Oaxaca-Blinder type Decomposition Methods for Duration Outcomes

Andres Garcia-Suaza
garciasuaza@gmail.com

Univerisidad del Rosario

Seminar of the PhD in Mathematical Engineering
Universidad EAFIT
Medellin, April 2017

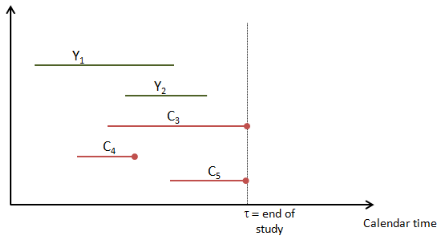
Decomposition Methods

- Study difference of distributional features between two populations:

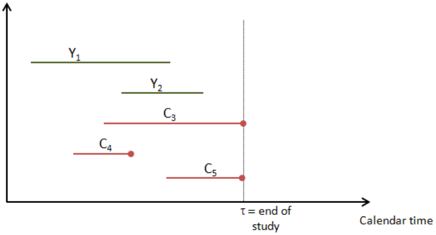
  Total difference = Structure Effect + Composition Effect

- Decomposition of the mean: Oaxaca (1973) and Blinder (1973) -OB-.
- Extensions OB decomposition:
  - Variance and Inequality: Juhn et. al. (1992).
  - Quantile: Machado and Mata (2005), Melly (2005).
  - Distribution and its functionals: DiNardo et. al. (1996), Chernozhukov et. al. (2013) -CFM-.
  - Non-linear models: Bauer and Sinning (2008).

Censored Data: Duration Outcomes



Censored Data: Duration Outcomes



- Some examples in economics: Unemployment duration, employment duration, firms lifetime, school dropout, ...
- Previous methods no valid for censored data.
- Decomposition of average hazard rate.

Oaxaca-Blinder type Decomposition Methods for Duration Outcomes

Our goals

- Propose decomposition methods for censored outcomes.
  - Mean decomposition.
  - Decomposition of other parameters through the estimation of the whole outcome distribution.
- Discuss the effect of neglecting the presence of censoring and the role of the censoring mechanism assumptions.
- Analyze factors explaining unemployment gender gaps using Spanish data for 2004-2007.

Outline

Outline

Suppose we are interested in the average difference of an outcome Y between two groups, denoted by $\ell \in \{0, 1\}$ (e.g. 0 men and 1 women).

$$\Delta_Y^{\mu} = \mu_Y^{(1)} - \mu_Y^{(0)}$$

Suppose we are interested in the average difference of an outcome Y between two groups, denoted by $\ell \in \{0, 1\}$ (e.g. 0 men and 1 women).

Let X be a vector of characteristics of the population. The difference can be expressed in terms of the best linear predictor:

$$\begin{aligned}\Delta_Y^{\mu} &= \mu_Y^{(1)} - \mu_Y^{(0)} \\ &= \beta_1^T \mu_X^{(1)} - \beta_0^T \mu_X^{(0)}\end{aligned}$$

Suppose we are interested in the average difference of an outcome Y between two groups, denoted by $\ell \in \{0,1\}$ (e.g. 0 men and 1 women).

Let X be a vector of characteristics of the population. The difference can be expressed in terms of the best linear predictor:

$$\begin{aligned} \Delta_Y^{\mu} &= \mu_Y^{(1)} - \mu_Y^{(0)} \\ &= \beta_1^T \mu_X^{(1)} - \beta_0^T \mu_X^{(0)} \end{aligned}$$

where, $\beta_\ell^T \mu_X^{(\ell)}$ is the best linear predictor of $\mu_Y^{(\ell)}$, and

$$\beta_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \quad \mathbb{E}\left(Y - b^T X \mid D = \ell\right)^2$$

OB Decomposition

Mean Difference Decomposition

Suppose we are interested in the average difference of an outcome Y between two groups, denoted by $\ell \in \{0, 1\}$ (e.g. 0 men and 1 women).

Let X be a vector of characteristics of the population. The difference can be expressed in terms of the best linear predictor:

$$\begin{aligned} \Delta_Y^{\mu} &= \mu_Y^{(1)} - \mu_Y^{(0)} \\ &= \beta_1^T \mu_X^{(1)} - \beta_0^T \mu_X^{(0)} \end{aligned}$$

Note that $\mathbb{E}\left(\beta^T X \mid D = \ell\right) = \beta_\ell^T \mu_X^{(\ell)}$, but it does not hold for the conditional hazard!.

Suppose we are interested in the average difference of an outcome Y between two groups, denoted by $\ell \in \{0,1\}$ (e.g. 0 men and 1 women).

Let X be a vector of characteristics of the population. The difference can be expressed in terms of the best linear predictor:

$$\begin{aligned}\Delta_Y^{\mu} &= \mu_Y^{(1)} - \mu_Y^{(0)} \\ &= \beta_1^T \mu_X^{(1)} - \beta_0^T \mu_X^{(0)}\end{aligned}$$

Rearranging terms:

$$\Delta_Y^{\mu} = \underbrace{(\beta_1 - \beta_0)^T \mu_X^{(1)}}_{\text{Structure effect}} + \underbrace{\beta_0^T \left(\mu_X^{(1)} - \mu_X^{(0)}\right)}_{\text{Composition effect}}$$

Identification

OB Decomposition

Estimation

Given a sample $\{Y_i, X_i, D_i\}_{i=1}^n$, OB decomposition can be estimated as:

$$\bar{\Delta}_Y^{\mu} = \left(\bar{\beta}_1 - \bar{\beta}_0\right)^T \bar{\mu}_X^{(1)} + \bar{\beta}_0^T \left(\bar{\mu}_X^{(1)} - \bar{\mu}_X^{(0)}\right)$$

where,

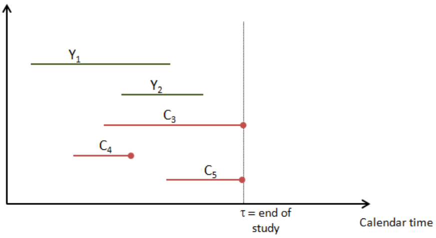
$$\bar{\mu}_X^{(\ell)} = n_\ell^{-1} \sum_{i=1}^{n} X_i 1_{\{D_i=\ell\}},$$

and

$$\bar{\beta}_\ell = \arg\min_{b \in \mathbb{R}^k} \sum_{i=1}^{n} \left(Y - b^T X\right)^2 1_{\{D_i=\ell\}}$$

with $n_\ell = \sum_{i=1}^{n} 1_{\{D_i=\ell\}}$.

OB Decomposition under Censoring
Data Structure



- Let Y a non-negative random variable denoting duration, e.g., unemployment duration.
- X is a set of relevant covariates.
- We observe $Z = \min(Y, C)$, and $\delta = 1_{\{Y \leq C\}}$.
- Hence, instead of (Y, X, D, C), we observe (Z, X, D, δ).

OB Decomposition under Censoring

Consider the joint distribution of (Y, X, D), given by

$$F(y, x, \ell) = \mathbb{P}(Y \leq y, X \leq x, D = \ell)$$

Note that:

$$\mu_Y^{(\ell)} = \int y dF(y, \infty, \ell) \qquad \text{and} \qquad \mu_X^{(\ell)} = \int x dF(\infty, x, \ell)$$

with,

$$\beta_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \int \left(y - b^T x\right)^2 dF(y, x, \ell).$$

OB Decomposition under Censoring

Consider the joint distribution of (Y, X, D), given by

$$F(y, x, \ell) = \mathbb{P}(Y \leq y, X \leq x, D = \ell)$$

Note that:

$$\mu_Y^{(\ell)} = \int y dF(y, \infty, \ell) \qquad \text{and} \qquad \mu_X^{(\ell)} = \int x dF(\infty, x, \ell)$$

with,

$$\beta_\ell = \arg\min_{b \in \mathbb{R}^k} \int \left(y - b^T x\right)^2 dF(y, x, \ell).$$

Therefore, $\bar{\Delta}_Y^\mu$ is the empirical analog of Δ_Y^μ when F is replaced by its sample version

$$\bar{F}(y, x, \ell) = n_\ell^{-1} \sum_{i=1}^{n} 1_{\{Y_i \leq y, X_i \leq x, D_i = \ell\}}$$

Consider the following probability

$$\begin{aligned}\mathbb{P}\left(y \leq Y < y + h, X \in B \mid Y \geq y, D = \ell\right) &= \frac{\mathbb{P}\left(y \leq Y \leq y + h, X \in B, D = \ell\right)}{\mathbb{P}\left(Y \geq y, D = \ell\right)} \\ &= \int_{\{X \in B\}} \frac{F\left(y + h, dx, \ell\right) - F\left(y-, dx, \ell\right)}{1 - F\left(y-, \infty, \ell\right)}\end{aligned}$$

Therefore, the cumulative hazard can be defined as

$$\Lambda\left(y, x, \ell\right) = \int_0^y \frac{F\left(d\bar{y}, x, \ell\right)}{1 - F\left(\bar{y}-, \infty, \ell\right)}$$

Consider the following probability

$$\begin{aligned}\mathbb{P}\left(y \leq Y < y+h, X \in B \mid Y \geq y, D=\ell\right) &= \frac{\mathbb{P}\left(y \leq Y \leq y+h, X \in B, D=\ell\right)}{\mathbb{P}\left(Y \geq y, D=\ell\right)} \\ &= \int_{\{X \in B\}} \frac{F\left(y+h, dx, \ell\right) - F\left(y-, dx, \ell\right)}{1 - F\left(y-, \infty, \ell\right)}\end{aligned}$$

Therefore, the cumulative hazard can be defined as

$$\Lambda\left(y, x, \ell\right) = \int_0^y \frac{F\left(d\bar{y}, x, \ell\right)}{1 - F\left(\bar{y}-, \infty, \ell\right)}$$

$$F\left(y, x, \ell\right) = 1 - \exp\left\{-\Lambda^C\left(y, x, \ell\right)\right\} \prod_{\bar{y} \leq y} \left[1 - \Lambda\left(\left\{\bar{y}\right\}, x, \ell\right)\right]$$

Multivariate Kaplan-Meier Estimator

Identification of the Multivariate Distribution

Assumption 2

a. $\mathbb{P}(Y \leq y, C \leq c \mid D = \ell) = \mathbb{P}(Y \leq y \mid D = \ell)\mathbb{P}(C \leq c \mid D = \ell)$.

b. $\mathbb{P}(Y \leq C \mid Y, X, D) = \mathbb{P}(Y \leq C \mid Y, D)$.

Assumption 2

a. $\mathbb{P}(Y \leq y, C \leq c \mid D = \ell) = \mathbb{P}(Y \leq y \mid D = \ell)\mathbb{P}(C \leq c \mid D = \ell)$.

b. $\mathbb{P}(Y \leq C \mid Y, X, D) = \mathbb{P}(Y \leq C \mid Y, D)$.

Proposition 1

Under Assumption 2, the joint cumulative hazard function can be writen as:

$$\Lambda(y, x, \ell) = \int_0^y \frac{H_{11}(d\bar{y}, x, \ell)}{1 - H(\bar{y}-, \ell)}$$

where,

$$H_{11}(y, x, \ell) = \mathbb{P}(Z \leq y, X \leq x, D = \ell, \delta = 1)$$

and,

$$H(y, \ell) = \mathbb{P}(Z \leq y, \delta = 1)$$

Details

Multivariate Kaplan-Meier Estimator

Equivalently, $\hat{F}$ can be written as:

$$\hat{F}(y, x, \ell) = \sum_{i=1}^{n_\ell} W_i^{(\ell)} 1_{\left\{ Z_{i:n_\ell}^{(\ell)} \leq y, X_{[i:n_\ell]}^{(\ell)} \leq x \right\}}$$

where,

$$W_i^{(\ell)} = \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell - R_i^{(\ell)} + 1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_{[j:n_\ell]}^{(\ell)}}{n_\ell - R_j^{(\ell)} + 1} \right)$$

with $Z_{i:n_\ell}^{(\ell)} = Z_j$ if $R_j^{(\ell)} = i$, and for any $\{\xi_i\}_{i=1}^{n_\ell}$, $\xi_{[i:n_\ell]}^{(\ell)}$ is the i-th concomitant of $Z_{i:n_\ell}^{(\ell)}$, i.e., $\xi_{[i:n_\ell]}^{(\ell)} = \xi_j$ if $Z_{i:n_\ell}^{(\ell)} = Z_j$.

Equivalently, $\hat{F}$ can be written as:

$$\hat{F}(y,x,\ell) = \sum_{i=1}^{n_\ell} W_i^{(\ell)} 1_{\left\{Z_{i:n_\ell}^{(\ell)} \leq y, X_{[i:n_\ell]}^{(\ell)} \leq x\right\}}$$

where,

$$W_i^{(\ell)} = \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell - R_i^{(\ell)} + 1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_{[j:n_\ell]}^{(\ell)}}{n_\ell - R_j^{(\ell)} + 1}\right)$$

In absence of censoring: $\delta_{[i:n_\ell]}^{(\ell)} = 1 \; \forall i$, and hence, $W_i^{(\ell)} = n_\ell^{-1}$.

An example

(Censored) OB Decompisition -*COB*-

Pluging-in $\hat{F}$ we have:

$$\hat{\Delta}_Y^{\mu} = \left(\hat{\beta}_1 - \hat{\beta}_0\right)^T \hat{\mu}_X^{(1)} + \hat{\beta}_0^T \left(\hat{\mu}_X^{(1)} - \hat{\mu}_X^{(0)}\right)$$

where $\hat{\mu}_X^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)}$ and

$$\hat{\beta}_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n_\ell} W_i^{(\ell)} \left(Z_{i:n_\ell}^{(\ell)} - b^T X_{[i:n_\ell]}^{(\ell)}\right)^2$$

Inference

- Inference: asymptotic variance and bootstrap are suitable. Bootstrap
- In absence of censoring, classical results are obtained.
- Monte Carlo simulations show a fairly good performance.
- What if mean is not informative?

Outline

Counterfactual Distributions

A convenient form to write the marginal distribution of the outcome for $D = \ell$ is given by:

$$F_Y^{(\ell,\ell)}(y) = \mathbb{E}\left[1_{\left\{Y_i^{(\ell)} \leq y\right\}}\right]$$

Counterfactual Distributions

A convenient form to write the marginal distribution of the outcome for $D = \ell$ is given by:

$$\begin{aligned} F_Y^{(\ell,\ell)}(y) &= \mathbb{E}\left[1_{\left\{Y_i^{(\ell)} \leq y\right\}}\right] \\ &= \mathbb{E}\left[F^{(\ell)}(y|x) \mid D = \ell\right] \\ &= \int F^{(\ell)}(y|x)\, dF^{(\ell)}(x) \end{aligned}$$

with $F^{(\ell)}(x) = \mathbb{P}(X \leq x \mid D = \ell)$.

Counterfactual Distributions

Define the counterfactual outcome $Y^{(i,j)}$, which follows a distribution function given by $F_Y^{(i,j)}(y)$. Consider the counterfactual *operator* distribution (Rothe -2010- and Chernozhuvok et. al. -2013-):

$$
\begin{aligned}
F_Y^{(i,j)}(y) &= \mathbb{E}\left[F^{(i)}(y|x) \mid D = j\right] \\
&= \int F^{(i)}(y|x)\, dF^{(j)}(x)
\end{aligned}
$$

Counterfactual Distributions

Define the counterfactual outcome $Y^{(i,j)}$, which follows a distribution function given by $F_Y^{(i,j)}(y)$. Consider the counterfactual *operator* distribution (Rothe -2010- and Chernozhuvok et. al. -2013-):

$$
\begin{aligned}
F_Y^{(i,j)}(y) &= \mathbb{E}\left[F^{(i)}(y|x) \mid D=j\right] \\
&= \int F^{(i)}(y|x)\, dF^{(j)}(x)
\end{aligned}
$$

Thus, if $F_Y^{(i,j)}(y)$ is identifiable, for any parameter $\theta\left(F_Y^{(i,j)}\right)$ we have:

$$\Delta_Y^{\theta} = \theta\left(F_Y^{(1,1)}\right) - \theta\left(F_Y^{(0,0)}\right)$$

Define the counterfactual outcome $Y^{(i,j)}$, which follows a distribution function given by $F_Y^{(i,j)}(y)$. Consider the counterfactual *operator* distribution (Rothe -2010- and Chernozhuvok et. al. -2013-):

$$\begin{aligned} F_Y^{(i,j)}(y) &= \mathbb{E}\left[F^{(i)}(y|x) \mid D=j\right] \\ &= \int F^{(i)}(y|x)\, dF^{(j)}(x) \end{aligned}$$

Thus, if $F_Y^{(i,j)}(y)$ is identifiable, for any parameter $\theta\left(F_Y^{(i,j)}\right)$ we have:

$$\begin{aligned} \Delta_Y^{\theta} &= \theta\left(F_Y^{(1,1)}\right) - \theta\left(F_Y^{(0,1)}\right) + \theta\left(F_Y^{(0,1)}\right) - \theta\left(F_Y^{(0,0)}\right) \\ &= \Delta_S^{\theta} + \Delta_C^{\theta} \end{aligned}$$

By replacing the multivariate empirical distribution of covariates, given by

$$\bar{F}^{(\ell)}(x) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{X_i \leq x, D_i = \ell\}}$$

we have:

$$\bar{F}_Y^{(i,j)}(y) = n_j^{-1}\sum_{l=1}^{n} \bar{F}^{(i)}(y|x_l)\, 1_{\{D_l = \ell\}}$$

By replacing the multivariate empirical distribution of covariates, given by

$$\bar{F}^{(\ell)}(x) = n_\ell^{-1} \sum_{i=1}^{n} 1_{\{X_i \leq x, D_i = \ell\}}$$

we have:

$$\bar{F}_Y^{(i,j)}(y) = n_j^{-1} \sum_{l=1}^{n} \bar{F}^{(i)}(y|x_l)\, 1_{\{D_l = \ell\}}$$

How identify and estimate $F^{(\ell)}(y|x)$ under censoring?

Conditional Distribution of Duration Outcomes

Identification and Estimation

$F^{(\ell)}(y|x)$ is identifiable under Assumption 2, but weaker conditions are needed. Censoring

Assumption 4 (Conditional independence)

For each $\ell = \{0, 1\}$, it is hold that $Y^{(\ell)} \perp C^{(\ell)} \mid X^{(\ell)}$

$F^{(\ell)}(y|x)$ is identifiable under Assumption 2, but weaker conditions are needed. Censoring

Assumption 4 (Conditional independence)

For each $\ell = \{0, 1\}$, it is hold that $Y^{(\ell)} \perp C^{(\ell)} \mid X^{(\ell)}$

- Classical methods, like distribution regression and quantile methods can be misleading or cumbersome to compute.
- Because of the link between hazards and distribution functions, it is specified a model for the conditional hazard.
  - ...but, parametric hazard models are very restrictive.
  - so, we use a semiparametric model.

Assume that hazard function (Cox 1972, 1975) is given by:

$$\lambda^{(\ell)}(y|x) = \lambda_0^{(\ell)}(y) \exp\left(\beta_\ell^T x\right)$$

Assume that hazard function (Cox 1972, 1975) is given by:

$$\lambda^{(\ell)}(y|x) = \lambda_0^{(\ell)}(y) \exp\left(\beta_\ell^T x\right)$$

Cox Model

- Conditional distribution is
  $F^{(\ell)}(y|x) = 1 - \exp\left(-\int_0^y \lambda_0^{(\ell)}(\bar{y})\, d\bar{y}\right)^{\exp\left(\beta_\ell^T x\right)}$
- β_ℓ is estimated by Partial Maximum Likelihood estimation.
- $\lambda_0^{(\ell)}(y)$ does not need to be specified, and is estimated nonparametrically (c.f. Kalbfleisch and Prentice, 1973; and Breslow, 1974) Baseline estimators

Decomposition of other Distributional Features ($CCOX$)

$$\hat{\Delta}_Y^{\theta} = \theta\left(\hat{F}_Y^{(1,1)}\right) - \theta\left(\hat{F}_Y^{(0,1)}\right) + \theta\left(\hat{F}_Y^{(0,1)}\right) - \theta\left(\hat{F}_Y^{(0,0)}\right)$$

where

$$\hat{F}_Y^{(i,j)}(y) = n_j^{-1} \sum_{l=1}^{n} \hat{F}^{(i)}(y|x_l)\, 1_{\{D_l=\ell\}}$$

and

$$\hat{F}^{(i)}(y|x) = 1 - \exp\left(-\int_0^y \hat{\lambda}_0^{(i)}(\bar{y})\, d\bar{y}\right)^{\exp\left(\hat{\beta}_i^T x\right)}$$

Decomposition of other Distributional Features (*CCOX*)

$$\hat{\Delta}_Y^{\theta} = \theta\left(\hat{F}_Y^{(1,1)}\right) - \theta\left(\hat{F}_Y^{(0,1)}\right) + \theta\left(\hat{F}_Y^{(0,1)}\right) - \theta\left(\hat{F}_Y^{(0,0)}\right)$$

where

$$\hat{F}_Y^{(i,j)}(y) = n_j^{-1}\sum_{l=1}^{n}\hat{F}^{(i)}(y|x_l)\,1_{\{D_l=\ell\}}$$

and

$$\hat{F}^{(i)}(y|x) = 1 - \exp\left(-\int_0^y \hat{\lambda}_0^{(i)}(\bar{y})\,d\bar{y}\right)^{\exp\left(\hat{\beta}_i^T x\right)}$$

- Validity follows the arguments in Chernozhuvok et. al. (2013).
- Inference can be performed based on bootstrapping techniques.

Asymptotics

Outline

Monte Carlo Simulations

Issues to evaluate:

- Compare proposed methods with classical methods for no-censored data.
- Performance under different censoring mechanims and distributional assumptions.
- Inference on decomposition components based on bootstrapping.

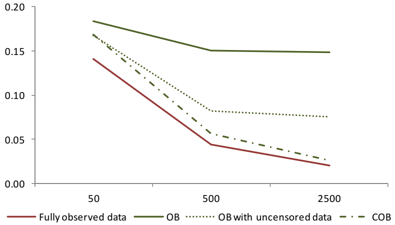
Parameters for the simulations:

- Sample sizes: 50, 500, 2500.
- Replications: 1000.
- Censorship levels: various levels depending on the exercises.

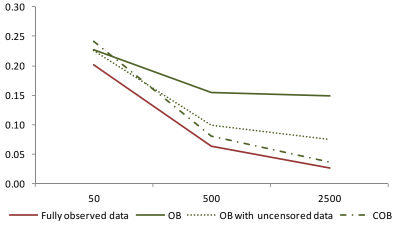
Simulation procedure:
Draw (Y, X, C), and then compute $Z = \min(Y, C)$ and $\delta = 1_{\{Y \leq C\}}$.

COB Decomposition

Montecarlo Exercises



Composition effect



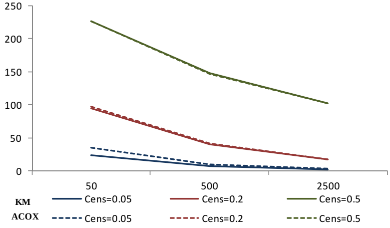
Structure Effect

y-axis measures the average of the absolute deviations across 1000 draws.

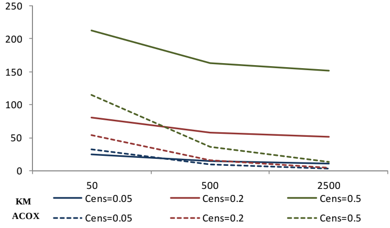
DGP

Censoring Mechanism

Montecarlo Exercises



$Y \perp C$



$Y \perp C|X$

y-axis measures the maximum distance: $MD = \max_{y} \left|\tilde{F}_Y(y) - \hat{F}_Y(y)\right|$. Values are multiplied by 1000 to facilitate comparisons. Simulated times follow Weibull distribution and baseline quantities computed by Breslow estimator.

DGP

Outline

Unemployment Duration Gender Gaps

Spain 2004-2007

- Spain had experienced one of the highest unemployment rates among OECD countries during the period 1995-2005: 14% in Spain, 5% in the US and 6.8% the average OECD.

- Also, more pronounced differences between gender are observed: 9 p.p in Spain and 0.04 p.p. in the US.

- Existing studies:
  - Differences in unemployment rates (Niemi, 1974; Jhonson, 1983; Azmat et. al., 2006).
  - Difference in the exit rate and the average hazard rate (Eusamio, 2004; Ortega, 2008; Baussola et. al., 2015).

A Decomposition Exercise

Gender Unemployment Duration Gap in Spain: 2004-2007

- Data: Survey of Income and Living Conditions 2004-2007.
  - Contain information about occupational status monthly.
  - Individual characteristics: age, educational level, tenure, marital status, household structure (household head and unemployed members) and region.
  - Population: workers older than 25 starting unemployment spell during 2004-2007.
- Target parameters: Average unemployment duration, the probability of being long term unemployed (12 and 24 months) and the Gini coefficient.
- Two dimensions of duration: duration until exit from unemployment and duration until getting a job.

Unemployment Duration in Spain: 2004-2007

Distributional Parameters of Duration to Exit from Unemployment

		Mean	LTU(12)	LTU(24)	Gini
Kaplan-Meier	**Women**	11.090	0.410	0.145	0.496
	Men	7.804	0.237	0.065	0.542
CCOX	**Women**	11.160	0.396	0.145	0.508
	Men	7.767	0.235	0.067	0.544
Only Uncensored	**Women**	7.456	0.292	0.045	0.446
	Men	5.466	0.153	0.014	0.485

Authors' calculations.

Decomposition Unemployment Duration Gender Gaps

Decomposition Distributional Statistics of Duration to Exit from Unemployment

			Total	Composition	Structure
COB	**Mean**	**Difference**	3.285	0.386	2.899
		CI 90%	[2.067 , 4.442]	[-0.478 , 2.425]	[0.408 , 4.219]
CCOX	**Mean**	**Difference**	3.392	0.537	2.855
		CI 90%	[2.098 , 4.491]	[-0.349 , 1.361]	[1.501 , 4.224]
	LTU(12)	**Difference**	0.161	0.012	0.148
		CI 90%	[0.115 , 0.206]	[-0.013 , 0.043]	[0.092 , 0.198]
	LTU(24)	**Difference**	0.078	0.014	0.064
		CI 90%	[0.043 , 0.109]	[-0.008 , 0.035]	[0.022 , 0.104]
	Gini	**Difference**	-0.036	0.006	-0.042
		CI 90%	[-0.071 , 0.000]	[-0.002 , 0.010]	[-0.075 , -0.005]

Authors' calculations.

Decomposition Unemployment Duration Gender Gaps

Decomposition Distributional Statistics of Duration from Unemployment to Employment

			Total	Composition	Structure
COB	**Mean**	**Difference**	7.224	5.698	1.525
		CI 90%	[4.804 , 9.020]	[3.816 , 8.678]	[-1.691 , 3.203]
CCOX	**Mean**	**Difference**	7.865	1.713	6.151
		CI 90%	[5.266 , 9.507]	[0.159 , 3.028]	[3.813 , 8.191]
	LTU(12)	**Difference**	0.184	0.036	0.148
		CI 90%	[0.137 , 0.231]	[0.000 , 0.071]	[0.095 , 0.202]
	LTU(24)	**Difference**	0.176	0.041	0.135
		CI 90%	[0.130 , 0.226]	[0.003 , 0.074]	[0.084 , 0.192]
	Gini	**Difference**	-0.040	-0.014	-0.025
		CI 90%	[-0.079 , -0.008]	[-0.029 , 0.001]	[-0.065 , 0.007]

Authors' calculations.

Other results

Outline

Concluding Remarks

- We propose a nonparametric Oaxaca-Blinder type decomposition method for the mean difference under censoring, using classical identification assumptions in survival analysis literature.
- Under weaker identification conditions, we present a decomposition method based on the estimation of the conditional distribution.
- Both methods work properly in finite samples.
- Unemployment duration gap by gender in Spain:
    - The structure effect plays the major role to explain the gender gaps of several distributional parameters.
    - The composition effect is statistically significant to explain gender gaps for the duration until getting a job.

Oaxaca-Blinder type Decomposition Methods for Duration Outcomes

Andres Garcia-Suaza
garciasuaza@gmail.com

Thank you!

Future Research

- The COB methods can be extended for:
    - Detailed decomposition: path dependece and omitted group problems (Oaxaca and Ransom 1999; Yun, 2005; Firpo et. al. 2007).
    - Decomposition of other parameters: RIF regression (Firpo et. al. 2009).
- Proportionality of hazard rates assumed by the Cox model might be unrealistic in some situations. More flexible semiparametric models, as Distributional Regression (Foresi and Peracchi, 1995; Chernozhukov et. al. 2013), are desirable.
- Empirical findings remark the relevance of institutional factors. We study whether unemployment benefits plays a role in the gender difference of the exit rate.

$$\Delta_Y^\mu = (\beta_1 - \beta_0)^T \mu_X^{(1)} + \beta_0^T \left(\mu_X^{(1)} - \mu_X^{(0)}\right)$$

- Target counterfactual statistic: $\mu_Y^{(0,1)} = \mathbb{E}\left(Y^{(0)}|X = \mathbb{E}(X \mid D = 1)\right) = \beta_0^T \mu_X^{(1)}$.

$$\Delta_Y^{\mu} = (\beta_1 - \beta_0)^T \mu_X^{(1)} + \beta_0^T \left(\mu_X^{(1)} - \mu_X^{(0)}\right)$$

Assumption 1

Let $\varepsilon^{(\ell)}$ be the best linear predictor error for subpopulation ℓ, i.e. $\varepsilon^{(\ell)} = (Y^{(\ell)} - \beta_\ell^T X^{(\ell)})$. The following conditions are satisfied:

a. Overlapping support: if $\mathcal{X} \times \mathcal{E}$ denotes the support of observables and unobservable characteristics of the underlying population, then $(X^{(0)}, \varepsilon^{(0)}) \cup (X^{(1)}, \varepsilon^{(1)}) \in \mathcal{X} \times \mathcal{E}$.

b. Simple counterfactual treatment.

c. Conditional independence of treatment and unobservables: $D \perp \varepsilon | X$.

Back

Define $\mathbb{P}(C \leq y|D = \ell) = G(y|D = \ell)$. Under Assumption 2:

$$\begin{aligned} H_{11}(y, x, \ell) &= \mathbb{P}(Z \leq y, X \leq x, D = \ell, \delta = 1) \\ &= \int_0^y [1 - G(\bar{y} - |D = \ell)] F(d\bar{y}, x, \ell) \end{aligned}$$

and,

$$\begin{aligned} H(y, \ell) &= \mathbb{P}(Z \leq y, \delta = 1) \\ &= [1 - G(\bar{y} - |D = \ell)][1 - F(y-, \infty, \ell)] \end{aligned}$$

Define $\mathbb{P}\left(C \leq y|D=\ell\right) = G\left(y|D=\ell\right)$. Under Assumption 2:

$$\begin{aligned} H_{11}\left(y,x,\ell\right) &= \mathbb{P}\left(Z \leq y, X \leq x, D=\ell, \delta=1\right) \\ &= \int_0^y \left[1-G\left(\bar{y}-|D=\ell\right)\right] F\left(d\bar{y},x,\ell\right) \end{aligned}$$

and,

$$\begin{aligned} H\left(y,\ell\right) &= \mathbb{P}\left(Z \leq y, \delta=1\right) \\ &= \left[1-G\left(\bar{y}-|D=\ell\right)\right]\left[1-F\left(y-,\infty,\ell\right)\right] \end{aligned}$$

Therefore,

$$\Lambda\left(y,x,\ell\right) = \int_0^y \frac{F\left(d\bar{y},x,\ell\right)\left[1-G\left(\bar{y}-|D=\ell\right)\right]}{1-F\left(\bar{y}-,\infty,\ell\right)\left[1-G\left(\bar{y}-|D=\ell\right)\right]} = \int_0^y \frac{H_{11}\left(d\bar{y},x,\ell\right)}{1-H\left(\bar{y}-,\ell\right)}$$

Thus, Λ is estimated by:

$$\hat{\Lambda}(y, x, \ell) = \int_0^y \frac{\hat{H}_{11}(d\bar{y}, x, \ell)}{1 - \hat{H}(\bar{y}-, \ell)} = \sum_{i=1}^{n_\ell} \frac{1_{\{Z_i \leq y, X_i \leq x, D_i = \ell, \delta_i = 1\}}}{n_\ell - R_i^{(\ell)} + 1}$$

where, $R_i^{(\ell)} = n_\ell \hat{H}(Z_i, \ell)$.

Thus, Λ is estimated by:

$$\hat{\Lambda}(y,x,\ell) = \int_0^y \frac{\hat{H}_{11}(d\bar{y},x,\ell)}{1-\hat{H}(\bar{y}-,\ell)} = \sum_{i=1}^{n_\ell} \frac{1_{\{Z_i \le y, X_i \le x, D_i=\ell, \delta_i=1\}}}{n_\ell - R_i^{(\ell)} + 1}$$

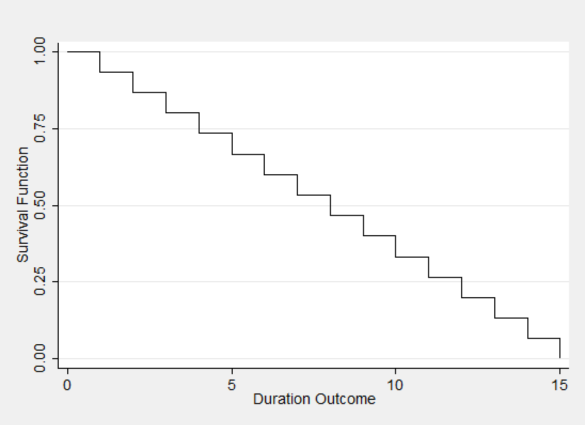
where, $R_i^{(\ell)} = n_\ell \hat{H}(Z_i,\ell)$.

And the joint distribution associated is

$$\hat{F}(y,x,\ell) = 1 - \prod_{Z_{i:n_\ell}^{(\ell)} \le y, X_{[i:n_\ell]}^{(\ell)} \le x} \left[1 - \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell - R_i^{(\ell)} + 1}\right]$$

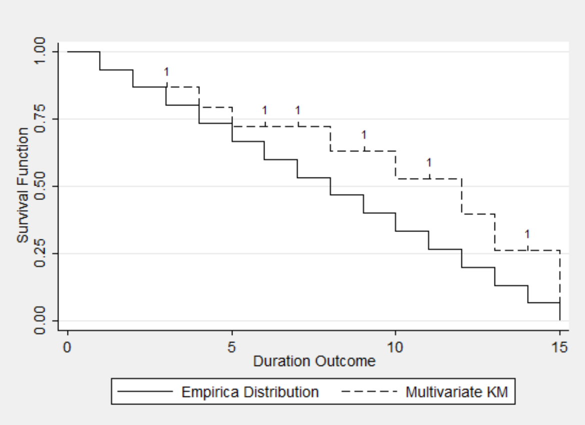
with $Z_{i:n_\ell}^{(\ell)} = Z_j$ if $R_j^{(\ell)} = i$, and for any $\{\xi_i\}_{i=1}^{n_\ell}$, $\xi_{[i:n_\ell]}^{(\ell)}$ is the i-th concomitant of $Z_{i:n_\ell}^{(\ell)}$, i.e., $\xi_{[i:n_\ell]}^{(\ell)} = \xi_j$ if $Z_{i:n_\ell}^{(\ell)} = Z_j$.

Example Multivariate KM



Z	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
δ	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1

Example Multivariate KM



Z	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
δ	1	1	0	1	1	0	0	1	0	1	0	1	1	0	1

(Censored) OB Decompisition *-COB-*

For a generic $J(y,\ell)$ define $\tau_J^{(\ell)} = \inf\{y : J(y,\ell) = 1\} \leq \infty$.

Assumption 3

For $\ell = \{0,1\}$, it holds that $\tau_{F_Y}^{(\ell)} \leq \tau_G^{(\ell)}$.

(Censored) OB Decompisition -*COB*-

For a generic $J(y,\ell)$ define $\tau_J^{(\ell)} = \inf\{y : J(y,\ell) = 1\} \leq \infty$.

Assumption 3

For $\ell = \{0,1\}$, it holds that $\tau_{F_Y}^{(\ell)} \leq \tau_G^{(\ell)}$.

Corollary 2

Assume $\mathbb{E}\left(X^{(\ell)} X^{(\ell)^T}\right)$ is positive definite and $\frac{n_\ell}{n} \to \rho_\ell$ with $\rho_0 + \rho_1 = 1$. Under Assumption 1-3, we have:

$$n^{1/2}\left(\hat{\Delta}_Y^{\mu} - \Delta_Y^{\mu}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\Delta_Y}\right)$$

$$n^{1/2}\left(\hat{\Delta}_S^{\mu} - \Delta_S^{\mu}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\Delta_S}\right)$$

$$n^{1/2}\left(\hat{\Delta}_C^{\mu} - \Delta_C^{\mu}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\Delta_C}\right)$$

Proposition 2

Assume $\mathbb{E}\left(X^{(\ell)}X^{(\ell)^T}\right)$ is positive definite and $\frac{n_\ell}{n} \rightarrow \rho_\ell$ with $\rho_0 + \rho_1 = 1$. Under Assumption 2-3, we have:

$$n^{1/2}\left[\left(\hat{\beta}_\ell - \beta_\ell\right), \left(\hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)}\right)\right] \xrightarrow{d} \mathcal{N}\left(0, \Sigma_{\beta\mu_X}^{(\ell)}\right)$$

where

$$\Sigma_{\beta\mu_X}^{(\ell)} = \left(\Sigma_{XX}^{(\ell)}\right)^{-1}\Sigma_0^{(\ell)}\left(\Sigma_{XX}^{(\ell)}\right)^{-1} = \left(\sigma_\beta^{(\ell)}, \sigma_{\beta\mu_X}^{(\ell)}; \sigma_{\beta\mu_X}^{(\ell)}, \sigma_{\mu_X}^{(\ell)}\right)$$

Let $\rho_0 = \rho$. V_{Δ_Y}, V_{Δ_S} and V_{Δ_C} are defined as:

$$V_{\Delta_Y} = \frac{1}{\rho} V_0 + \frac{1}{1-\rho} V_1$$

where $V_\ell = \mu_X^{(\ell)^T} \sigma_\beta^{(\ell)} \mu_X^{(\ell)} + \beta_\ell^T \sigma_{\mu_X}^{(\ell)} \beta_\ell + 2\beta_\ell^T \sigma_{\beta\mu_X}^{(\ell)} \mu_X^{(\ell)}$,

$$\begin{aligned} V_{\Delta_S} &= \frac{1}{1-\rho} \Delta_\beta^T \sigma_{\mu_X}^{(1)} \Delta_\beta + \frac{2}{1-\rho} \Delta_\beta^T \sigma_{\beta\mu_X}^{(1)} \mu_X^{(1)} + \frac{1}{\rho(1-\rho)} \mu_X^{(1)^T} \sigma_\beta \mu_X^{(1)} \\ V_{\Delta_C} &= \frac{1}{\rho} \Delta_{\mu_X}^T \sigma_\beta^{(0)} \Delta_{\mu_X} + \frac{2}{\rho} \beta_0^T \sigma_{\beta\mu_X}^{(0)} \Delta_{\mu_X} + \frac{1}{\rho(1-\rho)} \beta_0^T \sigma_{\mu_X} \mu_X^{(1)} \end{aligned}$$

with $\Delta_\beta^T = \beta_1 - \beta_0$, $\Delta_{\mu_X}^T = \mu_X^{(1)} - \mu_X^{(0)}$, $\sigma_\beta = \rho\sigma_\beta^{(1)} + (1-\rho)\,\sigma_\beta^{(0)}$ and $\sigma_{\mu_X} = \rho\sigma_{\mu_X}^{(1)} + (1-\rho)\,\sigma_{\mu_X}^{(0)}$.

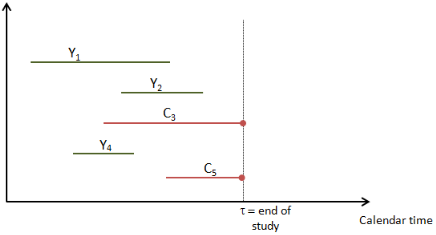
Back

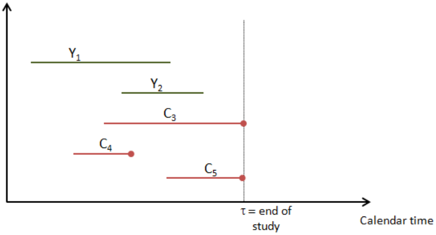
- Sampling methods (Efron -1981-).
  - Simple method: draw $(Z_i^*, \delta_i^*, X_i^*)$ for $i = 1, \ldots, n$ from (Z_i, δ_i, X_i).
  - Obvious method: draw $Y_i^* \sim \widehat{F}(y|X_i^*)$ and $C^* \sim \widehat{G}(y|X_i^*)$. Define $Z^* = \min(Y^*, C^*)$ and $\delta^* = 1_{\{Y^* \leq C^*\}}$.

- Usual methods for constructing confidence bands can be used: percentile, hybrid, boot-t.

Back

Censoring Mechanism





Back

Estimation Baseline Quantities

Breslow estimator:

$$\hat{\Lambda}_{0B}^{(\ell)}(y) = \sum_{i=1}^{y} \frac{1}{\sum_{j \in r(y_i)} e^{\hat{\beta}_\ell^T x_j^{(\ell)}}}$$

Kalbfleisch and Prentice estimator:

$$\hat{\Lambda}_{0KP}^{(\ell)}(y) = \sum_{i=1}^{n_\ell} \left(1 - \hat{\alpha}_i^{(\ell)}\right) 1_{\{y_i \leq y\}}$$

where the hazard probabilities $\hat{\alpha}_i^{(\ell)}$ solve:

$$\sum_{j \in d^{(\ell)}(y_i)} e^{\hat{\beta}_\ell^T x_j^{(\ell)}} \left[1 - \hat{\alpha}_i^{\exp\left(\hat{\beta}_\ell^T x_j^{(\ell)}\right)}\right]^{-1} = \sum_{l \in r^{(\ell)}(y_i)} e^{\hat{\beta}_\ell^T x_l^{(\ell)}}$$

with $r^{(\ell)}(y_i)$ the pool risk and the $d^{(\ell)}(y_i)$ set of individuals reporting failure.

Back

Validity CCOX Method

Consistency

Under Assumptions 1, 3 and 4, and $\frac{n_\ell}{n} \to \rho_\ell$ with $\rho_0 + \rho_1 = 1$, we have:

$$n^{1/2}\left(\hat{F}_Y^{(i,j)}(y) - F_Y^{(i,j)}(y)\right) \Longrightarrow \bar{M}^{(i,j)}(y)$$

where $\bar{M}^{(i,j)}(y)$ tight zero-mean Gaussian process with uniform continuous path on $\mathcal{S}upp\,(y)$, define as:

$$\bar{M}^{(i,j)}(y) = \rho_i^{1/2} \int M^{(i)}(y,x)\, dF^{(j)}(x) + \rho_j^{1/2} N^{(j)}\left(F^{(i)}(y|.)\right)$$

and,

$$n_\ell^{1/2}\left(\hat{F}^{(\ell)}(y|x) - F^{(\ell)}(y|x)\right) \Longrightarrow M^{(\ell)}(y,x)$$

$$n_\ell^{1/2} \int fd\left(\hat{F}^{(\ell)}(x) - F^{(\ell)}(x)\right) \Longrightarrow N^{(\ell)}(f)$$

Validity CCOX Method

- Previous result lies on the fact that $n^{1/2}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbb{V}_{\beta}\right)$ and $n^{1/2}\left(\hat{\Lambda}_0\left(y\right)-\Lambda_0\left(y\right)\right)\Longrightarrow\mathcal{C}_{\infty}$ (Tsiatis, 1981: Andersen and Gill, 1982).
- Since $F\left(y|.\right)$ is Hadamard differentiable with respect to β and $\Lambda^0\left(.\right)$ (see c.f. Freitag and Munk, 2005), the CCOX is Hadamard differentiable, and the related smooth functionals also obeys a central limit theorem.
- Given that bootstrap techniques are valid to make inference on Cox estimator of the conditional distribution (Cheng and Huang, 2010); then, resampling methods are also valid for the CCOX.
- By Hadamard differentiability this result also applies to smooth functionals.

Back

Validity CCOX Method

- Previous result lies on the fact that $n^{1/2}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbb{V}_{\beta}\right)$ and $n^{1/2}\left(\hat{\Lambda}_0\left(y\right)-\Lambda_0\left(y\right)\right)\Longrightarrow\mathcal{C}_{\infty}$ (Tsiatis, 1981: Andersen and Gill, 1982).
- Since $F\left(y|.\right)$ is Hadamard differentiable with respect to β and $\Lambda^0\left(.\right)$ (see c.f. Freitag and Munk, 2005), the CCOX is Hadamard differentiable, and the related smooth functionals also obeys a central limit theorem.
- Given that bootstrap techniques are valid to make inference on Cox estimator of the conditional distribution (Cheng and Huang, 2010); then, resampling methods are also valid for the CCOX.
- By Hadamard differentiability this result also applies to smooth functionals.

Back

Validity CCOX Method

- Previous result lies on the fact that $n^{1/2}\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{V}_{\beta}\right)$ and $n^{1/2}\left(\hat{\Lambda}_0\left(y\right)-\Lambda_0\left(y\right)\right) \Longrightarrow \mathcal{C}_{\infty}$ (Tsiatis, 1981: Andersen and Gill, 1982).
- Since $F\left(y|.\right)$ is Hadamard differentiable with respect to β and $\Lambda^0\left(.\right)$ (see c.f. Freitag and Munk, 2005), the CCOX is Hadamard differentiable, and the related smooth functionals also obeys a central limit theorem.
- Given that bootstrap techniques are valid to make inference on Cox estimator of the conditional distribution (Cheng and Huang, 2010); then, resampling methods are also valid for the CCOX.
- By Hadamard differentiability this result also applies to smooth functionals.

Back

Validity CCOX Method

- Previous result lies on the fact that $n^{1/2}\left(\hat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbb{V}_{\beta}\right)$ and $n^{1/2}\left(\hat{\Lambda}_0\left(y\right)-\Lambda_0\left(y\right)\right)\Longrightarrow\mathcal{C}_{\infty}$ (Tsiatis, 1981: Andersen and Gill, 1982).
- Since $F\left(y|.\right)$ is Hadamard differentiable with respect to β and $\Lambda^0\left(.\right)$ (see c.f. Freitag and Munk, 2005), the CCOX is Hadamard differentiable, and the related smooth functionals also obeys a central limit theorem.
- Given that bootstrap techniques are valid to make inference on Cox estimator of the conditional distribution (Cheng and Huang, 2010); then, resampling methods are also valid for the CCOX.
- By Hadamard differentiability this result also applies to smooth functionals.

Back

Simulation Setup

$\ell = \mathbf{0}$	$Y^{(0)} = 5 + X^{(0)} + \varepsilon_Y^{(0)} \quad \varepsilon_Y^{(0)} \sim \mathcal{N}(0,1)$
	$C^{(0)} = 5 + \varepsilon_C^{(0)} \quad \varepsilon_C^{(0)} \sim \mathcal{N}(\upsilon_0, 1.5)$
$\ell = \mathbf{1}$	$Y^{(1)} = 5 + X^{(1)} + \varepsilon_Y^{(1)} \quad \varepsilon_Y^{(1)} \sim \mathcal{N}(0,1)$
	$C^{(1)} = 5 + \varepsilon_C^{(1)} \quad \varepsilon_C^{(1)} \sim \mathcal{N}(\upsilon_1, 1.5)$

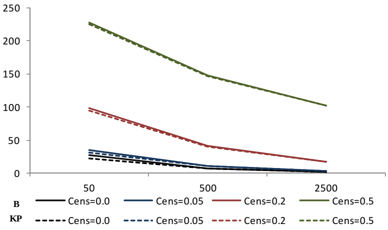
- $X^{(0)} \sim \mathcal{N}(1.5, 0.5)$ and $X^{(1)} \sim \mathcal{N}(1, 0.5)$.
- $(\upsilon_0, \upsilon_1) = (2.5, 2)$, hence, censoring level is 30%.

Back

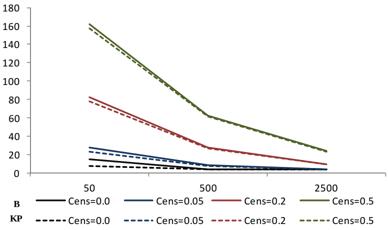
Simulation Setup

Assumption		DGP
$Y \perp C$	Weibull	$Y \sim WB\left(e^{2-x}, 5\right)$ $C \sim WB\left(e^{2+\upsilon}, 5\right)$ $\upsilon = (0.25, -0.2, -0.5)$
	Normal	$Y = 5 + X + \varepsilon_Y, \quad \varepsilon_Y \sim N(0,1)$ $C = 5 + \varepsilon_C, \quad \varepsilon_C \sim N(\upsilon, 1)$ $\upsilon = (3, 1.5, 0.5)$
$Y \perp C \mid X$	Weibull	$Y \sim WB\left(e^{2-x}, 5\right)$ $C \sim WB\left(e^{2-x+\upsilon}, 7\right)$ $\upsilon = (0.45, 0.2, -0.02)$
	Normal	$Y = 5 + X + \varepsilon_Y, \quad \varepsilon_Y \sim N(0,1)$ $C = 5 + X + \varepsilon_C, \quad \varepsilon_C \sim N(\upsilon, 1)$ $\upsilon = (2.5, 1, 0)$

- $X \sim \mathcal{U}(0,1)$.
- Censoring levels: 0%, 5%, 25% and 50%.
- We compare the estimation of the unconditional distribution using the counterfactual operator with the KM estimator.



Weibull

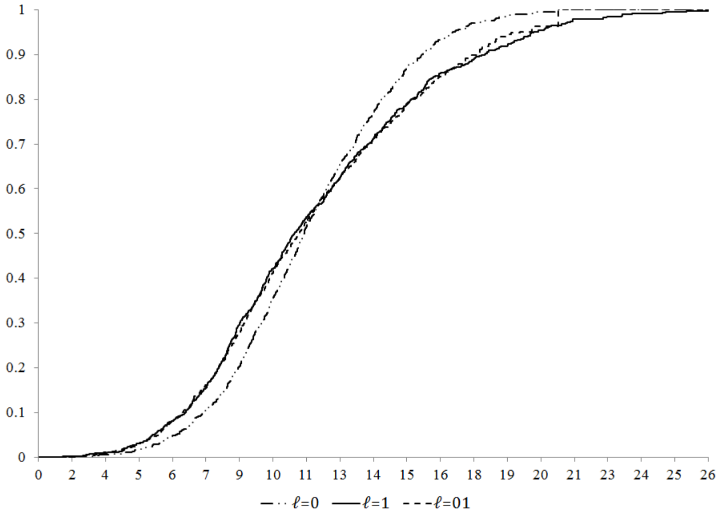


Normal

y-axis measures the maximum distance: $MD = \max_y \left|\tilde{F}_Y(y) - \hat{F}_Y(y)\right|$. Values are multiplied by 1000 to facilitate comparisons. Survival times generated under the assumption $Y \perp C$.

Back

- $Y^{(\ell)} \sim \mathcal{WB}\left(e^{3-X^{(\ell)}}, 5\right)$ and $C^{(\ell)} \sim \mathcal{WB}\left(e^{3.17-X^{(\ell)}}, 5\right)$.
- $X^{(0)} \sim \mathcal{U}(0,1)$ and $X^{(1)} = \sum_{i=1}^{3} \mathcal{U}(0,1)$.
- Censoring levels: 0% and 30%.
- $n = 500,\ B = 1000$.



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 2 4 5 6 7 9 10 11 13 14 15 16 18 19 20 21 23 24 25 26
ℓ=0
ℓ=1
ℓ=01

Decomposition Exercise and Inference

Montecarlo Exercises

Confidence Level	Censoring Levels		Truncated Mean		Q(0.50)	
	$\Pr\left(\delta^{(0)}=0\right)$	$\Pr\left(\delta^{(1)}=0\right)$	Percentile	Hybrid	Percentile	Hybrid
95	0.0	0.0	0.961	0.962	0.958	0.953
	0.0	0.3	0.954	0.963	0.952	0.940
	0.3	0.0	0.963	0.972	0.958	0.944
	0.3	0.3	0.952	0.966	0.968	0.944
90	0.0	0.0	0.907	0.913	0.917	0.911
	0.0	0.3	0.902	0.911	0.915	0.903
	0.3	0.0	0.915	0.923	0.915	0.897
	0.3	0.3	0.912	0.917	0.907	0.895

Mean is truncated at $Y = 15$.

Back

Decomposition Unemployment Duration Gender Gaps

Duration until getting a Job: Censoring Mechanism

Dependence of Censoring on Covariates

	Exit from Unemp.		Unemp. to Emp.	
	Women	Men	Women	Men
Linear Prob. Model	0.046	0.083	0.177	0.123
Logit	0.070	0.128	0.166	0.167

Authors' calculations.

Decomposition of Mean Difference Ignoring Censoring

	Exit from Unemp.		Unemp. to Emp.	
	COB	CCOX	COB	CCOX
Total	2.085	2.040	0.945	0.876
Composition	0.438	0.471	0.089	0.029
Structure	1.647	1.569	0.857	0.847

Authors' calculations.

Back