Quantile Regression Estimation of Panel Duration Models with Censored Data*

Matthew Harding† and Carlos Lamarche‡

August 2, 2012

Abstract

This paper studies the estimation of quantile regression panel duration models. We allow for the possibility of endogenous covariates and correlated individual effects in the quantile regression models. We propose a quantile regression approach for panel duration models under conditionally independent censoring. The procedure involves minimizing ℓ_1 convex objective functions and is motivated by a martingale property associated with survival data in models with endogenous covariates. We carry out a series of Monte Carlo simulations to investigate the small sample performance of the proposed approach in comparison with other existing methods. An empirical application of the method to the analysis of the effect of unemployment insurance on unemployment duration illustrates the approach.

JEL: C23, C33

Keywords: Quantile Regression; Duration Models; Panel Data; Unemployment Insurance

*We are grateful to Badi Baltagi, Jerry Hausman, Roger Koenker, Shakeeb Khan, Antonio Galvao, and participants at the 11th Advances in Econometrics Conference for insightful conversations and comments.

†Corresponding author: Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: `mch@stanford.edu`

‡Department of Economics, University of Oklahoma, 729 Elm Avenue, Norman, OK 73019; Phone: (405) 325-5857; Email: `lamarche@ou.edu`

1. Introduction

This paper focuses on econometric methods for longitudinal, or panel data durations consisting of repeated measurements on the same subject, which are commonly encountered in applied economics research. Han and Hausman's (1990) work on duration and competing risk models introduced flexible approaches to duration analysis, while also emphasizing the importance of controlling for unobserved heterogeneity. The recent econometric literature has built on these two fundamental ideas. On the one hand, there has been a growing literature on flexible approaches including semiparametric duration and censored models (e.g., Chen and Khan 2001, Khan and Tamer 2007, Khan and Tamer 2009, Horowitz and Lee 2004, Lee 2008, Honoré and de Paula 2010, Wang and Fygenson 2009, among others). Endogenous treatments in a hazard model were investigated in Eberwein, Ham and LaLonde (1997) and Abbring and van der Berg (2003). At the same time new approaches to controlling for unobserved heterogeneity in more general settings are being explored (Hausman and Woutersen 2005, Burda, Harding and Hausman 2012).

However, existing quantile regression approaches have not been developed for survival data when the covariates are endogenous and the latent unobserved heterogeneity is associated with the independent variables. For example, this situation arises in studies of firms' entry and exit of a market (e.g., Audretsch 1991, Geroski 1995, Audretsch and Mahmood 1995, De Silva, Kosmopoulou and Lamarche 2009). These studies include a small number of firms in business for a relatively long period of time. In these papers, it is important to account for the presence of a censored duration, unobserved heterogeneity, and the possibility that market duration affects non-duration explanatory variables.

The quantile regression model considered in this paper represents a more general version of several specifications recently proposed in the literature. Several econometric approaches exist for the case of one measure for each subject in the sample and exogenous covariates. While a quantile regression model with uncensored data can be estimated following Koenker and Geling (2001), a quantile regression model with censored data can be estimated using Portnoy (2003) or Peng and Huang (2008). Chernozhukov and Hansen (2005, 2006, 2008) might be used when the model includes endogenous covariates and a monotone transformation for the dependent variable. This estimator, however, does not accommodate for the possibility of censored data. In cases with more than one measure per subject and no censor data, the Koenker (2004) fixed effects estimator offers the possibility of estimating a quantile regression model with correlated individual effects. Alternatively, the practitioner can employ the Harding and Lamarche (2009) panel data estimator in a quantile regression model with potentially endogenous covariates. However, these methods are not designed to estimate a censored quantile regression model that simultaneously consider endogenous covariates and correlated individual effects.

This paper proposes a new approach to quantile regression survival analysis for longitudinal data. We allow for the possibility of endogenous covariates and correlated individual effects in the quantile regression model under conditionally independent censoring. The procedure involves minimizing ℓ_1 convex objective functions and is motivated by a martingale property associated with survival data in models with endogenous covariates. The proposed new estimator is similar in spirit to the Peng and Huang (2008) estimator, although their estimator fails to accommodate for correlated individual effects and endogenous covariates. Moreover, the performance of Peng and Huang (2008) estimator in small samples dramatically deteriorates when the model includes a large number of parameters and the degree of censoring is high. Our approach is also related to Chernozhukov and Hansen (2006, 2008), Harding and Lamarche (2009), and Koenker (2008). We carry out simulations to investigate the small sample performance of the proposed approach in comparison with other existing methods. Considering several Monte-Carlo designs, we find that the finite sample performance of the proposed method is satisfactory in all the variants of the models.

The method is applied to a panel duration model of unemployment to investigate how unemployment insurance benefits affect the duration of unemployment. We use data on unemployed workers who participated in a federal unemployment program from 1997 until 2001. Our dataset offers the possibility of observing multiple spells, where the number of spells per worker ranges from 1 to 6. Motivated by the early work of Jerry Hausman on duration data, we consider that unobserved individual heterogeneity could be potentially correlated with whether the worker received unemployment benefits and other important covariates in the model. All the variants of the quantile regression models estimated in this paper suggest that workers receiving unemployment benefits tend to be unemployed for a longer period of time than workers not receiving unemployment benefits. The estimated effect is largest at the lower tail of the conditional duration distribution, suggesting that workers' incentives to seek jobs are mostly influenced by the policy intervention among workers with low durations. Moreover, our approach offers different policy prescriptions relative to other competing approaches. The application illustrates the importance of controlling for unobserved heterogeneity in a quantile regression model when practitioners use multiple spells duration data.

The paper is organized as follows. The next section introduces the model and the proposed quantile regression estimator. Section 3 investigates the large sample performance of the proposed estimator. Section 4 offers Monte-Carlo evidence. Section 5 demonstrates how the estimator can be used in an empirical application. Section 6 concludes.

2. Model and Method

This paper investigates the estimation of quantile regression panel duration models. Specifically, we consider the following model:

$$h(T_{ij}) = \boldsymbol{d}'_{ij}\boldsymbol{\gamma} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \alpha_i + u_{ij}, \qquad i = 1, \ldots, n;\ j = 1, \ldots, m_i \tag{2.1}$$

$$\boldsymbol{d}_{ij} = \boldsymbol{w}'_{ij}\boldsymbol{\pi}_0 + \boldsymbol{x}'_{ij}\boldsymbol{\pi}_1 + \boldsymbol{v}_{ij}, \tag{2.2}$$

$$Y_{ij} = C_{ij} \wedge T_{ij}, \tag{2.3}$$

where $h(\cdot)$ is a known monotone transformation and T_{ij} is the (potentially) latent j-th response for the i-th subject. In this paper, we will adopt the logarithmic transformation, $h(T) = \log(T)$. In the first equation, the variable $\boldsymbol{d}_{ij}$ is a vector of k_1 endogenous variables, $\boldsymbol{x}_{ij}$ is a vector of k_2 exogenous independent variables, α_i is an individual effect potentially correlated with the independent variable $\boldsymbol{d}_{ij}$, and u_{ij} is the error term. We allow for dependence between the endogenous variable $\boldsymbol{d}_{ij}$ and the individual effect α_i. The parameter of interest is $\boldsymbol{\gamma}$. The second equation indicates that $\boldsymbol{d}_{ij}$ is correlated with a vector of $k_w \geq k_1$ instruments $\boldsymbol{w}_{ij}$, the exogenous variables $\boldsymbol{x}_{ij}$, and a variable $\boldsymbol{v}_{ij}$ that is stochastically dependent on u_{ij}. Although we can consider the case of over-identification, for simplicity we concentrate on the case of exact-identification. Thus, $k_w = k_1$ in what it follows. The last equation simply indicates that Y_{ij} represents observed values, with C_{ij} indicating censoring times. Additionally, we define $\delta_{ij} = I(T_{ij} \leq C_{ij})$, taking the value 1 when the event $T_{ij} \leq C_{ij}$ is true. We focus on the case where n and $\min(m_1, ..., m_n)$ are both large.

We begin by describing three examples of interest in the theoretical and empirical literature which can be thought of as special cases of the model in equations 2.1-2.3. We then consider the general version of the model and the identification strategy.

Example 1. Cross-sectional quantile regression models are investigated in Koenker and Geling (2001), Portnoy (2003) and Peng and Huang (2008). This case arises by setting $m_i = 1$ for all i

and assuming that u and v are stochastically independent. Model 2.1-2.3 reduces to:

$$h(T_i) = \boldsymbol{d}_i'\boldsymbol{\gamma} + \boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \qquad i = 1, \ldots, n \tag{2.4}$$

$$Y_i = C_i \wedge T_i. \tag{2.5}$$

While Portnoy (2003) and Peng and Huang (2008) propose approaches for the censored data case, Koenker and Geling (2001) consider the case of $\delta_i = I(T_i \leq C_i) = 1\ \forall i$.

Example 2. Chenozhukov and Hansen (2006, 2008) discuss the importance of addressing the endogeneity of d_{ij} when $m_i = 1$ for all i and $Y_{ij} = T_{ij}$ for all (i, j). They developed an approach that can be applied to the estimation of the following model:

$$h(T_i) = \boldsymbol{d}_i'\boldsymbol{\gamma} + \boldsymbol{x}_i'\boldsymbol{\beta} + u_i, \qquad i = 1, \ldots, n \tag{2.6}$$

$$\boldsymbol{d}_i = \boldsymbol{w}_i'\boldsymbol{\pi}_0 + \boldsymbol{x}_i'\boldsymbol{\pi}_1 + \boldsymbol{v}_i. \tag{2.7}$$

This model can be consistently estimated in the case of no censoring (e.g., $\delta_i = I(T_i \leq C_i) = 1$ for all subjects in the sample).

Example 3. Panel duration models arise in the case of $m > 1$. Under the assumption that u and v are stochastically independent and no censoring, we can apply the procedure in Koenker (2004) to estimate the following model:

$$h(T_{ij}) = \boldsymbol{d}_{ij}'\boldsymbol{\gamma} + \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \alpha_i + u_{ij}, \qquad i = 1, \ldots, n;\ j = 1, \ldots, m_i \tag{2.8}$$

$$Y_{ij} = T_{ij} = C_{ij} \wedge T_{ij}. \tag{2.9}$$

This approach estimates the quantile regression version of the previous model under large n, large m conditions. The case of dependence between u and v can be addressed by the method proposed by Harding and Lamarche (2009).

2.1. Estimating a Quantile Regression Duration Model

This paper develops an estimation procedure for the following quantile regression model,

$$Q_{T_{ij}}(\tau|\boldsymbol{d}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \exp(\boldsymbol{d}_{ij}'\boldsymbol{\gamma}(\tau) + \boldsymbol{x}_{ij}'\boldsymbol{\beta}(\tau) + \boldsymbol{z}_{ij}'\boldsymbol{\alpha}(\tau)), \tag{2.10}$$

where the quantile $\tau \in (0, 1)$, the variable $\boldsymbol{z}_{ij} = (0, ..., 1, ..., 0)'$ is a $n \times 1$ indicator for the individual effect α_i, and $\boldsymbol{\alpha}(\tau) = (\alpha_1(\tau), ..., \alpha_n(\tau))'$ is a $n \times 1$ vector of nuisance parameters. Given the logarithmic transformation $h(T) = \log(T)$, model 2.10 is equivalent to $Q_{\log T_{ij}}(\tau|\boldsymbol{d}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \boldsymbol{d}_{ij}'\boldsymbol{\gamma}(\tau) + \boldsymbol{x}_{ij}'\boldsymbol{\beta}(\tau) + \boldsymbol{z}_{ij}'\boldsymbol{\alpha}(\tau)$. Then, model 2.10 can be seen as the quantile regression version of equation 2.1 after the logarithmic transformation is adopted and α_i is replaced by $\boldsymbol{z}_{ij}'\boldsymbol{\alpha}$. The alternative notation is used in this section because it is convenient for introducing conditional quantile functions and estimating equations.

We assume that $P(T_{ij} \leq Q_{T_{ij}}(\tau|\boldsymbol{d}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij})|\boldsymbol{w}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij})$ is equal to the quantile τ. This assumption is equivalent to require that $u_{ij}(\tau) \equiv T_{ij} - Q_{T_{ij}}(\tau|\boldsymbol{d}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij})$ has zero τ-th quantile conditional on $\boldsymbol{w}_{ij}$, $\boldsymbol{x}_{ij}$, and $\boldsymbol{z}_{ij}$. Our parameter of interest, $\boldsymbol{\gamma}(\tau)$, is the effect of the endogenous variable on the quantile τ of the conditional distribution of the latent response variable $\log(T)$. The parameter $\boldsymbol{\gamma}(\tau)$ provides an opportunity for investigating how the endogenous factors influence the location, scale, and shape of the conditional distribution of the response variable.

The model also includes exogenous variables and individual effects. We assume that the individual effect, $\alpha_i(\tau)$, represents a distributional shift since we are focusing on estimating large m_i survival models. In other circumstances, when m_i is relatively small, it is appropriate to impose the condition that $\alpha_i(\tau) = \alpha_i$ for all $\tau \in (0, 1)$ (see, e.g., Koenker 2004). The individual specific effect is then a location shift, implying that the conditional quantiles of T for each subject can have different locations and equal shapes. An additional advantage of this condition is that the assumption facilitates the interpretation of α_i as a fixed effect. Our specification allows for different locations and shapes, which appears to be a more flexible model.

Our method estimates a quantile regression model 2.10 considering *iid* samples, $\{(Y_{ij}, \delta_{ij}, \boldsymbol{d}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{w}_{ij}) : i = 1, \ldots, n; j = 1, \ldots, m_i\}$. Consider for simplicity the case of balanced designs with $m_i = m$ for all i. We solve a linear programming problem corresponding to the following estimating equations,

$$\begin{aligned} E\Big\{&(nm)^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{m} (\boldsymbol{x}'_{ij}, \boldsymbol{z}'_{ij}, \boldsymbol{w}'_{ij})' N_{ij} (\exp(\boldsymbol{d}'_{ij}\boldsymbol{\gamma}(\tau) + \boldsymbol{x}'_{ij}\boldsymbol{\beta}(\tau) + \boldsymbol{z}'_{ij}\boldsymbol{\alpha}(\tau) + \boldsymbol{w}'_{ij}\boldsymbol{\lambda}(\tau))) \\ &- \int_0^{\tau} I(Y_{ij} \geq \exp((\boldsymbol{d}'_{ij}\boldsymbol{\gamma}(\upsilon) + \boldsymbol{x}'_{ij}\boldsymbol{\beta}(\upsilon) + \boldsymbol{z}'_{ij}\boldsymbol{\alpha}(\upsilon) + \boldsymbol{w}'_{ij}\boldsymbol{\lambda}(\upsilon)))dH(\upsilon)\Big\} = 0, \end{aligned} \tag{2.11}$$

where $H(\upsilon) = -\log(1 - \upsilon)$ for $\upsilon \in [0, 1)$ and the counting process $N_{ij} = I(Y_{ij} \leq t, \delta_{ij} = 1)$. It is possible to show that solving equation 2.11 is equivalent to finding the argument that minimizes an ℓ_1 convex objective function. The left hand side of equation 2.11 can be interpreted as the gradient of the convex objective function defined below in equation 2.12. (See Remark 2 below for details on related ideas for cross-sectional data).

Our convex objective function is defined as the sum of convex functions of the form,

$$Q_{ij}(\tau, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \varrho_\tau \left(h(y_{ij}) - \boldsymbol{d}'_{ij}\boldsymbol{\gamma} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{\alpha} - \boldsymbol{w}'_{ij}\boldsymbol{\lambda}\right), \tag{2.12}$$

where the function $\varrho_\tau(u) = u(\vartheta(\tau) - \delta \times I(u \leq 0))$, and

$$\vartheta_{ij}(\tau) = \sum_{k=0}^{q-1} I(h(y_{ij}) - \boldsymbol{d}'_{ij}\boldsymbol{\gamma} \geq \boldsymbol{x}_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\hat{\boldsymbol{\alpha}} + \boldsymbol{w}'_{ij}\hat{\boldsymbol{\lambda}})(H(\tau_{k+1}) - H(\tau_k)). \tag{2.13}$$

First we minimize the objective function above for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\lambda}$ as functions of τ and $\boldsymbol{\gamma}$,

$$\{\hat{\boldsymbol{\beta}}(\tau,\boldsymbol{\gamma}), \hat{\boldsymbol{\alpha}}(\tau,\boldsymbol{\gamma}), \hat{\boldsymbol{\lambda}}(\tau,\boldsymbol{\gamma})\} = \underset{\beta,\alpha,\lambda \in \mathcal{B}\times\mathcal{A}\times\mathcal{L}}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} Q_{ij}(\tau,\boldsymbol{\gamma};\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\lambda}). \tag{2.14}$$

Then we estimate the coefficient on the endogenous variable by finding the value of $\boldsymbol{\gamma}$, which minimizes a weighted distance function defined on $\boldsymbol{\lambda}$:

$$\hat{\gamma}(\tau) = \underset{\gamma \in \mathcal{G}}{\operatorname{argmin}} \left\{ \hat{\boldsymbol{\lambda}}(\tau,\boldsymbol{\gamma})'\hat{\boldsymbol{A}}(\tau)\hat{\boldsymbol{\lambda}}(\tau,\boldsymbol{\gamma}) \right\} \tag{2.15}$$

for a positive definite matrix $\boldsymbol{A}$. The parameter estimates are then given by,

$$\hat{\boldsymbol{\theta}}(\tau) \equiv \left(\hat{\boldsymbol{\gamma}}(\tau), \hat{\boldsymbol{\beta}}(\tau), \hat{\boldsymbol{\alpha}}(\tau)\right) = \left(\hat{\boldsymbol{\gamma}}(\tau), \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}}(\tau),\tau)), \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\gamma}}(\tau),\tau))\right). \tag{2.16}$$

Remark 1. To address the possibility of endogenous covariates, the estimator could be alternatively defined by integrating the approaches considered in Koenker (2004), Blundell and Powell (2007), and Peng and Huang (2008). The control function approach proposed in Blundell and Powell (2007) offers a convenient alternative for addressing the possibility of endogenous regressors. Our approach can be seen within the recent developments on quantile regression for the structural equations model (see, e.g., Chernozhukov and Hansen (2005) and Harding and Lamarche (2009)) to estimate the conditional quantile function. Alternatively, it can be seen as an extension of the ideas behind existing methods developed for the classical case of endogenous covariates in duration models (see, e.g., Lancaster (1985), Olsen and Farkas (1988), and Chesher (2002)).

Remark 2. The condition that leads to the modified version of the quantile regression check function, denoted by $\varrho_\tau(\cdot)$, is motivated by the martingale property of counting processes presented in Fleming and Harrington (1991) and employed by Peng and Huang (2008). Peng and Huang's estimator however fails to accommodate for endogenous individual effects and endogenous covariates. In the cross-sectional case with no endogenous covariates and no unobserved heterogeneity, Peng and Huang's estimating equations are,

$$\mathbb{E}\left\{ n^{-1/2}\sum_{i=1}^{n} \boldsymbol{x}'_i N_i(\exp(\boldsymbol{x}'_i\boldsymbol{\beta}(\tau))) - \int_0^\tau I(Y_i \geq \exp(\boldsymbol{x}'_i\boldsymbol{\beta}(s)))dH(s) \right\} = 0. \tag{2.17}$$

As pointed by Peng and Huang (2008), the monotonicity of (2.17) facilitates the computation of the estimator. Koenker (2008) provides important insights on the implementation of the approach, and Koenker (2010) offers an efficient estimation method based on simplex and interior point methods.

Remark 3. It is interesting to note that when there is no censoring, the method is similar to the panel data approach proposed by Harding and Lamarche (2009). In the one sample case considered in Peng and Huang (2008), the second term in equation (2.17) is $-\log(1-\tau) \approx \tau$, and therefore it follows that

$$Q_{ij}(\tau, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = \varrho_\tau \left(h(y_{ij}) - \boldsymbol{d}'_{ij}\boldsymbol{\gamma} - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{\alpha} - \boldsymbol{w}'_{ij}\boldsymbol{\lambda} \right), \tag{2.18}$$

where $\varrho_\tau(u) = \rho_\tau(u) = u(\tau - I(u \leq 0))$, the standard quantile regression check function (Koenker 2005).

2.2. Panel Implementation

The implementation of the estimator defined in 2.14 and 2.15 relies heavily on the sparsity of the design. Koenker (2008) shows that the solution of Peng and Huang's (2008) generalized equations 2.17 can be obtained as a solution of a linear programming problem. It is natural then to use existing simplex methods, and/or interior point methods. The key idea behind the implementation is a dual formulation of the linear minimization problem. However, finding the solution of this problem can be challenging when the number of estimated parameters is large and the degree of censoring is high. To overcome these difficulties, we use sparse matrix methods to take advantage of the sparsity of the design. This dramatically increases the algorithm's speed of convergence and significantly reduces memory requirements. The practitioner interested in single spell duration data should consider Koenker (2010), which offers several procedures for estimating a censored quantile regression model with cross-sectional data.

2.3. Inference and Asymptotic Considerations

We propose to use the bootstrap to provide inference about $\hat{\boldsymbol{\theta}}$. Peng and Huang (2008) highlight that the bootstrap seems to have advantages over the estimation of the covariance matrices of the limiting process $(nm)^{-1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. It might be possible to follow the quantile regression literature by considering the pair bootstrap by replacing pairs $\{(h(\boldsymbol{Y}_{i\bullet}), \boldsymbol{d}'_{i\bullet}, \boldsymbol{x}'_{i\bullet}, \boldsymbol{w}'_{i\bullet}) : 1 = 1, \ldots n\}$ over subject units i.[1] In the case of panel duration data we proceed by drawing a sample (with replacement) of

[1]Although the wild bootstrap has been used and extensively analyzed by researchers, the vast majority of the existing theory is associated with linear estimators. To the best of our knowledge, the exception is Feng, He and Wu (2011) which propose a wild bootstrap approach for the quantile regression estimator (Koenker and Bassett 1978). They find that many choices of the weight distribution commonly considered in the literature do not work for estimators with non-linear score functions. While the pair bootstrap works in the presence of heteroskedasticity, there is some evidence that the wild bootstrap may outperform the

n subjects including their m_i observations. Using these new pairs $(h(\boldsymbol{Y}_{i\bullet}^*), \boldsymbol{d}_{i\bullet}^*, \boldsymbol{x}_{i\bullet}^*, \boldsymbol{w}_{i\bullet}^*))$, we obtain $\boldsymbol{\theta}^*$ as the argument that minimizes 2.14 and 2.15. We reiterate this procedure B times to obtain a large sample of realizations $\{\boldsymbol{\theta}_b^*\}_{b=1}^B$. For a given quantile, we can obtain the variance of $\hat{\boldsymbol{\theta}}(\tau)$ as the sample variance of $\{\boldsymbol{\theta}_b^*\}_{b=1}^B$.

In this paper, we follow the resampling method described in Section 4.1 of Peng and Huang (2008), which was initially proposed by Jin, Ying and Wei (2001) and considered in Ma and Kosorok (2005). Peng and Huang propose a simple resampling approach designed to perturb the objective function 2.14. They suggest drawing $\zeta_{11}, \ldots, \zeta_{nm}$ independent variables from a non-negative distribution with mean 1 and variance 1. We consider the exponential distribution, drawing $\{\boldsymbol{\zeta}_{i\bullet} : i = 1, \ldots n\}$ over subjects. One can then obtain $\boldsymbol{\theta}^*$ as the argument that minimizes a perturbed equation 2.14 (See Section 4.1 of Peng and Huang (2008) for additional details). As before, we reiterate this procedure B times to obtain a large sample of realizations $\{\boldsymbol{\theta}_b^*\}_{b=1}^B$ and compute its sample variance. Moreover, a $100(1-2q)$ confidence interval can be obtained by constructing the qth quantile and $(1-q)$th quantile of $\{\boldsymbol{\theta}_b^*\}_{b=1}^B$. This procedure, which is applied in the empirical application considered in Section 4.2, works under fairly general conditions. Because it maintains the endogenous structure of the model, it can be accommodated to include the use of instrumental variables.

It is possible also to employ a Wald-type statistic (see, e.g., Koenker and Bassett, (1982) and Koenker, (2005)) for testing a basic general linear hypothesis on a vector $\boldsymbol{\xi}$ of the form $H_0 : \boldsymbol{R}\boldsymbol{\xi} = \boldsymbol{r}$, where $\boldsymbol{R}$ is a matrix that depends on the type of restrictions imposed. We evaluate the null hypothesis of equality of effects across quantiles considering a vector $\boldsymbol{\xi} = (\boldsymbol{\theta}(\tau_1), ..., \boldsymbol{\theta}(\tau_J))'$. In the special case of $\delta_{ij} = 1$ and $\hat{\mathbf{v}}_{ij} = c$ for all (i, j), it is possible to evaluate the vector over a range of quantiles by extending the framework developed in Koenker and Xiao (2002). The null hypothesis is $H_0 : \theta_k(\tau) = \mu_k + \sigma_k \theta(\tau)$, where k indicates the covariate, μ_k is the location parameter and σ_k is the scale parameter. Note that $\sigma_k = 0$ implies that the covariate effect affects only the location of the conditional distribution of survival time. Alternatively, if $\sigma_k > 0$, the covariate effect affects both the location and scale of the conditional distribution of the response.

The derivation of the asymptotic results including asymptotic covariance matrices raises several issues that are worth mentioning. First, the existence of a vector of dimension n that tends to infinity was noted by Koenker (2004). This issue is typically avoided by concentrating out the

pair bootstrap in linear models (Flachaire, 2005). Their relative performance in a quantile regression setting is not known however.

Bahadur representation of the individual effects in the objective function (see, e.g., Koenker 2004, Lamarche 2010). Second, the derivation of asymptotic results is facilitated by letting n and m tend to infinity, although it is possible that the estimator is biased in cases of low, fixed m (see Graham, Hahn, and Powell 2009 for a discussion when $\delta_{ij} = 1$ for all (i, j) and the independent variables are exogenous). We emphasize this point because m_i ranges between 3 and 6 unemployment spells in the application considered in Section 4. Let τ_U be a deterministic constant subject to some identifiability restrictions by the presence of censoring (Peng and Huang 2008). If regularity conditions are met, for $\tau \in (0, \tau_U)$, it is possible to show that the quantile regression estimator $\hat{\boldsymbol{\theta}}(\tau)$ converges weakly to a Gaussian process. The result can be shown using results and arguments in Chernozhukov and Hansen (2006, 2008), Koenker (2004) and Peng and Huang (2008).

3. Monte Carlo Evidence

This section reports the results of several simulation experiments designed to evaluate the performance of the proposed method in finite samples.

We generate the dependent variable considering the following equations:

$$T_{ij} = \beta_0 + \beta_1 d_{ij} + \beta_2 x_{ij} + \alpha_i + (1 + \delta d_{ij}) u_{ij}, \tag{3.1}$$

$$d_{ij} = \pi_0 + \pi_1 w_{ij} + v_{ij}, \tag{3.2}$$

$$\alpha_i = \gamma_1 \bar{d}_i + \gamma_2 \epsilon_i, \tag{3.3}$$

$$w_{ij} = \mu \eta_i + e_{ij}, \tag{3.4}$$

and $Y_{ij} = C_{ij} \wedge T_{ij}$, where C_{ij} indicates censoring times and $\bar{d}_i = m^{-1} \sum_j d_{ij}$ is the individual specific sample mean of d. The model presented in equations 3.1-3.4 is estimated using different sample sizes $n = \{10, 20, 50\}$ and $m = \{5, 20, 50\}$. In models with $m = 5$, we consider the following fixed and random censoring times: $C_{im} = 4$ and $C_{im} = 4 + \xi_{im}$, where ξ_{im} is a uniform random variable. When $m > 5$, we assume that the fixed censoring point is $C_{ij} = 7.5$ and the random censoring point is $C_{ij} = 6.5 + 0.25\xi_{ij}$, where ξ_{ij} is a Gaussian random variable. As a result of these variants of the model, the proportion of censoring in the simulations ranges from 6% to 25%. The random variable $\epsilon_i \sim \mathcal{N}(0, 0.5)$, and the variables $(e, x, \eta)'$ are Gaussian independent random variables. The error terms in the first two equations 3.1 and 3.2 are $(u_{it}, v_{it})' \sim \mathcal{N}(0, \boldsymbol{\Omega})$, where $\Omega_{11} = \Omega_{22} = 1$. The parameters are assumed to be: $\beta_0 = 5$, $\beta_1 = \beta_2 = \pi_1 = 1$, $\pi_0 = 0$, $\gamma_2 = -0.5$, and $\mu = 0.25$.

We consider four basic variations of the model:

Design I: The endogenous variable d_{ij} is not correlated with the α_i's, and the variables u_{ij} and v_{ij} are independent Gaussian variables. We assume $\gamma_1 = 0$ and $\Omega_{12} = \Omega_{21} = 0$.

Design II: The endogenous variable d_{ij} is not correlated with the α_i's, and the variables u_{ij} and v_{ij} are not independent. We assume $\gamma_1 = 0$ and $\Omega_{12} = \Omega_{21} = -0.5$.

Design III: The endogenous variable d_{ij} is correlated with the α_i's, and the variables u_{ij} and v_{ij} are independent Gaussian variables. We assume $\gamma_1 = -0.5$, and $\Omega_{12} = \Omega_{21} = 0$.

Design IV: The variables u_{it} is correlated with v_{it} and the individual effect α_i is correlated with the independent variable d_{ij}. We assume that $\gamma_1 = -0.5$, and $\Omega_{12} = \Omega_{21} = -0.5$.

Finally, we evaluate the method under fixed and random censoring in the location-shift model ($\delta = 0$) and the location-scale shift model ($\delta = 0.1$). By allowing δ to be different than zero, we consider models where the slope parameter β_1 changes across the quantiles of the conditional distribution of the response variable T. We report results at two different quantiles $\tau = \{0.25, 0.50\}$.

Tables 3.1-3.4 present the percentage of censored data, bias and root mean square error (RMSE) of the simulation experiments. For instance, Table 3.1 shows results for model 3.1-3.4 under fixed censoring when $\delta = 0$. While the upper panel of the table presents results for the case that $\tau = 0.5$, the lower panel of the table presents results for the case of $\tau = 0.25$. The tables show results from (i) a quantile regression approach that uses the latent variable T (Omni), (ii) a quantile regression approach that uses the variable Y (Naive), (iii) the survival approach for a censored quantile regression model (Peng-Huang) proposed by Peng and Huang (2008), and (iv) the quantile regression estimator for a panel duration model with individual effects (Panel). Note that the first three procedures being compared are in effect cross-sectional methods.

We can compare the different estimators for Model I in the absence of endogeneity or correlated individual effects. Our estimator performs very similarly to the Peng-Huang estimator and both are a substantial improvement relative to the Naive estimator. As we introduce endogeneity in Models II to IV however the performance of the Peng-Huang estimator decreases substantially and becomes comparable to that of the naive estimator. Our proposed estimator, by contrast, continues to perform very well. This indicates that, relatively to each other, endogeneity and correlated individual effects induce biases, which are an order of magnitude larger than the biases induced by censoring which is unaccounted for.

Our estimator performs exceptionally well in Design III, which has correlated individual specific effects but no additional source of endogeneity. This highlights the effectiveness of using a panel data approach to controlling for unobserved heterogeneity when such data is in fact available. Removing additional endogeneity using instrumental variables induces a small amount of bias (less than 5% in almost all cases, and less than 2% in most cases). In our simulations we have also considered the case where $n = 10$ and $m = 50$. Strictly speaking this corresponds to a multiple time-series scenario rather than the more common longitudinal data format encountered in microeconometrics. Notice that in this setting under Design III, the Peng-Huang estimator also performs quite well in the sense that it has substantially lower bias than the Naive estimator. This might be explained by noting that Naive does not account for censoring and $\gamma_1/m \approx -0.01$ generates a small correlation between α_i and the zero-mean variables d_{ij}'s. Nevertheless, the bias of the Peng-Huang estimator is larger than that of the Panel estimator proposed in this paper. In terms of RMSE however it performs better than the Panel estimator. This is to be expected since the Panel estimator includes individual effects and also performs an additional IV step and is thus not efficient. The multiple time-series case however is unlikely to be encountered in applied econometrics as typically in duration analysis data on multiple individuals is easier to obtain than data on a few individuals over a long period of time.

The good performance of the Panel estimator relative to the other competing estimators is also observed for sample size (n, m) equal to $(10, 20)$ and $(20, 50)$. The results were qualitatively similar to the ones reported in Tables 3.1-3.4, and they are not reported in the paper. They are, however, available upon request from the authors.

The simulations indicate that the proposed Panel estimator performs very well under both random and fixed censoring and is thus to be preferred over the Peng-Huang estimator when endogeneity and correlated individual effects are of concern.

Sample n	m	Design	Censoring	Omni Bias	Omni RMSE	Naive Bias	Naive RMSE	Peng-Huang Bias	Peng-Huang RMSE	Panel Bias	Panel RMSE
					$\tau = 0.5$ quantile						
50	5	I	0.106	-0.005	0.075	-0.086	0.116	-0.003	0.078	0.011	0.102
		II	0.104	-0.260	0.269	-0.323	0.331	-0.257	0.267	0.010	0.100
		III	0.118	-0.165	0.175	-0.243	0.251	-0.168	0.179	-0.002	0.088
		IV	0.097	-0.394	0.398	-0.456	0.460	-0.392	0.396	0.003	0.092
20	5	I	0.120	-0.011	0.115	-0.090	0.152	-0.012	0.118	-0.005	0.156
		II	0.123	-0.262	0.283	-0.322	0.342	-0.259	0.281	0.016	0.159
		III	0.153	-0.150	0.181	-0.220	0.243	-0.149	0.182	-0.019	0.155
		IV	0.128	-0.406	0.416	-0.461	0.473	-0.407	0.418	-0.009	0.149
20	20	I	0.097	-0.004	0.057	-0.087	0.102	-0.002	0.059	0.002	0.071
		II	0.063	-0.240	0.246	-0.287	0.292	-0.243	0.249	0.023	0.079
		III	0.110	-0.053	0.071	-0.148	0.154	-0.049	0.071	0.006	0.075
		IV	0.079	-0.302	0.304	-0.346	0.348	-0.298	0.301	0.035	0.086
10	50	I	0.099	0.005	0.058	-0.086	0.102	0.007	0.062	0.000	0.062
		II	0.086	-0.230	0.236	-0.287	0.291	-0.228	0.234	0.029	0.070
		III	0.113	-0.023	0.046	-0.130	0.135	-0.025	0.052	0.003	0.065
		IV	0.083	-0.264	0.267	-0.319	0.321	-0.263	0.266	0.034	0.075
					$\tau = 0.25$ quantile						
50	5	I	0.106	-0.005	0.073	-0.166	0.185	-0.002	0.075	0.006	0.112
		II	0.104	-0.256	0.266	-0.394	0.402	-0.253	0.262	0.010	0.102
		III	0.118	-0.161	0.173	-0.386	0.395	-0.163	0.175	-0.005	0.101
		IV	0.097	-0.391	0.395	-0.585	0.589	-0.388	0.393	0.002	0.098
20	5	I	0.120	-0.008	0.123	-0.135	0.199	-0.007	0.128	-0.014	0.172
		II	0.123	-0.261	0.288	-0.378	0.400	-0.262	0.289	0.002	0.160
		III	0.153	-0.147	0.186	-0.301	0.329	-0.153	0.191	-0.009	0.166
		IV	0.128	-0.407	0.421	-0.525	0.538	-0.407	0.421	-0.006	0.161
20	20	I	0.097	0.001	0.060	-0.062	0.081	0.001	0.060	-0.002	0.078
		II	0.063	-0.242	0.249	-0.271	0.276	-0.243	0.250	0.017	0.083
		III	0.110	-0.037	0.060	-0.119	0.127	-0.036	0.061	0.003	0.069
		IV	0.079	-0.285	0.289	-0.333	0.336	-0.286	0.290	0.027	0.081
10	50	I	0.099	-0.001	0.056	-0.075	0.091	0.000	0.058	-0.001	0.059
		II	0.086	-0.230	0.236	-0.265	0.270	-0.228	0.235	0.020	0.068
		III	0.113	-0.033	0.057	-0.104	0.112	-0.033	0.059	0.004	0.075
		IV	0.083	-0.304	0.307	-0.331	0.333	-0.304	0.307	0.018	0.069

TABLE 3.1. *Monte carlo results for model 3.1-3.4 under fixed censoring in the location shift model.*

Sample *n*	*m*	Design	Censoring	Omni Bias	Omni RMSE	Naive Bias	Naive RMSE	Peng-Huang Bias	Peng-Huang RMSE	Panel Bias	Panel RMSE
					$\tau = 0.5$ quantile						
50	5	I	0.085	0.001	0.074	-0.076	0.108	-0.001	0.075	0.009	0.086
		II	0.085	-0.255	0.264	-0.318	0.326	-0.253	0.263	0.016	0.088
		III	0.100	-0.194	0.204	-0.259	0.266	-0.192	0.203	-0.005	0.098
		IV	0.097	-0.428	0.432	-0.480	0.483	-0.426	0.430	0.002	0.093
20	5	I	0.104	-0.013	0.115	-0.080	0.140	-0.007	0.116	-0.016	0.147
		II	0.103	-0.262	0.283	-0.312	0.333	-0.256	0.279	0.005	0.142
		III	0.125	-0.144	0.178	-0.215	0.239	-0.146	0.180	0.009	0.159
		IV	0.128	-0.399	0.411	-0.450	0.460	-0.403	0.416	0.022	0.170
20	20	I	0.186	-0.007	0.061	-0.185	0.194	-0.009	0.071	0.002	0.076
		II	0.175	-0.245	0.251	-0.370	0.373	-0.242	0.249	0.064	0.114
		III	0.226	-0.060	0.074	-0.273	0.276	-0.058	0.077	0.001	0.078
		IV	0.187	-0.300	0.303	-0.430	0.432	-0.296	0.300	0.066	0.108
10	50	I	0.217	-0.004	0.050	-0.227	0.234	-0.004	0.063	0.003	0.071
		II	0.152	-0.241	0.246	-0.342	0.346	-0.237	0.244	0.052	0.093
		III	0.233	-0.036	0.058	-0.270	0.273	-0.026	0.060	0.010	0.078
		IV	0.196	-0.288	0.291	-0.429	0.431	-0.286	0.289	0.073	0.109
					$\tau = 0.25$ quantile						
50	5	I	0.085	0.009	0.079	-0.128	0.154	0.008	0.080	0.011	0.097
		II	0.085	-0.252	0.262	-0.362	0.371	-0.253	0.264	0.016	0.106
		III	0.100	-0.181	0.193	-0.320	0.329	-0.179	0.193	0.005	0.106
		IV	0.097	-0.419	0.424	-0.531	0.535	-0.420	0.424	0.005	0.097
20	5	I	0.104	-0.004	0.120	-0.131	0.181	-0.004	0.126	-0.022	0.168
		II	0.103	-0.259	0.280	-0.364	0.381	-0.258	0.280	-0.005	0.161
		III	0.125	-0.144	0.179	-0.263	0.286	-0.144	0.181	0.009	0.181
		IV	0.128	-0.400	0.412	-0.492	0.503	-0.402	0.414	0.020	0.178
20	20	I	0.186	0.013	0.068	-0.130	0.142	0.014	0.071	0.004	0.077
		II	0.175	-0.223	0.231	-0.302	0.306	-0.222	0.231	0.043	0.090
		III	0.226	-0.044	0.068	-0.239	0.243	-0.041	0.071	0.006	0.084
		IV	0.187	-0.299	0.303	-0.397	0.399	-0.297	0.301	0.057	0.108
10	50	I	0.217	-0.005	0.066	-0.139	0.150	-0.003	0.067	0.002	0.079
		II	0.152	-0.254	0.262	-0.325	0.329	-0.253	0.261	0.048	0.094
		III	0.233	-0.042	0.062	-0.231	0.234	-0.040	0.066	0.004	0.073
		IV	0.196	-0.287	0.290	-0.398	0.400	-0.286	0.289	0.055	0.100

TABLE 3.2. *Monte carlo results for model 3.1-3.4 under under random censoring in the location shift model.*

Sample n	m	Design	Censoring	Omni Bias	Omni RMSE	Naive Bias	Naive RMSE	Peng-Huang Bias	Peng-Huang RMSE	Panel Bias	Panel RMSE
$\tau = 0.5$ quantile											
50	5	I	0.105	-0.005	0.074	-0.095	0.123	-0.001	0.076	0.007	0.100
		II	0.102	-0.253	0.263	-0.319	0.326	-0.245	0.254	0.016	0.099
		III	0.116	-0.164	0.174	-0.261	0.269	-0.166	0.179	0.000	0.086
		IV	0.114	-0.429	0.432	-0.510	0.513	-0.419	0.423	0.003	0.098
20	5	I	0.118	-0.007	0.111	-0.100	0.154	-0.010	0.116	-0.004	0.154
		II	0.120	-0.253	0.276	-0.327	0.345	-0.249	0.272	0.017	0.157
		III	0.152	-0.154	0.184	-0.236	0.259	-0.152	0.185	-0.024	0.163
		IV	0.139	-0.406	0.415	-0.467	0.477	-0.410	0.420	-0.030	0.154
20	20	I	0.090	0.002	0.057	-0.083	0.099	0.002	0.058	-0.002	0.065
		II	0.062	-0.232	0.238	-0.274	0.279	-0.231	0.237	0.022	0.072
		III	0.112	-0.039	0.060	-0.140	0.145	-0.038	0.060	0.002	0.065
		IV	0.078	-0.272	0.275	-0.319	0.322	-0.269	0.272	0.039	0.080
10	50	I	0.115	0.001	0.053	-0.102	0.114	0.002	0.056	0.001	0.059
		II	0.085	-0.225	0.231	-0.278	0.282	-0.221	0.227	0.027	0.067
		III	0.107	-0.035	0.055	-0.129	0.134	-0.032	0.055	0.002	0.065
		IV	0.069	-0.288	0.291	-0.326	0.328	-0.284	0.287	0.029	0.077
$\tau = 0.25$ quantile											
50	5	I	0.105	0.009	0.071	-0.154	0.172	0.012	0.075	0.025	0.106
		II	0.102	-0.231	0.241	-0.358	0.364	-0.223	0.233	0.035	0.107
		III	0.116	-0.153	0.163	-0.380	0.388	-0.152	0.164	0.014	0.100
		IV	0.114	-0.373	0.377	-0.552	0.555	-0.370	0.375	0.025	0.106
20	5	I	0.118	0.001	0.126	-0.128	0.187	0.004	0.127	0.007	0.174
		II	0.120	-0.240	0.269	-0.347	0.369	-0.242	0.270	0.023	0.167
		III	0.152	-0.143	0.185	-0.298	0.324	-0.148	0.187	0.001	0.168
		IV	0.139	-0.394	0.406	-0.497	0.508	-0.397	0.409	0.010	0.166
20	20	I	0.090	0.013	0.061	-0.039	0.065	0.014	0.061	0.003	0.072
		II	0.062	-0.213	0.220	-0.234	0.239	-0.213	0.219	0.018	0.079
		III	0.112	-0.037	0.060	-0.104	0.113	-0.035	0.059	0.004	0.066
		IV	0.078	-0.274	0.277	-0.301	0.304	-0.272	0.276	0.021	0.073
10	50	I	0.115	0.010	0.056	-0.052	0.073	0.012	0.057	0.001	0.058
		II	0.085	-0.208	0.215	-0.233	0.238	-0.206	0.213	0.018	0.067
		III	0.107	-0.031	0.055	-0.090	0.099	-0.029	0.054	0.005	0.071
		IV	0.069	-0.280	0.283	-0.298	0.300	-0.278	0.281	0.019	0.079

TABLE 3.3. *Monte carlo results for model 3.1-3.4 under under fixed censoring in the location-scale shift model.*

Sample n	m	Design	Censoring	Omni Bias	Omni RMSE	Naive Bias	Naive RMSE	Peng-Huang Bias	Peng-Huang RMSE	Panel Bias	Panel RMSE
					$\tau = 0.5$ quantile						
50	5	I	0.087	0.001	0.075	-0.086	0.115	0.003	0.074	0.010	0.086
		II	0.082	-0.251	0.261	-0.317	0.326	-0.245	0.256	0.019	0.088
		III	0.098	-0.196	0.205	-0.273	0.280	-0.192	0.202	-0.005	0.095
		IV	0.094	-0.425	0.429	-0.486	0.489	-0.419	0.423	0.006	0.097
20	5	I	0.102	-0.016	0.112	-0.090	0.148	-0.010	0.117	-0.013	0.148
		II	0.101	-0.259	0.281	-0.319	0.339	-0.251	0.276	0.008	0.152
		III	0.123	-0.145	0.179	-0.227	0.251	-0.147	0.183	0.002	0.161
		IV	0.124	-0.395	0.407	-0.454	0.465	-0.396	0.409	0.023	0.169
20	20	I	0.181	0.006	0.057	-0.178	0.186	0.009	0.062	0.000	0.073
		II	0.169	-0.241	0.247	-0.356	0.360	-0.232	0.239	0.056	0.105
		III	0.248	-0.061	0.075	-0.296	0.299	-0.055	0.075	-0.004	0.078
		IV	0.179	-0.288	0.291	-0.409	0.411	-0.276	0.280	0.055	0.096
10	50	I	0.189	-0.009	0.059	-0.198	0.204	-0.008	0.066	-0.007	0.064
		II	0.146	-0.233	0.239	-0.327	0.330	-0.225	0.232	0.043	0.081
		III	0.218	-0.024	0.046	-0.247	0.250	-0.023	0.049	0.002	0.066
		IV	0.188	-0.277	0.280	-0.409	0.411	-0.267	0.270	0.064	0.097
					$\tau = 0.25$ quantile						
50	5	I	0.087	0.021	0.081	-0.115	0.141	0.022	0.083	0.027	0.103
		II	0.082	-0.231	0.241	-0.331	0.339	-0.226	0.238	0.039	0.112
		III	0.098	-0.173	0.184	-0.314	0.322	-0.172	0.185	0.018	0.110
		IV	0.094	-0.403	0.408	-0.506	0.509	-0.400	0.405	0.027	0.107
20	5	I	0.102	0.006	0.118	-0.118	0.169	0.010	0.121	-0.009	0.180
		II	0.101	-0.238	0.260	-0.332	0.348	-0.236	0.259	0.016	0.164
		III	0.123	-0.139	0.174	-0.257	0.279	-0.140	0.175	0.028	0.183
		IV	0.124	-0.381	0.393	-0.469	0.478	-0.381	0.393	0.038	0.182
20	20	I	0.181	0.012	0.065	-0.110	0.122	0.011	0.066	0.004	0.073
		II	0.169	-0.194	0.203	-0.259	0.264	-0.192	0.201	0.040	0.086
		III	0.248	-0.055	0.072	-0.196	0.200	-0.053	0.071	-0.002	0.069
		IV	0.179	-0.279	0.282	-0.354	0.356	-0.271	0.276	0.050	0.101
10	50	I	0.189	0.010	0.056	-0.140	0.150	0.012	0.056	0.009	0.068
		II	0.146	-0.229	0.237	-0.281	0.287	-0.226	0.235	0.040	0.086
		III	0.218	-0.021	0.048	-0.168	0.172	-0.019	0.052	0.007	0.073
		IV	0.188	-0.270	0.273	-0.356	0.357	-0.263	0.265	0.046	0.090

Table 3.4. *Monte carlo results for model 3.1-3.4 under under random censoring in the location-scale shift model.*

Number of Spells	Duration (in days)	Number of: Workers	Observations
1	442	2213	2213
2	186	722	1444
3	118	271	813
4	97	78	312
5	82	37	185
6	66	8	48
Total	-	3329	5015

TABLE 4.1. *Distribution of Spells and Durations.*

4. An Empirical Application

In this section, we analyze data for unemployed workers who received unemployment insurance, to study how insurance benefits affects the duration of unemployment.[2] The effect of unemployment insurance on the duration of unemployment has also been investigated by Han and Hausman (1990), Ham and Rea (1987), Meyer (1990) and Hausman and Woutersen (2005). This paper is however the first one to estimate the impact of unemployment insurance using repeated unemployment spells. Our results indicate that workers receiving unemployment benefits tend to be unemployed for a longer period of time than workers not receiving unemployment benefits. The quantile approach reveals that the estimated effect is largest at the lower tail of the conditional duration distribution, suggesting that workers' incentives to seek jobs among workers with low durations are more susceptible to the effect of unemployment insurance. The application illustrates the importance of controlling for unobserved heterogeneity in (quantile) survival analysis.

4.1. Data

The data used here includes unemployed workers who were eligible for receiving unemployment insurance during the period between 1997 to 2001. The sample includes information obtained from administrative records and telephone interviews with unemployed workers eligible for unemployment insurance. The data contains information on the recipient's exhaustee status, the date when the unemployment spell began and the date when the unemployment spell ended. With this information, we construct the duration of repeated spells. Table 4.1 shows the duration of the

[2]The data was originally collected by the Upjohn Institute and is available for purchase from their website.

Variable	Description	Mean	Std.dev.	Min	Median	Max
Duration	Length of the spell in days	273.98	300.52	10.00	145.00	1365.00
UI	= 1 if received unemployment benefit	0.69	0.46	0.00	1.00	1.00
Age	age of the unemployed worker	40.25	10.90	21.00	39.84	63.00
Perm	= 1 if permanent job before unemployment	0.78	0.41	0.00	1.00	1.00
Unemp	state rate of unemployment	4.52	0.92	2.20	4.60	6.90
Censored	= 1 if the observation is not censored	0.74	0.44	0.00	1.00	1.00

TABLE 4.2. *Variable description and summary statistics*

unemployment spells and the frequency distribution of spells, which ranges from 1 to 6 spells per worker. The sample offers a range of unemployment durations. While the average unemployment duration is more than one year for workers who were unemployed once in the period of analysis, it reduces to approximately two months for workers who have a maximum of six unemployment spells between 1997 and 2001.

The sample consists of 3329 workers and 5015 observations. In the regression analysis that we will perform later in this section, we restrict attention to workers who were unemployed between 3 and 6 times in the period of analysis. This leads to a data set of 1358 observations on multiple unemployment spells of 394 workers. Moreover, we limit the sample to include workers between the ages of 20 and 65. We have information on the duration of the spells, whether the worker received unemployment benefits, the age of the worker at the beginning of the unemployment spell, whether the worker had a permanent employment before unemployment, and the state where the worker collects her benefits. Variable definitions and sample means are introduced in Table 4.2.

Workers generally receive unemployment insurance in their first unemployment spell, and the proportion of workers receiving benefits decreases dramatically after the first spell. The workers considered in this sample do not appear to necessarily be part-time workers, since approximately 78 percent lost a permanent job before unemployment. We also have information on whether the observation is censored. As shown in Table 4.2, the sample includes 26 percent of censored observations.

Preliminary evidence on the effect of unemployment benefits on the duration of unemployment is offered in Figure 4.1. We classify states into groups according to the definition of the U.S. Census Bureau. The evidence presented by regions allows us to examine if important regional

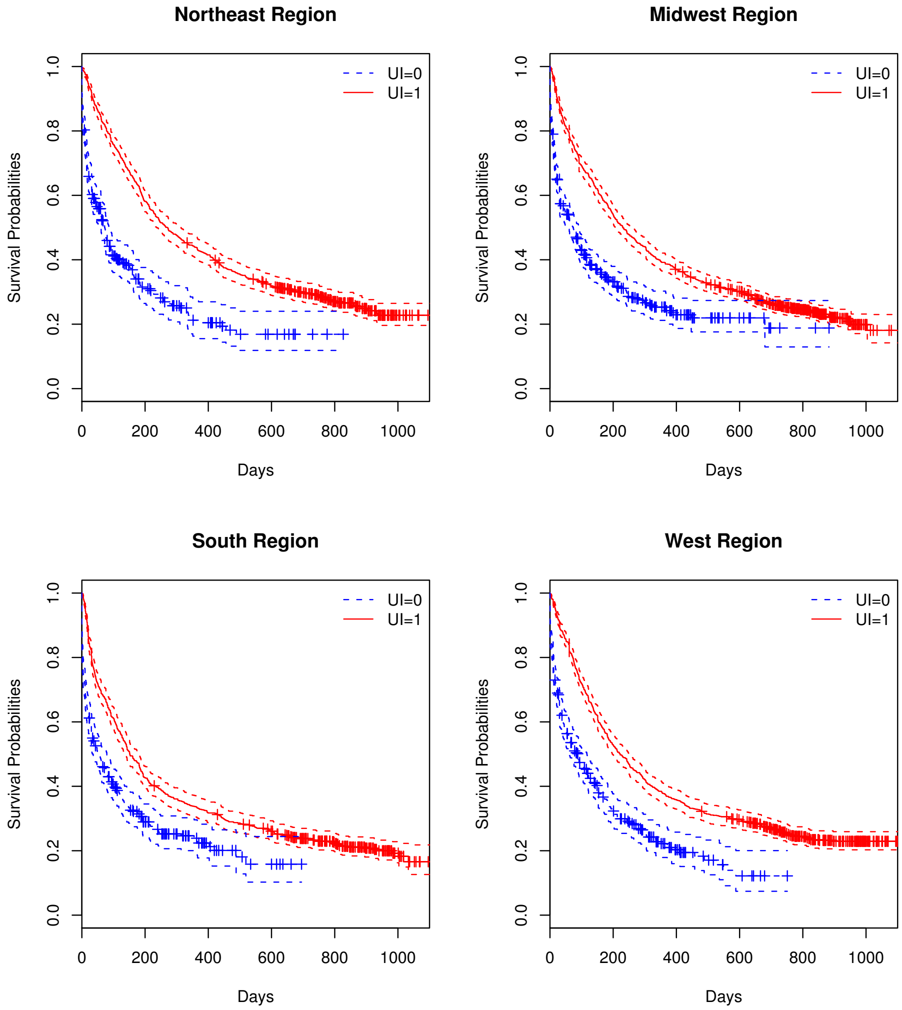

FIGURE 4.1. *Survival functions for the duration of unemployment. The dashed lines represent 95% confidence intervals.*

differences of the effect of unemployment benefits on the duration of unemployment may exist across regions. The figures present Kaplan-Meier estimators of survival functions for the number of days of unemployment, separately for workers who receive unemployment insurance ($UI = 1$) and workers who do not receive unemployment insurance ($UI = 0$). The effect of unemployment benefit on short unemployment durations appears to be important and significantly different for workers not receiving unemployment insurance and workers receiving unemployment insurance.

The effect of unemployment benefits tends to decrease at higher unemployment durations. The effect, however, remains statistically significant at standard levels. The evidence also suggests that regional differences in terms of the duration of unemployment and the effect of unemployment benefits are relatively minor quantitatively. The South exhibits the smallest difference in the estimated distributions between workers who receive unemployment insurance and those who do not.

4.2. **Empirical Results**

We estimate the following panel data duration model:

$$\log(T_{ij}) = \boldsymbol{d}_{ij}'\boldsymbol{\gamma} + \boldsymbol{x}_{ij}'\boldsymbol{\beta} + v_{ij}. \tag{4.1}$$

The vector $\boldsymbol{d}_{ij}$ includes an indicator variable for whether the worker received unemployment benefits (UI), the age of the worker at the beginning of the unemployment spell (Age), and an indicator variable for whether the worker had a permanent job before unemployment (Perm). The vector $\boldsymbol{x}_{ij}$ includes the rate of unemployment in the state where the worker collects unemployment benefits and controls for seasonal effects (i.e., variables indicating the month of the year). In models without individual effects, the vector $\boldsymbol{x}_{ij}$ includes indicator variables to control for the geographical region where the unemployment benefits were collected. These variables were constructed following the definition of the U.S. Census Bureau. In some applications it may also be advisable to introduce region and/or state effects but here it creates challenges for estimating models since we are faced with both state-invariant variables and individual-specific variables and in our sample, workers collected benefits in only one geographic location during the period of analysis.

Table 4.3 presents regression results derived from estimating several different econometric models. The statistical baseline is given by the accelerated failure time (AFT) model, where $v_{ij} = \alpha_i + u_{ij}$ and u_{ij} is distributed as Weibull. We assume that α_i models unobserved heterogeneity which is distributed as Gamma.

Table 4.3 also presents the results from three quantile regression models that were estimated at the conditional median. The method labeled NAIVE is a quantile regression estimator that ignores individual heterogeneity possibly correlated with the independent variables $\boldsymbol{d}_{ij}$ and censoring. The method labeled Peng-Huang is the quantile regression approach proposed by Peng and Huang (2008) for the survival model, and the method labeled Panel is the panel quantile approach developed in the previous sections.

Variables	AFT	NAIVE	Peng-Huang		Panel		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Receiving unemployment insurance	0.203	0.696	0.448	0.451	1.832	1.019	0.550
	(0.137)	(0.178)	(0.152)	(0.204)	(1.332)	(0.752)	(0.196)
Age	0.029	0.019	0.019	0.017	0.013	0.017	0.235
	(0.005)	(0.008)	(0.010)	(0.010)	(0.009)	(0.005)	(0.113)
Permanent employee before unemployment	0.382	0.144	0.352	0.375	0.301	0.235	0.783
	(0.129)	(0.169)	(0.048)	(0.180)	(0.331)	(0.172)	(0.273)
Unemployment	0.483	0.227	0.214	0.185	0.255	0.052	0.360
	(0.050)	(0.105)	(0.130)	(0.181)	(0.379)	(0.059)	(0.226)
Instruments	No	No	No	No	Yes	Yes	No
Time Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Effects	No	No	No	Yes	Yes	No	No
Individual Effects	No	No	No	No	No	No	Yes

Table 4.3. *Results based on the accelerated failure time (AFT) model and quantile regression models estimated at the median. Standard errors are in parentheses.*

We estimate two versions of the model using Peng-Huang estimator. In column (3), we estimate a model without controlling for workers' latent heterogeneity. In column (4), we estimate a similar model which includes time and region effects to control for the small regional differences suggested in Figure 4.1. As expected, columns (3) and (4) do not offer significantly different point estimates. In columns (5), (6) and (7), we employ the proposed approach to estimate a model under different assumptions. Column (5) reports results from a regression that uses unemployment insurance eligibility to instrument the potentially endogenous variable corresponding to whether a worker received unemployment insurance or not. Workers are eligible if they can demonstrate they have lost their jobs through no fault of their own. In column (6), we use a Hausman-Taylor (1981) type instrument obtained within the model, by simply considering differences from worker-specific mean of the indicator variable for whether the worker received unemployment insurance. If the individual effects represent a location shift and u_{ij} and v_{ij} in equations 3.1 and 3.2 are independent, this strategy may successfully remove the source of endogeneity. Lastly in column (7), instead of using instruments, we estimate the model augmented by individual effects.

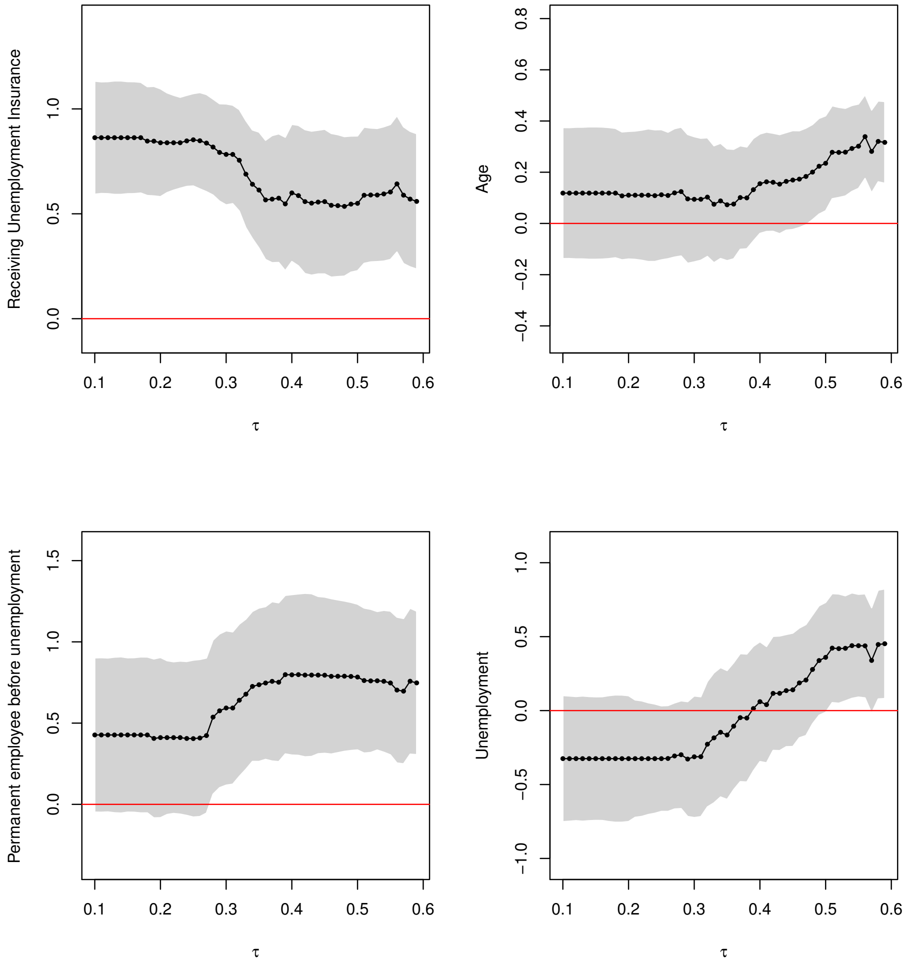

FIGURE 4.2. *Results for the panel duration model with fixed effects based on the quantile regression model. The grey areas represent 95% (pointwise) confidence intervals.*

All the variants of the models estimated in the table suggest that workers receiving unemployment benefits tend to be unemployed for a longer period of time than workers not receiving unemployment benefits. It is also interesting to see that the methods that treat the age of the worker and whether the worker had a permanent job before unemployment as independent of unobserved heterogeneity α, tend to report considerably different effects than the panel approach presented in the last column.

Using Figure 4.2, we extend the evidence presented in the last column of Table 4.3 beyond the median quantile. To obtain the results, we apply the method to estimate a panel-duration quantile version of equation 4.1:

$$Q_{\log(T_{ij})}(\tau|\boldsymbol{d}_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = \boldsymbol{d}'_{ij}\boldsymbol{\gamma}(\tau) + \boldsymbol{x}'_{ij}\boldsymbol{\beta}(\tau) + \boldsymbol{z}'_{ij}\boldsymbol{\alpha}(\tau). \quad (4.2)$$

for several quantiles τ in the grid (τ_0, τ_U). The quantile τ_0 was set to be zero and the quantile τ_U is a deterministic constant (Peng and Huang 2008), which is subject to some identifiability restrictions by the presence of censoring. In our application, the constant $\tau_U = 0.62$.

Our model is similar to equation 2.10 under the assumption that the independent variables are not correlated with the error term u_{ij}. Our quantile regression model conditions on individual heterogeneity, and therefore, we follow the traditional approach of allowing for dependence between unobserved heterogeneity α_i and the independent variable. This is an important concern in this application since individual effects may be correlated with receiving unemployment insurance, losing a permanent job, and the age at the time of unemployment. The figure displays estimates of the effects of interest as a function of the quantiles of the conditional distribution of unemployment duration. We report results for $\tau \in [0.1, 0.6]$.

Figure 4.2 indicates that the effect of receiving unemployment benefits is larger at the lower tail of the conditional unemployment duration distribution, and it tends to decrease as we go across quantiles. It is interesting to see that the effect of the age of the worker at the time of being unemployed and whether the worker had a permanent job before unemployment are significant factors at long (conditional) unemployment durations, while they do not appear to affect short (conditional) unemployment durations. On the other hand, the effect of state unemployment has different signs at the lower and upper tails of the conditional distribution of unemployment duration. While the evidence suggests that an increase in unemployment decreases short unemployment durations, it increases long unemployment durations. It should be noticed, however, that the effect is only statistically significant at standard levels at the upper tail of the conditional distribution.

5. Conclusion

This paper introduces a new quantile regression estimator for survival analysis when the dependent variable is censored. One of the main confounding factors in duration models is the presence of unobserved individual heterogeneity. The estimator proposed in this paper overcomes this problem by using data on multiple spells and using the panel data structure to account for correlated individual effects.

In the analysis of economic durations the applied economist is also often faced with the presence of additional variables which may be endogenous due to inherent selection issues, such as participation in a government program. Our proposed estimator shows how a two-step procedure using instrumental variables can be used to address this problem.

The estimator introduced in this paper has excellent finite sample performance and easily outperforms other estimators such as the Peng and Huang (2008) estimator which do not account for unobserved heterogeneity. The paper also discusses a realistic empirical application of the proposed methodology to the estimation of the effect of unemployment insurance on the duration of unemployment in the presence of unobserved individual effects. Our results indicate that controlling for unobserved individual effects produces qualitatively similar results, consistent with economic theory, but the estimated coefficients are quantitatively sufficiently different and may lead to different policy prescriptions.

References

Abbring, J. H., and G. J. Van Den Berg (2003): "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71(5), 1491–1517.

Audretsch, D., and T. Mahmood (1995): "New Firm Survival: New Results Using a Hazard Function," *Review of Economics and Statistics*, 77, 97–103.

Audretsch, D. B. (1991): "New-Firm Survival and the Technological Regime," *Review of Economics and Statistics*, 60, 441–450.

Blundell, R., and J. L. Powell (2007): "Censored Regression Quantiles with Endogenous Regressors," *Journal of Econometrics*, 141, 65–83.

Burda, M., M. Harding, and J. Hausman (2012): "Bayesian Nonparametric Duration Models," mimeo.

Chen, S., and S. Khan (2001): "Semiparametric Estimation of a Partially Linear Censored Regression Model," *Econometric Theory*, 17, 567–590.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–262.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491–525.

——— (2008): "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 142(1), 379–398.

Chesher, A. (2002): "Semiparametric Identification in Duration Models," CeMMAP Working Paper CWP20/02.

De Silva, D. G., G. Kosmopoulou, and C. Lamarche (2009): “The effect of information on the bidding and survival of entrants in procurement auctions,” *Journal of Public Economics*, 93, 56–72.

Eberwein, C., J. C. Ham, and R. J. Lalonde (1997): “The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data,” *The Review of Economic Studies*, 64(4), pp. 655–682.

Feng, X., X. He, and J. Hu (2011): “Wild bootstrap for quantile regression,” *Biometrika*, 98(4), 995–999.

Fitzenberger, B., and R. Wilke (2006): “Using Quantile Regression for Duration Analysis,” *Allgemeines Statistisches Archive*, 90(1), 105–120.

Flachaire, E. (2005): “Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap,” *Computational Statistics and Data Analysis*, 49(2), 361–376.

Fleming, T. R., and D. P. Harrington (1991): *Counting Processes and Survival Analysis*. John Wiley & Sons.

Geroski, P. A. (1995): “What do we know about entry,” *International Journal of Industrial Organization*, 13, 421–440.

Graham, B. S., J. Hahn, and J. L. Powell (2009): “The Incidental Parameter Problem in a Non-differentiable Panel Data Model,” *Economics Letters*, 105, 181–182.

Ham, J. C., and S. Rea (1987): “Unemployment Insurance and Male Unemployment Duration in Canada,” *Journal of Labor Economics*, 5(3), 325–353.

Han, A., and J. A. Hausman (1990): “Flexible parametric estimation of duration and competing risk models,” *Journal of Applied Econometrics*, 5(1), 1–28.

Harding, M., and C. Lamarche (2009): “A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables,” *Economics Letters*, 104, 133–135.

Hausman, J. A., and W. E. Taylor (1981): “Panel Data and Unobservable Individual Effects,” *Econometrica*, 49(6).

Hausman, J. A., and T. M. Woutersen (2005): “Estimating a Semiparametric Duration Model without Specifying Heterogeneity,” CeMMAP Working Paper CWP11/05.

Honoré, B. E., and A. de Paula (2010): “Interdependent Durations,” *Review of Economic Studies*, 77(3), 1138 – 1163.

Horowitz, J. L., and S. Lee (2004): “Semiparametric Estimation of a Panel Data Proportional Hazard Model with Fixed Effects,” *Journal of Econometrics*, 119, 155–198.

Jin, Z., Z. Ying, and L. J. Wei (2001): “A simple resampling method by perturbing the minimand,” *Biometrika*, 88(2), 381–390.

Khan, S., and E. Tamer (2007): “Partial Rank Estimation of Transformation Models with General Forms of Censoring,” *Journal of Econometrics*, 136, 251–280.

——— (2009): “Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities,” *Journal of Econometrics*, 152, 104–119.

Koenker, R. (2004): “Quantile Regression for Longitudinal Data,” *Journal of Multivariate Analysis*, 91, 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press.

——— (2008): “Censored Quantile Regression Redux,” *Journal of Statistical Software*, 27, 1–14.

——— (2010): "quantreg: An R Package for Quantile Regression," Version 4.57, http://CRAN.R-project.org/package=quantreg.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Koenker, R., and G. Bassett (1982): "Tests of Linear Hypothesis and ℓ_1 Estimation," *Econometrica*, 50, 1577–1584.

Koenker, R., and O. Geling (2001): "Reappraising medfly longevity: A quantile regression approach," *Journal of the American Statistical Association*, 96, 458–468.

Koenker, R., and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

Lancaster, T. (1985): "Simultaneous Equations Models in Applied Search Theory," *Journal of Econometrics*, 28, 113–126.

Lancaster, T. (1990): *The Econometric Analysis of Transition Data*. Cambridge University Press.

Lee, S. (2008): "Estimating Panel Data Duration Models with Censored Data," *Econometric Theory*, 24.

Ma, S., and M. R. Kosorok (2005): "Robust semiparametric M-estimation and the weighted bootstrap," *Journal of Multivariate Analysis*, 96(1), 190 – 217.

Meyer, B. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58(4), 757–782.

Olsen, R. J., and G. Farkas (1988): "Endogenous Covariates in Duration Models and the Effect of Adolescent Childbirth on Schooling," *Journal of Human Resources*, 24, 39–53.

Peng, L., and Y. Huang (2008): "Survival Analysis With Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.

Portnoy, S. (2003): "Censored Regression Quantiles," *Journal of The American Statistical Association*, 98, 1001–1010.

Wang, H. J., and M. Fygenson (2009): "Inference for Censored Quantile Regression in Longitudinal Studies," *Annals of Statistics*, 37, 756–781.