Testing Censoring Point Independence

Brigham R. Frandsen

Department of Economics
Brigham Young University

June 16, 2016

Abstract

Identification in censored regression analysis and hazard models of duration outcomes relies on the condition that censoring points are conditionally independent of latent outcomes, an assumption which may be questionable in many settings. This note proposes a test for this assumption based on a Cramer-von Mises-like test statistic comparing two different nonparametric estimators for the latent outcome cdf: the Kaplan-Meier estimator, and the empirical cdf conditional on the censoring point exceeding (for right-censored data) the cdf evaluation point. The test is consistent and has power against a wide variety of alternatives. Applying the test to unemployment duration data from the NLSY, the PSID, and the SIPP suggests the assumption is frequently suspect.

keywords: survival analysis, duration data, Kaplan-Meier estimator, censored regression, hazard models

1 Introduction

Censored outcomes occur frequently in empirical work. A few of many possible examples include survival times, unemployment durations, test scores, hours worked, and reported incomes. Censoring poses a challenge for empirical analysis, as has been well known since Stevens (1937). Restricting analysis to uncensored observations or ignoring censoring altogether leads to biased and inconsistent estimators.

Standard censored regression and hazard models address the censoring problem by assuming the censoring point is conditionally uninformative, or independent of the latent outcome conditional on observed covariates. While ubiquitous, this assumption is not innocuous. It may fail whenever censoring points vary across observations, and is especially likely to fail when censoring is due to attrition, competing risks, or follow-up intervals that depend on durations in unobserved ways. When it fails standard estimators are inconsistent and point identification in general is lost.

Despite this assumption's fragility, it is rarely tested. In fact, it cannot be tested when the observed data include only the outcome and a censoring indicator: one can always construct a distribution of latent outcomes that reconciles the assumption of censoring point independence with the observed data (Tsiatis, 1975). The few tests that have been developed rely on extra information that is not typically available to the researcher or impose equally tenuous auxiliary assumptions. For example, Lee and Wolfe's (1998) test relies on post-study data collected on censored units and Huang et al.'s (2004) test requires a specific clustered correlation structure among units and imposes uninformative censoring within clusters for power.

This article proposes a test for censoring point conditional independence that relies only on data commonly available to researchers and imposes no auxiliary assumptions for validity. The test's null hypothesis, H_0, is that censoring points are conditionally independent of latent outcomes, given a set of observed covariates. The alternative hypothesis against which the test is shown to have power is that censoring points and latent outcomes or either positively or negatively dependent in a sense defined in Section 3 below. The test requires the observed data to include the (possibly censored) outcome, censoring points for all observations, and the covariates upon which the censoring points are hypothesized to be conditionally independent of latent outcomes. The requirement that censoring points be observed even for uncensored observations does rule out certain settings, but it covers many common scenarios, such as when the censored outcome is a duration and the censoring point is a follow-up time. Clinical analyses of patient survival times, labor market studies of unemployment durations, and analyses of firm survival times are examples where the test's data requirements are typically met. Many censored estimation methodologies also require censoring points to be observed for all observations (Powell, 1986; Chernozhukov and Hong, 2002; Hong and Tamer, 2003; Blundell and Powell, 2007; Chernozhukov et al., 2011; Frandsen, 2015).

The test is based on a Cramer-von Mises-like comparison of two nonparametric estimators for the latent outcome cdf: the Kaplan-Meier estimator (Kaplan and Meier, 1958), and the empirical cdf of observed outcomes conditional on the censoring point exceeding the cdf evaluation point (the "reduced-sample" estimator). The test statistic therefore has a well-known limiting distribution and is straightforward to compute. Simulations show the test has good size and power properties in finite samples. Applying the test to unemployment duration data from the National Longitudinal Survey of Youth (NLSY), the Panel Study of Income Dynamics (PSID), and the Survey of Income and Program Participation (SIPP) suggests the assumption frequently fails.

In addition to testing the conditional independence assumption, the procedure also provides guidance on the nature of the violation if the assumption is rejected. When latent outcomes and censoring points are positively dependent, the reduced-sample estimator's probability limit first-order stochastically dominates that of the Kaplan-Meier estimator, and the reverse when latent outcomes and censoring points are negatively dependent. The difference between the two estimators therefore indicates not only the presence but also the direction of the violation, and can be used to deduce the direction of bias of estimators that rely on independence.

2 Statistical Framework

Consider a setting with an outcome variable Y^* that is right-censored at a (possibly random) censoring point T with support $\mathcal{T} \subseteq \mathbb{R}$. For example, Y^* may be unemployment duration, and T the elapsed time from the start of the unemployment spell to the survey or interview date. The observed outcome is then $Y := \min\{Y^*, T\}$. Let X be a vector of observed covariates with support $\mathcal{X}$. The quantities of interest to researchers typically include features of the conditional distribution of latent outcomes, $F_{Y^*|X}$, such as comparisons across a treatment and control group.

The standard approach to accounting for censoring in empirical analysis assumes the censoring point, T, is unrelated to latent outcomes conditional on X, or, formally:

Condition 1 *Y^* is independent of T conditional on X.*

This independence condition has been standard in the censoring literature from its introduction in Tobin's (1958) original paper through the most recent literature (Chernozhukov and Hong, 2002; Hong and Tamer, 2003; Chernozhukov et al., 2011; Frandsen, 2015). It is trivially satisfied by the fixed censoring assumptions in, for example, Powell (1986) and Blundell and Powell (2007). In the remainder, therefore, assume T is random.

When Condition 1 is satisfied the (conditional) distribution of latent outcomes, $F_{Y^*|X}$, can be identified from the observed outcomes, censoring indicators, and covariates. In the simplest case where covariates are not required, the distribution is consistently estimated by the Kaplan-Meier estimator:

$$\hat{F}_{Y^*}^{KM}(y) = 1 - \prod_{Y_i \leq y} \left(\frac{n-i}{n-i+1} \right)^{1(Y_i \leq T_i)}, \tag{1}$$

where n is the sample size and the data are assumed to have been sorted on the observed outcome, and ties within uncensored outcomes or censored outcomes are ordered arbitrarily and ties among uncensored and censored outcomes are ordered as if the former precedes the latter. The Kaplan-Meier estimator is the nonparametric maximum likelihood estimate for the latent outcome distribution (Kaplan and Meier, 1958). Under Condition 1 and regularity conditions it is consistent and asymptotically normal uniformly in y (Breslow and Crowley, 1974; Gill, 1983; Stute and Wang, 1993). Efron (1981), Lo and Singh (1986), and Akritas (1986) show conditions under which bootstrap inference is valid for the Kaplan-Meier estimator. The paper is widely used in biostatistics, and has found many applications in economic settings, especially regarding unemployment durations (Abbring and Berg, 2005; Sant'Anna, 2014; Frandsen, 2015).

Covariates can be incorporated by constructing Kaplan-Meier estimators within each stratum defined by the covariates, as in Amato (1988) and Cupples et al. (1995):

$$\hat{F}_{Y^*}^{KM}(y|x) = 1 - \prod_{Y_i \le y, X = x} \left(\frac{n(x) - i}{n(x) - i + 1} \right)^{1(Y_i \le T_i)}, \tag{2}$$

where $n(x)$ is the number of observations with $X = x$. If the covariates have continuous components or high dimension, X may be grouped into discrete categories.

The censoring point independence assumption underpinning the Kaplan-Meier and related estimators is not innocuous. It is trivially satisfied when the censoring point is fixed, but otherwise it deserves scrutiny. For example, in a duration or survival setting, the condition is violated if individuals are lost to follow-up due to factors that are correlated with durations or survival times, or in field settings where the difficulty in tracking subjects may induce correlation between follow-up times and unobserved determinants of the outcome (Finkelstein et al., 2012, for example).

Violations of censoring independence, or endogenous censoring, can severely bias standard estimators. Further, under endogenous censoring the parameters of interest are in general (point) unidentified, although under some conditions set identification can still be achieved (Khan and Tamer, 2009).

Unfortunately, without further restrictions censoring point independence cannot be tested when only the outcome, a censoring indicator, and covariates are observed. Tsiatis (1975) showed that in this case one can always construct a distribution of latent outcomes that reconciles the censoring point independence assumption with the observed data.

3 Censoring point independence test

The censoring point independence condition can be tested when the observed data include the censoring point for each observation (censored and uncensored) in addition to the outcome and covariates. In this case Condition 1 imposes testable restrictions on the observed data. It implies that the conditional cdf of latent outcomes is equal to the conditional cdf of realized outcomes, conditional on the censoring point exceeding the evaluation point of the cdf, as the following theorem establishes.

Theorem 2 (Conditional cdf equivalence) *Suppose Condition 1 holds. Then for all $y < \bar{t}(x) := \sup\{\mathcal{T}(x)\}$ and $x \in \mathcal{X}$ the following holds: $F_{Y|X,T>y}(y|x) = F_{Y^*|X}(y|x)$, where $F_{Y^*|X}(y|x)$ and $F_{Y|X,T>y}(y|x)$ are the conditional cdfs of latent and observed outcomes, respectively.*

Proof. All proofs are collected in the Appendix. ■

3.1 Test Statistic

Theorem 2 suggests the following alternative to the well-known Kaplan-Meier estimator for identifying the latent outcome conditional cdf, $F_{Y^*|X}$:

$$\hat{F}^{RS}(y|x) = \frac{\sum_{i=1}^{n} 1(Y_i \leq y)\, 1(T_i > y, X_i = x)}{\sum_{i=1}^{n} 1(T_i > y, X_i = x)}. \tag{3}$$

This "reduced-sample" estimator is the empirical analog to the conditional cdf in Theorem 2; it is simply the empirical cdf of observed outcomes in the subsample where $T_i > y$. It is consistent and, conditional on the censoring points, unbiased for the latent outcome cdf, $F_{Y^*|X}(y|x)$, and therefore under Condition 1 converges in probability to the latent outcome cdf.

The censoring point independence condition can be tested by comparing the Kaplan-Meier and reduced-sample estimators. The proposed test uses a Cramer-von Mises-type statistic to quantify how the estimated cdfs differ:

$$\hat{\Delta} = \int_{\mathcal{X}} \int_{y<\bar{t}(x)} \hat{\Delta}(y,x)^2 \, d\hat{F}^{KM}(y|x) \, d\hat{F}_X(x), \tag{4}$$

where $\hat{\Delta}(y,x) := \hat{F}^{KM}(y|x) - \hat{F}^{RS}(y|x)$ is the difference in the estimated conditional cdfs, and $\bar{t}(x)$ is an upper bound whose choice is discussed below.

The test statistic corresponds to the integrated squared distance between the two estimators, and, like the Cramer-von Mises test statistic, is sensitive to differences in location and shape, and thus leads to a test that has power against a wide variety of departures from Condition 1. Other measures of cdf differences, such as the Kolmogorov-Smirnov sup-norm, could also be used for the test statistic, and would likely have similar properties. The test's power relies on the fact that the two estimators differ in their sensitivity to dependence between latent outcomes and censoring points. Under departures from the censoring point independence assumption, the reduced-sample estimator (3) remains a pointwise unbiased (conditional on T) estimator of the conditional cdf of the latent outcome, conditional on $T > y$, while the Kaplan-Meier estimator does not. The test's power derives from this difference. The space over which the integral is taken is bounded from above to ensure that the test statistic has a bounded limiting variance.

3.2 Asymptotic Distribution

The test statistic converges to the integrated square of a Gaussian process, as the following theorem establishes for the case of iid data. The Appendix gives a similar result for non-iid data.

Theorem 3 (Test statistic limiting distribution) *Suppose Condition 1 holds and observed data* $\{Y_i, T_i, X_i\}_{i=1}^n$ *are iid conditional on* X_i*. Then* $\hat{\Delta}$ *converges in distribution to the following:*

$$\sqrt{n}\hat{\Delta} \underset{d}{\to} \psi\left(\mathbb{G}_\Delta(y|x)\right),$$

where ψH is the Hadamard-differentiable map $\psi = \int\int H(y,x)^2 \, dF_{Y^*|x}(y|x) \, dF_X(x)$, and

$$\mathbb{G}_\Delta(y|x) := J_\Delta' \begin{bmatrix} d\phi^{KM}(W(y,x)) \\ d\phi^{RS}(W(y,x)) \end{bmatrix} \cdot \mathbb{G}(y,x),$$

where $J_\Delta = (1,-1)'$, the derivative maps $d\phi^{KM}(W(y,x))$ and $d\phi^{RS}(W(y,x))$ are defined in the proof, and $\mathbb{G}(y,x)$ is a Gaussian process indexed by

$$\mathcal{Y}(x) = \{y : y < \bar{t}(x)\} \subset \mathbb{R}$$

with zero mean function and covariance function

$$C_{\mathbb{G}}(y,\tilde{y}|x) := E\left[W_i(y,x) W_i(\tilde{y},x)'\right] - W(y,x) W(\tilde{y},x)', \tag{5}$$

where

$$W_i(y,x) = \begin{pmatrix} 1(Y_i \le y, X_i = x) \\ 1(Y_i \le y, T_i \ge Y_i, X_i = x) \\ 1(Y_i \le y, T_i > y, X_i = x) \\ 1(T_i \le y, X_i = x) \end{pmatrix}$$

and $W(y,x) := E[W_i(y,x)]$.

3.3 Test Implementation

Theorem 3 suggests a straightforward procedure for testing Condition 1:

1. compute the test statistic $\hat{\Delta}$ via equation (4) by taking a weighted sample average of $\hat{\Delta}(Y_i, X_i)^2$ using the Kaplan-Meier weights in each X_i-cell;

2. find the test statistic's p-value by simulating $\Pr\left(\psi(\mathbb{G}_\Delta(y|x)) > \hat{\Delta}\right)$ using a recentered bootstrap.

The bootstrap simulation procedure, shown in the Appendix to be valid in this setting, consists of the following steps:

1. for $b \in \{1, ..., B\}$ compute the bootstrapped test statistic

$$\hat{\Delta}_b^* = \int_{\mathcal{X}} \int_{y<\bar{t}(x)} \left(\hat{\Delta}_b \left(y, x\right) - \hat{\Delta} \left(y, x\right) \right)^2 d\hat{F}^{KM} \left(y|x\right) d\hat{F}_X \left(x\right), \tag{6}$$

   where $\hat{\Delta}_b (y, x)$ is equivalent to $\hat{\Delta} (y, x)$ defined following equation (4), but computed from the b-th bootstrap sample. Note that the bootstrapped test statistic involves integrating the square of the *recentered* difference $\hat{\Delta}_b (y, x) - \hat{\Delta} (y, x)$. The reason for this is to ensure that the bootstrapped statistic reflects the null hypothesis that $\hat{\Delta}_b^*$ is the square integral of a mean-zero quantity.

2. compute the p-value:

$$p\text{-value} = \frac{1}{B} \sum_{b=1}^{B} 1 \left(\hat{\Delta}_b^* > \hat{\Delta}^* \right).$$

   A level-α test rejects if the p-value is less than or equal to α.

Finally, implementing the test requires choosing a value for $\bar{t}(x)$, the upper bound value of y within each covariate cell over which to integrate the squared difference in the test statistic. Choosing $\bar{t}(x)$ involves a tradeoff: as $\bar{t}(x)$ increases, the maximal variance of the process increases without bound and the test loses power in general; on the other hand, setting $\bar{t}(x)$ to be small can prevent the test from having power against alternatives where latent outcomes Y^* are related to censoring points T when T is large but not when T is small. Which alternatives are most important to maintain power against, and therefore the most appropriate choice of $\bar{t}(x)$, will necessarily depend on the application. Setting $\bar{t}(x)$ near the 75th percentile of the censoring point distribution within each covariate cell is perhaps a reasonable starting point, and is used in the simulations and applications below. Note that in the asymptotic theory and in the simulations $\bar{t}(x)$ is considered a fixed choice that does not vary in repeated samples.

3.4 Test Power

Given Theorem 3, the testing procedure described above will clearly have correct size asymptotically. What of its power? The test's power relies on the Kaplan-Meier estimator (2) and the reduced-sample estimator (3) having different probability limits when Condition 1 fails. This section shows that the estimators do indeed have different limits when the censoring point and the latent outcome are positively (or negatively) dependent. This section suppresses dependence on X to spare notation and assumes continuously distributed outcomes and censoring points for simplicity. The following result shows what the Kaplan-Meier and reduced-sample estimators converge to without imposing Condition 1:

Lemma 4 *Suppose Y^* and Y are continuously distributed. Regardless of whether condition 1 holds, (i) the Kaplan-Meier estimator (1) converges in probability to the following:*

$$\hat{F}^{KM}(a) \underset{p}{\to} S(a) := 1 - \left(1 - F_{Y|T}(a|T > a)\right) \exp\left\{ \int_{-\infty}^{a} \frac{\frac{d}{dr} F_{Y|T}(s|T > r = s)}{1 - F_{Y|T}(s|T > s)} ds \right\},$$

where $F_{Y|T}$ is the conditional cdf of Y given T and $f_{Y|T}$ is the corresponding density; and (ii) the reduced-sample estimator (3) converges in probability to the following:

$$\hat{F}^{RS}(a) \underset{p}{\to} F_{Y|T}(a|T > a).$$

The Lemma shows the probability limit of the Kaplan-Meier estimator when censoring point independence is not imposed. The result may be of interest in its own right, but for the purposes of the proposed test, it shows when Kaplan-Meier will differ in probability from the reduced-sample estimator, and therefore when the test will have power. Specifically, it shows that the test will have power against alternatives under which Y^* and T are positively (or negatively) dependent in the following sense::

Condition 5 (Positive (or negative) dependence) $\Pr(Y^* \le y|T \ge t, X)$ *is almost surely nonincreasing (or nondecreasing) in t along the path $\{(y,t) : y = t\} \subset supp(Y^*) \times \mathcal{T}$ and is* *strictly decreasing (or strictly increasing) over a subset of that path with positive F_Y-measure.*

Condition (5) means that conditional on X, when the censoring point exceeds a larger value, the latent outcome distribution stochastically shifts upward (or downward) over a measurable region. Lehmann (1966) described this notion and showed it implies positive (or negative) correlation, and is itself implied by positive (or negative) regression dependence.

As the following theorem formalizes, when Y^* and T are dependent in the sense of Condition 5, then the two estimators have different probability limits, and the test has power:

Theorem 6 *Suppose Condition 5 holds and Y^* and Y are continuously distributed. Then the Kaplan-Meier estimator (2) and Reduced-sample estimator (3) have different probability limits.*

Against fixed alternatives of positive or negative dependence between Y^* and T, then, the test will have asymptotic power of one. Consider a sequence of local alternatives converging to the null hypothesis of the form

$$S\left(a\right) - F_{Y|T}\left(a|T > a\right) = \gamma n^{-1/2}.$$

Since, by Theorem 3, each component of the difference which defines the test statistic $\hat{\Delta}$ converges at rate $n^{1/2}$, the test statistic's integrand converges in distribution to the following non-centered Gaussian process:

$$\sqrt{n}\hat{\Delta}\left(y|x\right) \underset{d}{\rightarrow} \mathbb{G}_{\Delta}\left(y|x\right) + \gamma,$$

where $\mathbb{G}_{\Delta}\left(y|x\right)$ is the Gaussian process that appears in the null distribution in Theorem 3. The limiting distribution is thus shifted by γ and the test will have nontrivial asymptotic power against local alternatives, a feature in common with other Cramer-von Mises tests (Neuhaus, 1976).

3.5 Implications if the Test Rejects

A rejection constitutes evidence against the censoring point independence condition, and implies that estimates based on standard censored regression procedures, such as Tobit or censored quantile regression, are biased and inconsistent. A rejection also implies the Kaplan-Meier estimator is inconsistent for the distribution of latent outcomes; in fact, the latent outcome distribution loses nonparametric point identification in general without censoring point conditional independence, although it may still be partially identified (Peterson, 1976; Abbring and Berg, 2005; Khan and Tamer, 2009). The test may therefore be used to motivate partial identification analysis.

The test not only indicates violations of the censoring point independence assumption, but also indicates the direction of the violation and thereby provides a basis for determining the direction of the bias of traditional methods based on censoring point independence, and can be used to provide bounds on the parameter of interest. Lemma 4 implies that the difference between the probability limits of the Kaplan-Meier estimator and that of the reduced-sample estimator is

$$\left(1 - F_{Y|T}\left(a|T > a\right)\right)\left(1 - \exp\left\{\int_{-\infty}^{a} \frac{\frac{d}{dr}F_{Y|T}\left(s|T > r = s\right)}{1 - F_{Y|T}\left(s|T > s\right)}ds\right\}\right).$$

The left-hand factor in the display in nonnegative; the right-hand factor's sign is determined by the dependence between latent outcomes and censoring points. Under independence it is naturally zero; under positive dependence in the sense of Condition 5 the sign is positive, since the exponential term will be less than one; and under negative dependence the sign is negative, since the exponential term will be greater than one. Thus finding that the Kaplan-Meier estimator of the cdf is greater than the reduced-sample estimator is consistent with positive correlation between Y^* and T, implying that upper-tail estimates of quantiles of latent outcomes based on, say, the reduced sample estimator will tend to be upward biased, and regression coefficients biased away from zero. Estimates in this case can be interpreted as upper bounds for the magnitude of the effect. If instead the Kaplan-Meier cdf estimator is less than the reduced-sample estimator, potential outcomes and T are instead likely negatively correlated, upper-tail estimates of latent outcomes will be downward biased, and regression coefficients will be biased toward zero. In this case estimates can be interpreted as lower bounds for the magnitude of the effect.

4 Simulations

Results in the previous section imply that the test will have correct size and good power for sufficiently large samples, but what of its finite-sample performance? Monte Carlo simulations show that the test has accurate size and good power properties over a wide range of sample sizes, censoring severity, and correlation between the censoring points and potential outcomes. The main set of simulations are based on a log-normal data generating process where latent outcomes and censoring points are jointly log normal:

$$
\begin{aligned}
Y^* &= \exp(Z_1) \\
T &= \exp(\mu_T + \sigma_T Z_2) \\
\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).
\end{aligned}
$$

The parameter μ_T captures censoring severity: the probability of censoring increases as μ_T decreases according to $1 - E[\Phi(\mu_T + \sigma_T Z_2)]$. The parameter σ_T captures the dispersion in censoring points. The parameter ρ captures correlation between potential outcomes Y^* and the censoring points T. The special case $\rho = 0$ corresponds to the independent censoring points condition 1. The simulations are based on 1000 replications of samples of size n drawn from this class of distributions. Unless stated otherwise, all simulation scenarios set $n = 500$, $\sigma_T = .5$, $\mu_T = 1$ and $\rho = 0$.

One set of simulations examines the test's performance when the support of the censoring variable is bounded, as may happen frequently in practice. In these simulations Y^* is generated as above, but the censoring variable is distributed independently of Y^* as $T \sim U(0, \tau)$ for $\tau \in [1, \exp(3)]$.

The first set of simulations examines the finite-sample size of the test for sample sizes ranging from $n = 100$ to $n = 1000$, with parameters $\mu_T = 1$, $\sigma_T = .5$, and $\rho = 0$. The simulations show that even for relatively small sample sizes the test maintains accurate size. Figure 1 plots the simulated rejection rate for tests of nominal size $\alpha = .05$ by sample size. The rejection rate hovers around five percent over the entire range.

The next set of simulations explores the test's performance when censoring becomes more severe. The simulations vary the probability of censoring from .1 to .9, corresponding to $\mu_T = 1.5$ to $\mu_T = -1.5$. The simulations show that even for severe censoring the test's size is very close to nominal. Figure 2 plots the simulated rejection rate for tests of nominal size $\alpha = .05$ by the censoring probability. The rejection rate is near five percent even for scenarios with a high censoring probability.

The next set of simulations assesses the test's size across differing degrees of variation in censoring points. The simulations vary the censoring point distribution's scale parameter, σ_T, from 0.1 to 2. The simulations show that the test maintains correct size across the entire range of censoring point variation. Figure 3 plots the simulated rejection rate for tests of nominal size $\alpha = .05$ by σ_T. The rejection rate is around five percent throughout the whole range.

The next set of simulations illustrates the test's performance when the censoring variable's support is bounded, as may occur frequently in practice. The simulations vary the upper bound of the censoring variable's support from one to $\exp(3) \approx 20.1$. The simulations show that the test's size is close to the nominal level over all ranges of censoring support tested. Figure 4 plots the simulated rejection rate for tests of nominal level $\alpha = .05$ as a function of τ, the upper bound of the censoring variable's support. The rejection rate is near five percent over the whole range.

The final set of simulations examines the test's power to detect correlation between censoring points and latent outcomes. The simulations vary the correlation parameter ρ from $-.9$ to $.9$. The simulations show that the rejection probability increases rapidly to 100 percent as the degree of correlation between latent outcomes and censoring points increases. Figure 5 plots the simulated rejection rate for tests of nominal size $\alpha = .05$ by ρ for sample size $n = 500$. At the center of the plot, where $\rho = 0$, the rejection rate is approximately .05, indicating correct size. As correlation is introduced, however, the rejection rate increases steeply. The test rejects over 80 percent of the time when $\rho = .2$ and essentially 100 percent of the time for ρ greater than .3 or so in absolute value.

5 Empirical Example: Unemployment Spells

This section applies the censoring point independence test to unemployment spells, and finds evidence that in commonly used datasets censoring points are not independent of latent spell durations. The data are gathered from the three most commonly used datasets in articles on unemployment spells published in prominent empirical research journals in the last ten years: the NLSY, the PSID, and the SIPP.

Unemployment spells offer an attractive setting for illustrating the censoring point independence test because (1) unemployment spell duration data are typically censored; (2) analysis in this setting—whether of the parametric variety (e.g., tobit) or nonparametric (e.g., proportional hazard models, censored quantile regression)—relies heavily on the independence assumption; and (3) censoring points here are typically observed for all units, making the test applicable. Unemployment spell observations are censored when a spell is ongoing as of the final date an individual is observed. In this setting the censoring point corresponds to the elapsed time from the beginning of the spell to the last date an individual is observed. The independence assumption fails if unemployment spell duration is related to survey dropout, or if the distribution of latent durations changes over time within the sample period.

5.1 Unemployment Durations in the PSID

The first example analyzes unemployment spells among working-age males in the PSID, a common dataset for studying employment dynamics. The sample is based on Low et al.'s (2010) study of consumption and labor supply dynamics, and includes males aged 22 to 61 for whom labor market status was observed. For the purposes of this example, unemployment is defined as out of the labor force or out of work but looking for a job, as these are not always distinguished in the NLSY. Respondents reported their labor market status monthly. Unemployment spells as measured in the PSID sample are censored for two reasons: spells are interrupted by retirement and spells are interrupted by the end of the sampling period. Consistent estimation of the distribution of unemployment spell durations requires that the censoring points be independent of latent durations.

The test applied to the PSID data on unemployment spells strongly rejects that censoring points are independent of spell durations. Figure 6 plots the cdf of unemployment spell durations estimated using the Kaplan-Meier estimator and the reduced sample estimator. The two cdfs coincide at short durations, but noticeably differ at longer durations. The p-value corresponding to the integrated squared difference is zero to three decimal places, strongly rejecting the null hypothesis of independence.

In this example, the Kaplan-Meier cdf estimator stochastically dominates the reduced-sample estimator, suggesting that durations are negatively correlated with censoring points, biasing the reduced-sample cdf estimates upward. A possible explanation is that individuals who become unemployed later in their career have longer latent unemployment spells—possibly because of higher reservation wages or depreciated human capital—but shorter censoring points because of looming retirement or the end of the study period. The result is that the analysis will tend to underestimate the degree to which unemployment spell lengths increase with age. The bias is potentially large: although the differences in the estimated cdfs in Figure 6 do not appear large, they correspond to dramatic differences in the upper quantiles of unemployment spell durations: the reduced-sample estimate of the 95th percentile of unemployment spell lengths is 41 months, while it is 65 months for the Kaplan-Meier estimator. Given the skewed nature of unemployment durations, these differences can cause large biases in estimated means.

5.2 Unemployment Durations in the NLSY

The next example draws from the NLSY79, a dataset used to study unemployment spells in Paserman (2008) and Engelhardt (2010), among others. Respondents to the NLSY reported their labor market status weekly. For this example, individuals are considered unemployed if they are not working, including out of the labor force. The sample includes black males whose employment status is observed from 1989 to 1993. This is the period used in Engelhardt (2010), and is included in the period studied by Paserman (2008). An unemployment spell duration is defined as the number of weeks elapsed from a job separation to the start of another job. Durations are censored when they are ongoing at the end of the study period. Thus, the censoring point can be defined for all spells (even uncensored ones) as the time from a job separation to the end of the study period (December 1993).

The testing procedure rejects the censoring point independence assumption. Figure 7 plots the cdf of unemployment spell durations estimated using the Kaplan-Meier estimator and the reduced-sample estimator. The two estimates give nearly identical results over much of the support of unemployment durations, but diverge significantly for the longest spells. The p-value on the difference is 0.005, providing evidence of dependence between latent outcomes and censoring points. The estimates suggest, however, that estimates of the distribution of unemployment spell duration based on the Kaplan-Meier or reduced-sample estimator are likely to be reliable for all but the highest quantiles.

5.3 Unemployment Durations in the SIPP

The final example analyzes unemployment spells in the 1996 SIPP. This dataset has been used frequently to study unemployment, including by Dey and Flinn (2005), whose sample is used here. The sample restricts to white males who were between the ages of 25 and 54 at the start of the study period, and who have at least a high school education. Unemployment is defined as being without a job all month. Unemployment spells are censored in this data if they are ongoing at the end of the sample period, but the censoring point (the number of months from the start of a spell to the end of the sample period) is defined for all spells, censored or not. This is therefore an appropriate setting for the censoring point independence test.

The censoring point independence test strongly rejects that censoring points are independent of latent spell durations. Figure 8 plots the two estimates of the unemployment spell duration cdf, Kaplan-Meier and reduced-sample. Beyond about 5 months, the two estimated cdfs differ appreciably, and the p-value of the difference is zero to the reported precision of three decimal places. As in the PSID, the difference in the cdfs appears modest, but it corresponds to large differences in the estimated quantiles of unemployment spell durations, especially in the upper tails. These differences suggest that estimates which require that latent spell durations are independent of censoring points are likely to be biased and inconsistent.

6 Conclusion

This paper proposed a test for the censoring point independence condition that underpins censored regression and duration methods. The test requires no auxiliary assumptions nor extra information beyond the observed (censored) outcome and censoring points. The testing procedure compares two different nonparametric estimators for the latent outcome cdf: the Kaplan-Meier estimator and the so-called reduced-sample estimator, both of which are consistent under censoring independence, but in general have different probability limits otherwise. Simulations show the test has good size and power properties in finite samples.

The test proved important in empirical examples based on unemployment duration data from recent articles in prominent empirical research journals. Censored unemployment spell data in all cases revealed strong evidence of dependence between censoring points and latent unemployment spell lengths. These examples underscore the need for researchers to include tests for censoring point independence as a standard part of duration or other censored outcome analysis.

The test relies on censoring points being observed even for uncensored outcomes. Many duration settings satisfy this requirement—for example, when censoring is due to the end of the study period—and it is a pre-requisite for most censored quantile methods in any case. It does rule out settings where only outcomes and indicators for the presence of censoring are observed, such as some competing risk settings. It also rules out settings with two-sided censoring, a scenario to which the Kaplan-Meier estimator is not easily adapted. Nevertheless, the test applies in a wide range of empirical settings, and the examples examined here suggest the test may be a useful addition to the analysis of durations and other censored outcomes.

References

Jaap Abbring and Gerard Van Den Berg. Social experiments and instrumental variables with duration outcomes. IFS Working Papers W05/19, Institute for Fiscal Studies, October 2005.

Michael G. Akritas. Bootstrapping the kaplanmeier estimator. *Journal of the American Statistical Association*, 81(396):1032–1038, 1986.

David A. Amato. A generalized kaplan-meier estimator for heterogenous populations. *Communications in Statistics - Theory and Methods*, 17(1):263–286, 1988.

Miguel A. Arcones. Weak convergence for the row sums of a triangular array of empirical processes under bracketing conditions. *Stochastic Processes and their Applications*, 73(2): 195 – 231, 1998.

Richard Blundell and James L Powell. Censored regression quantiles with endogenous regressors. *Journal of Econometrics*, 141(1):65–83, 2007.

N. Breslow and J. Crowley. A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.*, 2(3):437–453, 05 1974.

Victor Chernozhukov and Han Hong. Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, 97(459):pp. 872–882, 2002.

Victor Chernozhukov, Iván Fernández-Val, and Amanda Kowalski. Quantile regression with censoring and endogeneity. unpublished manuscript, April 2011.

L. Adrienne Cupples, David R. Gagnon, Ratna Ramaswamy, and Ralph B. D'Agostino. Age-adjusted survival curves with application in the framingham study. *Statistics in Medicine*, 14(16):1731–1744, 1995.

Matthew S. Dey and Christopher J. Flinn. An equilibrium model of health insurance provision and wage determination. *Econometrica*, 73(2):pp. 571–627, 2005.

Bradley Efron. Censored data and the bootstrap. *Journal of the American Statistical Association*, 76(374):312–319, 1981.

Bryan Engelhardt. The effect of employment frictions on crime. *Journal of Labor Economics*, 28(3):pp. 677–718, 2010.

Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. The oregon health insurance experiment: Evidence from the first year*. *The Quarterly Journal of Economics*, 127(3):1057–1106, 2012.

Brigham R. Frandsen. Treatment effects with censoring and endogeneity. *Journal of the American Statistical Association*, 110(512):1745–1752, 2015.

Richard Gill. Large sample behaviour of the product-limit estimator on the whole line. *Ann. Statist.*, 11(1):49–58, 03 1983.

Richard D Gill. Lectures on survival analysis. In *Lectures on Probability Theory*, pages 115–241. Springer, 1994.

Han Hong and Elie Tamer. Inference in censored models with endogenous regressors. *Econometrica*, 71(3):905–932, 2003.

Xuelin Huang, Robert A Wolfe, and Chengcheng Hu. A test for informative censoring in clustered survival data. *Statistics in medicine*, 23(13):2089–2107, 2004.

E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):pp. 457–481, 1958.

Shakeeb Khan and Elie Tamer. Inference on endogenously censored regression models using conditional moment inequalities. *Journal of Econometrics*, 152(2):104–119, 2009.

Seung-Yeoun Lee and Robert A Wolfe. A simple test for independent censoring under the proportional hazards model. *Biometrics*, pages 1176–1182, 1998.

E. L. Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, 37 (5):pp. 1137–1153, 1966.

Shaw-Hwa Lo and Kesar Singh. The product-limit estimator and the bootstrap: Some asymptotic representations. *Probability Theory and Related Fields*, 71(3):455–465, 1986.

Hamish Low, Costas Meghir, and Luigi Pistaferri. Wage risk and employment risk over the life cycle. *The American Economic Review*, 100(4):pp. 1432–1467, 2010.

Georg Neuhaus. Asymptotic power properties of the cramr-von mises test under contiguous alternatives. *Journal of Multivariate Analysis*, 6(1):95 – 110, 1976.

M. Daniele Paserman. Job search and hyperbolic discounting: Structural estimation and policy evaluation. *The Economic Journal*, 118(531):pp. 1418–1452, 2008.

Arthur V. Peterson. Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Sciences*, 73(1):11–13, 1976.

James L. Powell. Censored regression quantiles. *Journal of Econometrics*, 32(1):143 – 155, 1986.

Pedro H. C. Sant'Anna. Nonparametric tests for conditional treatment effects in duration outcomes. 2014.

W. L. Stevens. The truncated normal distribution. *Annals of Applied Biology*, 24(4):847–850, 1937.

W. Stute and J.-L. Wang. The strong law under random censorship. *Ann. Statist.*, 21(3): 1591–1607, 09 1993.

James Tobin. Estimation of relationships for limited dependent variables. *Econometrica*, 26 (1):pp. 24–36, 1958.

Anastasios Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72(1):20–22, 1975.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, New York, 1998.

A.W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer-Verlag, New York, 1996.

Jon A. Wellner. A Glivenki-Cantelli theorem for empirical measures of independent but non-identically distributed random variables. 11:309–312, 1981.

Appendix

Proofs

Proof of Theorem 2. Since conditional on $T > y$, the event $\{Y \leq y\}$ is equivalent to $\{Y^* \leq y\}$, we have $\Pr(Y \leq y|X = x, T > y) = \Pr(Y^* \leq y|X = x, T > y)$. By Condition 1, $\Pr(Y^* \leq y|X = x, T > y) = \Pr(Y^* \leq y|X = x)$, which, by the definition of a cumulative distribution function is $F_{Y^*|X}(y|x)$. ■

Proof of Theorem 3. The cdf difference of which the test statistic $\hat{\Delta}$ is the square integral, $\hat{\Delta}(y, x)$, is a Hadamard differentiable function of $\hat{F}^{KM}(y|x)$ and $\hat{F}^{RS}(y|x)$ with Jacobian $J_\Delta = (1, -1)'$. $\hat{F}^{KM}(y|x)$ and $\hat{F}^{RS}(y|x)$, in turn, are each Hadamard-differentiable functions of $\bar{W}_n := n^{-1}\sum_{i=1}^{n} W_i(y, x)$. By Glivenko-Cantelli $\bar{W}_n$ converges in probability uniformly in $y \in \mathcal{Y}(x)$ for each x in the (finite) support of X:

$$\bar{W}_n(y, x) \underset{p}{\rightarrow} W(y, x) = \begin{pmatrix} F_{YX}(y, x) \\ F^1(y, x) \\ p(y, x) \\ F_{TX}(y, x) \end{pmatrix},$$

where $F_{YX}(y, x) := \Pr(Y \leq y, X = x)$, $F^1(y, x) := \Pr(Y \leq y, T \geq Y, X = x)$, $p(y, x) := \Pr(Y \leq y, T \geq y, X = x)$, and $F_{TX}(y, x) := \Pr(T \leq y, X = x)$. By the Donsker theorem $\hat{G}(y, x) := \sqrt{n}\left(\bar{W}_n(y, x) - W(y, x)\right)$ converges to a mean-zero Gaussian process indexed by $y \in \mathcal{Y}(x)$:

$$\hat{G}(y, x) \underset{d}{\rightarrow} \mathbb{G}(y, x)$$

with covariance function

$$C_{\mathbb{G}}(y, \tilde{y}|x) := E\left[W_i(y, x) W_i(\tilde{y}, x)'\right] - W(y, x) W(\tilde{y}, x)'.$$

The (vector of) estimators $\begin{pmatrix} \hat{F}^{KM}(y|x) & \hat{F}^{RS}(y|x) \end{pmatrix}'$ is a Hadamard differentiable function of $\bar{W}_n$: $\hat{F}^{RS}$ via the map

$$\phi^{RS}\left(\bar{W}_n\right) = \frac{\bar{W}_{n,[3]}}{\bar{W}_{n,[4]}},$$

and, shown by Gill (1994), $\hat{F}^{KM}$ via the map

$$\phi^{KM}\left(\bar{W}_n\right) = 1 - \prod\left(1 - \frac{d\bar{W}_{n,[2]}}{1-\bar{W}_{n,[1]}}\right),$$

where Π denotes product integration over intervals $(-\infty, y]$ and the subscript brackets denote the elements of $\bar{W}_n$. The map ϕ^{RS} trivially has a Hadamard derivative whose probability limit applied to applied to $\hat{G}(y,x)$ is

$$d\phi^{RS}(W)\cdot\hat{G} = J'_{RS}\hat{G},$$

where the Jacobian $J_{RS}(y,x) := (1-F_{TX}(y,x))^{-1}\begin{bmatrix} 0 & 0 & 1 & F_{Y^*X}(y,x) \end{bmatrix}^{\top}$. Gill (1994) showed the map ϕ^{KM} has a Hadamard derivative whose probability limit applied to $\hat{G}(y,x)$ is

$$d\phi^{KM}(W)\cdot\hat{G} = \left(1-F_{Y^*|X}\right)\int\frac{d\hat{G}_{[2]}+\hat{G}_{[1]-}d\Lambda}{(1-F_{YX-})(1-\Delta\Lambda)},$$

where $\Lambda(y,x)$ is the conditional hazard function of Y^* given $X=x$, $\Delta\Lambda$ denotes the atoms (if any) of Λ, and subscript minus signs denote left limits. By the functional delta method (van der Vaart and Wellner, 1996, Theorem 3.9.4), therefore, $\sqrt{n}\hat{\Delta}(y,x)$ weakly converges to the following zero-mean Gaussian process in $\ell^{\infty}(\mathcal{Y}(x))$:

$$\sqrt{n}\hat{\Delta}(y,x)\underset{d}{\rightarrow} J'_{\Delta}\begin{bmatrix} d\phi^{KM}(W(y,x)) \\ d\phi^{RS}(W(y,x)) \end{bmatrix}\cdot\mathbb{G}(y,x).$$

The result then follows from the continuous mapping theorem. ■

Proof of Lemma 4. Take part (i) first. Stute and Wang (1993, equation 2.5) show that the probability limit of the Kaplan-Meier estimator, without imposing independence, is

$$S(a)=\int_{-\infty}^{a} m(y')\exp\left\{\int_{-\infty}^{y'}\frac{1-m(y)}{1-F_Y(y)}f_Y(y)\,dy\right\}f_Y(y')\,dy',$$

where $m(y) := \Pr(T < Y|Y = y)$. Denoting the joint pdf of Y^* and T as $f(y,t)$, we can write $m(y)$ as $\int_y^\infty f(y,t)\,dt/f_Y(y)$ and $1-F_Y(s)$ as $\int_s^\infty\int_s^\infty f(y,t)\,dtdy$. Making these substitutions we obtain

$$S(a)=\int_{-\infty}^{a}\int_{y'}^{\infty} f(y',t)\,dt\exp\left\{\int_{-\infty}^{y'}\frac{\int_s^\infty f(y,s)\,dy}{\int_s^\infty\int_s^\infty f(y,t)\,dtdy}ds\right\}dy'.$$

Repeated application of Leibniz' rule and integration by parts yield

$$\begin{aligned}S(a) &= -\int_{-\infty}^{a}\frac{d}{dy'}\exp\left\{-\int_{-\infty}^{y'}\frac{\int_s^\infty f(s,t)\,dt}{\int_s^\infty\int_s^\infty f(y,t)\,dtdy}ds\right\}dy'\\ &= 1-\exp\left\{\int_{-\infty}^{a}-\frac{f_{Y|T}(s|T>s)}{1-F_{Y|T}(s|T>s)}ds\right\}.\end{aligned}$$

Note that by the chain rule

$$\frac{d}{ds}\ln\left(1-F_{Y|T}(s|T>s)\right)=-\frac{f_{Y|T}(s|T>s)}{1-F_{Y|T}(s|T>s)}-\frac{\frac{\partial}{\partial r}F_{Y|T}(s|T>r=s)}{1-F_{Y|T}(s|T>s)},$$

so

$$-\frac{f_{Y|T}(s|T>s)}{1-F_{Y|T}(s|T>s)}=\frac{d}{ds}\ln\left(1-F_{Y|T}(s|T>s)\right)+\frac{\frac{\partial}{\partial r}F_{Y|T}(s|T>r=s)}{1-F_{Y|T}(s|T>s)}.$$

Substituting this in and simplifying, we get

$$\begin{aligned}S(a) &= 1-\exp\left\{\int_{-\infty}^{a}\frac{d}{ds}\ln\left(1-F_{Y|T}(s|T>s)\right)ds\right\}\exp\left\{\int_{s=-\infty}^{a}\frac{\frac{\partial}{\partial r}F_{Y|T}(s|T>r=s)}{1-F_{Y|T}(s|T>s)}ds\right\}\\ &= 1-\left(1-F_{Y|T}(a|T>a)\right)\exp\left\{\int_{-\infty}^{a}\frac{\frac{\partial}{\partial r}F_{Y|T}(s|T>r=s)}{1-F_{Y|T}(s|T>s)}ds\right\},\end{aligned}$$

where the second line follows by the fundamental theorem of calculus. Part (ii) follows from the law of large numbers and Bayes' rule. ■

Proof of Theorem 6. From Lemma 4, the probability limits are equal if and only if

$$\int_{-\infty}^{a}\frac{\frac{\partial}{\partial r}F_{Y|T}\left(s|T>r=s\right)}{1-F_{Y|T}\left(s|T>s\right)}ds=0.$$

The denominator is nonnegative along the entire integral path, and by Condition 5 the numerator is nonpositive (or nonnegative) along the entire integral path, and strictly negative (or positive) along a measurable portion of the path. The left-hand side of the display is therefore strictly negative (or positive) and the probability limits are not equal. ■

Test with Non-iid Data and Bootstrap Validity

The main text develops the test in the context of iid samples. The test may be adapted with minimal modification for many non-iid sampling settings. This appendix derives the limiting distribution of the test statistic in a clustered sampling setting and also establishes the validity of the empirical bootstrap for estimating the limiting distribution of the test statistic.

Suppose that observed data consist of observations on (Y_{ij}, T_{ij}), where $i = 1 \ldots n$ indexes clusters (perhaps individuals in a panel setting) and $j = 1 \ldots n_i$ indexes obsevations within each cluster i (perhaps time-series observations on individual i). Observations are independent (but not necessarily identically distributed) across clusters, but arbitrarily correlated within clusters. Similar to the main text, define the vector of indicators

$$W_{ij}\left(y,x\right)=\begin{pmatrix}1\left(Y_{ij}\leq y,X_{ij}=x\right)\\1\left(Y_{ij}\leq y,T_{ij}\geq Y_{i},X_{ij}=x\right)\\1\left(Y_{ij}\leq y,T_{ij}>y,X_{ij}=x\right)\\1\left(T_{ij}\leq y,X_{ij}=x\right)\end{pmatrix},$$

for $y \in \mathcal{Y}(x)$. Note that because these indicators are bounded, they automatically satisfy a Lyapounov bounded absolute third moment condition uniformly in $y \in \mathcal{Y}(x)$. Assume in addition:

Assumption 7 *The matrix* $\Omega(y, \tilde{y}, x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} \left(E\left[W_{ik}(y, x) W_{ij}(\tilde{y}, x)'\right]\right.$ $\left.- E\left[W_{ik}(y, x)\right] E\left[W_{ij}(\tilde{y}, x)\right]'\right)$ *exists.*

Assumption 7 will be satisfied if the cluster sizes n_i are fixed as the number of clusters increases. This is analogous to the fixed T, large n asymptotics standard in panel data analysis.

Theorem 8 *Suppose Assumption 7 holds. Then (i)* $\bar{W}_n(y, x) \underset{p}{\to} E\left[W_{ij}(y, x)\right]$*; and (ii)*

$$\sqrt{n}\left(\bar{W}_n(y, x) - E\left[W_{ij}(y, x)\right]\right) \underset{d}{\to} \mathbb{G}(y, x),$$

where $\mathbb{G}(y, x)$ *has mean zero and covariance function*

$$C_{\mathbb{G}}^{inid}(y, \tilde{y}, x) = \Omega(y, \tilde{y}, x).$$

Proof. Part (i) follows from Wellner's (1981) Glivenko-Cantelli theorem for independent but non-identically distributed random variables. Part (ii) follows from Arcones's (1998) Donsker theorem for triangular arrays. ■

The limiting distribution of the test statistic (and its derivation) is now identical to that given in Theorem 3, but using covariance function $C_{\mathbb{G}}^{inid}$ instead of $C_{\mathbb{G}}$.

The following Corollary to Theorem 3 establishes that the empirical bootstrap is valid for estimating the distribution of the test statistic.

Corollary 9 *The bootstrapped test statistic (6) converges in distribution to the limiting distribution of the original sample test statistic (4).*

Proof. Note that the vector of indicators $W_i(y, x)$ (being bounded) satisfies the Lindeberg condition, and therefore by van der Vaart (1998, Theorem 23.4), conditional on the original sample observations the bootstrap vector of sample means $\bar{W}_n^*(y,x)$ converges in distribution to the same centered Gaussian process as $\bar{W}_n(y,x)$. The test statistic being a Hadamard differential functional of $\bar{W}_n^*(y,x)$ (with Hadamard derivative shown in Theorem 3), the result follows from the bootstrap functional delta method of van der Vaart (1998, Theorem 23.5). ■

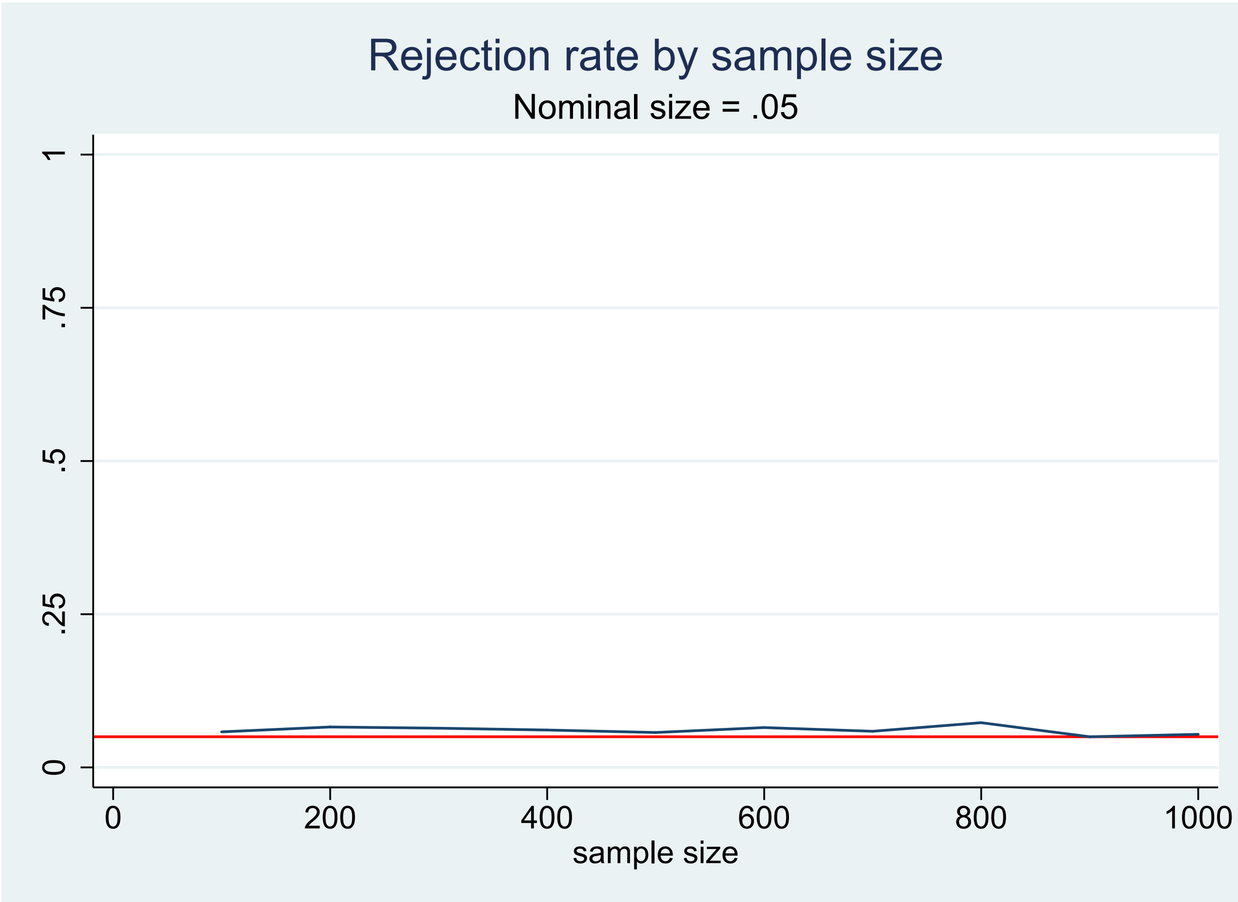

Figure 1: Monte Carlo simulation rejection rates from the censoring independence test as a function of the sample size (x-axis). The latent outcome was generated as a log-normal with parameters $\mu = 0$ and $\sigma = 1$. The censoring point was generated as a log-normal independent of the latent outcome with parameters $\mu_T = 1$ and $\sigma_T = .5$. The nominal size of the tests is .05. Based on 1000 iterations.

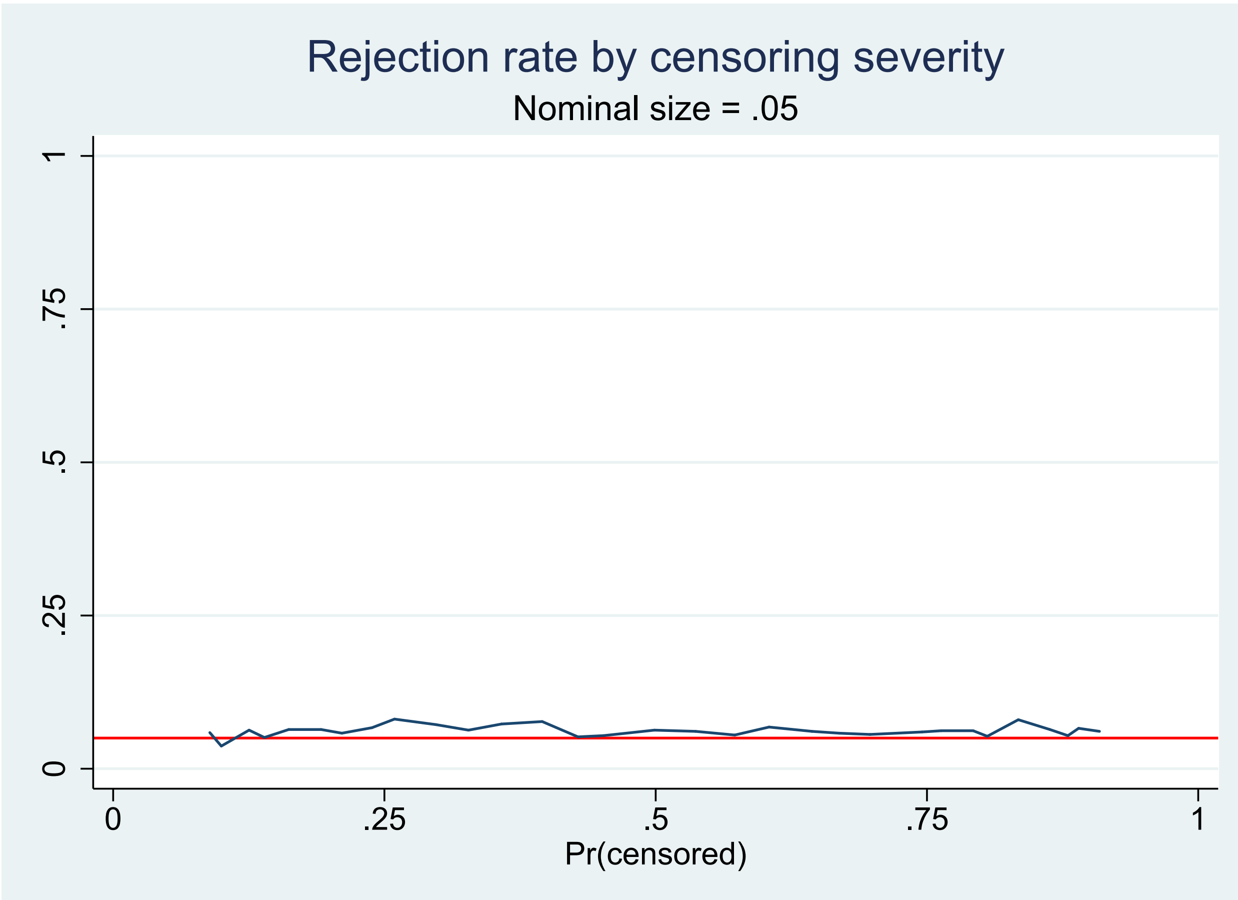

Figure 2: Monte Carlo simulation rejection rates from the censoring independence test as a function of the probability of censoring (x-axis). The latent outcome was generated as a log-normal with parameters $\mu = 0$ and $\sigma = 1$. The censoring point was generated as a log-normal independent of the latent outcome with parameters $\sigma_T = .5$ and μ_T ranging from -1.5 to 1.5, corresponding to censoring probabilities ranging from 0.1 to 0.9, with a sample size of $n = 500$. The nominal size of the tests is .05. Based on 1000 iterations.

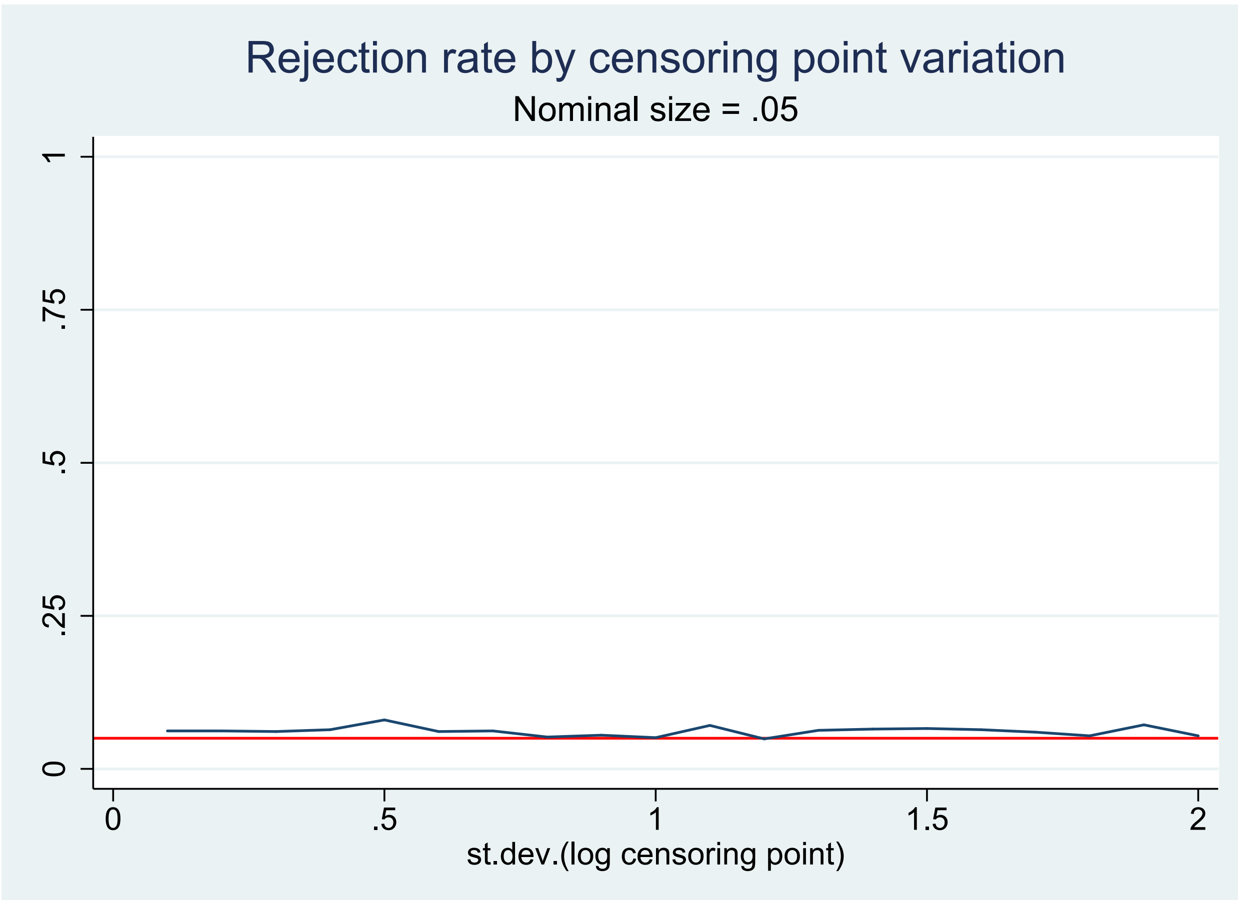

Figure 3: Monte Carlo simulation rejection rates from the censoring independence test as a function of the standard deviation of the censoring point (x-axis). The latent outcome was generated as a log-normal with parameters $\mu = 0$ and $\sigma = 1$. The censoring point was generated as a log-normal independent of the latent outcome with parameters $\mu_T = 1$ and σ_T ranging from 0.1 to 2, with a sample size of $n = 500$. The nominal size of the tests is .05. Based on 1000 iterations.

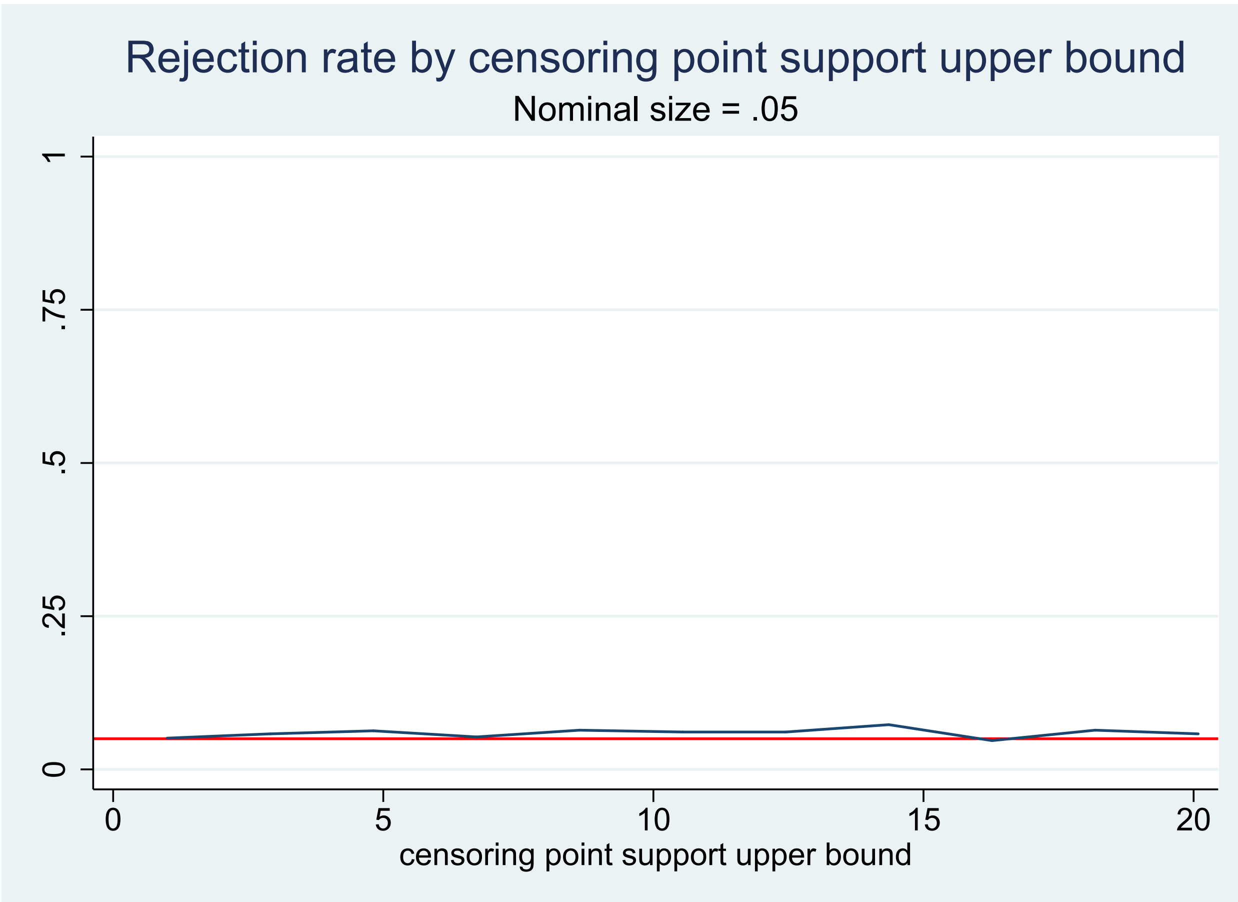

Figure 4: Monte Carlo simulation rejection rates from the censoring independence test as a function of the upper bound of the censoring point support (x-axis). The latent outcome was generated as a log-normal with parameters $\mu = 0$ and $\sigma = 1$. The censoring point was generated as a uniform from zero to the support upper bound indicated on the x-axis, with a sample size of $n = 500$. The nominal size of the tests is .05. Based on 1000 iterations.

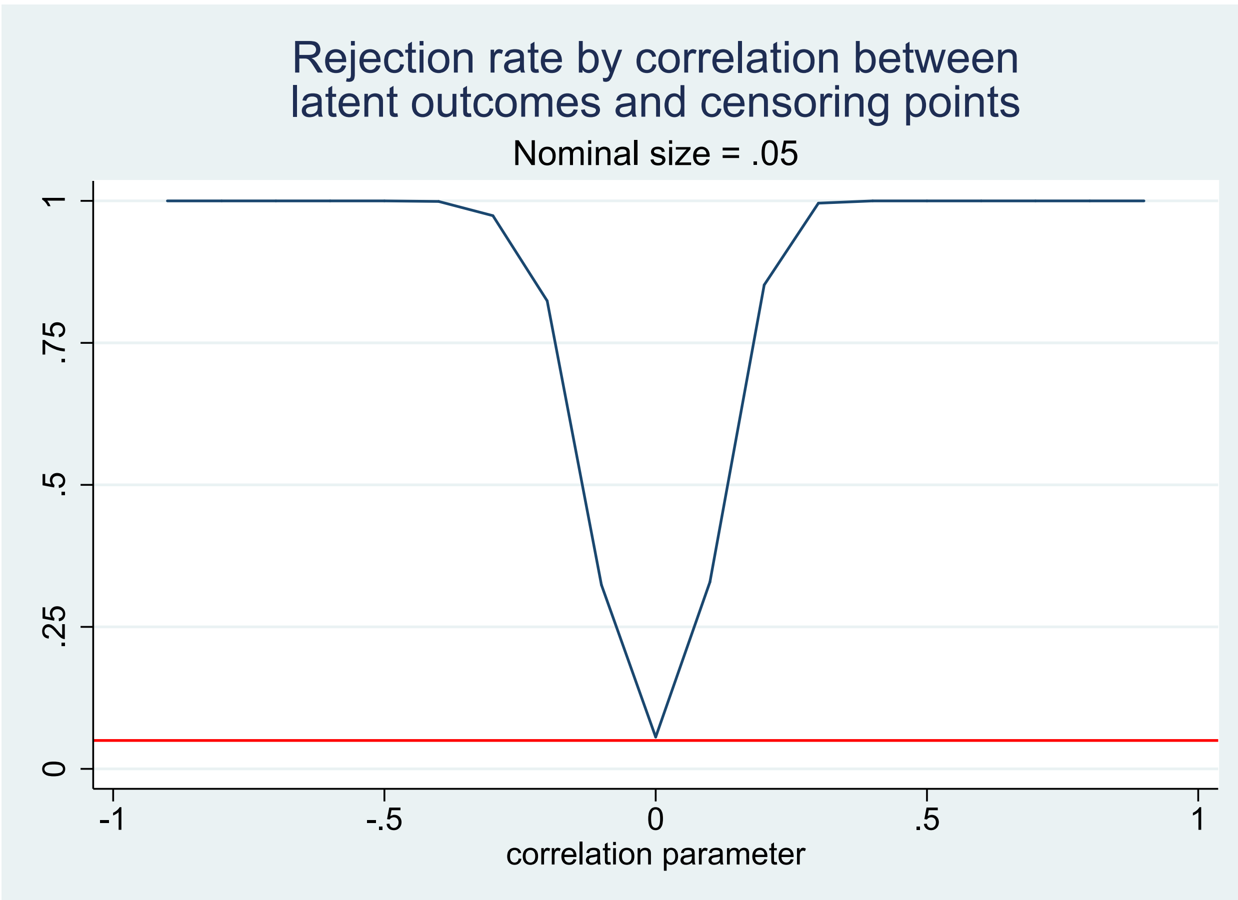

Figure 5: Monte Carlo simulation rejection rates from the censoring independence test as a function of the correlation between latent outcomes and censoring points (x-axis). The latent outcome and censoring point were generated as bivariate log-normals with correlation parameter ρ ranging from -.9 to .9. The latent outcome parameters were $\mu = 0$ and $\sigma = 1$. The censoring point parameters were $\mu_T = 1$ and $\sigma_T = .5$, with a sample size of $n = 500$. The nominal size of the tests is .05. Based on 1000 iterations.

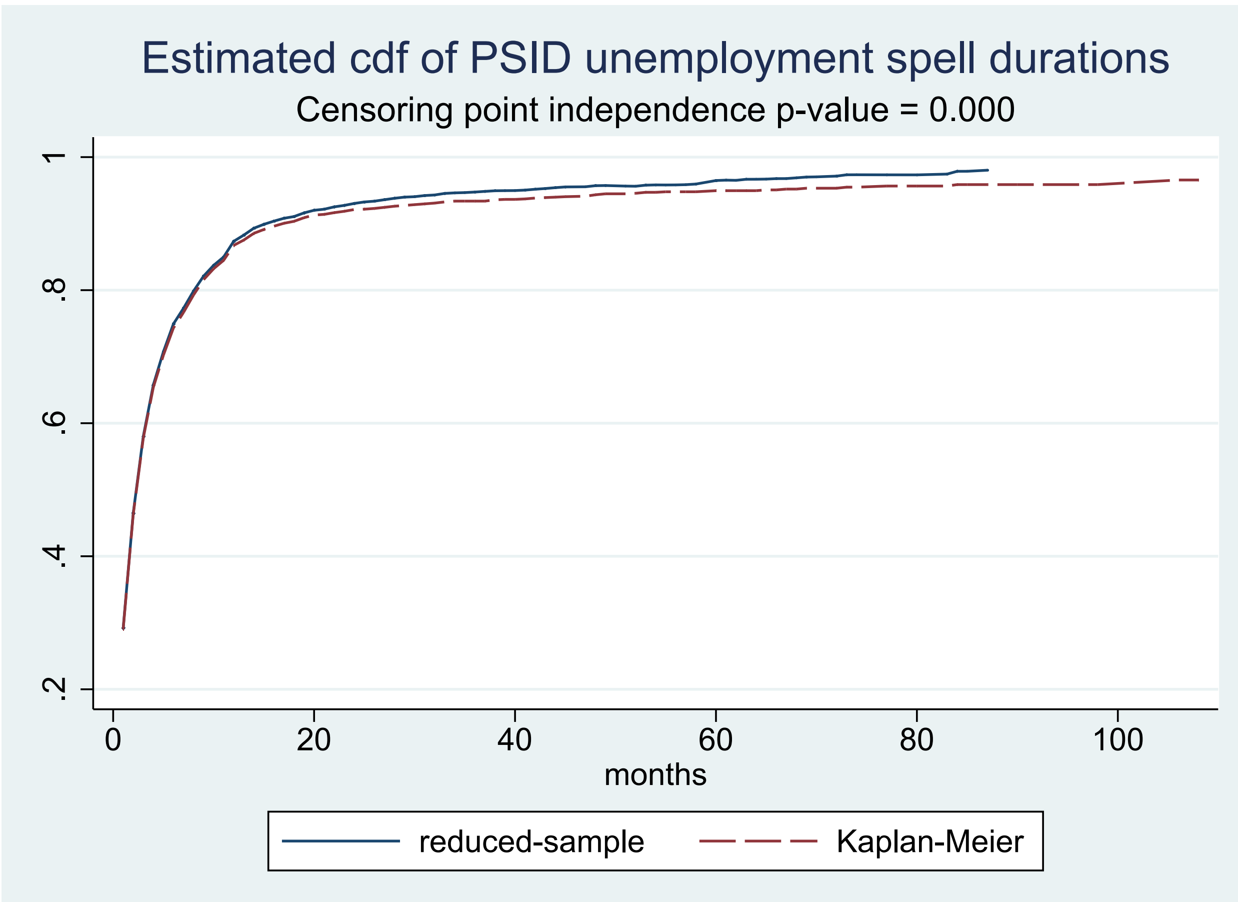

Figure 6: Reduced-sample and Kaplan-Meier estimates of the cdf of unemployment spell durations of working-age males. Data are from the PSID as analyzed in Low, et al (2010). The reported p-value is from the proposed censoring point independence test.

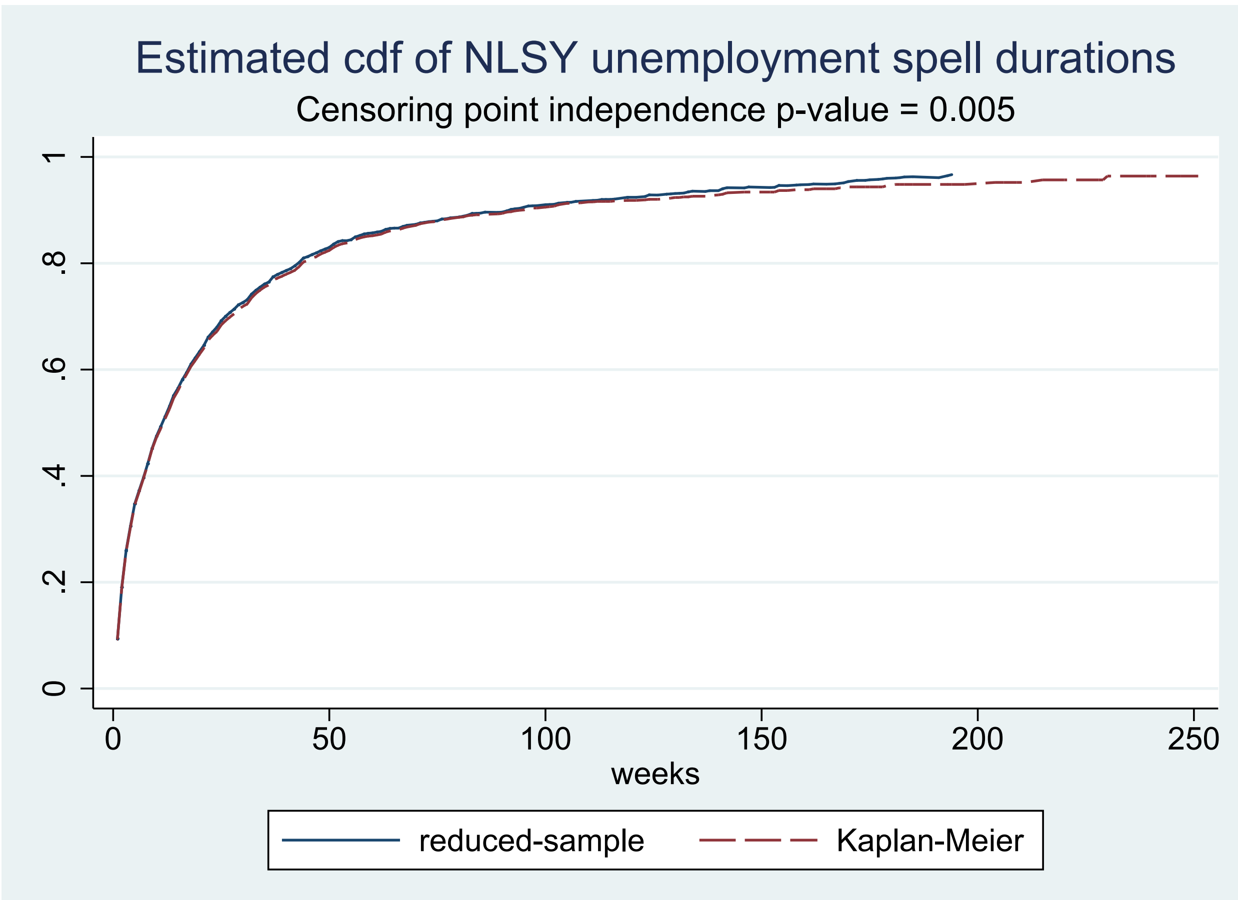

Figure 7: Reduced-sample and Kaplan-Meier estimates of the cdf of unemployment spell durations of NLSY respondents. Data are from the NLSY79 from the period 1989 to 1993. The reported p-value is from the proposed censoring point independence test.

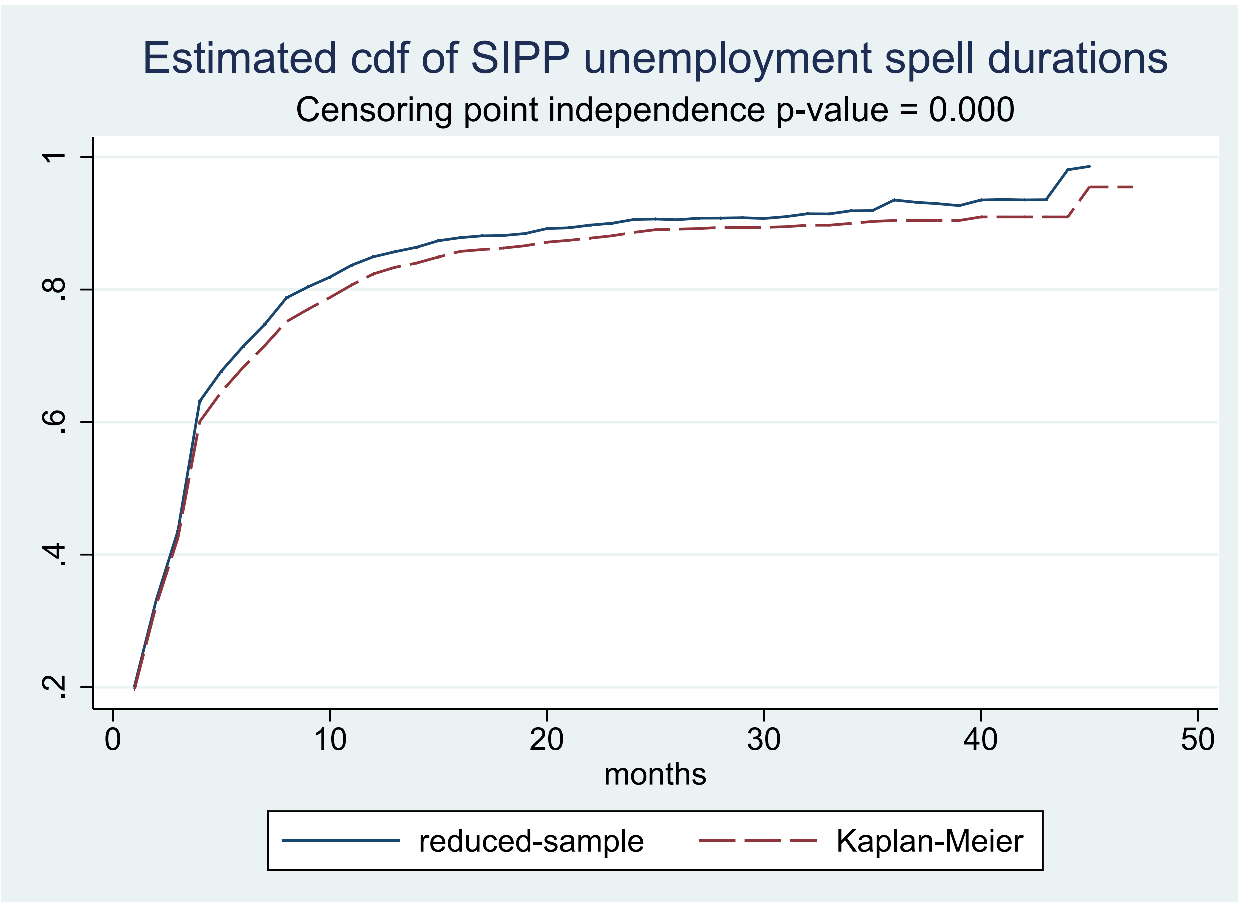

Figure 8: Reduced-sample and Kaplan-Meier estimates of the cdf of unemployment spell durations of working-age white males with at least a high school education. Data are from the 1996 SIPP as analyzed in Dey and Flinn (2005). The reported p-value is from the proposed censoring point independence test.