Distribution Regression in Duration Analysis: an Application to Unemployment Spells

Miguel A. Delgado[†], Andrés García-Suaza[‡] and Pedro H.C. Sant'Anna[††]

[†] *Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain*
E-mail: `delgado@est-econ.uc3m.es`
[‡] *Universidad del Rosario, Calle 14 625, Bogotá, Colombia.*
E-mail: `andres.garcia@urosario.edu.co`
[‡] *Microsoft and Vanderbilt University, 415 Calhoun Hall, Nashville, TN 37240, USA.*
E-mail: `pedro.h.santanna@vanderbilt.edu`

Summary   This article proposes inference procedures for distribution regression models in duration analysis using randomly right-censored data. This generalizes classical duration models by allowing situations where explanatory variables' marginal effects freely vary with duration time. The article discusses applications to testing uniform restrictions on the varying coefficients, inferences on average marginal effects, and others involving conditional distribution estimates. Finite sample properties of the proposed method are studied by means of Monte Carlo experiments. Finally, we apply our proposal to study the effects of unemployment benefits on unemployment duration.

Keywords: Conditional distribution; Duration models; Random Censoring; Unemployment duration; Varying coefficients model.
JEL Codes: C14, C24, C41, J64.

1. INTRODUCTION

Existing semiparametric duration models can be broadly classified into two groups: those based on the conditional hazard and those based on the quantile regression. The former includes the proportional hazard model (Cox, 1972, 1975), the proportional odds model (Clayton, 1976; Bennett, 1983; Murphy et al., 1997), and the accelerated failure time model (Kalbfleisch and Prentice, 1980); see Guo and Zeng (2014) for an overview. In these models, the conditional hazard identifies the conditional cumulative distribution expressed in terms of the error's marginal distribution of a transformed failure time regression model (see, e.g., Hothorn et al., 2014). Censored quantile regression proposals are relatively more recent, see, e.g., Ying et al. (1995), Honore et al. (2002), Portnoy (2003), Peng and Huang (2008), and Wang and Wang (2009).

Conditional hazard and quantile regression models are alternative modeling strategies with advantages and drawbacks. Classical conditional hazard specifications impose identification conditions that are difficult to justify in some circumstances. For instance, the proportional hazard specification rules out important forms of heterogeneity, see, e.g., Portnoy (2003) and the discussion in Section 2. On the other hand, models based on quantile regression specifications avoid this problem, but impose that the underlying conditional cumulative distribution of duration time is absolutely continuous, which rules out other sampling schemes such as discrete outcomes.

The distribution regression approach was proposed by Foresi and Peracchi (1995) and

formalized by Chernozhukov et al. (2013); see also Rothe and Wied (2013, 2019), Chernozhukov et al. (2019) and Chernozhukov et al. (2020) for later developments. One attractive feature of this modeling strategy is that marginal effects associated with distribution regression models are more flexible than those obtained from classical conditional hazard specifications, as they can accommodate richer forms of heterogeneity. In contrast to quantile regression, distribution regression can accommodate discrete, continuous and mixed duration data in a unified manner under fairly weak regularity conditions. Chernozhukov et al. (2013) show that distribution regression encompasses the Cox (1972) model as a special case and represents a useful alternative to quantile regression.

The distribution regression model specifies the distribution of the duration outcome variable by means of a generalized linear model with known link function and nonparametric varying coefficient depending on duration time. When data is uncensored, the varying coefficients at each duration value are consistently estimated by binary regressions. This method is also valid using censored data with known censoring time, which is an unlikely situation in duration studies, e.g., unemployment duration spells. When duration and censoring times are potentially unobserved, the standard binary regression approach lead to inconsistent estimators.

In this article, we propose a weighted binary regression procedure using Kaplan and Meier (1958) weights. The varying coefficients are identified by means of a set of moment restrictions of the duration time and covariates' joint distribution. These moments are consistently estimated by their corresponding Kaplan-Meier integrals when the duration time is subjected to censoring; if censoring is not an issue, the Kaplan-Meier integrals reduce to the sample analogue of the moment restrictions. The resulting estimator admits a representation as a non-degenerate U-statistic of order three, whose Hájek projection is asymptotically distributed as a normal with an asymptotic variance that can be consistently estimated from the data. The resulting inference procedures on the coefficients do not rely on choosing tuning parameters such as bandwidths, imposing smoothness conditions on the censoring random variable, or using truncation arguments.

We establish the consistency and finite-dimensional distribution convergence of the varying coefficients under weak regularity conditions. Under more restrictive assumptions we justify the convergence in distribution of the estimated coefficient function as a random element of a suitable metric space. We also justify bootstrap-based inference procedures.

In order to justify the asymptotic inferences based on our proposed method, we impose restrictions on the joint distribution of the duration time, censoring time and covariates. Stute (1993, 1996) provided strong consistency and a central limit theorem for Kaplan-Meier integrals that estimate moments of (partially observed) durations time and covariates under the following identification conditions: (a) the duration and censoring times are independent, and (b) the event of being censored is conditionally independent of the covariates given the duration time. Stute shows that the two conditions suffice to identify the joint distribution of duration time and covariates from the observed censored data, and, hence, to identify corresponding moments of these conditional distributions. Of course, the identification of alternative models may require alternative regularity conditions. For instance, identification of the Cox (1972)'s hazard model, Aalen (1980)'s additive hazard model, or censored quantile regressions models (Portnoy, 2003) only requires that duration and censoring times are independent given covariates. However, as we mention in Section 2, these models usually exclude important types of heterogeneity, or restrict their attention to absolutely continuous duration data. Alternatively, one can potentially estimate the moment equations using inverse probability weighting

(IPW) methods, as suggested by Robins and Rotnitzky (1992, 1995) and later extended by Wooldridge (2007) to general missing data problems. Although this path allows one to only assume that duration and censoring times are independent given covariates, it also requires that the conditional distribution of the censoring times given covariates is (uniformly) bounded away from one, which rules out discrete distributions and the most popular distribution functions used in survival analysis such as exponential, Gamma, log-normal and Weibull, among others. If one is not comfortable with these additional restrictions, truncation/trimming arguments need to be carefully introduced, and the choice of tuning parameters becomes a first-order concern. By adopting the Kaplan-Meier integrals approach, we avoid these alternative restrictions, but rely on conditions (a) and (b) above.

We illustrate the relevance of our proposal by assessing how changes in unemployment insurance benefits affect, on average, the distribution of unemployment duration, using data from the Survey of Income and Program Participation for the period 1985-2000. We find that, by allowing the distributional marginal effects to vary over the unemployment spell, our proposed method can reveal interesting insights when compared to traditional hazard models as those used by Chetty (2008). For instance, our results suggest a non-monotone marginal effect of an increase in unemployment insurance on the unemployment duration distribution. Such a finding is in sharp contrast with those obtained using classical proportional hazard models. On the other hand, our results agree with Chetty (2008) in that increases in unemployment insurance have larger effects on liquidity-constrained workers, suggesting that unemployment insurance affects unemployment duration not only through a moral hazard channel but also through a liquidity effect channel.

The rest of the article is organized as follows. Section 2 introduces the basic notation and motivates distribution regression models using duration data as an alternative to classical conditional hazard modeling. Section 3 describes our estimation procedure, whereas Section 4 introduces regularity conditions needed to justify inferences on the distribution regression varying coefficients. All the results are proved in the Supplementary Appendix. Application of the results to different contexts are placed in Section 5. Section 6 briefly summarizes the results of the Monte Carlo simulations – detailed discussion is presented in the Supplementary Appendix[1]. Finally, we apply the proposed techniques to investigate the effect of unemployment benefits on unemployment duration in Section 7. All our replication files are available at `https://github.com/pedrohcgs/KMDR-replication`.

2. DISTRIBUTION REGRESSION WITH DURATION OUTCOMES

Consider the $\mathbb{R}^{1+k} - valued$ random vector (T, X) defined on $(\Omega, \mathcal{F}, \mathbb{P})$, where T is the duration outcome of interest and X is a $k-$dimensional vector of time-invariant covariates, with supports $\mathcal{T}$ and $\mathcal{X}$, respectively. Henceforth, let $\mathbf{X} = (1, X')'$ and $\mathbf{x} = (1, x')'$. We assume that the conditional cumulative distribution function (CDF) of T given X follows the distribution regression (DR) model,

$$F_{Y|X}\left(t\,|\,X\right) = \Lambda\left(\varphi\left(t\right) + X'\beta_0\left(t\right)\right) \text{ for some } \left(\varphi\left(t\right), \beta_0\left(t\right)\right) \in \Theta \text{ a.s.} \qquad (2.1)$$

---

[1]The Supplementary Appendix is available at `https://pedrohcgs.github.io/files/Delgado_GarciaSuaza_SantAnna_2021_EJ-Appendix.pdf`

where $\theta_0\left(\cdot\right) = \left(\varphi\left(\cdot\right), \beta_0\left(\cdot\right)'\right)' \mapsto \Theta \subseteq \mathbb{R}^{k+1}$ is a vector of nonparametric functions, and Λ is a known link function.

Chernozhukov et al. (2013) point out that the DR model is a flexible alternative to classical duration models as it allows coefficients to vary with duration time. In particular, they show that it nests the Cox (1972) proportional hazard (PH) model. Model (2.1) also generalizes other traditional models commonly adopted in duration analysis such as Kalbfleisch and Prentice (1980) accelerated failure time (AFT) model, and Clayton (1976) proportional odds (PO) model.

We note that all the aforementioned classical duration models are special cases of the linear transformation model

$$\varphi\left(T\right) = -X'\beta_0 + \Lambda^{-1}(U)\ a.s., \tag{2.2}$$

where $\beta_0 \in \mathbb{R}^k$ is a vector of unknown parameters, φ is a (potentially unknown) monotonically increasing transformation function, Λ^{-1} is the $\Lambda's$ quantile function, and U is uniformly distributed in $[0,1]$ independently of X. For instance, (a) the PH model corresponds to model (2.2) with Λ the complementary log-log (cloglog) link function, $\Lambda\left(u\right) = 1 - \exp\left(-\exp\left(u\right)\right)$; ($b$) the AFT model corresponds to (2.2) with Λ commonly assumed to be the link function of a log-logistic or of a Weibull distribution; and (c) the PO model corresponds to (2.2) with Λ the logistic link function, $\Lambda\left(u\right) = \exp\left(u\right)/\left(1 + \exp\left(u\right)\right)$; see, e.g. Doksum and Gasko (1990), Cheng et al. (1995) and Hothorn et al. (2014). Indeed, the CDF associated with (2.2) is given by

$$F_{T|X}\left(t|X\right) = \Lambda\left(\varphi\left(t\right) + X'\beta_0\right). \tag{2.3}$$

Therefore, (2.1) is a natural generalization of (2.3) that allows all the slope coefficients varying with t. Other duration models with varying slope coefficients are also special cases of (2.1). For instance, Aalen (1980) semi-parametric additive hazard model reduces to (2.1) with $\Lambda\left(u\right) = 1 - \exp\left(-u\right)$. The non-proportional odds model considered by McCullagh (1980) and Armstrong and Sloan (1989) can also be expressed in terms of the DR specification (2.1) when Λ is the logistic link function.

We conclude this section by emphasizing that allowing for varying slope coefficients is particularly important to capture richer heterogeneous effects of covariates across the distribution of the duration outcome. Indeed, traditional duration specifications, such as the PH, PO or AFT models, implicitly impose that all covariates affect the CDF of T in a proportional, and monotonic form. If Λ is differentiable with Lebesgue density λ, partial effects for these models have the form

$$\frac{d}{dx}F_{T|X}\left(t|x\right) = \beta_0\lambda\left(\varphi\left(t\right) + x'\beta_0\right).$$

Thus, the sign of the marginal effect of any X's component is not allowed to vary with t, which may be too restrictive in some applications. For instance, in the context of the empirical application in Section 7, (2.3) does not allow a non-monotonic effect, e.g. U-shaped, ruling out non-stationary search models in the spirit of van den Berg (1990). The DR model (2.1) bypass such limitations since

$$\frac{d}{dx}F_{T|X}\left(t|x\right) = \beta_0(t)\lambda\left(\varphi\left(t\right) + x'\beta_0\left(t\right)\right),$$

which sign is allowed to vary with t. We view this added flexibility as an attractive feature of the DR modeling approach.

3. THE KAPLAN-MEIER DISTRIBUTION REGRESSION

The main challenge when dealing with duration data is that the outcome of interest T is subject to censoring according to a variable C. As so, inferences must be based on observations $\{Y_i, X_i, \delta_i\}_{i=1}^n$, with $Y_i = \min(T_i, C_i)$ the realized (potentially censored) outcome, $\delta_i = 1_{\{T_i \leq C_i\}}$ indicates whether or not T is observed, and $\{T_i, C_i, X_i\}_{i=1}^n$ are independent and identically distributed (*iid*) as (T, C, X). Henceforth, $1_{\{A\}}$ is the indicator function of the event A.

Suppose for the moment that $\delta_i = 1$, i.e., $T_i = Y_i$, for all $i = 1, \ldots, n$. In this case, the joint $(T, X')'$ CDF, $F(t, x)$, is consistently estimated by its sample version $\tilde{F}_n(t, x) = n^{-1} \sum_{i=1}^n 1_{\{Y_i \leq y, X_i \leq x\}}$, where inequalities are coordinate-wise. Therefore, following Foresi and Peracchi (1995) and Chernozhukov et al. (2013), $\theta_0(t)$ can be consistently estimated by $\tilde{\theta}_n(t)$, the maximizer of the conditional likelihood function of $\{1_{\{Y_i \leq t\}}, X_i\}_{i=1}^n$

$$\tilde{Q}_n(\theta, t) = \int_{\mathbb{R}^{1+k}} \ln p_{\theta,t}(1_{\{y \leq t\}}, x) \tilde{F}_n(dy, dx) = \frac{1}{n} \sum_{i=1}^n \ln p_{\theta,t}(1_{\{Y_i \leq t\}}, X_i), \qquad (3.1)$$

with

$$p_{\theta,t}(d, x) = \Lambda(\boldsymbol{x}'\theta)^d (1 - \Lambda(\boldsymbol{x}'\theta))^{1-d}.$$

However, when T is subject to right-censoring, i.e., $T \neq Y$, $\tilde{F}_n$ is no longer a consistent estimator of F and, hence, neither is $\tilde{\theta}_n(t)$.

Since T is not always observed, F and $\theta(t)$ must be identified from the joint distribution of the observed data on (Y, X, δ). However, this is not always possible without additional information. Henceforth, for any random variable ξ, which can be T, C or Y, define $F_\xi(t) \equiv \mathbb{P}(\xi \leq t)$ and $\tau_\xi \equiv \inf(t : \mathbb{P}(\xi \leq t) = 1)$, and let A denote the (possibly empty) set of F_T jumps. Also, for any generic function g, $g(t-) \equiv \lim_{s \uparrow t} g(s)$. If no information about T beyond τ_Y is available from the data, identification of

$$F_*(t, x) \equiv \begin{cases} F(t, x) \text{ for } t < \tau_Y \\ F(\tau_Y-, x) + 1_{\{\tau_Y \in A\}} [F(\tau_Y, x) - F(\tau_Y-, x)] \text{ for } t \geq \tau_Y. \end{cases}$$

is the best that one can hope for; see, e.g., Tsiatis (1975, 1981), and Stute (1993). In order to identify F_* we impose the following assumptions.

ASSUMPTION 3.1. *T and C are independent.*

ASSUMPTION 3.2. *δ and X are conditionally independent given T.*

Assumption 3.1 is introduced to identify $F_*(\cdot, \infty, \ldots, \infty)$ (see Tsiatis, 1975, 1981 for discussion), whereas Assumption 3.2 is introduced by Stute (1993) and allows us to incorporate covariates into the analysis. Taken together, these two assumptions imply that covariates should have no effect on the probability of being censored once T is known. Of course, Assumptions 3.1 and 3.2 hold if C is independent of (T, X), but can also hold under more general circumstances. See Stute (1993, 1996, 1999) for a discussion on Assumptions 3.1 and and 3.2, and the identification of F_* and its moments. Identification can be achieved under alternative conditions, in the context of alternative conditional survival models, or alternative estimation procedures. See Introduction for further comments.

REMARK 3.1. Under Assumptions 3.1 and 3.2, $\tau_Y = \min(\tau_T, \tau_C)$. Therefore, $F = F_*$ if either $\tau_T < \tau_C$, or $\tau_C = \infty$ irrespective of whether τ_T is finite or not. For most duration distributions considered in the literature, $\tau_T = \infty = \tau_C$; hence, $\tau_Y = \infty$. When $\tau_C < \tau_T$, $F \neq F_*$ in general, and F cannot be consistently estimated as T is not observed beyond τ_C. When $\tau_T = \tau_C < \infty$, $F = F_*$ depending on the local structure of F_T and F_C around the common endpoint. Notice that the above endpoint conditions for $F = F_*$ can be satisfied with discontinuous F.

In order to identify $\theta_0(t)$, we ensure that $F = F_*$ by imposing the following condition; see Stute (1999) for a similar assumption in the case of nonlinear regression with randomly censored outcomes.

ASSUMPTION 3.3. $\tau_T < \tau_C$ or $\tau_C = \infty$.

Thus, $\theta_0(t)$ is identified as the parameter value that maximizes $Q(\theta, t)$ under standard conditions in binary regression, e.g., Amemiya (1985, Section 9.2.2), or more recently, van der Vaart (1998, Example 5.40).

Henceforth, $Y_{1:n} \leq ... \leq Y_{n:n}$ are the ordered $Y-values$, where ties within outcomes T or within censoring times C are ordered arbitrarily and ties among T and C are treated as if the former precedes the latter. For observations $\{\xi_i\}_{i=1}^n$ of a random variable ξ, which may be δ or X, $\xi_{[i:n]}$ is the $i-th$ $\xi-$concomitant of the order statistics $\{Y_{i:n}\}_{i=1}^n$, i.e., $\xi_{[i:n]} = \xi_j$ if $Y_{i:n} = Y_j$.

In the absence of covariates, $F_{*T}(t) = F_*(t, \infty, ..., \infty)$ is consistently estimated by the Kaplan and Meier (1958) product-limit estimator,

$$\hat{F}_{Tn}(t) = 1 - \prod_{Y_{i:n} \leq t} \left(1 - \frac{\delta_{[i:n]}}{n-i+1}\right), \ t \in \mathbb{R}^+.$$

By noticing that the jumps of $\hat{F}_{Tn}$ at $Y_{i:n}$ are

$$W_{in} = \hat{F}_{Tn}(Y_{i:n}) - \hat{F}_{Tn}(Y_{i-1:n}) = \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_{[j:n]}}{n-j+1}\right), \qquad (3.2)$$

we can express $\hat{F}_{Tn}$ in an additive form as

$$\hat{F}_{Tn}(t) = \sum_{i=1}^n W_{in} \cdot 1_{\{Y_{i:n} \leq t\}}.$$

When covariates are present, the natural $F_*(t, x)$ estimator is

$$\hat{F}_n(t, x) = \sum_{i=1}^n W_{in} \cdot 1_{\left\{Y_{i:n} \leq t, X_{[i:n]} \leq x\right\}}, \qquad (3.3)$$

see, Stute (1993, 1996). Stute (1993) showed that, under Assumptions 3.1-3.3,

$$\lim_{n \to \infty} \sup_{(t,x) \in \mathcal{T} \times \mathbb{R}^k} \left|\left(\hat{F}_n - F\right)(t, x)\right| = 0 \ a.s.$$

Thus, (3.3) is indeed a natural candidate to replace $\tilde{F}_n$ in (3.1) under censoring. This

suggests estimating $Q_* (t, \theta)$ by the Kaplan-Meier integral,

$$\hat{Q}_n (\theta, t) = \int \ln p_\theta(1_{\{y \le t\}}, x) \hat{F}_n(dy, dx) = \sum_{i=1}^{n} W_{in} \cdot \ln p_\theta(1_{\{Y_{i:n} \le t\}}, X_{[i:n]}),$$

which is a weighted version of $\tilde{Q}_n (\theta, t)$.

The Kaplan-Meier distribution regression (KMDR) estimator of $\theta_0(t)$ is then given by

$$\hat{\theta}_n (t) = \arg\max_{\theta \in \Theta} \hat{Q}_n (\theta, t).$$

Notice that, to compute the KMDR estimators, one simply needs to run a weighted binary regression, where the weights are the (random) Kaplan-Meier weights W_{in}. In the absence of censoring, i.e., $Y_i = T_i$, $\delta_i = 1$, $i = 1, .., n$, we have that the Kaplan-Meier weights $W_{in} = n^{-1}$, and $\hat{\theta}_n (t)$ reduces to $\tilde{\theta}_n (t)$.

4. ASYMPTOTIC THEORY

In this section, we present the asymptotic properties of our proposed KMDR estimator $\hat{\theta}_n(t)$ for $\theta_0 (t)$.

In order to establish consistency, we impose the following fairly weak assumption which is compatible with discrete, continuous and mixed duration outcomes.

ASSUMPTION 4.1. *The CDF specification* (2.1) *holds, where* $\Lambda : \mathbb{R} \mapsto [0, 1]$ *is a continuously differentiable monotone function, the distribution of X is not concentrated on an affine subspace of $\mathbb{R}^{k-1}$, and $\mathbb{E} \|X\|^2 < \infty$.*

Next theorem, like any other result in the paper, is proved in the Supplemental Appendix. The proof applies the arguments in Example 5.40 in van der Vaart (1998) to show that

$$\sup_{\theta \in \Theta} \left| \left(\hat{Q}_n - Q \right) (\theta, t) \right| = o_p(1),$$

$$\sup_{\theta : \|\theta - \theta_*(t)\| \ge \varepsilon} Q (\theta, t) < Q (\theta_0(t), t) \text{ for all } \varepsilon > 0 \text{ and each } t \in \mathcal{T}.$$

THEOREM 4.1. *Under Assumptions 3.1, 3.2, 3.3 and 4.1, any sequence of estimators* $\left\{ \hat{\theta}_n(t) \right\}_{n \ge 1}$ *that satisfies*

$$\hat{Q}_n \left(\hat{\theta}_n(t) \right) \ge \hat{Q}_n (\theta_0(t)) - o_p (1),$$

converges in probability to $\theta_0(t)$, as $n \to \infty$.

In order to provide the finite dimensional asymptotic distribution of $\hat{\theta}_n(t)$, we need the following assumption.

ASSUMPTION 4.2. *$\theta_0 (t)$ is an inner point of Θ, which is a compact subset of $\mathbb{R}^{1+k}$. Also, let $\mathcal{V} = \{\boldsymbol{x}'\theta : \boldsymbol{x} = (1, x')', x \in \mathcal{X}, \theta \in \Theta\}$; for $v \in \mathcal{V}$, $\Lambda (v)$ is bounded away from zero and one, and admits a continuously differentiable Lebesgue density $\lambda(v)$.*

The restriction on Λ in Assumption 4.2 is a classical assumption, which can be found,

for instance, in Amemiya (1985) Assumption 9.2.1. The assumption is satisfied by the normal and logistic distribution (probit and logit), but also for many other distributions. In the general case, Assumption 4.2 essentially rules out extremes t, since, in these cases, Λ may be near zero or one. This assumption can be relaxed at the cost of more involved proofs.

The score function is given by

$$\hat{\Psi}_n\left(\theta, t\right) = \frac{\partial}{\partial \theta} \hat{Q}_n\left(\theta, t\right) = \sum_{i=1}^{n} W_{ni} \cdot \psi_{\theta t}\left(Y_{i:n}, X_{[i:n]}\right),$$

where

$$\psi_{\theta t}\left(y, x\right) = \frac{1_{\{y \leq t\}} - \Lambda(\mathbf{x}'\theta)}{\Lambda(\mathbf{x}'\theta)\left[1 - \Lambda(\mathbf{x}'\theta)\right]} \lambda(\mathbf{x}'\theta)\mathbf{x}.$$

The conditional Fisher information that the binary variable $1_{\{T \leq t\}}$ contains about $\theta_0(t)$ given X is $\mathcal{I}_0\left(t\right) = \mathcal{I}\left(\theta_0\left(t\right), t\right)$, where

$$\mathcal{I}\left(\theta, t\right) = \int_{\mathbb{R}^{1+k}} \psi_{\theta t}\left(y, x\right) \psi_{\theta t}'\left(y, x\right) F\left(dy, dx\right). \tag{4.1}$$

Notice that, under our assumptions, $\theta \mapsto \mathcal{I}\left(\theta, \cdot\right)$ is continuously differentiable.

We first pay attention to the asymptotic distribution of the score $\hat{\Psi}_n^0\left(t\right) = \hat{\Psi}_n\left(\theta_0(t), t\right)$. Applying Stute (1995, 1996) lemmata, $\hat{\Psi}_n\left(\theta, t\right)$ can be expressed as an $U-$statistic of order three with Hájek projection $\hat{U}_n^0(t) = \hat{U}_n(\theta_0(t), t)$, where

$$\hat{U}_n(\theta, t) = \frac{1}{n} \sum_{i=1}^{n} \zeta_{\psi_{\theta t}}(Y_i, X_i, \delta_i),$$

and for any function $(y, x) \mapsto \varphi\left(y, x\right)$,

$$\zeta_\varphi(y, x, d) = \varphi(y, x)\gamma^{(0)}(y)d + \gamma_\varphi^{(1)}(y)(1 - d) - \gamma_\varphi^{(2)}(y), \tag{4.2}$$

$$\gamma^{(0)}(y) = \exp\left\{\int_{-\infty}^{y-} \frac{F_Y^0(dw)}{1 - F_Y(w)}\right\}, \tag{4.3}$$

$$\gamma_\varphi^{(1)}(y) = \frac{1}{1 - F_Y(y)} \int 1_{\{y < w\}} \varphi(w, x)\gamma^{(0)}(w)F^1(dw, dx), \tag{4.4}$$

$$\gamma_\varphi^{(2)}(y) = \int \int \frac{1_{\{v < y, v < w\}} \varphi(w, x)\gamma^{(0)}(w)}{\left[1 - F_Y(v)\right]^2} F_Y^0(dv)F^1(dw, dx), \tag{4.5}$$

where $F_Y^0(t) = \mathbb{P}\left(Y \leq t, \delta = 0\right)$ and $F^1(t, x) = \mathbb{P}\left(Y \leq t, X \leq x, \delta = 1\right)$.

To derive the asymptotic normality of the score, we need the following extra moment conditions.

ASSUMPTION 4.3. *For each $t \in \mathcal{T}$,*

$$(a)\ \mathbb{E}\left[\left\|\psi_{\theta_0(t)t}(Y, X)\gamma^{(0)}(Y)\delta\right\|^2\right] < \infty$$

$$(b)\ \int \left\|\psi_{\theta_0(t)t}(y, x)\right\| \sqrt{S(y)}F(dy, dx) < \infty,$$

with

$$S\left(y\right) = \int_0^{y^-} \frac{F_C\left(d\bar{y}\right)}{\left[1 - F_Y\left(\bar{y}\right)\right]\left[1 - F_C\left(\bar{y}\right)\right]}.$$

Assumption 4.3(a) guarantees that the variance of the leading term in $\zeta_{\psi_{\theta_0(t)t}}$ is bounded, and implies that $\mathbb{E}\left\|\zeta_{\psi_{\theta t}}(Y, X, \delta)\right\|^2$ is finite. The bias of the Kaplan-Meier integral is not necessarily $o\left(n^{-1/2}\right)$ for any integrand function, and may decrease to zero at a polynomial rate depending on the degree of censoring, which is characterized by the function S. Assumption 4.3(b) on $\psi_{\theta_0(t)t}$ guarantees that the $\hat{\Psi}_n^0\left(t\right)$ bias is of order $o\left(n^{-1/2}\right)$; see Stute (1994), Chen and Ying (1996), and Chen and Lo (1997).

Let $\{Z\left(t\right)\}_{t=t_1}^{t_m}$ be a $k+1$ Gaussian random vector with zero mean and covariances

$$Cov\left(Z\left(t_j\right)Z\left(t_m\right)\right) = \mathbb{E}\left[\zeta_{\psi_{\theta_0(t_j)t_j}}\left(Y, X, \delta\right)\zeta'_{\psi_{\theta_0(t_m)t_m}}\left(Y, X, \delta\right)\right] = \Omega_0\left(j, m\right), \qquad (4.6)$$

with $j, m = 1, ..., k+1$.

With Assumption 4.3, we show that $\hat{\Psi}_n^0\left(t\right)$ is asymptotically equivalent to a $U-$ *statistic* of order 3 with Hájek projection $\hat{U}_n^0\left(t\right)$. Thus, $\sqrt{n}\hat{U}_n^0$ is asymptotically normal with finite dimensional covariance $\Omega_0\left(j, m\right)$. Theorem 4.2 below follows by applying Theorem 5.21 in van der Vaart (1998).

THEOREM 4.2. *Under Assumptions 3.1, 3.2, 3.3, 4.1, 4.2, and 4.3, for each* $t \in \mathcal{T}$,

$$\hat{\theta}_n\left(t\right) = \theta_0\left(t\right) + \mathcal{I}_0^{-1}\left(t\right) \cdot \hat{U}_n^0\left(t\right) + o_p\left(n^{-1/2}\right) \qquad (4.7)$$

and

$$\left\{\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)\left(t\right)\right\}_{t=t_1}^{t_m} \to_d \left\{\mathcal{I}_0^{-1}\left(t\right)Z(t)\right\}_{t=t_1}^{t_m}.$$

In order to conduct inferences on $\theta_0(\cdot)$, $\Omega_0\left(j, m\right)$ is estimated by

$$\hat{\Omega}_n\left(j, m\right) = \frac{1}{n}\sum_{i=1}^n \hat{\zeta}_i(t_j)\hat{\zeta}'_i(t_m),$$

where

$$\hat{\zeta}_i(t) = \hat{\psi}_i(t)\hat{\gamma}_i^{(0)}\delta_i + \hat{\gamma}_i^{(1)}(t)(1 - \delta) - \hat{\gamma}_i^{(2)}(t), \qquad (4.8)$$

$$\hat{\gamma}_i^{(0)} = \exp\left\{\frac{1}{n}\sum_{j=1}^n 1_{\{Y_j < Y_i\}}\frac{1 - \delta_j}{1 - \hat{F}_{nY}(Y_j)}\right\},$$

$$\hat{\gamma}_i^{(1)}(t) = \frac{1}{1 - \hat{F}_Y(Y_i)}\frac{1}{n}\sum_{j=1}^n 1_{\{Y_i < Y_j\}}\hat{\psi}_j(t)\hat{\gamma}_j^{(0)}\delta_j,$$

$$\hat{\gamma}_i^{(2)}(t) = \frac{1}{n^2}\sum_{j=1}^n\sum_{\ell=1}^n \frac{1_{\{Y_j < Y_i, Y_j < Y_\ell\}}\hat{\psi}_\ell(t)\hat{\gamma}_\ell^{(0)}}{\left[1 - \hat{F}_Y(Y_j)\right]^2}\left(1 - \delta_j\right)\delta_\ell,$$

$$\hat{\psi}_i(t) = \psi_{\hat{\theta}_n(t)t}\left(Y_i, X_i\right), \ \hat{F}_Y(t) = \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i \leq t\}},$$

and $\mathcal{I}_0(t)$ is estimated by

$$\widehat{\mathcal{I}}_n(t) = \sum_{i=1}^{n} W_{ni} \cdot \hat{\psi}_{i:n}(t)\hat{\psi}'_{i:n}(t), \tag{4.9}$$

where $\hat{\psi}_{i:n}(t) = \psi_{\hat{\theta}_n(t)t}\left(Y_{i:n}, X_{[i:n]}\right)^2$.

COROLLARY 4.1. *Under the assumptions of Theorem 4.2,* $\hat{\Omega}_n(j,m) \to_p \Omega_0(j,m)$ *and* $\widehat{\mathcal{I}}_n(t_j) \to_p \mathcal{I}_0^{-1}(t_j)$ *for all* $t_j, t_m \in \mathcal{T}$.

We can also justify inferences with the assistance of a multiplier bootstrap technique in the spirit of Theorem 2.3 in Stute et al. (2000) and Theorem 4 in Sant'Anna (2020), using an external resample of $\left\{\hat{\zeta}_i(t)\right\}_{i=1}^{n}$. Once we generate *iid* random numbers $\{V_i\}_{i=1}^{n}$ independently of the data with mean 0, variance 1, and finite third moment, the "resampled" $\left\{\hat{\zeta}_i^*(t)\right\}_{i=1}^{n}$, with $\hat{\zeta}_i^*(t) = \hat{\zeta}_i(t)V_i$, forms a basis to compute the bootstrap analog of $\hat{\theta}_n(t)$ based on the asymptotic linearization (4.7),

$$\hat{\theta}_n^*(t) = \hat{\theta}_n(t) + \widehat{\mathcal{I}}_n^{-1}(t)\hat{U}_n^*(t), \tag{4.10}$$

where

$$\hat{U}_n^*(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{\zeta}_i^*(t)$$

is the bootstrap version of $\hat{U}_n^0(t)$.

THEOREM 4.3. *Let the assumptions of Theorem 4.2 hold. Then, for* $t_1, ..., t_m \in \mathcal{T}$*, we have that* $\left\{\sqrt{n}\left(\hat{\theta}_n^* - \hat{\theta}_n\right)(t)\right\}_{t=t_1}^{t_m}$ *converges in distribution under the bootstrap law to* $\left\{\mathcal{I}_0^{-1}(t)Z(t)\right\}_{t=t_1}^{t_m}$*, with probability 1.*

The theorem follows by checking the conditional version of the Linderberg-Levy conditions to show that, with probability 1, $\hat{U}_n^*(t)$ converges in distribution to Z under the bootstrap law, i.e., conditional on the data, with probability 1.

Next, we strengthen the pointwise results in Theorems 4.2 and 4.3 to hold uniformly in t. Towards this end, let the space $\ell^\infty(\mathcal{G})^q$ be the set of all $q \times 1$ vectors of bounded functions on $\mathcal{G}$, that is, all the functions' vectors $g : u \mapsto \mathbb{R}$ such that $\sup_{u \in \mathcal{G}} \|g(u)\| < \infty$. $\mathcal{G}$ can be either an Euclidean or a functional space. We interpret the multivariate empirical process indexed by functions in $\mathcal{G}$ as a random element in the metric space $\ell^\infty(\mathcal{G})^q$ endowed with the sup-norm. The empirical process $\hat{\Psi}_n$ is indexed by $(\theta, t) \in \Theta \times \mathcal{T}$; in this case $\mathcal{G} = \Theta \times \mathcal{T}_0$, where $\mathcal{T}_0$ is the compact interval of $\mathcal{T}$ of interest. But $\hat{\Psi}_n$ can also be interpreted as an empirical process indexed by the class $\mathcal{F}$ of functions $\psi_{\theta t} : \mathcal{X} \times \mathcal{T} \times [0,1] \mapsto \mathbb{R}^{1+k}$; in this case $\mathcal{G} = \mathcal{F}$. The random process $\hat{\Psi}_n$ is viewed as a random element of the metric space $\ell^\infty(\Theta \times \mathcal{T}_0)^{k+1}$ or $\ell^\infty(\mathcal{F})^{k+1}$, where $\mathcal{T}_0 \in \mathcal{T}$ is the compact subset of $\mathcal{T}$ we are interested in.

In order to establish the weak convergence of $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ as a random element of

---

[2]Alternatively, one can estimate $-\mathcal{I}_0(t)$ using the Hessian of $\hat{Q}_n(\hat{\theta}_n, t)$. We follow this path in our simulations.

the metric space $\ell^\infty (\mathcal{T}_0)^{k+1}$, we impose the following additional assumption, which is also assumed in Chernozhukov et al. (2013).

ASSUMPTION 4.4. *The duration interval of interest $\mathcal{T}_0$ is a compact subset of $\mathbb{R}^+$, and the conditional distribution function $F_{Y|X} (y| X)$ admits a Lebesgue density $f_{Y|X} (y| X)$ that is uniformly bounded and uniformly continuous in $\mathcal{T}_0$ with probability 1.*

Notice that $f_{T|X} (t| X) = \dot\theta_0' (t) \cdot \boldsymbol{X} \cdot \lambda (\theta_0' (t) \boldsymbol{X})$ a.s., where $\dot\theta_0 (t) = d\theta_0 (t) / dt$. Therefore, assuming that $f_{T|X}$ is a proper probability density may not be innocuous. In practice, $\mathcal{T}_0$ should be bounded by the minimum and maximum value of the observed uncensored duration.

Let $\mathbb{Z}$ be a centered tight Gaussian element of $\ell^\infty (\mathcal{T}_0)^{k+1}$ with matrix of variance and covariance functions

$$\mathbb{E} [\mathbb{Z} (t_1) \mathbb{Z}' (t_2)] = \Omega_0 (t_1, t_2), \ t_1, t_2 \in \mathcal{T}_0,$$

with $\Omega_0 (t_1, t_2)$ defined in (4.6).

THEOREM 4.4. *Given a compact subset $\mathcal{T}_0$ of $\mathcal{T}$, under Assumptions 3.1, 3.2, 3.3, 4.1, 4.2, 4.3, and 4.4, $\left\{ \sqrt{n} \left(\hat\theta_n - \theta_0 \right) (t) \right\}_{t \in \mathcal{T}_0}$ converges weakly to $\left\{ \mathcal{I}_0^{-1} (t) \mathbb{Z} (t) \right\}_{t \in \mathcal{T}_0}$ in $\ell^\infty (\mathcal{T}_0)^{k+1}$.*

This functional CLT is proved extending Chernozhukov et al. (2013) (CFM henceforth) strategy to our setup. To this end, we first show, using Stute (1995, 1996) lemmata, that the score function $\hat\Psi_n$ and U_n are asymptotically equivalent in the metric space $\ell^\infty (\Theta, \mathcal{T}_0)^{k+1}$, i.e.,

$$\sup_{\theta \in \Theta, t \in \mathcal{T}_0} \left\| \hat\Psi_n (\theta, t) - U_n (\theta, t) \right\| = o_p \left(n^{-1/2} \right),$$

where U_n is a $U - process$ with Hájek projection $\hat U_n$. Then, we show that

$$\sup_{\theta \in \Theta, t \in \mathcal{T}_0} \left\| U_n (\theta, t) - \hat U_n (\theta, t) \right\| = o_p \left(n^{-1/2} \right)$$

applying asymptotic results for U-process in Arcones and Giné (1993, 1995). Since the class $\mathcal{E}$ of functions $(y, x, d) \mapsto \zeta_{\psi_{\theta t}} (y, x, d)$ is Donsker,

$$\left\{ \sqrt{n} \left(\hat U_n (\theta, t) - \mathbb{E} [\zeta_{\psi_{\theta t}} (Y, X, \delta)] \right) : \theta \in \Theta, t \in \mathcal{T}_0 \right\}$$

converges weakly in $\ell^\infty (\Theta \times \mathcal{T}_0)^{k+1}$. Then, noticing that Assumptions 4.1 - 4.4 imply Assumption DR in CFM, the functional $\theta \mapsto \Psi_0 (\theta, t)$, with $\Psi_0 (\theta, t) = \mathbb{E} [\zeta_{\psi_{\theta t}} (Y, X, \delta)]$, is continuously differentiable for each $t \in \mathcal{T}_0$, with $d\Psi_0 (\theta, t) / d\theta' |_{\theta = \theta_0 (t)} = -\mathcal{I}_0 (t)$, which is invertible. Then, according to CFM's Lemma E.3,

$$\sqrt{n} \left(\hat\theta_n - \theta_0 \right) (t) = \mathcal{I}_0^{-1} (t) \cdot \hat\Psi_n (\theta_0 (t), t) + \hat r_n (t),$$

with $\sup_{t \in \mathcal{T}_0} \| \hat r_n (t) \| = o_p (1)$. This result and $\mathcal{E}$'s Donskerness prove the theorem.

A bootstrap version of this theorem follows by applying the multiplier CLT (see sections 2.6 and 3.6 in van der Vaart and Wellner, 1996).

THEOREM 4.5. *Suppose the assumptions of Theorem (4.4) hold. Then, with probability 1* $\left\{ \sqrt{n} \left(\hat{\theta}_n^* - \hat{\theta}_n \right)(t) \right\}_{t \in \mathcal{T}_0}$ *converges weakly under the bootstrap law to* $\left\{ \mathcal{I}^{-1}(t) \mathbb{Z}(t) \right\}_{t \in \mathcal{T}_0}$ *in* $\ell^\infty (\mathcal{T}_0)^{k+1}$.

In next section we discuss applications of these results to test restrictions on θ_0, as well as to make inferences on counterfactual distributions and on average distribution marginal effects.

5. APPLICATIONS

5.1. Testing Linear and Nonlinear Hypothesis

A natural class of hypotheses to be tested is those of the form

$$H_0 : \sup_{t \in \mathcal{T}_0} \| \Phi \left(\theta_0(t) \right) \| = 0 \text{ against } H_1 : \sup_{t \in \mathcal{T}_0} \| \Phi \left(\theta_0(t) \right) \| > 0,$$

for some compact subset $\mathcal{T}_0$ of $\mathcal{T}$, where $\theta \mapsto \Phi(\theta) \in \mathbb{R}^q$, $q \leq k+1$, is a continuously differentiable map with derivatives $\dot{\Phi}(\theta) = \partial \Phi(\theta) / \partial \theta'$, such that $rank \left(\dot{\Phi}(\theta) \right) = q$ for all θ in a neighborhood of $\theta_0(t)$. For instance, when $\Phi(\theta) = \theta$, the null and alternative reads

$$H_0 : \sup_{t \in \mathcal{T}_0} \| \theta_0(t) \| = 0 \text{ against } H_1 : \sup_{t \in \mathcal{T}_0} \| \theta_0(t) \| > 0,$$

which is a significance test for varying coefficients. We can also test that a linear combination of coefficients is satisfied, e.g., $\Phi(\theta) = R\theta - r$, where R and r are known and $rank(R) = k+1$.

A natural statistic is,

$$\hat{\omega}_n = n \cdot \sup_{t \in \mathcal{T}_0} \left\| \Phi \left(\hat{\theta}_n(t) \right) \right\|^2.$$

By the delta-method, and applying Theorem 4.4, uniformly in $t \in \mathcal{T}_0$,

$$\sqrt{n} \left[\Phi \left(\hat{\theta}_n(t) \right) - \Phi \left(\theta_0(t) \right) \right] = \dot{\Phi} \left(\theta_0(t) \right) \sqrt{n} \left(\hat{\theta}_n - \theta_0 \right)(t) + o_p(1)$$

and under H_0,

$$\hat{\omega}_n \rightarrow_d \omega_\infty = \sup_{t \in \mathcal{T}_0} \left\| \dot{\Phi} \left(\theta_0(t) \right) \mathcal{I}_0^{-1}(t) \mathbb{Z}(t) \right\|^2.$$

Since the $\omega'_\infty s$ distribution depends on $\theta_0(\cdot)$ and other unknown features of the underlying data generating process, analytical critical values seems unfeasible, at least with this level of generality. However, critical values can be estimated using the bootstrapped estimator of $\theta_0(t)$

Applying Theorem 4.5, uniformly in $t \in \mathcal{T}_0$, for almost every sample, it follows that

$$\sqrt{n} \left[\Phi \left(\hat{\theta}_n^*(t) \right) - \Phi \left(\hat{\theta}_n(t) \right) \right] = \dot{\Phi} \left(\hat{\theta}_n(t) \right) \sqrt{n} \left(\hat{\theta}_n^* - \hat{\theta}_n \right)(t) + o_p(1).$$

Therefore, we can use the bootstrap test statistic

$$\hat{\omega}_n^* = n \cdot \sup_{t \in \mathcal{T}_0} \left\| \Phi \left(\hat{\theta}_n^*(t) \right) - \Phi \left(\hat{\theta}_n(t) \right) \right\|^2. \tag{5.1}$$

By Theorem 4.5, under H_0, with probability 1, $\hat{\omega}_n^*$ converges in distribution under the bootstrap law to ω_∞ . Under H_1, $\hat{\omega}_n^* = O(1)$ with probability 1.

We now describe a practical bootstrap algorithm to conduct such types of tests.

ALGORITHM 5.1. (BOOTSTRAPPED-BASED HYPOTHESIS TESTING)

STEP *1*. Generate *iid* $\{V_i\}_{i=1}^n$ from a distribution with mean 0, variance 1 and finite third moment, e.g., Rademacher distribution.

STEP *2*. For a grid $t = t_1, \ldots, t_p$, compute $\left\{\hat{\theta}_n^*(t)\right\}_{t=t_1}^{t_p}$ as in (4.10), using the same $\{V_i\}_{i=1}^n$ for all t's.

STEP *3*. Compute the bootstrap test statistics $\hat{\omega}_n^*$ in (5.1)

STEP *4*. Repeat steps 1-3 B times.

STEP *5*. Reject H_0 at the $\alpha - level$ of significance if either $\hat{\omega}_n > \hat{\omega}_{n,\alpha}^*$, where $\hat{\omega}_{n,\alpha}^*$ is the empirical $(1-\alpha)$-quantile of the B bootstrap draws of $\hat{\omega}_n^*$, or if p-val* $> \alpha$, where

$$p\text{-}val^* = \frac{1}{B}\sum_{j=1}^B 1_{\left\{\hat{\omega}_n^{*(j)} > \hat{\omega}_n\right\}}.$$

An interesting application of this type of test is to testing constancy of some functional of θ_0. For instance, one can set

$$\Phi(\theta) = \left[\mathbf{0} \vdots I_K\right]\theta$$

to test that all the slope coefficients are constant. A test statistic for this hypothesis is

$$\tilde{\omega}_n = n \cdot \sup_{t \in \mathcal{T}_0}\left\|\Phi\left(\hat{\theta}_n(t)\right) - \bar{\bar{\Phi}}_n\right\|^2,$$

with $\bar{\bar{\Phi}}_n = \int_{s \in \mathcal{T}_0}\Phi\left(\hat{\theta}_n(s)\right)\Psi(ds)/\int_{s \in \mathcal{T}_0}\Psi(ds)$, Ψ being a known probability measure.[3] Bootstrap critical values can be obtained using the bootstrapped statistic,

$$\tilde{\omega}_n^* = n \cdot \sup_{t \in \mathcal{T}_0}\left\|\left[\Phi\left(\hat{\theta}_n^*(t)\right) - \bar{\bar{\Phi}}_n^*\right] - \left[\Phi\left(\hat{\theta}_n(t)\right) - \bar{\bar{\Phi}}_n\right]\right\|^2,$$

with $\bar{\bar{\Phi}}_n^* = \int_{s \in \mathcal{T}_0}\Phi\left(\hat{\theta}_n^*(s)\right)\Psi(ds)/\int_{s \in \mathcal{T}_0}\Psi(ds)$.

5.2. Inferences on $F_{T|X}$

Theorems 4.4 and 4.5 can also be applied to make inferences on $F_{T|X}(t|x)$. The DR $F_{T|X}(t|x)'s$ estimator is

$$\hat{F}_{n\,T|X}(t|x) = \Lambda\left(\boldsymbol{x}'\hat{\theta}_n(t)\right),$$

with bootstrap analog,

$$\hat{F}_{n\,T|X}^*(t|x) = \Lambda\left(\boldsymbol{x}'\hat{\theta}_n^*(t)\right).$$

---

[3]In practice, one can set Ψ to be equal to the empirical distribution of the censored outcome Y, though, in this case, the bootstrap procedure needs to be adjusted to account for this additional source of randomness. See Section 5.3 for related results.

The asymptotic distribution can be obtained using the delta-method. Applying Theorem 4.4, we have that

$$\left\{ \sqrt{n} \left(\hat{F}_{n\,T|X} - F_{T|X} \right) (t|\,x) \right\}_{(t,x) \in \mathcal{A}_0} \to_d \left\{ \lambda \left(\theta_0'(t)\,x \right) \cdot x \cdot \mathcal{I}_0^{-1}(t)\,\mathbb{Z}(t) \right\}_{(t,x) \in \mathcal{A}_0},$$

where $\mathcal{A}_0 = \mathcal{T}_0 \times \mathcal{X}_0$ is a compact subset of $\mathcal{T}_0 \times \mathcal{X}_0$. Likewise, applying Theorem 4.5, with probability 1,

$$\left\{ \sqrt{n} \left(\hat{F}_{n\,T|X}^* - \hat{F}_{n\,T|X} \right) (t|\,x) \right\}_{(t,x) \in \mathcal{A}_0} \to_d \left\{ \lambda \left(\theta_0'(t)\,x \right) \cdot x \cdot \mathcal{I}_0^{-1}(t)\,\mathbb{Z}(t) \right\}_{(t,x) \in \mathcal{A}_0} \text{ in } \ell^\infty(\mathcal{A}_0).$$

Given two samples $\{Y_i^{(j)}, \delta_i^{(j)}, X_i^{(j)}\}_{i=1}^{n_j}$, $j = 1, 2$, under some standard overlap conditions, we can use the conditional CDF estimator computed using sample 1, $\widehat{F}_{T|X,n}^{(1)}$, to compute the marginal counterfactual CDF of population 1 with respect to population 2,

$$\widehat{F}_{T,n}^{(1,2)}(t) = \int_{\mathcal{X}} \widehat{F}_{T|X,n}^{(1)}(t|x)\tilde{F}_{X,n}^{(2)}(dx) = \frac{1}{n_2} \sum_{i=1}^{n_2} \left(\widehat{F}_{T|X,n}^{(1)}(t|X_i^{(2)}) \right).$$

5.3. Average Distribution Marginal Effects

Although $\hat{\theta}_n(t)$ provides useful information about the direction/sign of the effect of changes in X on $F_{T|X}(t|X)$, in general, $\hat{\theta}_n(t)$ may not have a clear economic interpretation. In this section, we argue that this potential limitation can be easily avoided by focusing on the average distribution marginal effects (ADME) of X,[4]

$$\eta_0(t) \equiv \mathbb{E}\left[\beta_0(t)\,\lambda\left(\mathbf{X}'\theta_0(t) \right) \right]. \tag{5.2}$$

The ADME is the distributional analogue of the popular average partial effects. From (5.2), it is clear that he natural estimator for $\eta_0(t)$ is

$$\hat{\eta}_n(t) = \hat{\beta}_n(t) \frac{1}{n} \sum_{i=1}^n \lambda\left(\mathbf{X}_i'\hat{\theta}_n(t) \right).$$

In what follows, we derive the asymptotic properties of $\hat{\eta}_n(t)$. As in Theorem 4.4, we first show that, uniformly in $t \in \mathcal{T}_0$,

$$(\hat{\eta}_n(t) - \eta_0(t)) = \frac{1}{n} \sum_{i=1}^n \zeta_{i,ADME}(t) + o_p\left(n^{-1/2} \right), \tag{5.3}$$

where

$$\begin{aligned}
\zeta_{i,ADME}(t) = {} & \beta_0(t) \cdot \left(\lambda(\mathbf{X}_i'\theta_0(t)) - \mathbb{E}\left[\lambda(\mathbf{X}'\theta_0(t)) \right] \right) \\
& + \mathbb{E}\left[\lambda(\mathbf{X}'\theta_0(t)) \right] \cdot H \cdot \mathcal{I}_0(t)^{-1} \cdot \zeta_i(t) \\
& + \beta_0(t) \cdot \mathbb{E}\left[\dot{\lambda}\left(\mathbf{X}'\theta_0(t) \right) \mathbf{X}' \right] \cdot \mathcal{I}_0(t)^{-1} \cdot \zeta_i(t)
\end{aligned} \tag{5.4}$$

with $\zeta_i(t)$ is as defined in (4.2), with $\varphi = \psi_{\theta t}$, $\dot{\lambda}(u) = d\lambda(u)/du$, $H \equiv [0_k, I_k]$, 0_k the $k \times 1$ vector of zero, and I_k the k-dimensional identity matrix.

---

[4]To avoid cumbersome notation, we consider the case where the covariates X are continuous, and enter the distribution regression model linearly. The asymptotic validity of all our results do not rely on this simplification.

Let $\mathbb{Z}_{ADME}$ be a centered tight Gaussian element of $\ell^{\infty}\left(\mathcal{T}_0\right)^k$ with matrix of variance and covariance functions

$$\mathbb{E}\left[\mathbb{Z}_{ADME}\left(t_1\right)\mathbb{Z}'_{ADME}\left(t_2\right)\right] = \mathbb{E}\left[\zeta_{ADME}\left(t_1\right)\zeta'_{ADME}\left(t_2\right)\right].$$

Consider the bootstrapped estimator for the ADME

$$\hat{\eta}_n^*\left(t\right) = \hat{\eta}_n(t) + \frac{1}{n}\sum_{i=1}^n V_i \cdot \hat{\zeta}_{i,ADME}(t) \tag{5.5}$$

where *iid* random numbers $\{V_i\}_{i=1}^n$ are *iid* random variables with mean 0 and variance 1, generated independently of the sample, and

$$
\begin{aligned}
\hat{\zeta}_{i,ADME}(t) = {} & \hat{\beta}_n\left(t\right) \cdot \left(\lambda(\boldsymbol{X}_i'\hat{\theta}_n\left(t\right)) - \frac{1}{n}\sum_{j=1}^n \lambda(\boldsymbol{X}_j'\hat{\theta}_n\left(t\right))\right) \\
& + \left(\frac{1}{n}\sum_{j=1}^n \lambda(\boldsymbol{X}_j'\hat{\theta}_n\left(t\right))\right) \cdot H \cdot \widehat{\mathcal{I}}_n\left(t\right)^{-1} \cdot \hat{\zeta}_i(t) \\
& + \hat{\beta}_n\left(t\right) \cdot \left(\frac{1}{n}\sum_{j=1}^n \dot{\lambda}\left(\boldsymbol{X}_j'\hat{\theta}_n\left(t\right)\right)\boldsymbol{X}_j'\right) \cdot \widehat{\mathcal{I}}_n\left(t\right)^{-1} \cdot \hat{\zeta}_i(t)
\end{aligned}
\tag{5.6}
$$

where $\hat{\zeta}_i(t)$ is as in (4.8) and $\widehat{\mathcal{I}}_n\left(t\right)$ is as in (4.9).

THEOREM 5.1. *Suppose that the conditions of Theorem 4.4 hold. Then,*

$$\left\{\sqrt{n}\left(\hat{\eta}_n\left(t\right) - \eta_0\left(t\right)\right)\right\}_{t\in\mathcal{T}_0} \to_d \left\{\mathbb{Z}_{ADME}(t)\right\}_{t\in\mathcal{T}_0} \; in \; \ell^{\infty}\left(\mathcal{T}_0\right)^k.$$

In addition, if follows that $\left\{\sqrt{n}\left(\hat{\eta}_n^* - \hat{\eta}_n\right)\left(t\right)\right\}_{t\in\mathcal{T}_0}$ *converges weakly under the bootstrap law to* $\left\{\mathbb{Z}_{ADME}(t)\right\}_{t\in\mathcal{T}_0}$ *in* $\ell^{\infty}\left(\mathcal{T}_0\right)^k$, *with probability one.*

We now describe a practical bootstrap algorithm to compute simultaneous confidence intervals for the ADME associated with a given covariate X_1

ALGORITHM 5.2. (BOOTSTRAPPED SIMULTANEOUS CONFIDENCE INTERVALS)

STEP *1.* Generate *iid* $\{V_i\}_{i=1}^n$ from a distribution with mean 0, variance 1 and finite third moment, e.g., Rademacher distribution.

STEP *2.* For a grid $t = t_1,\ldots,t_p$, compute $\{\hat{\eta}_n^*\left(t\right)\}_{t=t_1}^{t_p}$ as in (5.5), using the same $\{V_i\}_{i=1}^n$ for all t's.

STEP *3.* Let $\widehat{\kappa}$ and $\widehat{\kappa}^*$ be the vectorized $\{\hat{\eta}_n\left(t\right)\}_{t=t_1}^{t_p}$ and $\{\hat{\eta}_n^*\left(t\right)\}_{t=t_1}^{t_p}$, respectively, denote their *j-th*-element by $\widehat{\kappa}\left(j\right)$ and $\widehat{\kappa}^*\left(j\right)$, and compute $\hat{R}^*\left(j\right) = \left(\widehat{\kappa}^* - \widehat{\kappa}\right)\left(j\right)$.

STEP *4.* Repeat steps 1-3 B times.

STEP *5.* For each bootstrap draw, compute $\bar{R}^* = \max_j \left(\left|\widehat{R}^*\left(j\right)\right|\right)$.

STEP *6.* Construct $\widehat{c}_{1-\alpha}$ as the empirical $(1-a)$-quantile of the B bootstrap draws of $\bar{R}^*$.

STEP *7.* Construct the bootstrapped simultaneous confidence band for $\hat{\eta}_n\left(t\right)$, as $\widehat{CB}\left(t\right) = \left[\hat{\eta}_n\left(t\right) \pm \widehat{c}_{1-\alpha}\right]$.

6. SIMULATION STUDIES

In this section, we briefly summarize the simulation results discussed in Section S1 of the Supplementary Appendix. In short, we compare the finite sample performance of the proposed KMDR estimators with those based on the proportional hazard (PH) model and on the proportional odds (PO) model. We consider three different data generating processes (DGPs): one DGP that satisfies the PH but not the PO assumption, one DGP that satisfies the PO but not the PH assumption, and one DGP where both PH and PO assumptions are violated, though it admits a DR specification with time-varying slope coefficient. We consider different levels of random censoring, and compare the KMDR, PH and PO models based on the conditional CDF, $F_{T|X}$, and the ADME as defined in (5.2). Both functionals have a clear economic interpretation.

Overall, the simulation results highlight that our proposed KMDR estimators performs nearly as well as the PH and PO model when these models are correctly specified, since KMDR nests both PH and PO specifications. On the other hand, when PH and PO models are misspecified, the proposed KMDR estimators performs better than these other popular specifications. Such gains are especially noticeable when one is interested in the $\text{ADME}(t)$; see Table 1. When comparing DR specifications with different link functions, we notice that estimators that use the cloglog link function tend to be more robust against model misspecifications than those based on the logit link function. This is perhaps because the cloglog link function is asymmetric and adapts better to the non-central parts of the distribution where there are many zeros (left tail) or many ones (right tail). Thus, we recommend favoring the cloglog specification in detriment of the logit one, though a more formal discussion about such choices are beyond the scope of this paper.

7. THE EFFECT OF UNEMPLOYMENT BENEFITS ON UNEMPLOYMENT DURATION

One of the main concerns of the design of unemployment insurance policies is their adverse effect on unemployment duration. The prevailing view of the economics literature is that increasing unemployment insurance (UI) benefits leads to higher unemployment duration driven by a moral hazard effect: higher UI increases the agent's reservation wage and reduces the incentive to job search, see, e.g., Krueger and Meyer (2002) and the references therein. Given that moral hazard leads to a reduction of social welfare, this argument has been used against increases in UI benefits.

In a seminal paper, Chetty (2008) challenges the traditional view that the link between unemployment benefits and duration is only because of moral hazard. He shows, among many other things, that distortions cause by UI on search behavior are mostly due to a "liquidity effect". In simple terms, UI benefits provide cash-in-hand that allows liquidity constrained agents to equalize the marginal utility of consumption when employed and unemployed. Such a liquidity effect reduces the pressure to find a new job, leading to longer unemployment spells. However, in contrast to the moral hazard effect, the liquidity effect is a socially beneficial response to the correction of market failures. Thus, if one finds support in favor of liquidity effects, increases in UI benefits may lead to improvements in total welfare.

In this section, we provide additional evidence of the existence of this liquidity effect by comparing the effect of UI on household that are liquidity constrained with those that are unconstrained. In contrast to Chetty (2008), we do not rely on the Cox proportional hazard model but rather use our proposed KMDR tools. In this context, the Cox

Table 1. Simulated finite-sample properties when DGP does not satisfy neither PH nor PO assumptions.

	n	No Censoring				10 % Censoring				30% Censoring			
		PH	$P0$	DR_{cll}	DR_l	PH	$P0$	DR_{cll}	DR_l	PH	$P0$	DR_{cll}	DR_l
Average	100	0.62	0.59	0.35	0.65	0.63	0.72	0.36	0.69	0.58	0.78	0.98	1.24
absolute bias	400	0.85	0.69	0.08	0.41	0.73	0.64	0.12	0.45	0.67	0.59	0.25	0.55
for $F(t\|X=0.5)$	1600	0.91	0.77	0.03	0.34	0.81	0.73	0.03	0.34	0.61	0.62	0.15	0.46
Average	100	4.29	4.42	4.29	4.32	4.42	4.58	4.55	4.57	4.86	4.97	5.36	5.35
RMSE	400	2.34	2.41	2.12	2.20	2.42	2.52	2.28	2.36	2.57	2.64	2.54	2.62
for $F(t\|X=0.5)$	1600	1.46	1.48	1.08	1.19	1.41	1.47	1.09	1.20	1.40	1.50	1.25	1.38
Average	100	17.13	18.09	0.29	0.70	16.50	18.31	0.56	0.74	16.16	19.15	3.14	2.85
absolute bias	400	17.44	18.42	0.14	0.67	16.72	18.68	0.50	0.58	16.15	19.42	1.40	1.07
for $ADME(t)$	1600	17.44	18.53	0.07	0.51	16.78	18.72	0.06	0.53	16.19	19.56	0.76	0.67
Average	100	21.51	23.32	14.69	15.10	21.05	23.39	16.00	16.33	21.42	24.29	20.60	20.57
RMSE	400	18.84	20.02	7.09	7.37	18.21	20.31	7.46	7.75	17.86	21.06	9.62	9.87
for $ADME(t)$	1600	17.90	19.04	3.56	3.79	17.26	19.23	3.77	4.01	16.71	20.10	4.73	4.89

Note: Simulations based on one thousand Monte Carlo experiments. "PH" stands for estimators based on the proportional hazard model. "PO" stands for estimators based on the proportional odds model. "DR_{cll}" and "DR_l" stand for estimators based on the proposed distribution regression mode with the *cloglog* and *logit* link functions, respectively.

hazard model might be too rigid as does not allow for the effect of UI benefits to vary depending on whether a worker is starting their unemployment claim or they have been unemployed for a while. As we argued before, our proposed tools allow for this richer types of heterogeneity.

As in Chetty (2008), our data comes from the Survey of Income and Program Participation (SIPP) for the period spanning 1985-2000. Each SIPP panel surveys households at four-month intervals for two-four years, collecting information on household and individual characteristics, as well as employment status. The sample consists of prime-age males who have experienced job separation and report to be job seekers, are not on temporary layoff, have at least three months of work history in the survey and took up unemployment insurance benefits within one month after job loss. These restrictions leave 4,529 unemployment spells in the sample, 21.3% of those being right-censored. Unemployment durations are measured in weeks while individuals' UI benefits are measured using the two-step imputation method described in Chetty (2008). For further details about the data, see Chetty (2008).

To analyze the effect of UI benefits on unemployment duration, we estimate the DR model

$$F_{T|UI,Z}(t|UI,Z) = \Psi\left(\alpha(t) + \beta(t)\ln UI + Z'\gamma(t)\right), \tag{7.1}$$

where $\Psi(u) = 1 - \exp\left(-\exp(u)\right)$ is the complementary log-log link function, UI is the worker's weekly unemployment insurance benefits, and Z is a vector of controls including the worker's age, years of education, marital status dummy, logged pre-unemployment annual wage, and total wealth. To control for local labor market conditions and systematic differences in risk and performance across sector types, Z also includes the state average unemployment rate and dummies for industry. We note that Chetty (2008) considers additional controls, including state, year and occupation fixed effects, resulting in a specification with almost 90 unknown parameters. For this reason, we adopt a more parsimonious specification described above. Let $\theta(t) = \left(\alpha(t), \beta(t), \gamma(t)'\right)'$, and $\mathbf{X} = (1, \ln UI, Z')'$.

Our main goal is to understand the effect of changes in UI on the probability of one finding a job in t weeks, where $t = 2, 3, \ldots 50$. Although the sign of $\beta(t)$ indicates the direction of such a change, its magnitude may not have a straightforward economic

interpretation. Thus, we focus on the ADME(t) of $\ln UI$,

$$ADME_{\ln UI}(t) = \mathbb{E}\left[\frac{\partial F_{T|UI,Z}(t\,|\,UI,Z)}{\partial \ln UI}\right], \tag{7.2}$$

which is easy to interpret. For instance, an estimate of $ADME_{lnUI}(t)$ equal to 0.1 suggests that, on average, raising unemployment benefits by one percent increases the probability of finding a job in t weeks or less by 0.1 percentage point. As discussed in Theorem 3, we can estimate $ADME_{\ln UI}(t)$ by

$$\widehat{ADME}_{n,\ln UI}(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_n(t)\cdot\psi\left(\mathbf{X}_i'\hat{\theta}_n(t)\right), \tag{7.3}$$

where $\psi(u) = d\Psi(u)/du$, and $\hat{\theta}_n(t) = \left(\hat{\alpha}_n(t), \hat{\beta}_n(t), \hat{\gamma}_n(t)'\right)'$ is the KMDR estimator of $\theta(t)$. When Ψ is the cloglog link function, $\psi(u) = \exp(u)\exp(-\exp(u))$.

In what follows, we multiply the estimators of $ADME_{\ln UI}(t)$ by 10, which is interpreted as the effect of a 10% increase in UI benefits. Figure 1 reports the estimates for the full sample (solid line), together with the 90 percent bootstrapped pointwise, and simultaneous confidence intervals (dark and light shaded area, respectively) computed using Algorithm 5.2.

The result reveals interesting effects. On average, a 10% increase in UI benefits appears to have no effect on the probability of a worker finding a job in the first seven weeks of the unemployment spell. Nonetheless, Figure 1 shows that a change in UI is associated with a reduction of the probability of finding a job in the first $t = \{8,\ldots,50\}$ weeks. Such an effect seems to be monotone until week 18, where a 10% increase in UI is associated with a two-percentage-point decrease in the average probability of finding a job until that week. After week 18, the effect of an increase in UI benefits on employment probabilities seems to weaken but remains statistically significant at the 10% level, except for $t = \{34, 35, 39\}$. Note that the bootstrap simultaneous confidence interval is slightly wider than the pointwise one. However, it is important to mention that the bootstrap uniform confidence interval is designed to contain the entire true path of the $ADME_{\ln UI}(t)$ 90% of the time, which is in sharp contrast to the bootstrap pointwise confidence interval. This highlights the practical appeal of using simultaneous instead of pointwise inference procedures to better quantify the overall uncertainty in the estimation of all ADMEs. Overall, Figure 1 shows that although an increase in UI benefits is close to zero at the beginning of the unemployment spell, they have an U-shaped effect on the unemployment duration distribution. Interestingly, estimates of $ADME_{\ln UI}(t)$ based on the Cox PH model with the same set of covariates as in (7.1) suggests that the effect of UI on unemployment duration distribution is monotone for $t \in [0, 50]$. However, once the proportional hazard specification is tested using either Grambsch and Therneau (1994) procedure or our bootstrap-based testing procedure for the null hypothesis that all $\hat{\theta}_n(t), t = \{2,\ldots,50\}$ are constant in t, as described in Section 5.1, the null of proportionality is rejected at the usual significance levels[5], implying that, indeed, a proportional hazard model may not be appropriate for this application. Our KMDR model does not rely on such an assumption.

Although the results in Figure 1 show that, on average, changes in UI have a U-shaped effect on the unemployment duration distribution, the analysis remains silent about the

---

[5]The p-value associated with our proposed test statistics is 0.004.
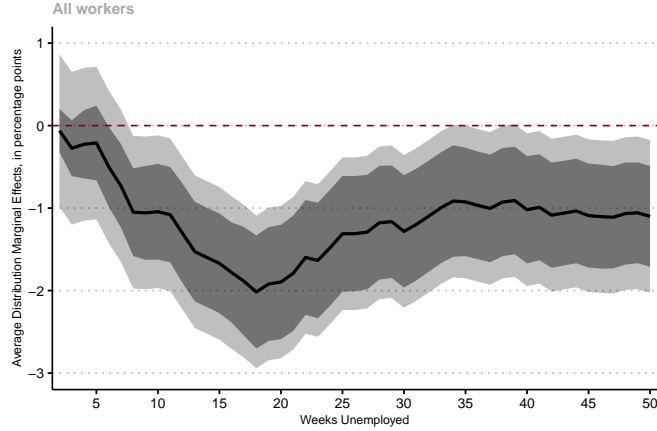
Figure 1. Estimated average effect of a 10% increase of unemployment benefits on the distribution of unemployment duration, full sample. The solid line represents the estimates while the dark and light shaded areas are the 90% bootstrapped based pointwise and simultaneous confidence interval based on 100,000 bootstrap draws, respectively. Sample size is 4,307.

liquidity effects. To shed light on the importance of the liquidity effect relative to the moral hazard effect, Chetty (2008) argues that one can compare the response to an increase in UI benefits of workers who are not financially constrained with those who are constrained. Given that unconstrained workers have the ability to smooth consumption during unemployment, liquidity effects are absent and UI benefits lengthen unemployment duration only via moral hazard effects for these subgroup of individuals. To pursue this logic, we follow Chetty (2008) and use two proxy measures of liquidity constraint: liquid net wealth at the time of job loss ("net wealth") and an indicator for having to make a mortgage payment. Chetty (2008) argues that workers with higher net wealth are less sensitive to UI benefit levels because they are less likely to be financially constrained. Similarly, workers that have to make mortgage payments before job loss have less ability to smooth consumption during unemployment because they are unlikely to sell their homes during the unemployment spell, whereas renters can adjust faster. To assess the role of liquidity effects, we divide the entire sample into four subsamples: (a) workers with net wealth below the median, (b) workers with net wealth above the median, (c) workers with a mortgage, and (d) workers without a mortgage. For each subsample, we estimate (7.1) using the KMDR approach, its corresponding estimator of $\text{ADME}_{\ln UI}(t)$, $\widehat{ADME}_{n,\ln UI}(t)$ as in (7.3), and construct 90% bootstrap pointwise and simultaneous confidence intervals.

Figure 2 reports the average effects of a 10% increase in UI benefits on the unemployment duration distribution in each of the four subsamples. The solid lines are the point estimates and the dark and light shaded areas are the 90% bootstrap pointwise and simultaneous confidence intervals, respectively.

The results show interesting heterogeneity of UI benefits effects with respect to liquidity constraint proxies. For those workers with net wealth above the median, an increase in UI benefits has no statistically significant effect on the unemployment duration distri-
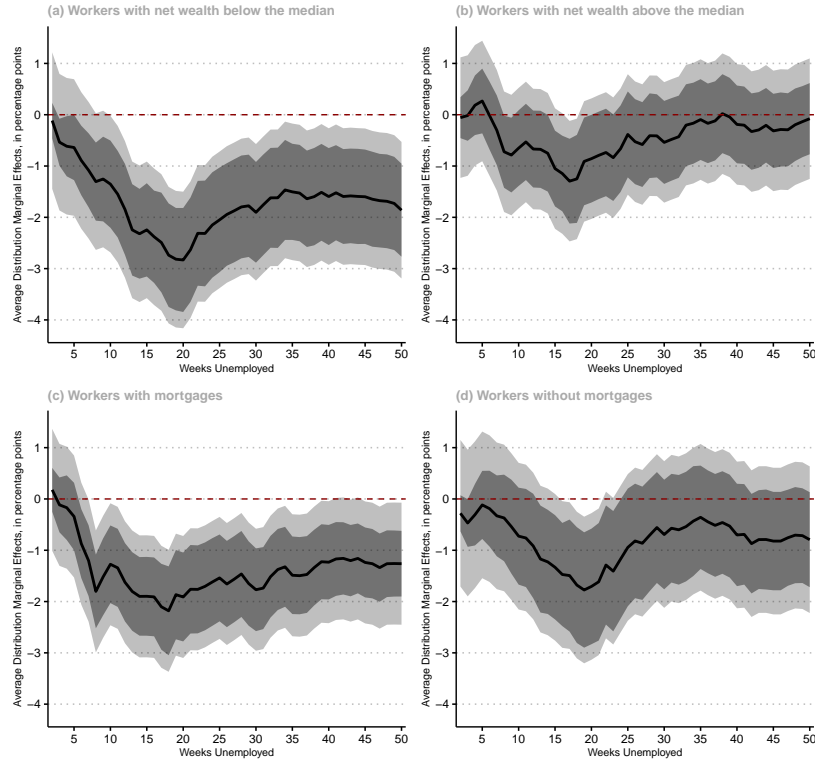
Figure 2. Estimated average effect of a 10% increase of unemployment benefits on the distribution of unemployment duration, in different subpopulations. The solid line represents the estimates while the dark and light shaded areas are the 90% bootstrapped based pointwise and simultaneous confidence interval based on 100,000 bootstrap draws, respectively. Sample sizes for Panel (a)-(d) are 2152, 2155, 1952 and 2355, respectively.

bution, except for $t = \{17, 18\}$. Thus, for those workers who are not constrained (and, therefore, for whom the liquidity effect is approximately zero), the moral hazard effect seems to be close to zero. On the other hand, for those workers with net wealth below the median, an increase in UI benefits is associated with lower probabilities of finding a job. The conclusion using mortgage as a proxy for liquidity constraint is qualitatively the same. Note that our analysis suggests that the ADME of an increase of UI is non-monotone across the unemployment duration distribution, highlighting the flexibility of the DR approach. Indeed, using our proposed test for all slope coefficients being constant, the proportional hazard specification is rejected at the 10% significance levels in each subsample.

In other to further highlight that the effect of UI benefits on unemployment duration is different depending on whether a worker is likely to be liquidity constrained or not, in Figure 3 we plot the difference of the estimated ADME between liquidity constrained and not-liquidity constrained workers, together with their pointwise and simultaneous confidence bands. Panel (a) suggests that UI benefits have a more negative effect on workers with net wealth below the median than on wealthier workers, and that this

difference is statistically significant at the 10% level. Panel (b) reveals a qualitatively similar pattern when one compared workers with mortgage with those without mortgage, though the magnitude of this difference is less pronounced (but we can still reject the null hypothesis that the effects are the same, at the 10% significance level.)
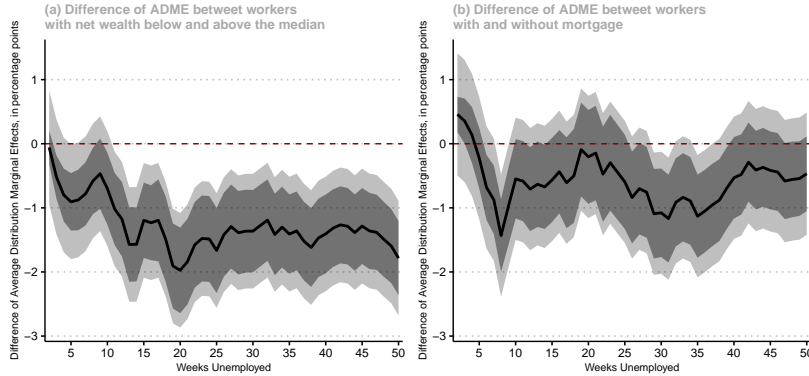

Figure 3. Difference of estimated average effect of a 10% increase of unemployment benefits on the distribution of unemployment duration among liquidity constrained and unconstrained workers . The solid line represents the estimates while the dark and light shaded areas are the 90% bootstrapped based pointwise and simultaneous confidence interval based on 100,000 bootstrap draws, respectively. Sample size for Panel (a) and (b) is 4,307.

Taken together, our results provide suggestive evidence that UI benefits have a non-monotone effect on the unemployment duration distribution and that such an effect varies whether workers are likely to be liquidity constrained or not. More precisely, our results suggest that the effect of UI on unemployment duration is larger for liquidity constrained workers. Through the lens of the results of Chetty (2008), our findings suggest that an increase in UI benefits affects unemployment duration not only through moral hazard but also because of a "liquidity effect."

ACKNOWLEDGEMENTS

REFERENCES

Aalen, O. O. (1980). Lecture Notes in Statistics – Proceedings. In W. Klonecki, A. Kozek, and J. Rosinski (Eds.), *Lecture Notes in Statistics*, Volume 2, pp. 1–25. New York: Springer.

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

Arcones, M. A. and E. Giné (1993, jul). Limit Theorems for U-Processes. *The Annals of Probability 21*(3), 347–370.

Arcones, M. A. and E. Giné (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Journal of Theoretical Probability 58*, 217–245.

Armstrong, B. G. and M. Sloan (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology 129*, 191–204.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine 2*, 273–277.

Chen, K. and S.-H. Lo (1997). On the rate of uniform convergence of the product-limit estimator: strong and weak laws. *Annals of Statistics 25*, 1050–1087.

Chen, K. and Z. Ying (1996). A counterexample to a conjecture concerning the Hall-Wellner band. *Annals of Statistics 24*, 641–646.

Cheng, S. C., L. J. Wei, and Z. Ying (1995). Analysis of transformation models with censored data. *Biometrika 82*, 835–845.

Chernozhukov, V., I. Fernández-Val, and S. Luo (2019). Distribution Regression with Sample Selection, with an Application to Wage Decompositions in the UK. *arXiv preprint arXiv: 1811.11603*, 1–50.

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

Chernozhukov, V., I. Fernández-Val, and M. Weidner (2020). Network and Panel Quantile Effects Via Distribution Regression. *Journal of Econometrics Forthcoming*, 1–28.

Chetty, R. (2008). Moral Hazard versus Liquidity and Optimal Unemployment Insurance. *Journal of Political Economy 116*, 173–234.

Clayton, A. D. G. (1976). An Odds Ratio Comparison for Ordered Categorical Data with Censored Observations. *Biometrika 63*, 405–408.

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

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 34*, 187–220.

Doksum, K. and M. Gasko (1990). On a correspondence between models in binary regression and survival analysis. *International Statistical Review 58*, 243–252.

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

Grambsch, P. and T. Therneau (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika 81*, 515–526.

Guo, S. and D. Zeng (2014). An overview of semiparametric models in survival analysis. *Journal of Statistical Planning and Inference 151-152*, 1–16.

Honore, B., S. Khan, and J. Powell (2002). Quantile regression under random censoring. *Journal of Econometrics 109*, 67–105.

Hothorn, T., T. Kneib, and P. Bühlmann (2014). Conditional transformation models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*, 3–27.

Kalbfleisch, J. D. and R. L. Prentice (1980). *The statistical analysis of failure time data* (2nd ed.). Hoboken, NJ: Wiley.

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

Krueger, A. B. and B. D. Meyer (2002). Labor supply effects of social insurance. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*, Volume 4, Chapter 33, pp. 2327–2392. Amsterdam: North-Holland: Elsevier.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological) 42*, 109–142.

Murphy, S. A., A. J. Rossini, and A. W. VanderVaart (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association 92*, 968–976.

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

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

Robins, J. M. and A. Rotnitzky (1992). Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. In *AIDS Epidemiology*, pp. 297–331. Boston, MA: Birkhäuser Boston.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association 90*, 122–129.

Rothe, C. and D. Wied (2013). Misspecification Testing in a Class of Conditional Distributional Models. *Journal of the American Statistical Association 108*, 314–324.

Rothe, C. and D. Wied (2019). Estimating Derivatives of Function-Valued Parameters in a Class of Moment Condition Models. *Journal of Econometrics Forthcoming*, 1–19.

Sant'Anna, P. H. C. (2020). Nonparametric Tests for Treatment Effect Heterogeneity with Duration Outcomes. *Journal of Business and Economic Statistics Forthcoming*, 1–17.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis 45*, 89 – 103.

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

Stute, W. (1995). The central limit theorem under random censorship. *Annals of Statistics 23*, 422–439.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics 23*, 461–471.

Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica 9*, 1089–1102.

Stute, W., W. González-Manteiga, and C. . Sánchez Sellero (2000). Nonparametric model checks in censored regression. *Communications in Statistics - Theory and Methods 29*, 1611 – 1629.

Tsiatis, A. A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America 72*, 20–22.

Tsiatis, A. A. (1981). A large sample study of Cox's regression model. *Annals of Statistics 9*, 93–108.

van den Berg, G. (1990). Nonstationarity in Job Search Theory. *Review of Economic Studies 57*, 255–277.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

Wang, H. J. and L. Wang (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association 104*, 1–32.

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics 141*, 1281–1301.

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