Endogenously Censored Median Regression with an Application to Benefit Elasticity of US Unemployment Duration [*]

Arkadiusz Szydłowski[†]

July 3, 2017

Abstract

We suggest a new inference method for median regression which is robust to assumptions on the censoring mechanism. Applying the method on the SIPP data we obtain that the benefit elasticity of median unemployment duration is between -1.1% and +7.9%.

Keywords: quantile regression, partial identification, unemployment duration, censoring

JEL codes: C21, J64, C41

1 Introduction

In economic duration analysis it is routinely assumed that the process which led to censoring of the observed duration is independent of unobserved characteristics. The objective of this paper is to examine the sensitivity of parameter estimates to this independence assumption in the context of a *median regression* for unemployment durations in the United States. We leave the distribution of censoring unrestricted, allowing it to be correlated both with observed and unobserved characteristics, which leads to partial identification of parameters of interest.

We observe a censored unemployment duration $Y_i = \min\{\widetilde{Y}_i, C_i\}$, where $\widetilde{Y}_i$ is the length of the uncensored spell and C_i is the censoring variable. Let $D_i = \mathbb{1}\{Y_i \leq C_i\}$ indicate observations which are not censored.

Consider a log-linear model of unemployment duration with median restriction:

$$\log \widetilde{Y}_i = X_i'\beta_0 + \varepsilon_i \qquad Med(\varepsilon_i|X_i) = 0. \tag{1}$$

where $X_i \in \mathbb{R}^K$. This model can be viewed as an alternative to the accelerated failure time (AFT) and mixed proportional hazard (MPH) models used in duration analysis. Latter models assume that the error term ε is statistically independent of the regressors X, whereas our model only imposes median independence, e.g. allowing for heteroskedastic errors.

Khan & Tamer (2009) analyze this model without putting exogeneity restrictions on the censoring mechanism. They allow C to be correlated both with X and ε, show that the slope coefficient β_0 in this model is in general not point-identified and characterize the identified set with conditional moment inequalities. However, they proceed with inference assuming that the support condition required for point identification is satisfied. The condition states that the support of the censoring variable is bounded below by $x'\beta_0$ for some values of $X = x$. This is a strong requirement. In our application, this would require that, for some levels of unemployment benefit, unemployment durations below the median cannot be censored.

Our paper does not put support conditions on the distribution of C and allows the slope coefficients to be partially identified. We apply the profiled procedure of Gandhi et al. (2013) (henceforth, GLS) to estimate confidence sets for the elasticity of unemployment duration with respect to unemployment benefit using SIPP data from the period 1985-2000. Similar inference procedure is used by Szydłowski (2017) to estimate the elasticity of unemployment *hazard rate* with respect to unemployment benefit using the same dataset but the MPH model for unemployment duration.[1]

Our confidence sets contain all values consistent with the log-linear model under a wide range of censoring distributions. In particular, they will contain the true parameter value even if the confidence interval from the random censoring model fails to do so. Combining estimates from different specifications we conclude that a 10% benefit hike will lead to a change in median unemployment duration between -1.1% to +7.9%. These estimates do not rule out a positive effect of benefits on unemployment durations. Such positive effect can be reconciled with a model where agents face liquidity constraints and job search is costly: a higher UI benefit may relax their liquidity

---

[1]This paper also argues why assuming exogenous censoring is problematic in the SIPP data.

constraint and induce more search, if this effect dominates the moral hazard effect (i.e. reduced incentive to search due to the conditional nature of unemployment benefits), the resulting change in the probability of finding a job will be positive.

2 Identification

Define:

$$Y_i^u = D_i Y_i + (1 - D_i) \times \infty$$

(with $0 \times \infty = 0$). Khan et al. (2011) show that the identified set for β_0 in this model is characterized by a set of moment inequalities. The moment inequalities are indexed by the values of the covariate vector $x \in \mathcal{X}$ and, to simplify the inference, we will assume that this set is discrete, i.e. $\mathcal{X} = \{x_1, x_2, \ldots, x_M\}$.[2] As a result, we obtain a finite set of moment inequalities:

$$\mu_m(\beta) \geq 0 \qquad m = 1, \ldots, 2M$$

where

$$\mu_m(\beta) = \begin{cases} x_m'\beta - Med(Y_i | X_i = x_m) & \text{for} \quad m = 1, \ldots, M \\ Med(Y_i^u | X_i = x_{m-M}) - x_{m-M}'\beta & \text{for} \quad m = M+1, \ldots, 2M \end{cases}$$

Let Θ_0 denote the set of β's satisfying these inequalities. Since the inequalities are linear in β, the identified set is convex.

3 Inference

We suggest a profiling procedure for building a confidence set for the coordinate(s) of the slope coefficient β. We focus on the elements of β, firstly, because the coordinates, not the full vector, are usually of interest in applications, and, secondly, because this leads to simpler computation.

---

[2]In application this will require discretizing the support of continuously distributed covariates. As shown by Szydłowski (2017), as long as the number of support points is relatively large, this should not have a significant effect on the size of the estimated confidence sets.

Partition the (compact and convex) parameter space $\Theta = \Theta_1 \times \Theta_{-1}$ and let:

$$\Theta_{0,1} = \{\beta_1 \in \Theta_1 : \quad \exists \beta_{-1} \in \Theta_{-1} \quad \text{such that} \quad (\beta_1, \beta_{-1}) \in \Theta_0\}.$$

Next, set:

$$\hat{\mu}_m(\beta) = \begin{cases} x'_m \beta - \widehat{Med}(Y_i|X_i = x_m) & \text{for} \quad m = 1,..,M \\ \widehat{Med}(Y_i^u|X_i = x_{m-M}) - x'_{m-M}\beta & \text{for} \quad m = M+1,..,2M \end{cases}$$

where:

$$\widehat{Med}(\cdot|X_i = x_m) = \inf\left\{y : \frac{\sum_{i=1}^n D_i \math1\{\cdot \le y, X_i = x_m\}}{\sum_{i=1}^n \math1\{X_i = x_m\}} \ge \frac{1}{2}\right\}$$

The profiled test statistic is given by:

$$T_n(\beta_1) = \min_{\beta_{-1} \in \Theta_{-1}} \sum_{m=1}^{2M} [\sqrt{n}\hat{\mu}_m(\beta)]_-^2 - \min_{\beta \in \Theta} \sum_{m=1}^{2M} [\sqrt{n}\hat{\mu}_m(\beta)]_-^2$$

where $[\cdot]_- = \min\{\cdot, 0\}$, and the corresponding bootstrap statistic is:

$$T_{n,r}^*(\beta_1) = \min_{\beta_{-1} \in \Theta_{-1}} \sum_{m=1}^{2M} [\sqrt{n}\bar{\mu}_m^*(\beta) + \sqrt{\kappa_n}\hat{\mu}_m(\beta)]_-^2 - \min_{\beta \in \Theta} \sum_{m=1}^{2M} [\sqrt{n}\bar{\mu}_m^*(\beta) + \sqrt{\kappa_n}\hat{\mu}_m(\beta)]_-^2.$$

where $\bar{\mu}_m^*(\beta) = \mu_m^*(\beta) - \hat{\mu}_m(\beta)$ and $\mu_m^*(\beta)$ is the estimate of $\mu_m(\beta)$ calculated on the bootstrap sample (we sample from $\{Y_i, D_i, X_i\}_{i=1}^n$ with replacement).

Further, let $c_\tau^*(\beta_1)$ be the $1 - \tau$ quantile of $T_{n,r}^*(\beta_1)$ across bootstrap samples, $r = 1, 2, \ldots, R$, and define:

$$Q(\beta) = \sum_{m=1}^{2M} [\mu_m(\beta)]_-^2.$$

We make the following assumptions in order to apply the results in GLS:

Assumption 1. *(a) Θ_0 is a strict subset of Θ.*

4

(b) For any $\beta_1 \in \Theta_{0,1}$ there exist constants $C, c > 0$ such that:

$$Q(\beta) \geq C \left\{ \inf_{b \in \Theta_0} \|\beta - b\|^2 \wedge c \right\}$$

for all $\beta \in \Theta$ such that $\beta = (\beta_1, \beta_{-1})$.

Assumption 2. $\kappa_n \to \infty$, $\kappa_n/n \to 0$.

Assumption 3. *For all m, the densities of Y and Y^u given $X = x_m$ exist and are positive and continuous in the neighborhoods of $y_{med} = Med(Y|X = x_m)$ and $y^u_{med} = Med(Y^u|X = x_m)$, respectively.*

A set of sufficient conditions for Assumption 3 is:

(i) $P(D_i = 1|X_i = x_m) \geq \frac{1}{2}$

(ii) ε_i's are absolutely continuously distributed given $X_i = x_m$ and the conditional density functions $f_\varepsilon(\cdot|X_i = x_m)$ are bounded away from zero in the neighborhoods of $\underline{\varepsilon} = Med(Y_i|X_i = x_m) - x'_m \beta_0$ and $\bar{\varepsilon} = Med(Y^u_i|X_i = x_m) - x'_m \beta_0$

for all m. The first condition requires limited amount of censoring and is satisfied in our application. The second condition imposes mild smoothness restrictions on the distribution of the error term.

The following theorem establishes that bootstrap provides a correct critical value for the test. Its proof is an application of Theorem 2 in GLS.

Theorem 1. *Let $\beta_1 \in \Theta_{0,1}$, $\{Y_i, D_i, X_i\}_{i=1}^n$ be an i.i.d. sample from (1) and Assumptions 1 - 3 hold. Then:*

$$\liminf_{n \to \infty} P(T_n(\beta_1) \leq c^*_\tau(\beta_1)) \geq 1 - \tau.$$

if $c^*_\tau(\beta_1)$ is the continuity point of the asymptotic distribution of $T_n(\beta_1)$.

Proof. First, note that our problem is simpler than the one in GLS, since we discretize $\mathcal{X}$ and do not have to integrate over the "boxes" of values of X. Since our moment conditions are linear in β, the smoothness conditions from Assumption C.2 in GLS are automatically satisfied. By Assumption 3

and Theorem 2.3.3.A in Serfling (1980), we have:

$$\sqrt{n}(\widehat{Med}(Y_i|X_i = x_m) - Med(Y_i|X_i = x_m)) \to^D N\left(0, \left(2\frac{\partial P(Y_i \le y|X = x_m)}{\partial y}\bigg|_{y=y_{med}}\right)^{-2}\right)$$

$$\sqrt{n}(\widehat{Med}(Y_i^u|X_i = x_m) - Med(Y_i^u|X_i = x_m)) \to^D N\left(0, \left(2\frac{\partial P(Y_i^u \le y|X = x_m)}{\partial y}\bigg|_{y=y_{med}^u}\right)^{-2}\right)$$

for all m and equivalent bootstrap version of this convergence result follows by standard arguments, which implies that a condition corresponding to Assumptions C.2(c)-(d) in GLS is satisfied. The rest of their assumptions are either trivially satisfied or are implied by our Assumptions 1-2. $\qquad\square$

4   Monte Carlo simulations

The aim of performing Monte Carlo simulations in our setup is twofold. Firstly, it gives us an insight about finite sample performance of our inference procedure. Secondly, it assists us in picking the tuning sequence κ_n needed for the application.

The data is generated from the following model:

$$\log \widetilde{Y}_i = \beta_0 X_i + (\log U_i - \log V_i)e^{\delta X_i}$$

$$\log C_i = c + \beta_0 X_i + \rho \log V_i e^{\delta X_i}$$

where V_i, U_i are exponentially distributed (mutually independent and independent of X), with $\beta_0 = 0.33$ and the following parameters: $\delta = \rho = 0$ in Design 1, $\delta = 0, \rho = 1$ in Design 2, $\delta = 0.25, \rho = 1$ in Design 3. Constant c is set to obtain around 22% censoring rate (same as in our SIPP data). The covariate X_i is drawn from a uniform distribution on $\mathcal{X} = \{0, 1/M, 2/M, \ldots, 59/M\}$ where $M = 60$ or $M = 30$. Censoring is exogenous in Design 1 so β is point-identified. Design 3 features heteroskedastic errors.

We set $\kappa_n = n/(\kappa \log(n)), \kappa = 0.5, 1, 1.5$ and consider different values of β to check if our test controls size correctly in the interior and at the boundary of the identified set and to examine its power properties. The boundary values of the identified set were calculated with $M = 60$ so simulations for $M = 30$ help us evaluate the power loss from introducing coarse discretization.

Table 1: Results of Monte Carlo simulations

M = 60

Design 1 (point-identified)

	n = 4000 coverage interior		boundary		1 - power		n = 8000 coverage interior		boundary		1 - power	
	β = 0.33				β = 0.5		β = 0.33				β = 0.5	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
κ = 0.5	0.97	0.99			0.40	0.63	0.97	0.99			0.09	0.21
κ = 1	0.95	0.98			0.21	0.40	0.94	0.98			0.03	0.08
κ = 1.5	0.94	0.97			0.18	0.31	0.92	0.97			0.02	0.05

Design 2 (partially-identified)

	interior		boundary		1 - power		interior		boundary		1 - power	
	β = 0.33		β = 0.334		β = 0.5		β = 0.33		β = 0.334		β = 0.5	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
κ = 0.5	0.98	0.99	0.98	0.99	0.61	0.80	0.97	0.99	0.97	0.99	0.31	0.54
κ = 1	0.96	0.99	0.96	0.99	0.40	0.59	0.95	0.99	0.95	0.98	0.15	0.29
κ = 1.5	0.95	0.98	0.95	0.98	0.34	0.51	0.95	0.98	0.94	0.98	0.11	0.21

Design 3 (partially-identified)

	interior		boundary		1 - power		interior		boundary		1 - power	
	β = 0.33		β = 0.345		β = 0.5		β = 0.33		β = 0.345		β = 0.5	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
κ = 0.5	0.97	0.99	0.97	0.99	0.57	0.77	0.98	1.00	0.97	0.99	0.37	0.59
κ = 1	0.94	0.99	0.92	0.97	0.31	0.50	0.97	0.99	0.94	0.98	0.13	0.28
κ = 1.5	0.93	0.97	0.90	0.95	0.22	0.39	0.95	0.98	0.92	0.96	0.08	0.17

M = 30

Design 1 (point-identified)

	interior		boundary		1 - power		interior		boundary		1 - power	
	β = 0.33		β = 0.33		β = 0.5		β = 0.33		β = 0.33		β = 0.5	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
κ = 0.5	0.97	1.00			0.40	0.63	0.97	0.99			0.09	0.22
κ = 1	0.94	0.98			0.22	0.40	0.94	0.98			0.03	0.09
κ = 1.5	0.93	0.97			0.19	0.33	0.93	0.97			0.03	0.06

Design 2 (partially-identified)

	interior		boundary		1 - power		interior		boundary		1 - power	
	β = 0.33		β = 0.334		β = 0.5		β = 0.33		β = 0.334		β = 0.5	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
κ = 0.5	0.97	0.99	0.97	0.99	0.68	0.83	0.96	0.99	0.96	0.99	0.36	0.58
κ = 1	0.96	0.99	0.96	0.98	0.48	0.69	0.94	0.97	0.95	0.97	0.20	0.35
κ = 1.5	0.95	0.98	0.95	0.98	0.41	0.61	0.93	0.97	0.94	0.97	0.16	0.27

Design 3 (partially-identified)

	interior		boundary		1 - power		interior		boundary		1 - power	
	β = 0.33		β = 0.345		β = 0.5		β = 0.33		β = 0.345		β = 0.5	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
κ = 0.5	0.97	0.99	0.98	1.00	0.67	0.83	0.97	0.99	0.98	0.99	0.46	0.67
κ = 1	0.96	0.98	0.94	0.98	0.43	0.63	0.96	0.98	0.95	0.98	0.21	0.40
κ = 1.5	0.94	0.98	0.93	0.97	0.33	0.53	0.95	0.98	0.94	0.97	0.15	0.28

Note: 1000 Monte Carlo simulations, 500 bootstrap replications. The column "1 - power" gives the probability that the value outside the identified set is included in the confidence set.

The results in Table 1 show that the coverage probabilities for values inside the identified set (Designs 2-3) are close to one and increase with the sample size. The coverage at the boundary is fairly close to the nominal values for $\kappa = 1.5$ and other values of κ lead to conservative inference, therefore we choose the former value for our application. Finally, the power of our test improves significantly with the sample size. We also see that power of the test decreases for $M = 30$, showing

that too coarse discretization may enlarge the estimated confidence sets.

5   Unemployment benefits and unemployment durations

We use an altered version of SIPP 1985-2000 data analyzed in Chetty (2008). We include following covariates: logarithm of the weekly unemployment benefit (defined as the average state unemployment benefit in the year of entry into the unemployment pool), pre-unemployment annual wage, average state unemployment rate in years 1985-2000, age and a marital status dummy.[3] Table 2 shows the estimated confidence sets under two different model specifications and compares them with confidence intervals obtained under the assumption of random censoring, where the latter were calculated by bootstrapping the RCLAD estimator of Honoré et al. (2002). Estimates from Model (1) imply that a 10% increase in the weekly unemployment benefit, which corresponds to a \$16 hike on average, raises the median unemployment duration by 0.6-7.9%. Thus, even allowing for endogenous censoring we can conclude that raising unemployment benefits leads to longer unemployment spells. However, the results from Model (2) do not confirm this conclusion and suggest that the elasticity of unemployment duration can be slightly negative. Combining estimates from both models by taking the largest confidence interval that contains estimates under both specifications, we can conclude that this elasticity falls between -0.11 and 0.79.

Table 2: Confidence sets for the elasticity of unemployment duration with respect to unemployment benefit, 90% level

	Model (1)	Model (2)
log UI benefit	**[0.06,0.79]**	**[-0.11,0.56]**
log annual wage	x	x
average unemployment rate	x	x
age		x
married		x
discretization	$10 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$
n	3986	3986
CI under random censoring	[0.08,0.51]	[0.19,0.53]

Note: The row "discretization" gives the number of discrete values of the variables included in the model in the order they appear in the rows of the table, e.g. $10 \times 2 \times 3$ means 10 values of log UI benefit, 2 values of log annual wage, 3 values of unemployment rate. The number of bootstrap replications is 500 and $\kappa = 1.5$.

---

[3]State unemployment rate controls for local labor market conditions. As shown in Szydłowski (2017) controlling for state unemployment rate is equivalent to including state dummies but leads to more parsimonious specification of conditional medians, thus more precise estimates. See that article for further motivation for our choice of covariates.

In both specifications the confidence sets contain the corresponding confidence intervals under random censoring. Thus, we cannot reject the hypothesis that censoring is exogenous in our dataset. Still, our estimates are robust to violations of this assumption. If censoring is endogenous, the true parameter value may lie outside the CI under random censoring but it will always be included in our confidence interval.

6   Conclusion

We focused on median unemployment duration above. It would be interesting to see if the effect of unemployment benefits differs for short- and long-term unemployed (e.g. first and third quartile). This can be done by adding moment inequalities corresponding to these quantiles to our profiled statistic and checking if there is a β_1 that satisfies these inequalities for every quartile.

References

Chetty, R. (2008), 'Moral hazard versus liquidity and optimal unemployment insurance', *Journal of Political Economy* **116**(2), pp. 173–234.

Gandhi, A., Lu, Z. & Shi, X. (2013), Estimating demand for differentiated products with error in market shares. Working paper.

Honoré, B., Khan, S. & Powell, J. L. (2002), 'Quantile regression under random censoring', *Journal of Econometrics* **109**(1), 67 – 105.

Khan, S., Ponomareva, M. & Tamer, E. (2011), 'Sharpness in randomly censored linear models', *Economics Letters* **113**(1), 23–25.

Khan, S. & Tamer, E. (2009), 'Inference on endogenously censored regression models using conditional moment inequalities', *Journal of Econometrics* **152**(2), 104–119.

Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Wiley, Chichester; New York.

Szydłowski, A. (2017), Endogenous censoring in the mixed proportional hazard model with an application to optimal unemployment insurance. Working Paper.