

# Estimating Structural Changes in Regression Quantiles\*

Tatsushi Oka<sup>†</sup>

National University of Singapore

Zhongjun Qu<sup>‡</sup>

Boston University

December 27, 2010

## Abstract

This paper considers the estimation of multiple structural changes occurring at unknown dates in one or multiple conditional quantile functions. The analysis covers time series models as well as models with repeated cross sections. We estimate the break dates and other parameters jointly by minimizing the check function over all permissible break dates. The limiting distribution of the estimator is derived and the coverage property of the resulting confidence interval is assessed via simulations. A procedure to determine the number of breaks is also discussed. Empirical applications to the quarterly US real GDP growth rate and the underage drunk driving data suggest that the method can deliver more informative results than the analysis of the conditional mean function alone.

**JEL Classification Number:** C14, C21, C22.

**Keywords:** breaks, conditional distribution, quantile regression, policy evaluation.

---

\*We thank Roger Koenker, Pierre Perron, Barbara Rossi, Ivan Fernandez-Val and participants at 2010 Econometric Society winter meeting for useful comments, Denis Tkachenko for research assistance and Tzuchun Kuo for suggesting the underage drunk driving data used in Section 8.

<sup>†</sup>Department of Economics, National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570 (oka@nsu.edu.sg)

<sup>‡</sup>Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).

## 1 Introduction

The issue of structural change has been extensively studied in a variety of applications. Recent contributions have substantially broadened the scope of the related literature. For example, Bai and Perron (1998, 2003) provided a unified treatment of estimation, inference and computation in linear multiple regression with unknown breaks. Bai et al. (1998), Bai (2000) and Qu and Perron (2007) extended the analysis to a system of equations. Hansen (1992), Bai et al. (1998) and Kejriwal and Perron (2008) considered regressions with integrated variables. Andrews (1993), Hall and Sen (1999) and Li and Müller (2009) considered nonlinear models estimated by Generalized Method of Moments. Kokoszka and Leipus (1999, 2000) and Berkes et al. (2004) studied parameter change in GARCH processes. One may refer to Csörgő and Horváth (1998) and Perron (2006) for a comprehensive review of the literature.

A main focus in the literature has been the conditional mean function, while, under many circumstances, structural change in the conditional quantile function is of key importance. For example, when studying income inequality, it is important to examine whether (and how) the wage differential between different racial groups, conditional on observable characteristics, has changed over time. An increase in inequality may increase the conditional dispersion of the differential, while leaving the mean unchanged. Thus, the conditional mean ceases to be informative and the conditional quantiles should be considered. As another example, consider a policy reform that aims at helping students with low test performance. In this case, attention should clearly be focused on the lower quantiles of the conditional distribution in order to understand the effect of such a policy. In both examples, it can be desirable to allow the break dates to be unknown and estimate them from the data. The reason is that in the former case it is often difficult to identify the source of the change *a priori*, while in the latter the policy effect may occur with an unknown time lag due to various reasons. To address such issues, Qu (2008) and Su and Xiao (2008) considered Wald and subgradient-based tests for structural change in regression quantiles, allowing for unknown break dates. However, they did not consider the issue of estimation and inference regarding break dates and other coefficients. This is the subject of the current paper.

Specifically, we consider the estimation of multiple structural changes occurring at unknown dates in conditional quantile functions. The basic framework is that of Koenker and Bassett (1978), with the conditional quantile function being linear in parameters in each individual regime. The analysis covers two types of models. The first is a time series model (e.g., the quantile autoregressive (QAR) model of Koenker and Xiao, 2006), which can be useful for studying structural change in

a macroeconomic variable. The second model considers repeated cross sections, which can be useful for the analysis of effects of social programs, laws and economic policy. For each model, we consider both structural change in a single quantile and in multiple quantiles. The joint analysis of multiple quantiles requires imposing stronger restrictions on the model, but is important, because it can potentially increase the efficiency of the break estimator and, more importantly, reveal the heterogeneity in the change, thus delivering a richer set of information.

We first assume that the number of breaks is known and construct estimators for unknown break dates and other coefficients. The resulting estimator is the global minimizer of the check function over all permissible break dates. When multiple quantiles are considered, the check function is integrated over a set of quantiles of interest. The underlying assumptions are mild, allowing for dynamic models. Also, they restrict only a neighborhood surrounding quantiles of interest. Other quantiles are left unspecified, thus being allowed to change or remain stable. The latter feature allows us to look at slices of the conditional distribution without relying on global distributional assumptions. Under these assumptions, we derive asymptotic distributions of the estimator following the methodology of Picard (1985) and Yao (1987). The distribution of the break date estimates depend on a two-sided Brownian motion, which is often encountered in the literature and the analytical properties of which have been studied by Bai (1997). It involves parameters that can be consistently estimated, thus the confidence interval can be constructed without relying on simulation.

We then discuss a testing procedure that allows to determine the number of breaks. It builds upon the subgradient-based tests proposed in Qu (2008). These tests do not require estimation of the variance (more precisely, the sparsity) parameter, thus having monotonic power even when multiple changes are present. This feature makes them suitable for our purposes.

We consider two empirical applications, one for each type of models studied in the paper. The first application is to the quarterly US GDP growth, whose volatility has been documented to exhibit a substantial decline since the early to mid-1980s (the so-called "Great Moderation"), see for example McConnell and Perez-Quiros (2000). The paper revisits this issue using a quantile regression framework. The result suggests that the moderation mainly affected the upper tail of the conditional distribution, with the conditional median and the lower quantiles remaining stable. Hence, it suggests that a major change in the GDP growth should be attributed to the fact that the growth was less rapid during expansions, while the recessions have remained just as severe when they occurred. The second application considers structural changes in young drivers' blood alcohol levels using a data set for the state of California over the period 1983-2007. Two structural changes

are detected, which closely coincide with the National Minimum Drinking Ages Act of 1984 and a beer tax hike in 1991. Interestingly, the changes are smaller in higher quantiles, suggesting that the policies are more effective for "light drinkers" than for "heavy drinkers" in the population.

The technical development in the paper relies heavily on Bai (1995, 1998). Bai (1995) developed asymptotic theory for least absolute deviation estimation of a shift in linear regressions, while Bai (1998) extended the analysis to allow for multiple changes. Recently, Chen (2008) further extended Bai's work to study structural changes in a single conditional quantile function. There, the regressors are assumed to be strictly exogenous. This paper is different from their studies in three important aspects. First, the assumptions allow for dynamic models. Therefore, the result has wider applicability. Secondly, we consider models with repeated cross sections, which is important for policy related applications. Finally, we consider structural change in multiple quantiles and a testing procedure for determining the number of breaks. From a methodological perspective, this paper is related to the literature of functional coefficient quantile regression models, see Cai and Xu (2008) and Kim (2007). Their model is suitable for modelling smooth changes and ours for sudden shifts. Finally, the paper is related to robust estimation of a change-point: see Hušková (1997), Fiteni (2002) and the references therein.

The paper is organized as follows. Section 2 discusses the setup and three simple examples to motivate the study. Section 3 discusses multiple structural changes in a pre-specified quantile in a time series model, together with the method of estimation and the limiting distributions of the estimates. Section 4 considers structural change in multiple quantiles. Section 5 considers models with repeated cross sections. In Sections 2 to 5, the number of breaks is assumed to be known. The issue of estimating the number of breaks is addressed in Section 6. Section 7 contains simulation results and Section 8 presents the empirical applications. All proofs are included in the appendix.

The following notation is used. The superscript 0 indicates the true value of a parameter. For a real valued vector  $z$ ,  $\|z\|$  denotes its Euclidean norm.  $[z]$  is the integer part of  $z$ .  $1(\cdot)$  is the indicator function.  $D_{[0,1]}$  stands for the set of functions on  $[0, 1]$  that are right continuous and have left limits, equipped with the Skorohod metric. The symbols " $\Rightarrow$ ", " $\rightarrow^p$ " and " $\rightarrow^{a.s.}$ " denote weak convergence under Skorohod topology, convergence in probability and convergence almost surely, and  $O_p(\cdot)$  and  $o_p(\cdot)$  is the usual notation for the orders of stochastic convergence.

## 2 Setup and examples

Let  $y_t$  be a real-valued random variable,  $x_t$  a  $p$  by 1 random vector, and  $Q_{y_t}(\tau|x_t)$  the conditional quantile function of  $y_t$  given  $x_t$ , where  $t$  corresponds to the time index or an ordering according

to some other criterion. Note that if  $t$  is the index for time, then  $Q_{y_t}(\tau|x_t)$  is interpreted as the quantile function of  $y_t$  conditional on the  $\sigma$ -algebra generated by  $(x_t, y_{t-1}, x_{t-1}, y_{t-2}, \dots)$ . Let  $T$  denote the sample size.

We assume the conditional quantile function is linear in parameters and is affected by  $m$  structural changes:

$$Q_{y_t}(\tau|x_t) = \begin{cases} x'_t \beta_1^0(\tau), & t = 1, \dots, T_1^0, \\ x'_t \beta_2^0(\tau), & t = T_1^0 + 1, \dots, T_2^0, \\ \vdots & \vdots \\ x'_t \beta_{m+1}^0(\tau), & t = T_m^0 + 1, \dots, T, \end{cases} \quad (1)$$

where  $\tau \in (0, 1)$  denotes a quantile of interest,  $\beta_j^0(\tau)$  ( $j = 1, \dots, m+1$ ) are the unknown parameters that are quantile dependent, and  $T_j^0$  ( $j = 1, \dots, m$ ) are the unknown break dates. A subset of  $\beta_j^0(\tau)$  may be restricted to be constant over  $t$  to allow for partial structural changes. The regressors  $x_t$  can include discrete as well as continuous variables. We now give three examples to illustrate our framework.

**Example 1** *Cox et al. (1985)* considered the following model for the short-term riskless interest rate:

$$dr_t = (\alpha + \beta r_t) dt + \sigma r_t^{1/2} dW_t, \quad (2)$$

where  $r_t$  is the riskless rate and  $W_t$  is the Wiener process. The process (2) can be approximated by the following discrete-time model if the sampling intervals are small (see Chan, et.al. 1992, p.1213 and the references therein for discussions on the issue of discretization):

$$r_{t+1} - r_t = \alpha + \beta r_t + (\sigma r_t^{1/2}) u_{t+1}$$

with  $u_{t+1} \sim i.i.d.N(0, 1)$ , implying that the quantiles of  $r_{t+1}$  given  $r_t$  are linear in parameters and satisfy

$$Q_{r_{t+1}}(\tau|r_t) = \alpha + (1 + \beta)r_t + (\sigma r_t^{1/2}) F_u^{-1}(\tau),$$

where  $F_u^{-1}(\tau)$  is the  $\tau$ th quantile of a standard normal random variable. The procedure developed in this paper can be used to estimate structural changes in some or all of the parameters ( $\alpha, \beta$  and  $\sigma$ ). More complicated models than (2) can be analyzed in the same way provided that, upon discretization, the conditional quantile function can be well approximated by a linear function.

**Example 2** *Chernozhukov and Umantsev (2001)* studied the following model for Value-at-Risk:<sup>1</sup>

$$Q_{y_t}(\tau|x_t) = x'_t \beta(\tau), \quad (3)$$

---

<sup>1</sup>See Taylor (1999) and Engle and Manganelli (2004) for related works.

where  $y_t$  is return on some asset,  $x_t$  is a vector of information variables affecting the distribution of  $y_t$ . For example,  $x_t$  may include returns on other securities, lagged values of  $y_t$ , and proxies to volatility (such as exponentially weighted squared-returns). They documented that such variables affect various quantiles of  $y_t$  in a very differential and nontrivial manner. Here, an interesting open issue is whether the risk relationship (3) undergoes substantial structural changes. Our method can be applied to address this issue without having to specify the dates of the changes a priori.

**Example 3** Pichl et al. (2003) applied the structural change methodology to evaluate the effect of the Boston Gun Project on youth homicide incidents. They allowed the break date to be unknown in order to capture an unknown time lag in policy implementation. Their focus was on structural change in the conditional mean. However, in many cases, the aim of a policy is to induce a distribution change rather than a pure location shift. For example, consider a public school reform aimed at improving the performance of students with low test scores. In this case, the lower quantiles of the conditional distribution are the targets of the reform. Another example is a public policy to reduce income inequality. In this case, the target is the dispersion of the distribution. In these cases, the standard structural change in mean methodology ceases to be relevant, while the methods developed in this paper can prove useful.

For now, assume that the number of structural changes is known. Later in the paper (Section 6), we will discuss a testing procedure that can be used to estimate the number of breaks. We first consider structural changes in a single conditional quantile function in a time series model.

### 3 Structural changes in a given quantile $\tau \in (0, 1)$

In the absence of structural change, the model (1) can be estimated by solving

$$\min_{b \in \mathbb{R}^p} \sum_{t=1}^T \rho_\tau(y_t - x_t' b), \quad (4)$$

where  $\rho_\tau(u)$  is the check function given by  $\rho_\tau(u) = u(\tau - 1(u < 0))$ ; see Koenker (2005) for a comprehensive treatment of related issues. Now suppose that the  $\tau$ th quantile is affected by  $m$  structural changes, occurring at unknown dates  $(T_1^0, \dots, T_m^0)$ . Then, define the following function for a set of candidate break dates  $T^b = (T_1, \dots, T_m)$ :

$$S_T(\tau, \beta(\tau), T^b) = \sum_{j=0}^m \sum_{t=T_j+1}^{T_{j+1}} \rho_\tau(y_t - x_t' \beta_{j+1}(\tau)), \quad (5)$$

where  $\beta(\tau) = (\beta_1(\tau)', \dots, \beta_{m+1}(\tau)')'$ ,  $T_0 = 0$  and  $T_{m+1} = T$ .<sup>2</sup> Motivated by Bai (1995, 1998), we estimate the break dates and coefficients  $\beta(\tau)$  jointly by solving

$$(\hat{\beta}(\tau), \hat{T}^b) = \arg \min_{\beta(\tau), T^b \in \Lambda_\varepsilon} S_T(\tau, \beta(\tau), T^b), \quad (6)$$

where  $\hat{\beta}(\tau) = (\hat{\beta}_1(\tau)', \dots, \hat{\beta}_{m+1}(\tau)')'$  and  $\hat{T}^b = (\hat{T}_1, \dots, \hat{T}_m)$ . Specifically, for a given partition of the sample, we estimate the coefficients  $\beta(\tau)$  by minimizing  $S_T(\tau, \beta(\tau), T^b)$ . Then, we search over all permissible partitions to find the break dates that achieve the global minimum. These break dates, along with the corresponding estimates for  $\beta(\tau)$ , are taken as final estimates. Note that in (6),  $\Lambda_\varepsilon$  denotes the set of possible partitions. It ensures that each estimated regime is a positive fraction of the sample. For example, it can be specified as

$$\Lambda_\varepsilon = \{(T_1, \dots, T_m) : T_j - T_{j-1} \geq \varepsilon T \ (j = 2, \dots, m), T_1 \geq \varepsilon T, T_m \leq (1 - \varepsilon)T\}, \quad (7)$$

where  $\varepsilon$  is a positive small number. The precise assumptions on  $\Lambda_\varepsilon$  will be stated later in the paper.

Let  $f_t(\cdot)$ ,  $F_t(\cdot)$  and  $F_t^{-1}(\cdot)$  denote the conditional density, conditional distribution and conditional quantile function of  $y_t$  given  $x_t$ . Let  $\mathcal{F}_{t-1}$  be the  $\sigma$ -algebra generated by  $(x_t, y_{t-1}, x_{t-1}, y_{t-2}, \dots)$  and  $u_t^0(\tau)$  be the difference between  $y_t$  and its  $\tau$ th conditional quantile, i.e.,

$$u_t^0(\tau) = y_t - x_t' \beta_j^0(\tau) \quad \text{for } T_{j-1}^0 + 1 \leq t \leq T_j^0 \ (j = 1, \dots, m+1).$$

We now state the assumptions needed for the derivation of asymptotic properties of the estimates.

**Assumption 1.**  $\{1(u_t^0(\tau) < 0) - \tau\}$  is a martingale difference sequence with respect to  $\mathcal{F}_{t-1}$ .

**Assumption 2.** The distribution functions  $\{F_t(\cdot)\}$  are absolutely continuous, with continuous densities  $\{f_t(\cdot)\}$  satisfying  $0 < L_f \leq f_t(F_t^{-1}(\tau)) \leq U_f < \infty$  for all  $t$ .

**Assumption 3.** For any  $\epsilon > 0$  there exists a  $\sigma(\epsilon) > 0$  such that  $|f_t(F_t^{-1}(\tau) + s) - f_t(F_t^{-1}(\tau))| < \epsilon$  for all  $|s| < \sigma(\epsilon)$  and all  $1 \leq t \leq T$ .

Assumptions 1 and 2 are familiar in the quantile regression literature; heteroskedasticity is allowed. Assumption 2 requires the densities to be uniformly bounded away from 0 and  $\infty$  at the quantile of interest. This assumption is local and therefore does not require the full density function to be bounded. Assumption 3 implies that the conditional densities are uniformly continuous in some neighborhood of the  $\tau$ th quantile. This Assumption, along with Assumption 2, implies that

---

<sup>2</sup>A more complete notation for  $\beta(\tau)$  should be  $\beta(\tau, T^b)$ . We have omitted  $T^b$  in order not to overburden the notation.

$f_t(u)$  is uniformly bounded away from 0 and  $\infty$  for all  $t$  and all  $u$  in some open neighborhood of  $F_t^{-1}(\tau)$ . Note that these two Assumptions entail stronger restrictions in dynamic models than in ordinary regression applications since the support of the explanatory variables is determined within the model. However, they are often not difficult to verify given a particular specification.

To illustrate these three assumptions, we can consider the following location-scale model (with no structural change):

$$y_t = z_t'\beta + (w_t'\eta)u_t, \quad (8)$$

where  $\{u_t\}$  is a sequence of *i.i.d.* errors independent of  $z_t$  and  $w_t$  ( $u_t$  can be correlated with  $z_{t+k}$  and  $w_{t+k}$  for  $k > 0$ , thus allowing for dynamic models). Then, Assumption 1 is satisfied due to the independence. Let  $f_u(\cdot)$  and  $F_u^{-1}(\tau)$  denote the density and the  $\tau$ th quantile of the errors  $u_t$ . Then,  $f_t(F_t^{-1}(\tau)) = f_u(F_u^{-1}(\tau))/(w_t'\eta)$ . Thus, Assumption 2 is satisfied if  $F_u(\cdot)$  is absolutely continuous with continuous density  $f_u(\cdot)$  satisfying  $\delta_u < f_u(F_u^{-1}(\tau)) < \infty$  and  $\delta_w < w_t'\eta < \infty$  for all  $t$  for some arbitrary strictly positive constants  $\delta_u$  and  $\delta_w$ . Note that  $f_u(F_u^{-1}(\cdot))$  can be unbounded at quantile different from  $\tau$ . Assumption 3 is satisfied if, in addition, the density  $f_u(\cdot)$  is continuous over an open interval containing the  $\tau$ th quantile.

**Assumption 4.**  $T_j^0 = [\lambda_j^0 T]$  ( $j = 1, \dots, m$ ) with  $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$ .

**Assumption 5.** (a) an intercept is included in  $x_t$ ; (b) for each  $j = 1, \dots, m + 1$ ,

$$\frac{1}{T} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[sT]} f_t(F_t^{-1}(\tau))x_t x_t' \rightarrow^p sH_j^0(\tau) \quad \text{and} \quad \frac{1}{T} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[sT]} x_t x_t' \rightarrow^p sJ_j^0 \quad (9)$$

hold uniformly in  $0 \leq s \leq \lambda_j^0 - \lambda_{j-1}^0$  as  $T \rightarrow \infty$ , where  $J_j^0$  and  $H_j^0(\tau)$  are non-random positive definite matrices; (c)  $E\|x_t\|^{4+\varphi} < L$  holds with some  $\varphi > 0$  and  $L < \infty$  for all  $t = 1, \dots, T$ ; (d) there exist  $M < \infty$  and  $\gamma > 2$  such that  $T^{-1} \sum_{t=1}^T E\|x_t\|^{2\gamma+1} < M$  and  $E(T^{-1} \sum_{t=1}^T \|x_t\|^3)^\gamma < M$  hold when  $T$  is large; (e) there exists  $j_0 > 0$ , such that the eigenvalues of  $j^{-1} \sum_{t=l}^{l+j} x_t x_t'$  are bounded from above and below by  $\lambda_{\max}$  and  $\lambda_{\min}$  for all  $j \geq j_0$  and  $1 \leq l \leq T - j$  with  $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$ .

Assumption 4 states that each regime occupies a non-vanishing proportion of the sample. Assumption 5 imposes some structure on the regressors  $x_t$ . The first part in (9) imposes some restriction on possible heteroskedasticity. Assumption 5(c) ensures weak convergence of the process  $T^{-1/2} \sum_{t=1}^{[sT]} x_t(1(F_t(y_t) \leq \tau) - \tau)$ . Assumption 5(d) is needed for stochastic equicontinuity of sequential empirical processes based on the estimated quantile regression residuals. It also implies  $\max_{1 \leq t \leq T} \|x_t\| = o_p(T^{1/2})$ , which is familiar in the literature on M-estimators. Assumption 5 rules



out trending regressors under which the rates of convergence can be verified, although the limiting distribution of the break estimates will be different. This situation requires a separate treatment.

**Assumption 6.** Let  $\Delta_{T,j}(\tau) = \beta_{j+1}^0(\tau) - \beta_j^0(\tau)$  ( $j = 1, \dots, m$ ). Assume  $\Delta_{T,j}(\tau) = v_T \Delta_j(\tau)$  for some  $\|\Delta_j(\tau)\| > 0$ , where  $\Delta_j(\tau)$  is a vector independent of  $T$ ,  $v_T > 0$  is a scalar satisfying  $v_T \rightarrow 0$  and  $T^{(1/2)-\vartheta} v_T \rightarrow \infty$  for some  $\vartheta \in (0, 1/2)$ .

Assumption 6 follows Picard (1985) and Yao (1987). The magnitudes of the shifts converge to zero as the sample size increases. Consequently, the breaks will be estimated at a slower convergence rate than under "fixed-break" asymptotics. A functional central limit theorem then applies, which permits us to obtain a limiting distribution invariant to the exact distribution of  $x_t$  and  $y_t$ . The setup is well suited to provide an adequate approximation to the exact distribution when the change is moderate but the resulting confidence interval can be liberal when the change is small. The quality of the resulting approximation will be subsequently evaluated using simulations.

To summarize, the assumptions have two important features. First, they allow for dynamic models, for example, the quantile autoregression (QAR) model of Koenker and Xiao (2006):  $y_t = \rho_0(U_t) + \rho_1(U_t)y_{t-1} + \dots + \rho_q(U_t)y_{t-q}$ , where  $\{U_t\}$  is a sequence of *i.i.d.* standard uniform random variables. Second, the assumptions are local, in the sense that they impose restrictions only on the  $\tau$ th quantile and a small neighborhood surrounding it. This allows us to look at slices of the distribution without making global distributional assumptions.

The following result establishes the convergence rates of the parameter estimates.

**Lemma 1** *Under Assumptions 1-6, we have  $v_T^2(\hat{T}_j - T_j^0) = O_p(1)$  for  $j = 1, \dots, m$  and  $\sqrt{T}(\hat{\beta}_j(\tau) - \beta_j^0(\tau)) = O_p(1)$  for  $j = 1, \dots, m + 1$ .*

The next result presents the limiting distributions of the estimates.

**Theorem 1** *Let Assumptions 1-6 hold. Then, for  $j = 1, \dots, m$ ,*

$$\left(\frac{\pi_j}{\sigma_j}\right)^2 v_T^2(\hat{T}_j - T_j^0) \rightarrow^d \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\sigma_{j+1}/\sigma_j)W(s) - (\pi_{j+1}/\pi_j)|s|/2 & s > 0 \end{cases},$$

where  $\pi_j = \Delta_j(\tau)' H_j^0(\tau) \Delta_j(\tau)$ ,  $\pi_{j+1} = \Delta_j(\tau)' H_{j+1}^0(\tau) \Delta_j(\tau)$ ,  $\sigma_j^2 = \tau(1-\tau) \Delta_j(\tau)' J_j^0 \Delta_j(\tau)$ ,  $\sigma_{j+1}^2 = \tau(1-\tau) \Delta_j(\tau)' J_{j+1}^0 \Delta_j(\tau)$  and  $W(s)$  is the standard two-sided Brownian motion. Also,

$$\sqrt{T}(\hat{\beta}_j(\tau) - \beta_j^0(\tau)) \rightarrow^d N(0, V_j)$$

with  $V_j = \tau(1-\tau)\Omega_j^0(\tau)/(\lambda_j^0 - \lambda_{j-1}^0)^2$  and  $\Omega_j^0(\tau) = (H_j^0(\tau))^{-1} J_j^0 (H_j^0(\tau))^{-1}$  for  $j = 1, \dots, m + 1$ .

The limiting distribution has the same structure as that of Bai (1995). An analytical expression for its cumulative distribution function is provided in Bai (1997). To construct confidence intervals, we need to replace  $H_j^0(\tau)$  and  $J_j^0$  by consistent estimates. These can be obtained by conditioning on the estimated break date  $\hat{T}_j$ . For example,

$$\hat{H}_1(\tau) = \hat{T}_1^{-1} \sum_{t=1}^{\hat{T}_1} \hat{f}_t(F_t^{-1}(\tau)) x_t x_t', \quad \hat{J}_1 = \hat{T}_1^{-1} \sum_{t=1}^{\hat{T}_1} x_t x_t'.$$

The estimates for the densities,  $\hat{f}_t(F_t^{-1}(\tau))$ , can be obtained using the difference quotient, as considered by Siddiqui (1960) and Hendricks and Koenker (1992). A detailed discussion can be found in Qu (2008, pp. 176-177).

#### 4 Structural changes in multiple quantiles

Structural change can be heterogeneous in the sense that different quantiles can change by different magnitudes. In such a context, it can be more informative to consider a range of quantiles as opposed to a single one.

Suppose that quantiles in the interval  $\mathcal{T}_\omega = [\omega_1, \omega_2]$  with  $0 < \omega_1 < \omega_2 < 1$  are affected by structural change. Then, a natural approach is to consider a partition of this interval and examine a set of quantiles denoted by  $\tau_h$ ,  $h = 1, \dots, q$ . After such a set of quantiles is specified, the estimation can be carried out in a similar way as in Section 3. Specifically, we define the following objective function for a set of candidate break dates  $T^b = (T_1, \dots, T_m)$  and parameter values  $\beta(\mathcal{T}_\omega) = (\beta(\tau_1)', \dots, \beta(\tau_q)')'$ :

$$S_T(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b) = \sum_{h=1}^q \sum_{j=0}^m \sum_{t=T_j+1}^{T_{j+1}} \rho_{\tau_h}(y_t - x_t' \beta_{j+1}(\tau_h)),$$

and solve

$$(\hat{\beta}(\mathcal{T}_\omega), \hat{T}^b) = \arg \min_{\beta(\mathcal{T}_\omega), T^b \in \Lambda_\varepsilon} S_T(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b), \quad (10)$$

where  $\Lambda_\varepsilon$  has the same definition as in (7). The minimization problem (10) consists of three steps. First, we minimize  $S_T(\tau_h, \beta(\tau_h), T^b)$  for a given partition of the sample and a given quantile  $\tau_h$ . Then, still conditioning on the same partition, we repeat the minimization for all quantiles  $\{\tau_h : h = 1, \dots, q\}$  to obtain  $S_T(\mathcal{T}_\omega, \hat{\beta}(\mathcal{T}_\omega), T^b)$ . Finally, we search over all possible partitions  $T^b \in \Lambda_\varepsilon$  to find the break dates that achieve the global minimum of  $S_T(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b)$ .

In practice, we need to choose  $\mathcal{T}_\omega$  and  $\tau_h$  ( $h = 1, \dots, q$ ). We view this as an empirical issue. The interval  $\mathcal{T}_\omega$  is often easy to determine given the question of interest. The choice of  $\tau_h$  is more delicate

and will require some judgement. Evidence from the empirical literature suggests that considering a coarse grid of quantiles often suffices to deliver desired information. The spacing can be between 5% and 15% depending on the question of interest. For example, Chamberlain (1994) studied a variety of issues including the changes in returns to education from 1979 to 1987. He considered quantile regression models of log weekly earnings for five quantiles  $\tau = 0.10, 0.25, 0.50, 0.75$  and  $0.90$ . There, the quantiles  $0.10$  and  $0.90$  were used to examine the tails of the distribution, the quantile  $0.50$  was used to measure the central tendency while  $0.25$  and  $0.75$  captured the intermediate cases. He showed that the returns to education are different across quantiles. They also appear to be different for 1979 and 1987. This choice of quantiles has now become a useful rule of thumb in the empirical labor literature. For example, Angrist, Chernozhukov and Fernández-Val (2006) considered the same quantiles when analyzing change in US wage inequality; see also Buchinsky (1994). Similar choices are also often made in other applied micro areas. For example, Eide and Showalter (1998) and Levin (2001) considered the effect of school quality and class size on student performance and Poterba and Rueben (1995) considered public-private wage differentials in the United States. Intuitively, a coarse grid often suffices because adjacent quantiles typically exhibit similar properties. Therefore, the incremental information gained from considering a finer grid is typically small. Once the grid is chosen, we suggest carrying out estimation using both multiple quantiles and individual quantiles and providing a full disclosure of the results.

In practice, there will be some arbitrariness associated with a particular grid choice. Therefore it is useful to experiment with different grids to examine result sensitivity. For each grid choice, the proposed method can be used to obtain point estimates and confidence intervals, which can then be compared to examine result sensitivity. This point will be illustrated using two empirical applications in Section 8.

We impose the following assumption on the  $q$  quantiles entering the estimation.

**Assumption 7.** The conditional quantile functions at  $\tau_h$  ( $h = 1, \dots, q$ ) satisfy Assumptions 1-5. There exists at least one quantile  $\tau_j$  ( $1 \leq j \leq q$ ) satisfying Assumption 6 (other quantiles can remain stable or satisfy Assumption 6). Also,  $q$  is fixed as  $T \rightarrow \infty$ .

The next Corollary gives the limiting distribution of estimated break dates. The distributions for  $\hat{\beta}(\tau_h)$  ( $h = 1, \dots, q$ ) are the same as in Theorem 1, and thus are not repeated here.

**Corollary 1** Under Assumption 7, for  $j = 1, \dots, m$ ,

$$\left(\frac{\pi_j^*}{\sigma_j^*}\right)^2 v_T^2(\hat{T}_j - T_j^0) \rightarrow^d \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\sigma_{j+1}^*/\sigma_j^*)W(s) - (\pi_{j+1}^*/\pi_j^*)|s|/2 & s > 0 \end{cases},$$

where  $W(s)$  is the standard two-sided Brownian motion,  $\pi_j^* = (1/q) \sum_{h=1}^q \Delta_j(\tau_h)' H_j^0(\tau_h) \Delta_j(\tau_h)$ ,  $\pi_{j+1}^* = (1/q) \sum_{h=1}^q \Delta_j(\tau_h)' H_{j+1}^0(\tau_h) \Delta_j(\tau_h)$ ,  $\sigma_j^{*2} = (1/q^2) \sum_{h=1}^q \sum_{g=1}^q (\tau_h \wedge \tau_g - \tau_h \tau_g) \Delta_j(\tau_h)' J_j^0 \Delta_j(\tau_g)$  and  $\sigma_{j+1}^{*2} = (1/q^2) \sum_{h=1}^q \sum_{g=1}^q (\tau_h \wedge \tau_g - \tau_h \tau_g) \Delta_j(\tau_h)' J_{j+1}^0 \Delta_j(\tau_g)$ .

## 5 Models with repeated cross sections

Suppose the data set contains observations  $(x'_{it}, y_{it})$ , where  $i$  is the index for individual and  $t$  for time. Assume  $i = 1, \dots, N$  and  $t = 1, \dots, T$ .

First, consider structural changes in a single conditional quantile function. Suppose the data generating process is

$$Q_{y_{it}}(\tau|x_{it}) = x'_{it}\beta_j^0(\tau) \quad \text{for } t = T_{j-1}^0 + 1, \dots, T_j^0, \quad (11)$$

where the break dates  $T_j^0$  ( $j = 1, \dots, m$ ) are common to all individuals. The estimation procedure is similar to that in Section 3. Specifically, for a set of candidate break dates  $T^b = (T_1, \dots, T_m)$ , define the following function

$$S_{NT}(\tau, \beta(\tau), T^b) = \sum_{j=0}^m \sum_{t=T_{j+1}}^{T_{j+1}} \sum_{i=1}^N \rho_\tau(y_{it} - x'_{it}\beta_{j+1}(\tau)),$$

where an additional summation " $\sum_{i=1}^N$ " is present to incorporate the cross-sectional observations.

Then, solve the following minimization problem to obtain the estimates:

$$(\hat{\beta}(\tau), \hat{T}^b) = \arg \min_{\beta(\tau), T^b \in \Lambda_\varepsilon} S_{NT}(\tau, \beta(\tau), T^b).$$

Let  $\mathcal{F}_{t-1}$  denote the  $\sigma$ -algebra generated by  $\{x_{it}, y_{i,t-1}, x_{i,t-1}, y_{i,t-2}, \dots\}_{i=1}^N$ . Let  $u_{it}^0(\tau)$  denote the difference between  $y_{it}$  and its  $\tau$ th conditional quantile, i.e.,

$$u_{it}^0(\tau) = y_{it} - x'_{it}\beta_j^0(\tau) \quad \text{for } T_{j-1}^0 + 1 \leq t \leq T_j^0 \quad (j = 1, \dots, m+1).$$

We make the following assumptions, which closely parallel the ones in Section 3.

**Assumption B1.** For a given  $i$ ,  $\{1(u_{it}^0(\tau) < 0) - \tau\}$  is a martingale difference sequence with respect to  $\mathcal{F}_{t-1}$ . Also,  $u_{it}^0(\tau)$  and  $u_{jt}^0(\tau)$  are independent conditional on  $\mathcal{F}_{t-1}$  for all  $i \neq j$ .

**Assumption B2.** Assumption 2 holds for  $\{F_{it}(\cdot)\}$  and  $\{f_{it}(\cdot)\}$ .

**Assumption B3.** For any  $\epsilon > 0$  there exists a  $\sigma(\epsilon) > 0$  such that  $|f_{it}(F_{it}^{-1}(\tau) + s) - f_{it}(F_{it}^{-1}(\tau))| < \epsilon$  for all  $|s| < \sigma(\epsilon)$  and all  $1 \leq i \leq N$  and  $1 \leq t \leq T$ .

**Assumption B4.** The break dates are common for all  $i$ ;  $T_j^0 = [\lambda_j^0 T]$  with  $0 < \lambda_1^0 < \dots < \lambda_m^0 < 1$ .

**Assumption B5.** (a) an intercept is included in  $x_{it}$ ; (b) for every  $j = 1, \dots, m + 1$ ,

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[sT]} f_{it}(F_{it}^{-1}(\tau)) x_{it} x'_{it} \rightarrow^p s \bar{H}_j^0(\tau) \quad \text{and} \quad \frac{1}{NT} \sum_{i=1}^N \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[sT]} x_{it} x'_{it} \rightarrow^p s \bar{J}_j^0$$

uniformly in  $N$  and  $0 \leq s \leq \lambda_j^0 - \lambda_{j-1}^0$ , where  $\bar{H}_j^0(\tau)$  and  $\bar{J}_j^0$  are non-random positive definite matrices; (c)  $E\|x_{it}\|^{4+\varphi} < L$  with some  $\varphi > 0$  and  $L < \infty$  for all  $i$  and  $t$ ; (d) there exist  $M < \infty$  and  $\gamma > 2$  such that  $(NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N E\|x_{it}\|^{2\gamma+1} < M$  and  $E((NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N \|x_{it}\|^3)^\gamma < M$  hold for all  $N$  and sufficiently large  $T$ ; (e) there exists  $j_0 > 0$  such that the eigenvalues of  $(jN)^{-1} \sum_{t=l}^{l+j} \sum_{i=1}^N x_{it} x'_{it}$  are bounded from above and below by  $\lambda_{\max}$  and  $\lambda_{\min}$  for all  $N$ , all  $j \geq j_0$  and  $1 \leq l \leq T - j$ ;  $0 < \lambda_{\min} \leq \lambda_{\max} < \infty$ .

Assumption B1 allows for dynamic models but rules out cross-sectional dependence in  $u_{it}^0(\tau)$ . B2 and B3 strengthen Assumptions 2 and 3, requiring them to hold for all  $i$  and  $t$ . B4 assumes that the individuals are affected by common breaks. B5 is essentially a re-statement of Assumption 5.

**Assumption B6.** Assume  $N$  and  $T$  satisfy one of the following two conditions: (1)  $N$  is fixed as  $T \rightarrow \infty$ , or (2)  $(N, T) \rightarrow \infty$  but  $\log N/T^{\vartheta/2} \rightarrow 0$ , where  $\vartheta$  is defined in Assumption 6. If the second condition holds, then assume the process  $\{x_{it}(1[u_{it}^0(\tau) < 0] - \tau)\}$  forms a stationary ergodic random field indexed by  $i$  and  $t$ .

Assumption B6 encompasses both small  $N$  and large  $N$  asymptotics. The condition  $\log N/T^{\vartheta/2} \rightarrow 0$  allows for both  $N/T \rightarrow 0$  and  $N/T \rightarrow \infty$ . The restriction is therefore quite mild. The second part of this assumption is used to ensure that the process  $\{x_{it}(1[u_{it}^0(\tau) < 0] - \tau)\}$  satisfies a functional central limit theorem as  $(N, T) \rightarrow \infty$ . It can be replaced by other conditions that serve the same purpose.

**Assumption B7.** Let  $\Delta_{NT,j}(\tau) = \beta_{j+1}^0(\tau) - \beta_j^0(\tau)$ . Assume  $\Delta_{NT,j}(\tau) = N^{-1/2} v_T \Delta_j(\tau)$  for some  $\|\Delta_j(\tau)\| > 0$  independent of  $T$  and  $N$ , where  $v_T > 0$  is a scalar satisfying  $v_T \rightarrow 0$  and  $T^{(1/2)-\vartheta} v_T \rightarrow \infty$  with  $\vartheta$  defined in Assumption 6.

Assumption B7 implies that, with the added cross-sectional dimension, the model can now handle break sizes that are of order equal to  $N^{-1/2}v_T$ , which is smaller than in the pure time series case given by  $O(v_T)$ . This assumption ensures that the estimated break dates will converge at the rate  $v_T^{-2}$ . If the breaks were of higher order than  $N^{-1/2}v_T$ , then the estimated breaks would converge faster than  $v_T^{-2}$ . In those cases, the confidence interval reported below will tend to be conservative. Thus, this assumption, as in the pure time series case, can be viewed as a strategy to deliver a confidence interval that has good coverage when the break size is moderate while being conservative when the break size is large.

The next two results present the rates of convergence and limiting distributions of the estimates.

**Lemma 2** *Under Assumptions B1-B7, we have  $v_T^2(\hat{T}_j - T_j^0) = O_p(1)$  for  $j = 1, \dots, m$  and  $\sqrt{NT}(\hat{\beta}_j(\tau) - \beta_j^0(\tau)) = O_p(1)$  for  $j = 1, \dots, m + 1$ .*

**Theorem 2** *Let Assumptions B1-B7 hold. Then, for  $j = 1, \dots, m$ ,*

$$\left(\frac{\bar{\pi}_j}{\bar{\sigma}_j}\right)^2 v_T^2(\hat{T}_j - T_j^0) \rightarrow^d \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\bar{\sigma}_{j+1}/\bar{\sigma}_j)W(s) - (\bar{\pi}_{j+1}/\bar{\pi}_j)|s|/2 & s > 0 \end{cases},$$

where  $\bar{\pi}_j = \Delta_j(\tau)' \bar{H}_j^0(\tau) \Delta_j(\tau)$ ,  $\bar{\pi}_{j+1} = \Delta_j(\tau)' \bar{H}_{j+1}^0(\tau) \Delta_j(\tau)$ ,  $\bar{\sigma}_j^2 = \tau(1-\tau) \Delta_j(\tau)' \bar{J}_j^0 \Delta_j(\tau)$ ,  $\bar{\sigma}_{j+1}^2 = \tau(1-\tau) \Delta_j(\tau)' \bar{J}_{j+1}^0 \Delta_j(\tau)$  and  $W(s)$  is the standard two-sided Brownian motion. Also,

$$\sqrt{NT}(\hat{\beta}_j(\tau) - \beta_j^0(\tau)) \rightarrow^d N(0, \bar{V}_j)$$

with  $\bar{V}_j = \tau(1-\tau) \bar{\Omega}_j^0(\tau) / (\lambda_j^0 - \lambda_{j-1}^0)^2$  and  $\bar{\Omega}_j^0(\tau) = (\bar{H}_j^0(\tau))^{-1} \bar{J}_j^0 (\bar{H}_j^0(\tau))^{-1}$  for  $j = 1, \dots, m + 1$ .

An equivalent way to express the limiting distribution in Theorem 2 is as follows

$$N \left\{ \frac{\left( \Delta_{NT,j}(\tau)' \bar{H}_j^0(\tau) \Delta_{NT,j}(\tau) \right)^2}{\tau(1-\tau) \Delta_{NT,j}(\tau)' \bar{J}_j^0 \Delta_{NT,j}(\tau)} \right\} (\hat{T}_j - T_j^0) \\ \rightarrow^d \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\bar{\sigma}_{j+1}/\bar{\sigma}_j)W(s) - (\bar{\pi}_{j+1}/\bar{\pi}_j)|s|/2 & s > 0 \end{cases},$$

where  $\Delta_{NT,j}(\tau)$  denotes the magnitude of the  $j$ th break for a given finite sample of size  $(N, T)$ . This representation clearly illustrates the effect of the cross section sample size  $N$  on the precision of the break estimates. Namely, if everything in the parentheses stays the same, increasing  $N$  will proportionally decrease the width of the confidence interval for break dates. Such a finding was first

reported by Bai et al. (1998) when considering the estimation of a common break in multivariate time series regressions.

We now extend the analysis to consider structural breaks in multiple quantiles. Define, for a given  $\mathcal{T}_\omega$  and  $T^b = (T_1, \dots, T_m)$ ,

$$S_{NT}(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b) = \sum_{h=1}^q \sum_{j=0}^m \sum_{t=T_{j+1}}^{T_{j+1}} \sum_{i=1}^N \rho(y_{it} - x'_{it} \beta_{j+1}(\tau_h))$$

and

$$(\hat{\beta}(\mathcal{T}_\omega), \hat{T}^b) = \arg \min_{\beta(\mathcal{T}_\omega), T^b \in \Lambda_\varepsilon} S_{NT}(\mathcal{T}_\omega, \beta(\mathcal{T}_\omega), T^b).$$

**Assumption B8.** The conditional quantiles for  $\tau_h$  ( $h = 1, \dots, q$ ) satisfy Assumptions B1-B6. There exists at least one quantile  $\tau_j$  ( $1 \leq j \leq q$ ) satisfying Assumption B7. Also,  $q$  is fixed as  $T \rightarrow \infty$ .

**Corollary 2** Under Assumption B8, for  $j = 1, \dots, m$ ,

$$\left( \frac{\bar{\pi}_j^*}{\bar{\sigma}_j^*} \right)^2 v_T^2 (\hat{T}_j - T_j^0) \rightarrow^d \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\bar{\sigma}_{j+1}^*/\bar{\sigma}_j^*)W(s) - (\bar{\pi}_{j+1}^*/\bar{\pi}_j^*)|s|/2 & s > 0 \end{cases},$$

where  $W(s)$  is the standard two-sided Brownian motion,  $\bar{\pi}_j^* = (1/q) \sum_{h=1}^q \Delta_j(\tau_h)' \bar{H}_j^0(\tau_h) \Delta_j(\tau_h)$ ,  $\bar{\pi}_{j+1}^* = (1/q) \sum_{h=1}^q \Delta_j(\tau_h)' \bar{H}_{j+1}^0(\tau_h) \Delta_j(\tau_h)$ ,  $\bar{\sigma}_j^{*2} = (1/q^2) \sum_{h=1}^q \sum_{g=1}^q (\tau_h \wedge \tau_g - \tau_h \tau_g) \Delta_j(\tau_h)' \bar{J}_j^0 \Delta_j(\tau_g)$  and  $\bar{\sigma}_{j+1}^{*2} = (1/q^2) \sum_{h=1}^q \sum_{g=1}^q (\tau_h \wedge \tau_g - \tau_h \tau_g) \Delta_j(\tau_h)' \bar{J}_{j+1}^0 \Delta_j(\tau_g)$ .

In summary, the method discussed in this section permits us to estimate structural breaks using individual level data. In this aspect, a closely related paper is Bai (2009), who considers common breaks in a linear panel data regression. A key difference is that Bai (2009) studies change in the mean or the variance while here we consider change in the conditional distribution. Hence, the results complement each other.

## 6 A procedure to determine the number of breaks

The following procedure is motivated by Bai and Perron (1998). It is built upon two test statistics,  $SQ_\tau$  and  $DQ$ , proposed in Qu (2008). We first give a brief review of these two tests.

The  $SQ_\tau$  test is designed to detect structural change in a given quantile  $\tau$ :

$$SQ_\tau = \sup_{\lambda \in [0,1]} \left\| (\tau(1-\tau))^{-1/2} \left[ H_{\lambda,T}(\hat{\beta}(\tau)) - \lambda H_{1,T}(\hat{\beta}(\tau)) \right] \right\|_\infty,$$

where

$$\begin{aligned} H_{\lambda,T}(\hat{\beta}(\tau)) &= \left( \sum_{t=1}^T x_t x_t' \right)^{-1/2} \sum_{t=1}^{\lfloor \lambda T \rfloor} x_t \psi_\tau(y_t - x_t' \hat{\beta}(\tau)) \text{ and} \\ \psi_\tau(u) &= \tau - \mathbf{1}(u < 0) \end{aligned}$$

if we have a single time series, and

$$H_{\lambda,T}(\hat{\beta}(\tau)) = \left( \sum_{t=1}^T \sum_{i=1}^N x_{it} x_{it}' \right)^{-1/2} \sum_{t=1}^{\lfloor \lambda T \rfloor} \sum_{i=1}^N x_{it} \psi_\tau(y_{it} - x_{it}' \hat{\beta}(\tau))$$

with repeated cross section,  $\hat{\beta}(\tau)$  is the estimate using the full sample assuming no structural change, and  $\|\cdot\|_\infty$  is the sup norm, i.e. for a generic vector  $z = (z_1, \dots, z_k)$ ,  $\|z\|_\infty = \max(z_1, \dots, z_k)$ .

The  $DQ$  test is designed to detect structural changes in quantiles in an interval  $\mathcal{T}_\omega$ :

$$DQ = \sup_{\tau \in \mathcal{T}_\omega} \sup_{\lambda \in [0,1]} \left\| H_{\lambda,T}(\hat{\beta}(\tau)) - \lambda H_{1,T}(\hat{\beta}(\tau)) \right\|_\infty.$$

These tests are asymptotically nuisance parameter free and tables for critical values are provided in Qu (2008). They do not require estimating the variance parameter (more specifically, the sparsity), thus having monotonic power even when multiple breaks are present. Qu (2008) provided a simple simulation study. The results show that these two tests compare favorably with Wald-based tests (c.f. Figure 1 in Qu, 2008).

We also need the following tests for the purpose of testing  $l$  against  $l + 1$  breaks, labelled as  $SQ_\tau(l + 1|l)$  test and  $DQ(l + 1|l)$  test. The construction follows Bai and Perron (1998). Suppose a model with  $l$  breaks has been estimated with the estimates denoted by  $\hat{T}_1, \dots, \hat{T}_l$ . These values partition the sample into  $(l + 1)$  segments, with the  $j$ th segment being  $[\hat{T}_{j-1} + 1, \hat{T}_j]$ . The strategy proceeds by testing each of the  $(l + 1)$  segments for the presence of an additional break. We let  $SQ_{\tau,j}$  and  $DQ_j$  denote the  $SQ_\tau$  and  $DQ$  test applied to the  $j$ th segment, i.e.,

$$\begin{aligned} SQ_{\tau,j} &= \sup_{\lambda \in [0,1]} \left\| (\tau(1 - \tau))^{-1/2} \left[ H_{\lambda, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) - \lambda H_{1, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) \right] \right\|_\infty, \\ DQ_j &= \sup_{\tau \in \mathcal{T}_\omega} \sup_{\lambda \in [0,1]} \left\| H_{\lambda, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) - \lambda H_{1, \hat{T}_{j-1}, \hat{T}_j}(\hat{\beta}_j(\tau)) \right\|_\infty, \end{aligned}$$

where

$$H_{\lambda, T_{j-1}, T_j}(\hat{\beta}_j(\tau)) = \left( \sum_{t=T_{j-1}+1}^{T_j} x_t x_t' \right)^{-1/2} \sum_{t=T_{j-1}+1}^{\lfloor \lambda(T_j - T_{j-1}) \rfloor} x_t \psi_\tau(y_t - x_t' \hat{\beta}_j(\tau))$$



if we have a single time series,

$$H_{\lambda, T_{j-1}, T_j}(\hat{\beta}_j(\tau)) = \left( \sum_{t=T_{j-1}+1}^{T_j} \sum_{i=1}^N x_{it} x'_{it} \right)^{-1/2} [\lambda(T_j - T_{j-1})] \sum_{t=T_{j-1}+1}^N \sum_{i=1}^N x_{it} \psi_{\tau}(y_{it} - x'_{it} \hat{\beta}_j(\tau))$$

if we have repeated cross sections,  $\hat{\beta}_j(\tau)$  is the estimate using the  $j$ th regime. Then,  $SQ_{\tau}(l+1|l)$  and  $DQ(l+1|l)$  equals to the maximum of the  $SQ_{\tau, j}$  and  $DQ_j$  over the  $l+1$  segments, i.e.,

$$\begin{aligned} SQ_{\tau}(l+1|l) &= \max_{1 \leq j \leq l+1} SQ_{\tau, j}, \\ DQ(l+1|l) &= \max_{1 \leq j \leq l+1} DQ_j. \end{aligned}$$

We reject in favor of a model with  $(l+1)$  breaks if the resulting value is sufficiently large.

Some additional notation is needed to present the limiting distributions of the  $SQ_{\tau}(l+1|l)$  and  $DQ(l+1|l)$  tests. Let  $B_p(s)$  be a vector of  $p$  independent Brownian bridge processes on  $[0, 1]$ . Also, let  $B_p(u, v) = (B_{(1)}(u, v), \dots, B_{(p)}(u, v))'$  be a  $p$ -vector of independent Gaussian processes with each component defined on  $[0, 1]^2$  having zero mean and covariance function

$$E(B_{(i)}(r, u)B_{(i)}(s, v)) = (r \wedge s - rs)(u \wedge v - uv).$$

The process  $B_{(i)}(r, u)$  is often referred to as the Brownian Pillow or tucked Brownian Sheet.

**Theorem 3** *Suppose  $m = l$  and that the model is given by (1) or (11) with Assumptions 1-6 or B1-B7 satisfied. Then,  $P(SQ_{\tau}(l+1|l) \leq x) \rightarrow G_p(x)^{l+1}$  with  $G_p(x)$  the distribution function of  $\sup_{s \in [0, 1]} \|B_p(s)\|_{\infty}$ . If these assumptions hold uniformly in  $\mathcal{T}_{\omega}$ , then  $P(DQ(l+1|l) \leq x) \rightarrow \bar{G}_p(x)^{l+1}$  with  $\bar{G}_p(x)$  the distribution function of  $\sup_{\tau \in \mathcal{T}_{\omega}} \sup_{s \in [0, 1]} \|B_p(\lambda, \tau)\|_{\infty}$ .*

The above limiting distributions depend on the number of parameters in the model ( $p$ ), the number of changes under the alternative hypothesis ( $l+1$ ) and the trimming proportion ( $\omega$ ) in the case of the DQ test (note that we assume  $\mathcal{T}_{\omega} = [\omega, 1 - \omega]$ ). Instead of reporting critical values for each case, we conduct extensive simulations and provide relevant information via response surface regressions. Specifically, we first simulate critical values for specifications with  $1 \leq p \leq 20$ ,  $0 \leq l \leq 4$  and  $0.05 \leq \omega \leq 0.30$ , with the increment being 0.01. Then, we estimate a class of nonlinear regression of the form:

$$CV_i(\alpha) = (z'_{1i}\beta_1) \exp(z'_{2i}\beta_2) + e_i,$$

where  $CV_i$  is a simulated critical value for a particular specification  $i$ ,  $z_{1i}$  and  $z'_{2i}$  indicate the corresponding  $p, l$  and  $\omega$ ,  $e_i$  is an error term, and  $\alpha$  is the nominal size. Regressors are selected such that the  $R^2$  is not smaller than 0.9999. The selected regressors are

- $SQ_\tau(l+1|l)$  test:  $z_1 = (1, p, l+1, 1/p, (l+1)p, )$  and  $z_2 = (1/(l+1))$ ;
- $DQ(l+1|l)$  test:  $z_1 = (1, p, l+1, 1/p, (l+1)p, (l+1)\omega)$  and  $z_2 = (1/(l+1), 1/(l+1)\omega, \omega)$ .

The estimated coefficients are reported in Table 1, which can then be used for a quick calculation of the relevant critical values for a particular application.

We now discuss a procedure that can be used to determine the number of breaks (we consider the interval  $\mathcal{T}_\omega$  and focus on the quantiles  $\tau_1, \dots, \tau_q \in \mathcal{T}_\omega$ ).

- **Step 1.** Apply the  $DQ$  test. If the test does not reject, conclude that there is no break and terminate the procedure. If it rejects, then estimate the model allowing one break. Save the estimated break date and proceed to Step 2.
- **Step 2.** Apply the  $DQ(l+1|l)$  tests starting with  $l = 1$ . Increase the value of  $l$  if the test rejects the null hypothesis. In each stage, the model is re-estimated and the break dates are the global minimizers of the objective function allowing  $l$  breaks. Continue the process until the test fails to reject the null.
- **Step 3.** Let  $\hat{l}$  denote the first value for which the test fails to reject. Estimate the model allowing  $\hat{l}$  breaks. Save the estimated break dates and confidence intervals.
- **Step 4.** This step treats the  $q$  quantiles separately and can be viewed as a robustness check. Specifically, for every quantile  $\tau_h$  ( $h = 1, \dots, q$ ), apply the  $SQ_\tau$  and  $SQ_\tau(l+1|l)$  tests. Carry out the same operations as in Steps 1 to 3. Examine whether the estimated breaks are in agreement with those from Step 3.

Since this is a sequential procedure, it is important to consider its rejection error. Suppose there is no break and a 5% significance level is used. Then, there is a 95% chance that the procedure will be terminated in Step 1, implying the probability of finding one or more breaks is 5% in large samples. If there are  $m$  breaks with  $m > 0$ , then, similarly, the probability of finding more than  $m$  breaks will be at most 5%. Of course, the probability of finding less than  $m$  breaks in finite samples will vary from case to case depending on the magnitude of the breaks.

## 7 Monte Carlo experiments

We focus on the following location-scale model with a single structural change:

$$y_{it} = 1 + x_{it} + \delta_N x_{it} 1(t > T/2) + (1 + x_{it}) u_{it}, \quad (12)$$

where  $x_{it} \sim \text{i.i.d.} \chi^2(3)/3$ ,  $u_{it} \sim \text{i.i.d.} N(0, 1)$ ,  $i = 1, \dots, N$  and  $t = 1, \dots, T$ . We set  $T = 100$  and consider  $N = 1, 50$  and  $100$ .  $\delta_N = 1.0/\sqrt{N}, 2.0/\sqrt{N}$  and  $3.0/\sqrt{N}$ , where the scaling factor  $\sqrt{N}$  makes the break sizes comparable across different  $N$ 's. Note that the powers of the DQ test against these three alternatives (at 5% nominal level, constructed with  $\mathcal{T}_\omega = [0.2, 0.8]$ ) are about 27%, 83% and 99% respectively. The powers of the sup-Wald test (Andrews, 1993) are similar. Thus,  $1.0/\sqrt{N}$  can be viewed as a small break and  $3.0/\sqrt{N}$  as a large break.

The computational detail is as follows. All parameters are allowed to change when estimating the model. The break date is searched over  $[0.15T, 0.85T]$ . The Bofinger bandwidth is used for obtaining the quantile density function. Finally, all simulation results are based on 2000 replications.

## 7.1 Coverage rates

We examine the coverage property of the asymptotic confidence intervals at 95% nominal level. Seven evenly spaced quantiles (0.2, 0.3, ..., 0.8) are considered in the analysis.

Table 2 presents coverage rates for the break date. The first seven columns are based on a single quantile function. The empirical coverage rates are between 86.6% and 92.0% when  $\delta_N = 1.0/\sqrt{N}$ , between 88.6% and 93.4% when  $\delta_N = 2.0/\sqrt{N}$  and between 92.6% and 97.0% when  $\delta_N = 3.0/\sqrt{N}$ . The values are quite stable across different  $N$ 's, suggesting that the framework developed in Section 5 provides a useful approximation. The last column in the table is based on all seven quantiles. The result is quite similar to the single quantile case.

Table 3 reports coverage rates for  $\delta_N$ . When the break size is small ( $\delta_N = 1.0/\sqrt{N}$ ) and the break date is estimated using a single quantile function, the confidence interval shows under-coverage, particularly for quantiles near the tail of the distribution. In contrast, when the break date is estimated using the seven quantile functions, the coverage rates are uniformly closer to the nominal rate, with the improvement being particularly important for more extreme quantiles. This suggests that even if one is only interested in a single quantile, say the 20th percentile, it may still be advantageous to borrow information from other quantiles when estimating the break date. Note that once the break size reaches  $2.0/\sqrt{N}$ , the coverage rate is satisfactory, being robust to different cross section sample sizes and to whether the break date is estimated based on a single quantile or multiple quantiles.

The above result suggests that the asymptotic framework delivers a useful approximation. A shortcoming to the shrinking break framework is that the confidence intervals are liberal when the true break is small. This problem can be alleviated to some extent by borrowing information across quantiles. It should be noted that a few studies have addressed this under-coverage issue in other

contexts and alternative inferential frameworks have been proposed, see Bai (1995) and Elliott and Müller (2007, 2010). A method that allows for multiple breaks has yet to be developed.

## 7.2 Empirical distribution of break date estimates

We compare estimates based on the median regression, the joint analysis of seven quantiles and the conditional mean regression. In addition to letting  $u_{it}$  being  $N(0,1)$ , we also consider a  $t$ -distribution with 2.5 degrees of freedom with other specifications unchanged.

Table 4 reports the mean absolute deviation (MAD) and the 90% inter-quantile range (IQR90) of the estimates. The upper panel corresponds to  $u_{it}$  being  $N(0,1)$ . It illustrates that the estimates based on the median and mean regression have similar properties, while the estimates based on multiple quantile functions have noticeably higher precision. The lower panel corresponds to  $u_{it}$  being  $t(2.5)$ . It shows that the estimates based on the median and multiple quantile functions are similar and are often substantially more precise than the conditional mean regression. Thus, there can be an important efficiency gain from considering quantile-based procedures in the presence of fat-tailed error distributions, even if the goal is to detect changes in the central tendency. Even though this is documented using a very simple model, the result should carry through to more general settings. Similar findings are reported in Bai (1998) in a median regression framework with i.i.d errors.

## 8 Empirical Applications

### 8.1 U.S. real GDP growth

It is widely documented that the volatility of the U.S. real GDP growth has declined substantially since the early to mid-1980s. For example, McConnell and Perez-Quiros (2000) considered an  $AR(1)$  model for the GDP growth and found a large break in the residual variance occurring in the first quarter of 1984. We revisit this issue using a quantile regression framework. The data set we use contains quarterly real GDP growth rates for the period 1947:2 to 2009:2. It is obtained from the web page of the St. Louis Fed (the GDPC96 series) and corresponds to the maximum sample period available at the time of writing our paper.

We consider the following model for the annualized quarterly growth rate  $y_t$ :

$$Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p}) = \mu_j(\tau) + \sum_{i=1}^p \alpha_{i,j}(\tau)y_{t-i} \quad (t = T_{j-1}^0 + 1, \dots, T_j^0),$$

where  $j$  is the index for the regimes and  $T_j^0$  corresponds to the last observation from the  $j$ th regime. The break dates and the number of breaks are assumed to be unknown. We consider

five equally spaced quantiles,  $\tau = 0.20, 0.35, 0.50, 0.65, 0.80$ , which are chosen to examine both the central tendency and the dispersion of the conditional distribution.

The Bayesian Information Criterion is applied to determine the lag order of the quantile autoregressions, with the maximum lag order set to  $\text{int}[12(T/100)^{1/4}]$ , where  $T$  is the sample size. The criterion selects 2 lags for the quantile  $\tau = 0.20$  and one lag for the other quantiles. We take a conservative approach and set  $p = 2$  for all five quantiles under consideration.

First, we study the five quantiles jointly. The results are summarized in Panel (a) of Table 5. The  $DQ$  test, applied to the interval  $[0.2, 0.8]$ , equals 0.994. This exceeds the 5% critical value, which is 0.906, suggesting at least one break is present. The  $DQ(2|1)$  test equals 0.612 and is below the 10% critical value. Therefore, we conclude that only one break is present. The break date estimated using all five quantiles is 1984:1 with a 95% confidence interval  $[77:3, 84:2]$ . This finding is consistent with McConnell and Perez-Quiros (2000).

Next, we analyze the quantiles separately. The results are summarized in Panel (b) of Table 5. The  $SQ_\tau$  test detects structural change only in the upper quantiles ( $\tau = 0.65, 0.80$ ), but not the median and the lower quantiles ( $\tau = 0.20, 0.35$ ). For  $\tau = 0.65$ , the estimated break date is 1984:2 and for  $\tau = 0.80$  the date is 1984:1. This confirms that the break is common to both quantiles.

Table 6 reports coefficient estimates conditional on the break date 1984:1. For both quantiles ( $\tau = 0.65, 0.80$ ), the structural change is characterized by a large decrease in the intercept and a small change in the sum of the autoregressive coefficients. Thus, overall the dispersion of the upper tail has decreased substantially. Overall, the results suggest that a major change in the GDP growth should be attributed to the fact that the growth was less rapid during expansions, while the recessions have remained just as severe when they occurred.

We also considered a finer partition including nine evenly spaced quantiles from 0.20 to 0.80. The results are reported in Table 5B. The application of the  $DQ$  test reports one break, estimated at 1984:2, with the 95% confidence interval being  $[1980:1, 1986:4]$ . The consideration of individual quantiles shows that only  $\tau = 0.65, 0.725$  and 0.80 are affected by the break, with the dates being 1984:2, 1984:2 and 1984:1, respectively. Overall, the results are consistent with the five-quantile specification considered earlier.

To further examine the robustness of the result, we repeated the analysis excluding observations from 2008-2009. The  $DQ(l+1|l)$  test detects one break and the estimated break date is 1984:2, confirming our findings using the full sample. We also studied the quantiles separately with the results summarized in Table 7. It shows that the upper quantiles are affected by structural change while the median and the 35th percentile are stable. The only difference from the full sample case

is that the 20th percentile also exhibits a break. However, the break date is 1958:1 and there is no statistically significant break during the 1980s. Thus, the general picture still holds.

## 8.2 Underage Drunk Driving

Motor vehicle crash is the leading cause of death among youth ages 15-20, a high proportion of which involves drunk driving. Here we study structural change in the blood alcohol concentration (BAC) among young drivers involved in traffic accidents. The study is motivated by the fact that the BAC level is an important measure of alcohol impairment, whose changes deliver important information on whether and how young driver's drinking behavior has changed over time.

The data set contains information on young drivers (less than 21-year-old) involved in motor vehicle accidents for the state of California over the period 1983-2007. It is obtained from National Highway Traffic Safety Administration (NHTSA), which reports the BAC level of the driver, his/her age, gender and whether the crash is fatal. For some observations, the BAC levels were not measured at the accident and their values were reconstructed using multiple imputations. They constitute about 26% of the sample. In these cases the first imputed value is used in our analysis.

The numbers of observations in each quarter vary between 108 and 314 with the median being 191. We start by constructing a representative random subsample, containing 108 observations in each quarter with 10,800 observations in total.<sup>3</sup> It should be noted that such a procedure does not introduce bias into our estimates. However, it does involve some arbitrariness and later in the paper we will report relevant results using the full sample to address this issue.

We consider the following model

$$Q_{y_{it}}(\tau|x_{it}) = \mu_j(\tau) + x'_{it}\gamma_j(\tau) \quad (t = T_{j-1}^0 + 1, \dots, T_j^0),$$

where  $y_{it}$  is the BAC level. The BAC levels below the 62th percentile are identically zero. Thus, we consider only the upper quantiles 0.70, 0.75, 0.80 and 0.85, all of which have a positive BAC in the aggregate. The consequence of such an action will be examined later in the paper.

A "general to specific" approach is adopted to determine which variables to include in the regression. We start with a regression that includes age, gender, and quarterly dummies. The dummy variable for a fatal crash is not included to avoid possible endogeneity. The model is estimated assuming there is no structural change, with insignificant regressors sequentially eliminated until the remaining ones are significant at 10% level. This leaves the age, gender and a dummy variable for the fourth quarter (labelled as the winter dummy) in the regression.

---

<sup>3</sup>We used the "surveyselect" procedure (SAS) with "Method" set to srs (simple random sampling without replacement) and "Seed" set to 2009.

We first analyze the four quantiles jointly. The results are summarized in Table 8. The  $DQ(l+1|l)$  test, applied to the interval  $[0.70, 0.85]$ , reports two breaks. Their dates are 1985:1 and 1992:2 with the 95% confidence intervals being  $[84:1, 86:1]$  and  $[91:2, 92:3]$ . We then consider the quantiles separately. The test suggests that the 70th, 75th and 80th percentiles are affected by two breaks, while the 85th percentile is only affected by the second break. The first estimated break is 1985:2 and the second is either 1992:2 or 1993:2. Although there is some local variation, overall these estimates are consistent with the ones based on multiple quantiles. It is interesting to point out that the confidence intervals for these two breaks include two historically important policy changes. Specifically, the National Minimum Drinking Age Act (MDA) was passed on July 17, 1984. The federal beer tax was doubled in 1991, while the California state beer tax experienced a four-fold increase in the same year.

To examine the result sensitivity, we consider a finer partition including seven evenly spaced quantiles from 0.70 to 0.85. The results are reported in Table 8B. The application of the DQ test reports two breaks, estimated at the same dates with identical 95% confidence intervals as the four-quantile case. The consideration of individual quantiles shows that the quantiles  $\tau = 0.70$  to 0.80 are affected by two breaks and  $\tau = 0.825$  and 0.85 are affected by one break. The estimated break dates and confidence intervals are consistent with the four-quantile case. Overall, the findings remain qualitatively similar.

Figures 1 and 2 report the changes in the quantile functions for representative values of  $x_{it}$  conditioning on break dates 1985:1 and 1992:2. They cover three age groups: 17, 18 and 19, which correspond to the 25th, 50th and 75th percentile of the unconditional distribution of  $x_{it}$ . Males and females are reported separately. The winter dummy is set to zero (setting it to one produces similar results and is omitted to save space). Figure 1 presents results for males. Each panel contains the changes and their pointwise 95% confidence intervals. Three interesting patterns emerge. First, the changes are all negative. They are also economically meaningful because BAC levels as low as 0.02 can affect a person's driving ability with the probability of a crash increasing significantly after 0.05 according to studies by NHTSA. Secondly, for the first break, the change becomes smaller as age increases while for the second the opposite is true. Finally and most importantly, the change is smaller for higher quantiles, suggesting that the policies are more effective for "light drinkers" than for "heavy drinkers" in the sample. This is unfortunate since heavy drinkers are more likely to cause an accident. Figure 2 presents results for females. The findings are qualitatively similar, except for the second break the change is more homogeneous across quantiles. It should also be noted that for the first break, the confidence intervals at the 85th percentile typically include zero

(except for the first figure in 1(a)). This is consistent with the findings in Table 8, where for this quantile only one break is detected.

### 8.2.1 Further robustness analysis

We focus on the following two issues: 1) the distribution of BAC levels has a mass at zero, and 2) the analysis has been conducted on a subsample.

To address the first issue, we apply the censored quantile regression of Powell (1996) conditioning on the break dates 1985:1 and 1992:2. The model is estimated using the *crq* procedure in *quantreg*. The estimated changes are reported in Figure 1 and 2 (the solid line with triangle). The results are very similar.

To address the second issue, the break dates are re-estimated using the full sample. The estimates are 1985:2 and 1992:3 using the four quantile functions. Conditioning on the break dates, the model is re-estimated using both quantile regression and censored quantile regression. The estimated changes are found to be very similar to the ones reported in Figures 1 and 2. Most importantly, the three patterns discussed above still hold. The details are not repeated here to save space.

Thus, the results remain qualitatively the same after accounting for these two issues. In summary, this empirical application, although quite simple, illustrates that rich information can be extracted from considering structural change in the conditional quantile function.

## 9 Conclusions

We have considered the estimation of structural changes in regression quantiles, allowing for both time series models and repeated cross-sections. The proposed method can be used to determine the number of breaks, estimate the break locations and other parameters, and obtain corresponding confidence intervals. A simple simulation study suggests that the asymptotic theory provides useful approximation in finite samples. The two empirical applications, to the "Great Moderation" and underage drunk driving, suggest that our framework can potentially deliver richer information than simply considering structural change in the conditional mean function.



## References

- Andrews, D.W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-56.
- Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006), "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica* 74, 539-563.
- Bai, J. (1995), "Least Absolute Deviation Estimation of a Shift," *Econometric Theory*, 11, 403-436.
- (1996), "Testing for Parameter Constancy in Linear Regressions: An Empirical Distribution Function Approach," *Econometrica*, 64, 597-622.
- (1997), "Estimation Of A Change Point In Multiple Regression Models," *The Review of Economics and Statistics*, 79, 551-563.
- (1998), "Estimation of Multiple-Regime Regressions with Least Absolutes Deviation," *Journal of Statistical Planning and Inference*, 74, 103-134.
- (2000), "Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance," *Annals of Economics and Finance*, 1, 303-339.
- (2009), "Common Breaks in Means and Variances for Panel Data," forthcoming at the *Journal of Econometrics*.
- Bai, J., Lumsdaine, R. L., and Stock, J. H. (1998), "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395-432.
- Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.
- (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics*, 18, 1-22.
- Berkes, I., Horváth, L., and Kokoszka, P. (2004), "Testing for Parameter Constancy in GARCH(p; q) Models," *Statistical Probability Letters*, 70, 263-273.
- Buchinsky, M. (1994), "Changes in U.S. Wage Structure 1963–1987: An Application of Quantile Regression," *Econometrica*. 62, 405–458.
- Cai, Z. W., and Xu, X. P. (2008), "Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models," *Journal of the American Statistical Association*, 1595-1608.
- Chamberlain, G. (1994), "Quantile Regression, Censoring and the Structure of Wages," *Advances in Econometrics*. Christopher Sims, ed. New York: Elsevier, p 171–209.

- Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *The Journal of Finance*, 47, 1209-1227.
- Chernozhukov, V., and Umantsev, L. (2001), "Conditional Value-at-Risk: Aspects of Modeling and Estimation," *Empirical Economics*, 26, 271-292.
- Chen, J. (2008), "Estimating and Testing Quantile Regression with Structural Changes," Working Paper, NYU Dept. of Economics.
- Chow, Y. S., and Teicher, H. (2003), *Probability Theory: Independence, Interchangeability, Martingales* (Third Edition), Springer.
- Csörgő, M., and Horváth, L. (1998), *Limit Theorems in Change-point Analysis*, Wiley.
- Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.
- Eide, E., and Showalter, M. (1998), "The Effect of School Quality on Student Performance: A Quantile Regression Approach," *Economics Letters*, 58, 345-350.
- Elliott, G., and Muller, U. K. (2007), "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions," *Journal of Econometrics*, 141, 1196-1218.
- (2010), "Pre and Post Break Parameter Inference," working paper, Department of Economics, Princeton University.
- Engle, R. F., and Manganelli, S. (2004), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles," *Journal of Business and Economic Statistics*, 22, 367-381.
- Fiteni, I. (2002), "Robust Estimation Of Structural Break Points," *Econometric Theory*, 18, 349-386.
- Hall, A. R., and Sen, A. (1999), "Structural Stability Testing in Models Estimated by Generalized Method of Moments," *Journal of Business and Economic Statistics*, 17, 335-48.
- Hansen, B. E. (1992), "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321-35.
- Hendricks, B., and Koenker, R. (1992), "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association*, 87, 58-68.
- Hušková, M. (1997), "Limit Theorems for Rank Statistics," *Statistics and Probability Letters*, 32, 45-55.
- Kejriwal, M., and Perron, P. (2008), "The Limit Distribution of the Estimates in Cointegrated Regression Models with Multiple Structural Changes," *Journal of Econometrics*, 146, 59-73.

- Kim, M. O. (2007), "Quantile regression with varying coefficients," *The Annals of Statistics*, 35, 92–108.
- Knight, K. (1998), "Limiting distributions for L1 regression estimators under general conditions," *Annals of Statistics*, 26 755–770.
- Kokoszka, P., and Leipus, R. (1999), "Testing for Parameter Changes in ARCH Models," *Lithuanian Mathematical Journal*, 39, 182-195.
- (2000), "Change-Point Estimation in ARCH Models," *Bernoulli*, 6, 513-539.
- Koenker, R., and Bassett, G. Jr. (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.
- Koenker, R., and Xiao, Z. (2006), "Quantile Autoregression," *Journal of the American Statistical Association*, 101, 980-990.
- Levin, J. (2001), "For Whom the Reductions Count: A Quantile Regression Analysis of Class Size on Scholastic Achievement," *Empirical Economics*, 26, 221–246.
- Li, H., and Müller, U. K. (2009), "Valid Inference in Partially Unstable Generalized Method of Moments Models," *Review of Economic Studies*, 76, 343-365.
- McConnell, M. M., and Perez-Quiros, G. (2000), "Output Fluctuations in the United States: What Has Changed since the Early 1980's?," *American Economic Review*, 90, 1464-1476.
- Perron, P. (2006), "Dealing with Structural Breaks," in *Palgrave Handbook of Econometrics*, Vol. 1: Econometric Theory, eds, K. Patterson and T.C. Mills, Palgrave Macmillan, 278-352.
- Picard, D. (1985), "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability*, 17, 841-867.
- Piehl, A. M., Cooper, S. J., Braga, A. A., and Kennedy, D. M. (2003), "Testing for Structural Breaks in the Evaluation of Programs," *The Review of Economics and Statistics*, 85, 550-558.
- Poghosyan, S., and Roelly, S. (1998), "Invariance principle for martingale-difference random fields," *Statistics and Probability Letters*, 38, 235–245.
- Poterba, J., and Rueben, K. (1995), "The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Regression Methods," *NBER Working Paper*, No. 4734.
- Powell, J. L. (1986), "Censored Regression Quantiles", *Journal of Econometrics* 32, 143-155.
- Qu, Z. (2008), "Testing for Structural Change in Regression Quantiles," *Journal of Econometrics*, 146, 170-184.
- Qu, Z., and Perron, P. (2007), "Estimating and Testing Structural Changes in Multivariate Regressions," *Econometrica*, 75, 459-502.

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

Siddiqui, M. (1960), "Distribution of quantiles from a bivariate population," *Journal of Research of the National Bureau of Standards*, 64B, 145-150.

Taylor, J. W. (1999), "A Quantile Regression Approach to Estimating the Distribution of Multi-period Returns," *The Journal of Derivatives*, 7, 64-78.

Yao, Y. C. (1987), "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables," *The Annals of Statistics*, 15, 1321-1328.

## Appendix

We provide detailed proofs for results in Section 5. They imply Lemma 1, Theorem 1 and Corollary 1 as special cases ( $N = 1$ ). All limiting results are derived using  $T \rightarrow \infty$  with  $N$  fixed, or  $(N, T) \rightarrow \infty$ . For a given  $\tau$  and  $\phi \in R^p$ , define

$$q_{\tau,it}(\phi) = \rho_{\tau}(u_{it}^0(\tau) - x'_{it}\phi) - \rho_{\tau}(u_{it}^0(\tau)) \quad \text{and} \quad Q_k(\tau, \phi) = \sum_{t=1}^k \sum_{i=1}^N q_{\tau,it}(\phi).$$

The following decomposition, due to Knight (1998), will be used repeatedly in the analysis:

$$Q_k(\tau, \phi) = W_k(\tau, \phi) + Z_k(\tau, \phi), \tag{A.1}$$

where

$$\begin{aligned} W_k(\tau, \phi) &= - \sum_{t=1}^k \sum_{i=1}^N \psi_{\tau}(u_{it}^0(\tau)) x'_{it}\phi \quad \text{with } \psi_{\tau}(u) = \tau - 1(u < 0), \\ Z_k(\tau, \phi) &= \sum_{t=1}^k \sum_{i=1}^N \int_0^{x'_{it}\phi} (1(u_{it}^0(\tau) < s) - 1(u_{it}^0(\tau) < 0)) ds. \end{aligned} \tag{A.2}$$

$W_k(\tau, \phi)$  is a zero-mean process for fixed  $\tau$  and  $\phi$ , while  $Z_k(\tau, \phi)$  is in general not. Define

$$\begin{aligned} b_{it}(\tau, \phi) &= x'_{it}1(u_{it}^0(\tau) < x'_{it}\phi) \\ \xi_{it}(\tau, \phi) &= \{b_{it}(\tau, \phi) - b_{it}(\tau, 0)\} - E_{t-1} \{b_{it}(\tau, \phi) - b_{it}(\tau, 0)\}, \end{aligned}$$

where  $E_{t-1}$  is taken with respect to the  $\sigma$ -algebra generated by  $\{x_{it}, y_{i,t-1}, x_{i,t-1}, y_{i,t-2}, \dots\}_{i=1}^N$ . Note that  $\xi_{it}(\tau, \phi)$  forms an array of martingale differences for given  $\tau$  and  $\phi$ .

The following Lemma provides upper and lower bounds for  $Z_k(\tau, \phi)$  in terms of  $b_{it}(\tau, \phi)$ .

**Lemma A.1** *Suppose there is no structural change. Then, for every  $k = 1, \dots, T$ ,*

$$0 \leq (1/2) \sum_{t=1}^k \sum_{i=1}^N \{b_{it}(\tau, \phi/2) - b_{it}(\tau, 0)\} \phi \leq Z_k(\tau, \phi) \leq \sum_{t=1}^k \sum_{i=1}^N \{b_{it}(\tau, \phi) - b_{it}(\tau, 0)\} \phi.$$

**Proof of Lemma A.1.** We consider the  $(i, t)$ th term in the summation (A.2). If  $x'_{it}\phi \geq 0$ , then it is bounded from below by

$$\begin{aligned} \int_{x'_{it}\phi/2}^{x'_{it}\phi} \{1(u_{it}^0(\tau) < s) - 1(u_{it}^0(\tau) < 0)\} ds &\geq \int_{x'_{it}\phi/2}^{x'_{it}\phi} \{1(u_{it}^0(\tau) < x'_{it}\phi/2) - 1(u_{it}^0(\tau) < 0)\} ds \\ &= (1/2) \{b_{it}(\tau, \phi/2) - b_{it}(\tau, 0)\} \phi \geq 0. \end{aligned}$$

If  $x'_{it}\phi < 0$ , then this term is equal to

$$\begin{aligned} \int_{-|x'_{it}\phi|}^0 \{1(u_{it}^0(\tau) < 0) - 1(u_{it}^0(\tau) < s)\} ds &\geq \int_{-|x'_{it}\phi|}^{-|x'_{it}\phi|/2} \{1(u_{it}^0(\tau) < 0) - 1(u_{it}^0(\tau) < -|x'_{it}\phi|/2)\} ds \\ &= (1/2) \{b_{it}(\tau, \phi/2) - b_{it}(\tau, 0)\} \phi \geq 0. \end{aligned}$$

Taking the summation yields the lower bound. The upper bound can be proved similarly.  $\blacksquare$

The next lemma will be used to study the asymptotic properties of the bounds derived in the previous lemma.

**Lemma A.2** *Suppose there is no structural change and Assumptions B1-B3 and B5-B6 hold.*

1. Let  $\Phi_1 = \{\phi \in R^p : \|\phi\| = A(NT)^{-1/2}\}$  with  $A$  being some arbitrary constant, then

$$\sup_{s \in [0,1]} \sup_{\phi \in \Phi_1} \left\| (NT)^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} \sum_{i=1}^N \xi_{it}(\tau, \phi) \right\| = o_p(1).$$

2. Let  $\Phi_2 = \{\phi \in R^p : \|\phi\| = A(Nk)^{-1/2}(\log NT)^{1/2}\}$  with  $A$  being some arbitrary constant. Then, for any  $\epsilon > 0$  and  $\eta > 0$ , there exists a  $T_L < \infty$ , such that for all  $T \geq T_L$ ,

$$P \left( \sup_{1 \leq k \leq T} \sup_{\phi \in \Phi_2} \left\| (Nk)^{-1/2} (\log NT)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi) \right\| > \eta \right) < \epsilon.$$

3. Let  $h_T$  and  $d_T$  be positive sequences such that as  $T \rightarrow \infty$ ,  $h_T$  is nondecreasing,  $h_T \rightarrow \infty$ ,  $d_T \rightarrow \infty$  and  $(h_T d_T^2)/T \rightarrow h$  with  $0 < h < \infty$ . Let  $\Phi_3 = \{\phi \in R^p : \|\phi\| = (NT)^{-1/2} d_T\}$ . Then, for any  $\epsilon > 0$ ,  $D > 0$  and  $B > 0$ , there exists  $T_L < \infty$ , such that for all  $T \geq T_L$ ,

$$P \left( \sup_{Bh_T \leq k \leq T} \sup_{\phi \in \Phi_3} \left\| k^{-1} N^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi) \right\| > D d_T / \sqrt{T} \right) < \epsilon.$$

**Proof of Lemma A.2.** Without loss of generality, we assume that the components of  $x_{it}$  are nonnegative. Otherwise, let  $x_{it,j}$  denote the  $j$ th component of  $x_{it}$  and we can write  $x_{it,j} = x_{it,j}^+ - x_{it,j}^- \equiv x_{it,j} 1(x_{it,j} \geq 0) - x_{it,j} 1(x_{it,j} < 0)$ . Then,  $x_{it,j}^+$  and  $x_{it,j}^-$  are nonnegative and satisfy the Assumptions stated in the paper. We will only prove the result for a fixed  $\phi$ , because the uniformity over  $\Phi_1, \Phi_2$  or  $\Phi_3$  follows from the compactness of these sets and the monotonicity of  $b_{it}(\tau, \phi)$  in  $x'_{it}\phi$ , which can be verified using the same argument as in Theorem A3(ii) in Bai (1996).

Consider the first result. For any  $\phi \in \Phi_1$ ,  $\xi_{it}(\tau, \phi)$  satisfies

$$\begin{aligned} E_{t-1} \|\xi_{it}(\tau, \phi)\|^2 &\leq \|x_{it}\|^2 \left| F_{it}(x'_{it}\beta^0(\tau) + (NT)^{-1/2} A \|x_{it}\|) - \tau \right| \\ &\leq (NT)^{-1/2} A U_f \|x_{it}\|^3, \end{aligned} \quad (\text{A.3})$$

where  $U_f$  is defined in Assumption B2. Applying the Doob inequality and the Rosenthal inequality as in Bai (1996, p.618), we have, for any  $N$  and  $T$ ,

$$\begin{aligned} &P \left( \sup_{s \in [0,1]} \left\| (NT)^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} \sum_{i=1}^N \xi_{it}(\tau, \phi) \right\| > \epsilon \right) \\ &\leq \frac{M_1}{\epsilon^{2\gamma}} \left\{ (NT)^{-\gamma} E \left( \sum_{t=1}^T \sum_{i=1}^N E_{t-1} \|\xi_{it}(\tau, \phi)\|^2 \right)^\gamma + (NT)^{-\gamma} E \left( \sum_{t=1}^T \sum_{i=1}^N E_{t-1} \|\xi_{it}(\tau, \phi)\|^{2\gamma} \right) \right\} \end{aligned}$$

where  $\gamma$  is defined in Assumption B5(d) and  $M_1$  is some constant that depends only on  $p$  and  $\gamma$ . The first term inside of the curly brackets, after applying (A.3), is bounded by

$$(NT)^{-\gamma/2} E \left( AU_f(NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N \|x_{it}\|^3 \right)^\gamma \leq (NT)^{-\gamma/2} (AU_f)^\gamma M \rightarrow 0 \quad \text{as } T \rightarrow \infty \quad (\text{A.4})$$

due to Assumption B5(d). The second term can be rewritten as

$$(NT)^{-\gamma+1/2} E \left( (NT)^{-1/2} \sum_{t=1}^T \sum_{i=1}^N E_{t-1} \|\xi_{it}(\tau, \phi)\|^{2\gamma-2} \|\xi_{it}(\tau, \phi)\|^2 \right).$$

Because  $\|\xi_{it}(\tau, \phi)\| \leq \|x_{it}\|$ , the preceding quantity is less than or equal to

$$\begin{aligned} & (NT)^{-\gamma+1/2} E \left( (NT)^{-1/2} \sum_{t=1}^T \sum_{i=1}^N \|x_{it}\|^{2\gamma-2} E_{t-1} \|\xi_{it}(\tau, \phi)\|^2 \right) \\ & \leq (NT)^{-\gamma+1/2} E \left( AU_f(NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N \|x_{it}\|^{2\gamma+1} \right) \\ & \leq (NT)^{-\gamma+1/2} AU_f M \rightarrow 0 \quad \text{as } T \rightarrow \infty, \end{aligned} \quad (\text{A.5})$$

where the first inequality uses (A.3) and the second uses Assumption B5(d). (A.4) and (A.5) imply that Lemma A.2.1 holds for any given  $\phi \in \Phi_1$ .

Consider the second result. Note that

$$P \left( \sup_{1 \leq k \leq T} \left\| \frac{\sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi)}{(Nk \log(NT))^{1/2}} \right\| > \eta \right) \leq \sum_{k=1}^T P \left( \left\| \frac{\sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi)}{(Nk \log(NT))^{1/2}} \right\| > \eta \right).$$

The rest of the proof is similar to the first result. Applying the Markov inequality followed by the Rosenthal inequality, we have, for any  $\eta > 0$  and some  $\gamma > 2$ ,

$$\begin{aligned} & \sum_{k=1}^T P \left( \left\| \frac{\sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi)}{(Nk \log(NT))^{1/2}} \right\| > \eta \right) \\ & \leq \sum_{k=1}^T \frac{M_2}{(Nk \log(NT))^{\gamma/2} \eta^{2\gamma}} \left\{ E \left( (Nk)^{-1} \sum_{t=1}^k \sum_{i=1}^N \|x_{it}\|^3 \right)^\gamma + \left( (Nk)^{-1} E \sum_{t=1}^k \sum_{i=1}^N \|x_{it}\|^{2\gamma+1} \right) \right\} \\ & \leq \left( \frac{2M_2 M}{(N \log(NT))^{\gamma/2} \eta^{2\gamma}} \right) \sum_{k=1}^T k^{-\gamma/2} \end{aligned}$$

where  $M_2$  is a constant that only depends on  $p$ ,  $U_f$ ,  $\gamma$  and  $A$ , and the second inequality uses Assumption B5(d). Because  $\gamma > 2$ , the summation  $\sum_{k=1}^T k^{-\gamma/2}$  is finite. The term inside the parentheses converges to zero. Thus, Lemma A.2.2 holds.

For the third result, applying the same argument as above and using

$$E_{t-1} \|\xi_{it}(\tau, \phi)\|^2 \leq (NT)^{-1/2} d_T U_f \|x_{it}\|^3$$

for any  $\phi \in \Phi_3$ , we have

$$\begin{aligned}
& P \left( \sup_{Bh_T \leq k \leq T} \left\| k^{-1} N^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi) \right\| > Dd_T/\sqrt{T} \right) \\
& \leq M_1 \sum_{k \geq Bh_T} \left\{ \left( \frac{T^{1/2}}{kd_T} \right)^\gamma \left( \frac{U_f}{N^{1/2}D^2} \right)^\gamma E \left( (Nk)^{-1} \sum_{t=1}^k \sum_{i=1}^N \|x_{it}\|^3 \right)^\gamma \right. \\
& \quad \left. + \left( \frac{T^{1/2}}{kd_T} \right)^{2\gamma-1} \left( \frac{U_f}{D^{2\gamma}} \right) E \left( (Nk)^{-1} \sum_{t=1}^k \|x_{it}\|^{2\gamma+1} \right) \right\} \\
& \leq M_3 \sum_{k \geq Bh_T} \left\{ \left( \frac{T^{1/2}}{kd_T} \right)^\gamma + \left( \frac{T^{1/2}}{kd_T} \right)^{2\gamma-1} \right\},
\end{aligned} \tag{A.6}$$

where  $M_3 = M_1 M \max \left\{ (U_f/(N^{1/2}D^2))^\gamma, U_f/D^{2\gamma} \right\}$ . Rewrite the preceding line as

$$\begin{aligned}
& \frac{M_3}{B^{\gamma-1}} \left( \frac{T}{h_T d_T^2} \right)^{\gamma-1} \left( \frac{d_T^2}{T} \right)^{\frac{\gamma}{2}-1} \left\{ \sum_{k \geq Bh_T} \frac{(Bh_T)^{\gamma-1}}{k^\gamma} \right\} \\
& + \frac{M_3}{B^{2\gamma-2}} \left( \frac{T}{h_T d_T^2} \right)^{2\gamma-2} \left( \frac{d_T^2}{T} \right)^{\gamma-\frac{3}{2}} \left\{ \sum_{k \geq Bh_T} \frac{(Bh_T)^{2\gamma-2}}{k^{2\gamma-1}} \right\}.
\end{aligned}$$

For the first term,  $(T/(h_T d_T^2))^{\gamma-1} \rightarrow h^{1-\gamma} < \infty$ ,  $(d_T^2/T)^{\gamma/2-1} \rightarrow 0$  because  $\gamma > 2$ , and the summation inside the curly brackets is finite by the Euler-Maclaurin formula. Thus, this term converges to zero. The second term also converges to zero by the same argument. ■

The next Lemma provides convergence rates for parameter estimates using subsamples.

**Lemma A.3** *Suppose that there is no structural change and that Assumptions B1-B3 and B5-B6 hold. Let  $\hat{\beta}_k$  be the quantile regression estimate of  $\beta^0(\tau)$  using observations  $t = 1, \dots, k$ , and  $\phi_k = \hat{\beta}_k - \beta^0(\tau)$ . Then:*

1.  $(\log NT)^{-1/2} (Nk)^{1/2} \|\phi_k\| = O_p(1)$  uniformly over  $k \in [1, T]$ .
2. For any  $0 < \delta < 1$ ,  $(NT)^{1/2} \|\phi_k\| = O_p(1)$  uniformly over  $k \in [\delta T, T]$ .

**Proof of Lemma A.3.** We only prove the first result; the proof of the second is similar and simpler. The proof is by contradiction, i.e., showing that otherwise the objective function  $Q_k(\tau, \phi_k)$  will be strictly positive with probability close to 1, implying that  $\hat{\beta}_k$  cannot be its minimizer.

Due to the convexity of  $Q_k(\tau, \phi_k)$ ,<sup>4</sup> it suffices to consider its property over  $\phi_k$  satisfying

$$(\log NT)^{-1/2} (Nk)^{1/2} \|\phi_k\| = B,$$

<sup>4</sup>If  $g(\theta)$  is convex, then for any  $\gamma \geq 1$ ,  $g(\gamma\theta) - g(0) \geq \gamma(g(\theta) - g(0))$ .



where  $B$  is an arbitrary positive constant. Apply the Knight identity (A.1) and study the terms  $Z_k(\tau, \phi_k)$  and  $W_k(\tau, \phi_k)$  separately. For  $Z_k(\tau, \phi_k)$ , by Lemma A.1,

$$\begin{aligned} Z_k(\tau, \phi_k) &\geq (1/2) \sum_{t=1}^k \sum_{i=1}^N \{b_{it}(\tau, \phi_k/2) - b_{it}(\tau, 0)\} \phi_k \\ &\geq \frac{1}{2} \sum_{t=1}^k \sum_{i=1}^N E_{t-1} \{b_{it}(\tau, \phi_k/2) - b_{it}(\tau, 0)\} \phi_k - \left\| \frac{1}{2} \sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi_k/2) \phi_k \right\| \\ &= (a) - \|(b)\|. \end{aligned} \tag{A.7}$$

Term (a) satisfies

$$(\log NT)^{-1}(a) \geq \frac{1}{4}(\log NT)^{-1}L_f \sum_{t=1}^k \sum_{i=1}^N \phi'_k x_{it} x'_{it} \phi_k \geq \frac{1}{4}L_f B^2 \lambda_{\min},$$

where  $\lambda_{\min}$  is the minimum eigenvalue of  $(Nk)^{-1} \sum_{t=1}^k \sum_{i=1}^N x_{it} x'_{it}$ , the first inequality is due to the mean value theorem, and the second inequality uses  $(\log NT)^{-1/2} (Nk)^{1/2} \|\phi_k\| = B$ . Term (b), after dividing by  $(\log NT)$ , is of order  $o_p(B)$  uniformly in  $k \in [1, T]$  by Lemma A.2.2. Thus,

$$(\log NT)^{-1} ((a) - \|(b)\|) \geq \frac{1}{8} L_f B^2 \lambda_{\min} \tag{A.8}$$

with probability close to 1 uniformly in  $k \in [1, T]$  for large  $T$ . Now consider  $W_k(\tau, \phi_k)$ .

$$\begin{aligned} (\log NT)^{-1} |W_k(\tau, \phi_k)| &\leq (\log NT)^{-1} \left\| \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_{it}^0(\tau)) x'_{it} \right\| \|\phi_k\| \\ &= B \left\| (Nk)^{-1/2} (\log NT)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_{it}^0(\tau)) x'_{it} \right\|. \end{aligned} \tag{A.9}$$

Applying the Hájek-Rényi inequality for martingales (see, e.g., Chow and Teicher, 2003, p. 255),

$$P \left( \sup_{1 \leq k \leq T} \left\| (\log NT)^{-1/2} (Nk)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_{it}^0(\tau)) x'_{it} \right\| > C \right) \leq \frac{1}{C^2} \sum_{i=1}^N \sum_{t=1}^T \frac{E \|x_{it}\|^2}{Nt \log(NT)},$$

where  $C$  is an arbitrary constant. Because  $E \|x_{it}\|^2 < \infty$  and  $\sum_{t=1}^T t^{-1} = (\log T)$ , the left hand side can be made arbitrarily small by choosing a large  $C$ . Therefore, if  $B$  is large, the term (A.9) will be dominated by (A.8) asymptotically, implying  $Q_k(\tau, \phi_k)$  will be strictly positive with probability close to 1 for large  $T$ . This contradicts the fact that  $\phi_k$  minimizes  $Q_k(\tau, \phi)$ , thus proving the first result. The second result can be proved along the same lines, by applying Lemma A.2.1 to term (b) and the Hájek-Rényi inequality  $W_k(\tau, \phi_k)$ . ■

The next Lemma shows that the objective function  $Q_k(\tau, \phi)$  can be bounded in various ways when the model is estimated using subsamples of various sizes. It is an extension of Lemma A.1 in Bai (1995) along the two directions, i.e., by allowing for time series dynamics and a cross-sectional dimension.

**Lemma A.4** *Suppose there is no structural change and Assumptions B1-B3 and B5-B6 hold.*

1. For any  $\delta \in (0, 1)$ ,  $\sup_{\delta T \leq k \leq T} |\inf_{\phi} Q_k(\tau, \phi)| = O_p(1)$ .
2.  $\sup_{1 \leq k \leq T} |\inf_{\phi} Q_k(\tau, \phi)| = O_p(\log NT)$ .
3. For any  $\delta \in (0, 1)$ ,  $\epsilon > 0$  and  $D > 0$  and  $T$  sufficiently large

$$P \left( \inf_{\delta T \leq k \leq T} \inf_{\|\phi\| \geq (NT)^{-1/2} \log NT} Q_k(\tau, \phi) < D \log NT \right) < \epsilon.$$

4. For any  $\delta \in (0, 1)$ ,  $\epsilon > 0$  and  $D > 0$ , there exists  $A > 0$  such that when  $T$  is sufficiently large

$$P \left( \inf_{\delta T \leq k \leq T} \inf_{\|\phi\| \geq A(NT)^{-1/2}} Q_k(\tau, \phi) < D \right) < \epsilon.$$

5. Let  $h_T$  and  $d_T$  be positive sequences such that  $h_T$  is nondecreasing,  $h_T \rightarrow \infty$ ,  $d_T \rightarrow \infty$  and  $(h_T d_T^2)/T \rightarrow h$  with  $0 < h < \infty$ . Then for each  $\epsilon > 0$  and  $D > 0$  there exists an  $A > 0$ , such that when  $T$  is large enough,

$$P \left( \inf_{Ah_T \leq k \leq T} \inf_{\|\phi\| \geq d_T(NT)^{-1/2}} Q_k(\tau, \phi) < D \right) < \epsilon.$$

6. Suppose the same conditions as in part (5) hold. Then, for any  $A > 0$ ,

$$\sup_{k \leq Ah_T} \left| \inf_{\|\phi\| \leq d_T(NT)^{-1/2}} Q_k(\tau, \phi) \right| = O_p(1).$$

**Proof of A.4.1.** By Lemma A.3,  $\sqrt{NT} \|\hat{\beta}_k - \beta^0(\tau)\| = O_p(1)$  uniformly over  $k \in [\delta T, T]$ . Thus, it suffices to prove

$$\sup_{1 \leq k \leq T} \sup_{\|\phi\| \leq A(NT)^{-1/2}} |Q_k(\tau, \phi)| = O_p(1) \quad \text{for any } A > 0. \quad (\text{A.10})$$

Note that the sup is taken over  $1 \leq k \leq T$  instead of  $k \in [\delta T, T]$ . Due to the convexity of  $Q_k(\tau, \phi)$ , it then suffices to show

$$\sup_{1 \leq k \leq T} \sup_{\|\phi\| = A(NT)^{-1/2}} |Q_k(\tau, \phi)| = O_p(1) \quad \text{for any } A > 0.$$

Apply the decomposition (A.1). For  $W_k(\tau, \phi)$ , we have

$$\sup_{1 \leq k \leq T} \sup_{\|\phi\| = A(NT)^{-1/2}} |W_k(\tau, \phi)| \leq \sup_{1 \leq k \leq T} A \left\| (NT)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \psi_{\tau}(u_{it}^0(\tau)) x'_{it} \right\| = O_p(1),$$

where the last inequality is due to the functional central limit theorem (or alternatively applying the Hájek-Rényi inequality). For  $Z_k(\tau, \phi)$ , apply Lemma A.1,

$$\begin{aligned} 0 \leq Z_k(\tau, \phi) &\leq \left| \sum_{t=1}^k \sum_{i=1}^N E_{t-1} \{b_{it}(\tau, \phi) - b_{it}(\tau, 0)\} \phi \right| \\ &\quad + \left\| A(NT)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \xi_{it}(\tau, \phi) \right\|. \end{aligned} \quad (\text{A.11})$$

The first term on the right hand side is bounded from above by  $U_f \sum_{t=1}^T \sum_{i=1}^N \phi' x_i x_i' \phi$  uniformly in  $\phi$  because of the mean value theorem, which is further bounded by  $A^2 U_f \lambda_{\max}$  because  $\phi = A(NT)^{-1/2}$ . The second term is  $o_p(1)$  by Lemma A.2.1. Thus,  $\sup_{1 \leq k \leq T} \sup_{\|\phi\|=A(NT)^{-1/2}} |Z_k(\tau, \phi)| = O_p(1)$ .

**Proof of A.4.2.** Let  $D_k = B(Nk)^{-1/2}(\log NT)^{1/2}$  with  $B$  an arbitrary constant. Because of the first result of Lemma A.3 and the convexity of  $Q_k(\tau, \phi)$ , it suffices to show

$$\sup_{1 \leq k \leq T} \sup_{\|\phi\|=D_k} \left| (\log NT)^{-1} Q_k(\tau, \phi) \right| = O_p(1) \quad \text{for each } B > 0.$$

Apply the decomposition (A.1). For  $W_k(\tau, \phi)$ ,

$$\begin{aligned} & (\log NT)^{-1} |W_k(\tau, \phi)| \\ & \leq (\log NT)^{-1} \left\| \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_{it}^0(\tau)) x_{it}' \right\| \|\phi\| \\ & \leq B \left\| (Nk)^{-1/2} (\log NT)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_{it}^0(\tau)) x_{it}' \right\| = O_p(1). \end{aligned}$$

For  $Z_k(\tau, \phi)$ , applying the same argument as in Lemma A.4.1. (c.f. (A.11)), we can show

$$(\log NT)^{-1} Z_k(\tau, \phi) = O_p(1).$$

This completes the proof.

**Proof of A.4.3.** Due to convexity, it suffices to consider  $\|\phi\| = (NT)^{-1/2} \log NT$  and show

$$P \left( \inf_{\delta T \leq k \leq T} \inf_{\|\phi\|=(NT)^{-1/2} \log NT} Q_k(\tau, \phi) < D \log NT \right) < \epsilon.$$

We have

$$\begin{aligned} & \inf_{\delta T \leq k \leq T} \inf_{\|\phi\|=(NT)^{-1/2} \log NT} Q_k(\tau, \phi) \\ & \geq \inf_{\delta T \leq k \leq T} \inf_{\|\phi\|=(NT)^{-1/2} \log NT} Z_k(\tau, \phi) - \sup_{\delta T \leq k \leq T} \sup_{\|\phi\|=(NT)^{-1/2} \log NT} |W_k(\tau, \phi)|. \end{aligned}$$

Applying similar arguments as in Lemma A.3, we can show

$$\inf_{\delta T \leq k \leq T} \inf_{\|\phi\|=(NT)^{-1/2} \log NT} (\log NT)^{-2} Z_k(\tau, \phi) \geq \frac{1}{8} \delta L_f \lambda_{\min}$$

in probability for large  $T$ , and

$$\begin{aligned} & \sup_{\delta T \leq k \leq T} \sup_{\|\phi\|=(NT)^{-1/2} \log NT} (\log NT)^{-2} |W_k(\tau, \phi)| \\ & \leq \sup_{\delta T \leq k \leq T} (\log NT)^{-1} \left\| (NT)^{-1/2} \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_{it}^0(\tau)) x_{it}' \right\| = o_p(1). \end{aligned}$$

The result follows by combining the above two results.

**Proof of A.4.4.** It is similar to A.4.3 and is omitted.

**Proof of A.4.5.** Due to convexity, it is sufficient to show

$$P \left( \inf_{Ah_T \leq k \leq T} \inf_{\|\phi\|=d_T(NT)^{-1/2}} Q_k(\tau, \phi) < D \right) < \epsilon.$$

First consider  $k^{-1}W_k(\tau, \phi)$ . Let  $C$  be an arbitrary constant, we have

$$\begin{aligned} & P \left( \sup_{Ah_T \leq k \leq T} \sup_{\|\phi\|=d_T(NT)^{-1/2}} |k^{-1}W_k(\tau, \phi)| > Cd_T^2/T \right) \\ & \leq P \left( \sup_{Ah_T \leq k \leq T} \left\| N^{-1/2}k^{-1} \sum_{t=1}^k \sum_{i=1}^N \psi_\tau(u_t^0(\tau))x'_{it} \right\| > Cd_T/\sqrt{T} \right) \\ & \leq \frac{T}{C^2d_T^2} \left( N^{-1}(Ah_T)^{-2} \sum_{t=1}^{Ah_T} \sum_{i=1}^N E \|x_{it}\|^2 + N^{-1} \sum_{t=Ah_T+1}^T \sum_{i=1}^N t^{-2} E \|x_{it}\|^2 \right) \\ & \leq \frac{TK}{C^2d_T^2} \left( (Ah_T)^{-1} + \sum_{t=Ah_T+1}^T t^{-2} \right) \leq \frac{3K}{C^2} \left( \frac{Ah_T d_T^2}{T} \right)^{-1} \end{aligned}$$

where the second inequality is due to the the Hájek-Rényi inequality, the third inequality is because of  $E \|x_{it}\|^2 < \infty$  by Assumption B5(c), and the last inequality is because of  $\sum_{t=Ah_T+1}^T t^{-2} \leq 2(Ah_T)^{-1}$ . Because  $h_T d_T^2/T \rightarrow h > 0$ , the quantity  $(Ah_T d_T^2/T)^{-1}$ , and thus the preceding display, can be made arbitrarily close to zero by choosing a large  $A$ .

Now, consider  $k^{-1}Z_k(\tau, \phi)$ . Applying the same argument as in the proof of Lemma A.3 (the discussion between the display (A.7) and (A.8)) but using Lemma A.2.3 instead of Lemma A.2.2, we have

$$P \left( \inf_{Ah_T \leq k \leq T} \inf_{\|\phi\|=d_T(NT)^{-1/2}} k^{-1}Z_k(\tau, \phi) < 2Cd_T^2/T \right) < \epsilon$$

for sufficiently large  $T$  and  $A$ . Thus,

$$P \left( \inf_{Ah_T \leq k \leq T} \inf_{\|\phi\|=d_T(NT)^{-1/2}} k^{-1}Q_k(\tau, \phi) < Cd_T^2/T \right) < \epsilon$$

for sufficiently large  $T$  and  $A$ . Because  $k \geq Ah_T$ , this implies  $Q_k(\tau, \phi)$  is greater than  $AC(h_T d_T^2/T)$  with probability arbitrarily close to 1 for sufficiently large  $T$  and  $A$ . However, because  $AC(h_T d_T^2/T) \rightarrow Ach$ , the quantity  $AC(h_T d_T^2/T)$  can be made greater than  $D$  by choosing a large  $A$ .

**Proof of A.4.6** uses the same argument as Bai (1995, p. 432). The detail is omitted.

The next lemma is similar to Lemma 9 in Bai (2000). It shows that when pooling data from two regimes, the estimated parameters are close to those of the dominating regime.

**Lemma A.5** Consider a sample of size  $T = k_1 + k_2$  with a structural change occurring at  $k_1$ :

$$Q_{y_{it}}(\tau|x_{it}) = \begin{cases} x'_{it}\beta_1^0(\tau), & t = 1, \dots, k_1, \\ x'_{it}\beta_2^0(\tau), & t = k_1 + 1, \dots, T. \end{cases}$$

Assume  $k_2 \leq [T^{1/2}v_T^{-1}]$  and suppose Assumptions B1-B3 and B5-B7 hold. Let  $\hat{\beta}(\tau)$  be the quantile regression estimate using the pooled sample ignoring the break and under the additional restriction that  $\|\hat{\beta}(\tau) - \beta_1^0(\tau)\| \leq (Nk_1)^{-1/2} \log(Nk_1)$ . Then,

$$\hat{\beta}(\tau) - \beta_1^0(\tau) = O_p((NT)^{-1/2}). \quad (\text{A.12})$$

**Proof of Lemma A.5.** The proof is based on analyzing the subgradient normalized by  $(NT)^{-1/2}$  :

$$(NT)^{-1/2} \sum_{t=1}^{k_1} \sum_{i=1}^N \{b_{it}(\tau, \hat{\beta}(\tau) - \beta_1^0(\tau))' - \tau x_{it}\} + (NT)^{-1/2} \sum_{t=k_1+1}^T \sum_{i=1}^N \{b_{it}(\tau, \hat{\beta}(\tau) - \beta_2^0(\tau))' - \tau x_{it}\}. \quad (\text{A.13})$$

We will proceed by contradiction, showing that the subgradient will be strictly positive with probability 1 if the condition (A.12) is violated.

Suppose  $\|\hat{\beta}(\tau) - \beta_1^0(\tau)\| > A(NT)^{-1/2}$  and study the two terms separately. Following the same argument as in Lemma 2 in Qu (2008, p.182), the first term in (A.13) can be made arbitrary large by choosing a large  $A$ . For the second term, rewrite it as

$$\begin{aligned} & (NT)^{-1/2} \sum_{t=k_1+1}^T \sum_{i=1}^N \left\{ b_{it}(\tau, \hat{\beta}(\tau) - \beta_2^0(\tau))' - b_{it}(\tau, \beta_1^0(\tau) - \beta_2^0(\tau))' \right\} \quad (\text{c}) \\ & + (NT)^{-1/2} \sum_{t=k_1+1}^T \sum_{i=1}^N \xi_{it}(\tau, \beta_1^0(\tau) - \beta_2^0(\tau))' \quad (\text{d}) \\ & + (NT)^{-1/2} \sum_{t=k_1+1}^T \sum_{i=1}^N x_{it} \{F_{it}(x_{it}'\beta_1^0(\tau)) - \tau\} \quad (\text{e}) \\ & + (NT)^{-1/2} \sum_{t=k_1+1}^T \sum_{i=1}^N x_{it} \{1(u_{it}^0(\tau) \leq 0) - \tau\} \quad (\text{f}). \end{aligned}$$

For (c), notice that the restriction  $\|\hat{\beta}(\tau) - \beta_1^0(\tau)\| \leq (Nk_1)^{-1/2} \log(Nk_1)$  and Assumption B6 imply  $\hat{\beta}(\tau) - \beta_1^0(\tau) = o_p(N^{-1/2}v_T)$ . This implies (c) =  $o_p(1)$  using the same argument as in the proof of Lemma A2. Similarly, term (d) =  $O_p(1)$ . Term (e) =  $O_p(1)$  after applying the mean value theorem and the strong law of large numbers. Term (f) =  $o_p(1)$  because of the functional central limit theorem and  $k_2/T \rightarrow 0$ . The above results all hold uniformly in  $0 \leq k_2 \leq T^{1/2}v_T^{-1}$ . Thus, by choosing a large  $A$ , the first term in (A.13) will dominate the second term. Consequently, the subgradient will be strictly positive with probability 1. ■

**Lemma A.6** Suppose there is no structural change. Under Assumptions B1-B3 and B5-B7, for any  $A < \infty, B < \infty$ , and  $\Delta \in R^p$ , we have

$$\sup_{k \leq Av_T^{-2}} \sup_{\|\phi\| \leq (NT)^{-1/2}B} \left| Q_k(\tau, N^{-1/2}v_T\Delta + \phi) - Q_k(\tau, N^{-1/2}v_T\Delta) \right| = o_p(1).$$

The proof is similar to that for Lemma A.4.1. and is omitted.

**Proof of Lemma 2.** For the ease of exposition, assume there are only two breaks, occurring at  $T_1^0$  and  $T_2^0$ . Let  $\hat{T}^b = (\hat{T}_1, \hat{T}_2)$  be their estimates,  $\hat{\beta}_j(\tau)$  be the coefficients estimates corresponding to the  $j$ th segment (i.e.,  $[\hat{T}_{j-1} + 1, \hat{T}_j]$ ), and  $\hat{\beta}(\tau) = (\hat{\beta}_1(\tau)', \hat{\beta}_2(\tau)', \hat{\beta}_3(\tau)')'$ . It consists of four steps that successively tighten the bounds on  $\hat{\beta}(\tau)$  and  $\hat{T}^b$ .

**Step 1.** (Prove  $P(|\hat{T}_j - T_j^0| \leq T^{1/2}v_T^{-1}) \rightarrow 1$  for  $j = 1, 2$  as  $T \rightarrow \infty$ .) The proof is by contradiction. Suppose the result does not hold and let  $K = \lceil T^{1/2}v_T^{-1} \rceil$ . Then, it suffices to consider the following three cases:

- (1)  $|\hat{T}_2 - T_2^0| > K$  and  $|\hat{T}_1 - T_2^0| > K$ ,
- (2)  $|\hat{T}_2 - T_2^0| > K$  and  $|\hat{T}_1 - T_2^0| \leq K$ ,
- (3)  $|\hat{T}_2 - T_2^0| \leq K$  and  $|\hat{T}_1 - T_1^0| > K$ .

First consider Case (1). Let  $T^{b^*}$  be the ordered version of  $\{\hat{T}_1, \hat{T}_2, T_1^0, T_2^0 - K, T_2^0 + K\}$ . Values in  $T^{b^*}$  partition the full sample into (at most) six segments, with the  $l$ th segment containing the portion of the sample that falls between the  $(l-1)$ th and  $l$ th largest values in  $T^{b^*}$ . For any such partition, we always have

$$\{S_{NT}(\tau, \hat{\beta}(\tau), \hat{T}^b) - S_{NT}(\tau, \beta^0(\tau), T^0)\} \geq \inf_{\beta^*(\tau)} \{S_{NT}(\tau, \beta^*(\tau), T^{b^*}) - S_{NT}(\tau, \beta^0(\tau), T^0)\}. \quad (\text{A.14})$$

Therefore, to reach a contradiction, it suffices to show that the right hand side is strictly positive.

Suppose the subsample  $[T_2^0 - K, T_2^0 + K]$  corresponds to the  $l$ th segment in the partition with coefficient estimates being  $\hat{\beta}_l^*(\tau)$ . It includes observations from both regimes 2 and 3. Note that

$$\begin{aligned} & \max \left\{ N^{1/2}K^{1/2} \left\| \hat{\beta}_l^*(\tau) - \beta_2^0(\tau) \right\|, N^{1/2}K^{1/2} \left\| \hat{\beta}_l^*(\tau) - \beta_3^0(\tau) \right\| \right\} \\ & \geq N^{1/2}K^{1/2} \|\beta_2^0(\tau) - \beta_3^0(\tau)\| / 2 \geq T^{v/2} \|\Delta_2(\tau)\| / 4 \geq \log(NK) \end{aligned}$$

for large  $T$ , where the second inequality is due to Assumption B7 and  $K = \lceil T^{1/2}v_T^{-1} \rceil$ , and the last inequality holds because Assumption B6 implies  $\log(NK)/T^{v/2} \rightarrow 0$  as  $T \rightarrow \infty$ . Therefore, without loss of generality, we can assume  $(NK)^{1/2} \|\hat{\beta}_l^*(\tau) - \beta_2^0(\tau)\| \geq \log(NK)$ . By Lemma A.4.3 (applied with  $T$  replaced by  $K$ ), the contribution of the sub-segment  $[T_2^0 - K, T_2^0]$  to the right side of (A.14) is greater than  $D \log(NK)$  with probability approaching 1. The contribution of the sub-segment  $[T_2^0 + 1, T_2^0 + K]$  is either nonnegative, or of order  $O_p(\log NK)$  when it is negative by Lemma A.4.2. Other segments are of order  $O_p(\log(NT))$  by Lemma A.4.2. By choosing  $D$  large enough, the term  $D \log(NK)$  dominates the rest and thus (A.14) is positive with probability approaching 1 as  $T \rightarrow \infty$ . Thus, we have reached a contradiction.

Next consider Case (2). By Assumption B7,  $K/T \rightarrow 0$  and therefore  $|\hat{T}_1 - T_1^0| > K$  for large  $T$ . Let  $T^{b^*}$  be the ordered version of  $\{T_1^0 - K, T_1^0 + K, T_2^0, \hat{T}_1, \hat{T}_2\}$ . Then,  $[T_1^0 - K, T_1^0 + K]$ , which forms a single segment in the partition, contains observations from both regimes 1 and 2. Repeat the same argument as in Case 1, starting at (A.14) but with  $T_2^0, \beta_2^0(\tau)$  and  $\beta_3^0(\tau)$  replaced by  $T_1^0, \beta_1^0(\tau)$  and  $\beta_2^0(\tau)$ . We can reach the same contradiction.

The analysis of Case (3) proceeds in the same way as Case (2), by considering the ordered version of  $\{\hat{T}_1, T_1^0 - K, T_1^0 + K, T_2^0, \hat{T}_2\}$  and noticing that  $[T_1^0 - K, T_1^0 + K]$  again forms a single segment.

**Step 2.** (Prove  $P\left(\|\hat{\beta}_j(\tau) - \beta_j^0(\tau)\| \leq (NT)^{-1/2} \log NT\right) \rightarrow 1$  for  $j = 1, 2$  and 3.) Suppose  $\hat{\beta}_2(\tau)$  does not satisfy this condition. Then  $|\hat{\beta}_2(\tau) - \beta_2^0(\tau)| > (NT)^{-1/2} \log NT$  with positive probability for any  $T$ . Consider a subset of the second segment, with boundary points  $\hat{T}_{1,1} = \max(\hat{T}_1, T_1^0)$  and  $\hat{T}_{1,2} = \min(\hat{T}_2, T_2^0)$ . Consider a new partition of the sample using the ordered version of  $\{\hat{T}_1, T_1^0, \hat{T}_2, T_2^0, \hat{T}_{1,1}, \hat{T}_{1,2}\}$ . Then, Step 1 implies that the segment  $[\hat{T}_{1,1}, \hat{T}_{1,2}]$  contains a positive fraction of the sample. Its contribution to (A.14) is positive and greater than  $D \log(NT)$  by Lemma A.4.3. Contributions from other segments are of order  $O_p(\log(NT))$  by Lemma A.4.2. Thus, the objective function (A.14) will be positive with probability approaching 1 as  $T \rightarrow \infty$ , and this contradicts the fact that  $\hat{\beta}(\tau)$  is its minimizer.

**Step 3.** (Prove  $\hat{\beta}_j(\tau) - \beta_j^0(\tau) = O_p((NT)^{-1/2})$  for  $j = 1, 2$  and 3.) The results from Steps 1 and 2 imply that we can restrict our attention to the following set:

$$\Phi = \left\{ \left\| \hat{\beta}_j(\tau) - \beta_j^0(\tau) \right\| \leq (NT)^{-1/2} \log NT \text{ (} j = 1, 2, 3 \text{) and } |\hat{T}_i - T_i^0| \leq T^{1/2} v_T^{-1} \text{ (} i = 1, 2 \text{)} \right\}.$$

Consider a partition of the sample using break dates  $\{\hat{T}_1, \hat{T}_2\}$ . Then, all segments are non-vanishing fragments of the sample. Consider the first segment. Then, it either: 1) only contains observations from the first regime, or 2) contains observations from both regimes but with less than  $T^{1/2} v_T^{-1}$  observations from the second regime. In the first case, apply the second result of Lemma A.3, and for the second case, apply Lemma A.5, leading to  $\hat{\beta}_1(\tau) - \beta_1^0(\tau) = O_p((NT)^{-1/2})$ . The parameter estimates corresponding to other segments can be analyzed similarly.

**Step 4.** (Prove  $v_T^2(\hat{T}_j - T_j^0) = O_p(1)$ .) From Step 3, it suffices to consider  $(\beta(\tau), T^b) \in \Phi$ . In what follows, we shall prove that for any  $\epsilon > 0$  and  $D > 0$ , there exists an  $A > 0$  such that

$$P \left( \inf_{(\beta(\tau), T^b) \in \Phi, |T_1 - T_1^0| > A v_T^{-2}} S_{NT}(\tau, \beta(\tau), T^b) - \inf_{(\beta(\tau), T^b) \in \Phi, T_1 = T_1^0} S_{NT}(\tau, \beta(\tau), T^b) < D \right) < \epsilon \quad (\text{A.15})$$

and

$$P \left( \inf_{(\beta(\tau), T^b) \in \Phi, |T_2 - T_2^0| > A v_T^{-2}} S_{NT}(\tau, \beta(\tau), T^b) - \inf_{(\beta(\tau), T^b) \in \Phi, T_2 = T_2^0} S_{NT}(\tau, \beta(\tau), T^b) < D \right) < \epsilon \quad (\text{A.16})$$

hold for all sufficiently large  $T$ . Because

$$\inf_{(\beta(\tau), T^b) \in \Phi, T_j = T_j^0} S_{NT}(\tau, \beta(\tau), T^b) \geq \inf_{(\beta(\tau), T^b) \in \Phi} S_{NT}(\tau, \beta(\tau), T^b) \text{ for } j = 1, 2,$$

(A.15) and (A.16) imply  $P\left(v_T^2 |\hat{T}_j - T_j^0| > A\right) < \epsilon$  for  $j = 1$  and 2 with large  $T$ .

We only prove (A.16) as the proof for (A.15) is similar. Without loss of generality, assume  $\hat{T}_2 > T_2^0$ . Let  $S_{NT}(\tau, \hat{\beta}(\tau), \hat{T}^b)$  and  $S_{NT}(\tau, \hat{\beta}^*(\tau), \hat{T}^{b*})$  denote the minimized values of the two

terms inside the parentheses, respectively. Then,

$$\begin{aligned}
& S_{NT}(\tau, \hat{\beta}(\tau), \hat{T}^b) - S_{NT}(\tau, \hat{\beta}^*(\tau), \hat{T}^{b*}) \\
& \geq S_{NT}(\tau, \hat{\beta}(\tau), \hat{T}^b) - S_{NT}(\tau, \hat{\beta}(\tau), \hat{T}^{b*}) \\
& = \sum_{t=T_2^0+1}^{\hat{T}_2} \sum_{i=1}^N q_{\tau, it}(\hat{\beta}_2(\tau) - \beta_3^0(\tau)) - \sum_{t=T_2^0+1}^{\hat{T}_2} \sum_{i=1}^N q_{\tau, it}(\hat{\beta}_3(\tau) - \beta_3^0(\tau)).
\end{aligned}$$

From Step 3,  $\|\hat{\beta}_j(\tau) - \beta_j^0(\tau)\| = O_p((NT)^{-1/2})$ , implying

$$\|\hat{\beta}_2(\tau) - \beta_3^0(\tau)\| \geq N^{-1/2} v_T \|\Delta_2(\tau)\| / 2.$$

By Lemma A.4.5 and (A.10), the above difference is greater than  $D$  with probability greater than  $1 - \epsilon$  when  $A$  is large. Note that Lemma A.4.5 is applied with  $h_T = v_T^{-2}$  and  $d_T = T^{1/2} v_T$ . Therefore (A.16) holds. ■

**Proof of Theorem 2.** By Lemma 2, we can restrict our attention to the set  $\mathcal{K} \times \Theta$ , where

$$\begin{aligned}
\mathcal{K} &= \{T_j : T_j = T_j^0 + [sv_T^{-2}] \text{ and } |s| \leq A < \infty, j = 1, \dots, m\}, \\
\Theta &= \{\beta_j : \sqrt{NT} \|\beta_j - \beta_j^0(\tau)\| \leq M, j = 1, \dots, m+1\}.
\end{aligned}$$

Adding and subtracting terms,

$$\begin{aligned}
& \inf_{T^b \in \mathcal{K}} \inf_{\beta(\tau) \in \Theta} S_{NT}(\tau, \beta(\tau), T^b) \tag{A.17} \\
& = \inf_{T^b \in \mathcal{K}} \inf_{\beta(\tau) \in \Theta} \left\{ S_{NT}(\tau, \beta(\tau), T^0) + \left( S_{NT}(\tau, \beta(\tau), T^b) - S_{NT}(\tau, \beta(\tau), T^0) \right) \right\}.
\end{aligned}$$

First, assume  $\hat{T}_j < T_j^0$  for all  $j = 1, \dots, m$ . The second term inside the curly brackets is equal to

$$\sum_{j=1}^m \sum_{t=\hat{T}_j+1}^{T_j^0} \sum_{i=1}^N \{ \rho_\tau(y_{it} - x'_{it} \beta_{j+1}^0(\tau)) - \rho_\tau(y_{it} - x'_{it} \beta_j^0(\tau)) \} + o_p(1)$$

uniformly in  $T^b \in \mathcal{K}$  and  $\beta(\tau) \in \Theta$  by Lemma A.6. Thus, minimizing (A.17) is asymptotically equivalent to solving

$$\inf_{\beta(\tau) \in \Theta} \{ S_T(\tau, \beta(\tau), T^0) \} + \inf_{T^b \in \mathcal{K}} \sum_{j=1}^m \sum_{t=T_j+1}^{T_j^0} \sum_{i=1}^N \{ \rho_\tau(y_{it} - x'_{it} \beta_{j+1}^0(\tau)) - \rho_\tau(y_{it} - x'_{it} \beta_j^0(\tau)) \}.$$

The first term depends only on  $\beta(\tau)$  but not on  $T^b$ , which delivers the asymptotic distribution of  $\hat{\beta}(\tau)$  as stated in the theorem. The second term only depends on  $T^b$  but not on  $\beta(\tau)$ , which delivers the limiting distribution for the break date estimate. Consider the  $j$ th break and rewrite



the summation involving  $\hat{T}_j$  as  $H_{j,2}(s) - H_{j,1}(s)$ , where

$$H_{j,1}(s) = (\beta_{j+1}^0(\tau) - \beta_j^0(\tau))' \sum_{t=T_j^0 + [sv_T^{-2}] + 1}^{T_j^0} \sum_{i=1}^N x_{it} \psi_\tau(u_{it}^0(\tau)),$$

$$H_{j,2}(s) = \sum_{t=T_j^0 + [sv_T^{-2}] + 1}^{T_j^0} \sum_{i=1}^N \int_0^{x'_{it}(\beta_{j+1}^0(\tau) - \beta_j^0(\tau))} \{1(u_{it}^0(\tau) \leq u) - 1(u_{it}^0(\tau) \leq 0)\} du.$$

First consider  $H_{j,1}(s)$ . If  $N$  is fixed, then we can apply a FCLT for martingale differences. If  $(N, T) \rightarrow \infty$ , we can apply a FCLT for random fields (e.g., Theorem 3 in Poghosyan and Roelly, 1998). In both cases,  $H_{j,1}(s) \Rightarrow \bar{\sigma}_j W(s)$ , where  $\sigma_j^2 = \tau(1 - \tau)\Delta_j(\tau)' \bar{J}_j^0 \Delta_j(\tau)$  and  $W(s)$  is a two-sided Wiener process satisfying  $W(0) = 0$ . Consider  $H_{j,2}(s)$ , its mean, for a given  $s$ , is equal to

$$\frac{1}{2} \Delta_j(\tau)' \bar{H}_j^0(\tau) \Delta_j(\tau) |s| + o_p(1) = \frac{\bar{\pi}_j}{2} |s| + o_p(1),$$

and the deviation from the mean is uniformly small. Similar arguments can be applied to analyze the case  $\hat{T}_j > T_j^0$ , leading to

$$v_T^2(\hat{T}_j - T_j^0) \Rightarrow \arg \max_s \begin{cases} \bar{\sigma}_j W(s) - \bar{\pi}_j |s|/2 & s \leq 0 \\ \bar{\sigma}_{j+1} W(s) - \bar{\pi}_{j+1} |s|/2 & s > 0 \end{cases},$$

Then, by a change of variables,

$$\left(\frac{\bar{\pi}_j}{\bar{\sigma}_j}\right)^2 v_T^2(\hat{T}_j - T_j^0) \Rightarrow \arg \max_s \begin{cases} W(s) - |s|/2 & s \leq 0 \\ (\bar{\sigma}_{j+1}/\bar{\sigma}_j) W(s) - (\bar{\pi}_{j+1}/\bar{\pi}_j) |s|/2 & 0 < s \end{cases}.$$

This completes the proof.  $\blacksquare$

Table 1. Estimated response surface regression

Test	Size (%)	Level regressors ( $x_1$ )					Exponentiated regressors ( $x_2$ )			
		1	$p$	$l+1$	$\frac{1}{p}$	$(l+1)p$	$(l+1)\omega$	$\frac{1}{l+1}$	$\frac{1}{(l+1)\omega}$	$\omega$
SQ( $l+1 l$ )	10	1.5432	0.0265	0.0388	-0.2371	-0.0021		-0.1168		
	5	1.6523	0.0242	0.0369	-0.2226	-0.0019		-0.0999		
	1	1.8703	0.0215	0.0331	-0.1742	-0.0017		-0.0770		
DQ( $l+1 l$ )	10	0.9481	0.0062	0.0166	-0.1386	-0.0004	0.0018	-0.0801	-0.0004	-0.0254
	5	0.9944	0.0058	0.0157	-0.1284	-0.0004	0.0017	-0.0716	-0.0005	-0.0203
	1	1.0929	0.0050	0.0134	-0.1134	-0.0002	0.0010	-0.0565	0.0000	-0.0062

*Note.*  $p$  denotes the number of parameters allowed to change,  $l$  is the number of breaks under the null hypothesis, and  $\omega$  is a trimming parameter determining the interval of quantiles being tested:  $[\omega, 1-\omega]$ .

Table 2. Coverage rates for the break date

(N, T)	$\sqrt{N}$ times Break Size	Quantile							
		0.2	0.3	0.4	0.5	0.6	0.7	0.8	All
(1, 100)	1.0	0.876	0.901	0.917	0.920	0.911	0.902	0.866	0.856
	2.0	0.895	0.930	0.933	0.934	0.923	0.916	0.882	0.908
	3.0	0.934	0.951	0.964	0.970	0.969	0.959	0.938	0.933
(50, 100)	1.0	0.897	0.904	0.900	0.899	0.901	0.904	0.903	0.869
	2.0	0.905	0.902	0.915	0.908	0.908	0.902	0.911	0.912
	3.0	0.926	0.944	0.943	0.948	0.940	0.934	0.941	0.953
(100, 100)	1.0	0.892	0.900	0.890	0.890	0.888	0.893	0.889	0.857
	2.0	0.892	0.897	0.905	0.900	0.910	0.901	0.892	0.904
	3.0	0.929	0.928	0.937	0.949	0.947	0.946	0.932	0.950

*Note.* The nominal size is 95%. Columns indicated by 0.2-0.8 include results based on a single quantile function. In the last column, the break date is estimated using all seven quantiles.

Table 3. Coverage rates for the break size parameter

(N, T)		$\sqrt{N}$ times		Quantile					
		Break Size	0.2	0.3	0.4	0.5	0.6	0.7	0.8
(1, 100)	Single	1.0	0.815	0.836	0.862	0.864	0.853	0.848	0.821
	Quantile	2.0	0.928	0.943	0.944	0.950	0.937	0.935	0.921
		3.0	0.926	0.930	0.933	0.941	0.937	0.939	0.933
		Multiple	1.0	0.879	0.879	0.888	0.880	0.884	0.888
	Quantiles	2.0	0.939	0.947	0.946	0.954	0.941	0.936	0.926
		3.0	0.936	0.932	0.932	0.942	0.940	0.942	0.942
(50, 100)	Single	1.0	0.794	0.829	0.846	0.834	0.842	0.826	0.797
	Quantile	2.0	0.948	0.946	0.951	0.952	0.948	0.948	0.935
		3.0	0.960	0.957	0.957	0.954	0.950	0.951	0.951
		Multiple	1.0	0.902	0.887	0.887	0.884	0.881	0.888
	Quantiles	2.0	0.956	0.952	0.958	0.952	0.949	0.951	0.945
		3.0	0.953	0.955	0.955	0.953	0.944	0.949	0.950
(100, 100)	Single	1.0	0.799	0.841	0.845	0.848	0.833	0.820	0.805
	Quantile	2.0	0.935	0.946	0.942	0.947	0.940	0.949	0.940
		3.0	0.950	0.961	0.959	0.956	0.952	0.955	0.960
		Multiple	1.0	0.903	0.898	0.887	0.891	0.878	0.887
	Quantiles	2.0	0.939	0.949	0.947	0.947	0.941	0.951	0.953
		3.0	0.948	0.954	0.957	0.953	0.951	0.948	0.955

*Note.* The nominal coverage rate is 95%. "Single quantile": the break date is estimated based on one quantile function. "Multiple quantiles": the break date is estimated using all seven quantiles. Conditioning on the estimated break date, no other restrictions are imposed across quantiles.

Table 4. Comparisons of different break estimators

Error ( $u_{it}$ )	(N, T)	$\sqrt{N}$ times Break Size	Median		Multiple Quantiles		Mean	
			MAD	IQR90	MAD	IQR90	MAD	IQR90
N(0,1)	(1, 100)	1.0	13.37	63	12.50	61	13.93	64
		2.0	5.55	31	4.93	27	5.69	33
		3.0	2.12	10	1.90	10	2.26	13
	(50, 100)	1.0	14.29	64	12.53	63	14.31	66
		2.0	6.69	39	4.93	27	6.36	36
		3.0	2.88	15	2.14	11	2.64	15
	(100, 100)	1.0	14.74	66	13.32	64	14.6	65
		2.0	6.87	39	5.09	28	6.44	35
		3.0	2.85	15	2.00	11	2.85	16
t(2.5)	(1, 100)	1.0	14.91	65	14.89	66	17.88	68
		2.0	7.43	42	7.43	43	11.48	61
		3.0	3.58	20	3.69	19	6.97	43
	(50, 100)	1.0	15.37	66	15.02	65	19.03	68
		2.0	7.88	46	6.95	39	14.25	65
		3.0	3.46	19	3.16	17	9.82	55
	(100, 100)	1.0	15.16	65	14.81	65	19.34	68
		2.0	7.68	43	7.11	39	13.94	65
		3.0	3.68	20	3.22	17	10.06	57

*Note.* MAD: Mean Absolute Deviation. IQR90: the distance between 90% and 10% quantiles of the empirical distribution. "Multiple quantiles": the estimation is based jointly on quantiles 0.2, 0.3,..., 0.8.

Table 5. Structural breaks in the U.S. real GDP growth rate

Panel (a). Joint analysis of multiple quantiles					
DQ(1 0)	0.994*				
DQ(2 1)	0.612				
Break Date	84:1				
95% C. I.	[77:3,84:2]				
Panel (b). Separate treatment of individual quantiles					
Quantile	0.20	0.35	0.50	0.65	0.80
SQ(1 0)	1.423	1.310	1.019	1.818**	2.078**
SQ(2 1)	-	-	-	0.964	1.116
Break Date	-	-	-	84:2	84:1
95% C. I.	-	-	-	[68:2,87:4]	[78:4,90:1]

*Note.* The sample period is 1947:2 to 2009:2. The model is a quantile autoregressive model with all parameters allowed to change. C.I. denotes the 95% confidence interval. \* and \*\* indicate statistical significance at 5% and 1% level, respectively.

Table 5B. Structural breaks in the U.S. real GDP growth rate (nine quantiles)

Panel (a). Joint analysis of multiple quantiles									
DQ(1 0)	0.994*								
DQ(2 1)	0.608								
Break Date	84:2								
95% C. I.	[80:1, 86:4]								
Panel (b). Separate treatment of individual quantiles									
Quantile	0.20	0.275	0.35	0.425	0.50	0.575	0.65	0.725	0.80
SQ(1 0)	1.423	1.409	1.310	1.006	1.019	1.097	1.818**	2.110**	2.078**
SQ(2 1)	-	-	-	-	-	-	0.964	1.165	1.116
Break Date	-	-	-	-	-	-	84:2	84:2	84:1
95% C. I.	-	-	-	-	-	-	[68:2,87:4]	[77.4,86.4]	[78:4,90:1]

*Note.* See Table 5.

Table 6. Coefficient estimates for the U.S. real GDP growth rate

Quantile	0.20	0.35	0.50	0.65	0.80
Break date	NA	NA	NA	84:1	84:1
$\mu_1(\tau)$	-0.928 (0.586)	0.697 (0.366)	1.938** (0.334)	4.507** (0.665)	6.129** (0.493)
$\alpha_{1,1}(\tau)$	0.288** (0.075)	0.282** (0.052)	0.335** (0.054)	0.411** (0.105)	0.374** (0.074)
$\alpha_{2,1}(\tau)$	0.236** (0.078)	0.159** (0.023)	0.049 (0.035)	-0.105 (0.104)	-0.091 (0.073)
$\mu_2(\tau) - \mu_1(\tau)$	-	-	-	-2.431** (0.739)	-3.089** (0.772)
$\alpha_{1,2}(\tau) - \alpha_{1,1}(\tau)$	-	-	-	-0.250 (0.148)	-0.211 (0.116)
$\alpha_{2,2}(\tau) - \alpha_{2,1}(\tau)$	-	-	-	0.425** (0.154)	0.405** (0.142)

*Note.* The sample period is 1947:2 to 2009:2. The model is

$$Q_{y_t}(\tau|y_{t-1}, y_{t-2}) = \begin{cases} \mu_1(\tau) + \alpha_{1,1}(\tau)y_{t-1} + \alpha_{2,1}(\tau)y_{t-2}, & \text{if } t \leq T_1 \\ \mu_2(\tau) + \alpha_{1,2}(\tau)y_{t-1} + \alpha_{2,2}(\tau)y_{t-2}, & \text{if } t > T_1 \end{cases}$$

Standard errors are in parentheses. \* and \*\* denote statistical significance at 5% and 1% level, respectively.

Table 7. Results for the U.S. real GDP growth rate based on a subsample

Quantile	0.20	0.35	0.50	0.65	0.80
Number of breaks	1	0	0	1	1
Break date	58:1	NA	NA	84:2	84:2
95% C.I.	[57:4,61:3]	NA	NA	[72:3,87:2]	[83:3,86:1]
$\mu_1(\tau)$	-3.616 (1.928)	0.954* (0.363)	2.213** (0.336)	4.517** (0.665)	6.129** (0.491)
$\alpha_{1,1}(\tau)$	0.879** (0.259)	0.242** (0.042)	0.283** (0.055)	0.420** (0.105)	0.374** (0.074)
$\alpha_{2,1}(\tau)$	-0.191 (0.274)	0.153** (0.024)	0.033 (0.039)	-0.105 (0.104)	-0.091 (0.073)
$\mu_2(\tau) - \mu_1(\tau)$	2.927 (1.992)	-	-	-1.305** (0.765)	-2.378** (0.537)
$\alpha_{1,2}(\tau) - \alpha_{1,1}(\tau)$	-0.606* (0.271)	-	-	-0.453 (0.141)	-0.441** (0.130)
$\alpha_{2,2}(\tau) - \alpha_{2,1}(\tau)$	0.424 (0.282)	-	-	0.287* (0.141)	0.427** (0.132)

*Note.* See Table 6. The sample period is 1947:2 to 2007:4.

Table 8. Structural breaks in the blood alcohol concentration

	Single Quantile			Multiple Quantiles
	0.70	0.75	0.80	
Tests ( $H_0$ vs $H_1$ )				
0 vs 1	5.176**	4.613**	3.503**	3.258**
1 vs 2	2.085**	2.036**	1.842*	1.048**
2 vs 3	1.339	1.259	1.303	0.616
Number of breaks	2	2	2	2
1st Break Date	85:2	85:2	85:2	85:1
95% C.I.	[84:2, 86:3]	[84:3, 86:2]	[84:1, 86:2]	[84:1, 86:1]
2nd Break Date	92:2	93:2	93:2	92:2
95% C.I.	[91:3, 92:3]	[92:1, 95:3]	[91:4, 95:4]	[91:1, 94:4] [91:2, 92:3]

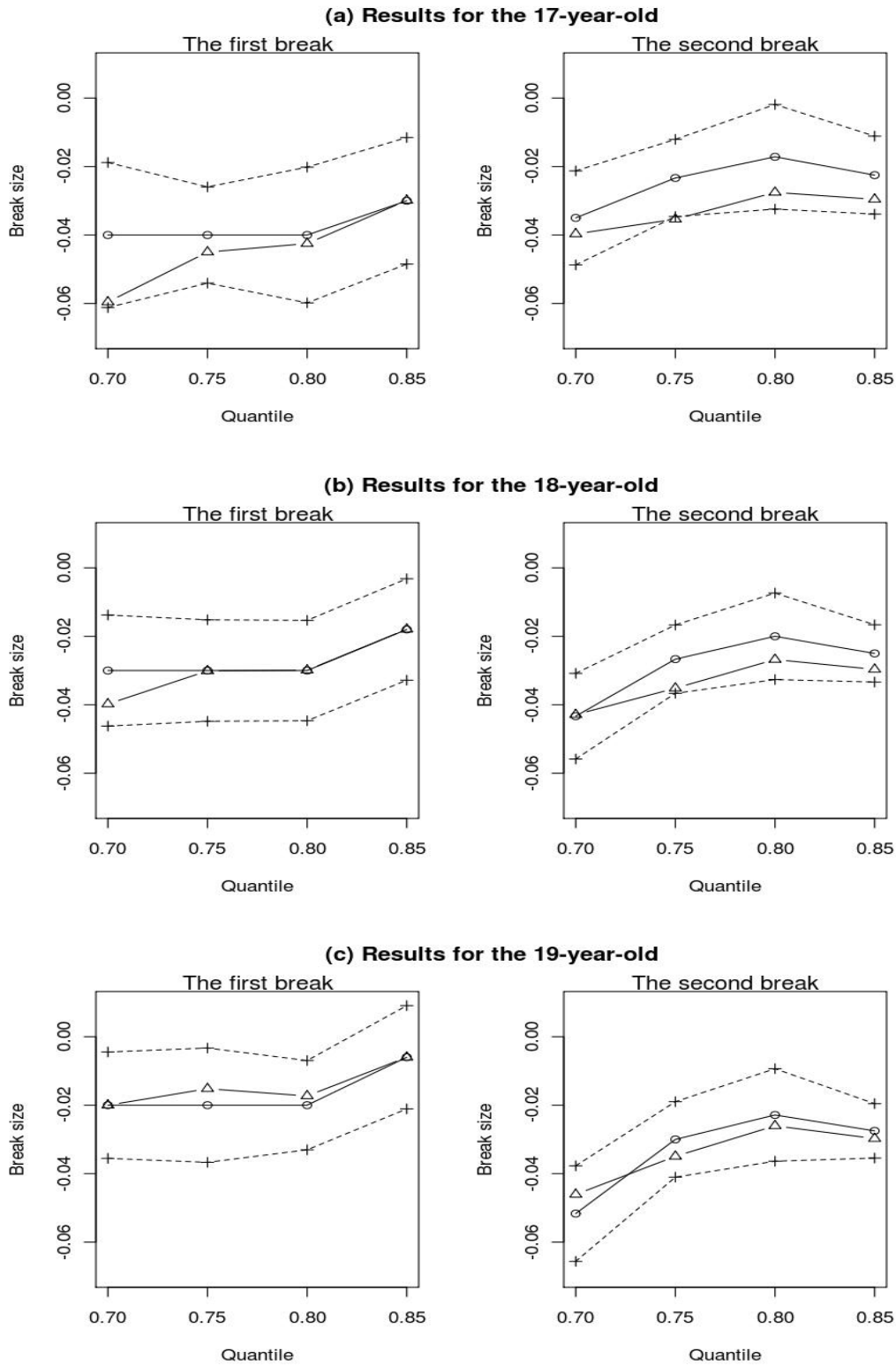
*Note.* The columns under “Single Quantile” report the  $SQ(l+1/l)$  test and the estimated break dates based on individual quantiles. The last column reports the  $DQ(l+1/l)$  test over the interval [0.70, 0.85] and break dates estimates based on four quantiles: 0.70, 0.75, 0.80 and 0.85. \* and \*\* indicate statistical significance at 5% and 1% level, respectively.

Table 8B. Structural breaks in the blood alcohol concentration (seven quantiles)

	Single Quantile							Multiple Quantiles
	0.70	0.725	0.75	0.775	0.80	0.825	0.85	
Tests ( $H_0$ vs $H_1$ )								
0 vs 1	5.176**	4.999**	4.613**	4.309**	3.503**	3.610**	3.258**	2.372**
1 vs 2	2.085**	2.088**	2.036**	1.958**	1.842*	1.230	1.027	1.048**
2 vs 3	1.339	1.318	1.259	1.314	1.303	-	-	0.616
Number of breaks	2	2	2	2	2	1	1	2
1st Break Date	85:2	85:2	85:2	85:2	85:2	-	-	85:1
95% C.I.	[84:2, 86:3]	[84:3, 86:2]	[84:3, 86:2]	[84:2, 86:3]	[84:1, 86:2]	-	-	[84:1, 86:1]
2nd Break Date	92:2	93:2	93:2	93:2	93:2	92:2	92:2	92:2
95% C.I.	[91:3, 92:3]	[92:1, 95:1]	[92:1, 95:3]	[91:2, 96:1]	[91:4, 95:4]	[91:1, 94:1]	[91:1, 94:4]	[91:2, 92:3]

*Note.* See Table 8. The last column reports the  $DQ(l+1/l)$  test over [0.70, 0.85] and break dates estimates based on seven quantiles.

Figure 1. Structural changes in young drivers' blood alcohol concentration (male)



*Note.*  $-\circ-$ : change in  $Q_{y_{it}}(\tau|x_{it})$ .  $\cdots + \cdots$ : pointwise 95% confidence interval.  $-\Delta-$ : results from censored quantile regression. To understand the magnitude of the changes, we note that the relevant unconditional quantiles of  $y_{it}$  are 0.13, 0.14, 0.16 and 0.18 for the first regime (1983.1 to 1985.1), 0.10, 0.12, 0.14 and 0.17 for the second regime (1985.2 to 1992.2), and 0.04, 0.08, 0.11 and 0.15 for the third regime (1992.3 to 2007.4).



Figure 2. Structural changes in young drivers' blood alcohol concentration (female)

