamma_\tau^0\right) | z_i\right) - f\left(\varsigma_{i\tau} | z_i\right) \right] dtds \right|$$

$$+ \frac{1}{2} \sup_{\|w_1\| \le M} \left| n^{-1} w_1' \sum_{i=1}^{n} \left[z_i\left(\gamma_\tau^0\right) z_i\left(\gamma_\tau^0\right)' f\left(\varsigma_{i\tau} | z_i\right) - \Omega_{1\tau} \right] w_1 \right|$$

$$\le \left| M^2 n^{-1} \sum_{i=1}^{n} \left\| z_i\left(\gamma_\tau^0\right) z_i\left(\gamma_\tau^0\right)' \right\| \int_0^1 \int_0^1 \sup_{\|w_1\| \le M} \left| f\left(\varsigma_{i\tau} + stn^{-1/2} w_1' z_i\left(\gamma_\tau^0\right) | z_i\right) - f\left(\varsigma_{i\tau} | z_i\right) \right| dtds \right|$$

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

$$= o_P(1) + o_P(1) = o_P(1),$$

where $\varsigma_{i\tau} \equiv \theta_{1\tau}^{0\prime} z_i\left(\gamma_\tau^0\right)$, 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* $\sup_{\|w_1\| \le M} |\bar{D}_{2n\tau}(w_1, w_2) - \bar{D}_{2n\tau}(0, w_2)| = o_P(1)$ *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 $1_i(r) \equiv 1\{\gamma_\tau^0 < r_i \le \gamma_\tau^0 + r\}$. Let $r \equiv \gamma_\tau - \gamma_\tau^0$ and $\Delta_{1\tau} \equiv \theta_{1\tau} - \theta_{1\tau}^0$, $a_1(x) \equiv \left(\mathbf{0}_{1 \times p}, x'\right)'$, $a_2(x) \equiv \left(x', \mathbf{0}_{1 \times p}\right)'$ and $a(x) \equiv a_1(x) - a_2(x)$, where $\mathbf{0}_{1 \times p}$ denotes a $1 \times p$ vector of zeros. Noting that $z_i(\gamma_\tau) - z_i\left(\gamma_\tau^0\right) = -a(x_i) 1_i(r)$ when $r > 0$, we observe that: (i) if $1_i(r) = 0$,

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

$$= \varepsilon_{i\tau} - \Delta_{1\tau}' z_i\left(\gamma_\tau^0\right) = y_i - \theta_{1\tau}' z_i\left(\gamma_\tau^0\right);$$

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

$$\rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right) - \rho_\tau(y_i - \theta_{1\tau}' z_i\left(\gamma_\tau^0\right))$$

$$= \left[\rho_\tau\left(y_i - \theta_{1\tau}' z_i(\gamma_\tau)\right) - \rho_\tau(y_i - \theta_{1\tau}' z_i\left(\gamma_\tau^0\right)) \right] 1_i(r)$$

$$= \left\{ \rho_\tau\left(\varepsilon_{i\tau} - \theta_{1\tau}^{0\prime} \left[z_i(\gamma_\tau) - z_i\left(\gamma_\tau^0\right) \right] - \Delta_{1\tau}' z_i(\gamma_\tau)\right) - \rho_\tau\left(\varepsilon_{i\tau} - \Delta_{1\tau}' z_i\left(\gamma_\tau^0\right)\right) \right\} 1_i(r)$$

$$= \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a(x_i) - \Delta_{1\tau}' a_2(x_i)\right) - \rho_\tau\left(\varepsilon_{i\tau} - \Delta_{1\tau}' a_1(x_i)\right) \right] 1_i(r). \tag{B.4}$$

(Similarly, if $r < 0$, then $\rho_\tau\left(y_i - \theta'_{1\tau}\dot{m}_{\gamma_\tau}(z_i)\right) - \rho_\tau(y_i - \theta'_{1\tau}\dot{m}_{\gamma^0_\tau}(z_i)) = [\rho_\tau(\varepsilon_{i\tau} - \theta^{0\prime}_{1\tau}a(x_i) - \Delta'_{1\tau}a_1(x_i)) - \rho_\tau\left(\varepsilon_{i\tau} - \Delta'_{1\tau}a_2(x_i)\right)]\bar{1}_i(r)$, where $\bar{1}_i(r) = 1\left\{\gamma^0_\tau + r < q_i \leq \gamma^0_\tau\right\}$.) By Knight's identity, we have

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

$$= D_{2n\tau}\left(\theta^0_{1\tau} + n^{-1/2}w_1, \gamma^0_\tau + n^{-1+2a}w_2\right) - D_{2n\tau}\left(\theta^0_{1\tau}, \gamma^0_\tau + n^{-1+2a}w_2\right)$$

$$= \sum_{i=1}^{n}\left\{[\rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau}a(x_i) - n^{-1/2}w'_1a_2(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})]\right.$$

$$- [\rho_\tau\left(\varepsilon_{i\tau} + \theta^{0\prime}_{1\tau}a(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})]$$

$$\left. - [\rho_\tau\left(\varepsilon_{i\tau} - n^{-1/2}w'_1a_1(x_i)\right) - \rho_\tau(\varepsilon_{i\tau})]\right\}1_i\left(n^{-1+2a}w_2\right)$$

$$= n^{-1/2}\sum_{i=1}^{n}\psi_\tau(\varepsilon_{i\tau})w'_1a(x_i)\,1_i\left(n^{-1+2a}w_2\right)$$

$$+ \sum_{i=1}^{n}\int_{-\theta^{0\prime}_{1\tau}a(x_i)}^{-\theta^{0\prime}_{1\tau}a(x_i)+n^{-1/2}w'_1a_2(x_i)}[1\left\{\varepsilon_{i\tau} \leq s\right\} - 1\left\{\varepsilon_{i\tau} \leq 0\right\}]\,ds\,1_i\left(n^{-1+2a}w_2\right)$$

$$- \sum_{i=1}^{n}\int_{0}^{n^{-1/2}w'_1a_1(x_i)}[1\left\{\varepsilon_{i\tau} \leq s\right\} - 1\left\{\varepsilon_{i\tau} \leq 0\right\}]\,ds\,1_i\left(n^{-1+2a}w_2\right)$$

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

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 $m_1(y, z; w_1, \gamma) = \psi_\tau(y - \theta^{0\prime}_{1\tau}z(\gamma))w'_1a(x)\,1\{q \leq \gamma\}$, $z(\gamma) = [x'1\{q \leq \gamma\}\ x'1\{q > \gamma\}]'$, and $\Theta_M = \{\omega_1 : \|\omega_1\| \leq M\}$. Let

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

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

$$\mathcal{F}_{1,3} = \left\{f_{1,3}(y, z; w_1) = w'_1a(x) : w_1 \in \Theta_M\right\}.$$

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 $\theta^{0\prime}_{1\tau}z(\gamma) = \beta^{0\prime}_\tau x + \delta^{0\prime}_\tau x1\{q \leq \gamma\}$ where $\delta^0_\tau = \alpha^0_\tau - \beta^0_\tau$, $\mathcal{F}_{1,2}$ 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

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

$$= o_P(1)$$

as $n^{-1+2a} \to 0$ as $n \to \infty$ by Assumption A7. Next we study $d_{2n\tau,3}(w_1)$. Write $d_{2n\tau,3}(w_1) = n^{-1/2}\sum_{i=1}^{n}w'_1 \times a_1(x_i)\int_0^1[1\{\varepsilon_{i\tau} \leq n^{-1/2}w'_1a_1(x_i)s\} - 1\{\varepsilon_{i\tau} \leq 0\}]ds\,1_i\left(n^{-1+2a}w_2\right)$. Let $m_2(y, z; w_1, \bar{w}_1) = w'_1a_1(x)\int_0^1 1\{y - \theta^{0\prime}_{1\tau}z(\gamma) \leq \bar{w}_1a_1(x)s\}1\{\gamma^0_\tau < q \leq \gamma^0_\tau + n^{-1+2a}w_2\}$. 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_{2,2}(y, z; \bar{w}_1) = $

$\int_0^1 1\{y - \theta_{1\tau}^{0\prime} z\left(\gamma_\tau^0\right) \leq \bar{w}_1' a_1\left(x\right) s\}ds : \bar{w}_1 \in \Theta_M\}$. 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 $1\left\{\gamma_\tau^0 < q \leq \gamma_\tau^0 + n^{-1+2a}w_2\right\}$, it also satisfies that Pollard's entropy condition and is stochastically equicontinuous with respect to the pseudometric defined by:

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

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

$$\sup_{\|w_1\| \leq M}\left|d_{2n\tau,3}\left(w_1\right)\right|$$

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

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

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

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

$$\leq M\left|n^{-1/2}\sum_{i=1}^n E\left\{\|a_1\left(x_i\right)\|\int_0^1\left[F\left(n^{-1/2}M\|a_1\left(x_i\right)\|s\mid z_i\right) - F\left(0|z_i\right)\right]ds\, 1_i\left(n^{-1+2a}w_2\right)\right\}\right|$$
$$+ o_P\left(1\right)$$

$$\leq Mn^{1/2}\left\|\|a_1\left(x_i\right)\|\int_0^1\left[F\left(n^{-1/2}M\|a_1\left(x_i\right)\|s\mid z_i\right) - F\left(0|z_i\right)\right]ds\right\|_2 \left\|1_i\left(n^{-1+2a}w_2\right)\right\|_2 + o_P\left(1\right)$$

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

where $\|A\|_2 \equiv \{E\|A\|^2\}^{1/2}$. By the same token, we can show that $\sup_{\|w_1\| \leq M}\left|d_{2n\tau,2}\left(w_1\right)\right| = o_P\left(1\right)$. This completes the proof of the lemma.

Proof of Theorem 2.1. Recall that $l^\infty\left(\mathcal{T} \times \Gamma\right)$ denotes the space of all bounded functions on $\mathcal{T} \times \Gamma$ equipped with the uniform topology. From Lemma B.1, we can easily obtain that under $\mathbb{H}_0$,

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

Note that $\hat{\delta}\left(\tau, \gamma\right) = \hat{\alpha}\left(\tau, \gamma\right) - \hat{\beta}\left(\tau, \gamma\right) = R\hat{\theta}_1\left(\tau, \gamma\right)$ and $\hat{V}(\tau, \gamma)$ is a uniformly consistent estimator of $V\left(\tau, \gamma\right)$ 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 $\hat{\theta}_\tau$ is also minimizing $\bar{S}_{n\tau}\left(\theta_\tau\right) \equiv n^{-1}\sum_{i=1}^n s_\tau\left(y_i, z_i; \theta_\tau\right)$, where $s_\tau\left(y_i, z_i; \theta_\tau\right) \equiv \rho_\tau(y_i - \theta_{1\tau}'z_i(\gamma_\tau)) - \rho_\tau\left(y_i - \theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0\right)\right)$. Let $\Delta\left(z_i, \theta_\tau\right) \equiv \theta_{1\tau}'z_i\left(\gamma_\tau\right) -$

$\theta_{1\tau}^{0\prime} z_i \left(\gamma_\tau^0\right)$ and $\varsigma_\tau \left(\theta_\tau\right) \equiv E\left[s_\tau \left(y_i, z_i; \theta_\tau\right)\right]$. Then by Knight's (1998) identity, the compactness of Θ_1 and Assumption A2, $E\left|s_\tau \left(y_i, x_i; \theta_\tau\right)\right| \leq 2E\left|\Delta \left(z_i, \theta_\tau\right)\right| \leq 2\left(\|\alpha_\tau\| + \|\beta_\tau\|\right) E\|x_i\| < \infty$. This ensures that $S_{n\tau} \left(\theta_\tau\right) = \varsigma_\tau \left(\theta_\tau\right) + o_{a.s.}$ (1) for each $\theta_\tau \in \Theta$ by the strong law of large numbers (LLN). By the proof of Lemma 2 in Galvao et al. (2011), the class of functions $\mathcal{F} \equiv \left\{s_\tau \left(y, z; \theta\right) : \theta \in \Theta\right\}$ is Glivenko–Cantelli. It follows that $\bar{S}_{n\tau} \left(\theta_\tau\right) = \varsigma_\tau \left(\theta_\tau\right) + o_{a.s.}$ (1) uniformly in $\theta_\tau \in \Theta$. Let $l_\tau \left(c\right) \equiv E\left[\rho_\tau \left(\varepsilon_{i\tau} - c\right) - \rho_\tau \left(\varepsilon_{i\tau}\right)\right]$. Knight's identity implies $l_\tau \left(c\right) > 0$ for any $c \neq 0$. Then by the law of iterated expectations and Assumption A6, we have

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

By Lemma B.2 $\varsigma_\tau \left(\theta_\tau\right)$ is continuous in θ_τ. It follows that θ_τ^0 is the unique minimizer of $\varsigma_\tau \left(\theta_\tau\right)$ 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

$$n^{1/2}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0) = O_P \left(1\right) \text{ and } n^{1-2a}(\hat{\gamma}_\tau - \gamma_\tau^0) = O_P \left(1\right). \tag{B.5}$$

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

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

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

$$P\left(\inf_{\theta_\tau \in N_{jb}} D_{n\tau} \left(\theta_{1\tau}, \gamma_\tau\right) > c_j\right) > 1 - \kappa \text{ for } n > n_0, j = 1, 2, \tag{B.6}$$

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

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

it suffices to analyze $D_{n\tau, 11}$, $D_{n\tau, 12}$, $D_{n\tau, 21}$, and $D_{n\tau, 22}$. We first analyze $D_{n\tau, 12}$. By Koul et al. (2003) it suffices to show that for all $\kappa \in \left(0, 1\right]$, $c_1 \in \left(0, \infty\right)$, there exists $c_0 < \infty$, $b_0 \in \left(0, \infty\right)$, $\sigma \in \left(0, 1\right)$ and n_0 such that $c_0 b_0 g \left(\gamma_\tau^0\right) / 2 > c_1$ and that

$$P\left(\inf_{\theta_\tau \in N_{jb}} \frac{D_{2n\tau} \left(\theta_{1\tau}, \gamma_\tau\right)}{nK \left(\left|\gamma_\tau - \gamma_\tau^0\right|\right)} > c_0\right) > 1 - \kappa/2 \text{ for all } n > n_0, \tag{B.7}$$

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

$$n^{-1}D_{2n\tau} \left(\theta_{1\tau}, \gamma_\tau\right) = n^{-1} \sum_{i=1}^{n} \left[\rho_\tau \left(y_i - \theta_{1\tau}' z_i \left(\gamma_\tau\right)\right) - \rho_\tau(y_i - \theta_{1\tau}' z_i \left(\gamma_\tau^0\right))\right]$$

$$= n^{-1} \sum_{i=1}^{n} \left[\rho_\tau \left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a \left(x_i\right) - \Delta_{1\tau}' a_2 \left(x_i\right)\right) - \rho_\tau \left(\varepsilon_{i\tau} - \Delta_{1\tau}' a_1 \left(x_i\right)\right)\right] 1_i \left(r\right)$$

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

where

$$d_{n\tau 1}\left(\theta_{1\tau}, r\right) = n^{-1}\sum_{i=1}^{n}\left[\rho_{\tau}\left(\varepsilon_{i\tau}\right) - \rho_{\tau}\left(\varepsilon_{i\tau} - \Delta_{1\tau}' a_1\left(x_i\right)\right)\right]1_i\left(r\right),$$

$$d_{n\tau 2}\left(\theta_{1\tau}, r\right) = n^{-1}\sum_{i=1}^{n}\left[\rho_{\tau}\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a\left(x_i\right) - \Delta_{1\tau}' a_2\left(x_i\right)\right) - \rho_{\tau}\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a\left(x_i\right)\right)\right]1_i\left(r\right),$$

$$d_{n\tau 3}\left(r\right) = n^{-1}\sum_{i=1}^{n}\left[\rho_{\tau}\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a\left(x_i\right)\right) - \rho_{\tau}\left(\varepsilon_{i\tau}\right) - p\left(r_i\right)\right]1_i\left(r\right),$$

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

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

$$d_{n\tau 6}\left(r\right) = p\left(\gamma_{\tau}^0\right) K\left(r\right),$$

$p\left(\gamma_{\tau}^0\right) = E\{\left[\rho_{\tau}\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a\left(x_i\right)\right) - \rho_{\tau}\left(\varepsilon_{i\tau}\right)\right]|r_i = \gamma_{\tau}^0\}$ and $K_n\left(r\right) = n^{-1}\sum_{i=1}^{n} 1_i\left(r\right)$. By Knight's identity, the law of iterated expectations, and Fubini's theorem, we can readily show that $p\left(\gamma\right)$ is strictly positive and continuous in γ 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 $K\left(r\right)$ 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 $d_{n\tau 4}\left(r\right)$ and $d_{n\tau 5}\left(r\right)$ because then $\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|d_{n\tau 4}\left(r\right)/K\left(r\right)\right| \leq \sup_{0\leq r\leq\sigma}\left|p\left(\gamma_{\tau}^0 + r\right) - p\left(\gamma_{\tau}^0\right)\right|$ $\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|K_n\left(r\right)/K\left(r\right)\right| \to 0$ and $\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|d_{n\tau 5}\left(r\right)/K\left(r\right)\right| \leq C\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|K_n\left(r\right)/K\left(r\right) - 1\right| \to 0$ and 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 $d_{n\tau 3}\left(r\right)/K\left(r\right)$ 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) $P\left(\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|\frac{K_n(r)}{K(r)} - 1\right| < c\right) > 1 - \kappa,$

(2) $P\left(\sup_{B/n^{1-2a}\leq r\leq\sigma}\left|\frac{R_n(r)}{K(r)}\right| < c\right) > 1 - \kappa,$
   where $R_n\left(r\right) = n^{-1}\sum_{i=1}^{n}\{J\left(x_i, \varepsilon_{i\tau}\right) - E\left[J\left(x_i, \varepsilon_{i\tau}\right)|r_i = r\right]\}1_i\left(r\right)$ and $J\left(x_i, \varepsilon_{i\tau}\right) \equiv \rho_{\tau}\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime} a\left(x_i\right)\right) - \rho_{\tau}\left(\varepsilon_{i\tau}\right).$

It follows that $n^{-1}D_{2n\tau}\left(\theta_{1\tau}, \gamma_{\tau}\right)/K\left(r\right) = p\left(\gamma_{\tau}^0\right) + o_P\left(1\right) > c > 0$ with probability approaching 1 as $n \to \infty$ uniformly in $\theta_{\tau} \in N_{1b}$, 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 $D_{n\tau}\left(\theta_{1\tau}, \gamma_{\tau}\right)$ by restricting our attention to the case where $n^{1/2}\left\|\theta_{1\tau} - \theta_{1\tau}^0\right\| \leq M$ and $n^{1-2a}\left|\gamma_{\tau} - \gamma_{\tau}^0\right| \leq M$ for some large but fixed positive number M. Let $w_{1n\tau} \equiv n^{1/2}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0)$ and $w_{2n\tau} \equiv n^{1-2a}\left(\hat{\gamma}_{\tau} - \gamma_{\tau}^0\right)$. Then by (B.3), $D_{n\tau}\left(\theta_{1\tau}^0 + n^{-1/2}w_1, \gamma_{\tau}^0 + n^{-1+2a}w_2\right) = D_{1n\tau}\left(w_1\right) + \bar{D}_{2n\tau}(w_1, w_2)$, where recall that $\bar{D}_{2n\tau}\left(w_1, w_2,\right) = D_{2n\tau}\left(\theta_{1\tau}^0 + n^{-1/2}w_1, \gamma_{\tau}^0 + n^{-1+2a}w_2\right)$. By Lemmas B.3 and B.4, we have

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

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

$$
\bar{D}_{2n\tau}(0, w_2) = \begin{cases}
\sum_{i=1}^{n} \left[\rho_\tau\left(\varepsilon_{i\tau} + \theta_{1\tau}^{0\prime}a\left(x_i\right)\right) - \rho_\tau\left(\varepsilon_{i\tau}\right)\right] 1_i\left(n^{-1+2a}w_2\right) & \text{if } w_2 > 0 \\[3mm]
\sum_{i=1}^{n} \left[\rho_\tau\left(\varepsilon_{i\tau} - \theta_{1\tau}^{0\prime}a\left(x_i\right)\right) - \rho_\tau\left(\varepsilon_{i\tau}\right)\right] \bar{1}_i\left(n^{-1+2a}w_2\right) & \text{if } w_2 \leq 0
\end{cases} \tag{B.8}
$$

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

$$
w_{1n\tau} = n^{-1/2}\Omega_{1\tau}^{-1}\sum_{i=1}^{n} z_i\left(\gamma_\tau^0\right)\psi_\tau\left(\varepsilon_{i\tau}\right) + o_P(1) \xrightarrow{d} N\left(\mathbf{0}_{2k\times 1}, \tau\left(1-\tau\right)\Sigma\left(\tau, \gamma_\tau^0\right)\right). \tag{B.9}
$$

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

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

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

$$
= \sum_{i=1}^{n}\left[\rho_\tau\left(\varepsilon_{i\tau} + \delta_\tau^{0\prime}\Delta x_i\left(w_2\right)\right) - \rho_\tau\left(\varepsilon_{i\tau}\right)\right],
$$

where $\Delta x_i(w_2) = x_i\left[1\left\{r_i \leq \gamma_\tau^0 + n^{-1+2a}w_2\right\} - 1\left\{r_i \leq \gamma_\tau^0\right\}\right]$ and $\delta_\tau^0 = \delta_{n\tau}^0 = \alpha_{n\tau}^0 - \beta_{n\tau}^0 = v_\tau n^{-a}$. By Knight's identity, $\bar{D}_{2n\tau}(0, w_2) = \sum_{i=1}^{n}\psi_\tau\left(\varepsilon_{i\tau}\right)\delta_\tau^{0\prime}\Delta x_i\left(w_2\right) + \sum_{i=1}^{n}\int_0^{-\delta_\tau^{0\prime}\Delta x_i(w_2)}[1\{\varepsilon_{i\tau} \leq s\} - 1\{\varepsilon_{i\tau} \leq 0\}]ds \equiv \bar{D}_{2n\tau,1}(w_2) + \bar{D}_{2n\tau,2}(w_2)$, say. Assume that $w_2 > 0$. Using arguments similar to those used in the proof of Lemma A.11 in Hansen (2000), we can readily show that

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

where $B_\tau(w_2)$ is a Brownian motion with variance $E[B_\tau(1)^2] = \tau\left(1-\tau\right)v_\tau'E[x_ix_i'|r_i = \gamma_\tau^0]v_\tau g\left(\gamma_\tau^0\right) \equiv \lambda_\tau$. Analogously to the proof of Lemma 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

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

$$
= \frac{1}{2}\delta_\tau^{0\prime}\sum_{i=1}^{n}\Delta x_i\left(w_2\right)\Delta x_i\left(w_2\right)'\delta_\tau^0 f\left(\theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0\right)|z_i\right) + o_P(1)
$$

$$
= \frac{1}{2}\delta_\tau^{0\prime}\sum_{i=1}^{n}x_ix_i'f\left(\theta_{1\tau}^{0\prime}z_i\left(\gamma_\tau^0\right)|z_i\right)\left[1\left\{r_i \leq \gamma_\tau^0 + n^{-1+2a}w_2\right\} - 1\left\{r_i \leq \gamma_\tau^0\right\}\right]^2\delta_\tau^0 + o_P(1)
$$

$$
= \frac{1}{2}\mu_\tau|w_2| + o_P(1),
$$

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

$$\bar{D}_{2n\tau}\left(0, w_2\right) \Rightarrow \frac{1}{2}\mu_\tau \left|w_2\right| - \sqrt{\lambda_\tau} W_1\left(w_2\right) \text{ if } w_2 > 0. \tag{B.10}$$

Similarly, we can show that

$$\bar{D}_{2n\tau}\left(0, w_2\right) \Rightarrow \frac{1}{2}\mu_\tau \left|w_2\right| - \sqrt{\lambda_\tau} W_2\left(-w_2\right) \text{ if } w_2 \leq 0, \tag{B.11}$$

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

$$
\begin{aligned}
w_{2n\tau} &\xrightarrow{d} \underset{-\infty < w_2 < \infty}{\arg\max} \; -\frac{1}{2}\mu_\tau \left|w_2\right| + \sqrt{\lambda_\tau} W\left(w_2\right) \\
&= \frac{\lambda_\tau}{\mu_\tau^2} \underset{-\infty < r < \infty}{\arg\max} \; -\frac{1}{2}\mu_\tau \left|\frac{\lambda_\tau}{\mu_\tau^2}r\right| + \sqrt{\lambda_\tau} W\left(\frac{\lambda_\tau}{\mu_\tau^2}r\right) \\
&= \frac{\lambda_\tau}{\mu_\tau^2} \underset{-\infty < r < \infty}{\arg\max} \; -\frac{\lambda_\tau}{2\mu_\tau}\left|r\right| + \frac{\lambda_\tau}{\mu_\tau} W\left(r\right) = \frac{\lambda_\tau}{\mu_\tau^2} \underset{-\infty < r < \infty}{\arg\max} \left\{-\frac{1}{2}\left|r\right| + W\left(r\right)\right\}, \quad \text{(B.12)}
\end{aligned}
$$

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

Proof of Theorem 2.4. Recall that $w_{1n\tau} \equiv n^{1/2}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0)$ and $w_{2n\tau} \equiv n^{1-2a}\left(\hat{\gamma}_\tau - \gamma_\tau^0\right)$. By ([B.3](#)), the relationship between $D_{2n\tau}$ and $\bar{D}_{2n\tau}$, and Lemma [B.4](#), we have $LR_{n\tau}\left(\gamma_\tau^0\right) = \bar{D}_{2n\tau}(w_{1n\tau}, 0) - \bar{D}_{2n\tau}(w_{1n\tau}, w_{2n\tau}) = -\bar{D}_{2n\tau}\left(0, w_{2n\tau}\right) + o_P\left(1\right)$. By the analysis of $\bar{D}_{2n\tau}\left(0, w_2\right)$ in the the proof of Theorem [2.3](#)(ii) and the CMT, implies that

$$LR_{n\tau}\left(\gamma_\tau^0\right) = -\bar{D}_{2n\tau}\left(0, w_{2n\tau}\right) + o_P\left(1\right) \xrightarrow{d} \sup_{w_2} \left\{-\frac{1}{2}\mu_\tau \left|w_2\right| + \sqrt{\lambda_\tau} W\left(w_2\right)\right\} \tag{B.13}$$

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

$$
\begin{aligned}
\sup_{w_2} \left\{-\frac{1}{2}\mu_\tau \left|w_2\right| + \sqrt{\lambda_\tau} W\left(w_2\right)\right\} &= \sup_{r} \left\{-\frac{1}{2}\mu_\tau \left|\frac{\lambda_\tau}{\mu_\tau^2}r\right| + \sqrt{\lambda_\tau} W\left(\frac{\lambda_\tau}{\mu_\tau^2}r\right)\right\} \\
&= \frac{\lambda_\tau}{\mu_\tau} \sup_{r} \left\{-\frac{1}{2}\left|r\right| + W\left(r\right)\right\}. \quad \text{(B.14)}
\end{aligned}
$$

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

Proof of Theorem 2.5.

(i) First, observe that $\hat{\theta}_1\left(\tau, \gamma\right)$ defined in ([2.9](#)) is also minimizing $\bar{S}_n\left(\theta_1, \tau, \gamma\right) \equiv n^{-1} \sum_{i=1}^{n} s\left(y_i, z_i; \theta_1, \tau, \gamma\right)$ with respect to θ_1, where $s\left(y_i, z_i; \theta_1, \tau, \gamma\right) \equiv \rho_\tau(y_i - \theta_1' z_i\left(\gamma\right)) - \rho_\tau\left(y_i - \theta_{1\tau}^{0\prime} z_i\left(\gamma^0\right)\right)$. Let $\varsigma\left(\theta_1, \gamma, \tau\right) \equiv E[s\left(y_i, z_i; \theta_1, \tau, \gamma\right)]$. By Lemma [B.2](#), $\varsigma\left(\theta_1, \tau, \gamma\right)$ is continuous in $\left(\theta_1, \gamma\right)$. It is straightforward to show that it is also continuous in τ. Thus $\varsigma\left(\theta_1, \tau, \gamma\right)$ is continuous over $\Theta_1 \times \mathcal{T} \times \Gamma$. By Lemma 2 in Galvao et al. ([2011](#)), $\{s\left(y, z; \theta_1, \gamma, \tau\right): \left(\theta_1, \tau, \gamma\right) \in \Theta_1 \times \mathcal{T} \times \Gamma\}$ is Glivenko–Cantelli. In conjunction with the pointwise convergence, this implies that $\sup_{(\theta_1, \tau, \gamma) \in \Theta_1 \times \mathcal{T} \times \Gamma} \left|\bar{S}_n\left(\theta_1, \tau, \gamma\right) - \varsigma\left(\theta_1, \tau, \gamma\right)\right| \xrightarrow{p} 0$. As remarked after Assumption [A9](#), $\theta_1^0\left(\tau, \gamma\right) = \arg\min_{\theta_1 \in \Theta_1} \varsigma\left(\theta_1, \tau, \gamma\right)$ is uniquely defined. It follows

from Lemma B.1 of Chernozhukov and Hansen (2006) that

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

Let $D_n\left(\gamma\right) \equiv \hat{S}_{n\pi}\left(\gamma\right) - \sum_{i=1}^n \int \rho_\tau\left(y_i - \theta_{1\tau}^{0\prime}z_i\left(\gamma^0\right)\right) d\Pi\left(\tau\right)$. (B.15) implies that uniformly over Γ

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

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

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

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

where $D\left(\gamma\right) = \int E\left[\rho_\tau\left(\varepsilon_{i\tau} - (\theta_1^0\left(\tau,\gamma\right) z_i\left(\gamma\right) - \theta_{1\tau}^{0\prime}z_i\left(\gamma^0\right))\right) - \rho_\tau\left(\varepsilon_{i\tau}\right)\right] d\Pi\left(\tau\right)$ and $\varepsilon_{i\tau} = y_i - \theta_{1\tau}^{0\prime}z_i\left(\gamma^0\right)$. Again, by the fact that $E\left[\rho_\tau\left(\varepsilon_{i\tau} - c\right) - \rho_\tau\left(\varepsilon_{i\tau}\right)\right] > 0$ for all $c \neq 0$, $D\left(\gamma\right)$ is minimized iff $\theta_1^0\left(\tau,\gamma\right) z_i\left(\gamma\right) = \theta_{1\tau}^{0\prime}z_i\left(\gamma^0\right)$ a.s., i.e., iff $\gamma = \gamma^0$ by Assumption A9(ii). By invoking Lemma B.1 of Chernozhukov and Hansen (2006) again, we have $\hat{\gamma} = \gamma^0 + o_P\left(1\right)$. This, in conjunction with (B.15) and the continuity of $\theta_1^0\left(\tau, \cdot\right)$, implies that $\hat{\theta}_1\left(\tau\right) = \hat{\theta}_1\left(\tau, \hat{\gamma}\right) = \theta_1^0\left(\tau, \hat{\gamma}\right) + o_P\left(1\right) = \theta_{1\tau}^0 + o_P\left(1\right)$.

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

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

$$= \mathbb{G}_n f_1\left(u_i; \hat{\theta}_1\left(\tau,\gamma\right), \tau, \gamma\right) + \sqrt{n}E\left[f_1\left(u_i; \theta_1, \tau, \gamma\right)\right]|_{\theta_1=\hat{\theta}_1(\tau,\gamma)}$$

$$= \mathbb{G}_n f_1\left(u_i; \theta_1^0\left(\tau,\gamma\right), \tau, \gamma\right) + o_P\left(1\right) + \sqrt{n}E\left[f_1\left(u_i; \theta_1, \tau, \gamma\right)\right]|_{\theta_1=\hat{\theta}_1(\tau,\gamma)}$$

$$= \mathbb{G}_n f_1\left(u_i; \theta_1^0\left(\tau,\gamma\right), \tau, \gamma\right) + o_P\left(1\right) - \Omega_1\left(\tau,\gamma\right)\sqrt{n}\left\{\hat{\theta}_1\left(\tau,\gamma\right) - \theta_1^0\left(\tau,\gamma\right)\right\}.$$

Thus we have the following uniform Bahadur representation

$$\sqrt{n}\left(\hat{\theta}_1\left(\tau,\gamma\right) - \theta_1^0\left(\tau,\gamma\right)\right) = \bar{\Omega}_1\left(\tau,\gamma\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\tau\left(y_i - \theta_1^0\left(\tau,\gamma\right)' z_i\left(\gamma\right)\right) z_i\left(\gamma\right) + o_P\left(1\right), \tag{B.16}$$

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

(iii) and (iv) Let $w_1\left(\tau\right) \equiv \sqrt{n}\left(\theta_{1\tau} - \theta_{1\tau}^0\right)$, $w_2 \equiv n^{1-2a}\left(\gamma - \gamma^0\right)$, $w_{1n}\left(\tau\right) \equiv \sqrt{n}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0)$ and $w_{2n} \equiv n^{1-2a}\left(\hat{\gamma} - \gamma^0\right)$. Let $\{\theta_{1\tau}\}$ denote $\{\theta_{1\tau}\}_{\tau\in\mathcal{T}}$, and similarly for $\{\hat{\theta}_{1\tau}\}$. As in the proof of Theorem 2.3,

we continue to use the decomposition for $D_{n\tau}(\theta_{1\tau}, \gamma)$ in (B.3), where the only difference is that γ and γ^0 are now τ-invariant. Noting that $\left(\{\hat{\theta}_{1\tau}\}, \hat{\gamma}\right) = \operatorname{argmin}_{\{\theta_{1\tau}\}, \gamma} \int D_{n\tau}(\theta_{1\tau}, \gamma) d\Pi(\tau)$, we have

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

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

$$\mathcal{D}_{1n}(w_1) = -n^{-1/2} \int w_1(\tau)' \sum_{i=1}^{n} z_i(\gamma^0) \psi_\tau(\varepsilon_{i\tau}) d\Pi(\tau) + \frac{1}{2} \int w_1(\tau)' \Omega_{1\tau} w_1(\tau) d\Pi(\tau) + o_P(1),$$
(B.17)

and

$$\bar{\mathcal{D}}_{2n}(w_1, w_2) = \bar{\mathcal{D}}_{2n}(0, w_2) + o_P(1) = \int \sum_{i=1}^{n} \left[\rho_\tau\left(\varepsilon_{i\tau} + \delta_\tau^{0\prime} \Delta x_i(w_2)\right) - \rho_\tau(\varepsilon_{i\tau})\right] d\Pi(\tau) + o_P(1).$$
(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 $w_{1n\tau} \equiv n^{1/2}(\hat{\theta}_{1\tau} - \theta_{1\tau}^0)$ and $w_{2n\tau} \equiv n^{1-2a}\left(\hat{\gamma}_\tau - \gamma_\tau^0\right)$. Let $\tilde{w}_{2n} \equiv n^{1-2a}\left(\tilde{\gamma} - \gamma^0\right)$. Then $LR_n = \int \left[\bar{D}_{2n\tau}(w_{1n\tau}, \tilde{w}_{2n}) - \bar{D}_{2n\tau}(w_{1n\tau}, w_{2n\tau})\right] d\Pi(\tau)$. The result in Lemma B.4 continues to hold when $\sup_{\|w_1\| \le M}$ is replaced by $\sup_{\|w_1\| \le M} \sup_{\tau \in \mathcal{T}}$. It follows that

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

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

$$\Upsilon(\tau, \bar{\tau}; w_2, \bar{w}_2) \equiv E[B(\tau, w_2) B(\bar{\tau}, \bar{w}_2)] = (\tau \wedge \bar{\tau} - \tau\bar{\tau}) \lambda(\tau, \bar{\tau}) b(w_2, \bar{w}_2),$$

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

$$-\int \bar{D}_{2n\tau}(0, \tilde{w}_{2n}) d\Pi(\tau) \xrightarrow{d} \frac{\lambda_0}{2\mu_0} \sup_r \{-|r| + 2W(r)\}.$$
(B.20)

This, in conjunction with (B.13)-(B.14), implies that $LR_n = c_{LR}\Xi + o_P(1)$, where $c_{LR} = \int \frac{\lambda_\tau}{2\mu_\tau} d\Pi(\tau) - \frac{\lambda_0}{2\mu_0}$ and $\Xi = \sup_{-\infty < r < \infty} \{-|r| + 2W(r)\}$. By the Cauchy–Schwarz inequality, one can easily verify that $c_{LR} \ge 0$ and the equality holds if and only if $\lambda_\tau = \mu_\tau^2$ a.s.-Π.

Proof of Theorem 2.7. Recall that $\tilde{w}_{2n} \equiv n^{1-2a}\left(\tilde{\gamma} - \gamma^0\right)$. Following closely the proof of Theorem 2.4, we obtain $\overline{LR}_n(\gamma^0) = -\int \bar{D}_{2n\tau}(0, \tilde{w}_{2n}) d\Pi(\tau) + o_P(1)$. The result then follows by arguments analogous to those used in the proof of Theorem 2.4.

Acknowledgments

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.

References

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In: Engle, R. F., McFadden, D. L., eds. *Handbook of Econometrics*. Vol. IV. New York: North-Holland, pp. 2248–2294.

Andrews, D. W. K., Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62:1383–1414.

Andrews, D. W. K., Ploberger, W. (1995). Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *Annals of Statistics* 23:1609–1629.

Bai, J. (1995). Least absolute deviation estimation of a shift. *Econometric Theory* 11:403–436.

Bolton, G. E., Katok, E., Ockenfels, A. (2004). How effective are online reputation mechanisms? An experimental study. *Management Science* 50:1587–1602.

Cabral, L., Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from eBay. *Journal of Industrial Economics* 58:54–78.

Cai, Y. (2010). Forecasting for quantile self exciting threshold autoregressive time series models. *Biometrika* 97:199–208.

Cai, Y., Stander, J. (2008). Quantile-self exciting threshold time series models. *Journal of Time Series Analysis* 29:187–202.

Caner, M. (2002). A note on least absolute deviation estimation of a threshold model. *Econometric Theory* 18:800–814.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21:520–533.

Chernozhukov, V., Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 127:491–525.

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.

Dellarocas, C., Wood, C. A. (2008). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science* 54:460–476.

Fan J., Yao, Q., Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* 83:189–206.

Galvao, A. F., Kato, K., Montes-Rojas, G., Olmo, J. (2014). Testing linearity against threshold effects: Uniform inference in quantile regression. *Annals of the Institute of Statistical Mathematics* 66:413–439.

Galvao, A. F., Montes-Rojas, G., Olmo, J. (2011). Threshold quantile autoregression models. *Journal of Time Series Analysis* 32:253–267.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64:413–430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* 68:575–605.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and Its Inference* 4:123–127.

Imbens, G. W., Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142:615–635.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis* 100:1816–1829.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics* 26:755–770.

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

Koenker, R., Bassett, W. G. (1978). Regression quantiles. *Econometrica* 46:33–50.

Koenker, R., Machado, W. G. (1999). Goodness of fit and related inference processes for quantile regressions. *Journal of the American Statistical Association* 94:1296–1310.

Koul, H. L., Qian, L., Surgailis, D. (2003). Asymptotics of m-estimators in two-phase linear regression models. *Stochastic Processes and Their Applications* 103:123–154.

Lanne, M., Saikkonen, P. (2002). Threshold autoregressions for strongly autocorrelated time series. *Journal of Business & Economic Statistics* 20:282–289.

Lee, S., Seo, M. H., Shin, Y. (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association* 106:220–231.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142:698–714.

Oka, T. (2010). Nonparametric quantile regression with discontinuities. *Working Paper*. Department of Economics, National University of Singapore.

Pakes, A., Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* 57:1027–1057.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7:186–199.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In: Barnett, W., Powell, J. L., Tauchen, G., eds., *Nonparametric and Semiparametric Methods in Econometrics*. Cambridge: Cambridge University Press, pp. 357–384.

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics* 146:170–184.

Resnick, P., Zeckhauser, R., Swanson, J., Lockwood, K. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics* 9:79–101.

Ruppert, D., Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75:828–838.

Su, L., Xiao, Z. (2008). Testing for parameter stability in quantile regression models. *Statistics & Probability Letters* 78: 2768–2775.

Tong, H. (1978). On a threshold model. In: Chen, C. H., eds., *Pattern Recognition and Signal Processing*, Amsterdam: Sijthoff & Noordhoff, pp. 575–586.

Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*. Lecture Notes in Statistics. New York: Springer-Verlag.

Tong, H. (2011). Threshold models in time series analysis – 30 years on. *Statistics and its Inference* 4:107–118.

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

Yu, P. (2012). Likelihood estimation and inference in threshold regression. *Journal of Econometrics* 167:274–294.

Yu, P. (2014). Integrated quantile threshold regression and distributional threshold effects. *Working Paper*, School of Economics and Finance, University of Hong Kong.