Decomposing the Unemployment Duration using Counterfactual Distributions with Random Censoring: The case of Spain 2004-2011.

Andrés García-Suaza*

January 29, 2015

-Preliminary version-

Abstract

This document discusses an estimator for counterfactual distributions when the variable of interest is censored, which is a characteristic of variables that measure duration. Under a semiparametric approach, we discuss the validity of the estimation procedure and its finite sample properties. Finally, a decomposition exercise of the unemployment duration in Spain for the period 2004-2011 suggests that both variations in socioeconomic characteristics and labor market circumstances play important roles, but the latter is more relevant to explain the long term unemployment.

Keywords: Censorship, Duration analysis, Decomposition Methods, Potential outcomes.

JEL Codes: C14, C24, C41.

*PhD student Universidad Carlos III (Madrid, Spain) and assistant professor Universidad del Rosario (Bogota, Colombia), email address: andres.garcia@urosario.edu.co. I thank Prof. Miguel A. Delgado for his guidance and support. I also thank the comments of Pedro Sant'Anna, Luis F. Gamboa and participants at LACEA-LAMES annual meeting 2014 (Sao Paulo). Research is funded by Spanish Plan Nacional de I+D+I grant number ECO2012-33053 .

1 Introduction

The central idea in economic analysis is to quantify whether a particular change in the economic enviroment affects some economic variables of interest. However, in some circumstances this analysis involves a kind of *what if...* questions based on hypothetical situations whose economic consequences are not directly observed in available data. In economics, experimental data are scarce since designing and running social experiments is costly in terms of funding and time, and depending on the objective, might be infeasible. As a consequence, policy analysis literature has been dedicated to estimate *potential* outcomes using information from observational studies. There are many ways of estimating potential outcomes when available data are completely observed, however the case of censored outcomes has received less attention. In this document, we deal with the estimation of counterfactual distributions when the variable of interest is censored, which is a usual feature of micro data related to duration or time-to-event.

From the seminal contributions of Rubin (1974) in causal inference, and Oaxaca (1973) and Blinder (1973) in decomposition analysis, a number of methodological approaches implicitly have considered the identification and estimation of counterfactual outcomes[1] (for a comprehensive discussion of different decomposition methods and policy evaluation issues see Fortin et al., 2011, and, Imbens and Wooldridge, 2009). Intuitively, a conterfactual outcome is the hypothetical (or synthetic) outcome resulting of changes in the economic environment (e.g. the introduction of a economic policy or social program) which is not directly observed. For instance, to contrast the gender wage discrimination hypothesis, we would like to compare men and women wages under equivalent circumstances. For that, we are interested in estimating the potential outcome defined as the prevailing women's wage if this population faces the men's wage schedule.

The first statistical approaches to estimate counterfactual outcomes focused on studying the average effect by following regression analysis framework[2]. To have a broader understanding of the economic policy effects, statistical inference has focused on functionals beyond the mean. Specifically, more general approaches have been concerned

[1]These techniques have been extensively applied in the context of microdata, e.g. in labor economics to study the gender discrimination hypothesis or the causal effect of human capital on labor outcomes. But some macro applications have been also addressed by using quasi-experimental data (see, Card and Krueger, 1993; Acemoglu et al., 2001; Pesaran and Smith, 2012).

[2]For a nonparametric approach see for instance Stock (1989).

about the effect over whole distribution as well as different features of the potential outcomes' distributions (DiNardo et al., 1995; Machado and Mata, 2005; Rothe, 2010; Firpo and Pinto, 2011; Rothe, 2012; Chernozhukov et al., 2013). By modelling the entire distribution allows to compute distributional effects and quantile effects, and makes feasible the estimation of a large class of related functionals such as order statistics or Gini index, and even function-valued statistics as the Lorenz curve.

In observational studies is usual to find that the available data is incomplete. Depending of the context, incompleteness might be due to different sources such that truncation, censoring or sample selection. In general, the implementation of the classical statistical methods might induce to misleading conclusions. Among these types of incomplete data, right censoring is particularly important mainly when the variable of interest represents time to event or duration. Right censoring (referred as *censoring* hereafter) appears in the context of follow-up studies when durations are partially observed for some individuals either because the event of interest has not occurred at the ending point or in the case that they have withdrawn during the study. Many examples of this situation can be found in economics, namely: unemployment spell, firm's lifetime, bank failure, school desertion, among others.

Therefore, our goal is to study a semiparametric estimator of the counterfactual distribution under censoring. To do so, we take into account a convenient representation of the marginal distribution as function of the conditional distribution of the outcome given some covariates (see for instance Rothe, 2010, and, Chernozhukov et al., 2013, CFM hereafter) as well as the Cox hazard model as natural estimator of the latter one. The contribution of this paper is twofold. First, under standard ignorability conditions, the proposed estimator is a flexible procedure to compute treatment effects when the variable of interest denotes duration, as well as to perform decomposition exercises in the spirit of Oaxaca-Blinder[3].

And secondly, we check the high level requirements studied by CFM under which the estimator of the counterfactual distribution with censoring is consistent and the inference through resampling methods is valid. Additionally, we provide evidence about the finite

[3]Some studies have considered the case of censored variables. For instance, using duration models, Powers and Yun (2009) extend the Blinder-Oaxaca decomposition to estimate the difference in the average hazard rate, but they do not focus in the estimation of the whole counterfactual distribution; and instead, due to the non-linearity of the hazard function, they implement Taylor approximations to express the difference in hazard rate as a weighted sum of observable quantities.

sample properties of the counterfactual operator with censoring in two scenarios: in the case that the censoring mechanism is purely random and when censoring is informative.

CFM proposed the Distribution Regression (DR) method to estimate the conditional distribution. This method encompasses the discrete transformation of the Cox model as link function; however, the presence of censoring requires a particular treatment. In specific, in the discrete version of the Cox model, the dichotomous dependent variable represents whether an observation is censored, while in the DR, this variable is the relative location of the observation across distribution function. In this manner, the Cox model is convenient to control the presence of censoring and, in comparison with DR procedure, provides some additional practical advantages. First, the distribution function is easily mapped from the hazard function. Second, the shape is estimated nonparametrically and all properties of the distribution function, including monotonicity, are naturally obtained. And lastly, the estimation procedure based on Cox model accommodates to the asymptotic results recently developed by CFM.

Our results suggest that this estimator covers more general situations than pure random censoring. In particular, it is useful not only to estimate counterfactual outcomes but also to estimate observable outcomes when the independence assumption of duration and censoring times fails. And also, the estimation procedure is well behaved for different levels and patterns of censoring, and for different data generating process. Lastly, considering a two sample problem to decompose differences in the mean lifetime and the quartiles, the results are favorable for inference based on bootstrapping techniques.

By performing a decomposition exercise of the unemployment duration distribution using data from Spain for the period 2004-2011, we obtain that both variations in socioeconomic characteristics and labor market structure play important roles. A detailed exercise to decompose the structure effect between factors associated to common risk of being unemployed and effect of the covariates on the probability of leaving from unemployment suggests that the former is quantitatively more relevant, mainly for explaining the particular transition from unemployment to employment. Moreover, by analyzing separately the transition from unemployment to employment or being out of the labor force, we found that structure effect explain the most of the variation of the target statistics, and factors associated to the common of being unemployment have particular importance to explain the transition unemployment-to-employment.

Besides this introduction, the rest of the document is organized as follows. The second section briefly presents a general framework for decomposition analysis as well as the basic idea of counterfactual distribution estimation. In turn, the third section discusses the identification assumptions needed when data are subject to be censored and briefly presents the Cox model. Some asymptotic results that guarantee validity of the estimation and inference procedures are checked in the fourth section , while the simulation results are presented in the fifth section. Finally, the decomposition exercise using data from Spain for the period 2004-2011 and some final remarks are exposed in the sixth and the seventh sections, respectively.

2 Conterfactual Distribution: A General Framework

2.1 Counterfactual Outcomes and Counterfactual Distributions

In sake of simplicity, to define what a counterfactual outcome is, we can revisit the original idea of Oaxaca-Blinder (OB hereafter) decomposition. Suppose we are interested in comparing the average wage, $\overline{Y_D}$, between two groups denoting by D, where $D \in \{0, 1\}$ (e.g. population might be defined by gender, 1 men and 0 women , or for a particular treatment). By assuming linear relation between the outcome and a set of covariates, and conditional independence between observables and unobservables, the average difference across groups can be written as follows:

$$\widehat{\Delta}\overline{Y} = \overline{X}_1' \left(\widehat{\beta}_1 - \widehat{\beta}_0 \right) + \left(\overline{X}_1 - \overline{X}_0 \right)' \widehat{\beta}_0 \tag{1}$$

where X_D is the set of observable factors of the underlying population including human capital variables and socioeconomic characteristics, $\widehat{\beta_D}$'s are the human capital returns estimated through a Mincer equation for each group. The first term in Equation (1) is associated to the difference in the human capital returns, while the second refers to the difference in human capital accumulation. In general, these components are known as *structure effect* and *composition effect*, respectively.

Despite its simplicity, OB decomposition shows the role of counterfactual outcomes and reveals some identification assumptions that rest behind. In particular, $\overline{X}_1' \widehat{\beta}_0$ represents the counterfactual average wage; that is, the prevailing average wage of propulation 1 if it was paid like population 0[4]. Intuitively, this term is obtained by applying to population 1 the structural form of the outcome for population 0. Such interpretation is only feasible if the structural form does not vary respect to the covariates, i.e., under the invariability of the conditional mean to variation in factors distribution and separability between observables and unobservables. A deep debate on the identifiability assumptions of the counterfactual outcome in a general framework is addressed by Fortin et al. (2011), which is summarized in Assumption 1.

Assumption 1. *a. Populations 0 and 1 are mutually exclusive groups.*

b. The observable outcomes $Y_0 = m_0(X_0, \varepsilon_0)$ and $Y_1 = m_1(X_1, \varepsilon_1)$ are properly defined in terms of observables X and unobservables ε.

c. Overlapping support: if $\mathcal{X} \times \mathcal{E}$ denotes the support of observables and unobservable characteristics of the underlying population, then $(X_0, \varepsilon_0) \cup (X_1, \varepsilon_1) \in \mathcal{X} \times \mathcal{E}$.

d. Simple counterfactual treatment.

e. Conditional independe of treatment and unobservables: $D \perp \varepsilon | X$.

f. Invariance of conditional distribution

In short, a. allows to interpret Y_0 and Y_1 as potential outcomes; that is, by keeping X and ε unaltered, an individual will get Y_0 if $D = 0$ and Y_1 if $D = 1$, but only one is observed. Part b. guarantees that $m_D(.)$ is a correctly specified structural form as function of observables and unobservables. In turn, c. implies that it is possible to identify (match) each profile defined by the set of covariates in both population with positive probability, which ensure the comparability across groups[5]. Meanwhile, d. rules out the possibility of getting a different outcome schedules rather other $m_0(.)$ and $m_1(.)$, and if e. holds, the effects of observable and unobservable are separable (unconfounded), .

Finally, f. enables to keep constant all the other components when covariates distribution varies, which validates the estimation of the counterfactual outcome by extrapolating

[4]Similarly, it can be also interpreted as the average wage that individuals population 0 would have if they have the characteristics of population 1.

[5]There are alternative methodologies to decompose the wage gap to the case where the populations do not have common support. See Nopo (2008).

the distribution of the covariates. Moreover, part e. also allows to establish a natural link between policy evaluation literature and decomposition analysis. Under this assumption, also known as ignorability condition, the distribution of unobservable is not informative about treatment assignment, implying that conditional distribution of unobservables is the same across groups. Therefore, as it is pointed out in Fortin et al. (2011, Section 2) and CFM (lemma 2), the second component in Equation (1) can be interpreted as causal effect.

There are many ways of making inference on counterfactual outcomes and, in the most of the cases, literature has focused on the average effect (RIF,QUANTILE TREATMENT). In that sense, instead of focusing on particular feature of the counterfactual outcome such as the mean or the variance, a appealing approach proposes to estimate the counterfactual distribution (Rothe, 2010, and Chernozhukov et al., 2013). According to this approach, the distribution of potential outcomes is estimated as the average of a conditional distribution over the probability measure of the covariates set. Lets denote the potential outcomes as $Y\langle i,j\rangle$ and $F_{Y\langle i,j\rangle}$ its corresponding distribution function, where j refers to the set of the characteristics and i is the schedule (payment structure) that such population faces. So, if $i=j$, the outcome is directly observed in the data, for instance, according to the OB decomposition, $Y\langle 0,0\rangle$ represents the wage of women.

In such context, we consider two information sets denoted by Ω_D which are composed of n_D replicates of the vector (Y_D, X_D), where X is a p-dimentional vector of economic factors or covariates. A convenient form to write the marginal distribution of the outcome for $D=i$, is given by:

$$F_{Y\langle i,i\rangle}(y) = \mathbb{E}\left[F_{Y_i|X_i}(y|X_i)\right] = \int_{\mathcal{X}_i} F_{Y_i|X_i}(y|x)\, dF_{X_i}(x) \tag{2}$$

that means that the distribution function can be obtained by integrating the conditional distribution from Ω_i over the support $\mathcal{X}_i \subseteq \mathbb{R}^p$.

Equation (2) provides an intuitive way for calculating the distribution of a counterfactual outcome. Since $F_{Y_i|X_i}$ contains all the information related with the structure or relationship between outcome and factors, it can be integrated over the factors distribution of the population j in order to obtain the potential outcome of population i if the economic environment defined by X_i, varies to X_j. Therefore, $F_{Y\langle i,j\rangle}$, is constructed by integrating the conditional distribution of the outcome for the information set Ω_i with respect to the distribution of the covariates under the information set Ω_j, which is given by:

$$F_{Y\langle i,j\rangle}(y) = \int_{\mathcal{X}_j} F_{Y_i|X_i}(y|x)\, dF_{X_j}(x) \tag{3}$$

The *counterfactual operator*, given by Equation (3), is well defined as long as $\mathcal{X}_1 \subseteq \mathcal{X}_0$ (Assumption 1. c.). Moreover, this representation accommodates to the presence of covariates of different type including binary factors and facilitates to consider the case when censoring is present since all of the censoring mechanism is control trough modelling $F_{Y_i|X_i}$.

2.2 General Decomposition

In economic analysis, one might be interested in making comparisons between observed outcomes and counterfactual outcomes. In that sense, setting $i = 0$ and $j = 1$, the *distributional effect* Δ can be decomposed as follows:

$$\begin{aligned} \Delta(y) &= F_{Y\langle 1,1\rangle}(y) - F_{Y\langle 0,1\rangle}(y) + F_{Y\langle 0,1\rangle}(y) - F_{Y\langle 0,0\rangle}(y) \\ &= \Delta_S(y) + \Delta_X(y) \end{aligned} \tag{4}$$

where Δ_X is the composition effect while Δ_S is the structure effect at a fixed y. Rather than focusing on the average effect, one might be curious about other features of the distribution such as quantiles, higher moments or more general functionals of the data. Once observed and counterfactual distributions are estimated, the target parameters of the potential outcome $Y\langle i,j\rangle$ can be estimated by empirical integrals or simply inverting the distribution function for the case of quantiles.

To do so, we consider the following general statistics:

$$Q_Y(t) = \inf\{y \in \mathbb{R} : F_Y(y) \geq t\}, \qquad \theta(F_Y) = \int \varphi(y)\, F(dy) \tag{5}$$

where Q_Y is the quantile function, θ is a large class of functionals covering higher moments, variance, among others; and, φ is an integrable function of the outcome. So, *quantile effect* and *θ-effect* can also be defined according to Equation (4). For instance for the case of the quantile, we can write:

$$\begin{aligned}\Delta^{Q}(t) &= Q_{Y\langle 1,1\rangle}(t) - Q_{Y\langle 0,1\rangle}(t) + Q_{Y\langle 0,1\rangle}(t) - Q_{Y\langle 0,0\rangle}(t) \\ &= \Delta^{Q}_{S}(t) + \Delta^{Q}_{X}(t)\end{aligned} \tag{6}$$

2.3 Estimation of Counterfactual Distributions

The crucial issue to compute these decompositions is how to estimate $F_{Y\langle 0,1\rangle}$. According to Equation (3), by plug-in principle, estimation procedure requires proper estimators of the covariates and conditional distributions. With respect to the covariates distribution, the natural estimator of the covariates distribution is the multivariate empirical distribution. For instance, for X_1 we have:

$$\widehat{F}_{X_1}(x) = n_1^{-1}\sum_{i=1}^{n_1} 1_{\{X_1 \leq x\}} \tag{7}$$

To deal with censored outcomes, the key point is the estimation of the conditional distribution. To estimate $F_{Y|X}$, CFM propose two methods: quantile regression and distribution regression. The former refers to the classical approach by Koenker and Basset (1978) where the conditional quantile is specified as a linear combination of the covariates. While Distribution Regression (DR) is a generalization of Foresi and Peracchi (1995) who model the conditional distribution through a serie of binary choice models defined by the location of a set of cut offs over the support of the outcome. Specifically, the conditional distribution at a fixed y can be written as:

$$F_{Y|X}(y|x) = \Psi\left(P(X)'\beta(y)\right) \tag{8}$$

where P is a vector of transformations of X, including non-linear relation between covariates, and $\Psi(.)$ is a link function such logit, probit or cloglog[6]. We adopt a similar approach by modelling the conditional distribution, instead of the quantiles, through the hazard function, as it will be presented in detail later.

[6] By comparing these approaches, CFM point out that distribution regression presents some advantages respect to quantile regression since it does not involve either inverting function, fine mesh approximation and/or trimming around the tails of the distribution, and also, it does not require smoothness (more details about asymptotic properties are obtained by Koenker et al., 2013).

So, the estimator of the counterfactual distribution is found by plugging-in the empirical counterparts of the covariates distribution and the conditional distribution in Equation (3):

$$\widehat{F}_{Y\langle 0,1\rangle}(y) = \int_{\mathcal{X}_1} \widehat{F}_{Y_0|X_0}(y|x)\, d\widehat{F}_{X_1}(x) = \frac{1}{n_1}\sum_{i=1}^{n_1}\widehat{F}_{Y_0|X_0}(y|x_i) \tag{9}$$

Once the counterfactual distribution is estimated, the same logic applies for *distribution effect*, *quantile effect* and *θ-effect.* CFM develop asymptotics properties of the counterfactual operator given by (9) and provide the conditions under which bootstrap methods are valid to perform statistical inference. Their asymptotic results require to estimate the conditional and covariate distribution at parametric rates and that such estimators satisfy a central limit theorem. Regarding the inferential procedure, it is needed the validity of the bootstrap procedure for estimating limit laws for both conditional distribution and covariate distribution.

Under certain considerations on the estimation of the conditional distribution and additional identification assumptions, the counterfactual operator can be naturally extended to the case when the variable of interest is subject to be censored. Essentially, quantile regression and distribution regression do not self-adjust to this feature of the data, but Cox hazard model turns out in a flexible alternative since censoring is automatically set and the hazard function is an identifier of the distribution function.

Additionally, the derived conditional distribution satisfies all properties of a distribution function and does not need any monotonicity arrangement. And, as it will be discuss later in the fourth section, by applying well known asymptotic results of Cox model, it is possible to extrapolate the asymptotic properties and inferential procedure of $\widehat{F}_{Y\langle i,j\rangle}$ to the case where censoring is present.

3 Notation and Set Up

3.1 Inference based on Censored Variables

In duration (or survival) analysis, the variable of interest represents the number of periods until an event occurs, e.g., the number of months (or quarters) that an individual is unemployed until get a new job or the number of periods that a firm stays operating before bankruptcy. This variable, denoted by Y, is a non-negative variable described by the distribution function F_Y and density function given by f_Y. Time-to-event is a dynamic variable that requires to collect information by following the units of analysis by a fixed period of time.

Due to the sampling scheme or large costs of conducting follow-up studies for long period of time, it is common that the time to event for some individuals is not available at the ending point. For instance, in the case unemployment duration, either some workers might be unemployed in the last follow-up period or they could withdraw during the study period. As consequence, complete unemployment spells are not directly observable. This feature is commonly known as right censoring.

In that context, inference is usually focused on the hazard function which is defined as the instantaneous probability of transition at y given that the duration in the current state is at least y, that is:

$$h_Y(y) = \frac{f_Y(y)}{S_Y(y)} = -\frac{\partial \ln S_Y(y)}{\partial y} \tag{10}$$

where $S_Y(y)$ denotes the survival distribution, that is $S_Y(y) = 1 - F_Y(y)$. The key point of the hazard function lies in that it allows to split the role of censored and uncensored observations such that censored observations only contribute to the survival function.

It is noticeable that survival analysis, in order to facilitate the interpretation of the results, inference is based on the survival function instead distribution function. Other function of interest that clarifies the link between the hazard function and survival distribution is through the cumulative hazard function, which is:

$$\Lambda_Y(y) = \int_0^y h_Y(s)\,ds = \int_0^y \frac{F_Y(ds)}{1 - F_Y(s-)} \tag{11}$$

where $F(s-) = \lim_{y \uparrow s} F(y)$.

To compute all of the aforementioned functions, information about the actual survival times is required. However, Y is not completely observed. In particular, what we observe is a set of *iid* replications from the random vector (T, δ), where $T = Y \wedge C$ are the observed times which follow the distribution function H, C represents the censoring times with distribution function G, and $\delta = 1_{\{Y \leq C\}}$ is an indicator function taking value 1 if Y is observed[7]. Hence, in order to estimate these quantities additional identifiability assumptions are needed (Tsiatis, 1975).

For instance, in the univariate analysis it is assume that the censoring mechanism is purely random, i.e. Y and C are independent[8]. In that case, the cumulative hazard function can be identified from the sub-distributions $H(y) = \Pr(T \leq y)$ and $H^1(y) = \Pr(T \leq y, \delta = 1)$. Specifically, if F and G do not have common jumps (see Peterson, 1977, for details) Equation (11) is equivalent to:

$$\Lambda_Y(y) = \int_0^y \frac{H^1(ds)}{1 - H(s-)} \tag{12}$$

A natural estimator of the cumulative hazard function is given by plugging-in the empirical analogs of H and H^1. This estimator, widely known in the literature as Nelson-Aalen estimator (Nelson, 1969; Aalen, 1978). Therefore, the survival function is obtained following Equations (10) and (11), $S_Y = exp(-\Lambda_Y)$. This estimator of the survival function is equivalent to the product-limit representation (see Gill, 1980, for details) provided by Kaplan and Meier (1958), which is given by:

$$S_Y^{KM}(y) \quad = \prod_{T_{i:n} \leq y} \left(\frac{n-i}{n-i+1} \right)^{\delta_{[i:n]}}, \qquad y < T_{n:n} \tag{13}$$

where the subindex $i:n$ denotes the *i-th* order statistic, and $[i:n]$ is the concomitant associated to $T_{i:n}$[9]. This estimator assigns zero mass whenever an observation is censored, $\delta = 0$, and self-adjusts the missing mass to the right. In absence of censorship, it is equivalent to the estimator of the empirical distribution where each observation has mass $1/n$.

As it is recognized in survival analysis literature, one difficulty of dealing with censored data is the identification of the upper tail because it is not possible to make inference

[7] "$\wedge$" represents the minimum operator.

[8] That imples $1 - H = (1 - F)(1 - G)$.

[9] Kaplan and Meier (1958) show that S_Y^{KM} is the nonparametric maximum likelihood estimator under random censoring. In turn, Efron (1967) makes a pedagogical explanation of the Kaplan-Meier estimator which is denominated redistribute-to-the-right estimator. According to this explanation, it is assigned a mass of $1/n$ to all observations. By following each order statistics until getting the first censored, a mass of $1/n$ is missed and re-assigned uniformly to the remaining individuals. And the process is repeated with the subsequent censored observations based on the updated weights.

on Y beyond the least upper bound τ_H of the support of H [10]. If $\tau_H = \tau_F < \tau_G$ consistency holds, otherwise relevant information about F on $(\tau_G, \tau_F]$ will be cut off (Stute and Wang, 1993). Therefore, in the case when $\tau_H = \tau_G < \tau_F$, the inference is restricted to $\left(0, \widetilde{T}\right]$, $\widetilde{T} \leq T_{n:n} < \tau_H$. One intuitive solution is to $\delta_{[n:n]} = 1$ (Efron, 1967). In practice, this correction does not affect the other properties of the distribution function estimator such as being nonnegative, nonincreasing and right continuous, and reduces the downward bias inherent to this kind of estimators (see Gill, 1980, and Mauro, 1985). Finally, another feature of duration variables is the prevalence of ties in the reported survival times. Both Nelson-Aalen and Kaplan-Meier estimators are conventionally adjust to the presence ties by ordering such that uncensored observations precede the censored observations, and ties within survival times or within censoring times are ordered arbitrarily.

3.2 Identification Assumptions and Survival Counterfactual Operator

Our populations of interest are described by the vector (Y_D, X_D, C_D). But, we observe (T_D, X_D, δ_D), for each $D \in \{0, 1\}$. By using the available information, the goal is to estimate $F_{Y\langle i,j \rangle}$. Because $Y\langle 0, 0\rangle$ and $Y\langle 1, 1\rangle$ are directly observed, one might be tempted to estimate their distribution function through the empirical analog which is given by the Kaplan-Meier estimator in the case of censored data. However, the presence of covariates requires to describe the relation between X and C. By keeping the independence assumption between Y and C, it is possible to admit relation between economic factors and censoring mechanism. To do so, we adopt the identification assumption proposed by Stute (1993).

Assumption 2.

a. Y_D and C_D are independent, and F_D and G_D have not common jumps.

b. $\Pr(\delta_D = 1 | Y_D, X_D) = \Pr(\delta_D = 1 | Y_D)$.

Assumption 2. a. is the classical independence assumption. In turn, 2. b. describes the relation of the set of covariates and censoring mechanism. It implies that all relation between covariates and censoring times is through the survival times; that is, given the

[10] $\tau_H = \inf\{y : H(y) = 1\} \leq \infty$. Similar definitions can be used with respect to F and G.

actual survival times, there is not information in X affecting C[11]. In this manner, Kaplan-Meier estimator is valid for $F_{Y\langle 0,0\rangle}$, $F_{Y\langle 1,1\rangle}$.

But, this approach restricts other channel whereby the covariates affect the censoring mechanism. Particularly, covariates might contain information about the probability of withdrew that is not in the actual survival times. For instance, the censoring produced by reaching to the ending point of the follow-up period might be considered independent of the occurrence of the event of interest, but it must not be the case if the withdrawal is consequence of outside options; e.g. in unemployment duration studies, the probability of going out from the workforce to inactivity or migrate can be influenced by individuals characteristics. In such case, a feasible assumption, introduced by Beran (1981)[12], is given by:

Assumption 3. $Y_D \perp C_D \mid X_D$

Under Assumption 3, inference based on Kaplan-Meier estimator is not valid; however, the conditional hazard function is identified. Once the conditional survival distribution is mapped, marginal survival distribution can be estimated as:

$$\widehat{S}_{Y\langle i,i\rangle}(y) = \mathbb{E}\left[\widehat{S}_{Y_i|X_i}(y|X_i)\right] = \frac{1}{n_i}\sum_{j=1}^{n_i}\widehat{S}_{Y_i|X_i}(y|x_j) \tag{14}$$

This is nothing but the empirical analog of Equation (2) for the survival distribution. The intuition behind is that the effects of informative censoring is adjusted by estimating the conditional survival for individuals with similar characteristics since they are facing the same censoring risk[13]. Hsu and Taylor (2010) study a similar estimator finding out that it provides efficiency gains respect to alternative methods and that is robustness to misspecification with respect to both distributional forms and ommitted covariates. Morover, Malani (1995) argues that even if the censoring mechanism is random, efficiency gains are derived of using covariates.

In this manner, under Assumption 3, the survival counterfactual operator allows to

[11] Stute (1999); Uña álvarez (2004); Sanchez-Sellero et al. (2005). Assumption 2 is also satisfied when C_D is independent of the vector (Y_D, X_D).

[12] It has been widely implemented in the discussion of numerous estimators of marginal and joint distributions based on (T, δ, X). See for instance Dabrowska (1987, 1989); Gonzalez-Manteiga and Cadarso-Suarez (1994); Akritas (1994); Leconte et al. (2002); Lopez (2011).

[13] Another class of estimators for the survival distribution base on the product-limit representation using covariates to recover the lost of information can be found in Satten et al. (2001).

estimate both observed and counterfactual outcomes in the case of informative censoring. Therefore, in general the counterfactual operator respect to the survival distribution is given by:

$$\widehat{S}_{Y\langle i,j\rangle}\left(y\right)=\int_{\mathcal{X}_j}\widehat{S}_{Y_i|X_i}\left(y|x\right)d\widehat{F}_{X_j}\left(x\right)=\frac{1}{n_j}\sum_{l=1}^{n_j}\widehat{S}_{Y_i|X_i}\left(y|x_l\right) \tag{15}$$

3.3 Conditional Survival Distribution

The key ingredient of the counterfactual operator is how to estimate the conditional distribution. For this, we consider the semiparamtric model by Cox (1975), which has been widely used in economic analysis and requires minimal distribution assumptions[14]. In particular, we implement the following specification for the conditional hazard function:

$$h_{Y|X}\left(y|x,\beta\right)=h\left(x,\beta\right)=h^0\left(y\right)\phi\left(x,\beta\right) \tag{16}$$

where h^0 is the baseline hazard that depends only on y and ϕ is a positive function representing the effect of the covariates on conditional hazard function. It is commonly specified as $\phi\left(x,\beta\right)=e^{x'\beta}$.

Cox (1975) propose to estimate $h_{Y|X}$ using the *partial likelihood* method, which does not require to specify h^0. To do so, denoting $r\left(y\right)$ the pool of individuals who are at risk of failing at period y. The contribution of each observation to the likelihood function, or the same the probability that an individual change of state at y given that it is in the set $r\left(y\right)$, will not depend on the nuisance parameter, that is:

$$\Pr\left(T_i=y|r\left(y\right),x\right)=\frac{h_i\left(x,\beta\right)}{\sum_{j\in r(y)}h_j\left(x,\beta\right)}=\frac{e^{x_i'\beta}}{\sum_{j\in r(y)}e^{x_j'\beta}}$$

It is noticeable that here the censoring is properly adapted since censored observations only contribute to the risk pool (the denominator of the conditional probability). Once β is estimated, the survival distribution can be computed as follows:

$$\widehat{S}_{Y|X}\left(y|x,\widehat{\beta}\right)\;=\;\widehat{S^0}\left(y\right)^{\exp\left(x'\widehat{\beta}\right)}\qquad\widehat{S^0}\left(y\right)=\exp\left(-\widehat{\Lambda^0}\left(y\right)\right) \tag{17}$$

[14]Despite the tractability of parametric models, they are at risk of distorting the information in the data by forcing inappropriate functional forms, and also, carrying out inconsistent estimates whenever some of these components are misspecified.

where S^0 is the baseline hazard.

Lastly, to estimate $\Lambda^0(y)$, there are two popular estimator in the literature. The first and the most commonly used, $\widehat{\Lambda}_B^0$, was proposed by Breslow (1974). While the second estimator, $\widehat{\Lambda}_{KP}^0$ proposed by Kalbfleisch and Prentice (1973), inspired in the cumulative hazard of discrete times, is estimated as the sum of the empirical hazard probability that satisfies the first order condition (or score) of the partial likelihood function. Such estimators are given by:

$$\widehat{\Lambda_B^0}(y) = \sum_{i=1}^{y} \frac{1}{\sum_{j \in r(y_i)} e^{x_j' \beta}} \qquad \widehat{\Lambda_{KP}^0}(y) = \sum_{i=1}^{n} (1 - \widehat{\alpha}_i) \, 1_{\{y_i \leq y\}} \tag{18}$$

where the hazard probabilities, $\widehat{\alpha}_i$, solve:

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

and d_i is the set of individuals changing state at period y_i.

Both $\widehat{\Lambda_B^0}$ and $\widehat{\Lambda_{KP}^0}$ self-adjust to the presence of ties in the failure time. And also, as result of their implementation, the conditional survival distribution given by Equation (17) is consistent and asymptotically normal (for details see Tsiatis, 1975; Andersen and Gill, 1982; Naes, 1982; Bailey, 1983, 1984, and Gill, 1984).

4 Estimation and Inference

4.1 Validity of the Counterfactual Operator based on Cox model

In order to establish validity of the estimation and inference procedure of the counterfactual operator given by Equation (15), we verified the fulfillment of the two high-level requirements studies in CFM, 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. To do so, we invoke well known results of he Cox estimator for the conditional survival distribution. As consequence, under requirement *i.,* the counterfactual operator satisfies a functional central limit theorem, while requirements *i.* and *ii.* guarantee that bootstrap 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.

In addition to the high-level requirements, the following regularity conditions are needed:

Condition 1. Condition c. in Assumption 1 holds, the sample size n_j is nondecreasing in n, and ${}^{n}/_{n_j} \to s_j$ as $n \to \infty$.

Condition 2. Let $\mathcal{F}$ be a class of bounded measurable functions under the metric λ_j defined as:

$$\lambda_j = \left[\int \left(f - \tilde{f} \right)^2 dF_{X_j} \right]^{1/2}$$

The following regularities hold:

- Define the empirical processes:

  $$\hat{Z}_j\left(y, x\right) = \sqrt{n_j} \left(\hat{F}_{Y_j|X_j}\left(y|x\right) - F_{Y_j|X_j}\left(y|x\right) \right) \text{ and } \hat{G}_j\left(f\right) = \sqrt{n_j} \int f d\left(\hat{F}_{X_j} - F_{X_j} \right)$$

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

  $$\left(\hat{Z}_j\left(y, x\right), \hat{G}_j\left(f\right) \right) \Rightarrow \left(Z_j\left(y, x\right), G_j\left(f\right) \right)$$

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

- The map $y \mapsto F_{Y_j|X_j}\left(y|.\right)$ is uniformly continuous with respect to the metric λ_j.

For the case of the Cox model, Condition 2 is verified following result from Tsiatis (1981) and Andersen and Gill (1982).

As discussed, under Assumption 3, $h_{Y|X}$ and $S_{Y|X}$ are identified (see Cox, 1975, for details). Consequently, by applying the survival counterfactual operator, $S_{Y\langle i,i \rangle}$ can be identified. Hence, jointly with Assumption 1 the counterfactual survival distribution $S_{Y\langle i,j \rangle}$ is also identified. The crucial point to establish the validity of the proposed procedure lies in the properties of the conditional distribution estimator. In similar way to

DR, the conditional distribution based on Cox model depends on the set of parameters β and on the functional parameter $\Lambda^0(y)$ which leads the shape of the distribution.

In this regard, Tsiatis (1981) shows that β and $\Lambda^0(.)$ are consistently estimated at parametric rate and satisfy a central limit theorem. In particular, $\sqrt{n}\left(\widehat{\beta}-\beta\right)$ converges in distribution to a normal random variable with zero mean, while the random function $\sqrt{n}\left(\widehat{\Lambda^0}(y)-\Lambda^0(y)\right)$ converges weakly to a Gaussian process[15] (Theorems 3.2 and 6.1, respectively). As result, in Lemma 6.2, Tsiatis (1981) states that:

$$\sqrt{n}\left(\widehat{\Lambda^0}(y)\exp\left(x'\widehat{\beta}\right)-\Lambda^0(y)\exp\left(x'\beta\right)\right)\Rightarrow\mathcal{V}_x(y)$$

$$\sqrt{n}\left\{\exp-\left(\widehat{\Lambda^0}(y)\exp\left(x'\widehat{\beta}\right)\right)-S_{Y|X}(y|x)\right\}\Rightarrow\mathcal{S}_x(y)$$

where $\mathcal{S}_x(y)$ is a Gaussian process with zero mean and covariance structure given by$S_{Y\langle i,i\rangle}$

$$\mathrm{Cov}\left(\mathcal{S}_x(y),\mathcal{S}_x(z)\right)=S_{Y|X}(y|x)\,S_{Y|X}(z|x)\,\mathrm{Cov}\left(\mathcal{V}_x(y),\mathcal{V}_x(z)\right),\quad 0\leq y\leq z\leq\tau_H$$

These results guarantee the achievement of requirement *i*. That implies that the survival counterfactual operator based on the Cox estimator (ACox, hereafter) satisfies a functional central limit theorem (that follows from CFM -Theorem 4.1-). In addition, since $S_{Y|X}$ given by Equation (17) is Hadamard differentiable with respect to β and $\Lambda^0(.)$ (see for details Freitag and Munk, 2005; McLain and Ghosh, 2009; Chen et al., 2010, and Hirose, 2011), by the chain rule of Hadamard differentiable maps (der Vaart and Wellner, 2004, Lemma 3.9.3), the counterfactual operator if 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[16], which includes the Cox model as particular case (BE SPECIFIC). This verifies the second high-level requirement. As Corollaries 5.3 and 5.4 in CFM, this shows that bootstrap consistently estimates the limit laws of the counterfactual operator for distributions of observable outcomes and counterfactual outcomes. Using the aforementioned argument,

[15] Analogous results using counting processes are discussed by Andersen and Gill (1982). Bailey (1983) shows that, for a fixed β, $\widehat{\Lambda^0_B}(y)$ is asymptotically equivalent to the maximum likelihood estimator of the cumulative hazard function.

[16] The result holds even when the nuisance parameter is not estimated at $\sqrt{n}$.

by Hadamard differentiability, this result holds for their smooth functionals.

4.2 Bootstraping Methods with Censored Data

Bootstrapping methods turn out in a practical alternative to perform inference in very general settings. In addition to the complicated forms of the variance for the real-valued and specially function-valued parameters (such as the survival distribution), the functionals based on counterfactual distributions have non-pivotal limit processes. To implement bootstrap methods, two ingredients have to be taken into account: the method of drawing the bootstrap sample and the method of forming the confidence intervals.

In the context of censored data, Efron (1981) presents two alternative resampling schemes that have been recognised in the literature as *simple bootstrap method* and *obvious bootstrap method*[17]. In short, in absence of covariates, simple method consists on drawing bootstrap samples (T^*, δ^*) by independent sampling of size n with replacement and assigning equal mass $1/n$ at each selected observation; while the obvious method requires to estimate the distribution of the survival times and censoring time[18], and then draw $Y^* \sim \widehat{F}$, $C^* \sim \widehat{G}$, and define $T^* = Y^* \wedge C^*$ and $\delta^* = 1_{\{Y^* \leq C^*\}}$. In absence of factor and under independence between Y and C, simple method and obvious method are equivalent[19].

To implement obvious method requires the estimation of distribution functions for survival and censoring times. Thus, simple method has important practical advantages because it does not require to impose any assumption on structure of the data when the usual assumptions about the censoring mechanism fail (Efron and Tibshirani, 1986). In such manner, the implementation of the simple method consists in drawing a random

[17]Many other resampling methods have been proposed in the censored data literature; however, by simplicity and asymptotic behavior, Efron's methods are the most used. For instance, in a comparative analysis, Akritas (1986) finds out that the alternative method proposed by Reid (1981), which is based on sampling the weights of uncersored observations, does not produce asymptotically correct confidence bands.

[18]For instance, using the Kaplan-Meier estimator. The Kaplan-Meier estimator of the censoring times can be computed exchanging the role of censored and uncensored observations in Equation (13).

[19]These resampling methods have been modified in order to consider other characteristics of the censored data. van Keilegom and Veraverbeke (1997) extend the obvious method including covariates to the case of forming confidence bands for a nonparametric estimator of the conditional distribution function. In turn, Wang (1991) generalizes the obvious method when data is also left truncated which in this context is not equivalent to the simple method. In this regard, Gross and Lai (1996); Bilker and Wang (1997); Iglesias-Pérez and Gonzalez-Manteiga (2003) indicate that the independence assumptions play a cricual role for validity and equivalence of the resampling methods.

sample of $(T_i^*, \delta_i^*, X_i^*)$ for $i = 1, \ldots, n$ by sampling with replacement and by assigning mass $1/n$ at each triplet.

In turn, for the construction of confidence bands, we consider classical methods such as *percentile* and *hybrid*[20,21] (see Hall (1988); Efron (1992) for a detailed comparison of coverage bands construction methods). There is not a general rule to select the proper method. For 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 computed combining both resampling and bands construction methods. The results reveal that there is no single winner and the pertinence of each method depends on the target parameter.

5 Monte Carlo Exercises

In order to assess the finite samples properties of the ACOX estimator, we carry out Monte Carlo exercises. Our target is to determine whether there is an effect of the censoring level on estimation procedure and how Cox model performs under different data generator process. To do so, we follow the standard procedures to generate survival times under censoring in the literature (see for instance Stute, 1993, and Uña álvarez, 2004). For that, the set of covariates, survival times and censoring times are simulated to define the observed vector (T, δ, X) under two different Data Generating Processes (DGP). In particular, we generate durations following the proportional hazard assumption by using the Weibull distribution and normal times as in Stute (1993). In terms of the censoring mechanism, we consider situations where censoring variable is independent of (Y, X), and where censoring is independent conditional to covariates.

[20]Describing briefly 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 $\widehat{\theta}^*$ and its distribution given by K. The percentile method sets the confidence interval as:

$$\left(K^{-1}(\alpha)\,,\;K^{-1}(1-\alpha)\right)$$

Instead of approximating the distribution of $\widehat{\theta}^*$, the hybrid method approximates the distribution of $\widehat{\theta} - \theta$ through the distribution of $\widehat{\theta}^* - \widehat{\theta}$, and define the coverage band as follows:

$$\left(2\widehat{\theta} - K^{-1}(1-\alpha)\,,\;2\widehat{\theta} - K^{-1}(\alpha)\right)$$

One important advantage of these methods is that estimation of variances is not needed.

[21]To the case of function-valued parameters, uniform bands built on Kolmogorov-Smirnov maximal t-statistic can be applied as well.

To control censoring levels, scale and shape parameters of censoring times are shifted to generate censoring levels of 5%, 20% and 50%. Regarding the simulation parameters, we consider sample sizes of 50, 500 and 2500, and the number of draws is set in 1000. Finally, by simplicity it is assumed a single covariate following an uniforme distribution in the $(0,1)$ interval. The information about the benchmark Monte Carlo exercises is summarized in Table 1.

Table 1: General Simulation Parameters

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

Once data is generated, we estimate the the empirical survival distribution of Y, denoted by $\widetilde{S}_Y$. This will be the benchmark for comparison purposes. The global performance is analyzed through three indicators based on the distance between the ACOX estimator and the empirical survival distribution. To be specific:

$$MD = \max_{y \in \Omega_y} \left| \widetilde{S}_Y(y) - \widehat{S}_Y(y) \right|, \qquad AD = \frac{1}{n}\sum_{i=1}^{n} \left| \widetilde{S}_Y(y) - \widehat{S}_Y(y) \right|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} \left(\widetilde{S}_Y(y) - \widehat{S}_Y(y) \right)^2$$

where Ω is the information set used in the estimation. First, MD is the maximum distance. The average distance AD measures the accuracy over all distribution, and, MSE computes the mean squared error. As results, we report all of the indicators multiplied by 1000 to facilitate comparisons. For MD and AD we report the average over the total draws, while for the latter, it is reported the squared root of the mean value.

5.1 Lack of Independence

The estimator given by Equation (14) exploits the relation between survival times and covariates to estimate properly the survival distribution when the independence assumption fails. To check the pertinence of implementing the ACOX estimator instead the Kaplan-Meier estimator when independence assumption is not held, our first simulation exercise consists on comparing the performance of these estimators. To do so, survival times and censoring times are assumed to follow Weibull distribution as in Table 1. We compute the conditional survival distribution according to Equation (17) and the baseline hazard quantitites using the Breslow's estimator. In the next section, particular simulations are performed in order to evaluate finite sample properties of Kalbfleisch-Prentice's estimator.

As expected, results reported in Table 5 suggest that, under independence assumption, Kaplan-Meier estimator outperforms the ACOX estimator in absence of censoring and even when censorship level is slight. But, with medium or heavy censoring levels there is not important differences between these competing estimators. Moreover, with heavy censoring level, the performance indicator are marginally smaller for the ACOX estimator.

When conditional independence respect to the factors is considered, it is noticeable that the bias of the Kaplan-Meier estimator clearly increases as censoring becomes more substantial and does not decrease importantly with the sample size. In contrast, ACOX estimator improves as sample size increases. Consequently, ACOX estimator is well behaved under multiple variations of the simulation setup.

5.2 Estimation Effect of Censoring

Because the performance of the ACOX estimator can be influenced by the features of the DGP, as a second stage, to provide some aditional evidence of the estimation effect of censoring, we implementing different estimators of the baseline hazard function and different DGP. In particular, we study finite sample properties of the classical estimators of the baseline hazard function proposed by Breslow (1974) -B- and Kalbfleisch and Prentice (1973) -KP-. And also, to explore flexibility the Cox model, we follow the underlying model in Stute (1993) by generating normal times. In these simulation exercises, we keep the two scenarios of dependence structure between Y and C and implement the DGP described in Table 1.

Table 6 presents the results for the three performance indicators for the case where survival times and censoring times are assumed independent. Three general results can be pointed out: *i.* the implementation Breslow's or Kalbfleisch-Prentice's estimators has no effect when sample is big enough, *ii.* censorship has effect but it is attenuated by increasing the sample size *iii.* ACOX estimator has an aceptable performance as estimator of the marginal survival distribution even when the proportional hazard assumption fails. The latter is also reported by Hsu and Taylor (2010).

Respect to the estimation of the baseline hazard, the KP estimator outperforms the B estimator, mainly in small samples. It is explained by the nature of this estimator since it is proposed in the context of discrete survival times. However, Such difference vanishes rapidly as sample size increases. It is importante to note that Breslow's estimator has some practical advantages in terms of implementability given that it does not require to solve complicated non-linear equations as KP estimator.

The performance of the estimator is affected by the censoring level even for large samples. But, considering loss of information in 50% of the observations, ACOX estimator shows good accurancy for estimating the unconditional survival distribution. All of these results also arise for the case of conditional independence (see Table 7).

5.3 Censoring Patterns

One peculiar result from the previous simulation exercise emerges by comparing the evaluation indicators across DGPs. In particular, it is observed a relative higher accurancy when data is generate from a Normal distribution. A quick reasoning leads to identify that censoring patterns might turn out in a relevant factor. A priori, one might think that a situation when the most of the lost information is located in the upper part of the distribution (long durations), affects much more the estimation respect to the case when the censored observations are uniformly located over durations distribution. That may depend on how long the follow up time is or on the reasons of withdrawal. Hence, censoring patterns may vary case to case.

To understand the role of censoring patterns, we take as reference data from Spanish labor market transitions collected by the Survey of Income and Living Conditions (SILC) which is carried on by European Commission. This survey allows to follow monthly the individual occupational status for a period of 4 years. We consider the cohorts 2004-2007 and 2008-2011 to contruct the unemployment durations. A particular fact to highlight is that during this period Spain suffered a deep depression inducing a rise in the unemployment rate from 11% in 2004 to 21.8% in 2011.

By using information of 1746 and 2454 observations with censoring levels of 17% and 27.3% respectively, we compute the cumulative proportion of censored observation as function of the quantiles of the observed times. That is, it is constructed a function representing the percentage of censored observation located below a given quantile of T. This function is also calculated for two typical draws of the DGPs under Weibull and Normal distributions. As reference we consider the case when Y and C are independent and the baseline survival distribution is estimated through Breslow's estimator. The resulting functions are presented in Figure 1.

Panels a and b show the patterns observed in SILC and in simulated data at censorship levels of 20% and 50% respectively. As result, according to SILC data, there is concentration of censored observations in the upper tail of the distribution. Below the median, only 26% of the total censored data are located. Respect to the simulated data, it is observed that both Weilbull and Normal distributed times generate similar patterns at 20% of censorship level, but differ specially for the case of Normal distribution for data heavily censored. Therefore, it is appealing to study the role of censoring patterns.

To do so, simulation parameters are re-configured to replicate censoring patterns in SILC data with levels around 30% (see Table 2). The resulting censoring patterns are presented in Panel c of Figure 1. In turn, Table 8 reproduces the previous results when censoring levels are 20% and 50%, and reports the results of simulating SILC censoring pattern at 30% of censoring level.

These results suggest that the concentration of censored observations affects the estimation of the distribution function. It is noticeable that for the case of Normal times, the improvement of the performance of the ACOX estimator as sample size increases, is significantly lower when censored observation are concentrated as SILC. It seems natural given that the estimation of the upper tail is highly inefficient, and that fact is exacerbated by the amount of missing mass at the least upper bound of the observable survival times.

Table 2: Simulation Parameters: Censoring Patterns

Weibull	$Y \sim WB\left(e^{2-x}, 5\right)$ $C \sim WB\left(e^{1.6}, 6\right)$
Normal	$Y = 15 + X + \varepsilon_Y, \;\; \varepsilon_Y \sim N\left(0, 4\right)$ $C = 15 + \varepsilon_C, \;\; \varepsilon_C \sim N\left(1.5, 2\right)$

5.4 Decomposition Exercise of Target Parameters

Since bootstrap techniques are valid to make inference on the survival counterfactual distribution and its functionals, we run a simulation exercise in order to study the finite sample performance of decomposing the mean lifetime and the quartiles. Since the misidentification problem in the upper tail, we truncate in the case of the mean lifetime. In sake of simplicity, DGP is set such that all difference between the two populations is due to the covariates distribution. To do so, we consider a covariate X_1 generated as uniform $(0, 1)$, while X_0 is the convolution of three independent uniform distributions in the interval $(0, 1/3)$. It is assumed that both survival times and censoring times follow Weibull distribution and are conditionally independent. The distribution of censoring times is shifted to achieve censoring levels of 30%. In this manner, we define $Y_D \sim WB\left(e^{3-x_D}, 5\right)$ and $C_D \sim WB\left(e^{3.17-x_D}, 5\right)$ for samples of 500 observations.

Once the vectors (T_D, δ_D, X_D) are computed, we estimate the unconditional survival distribution for each sample according to Equation (14), and the counterfactual distribution $S_{Y\langle 0,1\rangle}$ which corresponds to the the survival distribution we would observe if population 0 has the same covariates profile as population 1. The conditional survival distribution is modelled by using the Breslow estimator. Because all the difference between the observed populations is given by different covariates profiles, we should observe no differences between $\widehat{S}_{Y\langle 1,1\rangle}$ and $\widehat{S}_{Y\langle 0,1\rangle}$ as Figure 2 illustrates (see Appendix).

In order to decompose the target parameters θ we follow the spitrit of the OB decomposition (see Equations -4- and -6-). That is:

$$\begin{aligned} \theta_{Y\langle 1,1\rangle} - \theta_{Y\langle 0,0\rangle} &= \theta_{Y\langle 1,1\rangle} - \theta_{Y\langle 0,1\rangle} + \theta_{Y\langle 0,1\rangle} - \theta_{Y\langle 0,0\rangle} \\ &= \Delta_S^{\theta} + \Delta_X^{\theta} \end{aligned} \tag{19}$$

Under the aforementioned parameterization, it is expected that composition effect be statistically significant and the structure effect equal to zero. In order to check these hypotheses we compute the coverage rate for the structure effect at zero. For that, we consider 1000 bootstrap repetitions and 1000 draws of the exercise. In turn, resampling is executed using the simple method, and coverage bands (at 95% and 90% of confidence level) are constructed according to percentile and hybrid methods.

In Table 9 we report the results. First, it is noticeable that percentile method tends to outperform the hybrid method, although the difference is quite small. In the case of no censoring, results are similar to those reported in CFM. In turn, when censoring is present no only the level is important but also the difference respect to the censoring level of the other population. Particularly, the coverage rate is close to its nominal value if the two populations exhibit similar censoring levels. This is explained by the lost of accuracy in the upper tail induced by the missing probability mass. In general, the performance of the ACOX estimator for decomposition exercises is satisfactory.

6 Decomposition Exercise of the Unemployment Duration in Spain

The recent Great Recession experienced in the most of the developed countries between the second half of 2008 and 2009 has left tremendous negative numbers, mainly expressed in the labor markets imbalances. In the US, Greece, Portugal, and specially, Spain, unemployment rate increased rapidly. In the case of Spain, unemployment rate doubled from 11% in 2004 to 21.8% in 2011 and reached 25% one year later, the highest levels in all European Union. This situation was even more dramatic for some subgroups, for instance, the unemployment rate of young population had more dramatic variations by changing from 22% up to 53.2% in the same period. This overreaction in the labor market performance has been attributed to the particular structure of the labor market in Spain which in the preceding years to the economic downturn , compared with other European developed countries, was characterized by high immigration and rigid labor markets institutions (see Aiginger et al., 2011, and, Aceleanu, 2013). That scenario motivates to quantify the which factors accounting the variation of the main labor markets indicators (see Elsby et al., 2011; Anderton et al., 2012; Bertola, 2013; Kroft et al., 2013; Tridico, 2013, and, Junankar, 2014 as reference of some related literature, mainly studying the case of the US).

To understand the dynamics in the labor market, search-matching theory provides a pertinent framework (Diamond, 1982; Pissarides, 1979, and, Mortensen, 1986). According to this theory, probability at which a workers escapes unemployment depends on the arrival offer rate and the probability of accepting that offer. Consequently, this process involves factors associated to characteristics of the unemployed workers such as human capital accumulation and reservation wage as well as the labor market situation. For instance, if a negative economic shock occurs, the pool of unemployed workers increases, among other reasons, due to the acceleration of the job displacement rate and the reduction of job creation rate. Moreover, there are new job searchers motivated by reduction of household income enter to unemployment situation (added worker effect). Therefore, the stagnation of the leaving rate from unemployment is exacerbated by variations in the average unemployed profile, which turns out in unemployment situations more persistent and prolonged.

In view of that complex puzzle, to study the variation in the unemployment profiles, we take into account changes in unemployed population composition and variations of factors associated to the labor market structure. In such way, using information from two cohorts of the SILC, 2004-2007 and 2008-2011, we perform decomposition exercises in order to study the relative role of individuals characteristics and labor market factors in the total variation of the unemployment duration profile. In particular, we examine variations of the average duration of unemployment and variations of the probability of being a long term unemployed. This analysis allows to identify whether the changes in unemployment duration profile are heterogeneous across the distribution.

To do so, we take into account a set of explanatory variables commonly use in unemployment duration studies such as gender, age, educational level, tenure, marital status, whether individual is head of the household and the number of unemployed (see for instance Foley, 1997; Addison and Portugal, 2003; Kuhn and Skuterud, 2004; Biewen and Wilke, 2004, and, Tansel and Tasci, 2010). While the first three variables control by human capital characteristics, the rest gives information about the opportunity cost of being unemployed and the reservation wage. In addition, we include city size and region to control by specific labor market characteristics[22].

As exploratory analysis, Table 10 presents the mean and standard deviation of the covariates for each cohort and the results of classical test for differences in the mean. These tests report differences statistically significant for all of the covariates but age, suggesting that variation in the population composition might be an important ingredient to explain changes in the unemployment duration distribution. In particular, it is worth mentioning that the more important difference is reported in tenure, which provides evidence in favor of added worker effect.

Taking into account this set of covariates, marginal distribution of unemployment duration is estimated for each cohort as well as the distribution prevailing if individuals from the cohort 2008-2011 would face the schedule of the cohort 2004-2007, which is estimated by implementing the counterfactual operator $\widehat{S}_{Y\langle 0,1\rangle}$. Once marginal distributions of the potential outcomes are estimated, we focus the analysis on the mean lifetime at 48 months and the probability of surviving as unemployment for long time; particularly, the probability of being unemployed longer than 12, 24 and 36 months.

Table 3 presents the estimates for the mean life time and the survival distribution at 12, 24 and 36 months. As general result, it is noticeable that both average duration and probability of long unemployment spells have increased. In comparison to the period 2004-2007, unemployment duration turns out around 4 months higher at average while the probability of being unemployed more than 12, 24 and 36 months increases by 8.2, 10.3 and 10.2 percentage points.

In practice, sometimes in unemployment duration studies, or in general about variables that are censored, statistical inference is made based on a subsample containing only uncensored observations. To illustrate the kind of misleading inference that this solution to censoring can induce, the bottom part in Table 3 presents the same estimates when only uncensored observations are considered. In this case, as expected, lower magnitudes are obtained because censoring observation tend to happen for longer durations. Moreover, by comparing across cohorts, it is concluded that the average duration and the probability of being unemployed for a long period are reduced, which contradicts the aforementioned results.

[22] These variables except age and tenure are dummy variables. The reference categories are female, primary school, other than married, other than household's head and village, respectively.

Table 3: Average Duration of Unemployment and Probability of being Long Term Unemployed

		Mean lifetime	S(12)	S(24)	S(36)
Full Sample	**2004-2007**	10.738	0.227	0.092	0.051
	2008-2011	14.699	0.309	0.195	0.153
	Counterfactual	12.654	0.285	0.132	0.079
Only Uncensored	**2004-2007**	7.764	0.135	0.029	0.004
	2008-2011	7.166	0.107	0.021	0.002
	Counterfactual	8.558	0.166	0.041	0.007

Authors' calculations.

6.1 Aggregate Decomposition

As first exercise, following Equation (19), we perform an aggregate decomposition to quantify the contribution of the labor market factors and composition factors in the variation of the unemployment duration profile. Estimates of this decomposition for the target statistics are contained in Table 4, which also presents confidence intervals at 90% built through 1000 bootstrap repetitions by using the percentile method.

Table 4: Aggregate Decomposition. Unemployment Duration in Spain 2004-2011.

		Total	Structure effect	Composition effect
Mean lifetime	**Difference**	3.9610	2.0440	1.9169
	CI 90%	[3.1958 , 4.7201]	[1.0394 , 2.8857]	[1.3350 , 2.5337]
S(12)	**Difference**	0.0811	0.0235	0.0575
	CI 90%	[0.0575 , 0.1073]	[-0.006 , 0.0509]	[0.0411 , 0.0745]
S(24)	**Difference**	0.1034	0.0636	0.0397
	CI 90%	[0.0830 , 0.1233]	[0.0373 , 0.0878]	[0.0263 , 0.0534]
S(36)	**Difference**	0.1018	0.0739	0.0279
	CI 90%	[0.0829 , 0.1205]	[0.0495 , 0.0958]	[0.0179 , 0.0392]

Authors' calculations.

Results reveal that both structure effect and composition effect, in general, play relevant role to explain the variation in the unemployment duration profile. In particular, it is found that around one half of the variation in the average duration of unemployment is due to changes in the individual characteristics of the unemployed workers. That implies that the entrance of less experienced workers, for whom the exit of unemployment takes longer time, increases the average duration of unemployment in almost two months. With respect to the structure effect, which can be associated to labor market circumstances such as labor market tightness and job destruction, turns out significant explaining the other half of the total variation.

Regarding the probability of having long duration unemployment, evidence is mixed. In the probability of being unemployed longer than 12 months, composition effect is approximately 70% of the total variation, while for the probabilities of having 24 and

36 months, structure effect dominates explaining 61% and 72% of the total variation, respectively. This is indicating that there is no a monotonic effect across unemployment duration distribution and that long term unemployed are more severely affected. This relevance of the duration dependence and less significant role of the composition effect to explain the recent variation of the unemployment indicators was also documented by Kroft et al. (2013) in the case of the U.S. in a similar period.

6.2 Decomposing Structure Effect

The convenient representation of the survival distribution under the Cox model admits a more detailed decomposition of the structure effect. As it was discussed, the conditional distribution is function of the baseline survival and a set of parameters. Such decomposition allow to separate the effect of the labor market conditions common for all workers and the structure effect due to the the marginal effect of the characteristics.

In other words, everything else constant, variations in the baseline survival represents changes in labor market circumstances that are independent of the individuals characteristics, e.g. labor market tightness. In turn, changes in the parameters are related to individual job search processes. That is, marginal effects show how much labor market appraises workers' characteristics and, given these characteristics, how they take decisions about offers and potential outside options. For instance, because the increasing number of candidates with low tenure, the relative value of having an additional year of experience varies, which also modifies the perception of the unemployed about different offers.

In this manner, the detailed decomposition of the structure affect follows the next representation:

$$\begin{aligned} S_{Y\langle 1,1,1\rangle} - S_{Y\langle 0,0,0\rangle} &= S_{Y\langle 1,1,1\rangle} - S_{Y\langle 0,1,1\rangle} + S_{Y\langle 0,1,1\rangle} - S_{Y\langle 0,0,1\rangle} + S_{Y\langle 0,0,1\rangle} - S_{Y\langle 0,0,0\rangle} \\ &= \Delta_{SCR} + \Delta_{SR} + \Delta_{X} \end{aligned}$$

where $S_{Y\langle i,j,k\rangle}$ is the survival distribution of the potential outcome considering the baseline quantity i with set of parameters j and the covariates set k. Therefore, using the same logic than in the aggregate decomposition, the difference between the underlying survival distributions can be summed up in three quantities, named: the composition effect, and the structure effect is splitted into common risk effect Δ_{SCR} and residual structure effect Δ_{SR}.

The results of the detailed decomposition of the structure effect presented in Table 11 shows the expected signs for common risk effect and residual effect, however, the latter turns out statistically non significant. This result implies that the economic downturn increases unemployment duration through macroeconomic channels, but it reduces by a sort selection effect because unemployed workers might be less selective (Groot and van der Klaauw, 2013).

Since intuitively the common risk effect can be interpreted as the net variation of the unemployment duration once all the effects associated to idiosyncratic variations are discounted, its relative importance states that the structure effect is mainly driven by aggregate (macroeconomic) factors affecting simultaneously all unemployed workers. In such way, rather than market selection according to individuals profile, the probability of leaving from unemployment is guided by a general downturn in the labor market performance. Lastly, with respect to the previous results, in the case of the probability of being long term unemployed, it is found that the common risk effect follows a similar increasing pattern across unemployment duration distribution.

6.3 Multiple transition Options: Competing Risk Approach

We have studied the unemployment duration taking as reference the time until leaving from unemployment with no distinction between possible destinations. This limits the posibility of exploring differences across potential transitions, e.g., the time until getting a job or the time until be out of labor force (OLF). Such distinction is specially important for assessing labor policies or studying the behavior of job searchers. For instance, in order to evaluate the impact of a labor intermediation program whose goal is find a job, to consider the two type of transitions as the same event might induce misleading conclusions.

Furthermore, there is a large literature supporting the evidence of spikes in the job finding rate (Boone and van Ours, 2012; Mortensen, 1977; Katz and Meyer, 1990, and, Groot and van der Klaauw, 2013, among others), which indicates that individuals have different considerations to choice between employment and being OLF. Therefore, the presence of particular patterns of leaving from unemployment might be identified by separating the possible transitions as different events.

In order to consider this double nature of the unemployment duration, we use a competing risk model (following the approach of Addison and Portugal, 2003 and Farber and Valletta, 2013) and compute the aggregate decomposition and detailed decomposition of the structure effect for each type of unemployment duration. According to the aggregate decomposition (see Table 12), the total difference follows the same qualitative results than the previous exercises. In particular, there is a increasing variation in the probability of being long term unemployed over duration level. Nevertheless, comparing the two possible transitions, contrasting results are found. While in the transition to employment, the composition effect plays a relevant role, the structure effect explains the most of the variation of the target statistics in the transition from unemployment to OLF.

In such context, one might argue that in the unemployment-to-employment transition, all the variation in the unemployment duration is consequence of the composition effect, or the same, of the adjustment in the pool of unemployed workers. But with respect to the detailed decomposition of the structure effect (see Table 13), it is noticeable that, rather than negligible, structure effect is the sum of two opposite forces where both the common risk effect and the residual structure effect are significant with similar magnitudes.

Although no significant the two components are positive to the case unemployment-to-OLF. And the common risk effect tends to be quantitative more important. Besides, and the common risk effect tends to be quantitative more important, which means that, during economic slump, unemployed workers delayed the decision of being out of the labor force as result of a higher opportunity cost.

7 Concluding Remarks

Many fields in economics have focused on studying the effect of policy interventions. The most recent statistical approaches allow for estimating the distribution of potential outcomes and computing distributional effects. There are many means for estimating distributional effects of policy where the related variable is always observable, however is usual to find that data is not completely observed. Considering the estimation of counterfactual outcomes based the classical Cox model, random censoring can be adapted. It allows for addressing questions related with economic outcomes denoting duration such as unemployment duration and firms' bankrupcy.

By combining decomposition analysis literature and identifiability conditions in the context of censored data, it is found out that the proposed estimator of the counterfactual distribution is robust to the independence structure between survival times and censoring times. Taking into account classical assumptions in decomposition analysis and asymptotic results of the Cox model, uniform convergence and validity of statistical inference based on bootstrap techniques are guaranteed for the survival counterfactual distribution and its functionals.

To assess the global performance in finite sample of this estimator, Monte Carlo exercises are conducted under for different scenarios of the distribution of survival times and independence structure respect to the censoring variable. In general, the estimator of the unconditional survival distribution performs properly. Simulation exercises reveal that not only censoring level has an estimation effect of the outcome distribution but also censoring pattern matters. In particular, when simulated data are configured to replicate observable censoring patterns, the performance of the estimator for the outcome distribution is favorable.

Lastly, we perform a decomposition exercise using unemployment data from Spain in order to analyze the variation of the average duration of unemployment and the probability of being unemployed long term. As result, structure effect explains a half of the variation of the average duration of unemployment and drives the the variation in the probability of long term unemployment. A detailed exercise to decompose the structure effect between factors associated to common risk of being unemployed and effect of the covariates on the probability of leaving from unemployment suggests that the former plays a more relevant role mainly for explaining the particular transition from unemployment to employment.

References

Aalen, O. (1978): "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6, 701–726.

Aceleanu, M. (2013): "The labour Market in the Post-Crisis Economy: the Case of Spain." *Theoretical & Applied Economics*, XX, 135–146.

Acemoglu, D., S. Johnson, and J. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *The American Economic Review*, 91, 1369–1401.

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

Aiginger, K., T. Horvath, and H. Mahringer (2011): "Why labor market performance differed across countries : the impact of institutions and labor," .

Akritas, M. (1986): "Bootstrapping the Kaplan-Meier Estimator," *Journal of the American Statistical Association*, 81, 1032–1038.

——— (1994): "Nearest Neighbor Estimation of a Bivariate Distribution under Random Censoring," *The Annals of Statistics*, 22, 1299–1327.

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

Anderton, R., M. Izquierdo, and T. Aranki (2012): "Euro Area Labour Markets and the Crisis," .

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

——— (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, 730–736.

Beran, R. (1981): "Nonparametric Regression with Randomly Censored Survival Data," Tech. rep.

Bertola, G. (2013): "Labor Market Policies and European Crises," *IZA Journal of Labor Policy*, 3, 1–25.

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

Bilker, W. B. and M.-C. Wang (1997): "Bootstrapping Left Truncated and Right Censored Data," *Communications in Statistics - Simulation and Computation*, 26, 141–171.

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

Boone, J. and J. C. van Ours (2012): "Why is There a Spike in the Job Finding Rate at Benefit Exhaustion?" *De Economist*, 160, 413–438.

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

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

Card, D. and A. Krueger (1993): "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania," *The American Economic Review*, 84, 772–792.

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

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

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

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

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

——— (1989): "Uniform Consistency of the Kernel Conditional Kaplan-Meier Estimate," *The Annals of Statistics*, 17, 1157–1167.

der Vaart, A. V. and J. Wellner (2004): "Weak Convergence and Empirical Processes: With Applications to Statistics. 1996," *Springer, New York*.

Diamond, P. (1982): "Wage determination and efficiency in search equilibrium," *The Review of Economic Studies*, 49, 217–227.

DiNardo, J., N. Fortin, and T. Lemieux (1995): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

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

——— (1981): "Censored Data and the Bootstrap," *Journal of the American Statistical Association*, 76, 312–319.

——— (1992): "Six Questions Raised by the Bootstrap," Tech. rep.

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

Elsby, M., B. Hobijn, and A. Sahin (2011): "The Labor Market in the Great Recession-An Update to September 2011," *Brookings Papers on . . .* , 353–384.

Farber, H. and R. Valletta (2013): "Do Extended Unemployment Benefits Lengthen Unemployment Spells? Evidence from Recent Cycles in the US Labor Market," .

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

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition Methods in Economics," in *Handbook of Labor Economics*, 1–102.

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

Gill, R. (1980): "Censoring and Stochastic Integrals," Tech. rep.

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

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

Groot, N. D. and B. van der Klaauw (2013): "The Effect of Reducing the Entitlement Period to Unemployment Insurance Benefits," .

Gross, S. and T. Lai (1996): "Bootstrap Methods for Truncated and Censored Data," *Statistica Sinica*, 6, 509–530.

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

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

Hsu, C.-H. and J. M. G. Taylor (2010): "A Robust Weighted Kaplan-Meier Approach for Data with Dependent Censoring using Linear Combinations of Prognostic Covariates." *Statistics in medicine*, 29, 2215–23.

Iglesias-Pérez, M. and W. Gonzalez-Manteiga (2003): "Bootstrap for the Conditional Distribution Function with Truncated and Censored Data," *Annals of the Institute of Statistical Mathematics*, 55, 331–357.

Imbens, G. and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Jorunal of Economic Literature*, 47, 5–86.

Junankar, P. (2014): "The Impact of the Global Financial Crisis on Youth Labour Markets," .

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

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

Katz, L. and B. Meyer (1990): "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment," *Journal of public economics*, 41, 45–72.

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

Koenker, R., S. Leorato, and F. Peracchi (2013): "On Distributional vs. Quantile Regression," .

Kroft, K., F. Lange, M. Notowidigdo, and L. Katz (2013): "Long-term Unemployment and the Great Recession: the role of Composition, Duration Dependence, and Non-participation," .

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

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

Lopez, O. (2011): "Nonparametric Estimation of the Multivariate Distribution Function in a Censored Regression Model with Applications," *Communications in Statistics - Theory and Methods*, 40, 2639–2660.

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

Malani, H. (1995): "A Modification of the Redistribution to the Right Algorithm using Disease Markers," *Biometrika*, 82, 515–526.

Mauro, D. (1985): "A combinatoric approach to the Kaplan-Meier estimator," *The Annals of Statistics*, 13, 142–149.

McLain, A. and S. Ghosh (2009): "Nonparametric Estimation of the Conditional Mean Residual Life Function based on Censored Data," Tech. rep.

Mortensen, D. (1977): "Unemployment Insurance and Labor Supply Decisions," .

Mortensen, D. T. (1986): "Chapter 15 Job search and labor market analysis," *Handbook of Labor Economics*, 2, 849–919.

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

Nelson, W. (1969): "Hazard Plotting for Incomplete Failure Data," *Journal of Quality Technology*, 1, 27–52.

Nopo, H. (2008): "Matching as a Tool to Decompose Wage Gaps," *The review of economics and statistics*, 90, 290–299.

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

Pesaran, M. and R. Smith (2012): "Counterfactual Analysis in Macroeconometrics: An Empirical Investigation into the Effects of Quantitative Easing," .

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

Pissarides, C. (1979): "Job matchings with state employment agencies and random search," *The Economic Journal*, 89, 818–833.

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

Reid, N. (1981): "Estimating the Median Survival Time," *Biometrika*, 68, 601–608.

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

——— (2012): "Partial Distributional Policy Effects," *Econometrica*, 80, 2269–2301.

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

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

Satten, G. a., S. Datta, and J. Robins (2001): "Estimating the Marginal Survival Function in the Presence of Time Dependent Covariates," *Statistics & Probability Letters*, 54, 397–403.

Stock, J. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association*, 84, 567–575.

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

——— (1999): "Nonlinear Censored Regression," *Statistica Sinica*, 9, 1089–1102.

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

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

Tridico, P. (2013): "The Impact of the Economic Crisis on EU Labour Markets: A Comparative Perspective," *International Labour Review*, 152, 175–190.

Tsiatis, A. (1975): "A Nonidentifiability Aspect of the Problem of Computing Risks," in *Proceedings of the National Academy of Sciences*, vol. 72, 20–22.

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

Uña Álvarez, J. D. (2004): "Nonparametric Estimation under Length-Biased Sampling and Type I Censoring: A Moment Based Approach," *Annals of the Institute of Statistical Mathematics*, 56, 667–681.

van Keilegom, I. V. and N. Veraverbeke (1997): "Estimation and bootstrap with censored data in fixed design nonparametric regression," *Annals of the Institute of Statistical Mathematics*, 49, 467–491.

Wang, M. (1991): "Nonparametric Estimation from Cross-Sectional Survival Data," *Journal of the American Statistical Association*, 86, 130–143.

Appendix

Table 5: Comparison between Kaplan-Meier and ACOX Estimators

$Y \perp C$

Censoring Level	Sample Size	Kaplan-Meier			ACOX		
		MD	**AD**	**MSE**	**MD**	**AD**	**MSE**
0	**50**	0.00	0.00	0.00	26.81	9.20	0.39
	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
0.05	**50**	23.54	5.54	0.32	34.92	11.75	0.50
	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
0.2	**50**	94.91	22.99	1.13	97.53	24.43	1.16
	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
0.5	**50**	226.17	54.30	2.62	226.79	53.66	2.61
	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			ACOX		
		MD	**AD**	**MSE**	**MD**	**AD**	**MSE**
0	**50**	0.00	0.00	0.00	26.77	9.20	0.39
	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
0.05	**50**	24.75	7.10	0.37	32.71	10.69	0.46
	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
0.2	**50**	80.70	29.71	1.30	54.24	15.98	0.70
	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
0.5	**50**	212.58	82.98	3.50	114.75	32.68	1.48
	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

Table 6: Global Performance ACOX 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

Table 7: Global Performance ACOX 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

Table 8: Global Performance ACOX Estimator: Censoring Patterns

Weibull Times				
Censoring Level	**Sample Size**	**MD**	**AD**	**MSE**
0.2	**50**	98.03	24.46	1.08
	500	41.18	7.69	0.62
	2500	17.73	2.68	0.38
0.5	**50**	227.41	53.73	1.62
	500	147.01	23.51	1.15
	2500	101.68	12.19	0.89
SILC (0.3)	**50**	231.78	42.79	2.36
	500	131.57	16.42	1.03
	2500	77.20	7.19	0.52

Normal Times				
Censoring Level	**Sample Size**	**MD**	**AD**	**MSE**
0.2	**50**	82.44	18.64	0.96
	500	27.45	5.07	0.27
	2500	9.74	1.84	0.10
0.5	**50**	162.23	36.83	1.83
	500	62.71	10.87	0.59
	2500	23.91	3.66	0.21
SILC (0.3)	**50**	183.07	32.75	1.84
	500	97.32	11.99	0.76
	2500	54.03	5.07	0.36

Table 9: Decomposition Exercise: Mean Lifetime and Quartiles

Confidence Level	**Censoring Levels**		**Truncated Mean**		**Q(0.50)**	
	$\Pr(\delta_0 = 0)$	$\Pr(\delta_1 = 0)$	**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)**	
	$\Pr(\delta_0 = 0)$	$\Pr(\delta_1 = 0)$	**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 10: Descriptive Statistics. Determinants of Unemployment Duration

Variable	**2004-2007**		**2008-2011**		**Diff.**
	Mean	**Stand. Dev.**	**Mean**	**Stand. Dev.**	
Gender (male)	0.363	0.012	0.448	0.010	***
Age	39.837	0.298	39.917	0.253	
Primary	0.285	0.011	0.250	0.009	**
Secondary	0.506	0.012	0.517	0.010	
University or higher	0.181	0.009	0.200	0.008	
Tenure	13.549	0.301	6.841	0.227	***
Marital staus (married)	0.641	0.011	0.559	0.010	***
Head of household	0.321	0.011	0.347	0.010	*
Number unemployed	0.133	0.011	0.160	0.008	**
n	1746		2454		

Authors' calculations. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Decomposition Structure Effect. Unemployment Duration in Spain 2004-2011.

		Structure effect	Common risk effect	Residual structure effect
Mean lifetime	**Difference**	2.0440	3.8504	-1.806
	CI 90%	[1.0394 , 2.8857]	[0.1629 , 6.7847]	[-4.865 , 1.8458]
S(12)	**Difference**	0.0235	0.0793	-0.055
	CI 90%	[-0.006 , 0.0509]	[-0.025 , 0.1689]	[-0.144 , 0.0482]
S(24)	**Difference**	0.0636	0.0950	-0.031
	CI 90%	[0.0373 , 0.0878]	[0.0171 , 0.1505]	[-0.086 , 0.0449]
S(36)	**Difference**	0.0739	0.0942	-0.020
	CI 90%	[0.0495 , 0.0958]	[0.0339 , 0.1317]	[-0.058 , 0.0363]

Authors' calculations.

Table 12: Competing Risk and Aggregate Decomposition. Unemployment Duration in Spain 2004-2011.

		From unemployment to employment		
		Total	**Structure effect**	**Composition effect**
Mean lifetime	**Difference**	3.6870	-0.701	4.3883
	CI 90%	[2.7042 , 4.7945]	[-1.812 , 0.6925]	[3.5388 , 5.1940]
S(12)	**Difference**	0.0569	-0.047	0.1046
	CI 90%	[0.0312 , 0.0828]	[-0.074 , -0.017]	[0.0854 , 0.1228]
S(24)	**Difference**	0.0814	-0.020	0.1016
	CI 90%	[0.0533 , 0.1104]	[-0.051 , 0.0164]	[0.0817 , 0.1208]
S(36)	**Difference**	0.1114	0.0201	0.0912
	CI 90%	[0.0821 , 0.1458]	[-0.013 , 0.0602]	[0.0711 , 0.1097]
		From unemployment to OLF		
		Total	**Structure effect**	**Composition effect**
Mean lifetime	**Difference**	5.5787	6.4541	-0.875
	CI 90%	[3.6466 , 7.3021]	[4.5196 , 8.4363]	[-1.997 , 0.2572]
S(12)	**Difference**	0.0564	0.0748	-0.018
	CI 90%	[0.0202 , 0.0927]	[0.0361 , 0.1169]	[-0.043 , 0.0039]
S(24)	**Difference**	0.1765	0.2014	-0.024
	CI 90%	[0.1214 , 0.2308]	[0.1425 , 0.2592]	[-0.056 , 0.0076]
S(36)	**Difference**	0.1892	0.2143	-0.025
	CI 90%	[0.1065 , 0.2617]	[0.1361 , 0.2908]	[-0.055 , 0.0087]

Authors' calculations.

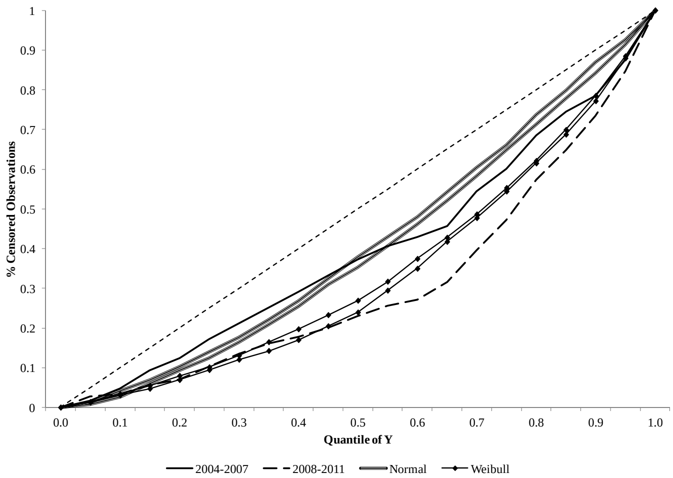
Table 13: Competing Risk and Decomposition Structure Effect. Unemployment Duration in Spain 2004-2011.

		From unemployment to employment		
		Structure effect	**Common risk effect**	**Residual structure effect**
Mean lifetime	**Difference**	-0.701	6.0383	-6.739
	CI 90%	[-1.812 , 0.6925]	[1.6904 , 10.142]	[-10.94 , -2.027]
S(12)	**Difference**	-0.047	0.1201	-0.167
	CI 90%	[-0.074 , -0.017]	[0.0044 , 0.2320]	[-0.281 , -0.047]
S(24)	**Difference**	-0.020	0.1295	-0.149
	CI 90%	[-0.051 , 0.0164]	[0.0340 , 0.2140]	[-0.234 , -0.046]
S(36)	**Difference**	0.0201	0.1495	-0.129
	CI 90%	[-0.013 , 0.0602]	[0.0699 , 0.2160]	[-0.198 , -0.041]
		From unemployment to OLF		
		Structure effect	**Common risk effect**	**Residual structure effect**
Mean lifetime	**Difference**	6.4541	5.2269	1.2272
	CI 90%	[4.5196 , 8.4363]	[-2.359 , 13.773]	[-7.393 , 8.3034]
S(12)	**Difference**	0.0748	0.0502	0.0245
	CI 90%	[0.0361 , 0.1169]	[-0.084 , 0.2634]	[-0.182 , 0.1548]
S(24)	**Difference**	0.2014	0.1666	0.0348
	CI 90%	[0.1425 , 0.2592]	[-0.041 , 0.4075]	[-0.207 , 0.2355]
S(36)	**Difference**	0.2143	0.1781	0.0361
	CI 90%	[0.1361 , 0.2908]	[-0.066 , 0.4120]	[-0.194 , 0.2529]

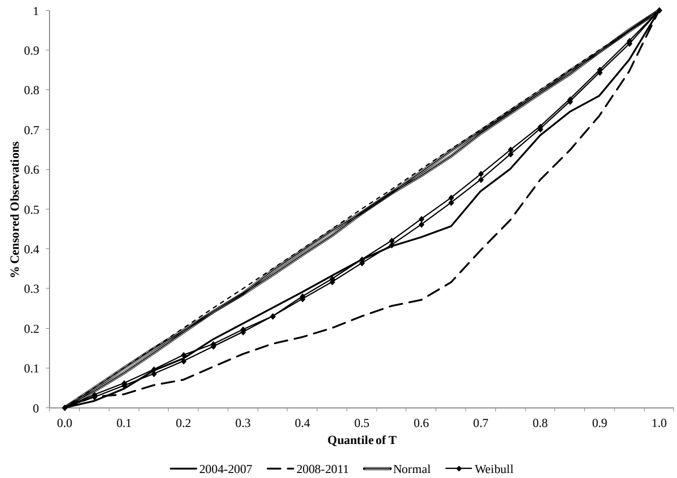
Authors' calculations.

Figure 1: Censoring Patterns in SILC and Simulated Data

a. Censoring Level 20%



b. Censoring Level 50%



c. Censoring Level 30% and SILC Pattern

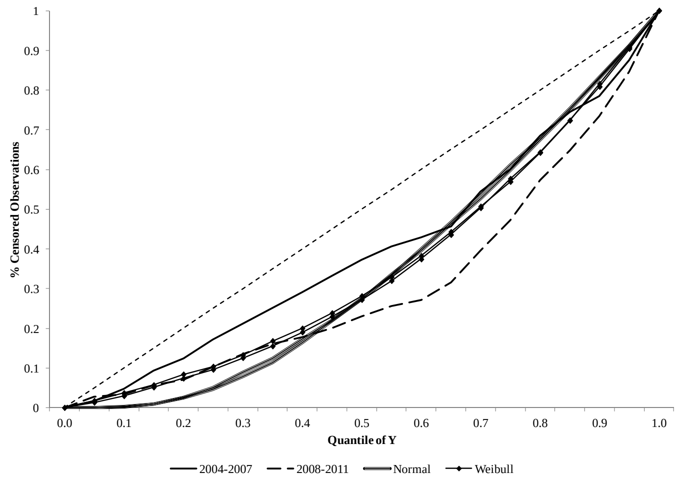


Figure 2: Decomposition Exercise: Simulated Survival Distributions