Universidad del Rosario
Facultad de Economía

Oaxaca-Blinder Type Counterfactual Decomposition Methods for Duration Outcomes

Andrés García-Suaza

SERIE DOCUMENTOS DE TRABAJO

No. 183

Enero de 2016

Oaxaca-Blinder type Counterfactual Decomposition Methods for Duration Outcomes

Andrés García-Suaza*

January 2016

Abstract

Existing inference procedures to perform counterfactual decomposition of the difference between distributional features, applicable when data is fully observed, are not suitable for censored outcomes. This may explain the lack of counterfactual analyses using target variables related to duration outcomes, typically observed under right censoring. For instance, there are many studies performing counterfactual decomposition of the gender wage gaps, but very few on gender unemployment duration gaps. We provide an Oaxaca-Blinder type decomposition method of the mean for censored data. Consistent estimation of the decomposition components is based on a prior estimator of the joint distribution of duration and covariates under suitable restrictions on the censoring mechanism. To decompose other distributional features, such as the median or the Gini coefficient, we propose an inferential method for the counterfactual decomposition by introducing restrictions on the functional form of the conditional distribution of duration given covariates. We provide

*Email address: andres.garcia@urosario.edu.co, agarcias@eco.uc3m.es. Assistant professor at Universidad del Rosario and PhD student at Universidad Carlos III de Madrid. I thank Professor Miguel A. Delgado for his guidance and support. I also thank comments from Pedro Sant'Anna, Juan Carlos Escanciano, Luis F. Gamboa, Jan Stuhler, Winfried Stute, Andrii Parkhomenko, Salvatore Lo Bello, Sebastian Panthöfer, Alfonso Alba and participants at the LACEA-LAMES annual meeting 2014 (Sao Paulo), the Enter Jamboree 2015 (Mannheim), the XVIII Meeting of Applied Economics -ALdE- 2015 (Alicante), the XI Jornadas de Economia Laboral (Barcelona) and the 2nd IAAE Annual Conference (Thessaloniki). This research is funded by Spanish Plan Nacional de I+D+I grants number ECO2012-33053 and ECO2014-55858-P.

formal justification for asymptotic inference and study the finite sample performance through Monte Carlo experiments. Finally, we apply the proposed methodology to the analysis of unemployment duration gaps in Spain. This study suggests that factors beyond the workers' socioeconomic characteristics play a relevant role in explaining the difference between several unemployment duration distribution features such as the mean, the probability of being long term unemployed and the Gini coefficient.

Keywords: Oaxaca-Blinder Decomposition, Right Censoring, Counterfactual Outcomes, Duration Data, Hazard Models, Unemployment Duration, Gender Gaps.

JEL Codes: C14, C24, C41, J64.

1 Introduction

From Oaxaca (1973) and Blinder (1973) contributions (OB henceforth), counterfactual decomposition technique became a popular research tool for economic analysis. This consists in decomposing the difference between the means of two subpopulations into counterfactual components based on observed characteristics. On some occasions, the decomposition of the mean may not be enough for studying the difference in the outcome of interest, so that the decomposition of other distributional features, such as the median or Gini coefficient, has also been proven useful. In this context, Freeman R.B. (1980); Juhn *et al.* (1991); DiNardo *et al.* (1996); Machado and Mata (2005) and Chernozhukov *et al.* (2013) developed further decomposition techniques going beyond the mean. See Fortin *et al.* (2011) for a comprehensive review.

The aforementioned procedures, designed for the case of fully observed data, have been widely used in the analysis of microdata, mainly in labor economics. For instance, there is a large literature devoted to studying the gender wage gap (see Oaxaca, 1973; Blinder, 1973; Cain, 1986; Cotton, 1988; Machado and Mata, 2005; O'Neill and O'Neill, 2006; Blau and Kahn, 1992) or the increase in the US wage dispersion in the 1980's (Juhn *et al.*, 1991, 1993; DiNardo *et al.*, 1996; Melly, 2005). However, other relevant outcomes, such as unemployment duration, have not received so much attention, possibly due to the absence of decomposition methods for censored variables.

Collecting duration data requires following individuals over time. In this context, censoring occurs because individuals either do not change their status during the follow-up period or withdraw before the end of the study. For instance, in the case of unemployment duration, it is not possible to observe the complete unemployment duration for those individuals still unemployed at the last follow-up period or for those who leave the labor force.

Existing literature concerning the decomposition of the unemployment duration gender gap has focused on the average hazard rate (c.f. Powers and Yun, 2009). However, this parameter does not correspond to the hazard function associated to the underlying duration distribution, and consequently, the decomposition is hard to interpret. Therefore, we consider decomposition methods suitable for several parameters related to the duration distribution that allow the usual interpretation as in the no censoring case. These methods are based on classical identification assumptions in the survival analysis literature, allow to deal with covariates of diverse nature (continuous and discrete) and are simple to implement in conventional software[1].

In particular, we provide a regression-based method, analogous to the classical OB, in a nonparametric context. Consistent estimation of the means in the two subpopulations, and the corresponding decomposition components, requires identification restrictions, mainly that the censoring mechanism be random. Additionally, we consider an alternative method to perform counterfactual decomposition beyond the mean under weaker identification assumptions, but which requires some knowledge on the functional form of the underlying conditional distribution of duration given covariates. Thus, we discuss a flexible decomposition method based on proportional hazard models, but other specifications, e.g., a quantile regression or a fully parametric model, are also possible.

The proposed methodologies are used to study unemployment duration gender gaps in Spain during the period 2004-2007 using data from the European Survey on Income and Living Conditions (SILC). Spain is one of the OECD countries with the highest and most persistent gender gaps in unemployment rates (Azmat *et al.*, 2006). In order to provide additional evidence about the unemployment gender gap, we investigate the duration in two dimensions: the duration until leaving unemployment and the duration until getting

[1] All codes used are available upon request.

a job.

To do so, we perform counterfactual decomposition for several parameters of the unemployment duration to quantify to what extent the gender gap is explained by socioeconomic factors, like individual and household background and local labor market characteristics. Particularly, we study the mean unemployment duration, the probability of being long term unemployed and the Gini coefficient. Our findings reveal the important role of the structure effect, i.e. the return to the characteristics, to explain the gender gap in the two types of durations; while the difference in the workers' characteristics is only important for the duration until getting a job.

The rest of the article is organized as follows. The next section introduces the implementation of the OB method for the censored data case in a full nonparametric context, and provides sufficient conditions to perform valid inferences on the counterfactual decomposition components. Section 3 discusses the decomposition of other distributional features under a semiparametric specification. Section 4 studies the finite sample properties through Monte Carlo simulations. Section 5 applies the proposed methodologies to analyze unemployment duration gender gaps in Spain. The last section is devoted to final remarks and suggestions for further work. Some mathematical details and further discussion on technical results are presented in the Appendix.

2 Nonparametric Oaxaca-Blinder Decomposition under Censoring

Consider a $\mathbb{R}^{+}\times\mathbb{R}^{k}\times\{0,1\}-valued$ random vector (Y,X,D) related to the population under study, where Y denotes duration outcome, X a $k\times 1$ vector of characteristics (including a constant) and D a dummy variable identifying two subpopulations. For instance, Y may be unemployment duration, X relevant socioeconomic characteristics and D a dummy variable for gender.

The difference between the means of the two subpopulations, denoted by $\Delta_Y^{\mu}=\mu_Y^{(1)}-\mu_Y^{(0)}$, with $\mu_Y^{(\ell)}=\mathbb{E}\left(Y\mid D=\ell\right)$, $\ell=\{0,1\}$, can be expressed in terms of the best linear predictors for each subpopulation ℓ. That is:

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

where, $\mu_X^{(\ell)} = \mathbb{E}(X \mid D = \ell)$,

$$\beta_\ell = \underset{b \in \mathbb{R}^k}{\arg\min} \quad \mathbb{E}\left(Y - b^T X \mid D = \ell\right)^2 = \mathbb{E}\left(XX^T \mid D = \ell\right)^{-1} \mathbb{E}\left(XY \mid D = \ell\right)$$

and $\mathbb{E}$ is the expectation operator and "T" denotes transpose.

Oaxaca (1973) and Blinder (1973) exploit this fact to rearrange Equation (1) as

$$\Delta_Y^\mu = (\beta_1 - \beta_0)^T \mu_X^{(1)} + \beta_0^T \left(\mu_X^{(1)} - \mu_X^{(0)}\right) = \Delta_S^\mu + \Delta_C^\mu. \tag{2}$$

This is the classical Oaxaca-Blinder decomposition (OB decomposition, henceforth), where the term Δ_S^μ , known as *the structure effect*, is interpreted as the difference explained by the effect (return) of the explanatory variables on Y, while Δ_C^μ, known as *the composition effect*, is the part of the mean difference explained by the difference in the characteristics.

The crucial ingredient in this counterfactual decomposition is $\beta_0^T \mu_X^{(1)}$, i.e. the best predictor of Y in subpopulation 0 given $X = \mathbb{E}(X \mid D = 1)$. Intuitively, this is the average of the counterfactual outcome $Y^{(0,1)}$ which represents the potential outcome in population 0 if individuals were endowed with characteristics of population 1. Indeed, the label *counterfactual* comes from the fact that this outcome cannot be directly observed in the data[2].

In order to identify the counterfactual average $\mu_Y^{(0,1)} = \beta_0^T \mu_X^{(1)}$, it is necessary to impose some restrictions. Assumption 1 below summarizes the identification conditions usually considered in the decomposition methods literature (see Fortin *et al.*, 2011 for further discussion).

Assumption 1 *Let $\varepsilon^{(\ell)}$ be the best linear predictor error for subpopulation ℓ, i.e. $\varepsilon^{(\ell)} = \left(Y^{(\ell)} - \beta_\ell^T X^{(\ell)}\right)$, with $Y^{(\ell)}$ and $X^{(\ell)}$ the outcome variable and covariates of the corres-*

[2]Counterfactual analysis, as a concept, has been used in a very philosophical way in many sciences (Lewis, 1973). In social science, and particularly in economics since the seminal contribution by Rubin (1974), counterfactual analysis has served to establish a natural framework for studying causal relation (For discussion on the use of counterfactuals in quantitative analysis see Dawid, 2000; Cartwright and Reiss, 2004; Reiss and Cartwright, 2004; Rubin, 2005; Höfler, 2005, and Pearl, 2009).

ponding subpopulation. 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 $\left(X^{(0)}, \varepsilon^{(0)}\right) \cup \left(X^{(1)}, \varepsilon^{(1)}\right) \in \mathcal{X} \times \mathcal{E}$.

b. The only possible counterfactual outcome for an individual that belongs to subpopulation i is $Y^{(j)}$, with $i, j \in \{0, 1\}$ and $i \neq j$.

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

Assumption 1.a is the classical common support condition. Assumption 1.b , known as *simple counterfactual treatment*, rules out the existence of other potential outcomes besides $Y^{(0)}$ and $Y^{(1)}$. Lastly, Assumption 1.c is the classical ignorability (*unconfoundedness*) condition, which ensures that the distribution of unobservables is the same across subpopulations[3]. Conditions 1.b-1.c imply that the conditional distributions of the outcome and the unobservables given covariates remain unaltered when the distribution of covariates varies. This invariance property is the key assumption for validating the estimation of counterfactual outcomes by combining estimates from the two subpopulations.

Given a random sample $\{Y_i, X_i, D_i\}_{i=1}^n$ from (Y, X, D), the OB decomposition is estimated by:

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

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

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

with $n_\ell = \sum_{i=1}^n 1_{\{D_i = \ell\}}$, $\ell = \{0, 1\}$ and $1_{\{A\}}$ denoting the indicator function of the event A.

However, in practice, these estimators are infeasible when Y is observed under censoring. In the context of duration analysis, censoring appears due to lack of follow-up of the individuals. If individuals are observed along a fixed period, complete durations are not always available because either the relevant event did not occur at the end of the observation period (administrative censoring), or the individual abandoned the study.

[3] Even though identification of decomposition factors is given by analogous assumptions to those used in the policy evaluation literature, the causal interpretation requires stringent conditions on the nature of the treatment D and the control variables X.

For instance, in an unemployment duration study some individuals are still unemployed at the end of the follow-up period, while others leave the labor force. Under these circumstances, the observed sample is $\{Z_i, X_i, \delta_i, D_i\}_{i=1}^n$ of the random vector (Z, X, δ, D), where $Z = \min(Y, C)$, $\delta = 1_{\{Y \leq C\}}$ and C denotes the censoring times. In this case, the estimator $\bar{\Delta}_Y^\mu$, based on observed durations Z, turns out biased.

Consider the joint distribution of (Y, X, D), $F(y, x, \ell) = \mathbb{P}(Y \leq y,\ X \leq x, D = \ell)$, where henceforth $\leq$ is coordinatewise. Notice that, for $\ell = \{0, 1\}$, we can express $\mu_Y^{(\ell)} = \int_{\mathbb{R}} y dF(y, \infty, \ell)$, $\mu_X^{(\ell)} = \int_{\mathbb{R}^k} x dF(\infty, x, \ell)$, and

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

In fact, in the absence of censoring, $\bar{\Delta}_Y^\mu$ is the sample analog of Δ_Y^μ, where F is replaced by the sample version

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

Under censoring, a consistent estimator of F can be obtained by exploiting its representation in terms of the cumulative, or integrated, hazard function. Consider, in the context of an unemployment study, the probability that an individual, taken at random at time y from the subpopulation ℓ that belongs to the group of individuals with characteristics $\{X \in B\}$, finds a job before $y + h$. This probability can be written as,

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

where for any generic function J, $J(y-) = \lim_{x \uparrow y} J(x)$, and $B \in \boldsymbol{\beta}^k$ the smallest sigma algebra in $\mathbb{R}^k$. For practical purposes, take $B = (-\infty, x]$.

Suppose that there exists a function λ such that

$$\frac{F(y + h, x, \ell) - F(y-, x, \ell)}{1 - F(y-, \infty, \ell)} = h\lambda(y, x, \ell) \text{ as } h \to 0.$$

This function $\lambda(., x, \ell)$ is the hazard function for individuals of subpopulation ℓ with $\{X \leq x\}$. In the context of unemployment, $\lambda(y, x, \ell)$ can be interpreted as the probability that an individual belonging to subpopulation ℓ with characteristics $\{X \leq x\}$, finds a job immediately after moment y.

The associated cumulative hazard can be defined as

$$\Lambda(y,x,\ell)=\int_0^y \frac{F(d\bar{y},x,\ell)}{1-F(\bar{y}-,\infty,\ell)}, \tag{3}$$

and if λ exists, $\Lambda(y,x,\ell)=\int_0^y \lambda(\bar{y},x,\ell)\,d\bar{y}$. Using the fact that any distribution function can be expressed in terms of the corresponding integrated hazard (see Gill, 1980 and Shorack and Wellner, 2009, p. 301d, for details), we have

$$1-F(y,x,\ell)=\exp\{-\Lambda^c(y,x,\ell)\}\prod_{\bar{y}\le y}\left[1-\Lambda(\{\bar{y}\},x,\ell)\right] \tag{4}$$

where Λ^c is the continuous part of Λ, and for any generic function J, $J\{y\}=J(y)-J(y-)$. Therefore, $F(y,x,\ell)$ can be estimated by plugging-in a proper estimator of Λ. Because of the presence of censoring, the identification of Λ requires to imposing restrictions on the censoring mechanism.

Assumption 2 *The following conditions are satisfied:*
a. $\mathbb{P}(Y\le y, C\le c \mid D=\ell)=\mathbb{P}(Y\le y\mid D=\ell)\,\mathbb{P}(C\le c\mid D=\ell)$.
b. $\mathbb{P}(Y\le C\mid Y,X,D)=\mathbb{P}(Y\le C\mid Y,D)$.

This assumption has been widely used in survival analysis (c.f. Uña-Álvarez and Rodríguez-Campos, 2004; Sanchez-Sellero *et al.*, 2005 and Sant'Anna, 2014). Assumption 2.a. is the classical independence assumption that guarantees identification of the marginal distribution of survival times (c.f. Peterson, 1977). In turn, Assumption 2.b. states the relation between the censoring mechanism and the covariates so that, given the actual survival times Y, the covariates do not provide any further information on whether censoring occurs (see Stute, 1993, 1996, 1999 for further discussion). In this framework, potential dependence between C and X is allowed, and of course, it is also held when C is independent of (Y,X), which is another common assumption in survival analysis literature.

Then, under Assumption 2 we can express Λ in terms of the following sub-distributions:

$$\begin{aligned} H\left(y,\ell\right) &= \mathbb{P}\left(Z\leq y,D=\ell\right), \text{ and} \\ H_{11}\left(y,x,\ell\right) &= \mathbb{P}\left(Z\leq y,X\leq x,D=\ell,\delta=1\right). \end{aligned}$$

Proposition 1 *Under Assumption 2, the joint cumulative hazard function can be written as:*

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

The sample analogs of $H\left(y,\ell\right)$ and $H_{11}\left(y,x,\ell\right)$ are given by

$$\hat{H}\left(y,\ell\right)=n_\ell^{-1}\sum_{i=1}^n 1_{\{Z_i\leq y,D_i=\ell\}} \quad\text{and}\quad \hat{H}_{11}(y,x,\ell)=n_\ell^{-1}\sum_{i=1}^n 1_{\{Z_i\leq y,X_i\leq x,\ D_i=\ell,\delta_i=1\}}$$

and hence, $\Lambda\left(y,x,\ell\right)$ is estimated by

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

where $R_i^{(\ell)}=n_\ell\hat{H}(Z_i,\ell)$ is the rank of Z_i provided that *i-th* individual belongs to sub-population ℓ.

As a consequence, the joint distribution can be estimated by

$$\hat{F}(y,x,\ell)=1-\prod_{\bar{y}\leq y}\left[1-\hat{\Lambda}\left(\{\bar{y}\},x,\ell\right)\right]=1-\prod_{Z_{i:n_\ell}^{(\ell)}\leq y,X_{[i:n_\ell]}^{(\ell)}\leq x}\left[1-\frac{\delta_{[i:n_\ell]}^{(\ell)}}{n_\ell-R_i^{(\ell)}+1}\right] \tag{5}$$

where $Z_{1:n_\ell}^{(\ell)}\leq Z_{2:n_\ell}^{(\ell)}\leq\ldots\leq Z_{n_\ell:n_\ell}^{(\ell)}$ are the order statistics of Z in subpopulation ℓ, i.e. $Z_{i:n_\ell}^{(\ell)}=Z_j$ if $R_j^{(\ell)}=i$, and for any $\{\xi_i\}_{i=1}^{n_\ell}$, $\xi_{[i:n_\ell]}^{(\ell)}$ is the *i-th* $\xi^{(\ell)}$-concomitant of $Z_{i:n_\ell}^{(\ell)}$, that is, $\xi_{[i:n_\ell]}^{(\ell)}=\xi_j$ if $Z_{i:n_\ell}^{(\ell)}=Z_j$. This is the version of the Kaplan-Meier estimator (Kaplan and Meier, 1958) taking into account covariates .

Analogous to the univariate case, this estimator admits an additive form which can be interpreted as a Inverse-Probability-Weighting estimator in the line of Horvitz and Thompson (1952).

Corollary 1 *Under Assumption 2, the estimator of the joint distribution $F(y,x,\ell)$ can* *be written as:*

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

where

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

The weights $\left\{W_i^{(\ell)}\right\}$ represent the mass attached to the *i-th* order statistic $Z_{i:n_\ell}^{(\ell)}$, which can be obtained multiplying the status $\delta_{[i:n_\ell]}^{(\ell)}$ by the inverse of the probability of observing a failure (see Robins and Rotnitzky, 1992; Satten and Datta, 2001, for further discussion in the univariate case). Note that $\hat{F}(y,x,\ell)$ assigns zero weight to censored observations and in absence of censoring, i.e. when $\delta_{[i:n_\ell]}^{(\ell)} = 1\ \forall i$, it reduces to the multivariate empirical distribution with $W_i^{(\ell)} = n_\ell^{-1}$. Asymptotic properties of $\hat{F}(y,x,\ell)$ and the associated empirical integrals (known as *Kaplan Meier integrals*) of the form $\int \varphi(y,x,\ell)\, d\hat{F}(y,x,\ell)$, with φ an integrable function, have been studied by Stute (1993, 1996).

In this way, the OB decomposition under censoring (we call *Censored Oaxaca-Blinder*, COB hereafter) can be computed replacing F by its sample analog $\hat{F}$. In particular, the total difference Δ_Y is estimated by:

$$\hat{\Delta}_Y^{\mu} = \hat{\mu}_Y^{(1)} - \hat{\mu}_Y^{(0)}$$

where $\hat{\mu}_Y^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} Z_{i:n_\ell}^{(\ell)}$, and the counterfactual decomposition components are

$$\hat{\Delta}_Y^{\mu} = \left(\hat{\beta}_1 - \hat{\beta}_0\right)^T \hat{\mu}_X^{(1)} + \hat{\beta}_0^T \left(\hat{\mu}_X^{(1)} - \hat{\mu}_X^{(0)}\right) \tag{6}$$

where $\hat{\mu}_X^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)}$, and $\hat{\beta}_\ell$ is estimated by the weighted least squares procedure given by

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

There are other alternative regression methods for censored data (see Buckley and James, 1979; Koul *et al.*, 1981; Miller and Halpern, 1982; Ritov, 1990; Heuchenne and Keilegom, 2007); but using Kaplan Meier integrals provides a parsimonious method. For instance, it is flexible to compute functions involving both the duration outcome and the covariates, and β_ℓ is simpler to compute, avoiding the use of iterative methods and smoothers.

As a general feature in the context of censored outcomes, consistency of the estimator in Equation (6) requires additional restrictions on the support of the duration outcome and censoring times (for details, see Stute and Wang, 1993; Stute, 1995; Sanchez-Sellero *et al.*, 2005). To do so, define the subdistribution $F_Y(y,\ell) = \mathbb{P}(Y \leq y, D = \ell)$, i.e. $F_Y(y,\ell) = F(y,\infty,\ell)$, and the distribution function of the censoring times as $G(y,\ell) = \mathbb{P}(C \leq y, D = \ell)$. Additionally, for a generic distribution $J(y,\ell)$ define the least upper bound as $\tau_J^{(\ell)} = \inf\{y : J(y,\ell) = 1\} \leq \infty$.

Assumption 3 *For $\ell = \{0,1\}$, it holds that $\tau_{F_Y}^{(\ell)} \leq \tau_G^{(\ell)}$.*

Hence, if $\tau_H^{(\ell)} = \tau_{F_Y}^{(\ell)} \leq \tau_G^{(\ell)}$, estimators above are consistent over all the support. But, in the case when $\tau_H^{(\ell)} = \tau_G^{(\ell)} < \tau_{F_Y}^{(\ell)}$, the inference is restricted to $\left(0, \tilde{Z}^{(\ell)}\right]$, $\tilde{Z}^{(\ell)} \leq Z_{n_\ell:n_\ell}^{(\ell)} < \tau_H^{(\ell)}$, and estimates are typically downward biased (c.f. Gill, 1980; Mauro, 1985; Stute, 1994). Efron (1967) proposed an intuitive solution to reduce the bias by setting $\delta_{[n_\ell:n_\ell]}^{(\ell)} = 1$.

Lastly, Proposition 2 and Corollary 2 provide the basis to perform statistical inference on the counterfactual decomposition components in Equation (6). First, define:

$$\gamma_0^{(\ell)}(y) = \exp\left\{\int_0^{y-} \frac{H_0(d\bar{y},\ell)}{1-H(\bar{y},\ell)}\right\}$$

$$\gamma_1^{(\ell)}\left(y;\varphi^{(\ell)}\right) = \frac{1}{1-H(y,\ell)}\int 1_{\{y<\bar{y}\}}\varphi^{(\ell)}(\bar{y},x)\gamma_0^{(\ell)}(\bar{y})\,dH_{11}(\bar{y},x,\ell)$$

$$\gamma_2^{(\ell)}\left(y;\varphi^{(\ell)}\right) = \int\int \frac{1_{\{\bar{s}<y,\bar{s}<\bar{y}\}}\varphi^{(\ell)}(\bar{y},x)\gamma_0^{(\ell)}(\bar{y})}{\left[1-H(\bar{s},\ell)\right]^2} H_0(d\bar{s},\ell)\,H_{11}(\bar{y},x,\ell)$$

Proposition 2 *Assume that $\mathbb{E}\left(X^{(\ell)}X^{(\ell)^T}\right)$ is positive semidefinite. Under Condition 1 (in Appendix 7.1) and Assumptions 2-3, for $\ell = \{0,1\}$:*

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

where

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

$$\Sigma_{XX}^{(\ell)} = \begin{pmatrix} \mathbb{E}\left(X^{(\ell)} X^{(\ell)^T}\right) & 0 \\ 0 & I_k^{(\ell)} \end{pmatrix} \quad \textit{and} \quad \Sigma_0^{(\ell)} = \begin{pmatrix} \boldsymbol{\sigma}_{11}^{(\ell)} & \\ \boldsymbol{\sigma}_{12}^{(\ell)} & \boldsymbol{\sigma}_{22}^{(\ell)} \end{pmatrix}$$

and

$\boldsymbol{\sigma}_{ij}^{(\ell)} = \mathbb{E}\left[\boldsymbol{\varphi}_i\left(Z^{(\ell)}, X^{(\ell)}\right) \boldsymbol{\varphi}_j\left(Z^{(\ell)}, X^{(\ell)}\right) \left(\gamma_0^{(\ell)}\left(Z^{(\ell)}\right)\right)^2 \delta^{(\ell)} - \gamma_1^{(\ell)}\left(Z^{(\ell)}; \boldsymbol{\varphi}_i\right) \gamma_1^{(\ell)}\left(Z^{(\ell)}; \boldsymbol{\varphi}_j\right) \left(1 - \delta^{(\ell)}\right)\right]$

$\boldsymbol{\varphi}_i(y, x) = (\varphi_{i1}, \ldots, \varphi_{ik})$, $\varphi_{1l}(y, x) = x_l\left(y - \beta^T x\right)$, $\varphi_{2l}(y, x) = x_l - \mu_X$ *for* $1 \leq l \leq k$.

Corollary 2 *Under the same conditions as in Proposition 2 and* $\frac{n_\ell}{n} \rightarrow \rho_\ell$ *with* $\rho_0 + \rho_1 = 1$, *we have:*

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

where V_{Δ_Y}, V_{Δ_S} *and* V_{Δ_C} *are defined in Appendix 7.1.*

Accordingly, a confidence interval of $100\,(1 - 2\alpha)\,\%$ for the structure effect and composition effect are given by

$$\hat{\Delta}_S^\mu \pm \mathcal{Z}_{1-\alpha} \frac{\hat{V}_{\Delta_S}}{n^{1/2}} \quad \text{and} \quad \hat{\Delta}_C^\mu \pm \mathcal{Z}_{1-\alpha} \frac{\hat{V}_{\Delta_C}}{n^{1/2}}$$

where,

$$\hat{V}_{\Delta_S} = \frac{1}{1-\rho} \hat{\Delta}_\beta^T \hat{\boldsymbol{\sigma}}_{\mu_X}^{(1)} \hat{\Delta}_\beta + \frac{2}{1-\rho} \hat{\Delta}_\beta^T \hat{\boldsymbol{\sigma}}_{\beta\mu_X}^{(1)} \hat{\mu}_X^{(1)} + \frac{1}{\rho(1-\rho)} \hat{\mu}_X^{(1)^T} \hat{\boldsymbol{\sigma}}_\beta \hat{\mu}_X^{(1)},$$
$$\hat{V}_{\Delta_C} = \frac{1}{\rho} \hat{\Delta}_{\mu_X}^T \hat{\boldsymbol{\sigma}}_\beta^{(0)} \hat{\Delta}_{\mu_X} + \frac{2}{\rho} \hat{\beta}_0^T \hat{\boldsymbol{\sigma}}_{\beta\mu_X}^{(0)} \hat{\Delta}_{\mu_X} + \frac{1}{\rho(1-\rho)} \hat{\beta}_0^T \hat{\boldsymbol{\sigma}}_{\mu_X} \hat{\beta}_0,$$

$\mathcal{Z}_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution, $\hat{\Delta}_\beta = \hat{\beta}_1 - \hat{\beta}_0$, $\hat{\boldsymbol{\sigma}}_\beta = \rho\hat{\boldsymbol{\sigma}}_\beta^{(1)} + (1-\rho)\,\hat{\boldsymbol{\sigma}}_\beta^{(0)}$, $\hat{\Delta}_{\mu_X} = \hat{\mu}_X^{(1)} - \hat{\mu}_X^{(0)}$, $\hat{\boldsymbol{\sigma}}_{\mu_X} = \rho\hat{\boldsymbol{\sigma}}_{\mu_X}^{(1)} + (1-\rho)\,\hat{\boldsymbol{\sigma}}_{\mu_X}^{(0)}$ and $\rho_0 = \rho$. In addition, $\hat{\boldsymbol{\sigma}}_\beta^{(\ell)}$ and $\hat{\boldsymbol{\sigma}}_{\mu_X}^{(\ell)}$ are the empirical analog of $\boldsymbol{\sigma}_\beta^{(\ell)}$ and $\boldsymbol{\sigma}_{\mu_X}^{(\ell)}$ defined in Proposition 2.

In absence of censoring, these results coincide with those proposed by Jann (2005, 2008). In general, computing the asymptotic variance of the decomposition components is cumbersome (c.f. Fortin *et al.*, 2011; Rothe, 2012), even more in the case of censoring when estimating $\gamma_0^{(\ell)}$ and $\gamma_1^{(\ell)}$ is needed. A practical alternative widely used in the decomposition methods literature, is the implementation of nonparametric bootstrap techniques. Proper resampling methods for censored data are briefly described in Section 6.

3 Decomposition based on Model Specification

When the mean difference is not informative, the counterfactural decomposition of other distributional features, such as the variance or the Gini coefficient, is compelling. Consider the counterfactual distribution of the subpopulation i given the characteristics of subpopulation j, say $F_Y^{(i,j)}$, which can be defined in terms of the conditional distribution

$$F^{(\ell)}(y \mid x) = \mathbb{P}(Y \leq y, D = \ell \mid X), \ \ell = \{0,1\},$$

then,

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

with $F^{(\ell)}(x) = \mathbb{P}(X \leq x \mid D = \ell)$, $\ell = \{0,1\}$.

The validity of this *counterfactual operator* (also discussed by Rothe, 2010; Chernozhukov *et al.*, 2013 and Donald and Hsu, 2014) lies in Assumption 1. That is, when varying the covariates distribution, the conditional distribution of unobservables is not affected, and the counterfactual distribution $F_Y^{(i,j)}$ is defined as the integral of $F^{(i)}(y \mid x)$ over the covariates distribution of subpopulation j. If $F_Y^{(i,j)}$ is identifiable, the parameter $\theta\left(F_Y^{(i,j)}\right)$ can be decomposed as

$$\Delta_Y^\theta = \theta\left(F_Y^{(1,1)}\right) - \theta\left(F_Y^{(0,1)}\right) + \theta\left(F_Y^{(0,1)}\right) - \theta\left(F_Y^{(0,0)}\right) = \Delta_S^\theta + \Delta_C^\theta \tag{7}$$

where Δ_S^θ and Δ_C^θ are the corresponding structure effect and composition effect.

An estimator of the counterfactual distribution $\bar{F}^{(i,j)}$ can be obtained by plugging-in the empirical analog of the multivariate distribution of covariates,

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

i.e., we have

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

As a consequence, the estimation procedure reduces to identify and estimate properly $F^{(\ell)}(y \mid x)$. Under Assumption 2, the conditional distribution is identified; however, this can be replaced by a weaker condition, at the cost of imposing restrictions on the functional form.

Assumption 4 *For each $\ell = \{0, 1\}$, it holds that $Y^{(\ell)} \perp C^{(\ell)} | X^{(\ell)}$.*

This assumption is appropriate whenever the strong independence fails. In fact, this assumption has been taken into consideration to propose numerous generalizations of the Kaplan-Meier estimator (c.f. Beran, 1981; Dabrowska, 1987, 1989; Gonzalez-Manteiga and Cadarso-Suarez, 1994; Akritas, 1994; Leconte *et al.*, 2002).

Because of the presence of censoring, classical methods to estimate the conditional distribution, such as quantile regression and distribution regression (c.f. Chernozhukov *et al.*, 2013; Koenker *et al.*, 2013), are not valid. There are quantile regression methods available for censored data (c.f. Ying *et al.*, 1995; Lipsitz *et al.*, 1997; Bang and Tsiatis, 2002; Portnoy, 2003; Peng and Huang, 2008; Wang and Wang, 2009; Gorfine *et al.*, 2014), but these procedures are computationally very demanding, and involve inverting functions to recover the conditional distribution, to make fine approximation around the tails of the distribution and to carry out arrangements to guarantee monotonicity.

Taking into account the relation between F and Λ (see Equation –3–), and that the hazard function self-adjusts to the presence of censoring, we consider the popular proportional hazard specification proposed by Cox (1972, 1975) which assumes the following specification for the conditional hazard function:

$$\lambda^{(\ell)}(y|x) = \lambda_0^{(\ell)}(y)\, \phi(x, \beta_\ell) \tag{8}$$

where $\lambda_0^{(\ell)}$ is the baseline hazard (common risk) depending only on y and ϕ is a positive function representing the effect of the covariates on the conditional hazard function, commonly specified as $\phi(x, \beta_\ell) = \exp(\beta_\ell^T x)$. In this context a fully parametric model is possible, but these models usually force the hazard function to be monotone. In fact, the Cox model does not require any shape assumption on $\lambda_0^{(\ell)}$. Moreover, this model is flexible for incorporating time-varying covariates and unobservable heterogeneity.

As a consequence of Equation (8), the conditional cumulative hazard is

$$\Lambda_Y^{(\ell)}(y|x) = \int_0^y \lambda^{(\ell)}(\bar{y}|x)\, d\bar{y}$$

and hence, the conditional distribution is given by:

$$F^{(\ell)}(y|x) = 1 - \exp(-\Lambda_Y^{(\ell)}(y|x)) = 1 - \left[\exp(-\Lambda_0^{(\ell)}(y))\right]^{\exp(\beta_\ell^T x)}$$

where $\Lambda_0^{(\ell)}(y) = \int_0^y \lambda_0^{(\ell)}(\bar{y})\, d\bar{y}$. In order to estimate $F^{(\ell)}(y|x)$, Cox (1975) proposed the *partial likelihood method* which directly estimates β_ℓ and allows the nonparametric component $\Lambda_0^{(\ell)}(y)$ to be estimated.

With respect to the latter, there are two popular estimators in the literature. The first and the most commonly used is the Breslow estimator, $\hat{\Lambda}_{0B}^{(\ell)}$, introduced by Breslow (1974). This is given by

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

In turn, the second was proposed by Kalbfleisch and Prentice (1973), which constructs a discrete cumulative hazard that is consistent with the first order condition (or score) of the partial likelihood function, i.e.:

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

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

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

with $r^{(\ell)}(y_i)$ the pool risk in subpopulation ℓ at period y_i and $d^{(\ell)}(y_i)$ the set of individuals in subpopulation ℓ changing state at period y_i. Of course, $\hat{\Lambda}_{0B}^{(\ell)}$ presents practical advantages since it does not involve solving auxiliary equations. Both estimators perform similarly in finite sample, as will be discussed in Section 4.

Thus, an estimator for the counterfactual components of the decomposition in Equation (7), hereafter *Counterfactual Cox* decomposition (CCOX), is given by:

$$\hat{\Delta}^{\theta} = \theta\left(\hat{F}_Y^{(1,1)}\right) - \theta\left(\hat{F}_Y^{(0,1)}\right) + \theta\left(\hat{F}_Y^{(0,1)}\right) - \theta\left(\hat{F}_Y^{(0,0)}\right) \tag{9}$$

where

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

and

$$\hat{F}^{(\ell)}(y|x) = 1 - \left[\exp(-\hat{\Lambda}_0^{(\ell)}(y))\right]^{\exp(\hat{\beta}_\ell^T x)}. \tag{10}$$

The validity of the CCOX is established in Proposition 3.

Proposition 3 *Consider that Assumptions 1, 3 and 4, Condition 2 (see Appendix 7.1) and Equation 8 hold, and* $\frac{n_\ell}{n} \to \rho_\ell$ *with* $\rho_0 + \rho_1 = 1$. *Then:*

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

where $\bar{M}_{ij}$ *is a tight zero-mean Gaussian process with uniform continuous path on* $Supp(Y)$*, define as:*

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

Moreover, since the limit process of $\hat{F}_Y^{(i,j)}$ is nonpivotal (see Chernozhukov *et al.*, 2013), resampling methods are suitable for making inference on the counterfactual components (see Appendix 7.1 for further discussion).

4 Monte Carlo Simulations

To study finite sample properties of COB and CCOX, we carry out Monte Carlo experiments. These exercises allow the proposed methods to be compared with other competing alternatives, and provide evidence on the performance under different censoring scenarios and distributional assumptions.

4.1 COB Decomposition

We study the performance of the COB procedure to estimate the structure and composition effects with respect to two alternatives: the classical OB neglecting the presence of censoring and OB when censored observations are dropped. To do so, we consider Data Generator Processes (DGPs) shown in Table 1. In addition, we assume a single covariate simulated as $X^{(0)} \sim \mathcal{N}(1.5, 0.5)$ and $X^{(1)} \sim \mathcal{N}(1, 0.5)$. To adjust the censoring level to 30%, censoring times distribution are shifted by $(\upsilon_0, \upsilon_1) = (2.5, 2)$. Finally, we consider sample size of 50, 500 and 2500 and evaluate the performance of $\hat{\Delta}_S$ and $\hat{\Delta}_C$ using the average of absolute deviations across 1000 simulation draws.

Table 1 Simulation Setup

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

Results in Table 6 show that there are important differences among alternative estimators when censoring is present. For instance, if censoring is ignored, the absolute bias is not reduced as sample size increases, and most importantly, COB outperforms these alternatives.

4.2 Censoring Mechanism and Distributional Assumption

In this exercise, we examine the performance of the counterfactual distribution operator based on the Cox model under different DGPs. This allows three relevant aspects to be studied: *i.* the censoring mechanism, *ii.* the distributional assumption on duration outcome, and *iii.* the estimator of the baseline cumulative hazard. The main parameters of the simulation are presented in Table 2. It is assumed a single covariate following a uniform distribution $\mathcal{U}(0, 1)$. The scale and the shape of censoring times are shifted to generate censoring levels of 5%, 20% and 50% and sample sizes are set at 50, 500 and 2500.

To evaluate the performance of the counterfactual operator, we make comparisons with the classical Kaplan-Meier (KM) estimator and use the empirical distribution $\tilde{F}_Y$ as a benchmark. To do so, we compute three measures: MD is the maximum distance, AD

Table 2 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	$Y = 5 + X + \varepsilon_Y,\ \ \varepsilon_Y \sim N(0,1)$ $C = 5 + \varepsilon_C,\ \ \varepsilon_C \sim N(\upsilon,1)$ $\upsilon = (3, 1.5, 0.5)$
$Y \perp C \vert X$	**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	$Y = 5 + X + \varepsilon_Y,\ \ \varepsilon_Y \sim N(0,1)$ $C = 5 + X + \varepsilon_C,\ \ \varepsilon_C \sim N(\upsilon,1)$ $\upsilon = (2.5, 1, 0)$

is the average distance and MSD is the mean squared distance. To be specific,

$$MD = \max_y \left|\tilde{F}_Y(y) - \hat{F}_Y(y)\right|, \quad AD = \frac{1}{n}\sum_{i=1}^{n}\left|\tilde{F}_Y(y) - \hat{F}_Y(y)\right|$$

$$MSD = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{F}_Y(y) - \hat{F}_Y(y)\right)^2.$$

For MD and AD we report the average over 1000 draws, while for the latter, the square root of the mean value is reported.

Results in Table 7 suggest that, under the independence assumption, the KM estimator outperforms the CCOX estimator when censorship level is low. But there are not important differences with medium or heavy censoring levels (20% and 50%). In turn, under conditional independence, it is noticeable that the performance measures decrease faster for CCOX than KM with the sample size, and this fact is more remarkable as censoring becomes more substantial.

Regarding the distributional assumptions (see Tables 8 and 9), we can observe that the CCOX estimator performs fairly well even if survival times follow a normal distribution. Regarding the estimators of the baseline cumulative hazard, results are roughly the same, except for a very small sample where $\hat{\Lambda}_{0KP}$ outperforms the $\hat{\Lambda}_{0B}$ estimator. This is explained by the nature of $\hat{\Lambda}_{0KP}$ since it is proposed in the context of discrete survival times.

4.3 Decomposition Exercise and Inference

To study the finite sample performance of the CCOX to compute counterfactual decompositions beyond the mean, we consider a simulation exercise where all the difference between the two populations is due to the shape covariates distribution. In particular, we consider that $X^{(1)}$ is uniform $(0, 1)$ and $X^{(0)}$ is the sum of three independent uniform distributions in the interval $\left(0, \frac{1}{3}\right)$. Hence, we decompose the truncated mean at 15 and the quartiles. The distribution of censoring times is shifted to achieve censoring levels of 30% so that, $Y^{(\ell)} \sim \mathcal{WB}\left(e^{3-X^{(\ell)}}, 5\right)$ and $C^{(\ell)} \sim \mathcal{WB}\left(e^{3.17-X^{(\ell)}}, 5\right)$. We consider $n_\ell = 500$ and 1000 draws (Figure 1 shows a typical draw).

We estimate the decomposition given by Equation (9) using the Breslow estimator for the baseline cumulative hazard and we test the hypothesis $\Delta_S^\theta = 0$ using 1000 bootstrap repetitions. The resampling procedure is executed using the *simple* method, and coverage intervals (at 95% and 90% confidence level) are constructed according to *percentile* and *hybrid* methods (see Section 6 for details).

Results in Table 10 suggest that the coverage rate is close to its nominal value and the accuracy improves if the two subpopulations exhibit similar censoring levels. Regarding the confidence intervals, the percentile method tends to outperform the hybrid method, although the difference is quite small.

5 Unemployment Duration Gender Gaps in Spain

Spain is an interesting case to study unemployment gender gaps. First, Spain has experienced one of the highest unemployment rates among OECD countries in the recent decades. According to official statistics (OECD, 2013), the average unemployment rate in OECD countries was around 6.8% and in the US 5% for the period 1995-2005, while in Spain it was 14%. Moreover, the difference in unemployment rates by gender has also been important. For the same period, women exhibited on average an unemployment rate 9 percentage points (p.p.) higher in Spain, while in the US such a difference was rather than slight (0.04 p.p.).

There are a number of studies exploring the gender gaps in the unemployment rate (see for instance, Niemi, 1974; Johnson, 1983; Azmat *et al.*, 2006; Queneau and Sen, 2007), but other aspects of unemployment have been neglected. Therefore, we provide additional evidence of the unemployment gap in Spain by analyzing the duration rather than the rate. In particular, we estimate the total gender gap and perform counterfactual decomposition analysis to examine to what extent this gap is explained by workers' socioeconomic characteristics, labor market features or other factors.

Literature devoted to studying unemployment duration gaps has focused exclusively in explaining the difference in the average hazard rate (see Ham *et al.*, 1999, for the Czech and Slovak Republic; Gonzalo and Saarela, 2000, for Finland; Eusamio, 2004, for Spain and Portugal; Ortega, 2008, for Argentina; Du and Dong, 2009, 2009, for China; Tansel and Tasci, 2010, for Turkey; and Baussola *et al.*, 2015, for Italy and UK). In these exercises interpretation is difficult since the average hazard rate is not a parameter of the duration distribution. Instead, we use the proposed methods to decompose several parameters associated to the unemployment duration distribution.

In particular, we study the gender gap in the average unemployment duration, the probability of being long term unemployed (12 and 24 months or longer) and the Gini coefficient. While the mean gives a broad picture of the difference in unemployment duration by gender, the other two parameters allow the difference in terms of severity to be analyzed. The latter, the Gini coefficient is interesting since, analogous to income, unemployment duration has normative implications on social welfare (c.f. Paul, 1992; Borooah, 2002; Sengupta, 2009; Shorrocks, 2009a,b).

We explore two dimensions of unemployment duration: the duration until exit from unemployment and the duration until getting a job. To analyze the latter case, we follow the competing risk approach (similar to Addison and Portugal, 2003), by considering as censored, transitions from unemployment to a destination other than getting a job. The distinction of the two types of duration is important for studying the dynamics of the labor market transitions. For instance, it might be the case that women and men have similar unemployment durations, but women might be more prone to transit to inactivity. In addition, it also allows the role of the identification assumption related to the censoring mechanism to the illustrated. In the case of duration to exit from unemployment, censoring can be considered as administrative; but when the transition unemployment-to-employment is studied, censoring might not be independent since workers' characteristics affect the decision of being employed or out of the labor force.

To do so, we use information from the Survey of Income and Living Conditions (SILC) for the period 2004-2007. This survey, managed by the European Commission, is a rotative household panel that collects information on socioeconomic characteristics, including the occupational status (monthly) for a period of 4 years. Our population consists of unemployed workers older than 25 starting their unemployment spell during the period 2004-2007. We take into account a set of explanatory variables commonly used in unemployment duration studies such as age, educational level, tenure, marital status, whether the individual is head of the household, and the number of unemployed in the household (see for instance, Foley, 1997; Addison and Portugal, 2003; Kuhn and Skuterud, 2004; Biewen and Wilke, 2004; Tansel and Tasci, 2010). The first three variables control by human capital characteristics, the rest give information about the opportunity cost of being unemployed and the reservation wage. In addition, we include city size and region to control for specific labor market characteristics.

We first focus on the duration until leaving unemployment. In this case, the censoring levels are 21.4% for women and 16.2% for men. Based on $\hat{F}(y, x, \ell)$ and $\hat{F}_{Y(\ell,\ell)}$, we compute the average duration, the probability of being long term unemployed (LTU) and the Gini coefficient (see Table 3). It is noticeable that estimates for both are very similar. To give some insight about the misleading conclusions produced by ignoring censoring, the bottom part of Table 3 includes the estimates when censored observations are dropped. As expected, estimates are lower in the case of the average, and for the LTU and the Gini coefficient remarkable differences are also found.

Following the COB and CCOX methods, we compute the total difference and the decomposition components[4]. Results are presented in Table 4 coupled with confidence intervals at 90% built through 1000 bootstrap repetitions by using the percentile method.

[4]In the case of the CCOX method, we check the validity of the proportional hazard assumption using the Schoenfeld residuals. The p-values (0.031 and 0.654 for women and men, respectively) suggest that in both cases the proportional hazard specification might be suitable.

Table 3 Distributional Parameters of Duration to Exit from Unemployment

		Mean	LTU(12)	LTU(24)	Gini
Kaplan-Meier integrals	**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.

For the case of the average unemployment duration, results across methods are qualitatively the same and quantitatively similar. In general, it is observed that women present higher average duration and higher survival probability, which is consistent with the argument according to which women are less attached to the labor market. In the case of the Gini coefficient, the difference is negative indicating that men's duration distribution is more unequal, which is consistent with the fact that men leave unemployment faster, on average, but they are also severely affected by LTU.

With respect to the decomposition factors, it is found that the structure effect is statistically different from zero and plays a major role in explaining the gender gap. Although not significant, the composition effect is always positive indicating that the difference in workers' characteristics slightly increases the severity of unemployment to the detriment of women. In turn, the structure effect is positive except in the case of the Gini coefficient, suggesting that factors others than workers' characteristics, i.e. institutional factors, labor market circumstances, behavioral aspects, among others, also increase the average duration and probability of being LTU. Lastly, an interesting finding is that the LTU(24) is lower than LTU(12), and this reduction is due to the decrease in the structure effect, implying that women are relatively less prone to experience long term unemployment, which agrees with the negative sign in the Gini coefficient difference.

In the second exercise, we study the gender gaps in unemployment duration until get a job[5]. Results of this decomposition are presented in Table 5. It can be observed that, in contrast to the previous exercise, the decomposition factors differ importantly between

[5] As before, we test the validy of the proportional hazard assumption, obtaining p-vaues of 0.283 and 0.410 for woman and men respectively. We also test the presence of unobservable heterogeneity at region level, but the hypothesis was rejected.

Table 4 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.

the COB and CCOX methods, which can be related to the validity of the identification assumptions. As mentioned previously, in this case the independence assumption between survival times and censoring times might be strong. Despite the fact that the identification assumptions cannot be tested because of the unobservability of complete durations, we provide some suggestive evidence on the relation between the probability of censoring and the covariates.

Table 11 presents measures of goodness-of-fit for different probability models for the censoring indicator on the covariates. In particular, we estimate linear probability models and logit medels, and report the corresponding R^2-adjusted and *p-seudo* R^2. Overall, we observe that the covariates are relatively more important for predicting the censoring indicator for the duration until getting a job. Therefore, Assumption 2 might not be appropriate. We provide some additional evidence by performing the decomposition eliminating the censored observations (see Table 12), obtaining similar results for the two kinds of transitions we study. This exercises is appealing since assumptions on the censoring mechanism have no role.

Hence, focusing on the CCOX methods, results are qualitatively the same as in the case of duration to leave unemployment. Moreover, we observe that the magnitudes of the total differences are higher, implying that inactivity is an important destination for women. This fact is also proven by the persistent difference in the LTU. That is, the total difference in the probability of being unemployed in the long term does not decrease over duration spells.

Table 5 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.

Likewise, the composition effect turns out to be statistically significant, except for the Gini coefficient, and the structure effect has the most relevant role in explaining the unemployment duration gender gaps. The latter result has been also reported by Ham *et al.* (1999); Gonzalo and Saarela (2000); Eusamio (2004) and Ortega (2008) who study the average hazard rate. It is remarkable that this *unexplained* component is associated to many factors involved in the job search process, i.e. the behavior of workers and employers and different circumstances of the labor market such as the labor market tightness, and discrimination, among others. Thus, these results point out the importance of deeply studying such factors to assess the differential gender effect of labor market policies.

6 Final Remarks and Further Research

We have proposed inferential tools to perform counterfactual decompositions under censoring. These tools encompass decompositions for the mean difference as well as the difference in other distributional features. For the mean difference decomposition, we provide asymptotic results useful to test statistically the significance of the decomposition components. However, the form of the variances tend to be cumbersome to compute. Additionally, in the case of the decomposition based on models, the limit processes are non-pivotal. Thus, bootstrapping methods turn out to be practical alternative for performing statistical inference. The implementation of statistical inference based on bootstrapping techniques is more accurate than first-order asymptotic approximation (Hall, 1992).

In the context of censored data, Efron (1981) presents two alternative resampling schemes that have been recognized in the literature as the *simple* bootstrap method and the *obvious* bootstrap method. In short, the simple method consists in drawing bootstrap samples $(Z^*, X^*, D^*, \delta^*)$ by independent sampling of size n with replacement and assigning equal mass n^{-1} at each selected observation. Instead, the obvious method requires estimating the distribution of the survival times and censoring times. In particular, for each draw (X_i^*, D_i^*), compute $Y_i^* \sim \hat{F}^{(\ell)}(y \mid x)$, and $C_i^* \sim \hat{G}^{(\ell)}(y \mid x)$ and define $Z_i^* = \min(Y_i^*, C_i^*)$ and $\delta_i^* = 1_{\{Y_i^* \leq C_i^*\}}$. Under independence between Y and C, these methods are equivalent (c.f. Efron and Tibshirani, 1986). The simple method has important practical convenience because it does not require imposing any assumption on the structure of the data and does not depend on the censoring mechanism.

To construct confidence bands, we consider classical methods such as *percentile* and *hybrid* (see Hall, 1988; Efron, 1992; Burr, 1994, for a detailed comparison of coverage bands construction methods). One important advantage of these methods is that estimation of variances is not needed. In order to describe the pivotal quantities, suppose we are interested in forming $100(1-2\alpha)\%$ confidence bands for the target parameter θ. Denote the estimated parameter from a bootstrap sample as $\hat{\theta}^*$ and its distribution given by J. The percentile method sets the confidence interval as:

$$\left(J^{-1}(\alpha), \quad J^{-1}(1-\alpha)\right)$$

Instead of approximating the distribution of $\hat{\theta}^*$, the hybrid method approximates the distribution of $\left(\hat{\theta} - \theta\right)$ through the distribution of $\left(\hat{\theta}^* - \hat{\theta}\right)$. Therefore, the coverage interval is defined as:

$$\left(2\hat{\theta} - J^{-1}(1-\alpha), \quad 2\hat{\theta} - J^{-1}(\alpha)\right)$$

There is not a general rule to select the proper method. For instance, in the particular case of censored data, considering real-valued and function-valued parameters estimated through the Cox model, Burr (1994) makes comparative analysis of bootstrap confidence intervals combining resampling and interval construction methods. The results suggest that there is no single winner and the pertinence of each method depends on the target parameter.

The COB method can be extended either to detailed decomposition or to approximate decomposition of other distributional features. With respect to the former, the usual procedure can be implemented by addressing the identification issues related to the path dependence (c.f. Firpo *et al.*, 2007; Firpo and Pinto, 2011; Rothe, 2012) and the omitted group problem (c.f. Oaxaca and Ransom, 1999; Gardeazabal and Ugidos, 2004; Yun, 2005). In turn, our weighted regression method can be adapted to estimate the conditional recentered influence function for several distributional statistics as proposed in Firpo *et al.* (2009) by applying the proper transformation in the dependent variable. A list of RIFs for relevant distributional parameters can be found in Firpo *et al.* (2007) and Essama-Nssah and Lambert (2011).

In the decomposition methods based on the specification of the conditional distribution, we consider a proportional hazard model. Naturally, in some applications the proportional hazard model might induce misspecification, and hence, more flexible models for the conditional distribution are needed. In this context the use of nonparametric models might be complicated since it is usual to deal with a large number of covariates, and establishing the limit process of the counterfactual operator would require further theoretical work (Rothe, 2010). In this respect, other semiparametric approaches such as distributional regression-type methods in the spirit of Foresi and Peracchi (1995) and Chernozhukov *et al.* (2013) are a compelling alternative. We pursue this approach in separate ongoing work.

References

Addison, J. and Portugal, P., 2003. Unemployment Duration Competing and Defective Risks, *Journal of Human Resources*, 38 (1), 157–191.

Akritas, M., 1994. Nearest Neighbor Estimation of a Bivariate Distribution Under Random Censoring, *The Annals of Statistics*, 22 (3), 1299–1327.

Andersen, P. and Gill, R., 1982. Cox's Regression Model for Counting Processes: a Large Sample Study, *The Annals of Statistics*, 10 (4), 1100–1120.

Azarang, L., Uña-Álvarez, J.D., and Stute, W., 2015. The Jackknife Estimate of Covariance of two Kaplan-Meier Integrals with Covariables, *Statistics: A Journal of Theoretical and Applied Statistics*, 49 (3), 1005–1025.

Azmat, G., Güell, M., and Manning, A., 2006. Gender Gaps in Unemployment Rates in OECD Countries, *Journal of Labor Economics*, 24 (1), 1–37.

Bailey, K., 1983. The Asymptotic Joint Distribution of Regression and Survival Parameter Estimates in the Cox Regression Model, *The Annals of Statistics*, 11 (1), 39–48.

Bailey, K., 1984. Asymptotic Equivalence Between the Cox Estimator and the General ML Estimators of Regression and Survival Parameters in the Cox Model, *The Annals of Statistics*, 12 (2), 730–736.

Bang, H. and Tsiatis, A.a., 2002. Median Regression with Censored Cost Data., *Biometrics*, 58 (3), 643–649.

Baussola, M., Mussida, C., Jenkins, J., and Penfold, M., 2015. Determinants of the Unemployment Gender Gap: A Comparative Investigation, *International Labour Review*, Forthcomin.

Beran, R., 1981. Nonparametric Regression with Randomly Censored Survival Data, Tech. rep., Univ. California, Berkeley.

Biewen, M. and Wilke, R., 2004. Unemployment Duration in West-Germany: Do the IAB Employment Subsample and the German Socio-Economic Panel Yield the Same Results?

Blau, F.D. and Kahn, L.M., 1992. The Gender Earnings Gap: Learning from International Comparisons, *The American Economic Review*, 82 (2), 533–538.

Blinder, A., 1973. Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources*, 8 (4), 436–455.

Borooah, V.K., 2002. A Duration-Sensitive Measure of the Unemployment Rate: Theory and Application, *Labour*, 16 (3), 453–468.

Breslow, N., 1974. Covariance Analysis of Censored Survival Data, *Biometrics*, 30 (1), 89–99.

Buckley, J. and James, I., 1979. Data with Censored Linear Regression, *Biometrika*, 66 (3), 429–436.

Burr, D., 1994. A comparison of Certain Bootstrap Confidence Intervals in the Cox Model, *Journal of the American Statistical Association*, 89 (428), 1290–1302.

Cain, G., 1986. The Economic Analysis of Labor Market Discrimination: A Survey, *in:* R. Ashenfelter, O and Layard, ed., *Handbook of Labor Economics*, Elsevier Science Publishers, vol. L, chap. 13, 1st ed.

Cartwright, N. and Reiss, J., 2004. Uncertainty in Econometrics: Evaluating Policy Counterfactuals, *in: Economic Policy Under Uncertainty: The Role of Truth and Accountability in Policy Advice.*, July, 1–29.

Chen, L., Lin, D.Y., and Zeng, D., 2010. Attributable Fraction Functions for Censored Event Times., *Biometrika*, 97 (3), 713–726.

Cheng, G. and Huang, J.Z., 2010. Bootstrap Consistency for General Semiparametric M-estimation, *The Annals of Statistics*, 38 (5), 2884–2915.

Chernozhukov, V., Fernández-Val, I., and Melly, B., 2013. Inference on Counterfactual Distributions, *Econometrica*, 81 (6), 2205–2268.

Cotton, J., 1988. On the Decomposition of Wage Differentials, *The Review of Economics and Statistics*, 70 (2), 236–243.

Cox, D., 1975. Partial Likelihood, *Biometrika*, 62 (2), 269–276.

Cox, D.R., 1972. Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, 34 (2), 187–220.

Dabrowska, D., 1987. Nonparametric Regression with Censored Survival Time Data, *Scandinavian Journal of Statistics*, 14 (3), 181–197.

Dabrowska, D., 1989. Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate, *The Annals of Statistics*, 17 (3), 1157–1167.

Dawid, A., 2000. Causal inference Without Counterfactuals, *Journal of the American Statistical Association*, 95 (450), 407–424.

DiNardo, J., Fortin, N., and Lemieux, T., 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, *Econometrica*, 64 (5), 1001–1044.

Donald, S.G. and Hsu, Y.C., 2014. Estimation and inference for Distribution Functions and Quantile Functions in Treatment Effect Models, *Journal of Econometrics*, 178 (3), 383–397.

Du, F.L. and Dong, X.Y., 2009. Why Women Have Longer Unemplyoment Durations than Men in Post-restructuring Urban China, *Cambridge Journal of Economics*, 33 (2), 233–252.

Efron, B., 1967. The Two Sample Problem with Censored Data, *in: Proceedings of the fifth Berkeley Symposium on Math, Statistics and Probability*, 831–853.

Efron, B., 1981. Censored Data and the Bootstrap, *Journal of the American Statistical Association*, 76 (374), 312–319.

Efron, B., 1992. Six Questions Raised by the Bootstrap, *Exploring the Limits of Bootstrap*, 99–126.

Efron, B. and Tibshirani, R., 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy, *Statistical science*, 1 (1), 54–75.

Essama-Nssah, B. and Lambert, P.J., 2011. Influence Functions for Distributional Statistics.

Eusamio, E., 2004. El Diferencial de las Tasas de Paro de Hombres y Mujeres en España (1994-1998).

Firpo, S., Fortin, N., and Lemieux, T., 2009. Unconditional Quantile Regressions, *Econometrica*, 77 (3), 953–973.

Firpo, S., Fortin, N.M., and Lemieux, T., 2007. Decomposing Wage Distributions Using Recentered Influence Function Regressions.

Firpo, S. and Pinto, C., 2011. Identification and Estimation of Distributional Impacts of Interventions using Changes in Inequality Measures.

Foley, M., 1997. Determinants of Unemployment Duration in Russia.

Foresi, S. and Peracchi, F., 1995. The Conditional Distribution of Excess Returns: An Empirical Analysis, *Journal of the American Statistical*, 90 (430), 451–466.

Fortin, N., Lemieux, T., and Firpo, S., 2011. Decomposition Methods in Economics, *in: Handbook of Labor Economics*, 1–102, 4th ed.

Freeman R.B., 1980. Unionism and the Dispersion of Wages, *Industrial and Labor Relations Review*, 34 (1), 3–23.

Freitag, G. and Munk, A., 2005. On Hadamard Hifferentiability in k-sample Semiparametric Models with Applications to the Assessment of Structural Relationships, *Journal of Multivariate Analysis*, 94 (1), 123–158.

Gardeazabal, J. and Ugidos, A., 2004. More on Identification in Detailed Wage Decompositions, *Review of Economics and Statistics*, 86 (4), 1034–1036.

Gill, R., 1980. Censoring and Stochastic Integrals, *Statistica Neerlandica*, 2, 124–124.

Gill, R., 1984. Understanding Cox's Regression Model: a Martingale Approach, *Journal of the American Statistical Association*, 79 (386), 441–447.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C., 1994. Asymptotic Properties of a Generalized Kaplan-Meier Estimator with some Applications, *Journal of Nonparametric Statistics*, 4 (1), 65–78.

Gonzalo, M.T. and Saarela, J., 2000. Gender Differences in Exit Rates fom Unemployment: Evidence from a Local Finnish Labor Market, *Finnish Economic Papers*, 13 (2), 129–139.

Gorfine, M., Goldberg, Y., and Ritov, Y., 2014. A Quantile Regression Model for Failure-time Data.

Hall, P., 1988. Theoretical Comparison of Bootstrap Confidence Intervals, *The Annals of Statistics*, 16 (3), 927–953.

Hall, P., 1992. *The Bootstrap and Edgeworth Expansion*, New York: Springer.

Ham, J.C., Svejnar, J., and Terrell, K., 1999. Women's Unemployment During Transition, *Economics of Transition*, 7 (1), 47–78.

Heuchenne, C. and Keilegom, I., 2007. Polynomial Regression with Censored Data based on Preliminary Nonparametric Estimation, *Annals of the Institute of Statistical Mathematics*, 59 (2), 273–297.

Hirose, Y., 2011. Asymptotic Linear Expansion of Profile Likelihood in the Cox Model, *Mathematical Methods of Statistics*, 20 (3), 224–231.

Höfler, M., 2005. Causal Inference Based on Counterfactuals, *BMC Medical Research Methodology*, 5, 28.

Horvitz, D. and Thompson, D., 1952. A Generalization of Sampling Without Replacement from a Finite Universe, *Journal of the American Statistical Association*, 47 (260), 663–685.

Jann, B., 2005. Standard Errors for the Blinder-Oaxaca Decomposition, *3rd German Stata Users' Group Meetings*.

Jann, B., 2008. The Blinder-Oaxaca Decomposition for Linear Regression Models, *Stata Journal*, 8 (4), 453–479.

Johnson, J.L., 1983. Sex Differentials in Unemployment Rates: A Case for No Concern, *Journal of Political Economy*, 91 (2), 293–303.

Juhn, C., Murphy, K.M., and Pierce, B., 1991. Accounting for the Slowdown in Black-White Wage Convergence, *in:* M.H. Kosters, ed., *Workers and their Wages*, Washington DC: American Enterprise Institute, 107–143.

Juhn, C., Murphy, K.M., and Pierce, B., 1993. Wage Inequality and the Rise in Returns to Skill, *Journal of Political Economy*, 101 (3), 410–442.

Kalbfleisch, J. and Prentice, R., 1973. Marginal Likelihoods based on Cox's Regression and Life Model, *Biometrika*, 60 (2), 267–278.

Kaplan, E. and Meier, P., 1958. Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, 53 (282), 457–481.

Koenker, R., Leorato, S., and Peracchi, F., 2013. Distributional vs. Quantile Regression.

Koul, H., Susarla, V., and Ryzin, J.V., 1981. Regression Analysis with Randomly Right-Censored Data, *The Annals of Statistics*, 9 (6), 1276–1288.

Kuhn, P. and Skuterud, M., 2004. Internet Job Search and Unemployment Durations, *American Economic Review*, 94 (1), 218–232.

Leconte, E., Poiraud-Casanova, S., and Thomas-Agnan, C., 2002. Smooth Conditional Distribution Function and Quantiles under Random Censorship., *Lifetime data analysis*, 8 (3), 229–46.

Lewis, D., 1973. *Counterfactuals*, Oxford: B. Blackwell.

Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G., and Zhao, L.P., 1997. Quantile Regression Methods for Longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46 (4), 463–476.

Machado, J..F. and Mata, J., 2005. Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression, *Journal of Applied Econometrics*, 20 (4), 445–465.

Mauro, D., 1985. A Combinatoric Approach to the Kaplan-Meier Estimator, *The Annals of Statistics*, 13 (1), 142–149.

McLain, A. and Ghosh, S., 2011. Nonparametric Estimation of the Conditional Mean Residual Life Function based on Censored Data, *Lifetime data analysis*, 14 (4), 514–532.

Melly, B., 2005. Decomposition of Differences in Distribution using Quantile Regression, *Labour Economics*, 12 (4), 577–590.

Miller, R. and Halpern, J., 1982. Regression with Censored Data, *Biometrika*, 69 (3), 521–531.

Naes, T., 1982. The Asymptotic Distribution of the Estimator for the Regression Parameter in Cox's Regression Model, *Scandinavian Journal of Statistics*, 9 (2), 107–115.

Niemi, B., 1974. The Female-Male Differential in Unemployment, *Industrial and Labor Relations Review*, 27 (3), 331–350.

Oaxaca, R., 1973. Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review*, 14 (3), 693–709.

Oaxaca, R.L. and Ransom, M.R., 1999. Identification in Detailed Wage Decompositions, *Review of Economics and Statistics*, 81 (1), 154–157.

OECD, 2013. http://stats.oecd.org/.

O'Neill, J. and O'Neill, D., 2006. What do Wage Differentials Tell us About Labor Market Discrimination?, *in: The Economics of Immigration and Social Policy*, 293–357.

Ortega, A., 2008. Gender Gaps in Unemployment Rates in Argentina, *Economica*, 54 (1-2), 161–202.

Paul, S., 1992. An Illfare Approach to the Measurement of Unemployment, *Applied Economics*, 24 (7), 739–743.

Pearl, J., 2009. Causal Inference in Statistics: An Overview, *Statistics Surveys*, 3, 96–146.

Peng, L. and Huang, Y., 2008. Survival Analysis with Quantile Regression Models, *Journal of the American Statistical Association*, 103 (482), 637–649.

Peterson, A., 1977. Expressing the Kaplan-Meier Estimator as a Function of Empirical Subsurvival Functions, *Journal of the American Statistical Association*, 72 (360), 854–858.

Portnoy, S., 2003. Censored Regression Quantiles, *Journal of the American Statistical Association*, 98 (464), 1001–1012.

Powers, D.a. and Yun, M.S., 2009. Multivariate Decomposition for Hazard Rate Models, *Sociological Methodology*, 39 (1), 233–263.

Queneau, H. and Sen, A., 2007. Evidence Regarding Persistence in the Gender Unemployment Gap Based on the Ratio of Female to Male Unemployment rate, *Economics Bulletin*, 5 (23), 1–10.

Reiss, J. and Cartwright, N., 2004. Uncertainty in Econometrics: Evaluating Policy Counterfactuals.

Ritov, Y., 1990. Estimation in a Linear Regression Model with Censored Data, *The Annals of Statistics*, 18 (1), 303–328.

Robins, J. and Rotnitzky, A., 1992. Recovery of Information and Adjustment for Dependent Censoring using Surrogate Markers, *AIDS Epidemiology*, 297–331.

Rothe, C., 2010. Nonparametric Estimation of Distributional Policy Effects, *Journal of Econometrics*, 155 (1), 56–70.

Rothe, C., 2012. Decomposing the Composition Effect.

Rubin, D.B., 1974. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies., *Journal of Educational Psychology*, 66 (5), 688–701.

Rubin, D.B., 2005. Causal Inference Using Potential Outcomes, *Journal of the American Statistical Association*, 100 (469), 322–331.

Sanchez-Sellero, C., Gonzalez-Manteiga, W., and van Keilegom, I., 2005. Uniform Representation of Product-Limit Integrals with Applications, *Scandinavian Journal of Statistics*, 32, 563–581.

Sant'Anna, P., 2014. Nonparametric Tests for Treatment Effects with Duration Outcomes.

Satten, G. and Datta, S., 2001. The Kaplan-Meier Estimator as an Inverse Probability of Censoring Weighted Average, *The American Statistician*, 55 (3), 207–210.

Sengupta, M., 2009. Unemployment Duration and the Measurement of Unemployment, *The Journal of Economic Inequality*, 7 (3), 273–294.

Shorack, G. and Wellner, J., 2009. *Empirical Processes with Applications to Statistics.*

Shorrocks, A., 2009a. On the Measurement of Unemployment, *Journal of Economic Inequality*, 7 (3), 311–327.

Shorrocks, A., 2009b. Spell Incidence, Spell Duration and the Measurement of Unemployment, *Journal of Economic Inequality*, 7 (3), 295–310.

Stute, W., 1993. Consistent Estimation under Random Censorship when Covariables are Present, *Journal of Multivariate Analysis*, 45, 89–103.

Stute, W., 1994. The Bias of Kaplan-Meier Integrals, *Scandinavian Journal of Statistics*, 21 (4), 475–484.

Stute, W., 1995. The Central Limit Theorem under Random Censorship, *The Annals of Statistics*, 23 (2), 422–439.

Stute, W., 1996. Distributional Convergence under Random Censorship when Covariables are Present, *Scandinavian Journal of Statistics*, 23 (4), 461–471.

Stute, W., 1999. Nonlinear Censored Regression, *Statistica Sinica*, 9, 1089–1102.

Stute, W. and Wang, J., 1993. The Strong Law under Random Censorship, *The Annals of Statistics*, 21 (3), 1591–1607.

Tansel, A. and Tasci, H., 2010. Hazard Analysis of Unemployment Duration by Gender in a Developing Country: The Case of Turkey, *Labour*, 24 (4), 501–530.

Tsiatis, A., 1981. A Large Sample Study of Cox's Regression Model, *The Annals of Statistics*, 9 (1), 93–108.

Uña-Álvarez, J. and Rodríguez-Campos, M.C., 2004. Strong Consistency of Presmoothed Kaplan-Meier Integrals when Covariables are Present, *Statistics*, 38 (6), 483–496.

van der Vaart, A. and Wellner, J., 2004. Weak Convergence and Empirical Processes: With Applications to Statistics, *Springer, New York*.

Wang, H.J. and Wang, L., 2009. Locally Weighted Censored Quantile Regression, *Journal of the American Statistical Association*, 104 (487), 1–32.

Ying, Z., Jung, S., and Wei, L., 1995. Survival analysis with median regression models, *Journal of the American Statistical Association*, 90 (429), 178–184.

Yun, M.S., 2005. A Simple Solution to the IdentificationProblem in Detailed Wage Decompositions, *Economic Inquiry*, 43 (4), 766–772.

7 Appendix

7.1 Some Theoretical Results

Proof of Proposition 1 and Estimation of the Joint Distribution

To achieve identification of the joint distribution we consider Assumption 2. Under Assumption 2.b. we have:

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

and by Assumption 2.a.

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

Thus, using Equation (3)

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

Using the sample version of H_{11} and H given by

$$\hat{H}(y,\ell) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{Z_i\le y, D_i=\ell\}} \quad and \quad \hat{H}_{11}(y,x,\ell) = n_\ell^{-1}\sum_{i=1}^{n} 1_{\{Z_i\le y, X_i\le x,\ D_i=\ell, \delta_i=1\}}$$

the jump $\hat{\Lambda}$ is defined as

$$\hat{\Lambda}(\{\bar{y}\},x,\ell) = \frac{\delta^{(\ell)}_{[i:n_\ell]}}{n_\ell - R_i^{(\ell)}+1}.$$

Therefore, $\hat{F}(y,x,\ell)$ can be estimated by:

$$\hat{F}(y,x,\ell) = 1-\prod_{\bar{y}\le y}\left[1-\hat{\Lambda}(\{\bar{y}\},x,\ell)\right] = 1-\prod_{Z^{(\ell)}_{i:n_\ell}\le y, X^{(\ell)}_{[i:n_\ell]}\le x}\left[1-\frac{\delta^{(\ell)}_{[i:n_\ell]}}{n_\ell - R_i^{(\ell)}+1}\right]$$

Proof of Corollary 1

Define $S(y,\ell) = 1-F_Y(y,\ell)$. The joint distribution can be written as:

$$\begin{aligned}F(y,x,\ell) &= \int_0^y F(d\bar{y},x,\ell) \\ &= \int_0^y \left[1-F_Y(\bar{y}-,\ell)\right]\frac{F(d\bar{y},x,\ell)}{\left[1-F_Y(\bar{y}-,\ell)\right]} \\ &= \int_0^y S(\bar{y},\ell)\,\Lambda(d\bar{y},x,\ell) \\ &= \int_0^y S(\bar{y},\ell)\,\frac{H_{11}(d\bar{y},x,\ell)}{\left[1-H(\bar{y}-,\ell)\right]}.\end{aligned}$$

H_{11} and H can be estimated from available data and S_Y is consistently estimated by the Kaplan-Meier estimator (by Assumption 2.a.). Thus,

$$\hat{F}(y,x,\ell) = \int_0^y \hat{S}(\bar{y},\ell) \frac{\hat{H}_{11}(d\bar{y},x,\ell)}{\left[1-\hat{H}(\bar{y}-,\ell)\right]}$$

where,

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

Thus, an estimator of the joint distribution is given by:

$$\begin{aligned}
\hat{F}(y,x,\ell) &= \sum_{i=1}^{n} \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{[j:n_\ell]}^{(\ell)}}{n_\ell - R_j^{(\ell)} + 1}\right] \frac{1_{\left\{Z_{i:n_\ell}^{(\ell)} \le y, X_{[i:n_\ell]}^{(\ell)} \le x, \delta_{[i:n_\ell]}^{(\ell)} = 1\right\}}}{n_\ell - R_i^{(\ell)} + 1} \\
&= \sum_{i=1}^{n} \left\{ \prod_{j=1}^{i-1} \left[1 - \frac{\delta_{[j:n_\ell]}^{(\ell)}}{n_\ell - R_j^{(\ell)} + 1}\right] \frac{\delta_{[i:n_\ell]}^{(\ell)}}{n - R_i^{(\ell)} + 1} \right\} 1_{\left\{Z_{i:n_\ell}^{(\ell)} \le y, X_{[i:n_\ell]}^{(\ell)} \le x\right\}} \\
&= \sum_{i=1}^{n} W_i^{(\ell)} 1_{\left\{Z_{i:n_\ell}^{(\ell)} \le y, X_{[i:n_\ell]}^{(\ell)} \le x\right\}}
\end{aligned}$$

where

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

Proof of Proposition 2

Before state the main result, we require some regularity conditions on the generic integrable functions $\varphi^{(\ell)}$.

Condition 1 *Consider that following integrability conditions holds:*

$$a.\ \int \left[\varphi^{(\ell)}(Z,X)\gamma_0^{(\ell)}(Z)\,\delta^{(\ell)}\right] d\mathbb{P} < \infty$$

$$b.\ \int \left|\varphi^{(\ell)}(\bar{y},\bar{x})\right| K^{1/2}(\bar{y},\ell)\, d\mathbb{P} < \infty$$

with

$$K(y,\ell) = \int_0^{y-} \frac{G(d\bar{y},\ell)}{\left[1-H(\bar{y},\ell)\right]\left[1-G(\bar{y},\ell)\right]}$$

Condition 1.a. generalizes the second order assumption on $\varphi^{(\ell)}$ so that when censoring is not presence, this condition states that the second moment is finite. In turn, Condition 1.b controls the bias of $\mathbb{E}\left[\varphi^{(\ell)}(y,x)\right]$ and guarantees that censoring effects does not dominate in the right tail.

Now define:

$$Q_{XX}^{(\ell)} = \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} X_{[i:n_\ell]}^{(\ell)^T}$$

By Theorem 1 in Stute and Wang (1993), we have $Q_{XX}^{(\ell)} \longrightarrow \mathbb{E}\left[X^{(\ell)} X^{(\ell)^T}\right]$.

To compute the joint distribution of $\left[\left(\hat{\beta}_\ell - \beta_\ell\right), \left(\hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)}\right)\right]$, note that we can write:

$$\hat{\beta}_\ell = Q_{XX}^{(\ell)^{-1}} \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} Z_{i:n_\ell}^{(\ell)}.$$

For all i, we know that $\delta_{[i:n_\ell]}^{(\ell)} Z_{i:n_\ell}^{(\ell)} = \delta_{[i:n_\ell]}^{(\ell)} Y_{i:n_\ell}^{(\ell)}$. Then, we have:

$$\hat{\beta}_\ell - \beta_\ell = Q_{XX}^{(\ell)^{-1}} \sum_{i=1}^{n_\ell} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} \varepsilon_{i:n_\ell}^{(\ell)}$$

Therefore,

$$\begin{aligned}
\begin{pmatrix} \hat{\beta}_\ell - \beta_\ell \\ \hat{\mu}_X^{(\ell)} - \mu_X^{(\ell)} \end{pmatrix} &= \begin{pmatrix} Q_{XX}^{(\ell)^{-1}} & 0 \\ 0 & I_K \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} \varepsilon_{[i:n_\ell]} \\ \sum_{i=1}^{n} W_i^{(\ell)} X_{[i:n_\ell]}^{(\ell)} - \mu_X^{(\ell)} \end{pmatrix} \\
&= \boldsymbol{Q}_{XX}^{(\ell)^{-1}} \begin{pmatrix} \sum_{i=1}^{n} W_i^{(\ell)} \boldsymbol{\varphi}_1^{(\ell)}(Z_i^{(\ell)}, X_i^{(\ell)}) \\ \sum_{i=1}^{n} W_i^{(\ell)} \boldsymbol{\varphi}_2^{(\ell)}(Z_i^{(\ell)}, X_i^{(\ell)}) \end{pmatrix} \\
&= \boldsymbol{Q}_{XX}^{(\ell)^{-1}} \boldsymbol{U}
\end{aligned}$$

where:

- $\boldsymbol{\varphi}_i\,(y,x) = (\varphi_{i1}, \ldots, \varphi_{ik})$
- $\varphi_{1l}\,(y,x) = x_l\left(y - \beta^T x\right) \quad \text{for } 1 \le l \le k$
- $\varphi_{2l}\,(y,x) = x_l - \mu_X^{(\ell)} \quad \text{for } 1 \le l \le k.$

From the SLLN (Stute, 1993), it follows that $\boldsymbol{U} \longrightarrow \mathbf{0}$. For a generic integrable $\varphi^{(\ell)}$, under Condition 1, $Q_{\varphi}^{(\ell)} = \int \varphi^{(\ell)}(y,x)dF$ admits the following representation (see Stute, 1996 for details):

$$
\begin{aligned}
Q_{\varphi}^{(\ell)} &= \sum_{i=1}^{n_\ell} W_i^{(\ell)} \varphi(Z_{i:n_\ell}^{(\ell)}, X_{[i:n_\ell]}^{(\ell)}) \\
&= \frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\left[\varphi^{(\ell)}(Z_i^{(\ell)}, X_i^{(\ell)})\gamma_0^{(\ell)}\left(Z_i^{(\ell)}\right)\delta_i^{(\ell)} + \gamma_1^{(\ell)}\left(Z_i^{(\ell)};\varphi^{(\ell)}\right)\left(1-\delta_i^{(\ell)}\right) - \gamma_2^{(\ell)}\left(Z_i^{(\ell)};\varphi^{(\ell)}\right)\right] \\
&\quad + o_{\mathbb{P}}\left(n_\ell^{-1/2}\right) \\
&= \frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\eta_i^{(\ell)}\left(Z_i^{(\ell)}, X_i^{(\ell)};\varphi^{(\ell)}\right) + o_{\mathbb{P}}\left(n_\ell^{-1/2}\right)
\end{aligned}
$$

with

$$
\begin{aligned}
\eta_i^{(\ell)}\left(Z_i^{(\ell)}, X_i^{(\ell)};\varphi^{(\ell)}\right) &= \varphi^{(\ell)}(Z_i^{(\ell)}, X_i^{(\ell)})\gamma_0^{(\ell)}\left(Z_i^{(\ell)}\right)\delta_i^{(\ell)} + \gamma_1^{(\ell)}\left(Z_i^{(\ell)};\varphi^{(\ell)}\right)\left(1-\delta_i^{(\ell)}\right) \qquad (11) \\
&\quad -\gamma_2^{(\ell)}\left(Z_i^{(\ell)};\varphi^{(\ell)}\right) \\
&= A_i^{(\ell)} + B_i^{(\ell)} - C_i^{(\ell)}
\end{aligned}
$$

a sum of *iid* quantities such that:

- $\mathbb{E}\left[A_i^{(\ell)}\right] = \mathbb{E}\left[\varphi^{(\ell)}(Z_i^{(\ell)}, X_i^{(\ell)})\right]$
- $\mathbb{E}\left[B_i^{(\ell)}\right] = \mathbb{E}\left[C_i^{(\ell)}\right]$

In such manner, we have the following result:

$$
n_\ell^{1/2}\boldsymbol{U} \xrightarrow{d} \mathcal{N}_{2k}\left(0, \Sigma_0^{(\ell)}\right)
$$

where

$$
\Sigma^{(\ell)} = \begin{pmatrix} \boldsymbol{\sigma}_{11}^{(\ell)} & \cdot \\ \boldsymbol{\sigma}_{12}^{(\ell)} & \boldsymbol{\sigma}_{22}^{(\ell)} \end{pmatrix}
$$

$$
\boldsymbol{\sigma}_{ij}^{(\ell)} = \mathbb{C}ov\left[\boldsymbol{\eta}^{(\ell)}\left(Z^{(\ell)}, X^{(\ell)};\boldsymbol{\varphi}_i^{(\ell)}\right), \boldsymbol{\eta}^{(\ell)}\left(Z^{(\ell)}, X^{(\ell)};\boldsymbol{\varphi}_j^{(\ell)}\right)\right]
$$

where each element of the vector $\boldsymbol{\eta}^{(\ell)}$ can be written as in Equation (11).

To simplify notation, let's omit "$^{(\ell)}$".Since each component of $\boldsymbol{\eta}$ has zero mean, the covariance can be written as:

$$\begin{aligned}
\mathbb{C}ov\left[\boldsymbol{\eta}\left(Z,X;\boldsymbol{\varphi}_i\right),\boldsymbol{\eta}\left(Z,X;\boldsymbol{\varphi}_j\right)\right] &= \mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{A}_j\right)+\mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{B}_j\right)-\mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{C}_j\right) \\
&\quad +\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{A}_j\right)+\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_j\right)-\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{C}_j\right) \\
&\quad -\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{A}_j\right)-\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{B}_j\right)+\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{C}_j\right)
\end{aligned}$$

Azarang *et al.* (2015) shown that:

- $\mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{C}_j\right)=\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_j\right)$
- $\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{C}_j\right)=\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{C}_j\right)+\mathbb{E}\left(\boldsymbol{C}_i\boldsymbol{B}_j\right)$

In addition, $\mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{B}_j\right)=\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{A}_j\right)=0$ so that the covariance becomes:

$$\begin{aligned}
\boldsymbol{\sigma}_{ij} &= \mathbb{E}\left(\boldsymbol{A}_i\boldsymbol{A}_j\right)-\mathbb{E}\left(\boldsymbol{B}_i\boldsymbol{B}_j\right) \\
&= \mathbb{E}\left[\boldsymbol{\varphi}_i\left(Z,X\right)\boldsymbol{\varphi}_j\left(Z,X\right)\left(\gamma_0\left(Z\right)\right)^2\delta-\gamma_1\left(Z;\boldsymbol{\varphi}_i\right)\gamma_1\left(Z;\boldsymbol{\varphi}_j\right)\left(1-\delta\right)\right]
\end{aligned}$$

Finally, as

$$\boldsymbol{Q}_{XX}^{(\ell)}\longrightarrow\Sigma_{XX}^{(\ell)}=\begin{pmatrix}\mathbb{E}\left(XX^T\right) & 0\\ 0 & I_K\end{pmatrix}$$

we get our result.

Note that an estimator of the variance is obtained by plugging-in the sample analogs of $\Sigma_{XX}^{(\ell)}$, $\gamma_0^{(\ell)}$ and $\gamma_1^{(\ell)}$. Moreover, in absence of censoring, $\gamma_0^{(\ell)}=\delta^{(\ell)}=1$ and it arrives to the classical result given by:

$$\boldsymbol{\sigma}_{ij}^{(\ell)}=\mathbb{E}\left[\boldsymbol{\varphi}_i^{(\ell)}\left(Z,X\right)\boldsymbol{\varphi}_j^{(\ell)}\left(Z,X\right)\right]$$

where $\boldsymbol{\sigma}_{12}^{(\ell)}=0$ as long as $\mathbb{E}\left(\varepsilon|X\right)=0$ holds.

Proof of Corollary 2

By Proporsition 2, we can write:

$$n^{1/2}\begin{pmatrix}\hat{\beta}_\ell\\ \hat{\mu}_X^{(\ell)}\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\left(\beta_\ell,\mu_X^{(\ell)}\right)^T,\frac{1}{\rho_\ell}\Sigma_{\beta\mu_X}^{(\ell)}\right)$$

where $\frac{n_\ell}{n}\to\rho_\ell$.

Then,

$$n^{1/2}\hat{\beta}_\ell\hat{\mu}_X^{(\ell)}\xrightarrow{d}\mathcal{N}\left(\beta_\ell\mu_X^{(\ell)},\frac{1}{\rho_\ell}V_\ell\right)$$

with

$$V_\ell=\mu_X^{(\ell)^T}\boldsymbol{\sigma}_\beta^{(\ell)}\mu_X^{(\ell)}+\beta_\ell^T\boldsymbol{\sigma}_{\mu_X}^{(\ell)}\beta_\ell+2\beta_\ell^T\boldsymbol{\sigma}_{\beta\mu_X}^{(\ell)}\mu_X^{(\ell)}$$

Denote $\rho_0=\rho$. Thus, for the total mean difference, it follows that:

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

and

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

Analogously, to compute the asymptotic distribution of $\hat{\Delta}_S=\left(\hat{\beta}_1-\hat{\beta}_0\right)^T\hat{\mu}_X^{(1)}$ we know that:

$$n^{1/2}\begin{pmatrix}\hat{\beta}_1-\hat{\beta}_0\\ \hat{\mu}_X^{(1)}\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(\begin{pmatrix}\beta_1-\beta_0\\ \mu_X^{(1)}\end{pmatrix},\begin{pmatrix}\frac{\boldsymbol{\sigma}_\beta}{\rho(1-\rho)} & \cdot\\ \frac{\boldsymbol{\sigma}_{\beta\mu_X}^{(1)}}{(1-\rho)} & \frac{\boldsymbol{\sigma}_{\mu_X}^{(1)}}{(1-\rho)}\end{pmatrix}\right)$$

where $\boldsymbol{\sigma}_\beta=\rho\boldsymbol{\sigma}_\beta^{(1)}+(1-\rho)\,\boldsymbol{\sigma}_\beta^{(0)}$.

Now, define $\Delta_\beta=\beta_1-\beta_0$. For the structure effect we have:

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

and

$$V_{\Delta_S}=\frac{1}{1-\rho}\Delta_\beta^T\boldsymbol{\sigma}_{\mu_X}^{(1)}\Delta_\beta+\frac{2}{1-\rho}\Delta_\beta^T\boldsymbol{\sigma}_{\beta\mu_X}^{(1)}\mu_X^{(1)}+\frac{1}{\rho\,(1-\rho)}\mu_X^{(1)^T}\boldsymbol{\sigma}_\beta\mu_X^{(1)}.$$

In a similar way, for the composition effect we get:

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

and

$$V_{\Delta_C} = \frac{1}{\rho}\Delta_{\mu_X}^T \boldsymbol{\sigma}_\beta^{(0)} \Delta_{\mu_X} + \frac{2}{\rho}\beta_0^T \boldsymbol{\sigma}_{\beta\mu_X}^{(0)} \Delta_{\mu_X} + \frac{1}{\rho\left(1-\rho\right)}\beta_0^T \boldsymbol{\sigma}_{\mu_X} \beta_0$$

with $\Delta_{\mu_X} = \mu_X^{(1)} - \mu_X^{(0)}$ and $\boldsymbol{\sigma}_{\mu_X} = \rho\boldsymbol{\sigma}_{\mu_X}^{(1)} + \left(1-\rho\right)\boldsymbol{\sigma}_{\mu_X}^{(0)}$.

Additionally, t-statistics can be constructed by plugging-in the empirical analogs of the corresponding variances. For instance:

$$t_{\Delta_Y} = \frac{\hat{\Delta}_Y}{\sqrt{\hat{\mathbb{V}}\Delta_Y}} \xrightarrow{d} \mathcal{N}\left(0,1\right)$$

with $\hat{\mathbb{V}}\Delta_Y = \frac{1}{n}\hat{V}_{\Delta_Y} = \frac{1}{n_0}\hat{V}_0 + \frac{1}{n_1}\hat{V}_1$.

Validity of the Counterfactual Operator based on Cox model

The validity of the estimation and inference procedure of the CCOX follows the arguments in Chernozhukov *et al.* (2013, CFM, hereafter). Consider the following regularity condition:

Condition 2 *Let $\mathcal{F}$ be a class of bounded measurable functions under the metric $\phi^{(\ell)}$ defined as:*

$$\phi^{(\ell)} = \left[\int \left(f - \tilde{f}\right)^2 dF^{(\ell)}\left(x\right)\right]^2$$

The following regularities hold:

a. Define the empirical processes:

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

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

with $f \in \mathcal{F}$. Then:

$$\left(\hat{M}^{(\ell)}(y,x), \hat{N}^{(\ell)}(f)\right) \Longrightarrow \left(M^{(\ell)}(y,x), N^{(\ell)}(f)\right)$$

where $\left(M^{(\ell)}(y,x), N^{(\ell)}(f)\right)$ is a zero mean tight Gaussian process, $M^{(\ell)}$ has uniformly continuous paths with respect to a standard metric on $\mathbb{R}^{1+k}$ and $N^{(\ell)}$ has uniformly continuous paths with respect to the metric $\phi^{(\ell)}$ on $\mathcal{F}$.

b. The map $y \mapsto F^{(\ell)}(y|.)$ is uniformly continuous with respect to the metric $\phi^{(\ell)}$.

To establish validity of the estimation and inference procedure based on bootstrapping methods, it is needed to verified the fulfillment of two high-level requirements, namely: *i.* the estimator of both conditional distribution and covariates distribution converge at parametric rate and satisfy a functional central limit theorem; and *ii.* bootstrapping methods are valid for estimating the limit laws of the conditional and the covariates distributions.

Under requirement *i.*, the counterfactual operator satisfies a functional central limit theorem, while requirements *i.* and *ii.* jointly guarantee that bootstrapping techniques are valid for making inference of the counterfactual operator and its smooth related functionals. The latter result is pertinent since the limit process of the counterfactual operator is nonpivotal.

Condition 2 is verified by Tsiatis (1981) and Andersen and Gill (1982). In particular, Tsiatis (1981) shows that $n^{1/2}\left(\hat{\beta}_\ell - \beta_\ell\right)$ converges in distribution to a normal random variable with zero mean, while the random function $n^{1/2}\left(\hat{\Lambda}_0^{(\ell)}(y) - \Lambda_0^{(\ell)}(y)\right)$ converges weakly to a Gaussian process (Theorems 3.2 and 6.1, respectively). These asymptotic results have also been documented by Naes (1982); Bailey (1983, 1984); Gill (1984). Following Tsiatis (1981, Lemma 6.2), we have that:

$$n_\ell^{1/2}\left(\hat{\Lambda}_0^{(\ell)}(y)\exp\left(\hat{\beta}_\ell^T x\right) - \Lambda_0^{(\ell)}(y)\exp\left(\beta_\ell^T x\right)\right) \Longrightarrow \mathcal{V}_x^{(\ell)}(y)$$

$$n_\ell^{1/2}\left(1 - \exp - \left(\hat{\Lambda}_0^{(\ell)}(y)\exp\left(\hat{\beta}_\ell^T x\right)\right) - F^{(\ell)}(y|x)\right) \Longrightarrow \mathcal{S}_x^{(\ell)}(y)$$

where $\mathcal{S}_x(y)$ is a Gaussian process with zero mean and covariance structure given by:

$$\mathbb{C}ov\left(\mathcal{S}_x^{(\ell)}(y), \mathcal{S}_x^{(\ell)}(z)\right) = F^{(\ell)}(y|x)\, F^{(\ell)}(z|x)\, \mathbb{C}ov\left(\mathcal{V}_x^{(\ell)}(y), \mathcal{V}_x^{(\ell)}(z)\right) \quad 0 \leq y \leq z \leq \tau_H^{(\ell)}$$

Consequently, CCOX estimator satisfies a functional central limit theorem (that follows from CFM, Theorem 4.1). In addition, since $F^{(\ell)}(y|x)$ is Hadamard differentiable with respect to β_ℓ and $\Lambda_0^{(\ell)}(.)$ (see for details Freitag and Munk, 2005; McLain and Ghosh, 2011; Chen *et al.*, 2010; Hirose, 2011), and hence, by the chain rule of Hadamard differentiable maps (van der Vaart and Wellner, 2004, Lemma 3.9.3), the counterfactual operator is Hadamard differentiable respect its arguments. Hence, the related smooth functionals also obey a central limit theorem (see Corollary 4.2 in CFM for details).

With respect to the inferential procedure, Cheng and Huang (2010) justify the validity of exchangeable resampling methods for general semiparametric $M-$estimators, which includes the Cox model as particular case. This verifies the second high-level requirement. As Corollaries 5.3 and 5.4 in CFM state, this shows that bootstrap consistently estimates the limit laws of the counterfactual operator based on the Cox model. Using the aforementioned argument, by Hadamard differentiability, this result holds for their smooth functionals.

7.2 Tables

Table 6 Performance OB Decomposition

Δ_C				
Censoring Level	**Sample Size**	**OB**	**OB - censored**	**COB**
	50	0.141	0.141	0.141
0.0	500	0.045	0.045	0.045
	2500	0.020	0.020	0.020
	50	0.184	0.167	0.168
0.3	500	0.151	0.082	0.056
	2500	0.149	0.075	0.026

Δ_S				
Censoring Level	**Sample Size**	**OB**	**OB - censored**	**COB**
	50	0.201	0.201	0.201
0.0	500	0.064	0.064	0.064
	2500	0.027	0.027	0.027
	50	0.228	0.226	0.241
0.3	500	0.155	0.098	0.080
	2500	0.148	0.074	0.036

Table 7 Comparison between Kaplan-Meier and CCOX Estimators

$Y \perp C$

Censoring Level	Sample Size	Kaplan-Meier			CCOX		
		MD	**AD**	**MSE**	**MD**	**AD**	**MSE**
	50	0.00	0.00	0.00	26.81	9.20	0.39
0	**500**	0.00	0.00	0.00	7.61	2.29	0.11
	2500	0.00	0.00	0.00	2.39	0.62	0.03
	50	23.54	5.54	0.32	34.92	11.75	0.50
0.05	**500**	7.82	1.81	0.09	10.36	3.00	0.13
	2500	2.52	0.53	0.03	3.23	0.83	0.04
	50	94.91	22.99	1.13	97.53	24.43	1.16
0.2	**500**	40.91	7.62	0.39	41.05	7.69	0.39
	2500	17.62	2.69	0.15	17.72	2.69	0.14
	50	226.17	54.30	2.62	226.79	53.66	2.61
0.5	**500**	147.41	24.28	1.34	147.06	23.52	1.32
	2500	101.99	12.53	0.79	101.98	12.21	0.78

$Y \perp C|X$

Censoring Level	Sample Size	Kaplan-Meier			CCOX		
		MD	**AD**	**MSE**	**MD**	**AD**	**MSE**
	50	0.00	0.00	0.00	26.77	9.20	0.39
0	**500**	0.00	0.00	0.00	7.62	2.29	0.11
	2500	0.00	0.00	0.00	2.39	0.62	0.03
	50	24.75	7.10	0.37	32.71	10.69	0.46
0.05	**500**	14.06	4.52	0.20	9.72	2.64	0.12
	2500	11.21	3.80	0.17	3.04	0.72	0.03
	50	80.70	29.71	1.30	54.24	15.98	0.70
0.2	**500**	57.23	21.80	0.92	15.67	4.01	0.18
	2500	51.56	18.81	0.84	4.93	1.14	0.05
	50	212.58	82.98	3.50	114.75	32.68	1.48
0.5	**500**	162.56	65.64	2.78	36.54	8.68	0.39
	2500	150.83	57.97	2.56	12.86	2.72	0.13

MD, AD and MSE are multiplied by 1000 to facilitate comparisons.

Table 8 Performance CCOX Estimator: $Y \perp C$

Weibull Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0	**50**	26.75	22.52	9.18	8.22	0.63	0.60
	500	7.61	7.48	2.30	2.23	0.32	0.32
	2500	2.39	2.38	0.62	0.61	0.17	0.17
0.05	**50**	34.84	31.29	11.71	10.51	0.71	0.68
	500	10.34	10.18	3.00	2.91	0.36	0.36
	2500	3.22	3.21	0.83	0.82	0.20	0.19
0.2	**50**	98.03	93.86	24.46	23.39	1.08	1.05
	500	41.18	40.45	7.69	7.59	0.62	0.62
	2500	17.73	17.46	2.68	2.66	0.38	0.38
0.5	**50**	227.41	225.05	53.73	53.09	1.62	1.61
	500	147.01	146.76	23.51	23.51	1.15	1.15
	2500	101.68	101.59	12.19	12.23	0.89	0.89

Normal Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0	**50**	14.62	7.46	4.99	2.59	0.22	0.13
	500	3.87	4.28	1.40	1.30	0.06	0.06
	2500	3.85	3.91	1.08	1.06	0.05	0.05
0.05	**50**	28.25	23.46	6.97	5.40	0.36	0.29
	500	8.61	8.18	2.01	1.99	0.10	0.09
	2500	4.17	4.20	1.17	1.15	0.05	0.05
0.2	**50**	82.44	78.41	18.64	17.79	0.96	0.90
	500	27.45	26.66	5.07	4.92	0.27	0.26
	2500	9.74	9.50	1.84	1.78	0.10	0.09
0.5	**50**	162.23	157.60	36.83	36.07	1.83	1.78
	500	62.71	61.33	10.87	10.63	0.59	0.57
	2500	23.91	23.18	3.66	3.53	0.21	0.20

MD, AD and MSE are multiplied by 1000 to facilitate comparisons.

Table 9 Performance CCOX Estimator: $Y \perp C|X$

Weibull Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0	**50**	26.79	22.55	9.20	8.22	0.39	0.36
	500	7.61	7.48	2.29	2.22	0.11	0.10
	2500	2.39	2.39	0.62	0.61	0.03	0.03
0.05	**50**	32.76	29.80	10.69	9.55	0.46	0.41
	500	9.70	9.54	2.63	2.54	0.12	0.12
	2500	3.04	3.04	0.72	0.72	0.03	0.03
0.2	**50**	54.32	51.78	16.02	14.85	0.70	0.65
	500	15.68	15.43	4.01	3.92	0.18	0.17
	2500	4.94	4.93	1.14	1.13	0.05	0.05
0.5	**50**	115.32	111.99	32.76	31.71	1.49	1.43
	500	36.56	36.08	8.69	8.61	0.39	0.39
	2500	12.87	12.75	2.72	2.71	0.13	0.13

Normal Times							
Censoring Level	**Sample Size**	**MD**		**AD**		**MSE**	
		B	**KP**	**B**	**KP**	**B**	**KP**
0	**50**	14.61	7.46	4.98	2.59	0.22	0.13
	500	3.88	4.28	1.40	1.31	0.06	0.06
	2500	3.85	3.91	1.08	1.06	0.05	0.05
0.05	**50**	26.56	21.89	6.78	5.11	0.34	0.27
	500	7.72	7.52	1.87	1.88	0.09	0.09
	2500	4.17	4.23	1.09	1.08	0.05	0.05
0.2	**50**	75.22	72.44	17.62	16.96	0.88	0.84
	500	24.07	23.83	4.70	4.65	0.24	0.24
	2500	8.30	8.30	1.65	1.64	0.08	0.08
0.5	**50**	149.78	146.90	35.71	35.27	1.74	1.70
	500	52.93	52.52	9.88	9.82	0.51	0.51
	2500	18.81	18.75	3.20	3.19	0.17	0.17

MD, AD and MSE are multiplied by 1000 to facilitate comparisons.

Table 10 Decomposition Exercise: Mean Lifetime and Quartiles

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

Confidence Level	Censoring Levels		Q(0.25)		Q(0.75)	
	$\Pr\left(\delta^{(0)}=0\right)$	$\Pr\left(\delta^{(1)}=0\right)$	**Percentile**	**Hybrid**	**Percentile**	**Hybrid**
95	**0.0**	**0.0**	0.946	0.928	0.957	0.935
	0.0	**0.3**	0.965	0.945	0.958	0.940
	0.3	**0.0**	0.968	0.942	0.964	0.931
	0.3	**0.3**	0.963	0.933	0.958	0.930
90	**0.0**	**0.0**	0.907	0.882	0.909	0.869
	0.0	**0.3**	0.926	0.897	0.920	0.896
	0.3	**0.0**	0.925	0.897	0.920	0.884
	0.3	**0.3**	0.916	0.886	0.909	0.884

Table 11 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.

Table 12 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.

7.3 Figures

Figure 1 Decomposition Exercise: Simulated Data