Endogenous Censoring in the Mixed Proportional Hazard Model with an Application to Optimal Unemployment Insurance [*]

Arkadiusz Szydłowski[†]
University of Leicester

May 16, 2019

Summary

We examine the sensitivity of estimates of the MPH model with respect to assumptions on the censoring mechanism in the context of an economic model of optimal unemployment insurance. We assume a parametric model for the duration of interest and leave the distribution of censoring unrestricted, allowing it to be correlated with observed and unobserved characteristics. We provide a practical characterization of the identified set with moment inequalities and suggest methods for estimating this set. We apply this approach to estimate the elasticity of unemployment exit rate with respect to unemployment benefit. Finally, we investigate welfare consequences of our estimates.

Keywords: Duration model, Partial identification, Unemployment insurance

1 Introduction

Duration models are a useful tool for analyzing the relationship between time spent in some state and observed characteristics. In practice data on the duration of interest are frequently censored. The standard approach in duration analysis is to assume that censoring is independent of unobserved heterogeneity (existing results allow censoring to be correlated with observed characteristics). The objective of this paper is to examine the sensitivity of parameter estimates to this independence assumption in the context of an economic model of optimal unemployment insurance.

Firstly, we discuss identification and provide methods for estimating the mixed proportional hazard (MPH) duration model when the distribution of censoring is left unrestricted. We show that if no assumptions are made on the censoring process, the parameters of the model are set- and not point-identified. We provide a practical characterization of the identified set and suggest a procedure to estimate confidence sets for the parameters of interest based on Kaido et al. (2019) (see Bugni et al. (2017) for alternative profiling procedure). For practical reasons, our inference procedure maintains a parametric form for the baseline hazard and the distribution of unobserved heterogeneity.

Secondly, we provide new set estimates of the elasticity of the exit rate from unemployment with respect to unemployment benefit from models that do not fully restrict the distribution of censoring but use a parametric model for the unemployment duration. These estimates are robust to violations of independence between censoring and unobserved heterogeneity. In particular, it is possible that the true parameter value does not lie in the confidence interval constructed under the assumption of independent censoring but it is contained in our confidence set (in Section 2.2 we show an example where this arises).

The elasticity of the unemployment exit rate with respect to unemployment benefit is of interest in the economic analysis of optimal unemployment insurance. Chetty (2008) shows that the welfare consequences of a change in unemployment benefit can be derived from a small set of estimated parameters among which this elasticity plays a crucial role. Thus,

the estimate of this elasticity can be used in conjunction with Chetty's welfare formula to judge if the current level of unemployment benefits in the US is optimal. We estimate this elasticity using a sample of unemployed persons from the Survey of Income and Program Participation (SIPP). In our baseline specification we find that a 10% increase in benefit leads to between 2.2% and 11.5% drop in unemployment exit rate.

When combined with the welfare formula, our baseline set estimates do not permit us to put an upper bound on the size of the welfare change due to an increase in the unemployment benefit. However, our tightest bounds, coming from a model that restrict the dependence between censoring and unemployment duration to half the level of dependence between censoring and observed covariates, imply that a hike in benefit level would have a modest negative effect on GDP in the US. Overall we conclude that given the SIPP data and the available theoretical formula one cannot credibly judge if the unemployment benefits in the US are close to the optimal level. Chetty (2008) uses his point estimates to deduce that the welfare gains from increasing unemployment benefits would be positive but rather small, which implies that the benefits in the US are set close to the optimal level. Allowing for correlated censoring, the empirical results are not as informative about the optimality of unemployment benefits.[1]

Our model can be interpreted as a competing risks model with dependent risks, where the risks are censoring and, for example, exit from unemployment. In this view there are several alternatives to our approach. Firstly, one does not need to pose any model for either unemployment duration nor censoring, in particular one can drop the proportionality assumption underlying the MPH model. Peterson (1976) showed that one can obtain informative bounds on the distributions of the risks without parametric assumptions on the joint distribution. However, these bounds are usually very wide (see Peterson (1976), Honoré and Lleras-Muney (2006)), which makes this approach unattractive to applied researchers.

Another possible choice is to assume mixed proportional hazards for both risks (exit

---

[1]Though, these results are in line with Chetty's results when interval estimates (incorporating standard errors) are used.

from unemployment and censoring), see e.g. Van den Berg et al. (1994). This restores point-identification under support conditions on the explanatory variables (cf. Heckman and Honoré (1989), Abbring and Van den Berg (2003)). However, justifying a parametric model for the censoring variable may be problematic.

Several authors have considered dependent censoring using partial identification approach in related models or under different assumptions. Partial identification of a log-linear duration model with dependent censoring under median restriction was analyzed by Khan and Tamer (2009). However, they proceed with inference assuming that their condition for point-identification is satisfied. Khan et al. (2016) investigate consequences of endogenous censoring in a panel data model under stationarity or conditional independence restrictions. They use a stochastic dominance test for inference and, unlike our article, do not consider profiled inference. Honoré and Lleras-Muney (2006) analyze an AFT competing risks model with interdependent risks and a binary covariate. Thus, they relax the assumption on the support of covariates required for point-identification (Heckman and Honoré (1989)) but still pose a semiparametric model for both risks. In this paper we do not require covariates to be continuously distributed nor put any semiparametric restrictions on the distribution of one of the risks.

The article is organized as follows. Section 2 develops the moment inequalities delineating the identified set. Section 3 discusses the inference procedure that uses these inequalities to obtain a confidence set for the parameters of interest. The Monte Carlo study in Section 4 verifies that our method works in practice and assists us in picking the tuning sequence needed for the application of the inference procedure. In Section 5 we apply the previously developed method to build a confidence set for the elasticity of unemployment exit rate with respect to the unemployment benefit and use this estimate to analyze optimal unemployment insurance. All proofs are relegated to the online appendix (Szydlowski (2019)).

2 Identification

This section provides moment conditions that partially identify the parameters of the mixed proportional hazard (MPH) model given the joint distribution of covariates and censored spells. In MPH model the hazard rate for a person who has stayed unemployed for y weeks is given by:

$$\lambda(y|X_i, V_i) = \lambda_0(y)e^{X_i'\beta}V_i \tag{1}$$

where $\lambda_0(\cdot)$ denotes the baseline hazard, X_i is a $[K \times 1]$ vector of observed characteristics (including a constant term) and V_i is a scalar unobserved heterogeneity term. Before proceeding to identification under endogenous censoring, it will be instructive to discuss identification of the single-spell mixed proportional hazard duration model without censoring.

The identification of the MPH model in (1) has been investigated by Elbers and Ridder (1982), Heckman and Singer (1984), Horowitz (1999) and Ridder and Woutersen (2003). These papers differ in the assumptions they impose on the distribution of V and the baseline hazard as well as types of normalizations used (see Hausman and Woutersen (2014), Abbring and Ridder (2015) for a review).

Let $\mathcal{X} \subset R^K$ denote the support of X_i and Λ_0 denote the integrated hazard. In this paper we make the following assumptions:

Assumption 2.1. *(a)* V_i *is a non-negative random variable with c.d.f.* F_v *and* $V_i \perp X_i$,

(b) $\mathcal{X}$ *is not contained in any proper linear subspace of* R^K,

(c) $\Lambda_0 : [0, \infty) \to [0, \infty)$ *is nondecreasing and differentiable almost everywhere,* $\Lambda_0(0) = 0$,
$\lim_{y \to \infty} \Lambda_0(y) = \infty$,

(d) $E(V_i) = 1$.

These assumptions correspond to Elbers and Ridder (1982). Under Assumptions 2.1(a)-

(d) and no censoring (Λ_0, β, F_v) are identified.[2]

2.1 Identification with arbitrary censoring

In practice we do not observe a full spell $\tilde{Y}_i$ for each person but rather a censored spell:

$$Y_i = \min\{\tilde{Y}_i, C_i\}$$

where C_i is the censoring variable. All the existing approaches assume independence between censoring C_i and unobserved heterogeneity V_i. However, it is often hard to justify this assumption. We provide several examples where this condition fails.[3]

Example 1 (survey attrition): Consider the 1996 panel of the Survey of Income and Program Participation (SIPP), which is a part of our estimation sample. As reported by Slud and Bailey (2006) 30% of individuals in the initial sample left the survey by the final wave of the interviews. Moreover, Bailey (2004), Table 3.2, shows that people leaving the survey are systematically different with respect to observed characteristics than responders. It is likely that unemployed leaving the survey differ from the remaining unemployed also with respect to unobserved characteristics. For example, individuals may fail to complete the survey because of alcohol or drug problems. These problems will also affect their chances to find a job. In other words, the same unobserved characteristics will affect attrition and unemployment duration, which violates independence between C_i and V_i.

Example 2 (administrative unemployment data): Independence between censoring and unobserved characteristics is questionable in the studies of unemployment duration based on administrative data (see e.g Meyer (1990)). With this type of data, we observe the unemployed person only as long as she receives benefits. Therefore, the unemployed who use the full length of benefits are no longer observed and their unemployment spells are censored.

---

[2]The same identification argument holds also for a model with completely random censoring and censoring correlated with observed covariates.

[3]Further examples are given in Szydlowski (2019).

Those who do not exhaust their benefits may also be censored if they do not accept a proposed job offer or refuse to participate in a reemployment services program. Moreover, in the US an unemployed person can often extend her benefits beyond the standard period at the cost of subjecting herself to stricter job search requirements, e.g. contacting a specific number of employers every week, reporting search effort etc. The extension can be canceled at any time if the person fails to satisfy these requirements (see e.g. IDES (2017)). In all these cases one can expect that the observed benefit period would be affected by unobserved characteristics like motivation to find a job or search skills. This would violate the assumption that censoring is independent of unobserved characteristics.[4]

These examples show that independent censoring implicitly entails strong economic assumptions. Thus, in this section we relax the assumption of independence between censoring and unobserved heterogeneity in the mixed proportional hazard model. Let $D_i = \mathbb{1}\{\tilde{Y}_i \leq C_i\}$ indicate observations which are not censored.

Although our identification results carry through to the general MPH model where Λ_0 and F_v are treated nonparametrically, for the ease of exposition, in what follows we will assume that the distribution of unobserved heterogeneity and the baseline hazard are known up to finitely dimensional parameters, i.e. $F_v(\cdot) = F_v(\cdot; \gamma), \gamma \in R^{d_\gamma}$ and $\Lambda_0(\cdot) = \Lambda(\cdot; \alpha), \alpha \in R^{d_\alpha}$.

Denote $\theta = (\alpha, \beta, \gamma)$ and let Θ be a subset of R^{d_θ}, $d_\theta = d_\alpha + d_\gamma + K$. Define $F_{vc}^x(v, c|x)$ to be the joint cumulative distribution function of (V, C) conditional on $X = x$. Let $\mathcal{F} = \{F_{vc}^x : E[V] = 1, F_v(\cdot) = F_v(\cdot; \gamma), \gamma \in R^{d_\gamma}\}$ (note that we do not assume that the joint distribution is continuous to allow the case of fixed censoring). Let $p(y, d, x) = p(Y = y, D = d|X = x)$ where p is the true density generating the data and $p_\theta(y, d, x; F_{vc}^x)$ denotes the respective density generated from the censored parametric MPH model with dependence between V, C and X governed by F_{cv}^x. The identified set for the true value θ_{true} is the set of all θ's for which there exists a cumulative distribution function $F_{cv}^x \in \mathcal{F}$ such that the model probabilities are

---

[4]This will not be a problem if the researcher is willing to censor all the spells at 26 weeks or lower. The existing tools will work well in such case.

consistent with the true probabilities, i.e.

$$\Theta_I = \{\theta : p(y, d, x) = p_\theta(y, d, x; F_{vc}^x) \text{ for some } F_{vc}^x \in \mathcal{F} \text{ and all } (y, d, x)\}.$$

In other words, we allow any form of statistical dependence between censoring C and unemployment duration $\tilde{Y}$ as long as the marginal distribution of $\tilde{Y}$ follows our MPH model.

Alternatively, the identified set can be described as the set of maximizers of the log-likelihood:

$$\Theta_I = \arg\max_{\theta \in \Theta} \sup_{F_{vc}^x \in \mathcal{F}} E[\log p_\theta(y, d, x; F_{vc}^x)],$$

where the expectation E is taken with respect to the true probability p. The latter characterization falls in the class of models analyzed by Chen et al. (2011), who suggest a sieve likelihood ratio test for doing inference. However, implementing their procedure in our context is difficult since it requires approximating the CDF F_{vc}^x by sieves. This function has $K + 2$ arguments, thus the number of sieve coefficients will be large and the resulting estimates may have a large bias in finite samples. Instead, we aim at providing an alternative characterization of the identified set that leads to a simpler inference procedure.

Let

$$S(y|x) = P(Y_i > y|X = x) \qquad \text{and} \qquad \tilde{S}(y|x) = P(\tilde{Y}_i > y|X = x)$$

denote the survival functions for observed and latent spells and define:

$$S^u(y|x) = 1 - E[D_i \mathbb{1}\{Y_i \le y\}|X = x].$$

Then, we have $Y_i \leq \tilde{Y}_i \leq D_i Y_i + (1 - D_i)\infty$, which implies:[5]

$$S(y|x) \leq \tilde{S}(y|x) \leq S^u(y|x). \tag{2}$$

These inequalities provide a starting point for deriving the moment inequalities which characterize the identified set. Let

$$\mathcal{L}_v(s; \gamma) = \int_0^\infty e^{-sv} dF_v(v; \gamma)$$

denote the Laplace transform of the distribution F_v. We have:

$$\tilde{S}(y|x) = \mathcal{L}_v(\Lambda(y; \alpha_{true})e^{x'\beta_{true}}; \gamma_{true}). \tag{3}$$

which, together with (2), implies that Θ_I is contained in a set defined by a collection of moment inequalities. Additionally, note that $S^u(y|x) - \tilde{S}(y|x) = P(Y_i \leq y, D_i = 0 | X = x)$, which implies that the right-hand side of this expression has to be non-decreasing in y. Proposition 1 shows that adding this requirement to inequalities in (2) is enough to provide sharp characterization of the identified set in our model.[6]

Proposition 1. *Let Θ_0 be the set of $\theta = (\alpha, \beta, \gamma)$'s satisfying:*

$$\mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta}; \gamma) - S(y|x) \geq 0; \tag{P1}$$

$$S^u(y|x) - \mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta}; \gamma) \geq 0; \tag{P2}$$

$$S^u(y_1|x) - \mathcal{L}_v(\Lambda(y_1; \alpha)e^{x'\beta}; \gamma) \geq S^u(y_2|x) - \mathcal{L}_v(\Lambda(y_2; \alpha)e^{x'\beta}; \gamma); \tag{M}$$

$$\forall y \in [0, \infty), x \in \mathcal{X}; \quad \forall y_1, y_2 \in [0, \infty), y_1 \geq y_2, x \in \mathcal{X}$$

and suppose Assumption 2.1 holds. Then, if $P(C_i \leq y | D_i = 0, X_i = x)$ is continuous and

---

[5]We assume $0 \times \infty = 0$ here.

[6]I am grateful to an anonymous referee for suggesting these additional inequalities and the argument leading to this sharpness result.

strictly increasing for $y \in (0, C^{max}), C^{max} > 0$:

$$\Theta_I = \Theta_0$$

Note that the moment inequalities (P1)-(P2) correspond to the bounds in Peterson (1976). His bounds are nonparametric - he does not pose a parametric model for either of the risks (employment and censoring). In that model the extreme joint distributions of $(\tilde{Y}, C)$ generating the bounds automatically satisfy condition (M). With a parametric model that does not have to be the case anymore as we are restricting the marginal distribution of $\tilde{Y}$ to the MPH class, which may make Peterson bounds not sharp. Thus, additional inequalities in (M) are needed to guarantee sharpness. When evaluated at θ_{true} the left hand side of (M) becomes $P(\tilde{Y} \leq y_1, D = 0 | X = x)$, i.e. the distribution of unobserved part of $\tilde{Y}$, so this inequality excludes joint distributions of $(\tilde{Y}, C)$ which are consistent with MPH model for $\tilde{Y}$ but at the cost of violating monotonicity in the unobserved part. Proof of Proposition 1 shows that this is enough to guarantee sharpness.

It is natural to ask under which conditions the identified set shrinks to a point. Proposition 2 provides a set of sufficient conditions.

Proposition 2. *If there exist $\epsilon > 0$ such that the set*

$$\mathcal{X}_{ID} = \{x \in \mathcal{X} : P(C_i \leq \epsilon, D_i = 0 | X_i = x) = 0\}$$

with $P(x \in \mathcal{X}_{ID}) > 0$ satisfies Assumption 2.1(b) and $\Lambda(\cdot; \alpha_{true})$ is an analytic function on $(0, \infty)$, then we have $\Theta_0 = \{\theta_{true}\}$.

Proposition 2 basically requires that short durations are never censored for some values of covariates, for example, short-term unemployed with large families never drop out of the survey before the end of their spell. In particular, point identification obtains in the case of fixed censoring and when censoring depends only on the observed covariates. The require-

ment for the integrated baseline hazard function to be analytic is satisfied by the Weibull specification used in our application or, for example, by any polynomial approximation.[7]

Note that Proposition 1 proves sharpness when censoring is continuously distributed with positive density everywhere on $(0, C^{max})$ (a sufficient condition for condition in Proposition 1 under our MPH model). When C has a discrete distribution with a finite number of points of support, Proposition 2 shows that the parameters are point-identified for most models of interest and, thus, naturally the set is sharp as well.

For given y and x the inequalities (P1)-(P2) in Proposition 1 can be rewritten as:

$$S(y|x) \leq \mathcal{L}_v(\Lambda(y;\alpha)e^{x'\beta};\gamma) \leq S^u(y|x).$$

We can see that the identified set Θ_0 is an intersection of the areas between two level sets of the function $\mathcal{L}_v(\Lambda(y;\alpha)e^{x'\beta};\gamma)$ for different values of y. This function is nonlinear, so the areas between the level sets will not be convex and convexity of Θ_0 is not guaranteed (additional inequalities in (M) do not help here).

2.2 Shape of the identified set

In this section we analyze an example MPH model with dependent censoring to get some insight about the shape of the identified set and potential consequences of assuming independent censoring (here, as in the rest of the article, this means censoring independent of unobserved heterogeneity, thus allowing for covariate-dependent censoring). Moreover, our inference procedure will require discretizing the support of X.[8] Since the moment inequalities are indexed by values of X, a reduction in the number of points of support of X will lead

---

[7]The analyticity assumption can be dropped if one is only interested in identifying the baseline hazard on $[0, \epsilon]$. Also, even if we cannot fully identify the baseline hazard, the coefficient β can be identified from knowledge of the survival function around zero.

[8]We also discretize the evaluation points for y but this is of less concern as the main cost of a finer grid for y is computational. Thus, if one is concerned about the loss of identifying power due to a finite number of evaluation points, one could introduce more points at the cost of longer computation. A finer grid for X means both longer computing times and less precision in estimating conditional CDFs.

to a decrease in the number of moment conditions and, in general, will increase the identified set. Thus, we also take a chance here to investigate how the coarseness of discretization will affect the size of Θ_0.

We investigate the following MPH model:

$$\alpha \log \tilde{Y} = -X\beta - \log V + \log U$$

$$\alpha \log C = c + \log V$$

where V, U have the unit exponential distribution (which implies $\gamma = 1$) and are mutually independent. We set $\alpha = 1.5, \beta = -0.5, c = 2.5$ and impose an upper bound on the observed durations equal to 5. This guarantees a censoring rate around 22%. The model implies that the lower the unobserved "ability" (V) the longer the unemployment spell and the more likely the spell will be censored, which is in line with the intuition that people with low ability or motivation will exit the unemployment records sooner than highly motivated individuals.

In this setup we can derive analytic expressions for the probabilities in (P1)-(M) (see Section E in Szydlowski (2019)). To check how the size of the identified set varies with the coarseness of the discretization of X we consider $x = -1, -1 + 2/M, -1 + 4/M, \ldots, 1 - 2/M$ for $M = 2, 10, 60$. Figure 1 portrays the identified set. In order to compare our sets with estimates obtained under assumption of independent censoring, we generate 1000 artificial samples of size 4000 (approx. the sample size in our application) from the model and estimate the Weibull-gamma model under this assumption for each of these samples. We report the median of the estimates together with the median confidence interval, where the latter is constructed using the median of the standard errors across the simulated samples.

In our numerical examples the identified set Θ_0 turns out to be convex despite the fact that it is an intersection of non-convex regions (the same holds in other examples not reported in this article). Figure 1 also shows that the estimates obtained under the erroneous assumption of independent censoring may be far from the true value which is, naturally, always included

12

in our identified set.

The identified set for β shrinks considerably when the number of x values goes from 2 to 10 but we see only minor tightening of the set when we increase M from 10 to 60. These observations suggest that discretization of covariates in our inference procedure should not lead to significant widening of the confidence set as long as the number of points of support is not extremely low. In our application x is multivariate and we discretize x so that $x'\beta$ takes around 30-60 values. The above results suggest that we should not lose a lot of identifying power with this discretization.

3 Inference

This part shows how to obtain confidence regions for the identified set defined by the moment inequalities (P1)-(M). In our model we are faced with infinitely many moment inequalities indexed by two continuous parameters y and x. One way to proceed is to take the supremum over these parameters. In practice this would require the dimension of x to be low and may involve a significant computational burden questioning the applicability of this approach (e.g. when $S(y|x)$ and $S^u(y|x)$ are estimated nonparametrically). Instead, we simplify the problem by assuming that the support of the covariate vector X is finite. Let M denote the cardinality of $\mathcal{X}$.

Assumption 3.1. $\mathcal{X}$ *contains finitely many values* $x_1, x_2, \ldots, x_M$ *and* $P(X = x_m) \geq \delta > 0$ *for all* $m = 1, 2, \ldots, M, 1 < M < \infty$ *and for some* $\delta > 0$.

If some of the covariates are continuously distributed, one can discretize them. The resulting confidence region will most likely differ from the confidence region without imposing discrete support. However, if the discretization is relatively fine, the two regions should be close to each other, as shown in the previous section.

In addition to assuming discrete support for X we will also consider only a finite number of y values: $y_1, y_2, \ldots, y_S$ (we assume $y_1 \leq y_2 \leq \ldots y_S$). In principle discretization of y

may lead to loss of identifying power, however in practice the observations on Y are often recorded on a discrete scale, e.g. unemployment durations are recorded in weeks, which naturally restricts the number of relevant y values. In this case one can check if the moment inequalities are satisfied for the points of support of Y recorded in the data.

3.1 KMS confidence sets

With discrete set of values for x and y our model turns into an unconditional moment inequality model.[9] In practice one is interested only in a component of the parameter vector θ. In our empirical application the object of interest is a single element of the β vector. Thus, for inference we focus on obtaining a marginal confidence set for a subvector θ_1 of θ.[10]

We employ the bootstrap calibrated projection procedure developed recently by Kaido et al. (2019) (KMS henceforth). Index the inequalities by $j = 1, 2, \ldots, J$ where $J = 2MS + M(S-1)$. Note that under Assumption 3.1 and discrete y's the inequalities (P1)-(M) are equivalent to:

$$E[\mu_j(Y_i, X_i, \theta)] \leq 0$$

where:

$$\mu_j(Y_i, X_i, \theta) = \begin{cases} \mathbb{1}\{Y_i > y_s, X_i = x_m\} - \mathcal{L}_v(\Lambda(y_s; \alpha)e^{x'_m\beta}; \gamma)\mathbb{1}\{X_i = x_m\} & \text{for} \quad j = (m-1)S + s \\[2ex] \mathcal{L}_v(\Lambda(y_s; \alpha)e^{x'_m\beta}; \gamma)\mathbb{1}\{X_i = x_m\} - 1 + D_i\mathbb{1}\{Y_i \leq y_s, X_i = x_m\} & \text{for} \quad j = MS + (m-1)S + s \\[2ex] [\mathcal{L}_v(\Lambda(y_{\tilde{s}+1}; \alpha)e^{x'_m\beta}; \gamma) - \mathcal{L}_v(\Lambda(y_{\tilde{s}}; \alpha)e^{x'_m\beta}; \gamma)]\mathbb{1}\{X_i = x_m\} & \\[1ex] \quad + D_i\mathbb{1}\{Y_i \leq y_{\tilde{s}+1}, X_i = x_m\} - D_i\mathbb{1}\{Y_i \leq y_{\tilde{s}}, X_i = x_m\} & \text{for} \quad j = 2MS + (m-1)(S-1) + \tilde{s} \end{cases}$$

with $m = 1, \ldots, M; s = 1, \ldots, S; \tilde{s} = 1, \ldots, S-1$.

---

[9]In principle, one could apply more elaborate method of passing from conditional to unconditional inequalities like the one in Andrews and Shi (2013). However, this would inevitably complicate, already computation-heavy, calculation of our confidence sets.

[10]Alternatively, one could build a confidence set for the whole vector θ and project it on the component of interest, θ_1. However, this approach would produce highly conservative confidence bounds on θ_1, as argued by e.g. Kaido et al. (2019).

Let $\bar{\mu}_j(\theta) = 1/n \sum_{i=1}^n \mu_j(Y_i, X_i, \theta)$ and $\hat{\sigma}_j(\theta)$ denote the estimated standard deviation of $\bar{\mu}_j(\theta)$. The KMS procedure calculates the confidence set for θ_1 as:

$$CS_n \equiv \left[\inf_{\theta \in \Theta} \theta_1 \quad \text{s.t.} \quad \sqrt{n}\frac{\bar{\mu}_j(\theta)}{\hat{\sigma}_j(\theta)} \le \hat{c}(\theta) \quad \forall j, \qquad \sup_{\theta \in \Theta} \theta_1 \quad \text{s.t.} \quad \sqrt{n}\frac{\bar{\mu}_j(\theta)}{\hat{\sigma}_j(\theta)} \le \hat{c}(\theta) \quad \forall j \right] \quad (4)$$

where $\hat{c}(\theta)$ is a critical value calibrated through a bootstrap procedure that iterates over linear programs (see Kaido et al. (2019) for details). The procedure uses GMS of Andrews and Soares (2010) which requires a choice of a tuning sequence κ_n satisfying:

Assumption 3.2. $\kappa_n \to \infty$, $\kappa_n/\sqrt{n} \to 0$.

Denote $\mathcal{L}_v(\Lambda(y_s; \alpha)e^{x_m'\beta}; \gamma)$ by $\mathcal{L}_{m,s}(\theta)$. In order to show validity of the KMS procedure we introduce the following assumptions:

Assumption 3.3. *(a)* Θ *is a compact hyprrectangle with nonempty interior.*

(b) $\{Y_i, D_i, X_i\}_{i=1}^n$ *is an i.i.d. sample.*

(c) For all $y_s \in \{y_1, \ldots, y_S\}$ *and* $x_m \in \mathcal{X}$ *the probabilities* $P(D_i = 1|X_i = x_m), P(Y_i > y_s, X = x_m)$ *and* $E[D_i \mathbb{1}\{y_s \le Y_i \le y_{s+1}, X_i = x_m\}]$ *are bounded away from 0 and 1.*

(d) $\mathcal{L}_{m,s}(\theta)$ *is differentiable with respect to* θ *and the elements of the Hessian matrix* $\nabla_\theta^2 \mathcal{L}_{m,s}(\theta)$ *are bounded above for all* $\theta \in \Theta$, $m = 1, \ldots, M$ *and* $s = 1, \ldots, S$.

Convexity of Θ is needed since KMS procedure uses mean value expansions of the standardized moments $E[\mu_j(Y_i, X_i, \theta)]/\sigma_j(\theta)$. Assumption 3.3(c) is needed to guarantee that the variance of moment conditions, $\sigma_j^2(\theta)$, is bounded away from zero. This condition also implies that the moment conditions cannot be perfectly correlated.

Assumptions 3.3(a) and 3.3(d) are sufficient conditions for the derivative of μ_j to be Lipschitz continuous. Note that for the Weibull baseline hazard $\Lambda(y, \alpha) = y^\alpha, \alpha > 0$ the second derivative $\frac{d^2 \mathcal{L}_m(\theta, y)}{d\alpha^2}$ may not be bounded around $y_1 = 0$. Therefore, in the Weibull model one should choose y_1 to be bounded away from zero. However, this is not worrisome

in practice because we rarely observe durations very close to zero. In our application the unemployment duration is given in weeks and we do not observe unemployment spells shorter than one week.

Partition the parameter space $\Theta = \Theta_1 \times \Theta_{-1}$ and let:

$$\Theta_{0,1} = \{\theta_1 \in \Theta_1 : \quad \exists \theta_{-1} \in \Theta_{-1} \quad \text{such that} \quad (\theta_1, \theta_{-1}) \in \Theta_0\}.$$

The following proposition describes the asymptotic size of the KMS procedure:

Proposition 3. *Let $0 \leq \tau \leq 1/2$. Then under Assumptions 2.1-3.3 we have:*

$$\liminf_{n \to \infty} \inf_{\theta_1 \in \Theta_{0,1}} P(\theta_1 \in CS_n) \geq 1 - \tau.$$

where CS_n is defined in (4).

The proposition shows that the KMS confidence set has correct coverage. This proposition is a consequence of Theorem 4.1 in Kaido et al. (2019). In Section B.3 of Szydlowski (2019) we verify that our assumptions are sufficient to apply their result. Theorem 3 can be strengthened to hold uniformly over all potential null distributions under additional assumptions as in Kaido et al. (2019).

4   Monte Carlo simulations

We investigate the performance of our testing procedure using the following designs. The unemployment duration is generated from the MPH model:

$$\alpha \log \tilde{Y}_i = -X_i \beta - \log V_i + \log U_i$$

where $\alpha = 1.5, \beta = -0.5$, V_i, U_i have the unit exponential distribution (which implies $\gamma = 1$) and are mutually independent as well as independent of X. The censoring process is described

by:

$$\alpha \log C_i = c_1 - X_i \beta \qquad \text{(design 1)}$$

$$\alpha \log C_i = c_2 + \log V_i \qquad \text{(design 2)}$$

$$\alpha \log C_i = c_3 - X_i \beta + \log V_i \qquad \text{(design 3)}$$

where $c_1 = 1.3, c_2 = c_3 = 2.5$. This guarantees a censoring rate around 22%. The covariate X_i is drawn from a discrete uniform distribution on $\mathcal{X} = \{-1, -0.8, -0.6, \ldots, 0.8\}$.

In the first design C_i depends only on X_i, hence the parameter vector (α, β, γ) is point identified. The second design is the one described in Section 2.2. The parameters are partially-identified in this setup, which is also the case in the third design. The slope parameter β is the object of interest. We use a grid $y_s = 0.5, 1, 1.5, \ldots, 5$ for support of Y_i. We set $\kappa_n = \sqrt{n/(\kappa \log(n))}$ and consider $\kappa = 0.5, 1, 1.5$. We consider different values of β to check if our test controls size correctly ($\beta = -0.5$, $\beta = -0.366$ and $\beta = -0.339$) and to examine power properties of our tests ($\beta = 0$). The results are reported in Table 1.

The simulations confirm the result in Proposition 3. Coverage probabilities are above nominal values for β's in the identified set. The conservativeness of the upper approximation to the critical value decreases with sample size towards the nominal values when sample increases from $n = 4000$ to $n = 8000$.

For the upper end of our confidence set ($\beta = -0.366$ or $\beta = -0.339$) we get coverage closest to the nominal values when $\kappa = 0.5$ with overcoverage for higher values of κ. Based on these results we choose $\kappa = 0.5$ for the empirical application. Additionally, with $n = 4000$ the test at the 90% level includes $\beta = 0$, which is outside the identified set, in at most 5% of the cases, which shows that the test has good power.

We note that our test is quite conservative in the point-identified design and design 2. Thus, our approach comes at a cost compared to the standard approach when censoring is in fact independent. However, compared to the cost of misspecifying the censoring mechanism

17

Table 1: Results of Monte Carlo simulations, $M = 10$

	$n = 4000$						$n = 8000$					
	coverage				1 - power		coverage				1 - power	
	interior		boundary				interior		boundary			
design 1 (point-identified)												
	$\beta = -0.5$		$\beta = -0.5$		$\beta = 0$		$\beta = -0.5$		$\beta = -0.5$		$\beta = 0$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	0.95	0.98	0.95	0.98	0.02	0.03	0.94	0.98	0.94	0.98	0.02	0.01
$\kappa = 1$	0.96	0.98	0.96	0.98	0.02	0.04	0.95	0.98	0.95	0.98	0.02	0.00
$\kappa = 1.5$	0.96	0.99	0.96	0.99	0.02	0.04	0.96	0.99	0.96	0.99	0.01	0.00
design 2 (partially-identified)												
	$\beta = -0.5$		$\beta = -0.366$		$\beta = 0$		$\beta = -0.5$		$\beta = -0.366$		$\beta = 0$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	1.00	1.00	0.98	1.00	0.26	0.47	1.00	1.00	0.97	0.99	0.02	0.06
$\kappa = 1$	1.00	1.00	0.99	1.00	0.30	0.52	1.00	1.00	0.98	0.99	0.02	0.08
$\kappa = 1.5$	1.00	1.00	0.99	1.00	0.33	0.54	1.00	1.00	0.98	0.99	0.03	0.10
design 3 (partially-identified)												
	$\beta = -0.5$		$\beta = -0.339$		$\beta = 0$		$\beta = -0.5$		$\beta = -0.339$		$\beta = 0$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	1.00	1.00	0.95	0.98	0.21	0.45	1.00	1.00	0.91	0.96	0.01	0.04
$\kappa = 1$	1.00	1.00	0.96	0.99	0.25	0.41	1.00	1.00	0.93	0.97	0.01	0.05
$\kappa = 1.5$	1.00	1.00	0.96	0.98	0.27	0.48	1.00	1.00	0.93	0.97	0.01	0.05

Note: 2000 Monte Carlo simulations, 500 bootstrap replications. The column "1 - power" gives the probability that the value outside the identified set is included in the confidence set. Boundary value is the value on the boundary of the marginal identified set (calculated numerically).

evidenced in Section 2.2 it seems worth incurring this power cost and using our approach instead of the standard one.

5 Optimal unemployment insurance

We re-investigate the question of optimal unemployment insurance using our novel approach which relaxes the assumption of independent censoring. As shown by Chetty (2008), welfare consequences of a change in unemployment benefits can be derived from a small set of estimated parameters. The crucial parameter in his welfare formula is the elasticity of unemployment exit rate with respect to unemployment benefit. All the previous studies (see e.g. Meyer (1990), Chetty (2008)) estimated this parameter assuming that censoring is independent of unobserved characteristics. Our goal is to find out what can be learned about this elasticity and, as a result, about optimality of unemployment benefits in the US

Table 2: Summary statistics

	mean	std. err.	min	max
unemployment duration in weeks	21.3	21.9	1	171
censored	0.22	.	.	.
benefit level in $	163	27	102	234
pre-unemployment annual wage in $	21 150	13 797	10	169 690
average unemployment rate	5.9	0.9	3.3	9.1
age	36.8	11.1	18.0	64.0
married	0.60	.	.	.

if one disposes of this assumption.

5.1 Data

We use a sample from SIPP 1985-2000 similar to Chetty (2008). The only difference is that we drop second and further unemployment spells for people who entered unemployment multiple times in our sample, which reduced the number of observations from 4529 to 3986.[11] This way we obtain an i.i.d. sample of single spells. The data consist of prime-aged males who receive benefits, search for a job, have at least 3 months of work history and are not on temporary layoff (see Appendix B in Chetty (2008) for a detailed description of the sample).

Our explanatory variables are: logarithm of unemployment benefit level, annual wage before unemployment, average state unemployment rate in years 1985-2000, age and a dummy indicating if the individual is married. Annual wage and average state unemployment rate are meant to control for observed productivity and local labor market conditions' differences, respectively. As in Chetty (2008), we let the unemployment benefit level equal the average state unemployment benefit level in the year of entry into the unemployment pool. Table 2 reports the summary statistics for our sample. An unemployed person spent around 21 weeks in unemployment on average and the longest unemployment spell goes over 2 years. The censoring rate equals 22%.

There are large differences in observed characteristics between censored and uncensored

---

[11]Chetty (2008) estimates a Cox model without unobserved heterogeneity component using all spells. He treats subsequent spells for the same person as separate observations.

observations. Individuals that are subject to censoring had earned $1700 less before they became unemployed, they are older (average age of 38.3) and less likely to be married (55% are married versus 60% in the whole sample). Thus, we should also not expect these two groups to be homogeneous when it comes to unobserved characteristics.

Our set of covariates is highly restricted when compared to Chetty (2008). He includes many other controls, in particular, a full set of year, occupation, industry and state dummies as well as high school completion dummy. However, as shown in Szydlowski (2019), Section C.1, one obtains almost identical estimates of the elasticity of exit rate from unemployment using our restricted set of covariates. This is not surprising given that almost none of the year, industry and occupation dummy variables included in Chetty (2008) are statistically significant. Only state dummies appear to be significant. This is because they control for local labor market conditions, which instead can also be captured by the average state unemployment rate. Adding other variables from his model (education, total household wealth, "seam effect" dummy) also does not significantly change the results (cf. Section C.1 in Szydlowski (2019)), which suggests that our set of covariates provides sufficient control of observed productivity differences.

We include a small set of covariates because our estimation of moment conditions involves estimating the conditional survival function $S(y|x)$. With too many covariates, the number of observations available to estimate $S(y|x)$ for each x will be small and the resulting estimate of poor quality, which would create difficulties for our estimation method. Thus, we do not include these remaining covariates in our model.

5.2 Results

In our first specification we assume that the baseline hazard has the Weibull form and that unobserved heterogeneity is distributed gamma with mean one, which implies:[12]

$$\mathcal{L}_v(\Lambda(y;\alpha)e^{x'\beta};\gamma) = \frac{1}{(1+\gamma y^\alpha e^{x'\beta})^{\frac{1}{\gamma}}}.$$

Next we consider a specification with a piecewise constant hazard with ten steps: $\Lambda(y,\alpha) = \sum_{l=1}^{10} \alpha_l \mathbb{1}\{y \geq c_l\}, \alpha_l > 0$, where $c_l = 1, 6, \ldots, 46.$ [13]

We choose $y_s = 5, 10, \ldots, 50$ (in Section C.5 of Szydlowski (2019) we consider alternative choices of support for y) and, based on MC simulations, pick $\kappa = 0.5$ for the bootstrap statistic. We employ the following discretization procedure. If the desired number of points in the support is greater than two, we divide the support of the covariates according to the quantiles and assign a value equal to the mean within each quantile group[14], e.g. if we want to have 4 points of support for log UI benefit, we divide the support by quartiles and for each quartile calculate the mean UI benefit within the quartile. For binary support, we use dummy variables - below/above median.

---

[12]This specification satisfies Assumption 3.3(d) if $\gamma > 0$ and α, β are bounded, subject to the caveat mentioned in the discussion after the statement of Assumption 3.3.

[13]Our estimates assume that the data generating process is obtained using these specifications. Testing for misspecification of our moment inequality model is beyond the scope of this paper.

[14]We consider alternative choice of values within quantiles in Szydlowski (2019), Section C.4.

Table 3: Confidence sets for the elasticity of unemployment exit rate with respect to unemployment benefit, 90% level

	Weibull hazard		c-depend.		Piecewise constant hazard	c-depend.	0.5 c-depend.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log UI benefit	[-16.27, 3.4]	[-10.03, 6.82]	[-8.06, 1.25]	[-0.97, -0.23]	[-1.64, 4.99]	[-1.15, -0.22]	[-1.15, -0.89]
log annual wage	x	x	x	x	x	x	x
average unemployment rate	x	x	x	x	x	x	x
age	x		x				
married	x		x				
discretization	$4 \times 2 \times 2 \times 2 \times 2$	$6 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$	$6 \times 2 \times 3$	$6 \times 2 \times 3$	$6 \times 2 \times 3$	$6 \times 2 \times 3$
n	3986	3986	3986	3986	3986	3986	3986
CI independent censoring	[-0.99,-0.4]	[-1.06,-0.49]					

Note: The row "discretization" gives the number of discrete values of the variables included in the model in the order they appear in the rows of the table, e.g. $6 \times 2 \times 3$ means 6 values of log UI benefit, 2 values of log annual wage, 3 values of unemployment rate. The number of bootstrap replications is 500 and $\kappa = 0.5$. Columns "c-depend." contain results after imposing the c-dependence condition from Section 5.2.1. Column "0.5 c-depend." uses half of the value of c.

Table 3 gives the results of our empirical study. For comparison we also report interval estimates from the model assuming independent censoring.[15] In the baseline specification (column (2)) we include log UI benefit, log annual wage and average unemployment rate as covariates. As in Chetty (2008) we also censor durations exceeding 50. This increases the censoring rate to 26%. Next we add the demographic controls to the model (column (1)). In Section C.3 of Szydlowski (2019) we compare our results to the previous results available from Meyer (1990) and Chetty (2008).

All our confidence sets with no restrictions on the censoring mechanism (columns (1),(2),(5)) are really wide and contain the interval estimate under independent censoring, thus suggest that we cannot rule out that censoring is independent of unobserved heterogeneity. On the other hand, our results are robust to this assumption. Though we obtain a larger set, we are more confident that the true value lies within this set.

Additionally, the bounds include a range of positive elasticities. Thus, without imposing independent censoring assumption we are not able to say whether the benefits would have positive or negative impact on the exit rate from unemployment. Positive estimates ought not to be viewed as an odd anomaly. For example, when Chetty (2008) estimates the elasticity of unemployment exit rate with respect to severance pay (Table 4, pp. 214 in his paper), as a by-product he also obtains an estimate of elasticity with respect to the UI benefit. The estimated elasticity equals 0.292 and is statistically different from zero (see Section C.2 in Szydlowski (2019)).

Certainly, one can think of an economic mechanism that would lead to positive effect of unemployment benefit on chances of finding a job. If the unemployed are liquidity constrained and job search is costly, then they may pick low (or no) search effort because of insufficient funds. In this case, a higher UI benefit may relax their liquidity constraint and induce more search. If this effect dominates the moral hazard effect (i.e. reduced incentive to search due

---

[15]These estimates use discretized variables, thus the differences between the first and the last row of Table 3 can be associated with different identification strategies and not dicretization. That discretization plays only a minor role for our results is also confirmed for the Cox model under independent censoring (see Szydlowski (2019)).

to the conditional nature of unemployment benefits), the resulting change in the probability of finding a job will be positive. We discuss the theoretical background for positive elasticity in Section D in Szydlowski (2019).

Comparing columns (2) and (5) we notice that the model with step hazard delivers smaller confidence interval than the Weibull model. Still both intervals are very wide. Thus, we conclude that in our application our model with unrestricted censoring is not very informative and one has to impose further restrictions in order to narrow the bounds. We do this by applying c-dependence of Masten and Poirier (2018), which naturally falls into our inference procedure.

5.2.1 Confidence sets under c-dependence

We restrict statistical dependence between censoring and unemployment duration by imposing:

$$\sup_{y_s} |P(D=1|\tilde{Y}=y_s, X=x_m) - P(D=1|X=x_m)| \leq c$$

$$\sup_{y_s} |P(D=1|C=y_s, X=x_m) - P(D=1|X=x_m)| \leq c$$

where c is calibrated as suggested in Masten and Poirier (2018) by taking sample equivalent of:

$$c = \sup_{k=1,2,...,K} \sup_{x_m^{-k}} \sup_{x_m^k} |P(D=1|X=(x_m^k, x_m^{-k})) - P(D=1|X^{-k}=x_m^{-k})|$$

where $x_m = (x_m^k, x_m^{-k})$, i.e. we are looking at maximal change in probability of duration being observed induced by removing the k-th covariate. Loosely speaking we proxy the dependence between censoring and unemployment duration by the dependence between censoring and observed covariates.

By Proposition 2 in Masten and Poirier (2018) such c-dependence restriction implies

(sharp) moment inequalities on the conditional CDF $P(\tilde{Y}_i \leq y_s | X = x_m) = 1 - \mathcal{L}_v(\Lambda(y_s; \alpha)e^{x'_m \beta}; \gamma)$. These inequalities can be naturally incorporated in our inference procedure.

Columns (3), (4) and (6) in Table 3 show that restricting dependence in such a way significantly narrows the estimated bounds. We can rule out positive elasticities in the Weibull model (column (4)) as well as in the step hazard model (column (6)). For example, results from column (6), which we take as our baseline estimates, imply that a 10% increase in the benefit would lead to between 2.2% and 11.5% decrease in the hazard rate.

5.3   Policy implications

In this section we ask what can be learned about optimality of unemployment insurance in the US given our set estimates. First, we endeavor to apply the welfare formula developed by Chetty (2008).

Let W denote social welfare and b the benefit level. Using a general theoretical model Chetty (2008) finds that the money-metric welfare gain at the benefit level equaling to half of the weekly wage can be written as:

$$\frac{dW}{db} = K_1 \left[\frac{f(\varepsilon_1, \varepsilon_2)}{1 - f(\varepsilon_1, \varepsilon_2)} + \frac{\varepsilon_1}{K_2} \right] \tag{5}$$

where:

$$f(\varepsilon_1, \varepsilon_2) = \frac{e^{\varepsilon_2} - 1}{e^{\varepsilon_1} - 1} K_3,$$

ε_1 is the elasticity of the unemployment exit rate with respect to the benefit level and ε_2 is the percentage change in unemployment exit rate associated with receipt of severance pay. K_1, K_2, K_3 are positive constants that are calibrated from the macro data.

The main idea of Chetty's approach is that ε_1 captures two effects. On one hand, additional cash in form of UI benefits creates moral hazard - it reduces the incentive to search for a job. On the other hand, it induces a liquidity effect - if the unemployed are liquidity

constrained, then receipt of UI benefits relaxes their financial constraint and allows them to achieve unconstrained optimum. Chetty's model implies that both the liquidity effect itself as well as the sum of the liquidity effect and the moral hazard effect are negative, i.e. the above welfare formula is valid only when:[16]

$$0 \leq f(\varepsilon_1, \varepsilon_2) \leq 1. \tag{6}$$

With the above welfare formula at hand we can now assess what the policy implications of our set estimate of ε_1 are. The welfare gain gives an average weekly rise in money-metric utility resulting from a 1\$ increase of the benefit at the benefit level equal to 50% of the average wage (the average replacement rate for our sample is close to 50%, see Chetty (2008) Table 1). We aggregate this over a year and over the whole population to obtain a total yearly gain and translate it into the percent of GDP. We fix the value of ε_2 at -0.233, the point estimate obtained by Chetty (2008), and use the same calibration for K_1, K_2 and K_3 as in his paper.

Figure 2 portrays the results. We use our baseline estimates from column (6) of Table 3 since estimates without imposing c-dependence are largely uninformative (note that using bounds from the Weibull model in column (4) would lead to similar conclusions). The black curve plots the welfare change as a function of the unemployment exit rate elasticity. The dashed lines correspond to our confidence sets. Unfortunately, condition (6) is not satisfied for all values of ε_1 in our confidence sets[17], thus in the figure we only include a relevant region where condition (6) holds.

Even in this restricted region we cannot put any upper bound on the welfare effect when we use our baseline estimates (red dashed line). The welfare gain diverges to infinity when ε_1 approaches the point for which $f(\varepsilon_1, \varepsilon_2) = 1$. Thus, we can only conclude that the welfare

---

[16]See formulas (6),(7),(11) in Chetty (2008). Our $\frac{f(\varepsilon_1,\varepsilon_2)}{1-f(\varepsilon_1,\varepsilon_2)}$ corresponds to $\frac{-\partial s_0/\partial A_0}{\partial s_0/\partial w_0}$ in formula (11) in Chetty (2008). Then the restriction on f follows from (6) and (7) in his article.

[17]The welfare formula in (5) is discontinuous at ε_1 for which $f(\varepsilon_1, \varepsilon_2) = 1$ and approaches $-\infty$ to the right of this point.

change from increasing the benefits is in the range $[-0.15, \infty]$% GDP. This only implies that a jump in the level of benefits would not lead to any major drop in output, but we cannot judge if such jump is in fact desirable or if currently benefits are close to the optimal level.

Chetty (2008) uses his point estimates to conclude that the welfare gain from increasing the benefit level would be around 0.04%, which means that in the 1990s the unemployment benefits in the US were set close to the optimal level. We argue that if one is concerned about dependent censoring and wants to stay robust to the assumptions on the censoring distribution, then such sharp conclusions cannot be drawn and a wide range of possible welfare effects are consistent with the model and the data.[18]

Since the baseline estimates do not produce decisive welfare implications, we consider estimates under more restrictive assumptions. Namely, we restrict the c-dependence parameter to half of its original value. Loosely speaking we assume that there is half as much statistical dependence between censoring and unemployment duration as there is between censoring and observed covariates. This produces estimates of the benefit elasticity between -1.15 and -0.89 (column (7) in Table 3) which translates into welfare effect in the range of [-0.15, -0.05]% GDP (dashed blue line in Figure 2). Hence, under half c-dependence a benefit hike would have modest negative effect on the output in the US.

We conclude that given the available data and the existing welfare formula one cannot credibly judge if the unemployment benefits in the US are close to the optimal level unless one is willing to make strong assumptions about the amount of dependence between censoring and unemployment duration. First, the estimates of the elasticity of unemployment exit rate with respect to UI benefit vary in a wide range both with our partial identification approach and with standard independent censoring approach. Secondly, the available welfare formula is not applicable to the whole range of plausible estimates so it is not known what their welfare implications are.

---

[18]We can use delta method to calculate the confidence interval for the welfare gain in Chetty (2008) (taking into account that both ε_1 and ε_2 are estimated from the data). The resulting interval is [-0.57, 1.03]% GDP. Thus, also his point estimate comes with a lot of estimation uncertainty.

6 Conclusion

We argue that the standard assumption in duration modeling that censoring is independent of the unobserved characteristics imposes strong economic assumptions on the underlying behavior that fail in many applications of interest. We show how to proceed with inference without this assumption. Our model does not parametrically restrict the distribution of censoring and partially identifies the parameters of interest.

We applied our model to estimate the elasticity of the unemployment exit rate with respect to the unemployment benefit. The estimates might be used to draw policy conclusions about the optimal level of benefits in the US. However, we found that given the available data the welfare formula obtained in the literature does not allow to draw any substantive policy conclusions unless one is ready to make strong (but still partially identifying) assumptions about dependence between censoring and unemployment duration. This calls for the need to obtain alternative formulas and richer datasets to sharpen the discourse about unemployment insurance policies.

References

Abbring, J. H. and Ridder, G. (2015). Regular variation and the identification of generalized accelerated failure-time models. *Econometric Theory*, 31(6):1229–1248.

Abbring, J. H. and Van den Berg, G. J. (2003). The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(3):pp. 701–710.

Andrews, D. W. K. and Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666.

Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157.

Bailey, L. (2004). Weighting alternatives to compensate for longitudinal nonresponse in the survey of income and program participation. Technical report, Census Bureau Internal Report.

Bugni, F. A., Canay, I. A., and Shi, X. (2017). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*, 8(1):1–38.

Chen, X., Tamer, E., and Torgovitsky, A. (2011). Sensitivity analysis in semiparametric likelihood models. Working paper.

Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy*, 116(2):pp. 173–234.

Doetsch, G. (1974). *The Laplace Transform as an Analytic Function*, pages 26–30. Springer Berlin Heidelberg, Berlin, Heidelberg.

Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies*, 49(3):403–09.

Hausman, J. A. and Woutersen, T. (2014). Estimating a semi-parametric duration model without specifying heterogeneity. *Journal of Econometrics*, 178:114 – 131.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320.

Heckman, J. J. and Honoré, B. E. (1989). The identifiability of the competing risks model. *Biometrika*, 76(2):325–330.

Honoré, B. E. and Lleras-Muney, A. (2006). Bounds in competing risks models and the war on cancer. *Econometrica*, 74(6):1675–1698.

Horowitz, J. L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica*, 67(5):1001–1028.

IDES (2017). Unemployment insurance benefits handbook. Technical report, The Illinois Department of Employment Security.

Kaido, H., Molinari, F., and Stoye, J. (2019). Confidence intervals for projections of partially identified parameters. *Econometrica*. forthcoming.

Khan, S., Ponomareva, M., and Tamer, E. (2016). Identification of panel data models with endogenous censoring. *Journal of Econometrics*, 194(1):57 – 75.

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

Masten, M. A. and Poirier, A. (2018). Identification of treatment effects under conditional partial independence. *Econometrica*, 86(1):317–351.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4):757–782.

Peterson, A. V. (1976). 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.

Ridder, G. and Woutersen, T. M. (2003). The singularity of the information matrix of the mixed proportional hazard model. *Econometrica*, 71(5):1579–1589.

Slud, E. V. and Bailey, L. (2006). Estimation of attrition biases in SIPP. *ASA Section on Survey Research Methods*.

Szydlowski, A. (2019). Online appendix to 'Endogenous Censoring in the Mixed Proportional Hazard Model with an Application to Optimal Unemployment Insurance'.

Van den Berg, G. J., Lindeboom, M., and Ridder, G. (1994). Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behaviour. *Journal of Applied Econometrics*, 9(4):421–435.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics.* Springer-Verlag, New York.

Figure 1: Marginal identified sets

	$M = 2$	$M = 10$	$M = 60$
α	[1.37, 1.62]	[1.37, 1.62]	[1.37, 1.62]
β	[-0.62,-0.32]	[-0.55,-0.37]	[-0.54,-0.38]
γ	[0.36, 1.01]	[0.36, 1.00]	[0.36, 1.00]
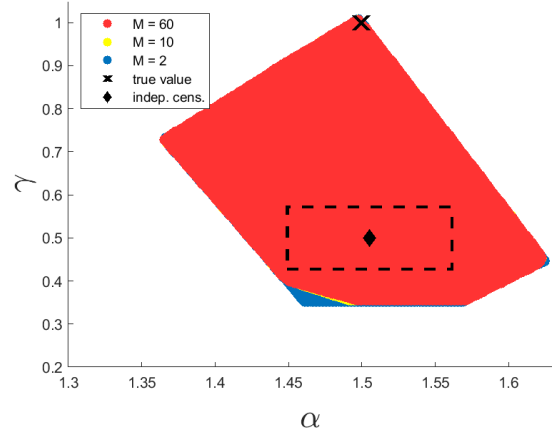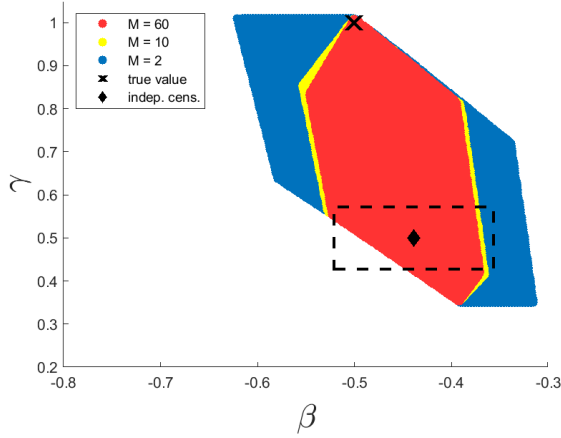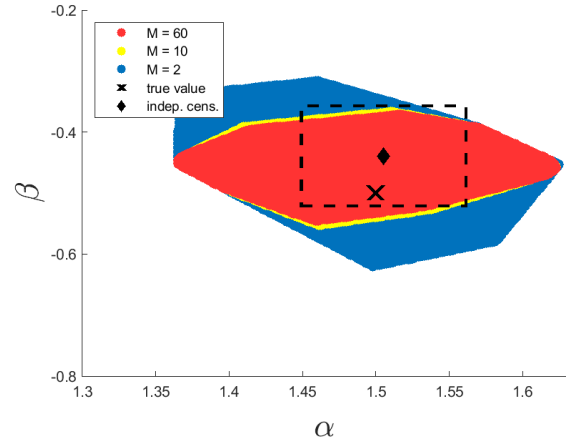
Note: The table gives the marginal identified sets, i.e. the projection of the 3-dimensional identified set on one of the dimensions. In the figures points for higher M are superimposed on points corresponding to lower M. The cross corresponds to the true value in the model. The diamond and the dashed line mark the median point estimate and median confidence interval obtained from estimating the model under the assumption of independent censoring on 1000 simulated samples with $n = 4000$. The median confidence interval is constructed using the median standard error across these simulations.

Figure 2: Yearly welfare change from a 1$ increase in weekly UI benefit as a percent of GDP
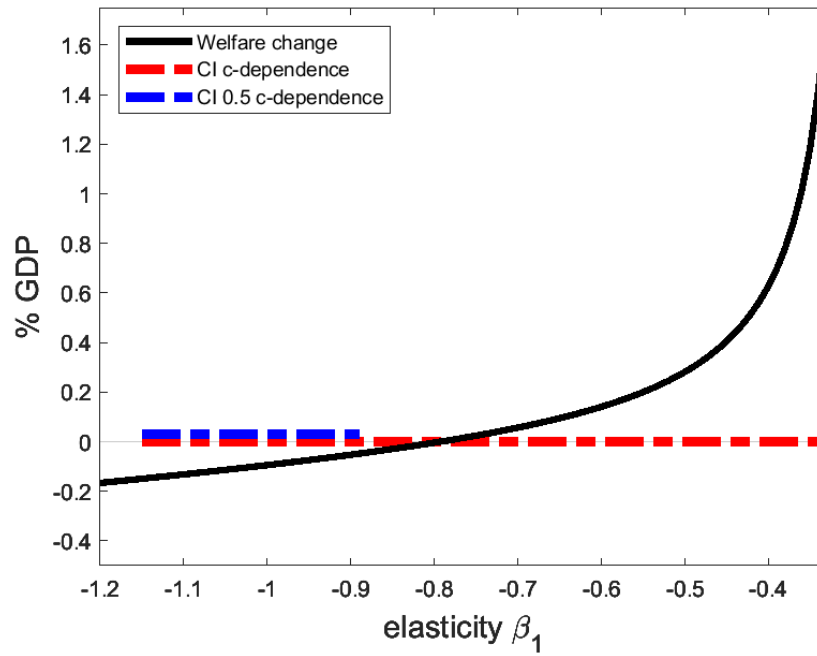
Note: The dashed lines mark the range of values in our confidence set from columns (6) and (7) in Table 3 that are consistent with condition (6).

Online Appendix

A   Further examples of endogenous censoring

Example 3 (censoring through competing risks): Suppose we are studying the improvements in the treatment of cancer over time (e.g. as in Honoré and Lleras-Muney (2006)). We observe the minimum of durations until death from various causes, e.g. if a person with cancer dies because of cardiovascular disease, her duration until death from cancer is censored. Individuals who have cancer may also possess risk factors that make them more prone to die from other causes, e.g. cardiovascular disease. Thus, the underlying observed and unobserved risk factors will be correlated both with the duration until death from cancer as well as with the censoring variable (here, the minimum of durations until death from other causes). Similar concerns arise in economic contexts. For example, suppose we investigate unemployment exit rates among people aged 55-65. People in this group face important health risks so they would often exit to disability or die and their unemployment spells are censored. Individuals with poor health will usually have more trouble finding a job. If health status is not observed perfectly, this would mean that unobserved characteristics are correlated both with employment risk and with competing risks (disability, death).

Example 4 (entry into unemployment during business cycle): Suppose there are two types of people - high (H) and low (L) - and H types leave unemployment faster (due to better motivation, search technology, higher unobserved productivity etc.). We observe a sample of individuals entering unemployment over a period of 10 quarters. Unemployment spells still running at the final date of observation are right-censored. Suppose that the distributions of entries into the unemployment pool differ between two types and are as in Figure 3. Here C_i is the time from entering unemployment until the end of the observation window. Thus, for the person starting her jobless spell at time t we will have $C_i = 10 - t$. Clearly, the distribution of censoring is not independent of the unobserved type. The

Figure 3: Distribution of entries into unemployment
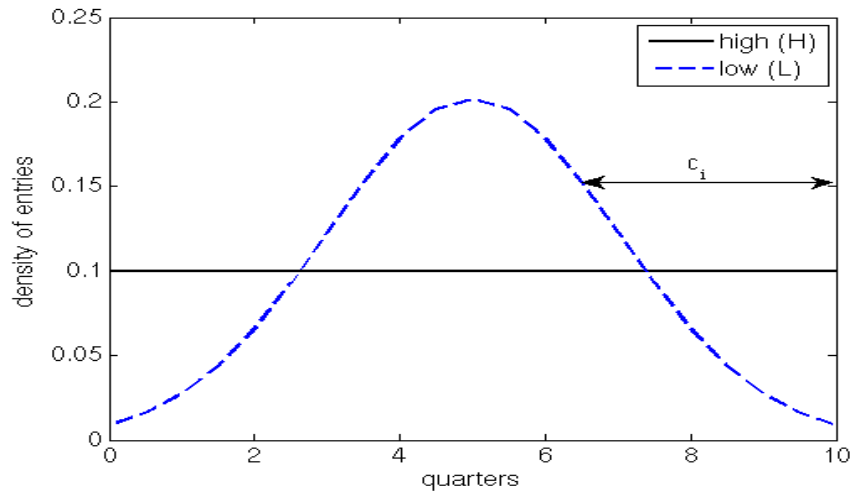
Note: Figure plots densities of entries into unemployment within the 10 quarter window for two groups of unemployed - H and L. C_i denotes a censoring time for an individual who entered the unemployment pool between the sixth and the seventh quarter.

pattern of entries presented in the graph may arise in applications if the employment of low productivity workers is more procyclical than that of high productivity types (the problem arises also if it is more anticyclical, just flip H and L in the figure).

B Proofs

B.1 Proof of Proposition 1

Proof. First, introduce the following notation:

$$p_x \equiv P(D = 1 | X = x)$$

$$\overline{\mathcal{L}}_x(y;\theta) \equiv 1 - \mathcal{L}_v(\Lambda(y;\alpha)e^{x'\beta};\gamma)$$

$$P_x(y;0) \equiv P(Y \leq y | D = 0, X = x)$$

Define:

$$\Psi_x(y;\theta) = \frac{\overline{\mathcal{L}}_x(y;\theta) - P(Y \leq y | D = 1, X = x)p_x}{1 - p_x}$$

In order to prove that inequalities (P1)-(M) provide sharp characterization of the identified set we need to show that for any $\theta = (\alpha, \beta, \gamma)$ satisfying the inequalities we can construct $\tilde{Y}$ such that:

(i) $\tilde{Y} = Y$ if $D = 1$ and $\tilde{Y} \geq Y$ if $D = 0$

(ii) $P(\tilde{Y} \leq y | X = x) = \overline{\mathcal{L}}_x(y;\theta)$ for any $x \in \mathcal{X}$ and any $y \in [0, \infty)$

If there is no censoring, i.e. $p_x = 1$, then construction of $\tilde{Y}$ is simple, just set $\tilde{Y} = Y$ when $X = x$. Condition (i) is trivially satisfied. Now note that $S^u(y|x) - S(y|x) = P(Y \leq y, D = 0 | X = x) = 0$ for such x, thus if θ satisfies inequalities (P1)-(P2), then $S(y|x) = \mathcal{L}_v(\Lambda(y;\alpha)e^{x'\beta};\gamma)$ and condition (ii) is satisfied.

Let's focus on the case $p_x < 1$. If θ satisfies inequality (M), then $\Psi_x(y;\theta)$ is non-decreasing. Further, it is continuous under our MPH model and we have:

$$\lim_{y \to -\infty} \Psi_x(y;\theta) = 0 \quad \lim_{y \to +\infty} \Psi_x(y;\theta) = 1$$

Thus, $\Psi_x(y;\theta)$ is a cumulative distribution function. Let $\Psi_x^{-1}(p;\theta)$ denote the inverse image of p

under $\Psi_x(\cdot; \theta)$. Now construct $\tilde{Y}$ as follows:

$$\tilde{Y} = \begin{cases} Y & \text{if } D = 1 \\ \\ \inf\{S : \Psi_x(S; \theta) \geq P_x(Y; 0)\} & \text{if } D = 0 \end{cases}$$

First we show that this construction satisfies condition (i). Continuity of Ψ_x gives $\Psi_x(\tilde{Y}; \theta) = P_x(Y; 0)$. Note that inequalities (P1)-(P2) imply:

$$0 \leq \Psi_x(y; \theta) \leq P_x(y; 0)$$

which together with invertibility of $P(\cdot; 0)$ gives

$$Y = P^{-1}(\Psi_x(\tilde{Y}; \theta); 0) \leq P^{-1}(P(\tilde{Y}; 0); 0) = \tilde{Y}$$

Second, in order to show (ii) consider the following derivation:

$$P(\tilde{Y} \leq y | X = x) = p_x P(Y \leq y | D = 1, X = x) + (1 - p_x) P(\inf\{S : \Psi_x(S; \theta) \geq P_x(Y; 0)\} \leq y | D = 0, X = x)$$

$$= p_x P(Y \leq y | D = 1, X = x) + (1 - p_x) P(P_x(Y; 0) \leq \Psi_x(y; \theta) | D = 0, X = x)$$

$$= p_x P(Y \leq y | D = 1, X = x) + (1 - p_x) \Psi_x(y; \theta)$$

$$= \overline{\mathcal{L}}_x(y; \theta)$$

where the second equality follows from weak monotonicity of Ψ_x and the final equality follows from definition of Ψ_x. $\qquad\square$

B.2 Proof of Proposition 2

Proof. Note that from proof of Proposition 1: $S^u(y|x) - S(y|x) = P(C \leq y, D = 0 | X = x)$. Thus, under our assumption bounds in (P1) and (P2) collapse and pin down $\tilde{S}(y|x)$ over $y \in [0, \epsilon]$. We assume that $\tilde{S}(y|x) = \mathcal{L}_v(\Lambda(y; \alpha_{true}) e^{x'\beta_{true}}; \gamma_{true})$ but $\mathcal{L}_v(\cdot; \gamma_{true})$ is an analytic function (see e.g. Doetsch (1974)). Together with the assumption that $\Lambda(\cdot; \alpha_{true})$ is analytic, this implies that $\tilde{S}(y|x)$ is analytic as a function of y and by analytic continuation it's uniquely pinned down on $y \in [0, \infty)$

37

for any $x \in \mathcal{X}_{ID}$. Thus, we can apply standard identification arguments for the MPH model (e.g. as in Elbers and Ridder (1982)) to show that $\theta = (\alpha, \beta, \gamma)$ is point-identified.

$\square$

B.3   Proof of Proposition 3

Proof. We will verify that Assumptions 4.1-4.5 needed for Theorem 4.1 in Kaido et al. (2019) to hold. As mentioned above, we abstract from uniformity issues and verify "non-uniform" version of these conditions.

Assumption 4.1(a)-(b)(ii) These are implied by our Assumptions 3.3(a)-(b).

Assumption 4.1(b)(iii) First note that $D_i[\mathbb{1}\{Y_i \leq y_{s+1}\} - \mathbb{1}\{Y_i \leq y_s\}] = D_i \mathbb{1}\{y_s \leq Y_i \leq y_{s+1}\}$ This condition is implied by Assumption 3.3(c) given that for all s and m:

$$E[D_i \mathbb{1}\{y_{s-1} \leq Y_i \leq y_s, X_i = x_m\}] \leq E[D_i \mathbb{1}\{Y_i \leq y_s, X_i = x_m\}] \leq 1 - P(Y_i > y_s, X_i = x_m)$$

so $E[D_i \mathbb{1}\{Y_i \leq y_s, X_i = x_m\}]$ is bounded away from 0 and 1.

Assumption 4.1(b)(iv) This assumption is required for uniformity, thus we skip it.

Assumption 4.2 We use the default GMS function in KMS procedure, which satisfies this assumption. Since we impose Assumption 4.3-(I) in KMS we only require $\kappa_n = o(n^{1/2})$ which follows from our Assumption 3.2.

Assumption 4.3-(I) Note that our Assumption 3.3(c) prevents the moment conditions from being perfectly correlated, in particular, it implies that for every value of x_m some fraction of observations on $\tilde{Y}_i$ will be censored with positive probability.

Assumption 4.4(i) Let ∇_θ denote the gradient with respect to θ. We have:

$$\nabla_\theta \frac{\mu_j(Y_i, X_i, \theta)}{\sigma_j(\theta)} = \frac{\nabla_\theta \mu_j(Y_i, X_i, \theta)\sigma_j(\theta) - \nabla_\theta \sigma_j(\theta)\mu_j(Y_i, X_i, \theta)}{\sigma_j(\theta)^2}$$

Now since $\min_j \inf_{\theta \in \Theta} \sigma_j(\theta)$ is bounded away from zero (note that since we are using Assumption 4.3-(I) in KMS we can replace Θ^ϵ with Θ in their Assumption 4.4), μ_j is bounded for all j and Assumption 3.3(d) holds, we conclude that $\nabla_\theta E[\mu_j(Y_i, X_i, \theta)/\sigma_j(\theta)]$ exists. Moreover, it can be con-

sistently estimated (uniformly over Θ) by replacing $P(X_i = x_m), P(Y_i > y_s, X_i = x_m), E[D_i \mathbb{1}\{Y_i \leq y_s, X_i = x_m\}]$ and $E[D_i \mathbb{1}\{y_s \leq Y_i \leq y_{s+1}, X_i = x_m\}]$ in the expressions for $\nabla_\theta \mu_j$ and $\nabla_\theta \sigma_j$ by their sample analogs.

Assumption 4.4(ii) Assumption 3.3 implies that there exists $\underline{\sigma} > 0$ such that $\min_j \inf_{\theta \in \Theta} \sigma_j(\theta) > \underline{\sigma}$. For $\theta, \tilde{\theta} \in \Theta$ consider the following derivation:

$$
\begin{aligned}
\left\| \nabla_\theta \frac{\mu_j(Y_i, X_i, \theta)}{\sigma_j(\theta)} - \nabla_\theta \frac{\mu_j(Y_i, X_i, \tilde{\theta})}{\sigma_j(\tilde{\theta})}) \right\| &\leq \frac{1}{\underline{\sigma}^2} \left\{ \left\| \nabla_\theta \mu_j(Y_i, X_i, \theta) \sigma_j(\theta) - \nabla_\theta \mu_j(Y_i, X_i, \tilde{\theta}) \sigma_j(\tilde{\theta}) \right\| \right. \\
&\left. + \left\| \nabla_\theta \sigma_j(\theta) \mu_j(Y_i, X_i, \theta) - \nabla_\theta \sigma_j(\tilde{\theta}) \mu_j(Y_i, X_i, \tilde{\theta}) \right\| \right\} \\
&\leq \frac{1}{\underline{\sigma}^2} \left\{ |\sigma_j(\theta)| \|\nabla_\theta \mu_j(Y_i, X_i, \theta) - \nabla_\theta \mu_j(Y_i, X_i, \tilde{\theta})\| \right. \\
&+ \|\nabla_\theta \mu_j(Y_i, X_i, \tilde{\theta})\| |\sigma_j(\theta) - \sigma_j(\tilde{\theta})| \\
&+ |\mu_j(Y_i, X_i, \tilde{\theta})| \|\nabla_\theta \sigma_j(\theta) - \nabla_\theta \sigma_j(\tilde{\theta})\| \\
&\left. + \|\nabla_\theta \sigma_j(\theta)\| \|\mu_j(Y_i, X_i, \theta) - \mu_j(Y_i, X_i, \tilde{\theta})\| \right\}
\end{aligned}
$$

Note that $|\sigma_j(\theta)|$ and $|\mu_j(Y_i, X_i, \tilde{\theta})|$ are bounded. Assumptions 3.3(a) and 3.3(d) imply that $\|\nabla_\theta \mu_j(Y_i, X_i, \tilde{\theta})\|$ and $\|\nabla_\theta \sigma_j(\theta)\|$ are bounded and that $\mu_j(Y_i, X_i, \theta), \nabla_\theta \mu_j(Y_i, X_i, \tilde{\theta}), \sigma_j(\theta), \nabla_\theta \sigma_j(\theta)$ are Lipschitz continuous in θ. Thus, there exists $M < \infty$ such that $\max_j \left\| \nabla_\theta \frac{\mu_j(Y_i, X_i, \theta)}{\sigma_j(\theta)} - \nabla_\theta \frac{\mu_j(Y_i, X_i, \tilde{\theta})}{\sigma_j(\tilde{\theta})}) \right\| \leq M \|\theta - \tilde{\theta}\|$. Moreover, Assumption 3.3(d) implies $\max_j \sup_{\Theta_0} \left\| \nabla_\theta \frac{\mu_j(Y_i, X_i, \theta)}{\sigma_j(\theta)} \right\| \leq \bar{M}$ for $\bar{M} < \infty$.

Assumption 4.5(i) This assumption is satisfied because moments μ_j are built up from indicator functions and indicator functions of measurable sets are measurable.

Assumption 4.5(ii) We will show that under Assumption 3.3 the class of functions:

$$
\mathcal{F}_j = \left\{ f_j : [0, \infty) \times \mathcal{X} \mapsto R : f_j(z_1, z_2) = \frac{\mu_j(z_1, z_2, \theta)}{\sigma_j(\theta)}; \theta \in \Theta \right\}
$$

is Donsker for every $j = 1, \ldots, J$. By Theorem 1.5.7 in Van der Vaart and Wellner (1996) (VW henceforth) this implies that the empirical process $\sqrt{n}(\bar{\mu}_j(\cdot) - E[\mu_j(\cdot)])/\sigma_j(\cdot)$ is asymptotically equicontinuous with respect to the variance semimetric.

Since $\min_j \inf_{\theta \in \Theta} \sigma_j(\theta) > \underline{\sigma}$, by Example 2.10.9 in VW it remains to show that the classes of functions entering the moments μ_j are Donsker. Firstly, the functions $\mathbb{1}\{\cdot \geq y_s, \cdot = x_m\}, \mathbb{1}\{\cdot = x_m\}$

and $D_i \mathbb{1}\{\cdot \leq y_s, \cdot = x_m\}$ are not indexed by θ, thus these are singleton classes and trivially Donsker. The function $\mathcal{L}_{m,s}(\theta)$ is constant in Y and X and, thus, Donsker as well. Since all the above mentioned classes are uniformly bounded for every m and s by Examples 2.10.7 and 2.10.8 in VW we conclude that $\mathcal{F}_j$ is Donsker for every j.

□

C   Application: additional results

C.1   Cox model estimates with a restricted set of covariates

Table 4: Cox model estimates

	(1)	(2)	(3)	(4)	(5)
log UI benefit	-0.504**	-0.520**	-0.455**	-0.499**	-0.410**
	[-0.795,-0.214]	[-0.795,-0.246]	[-0.727,-0.183]	[-0.811,-0.188]	[-0.726,-0.0930]
log annual wage	0.0522**	0.103***	0.0948***	0.0852**	0.113***
	[0.0153,0.0892]	[0.0663,0.139]	[0.0480,0.142]	[0.0285,0.142]	[0.0605,0.166]
av. unempl. rate	-0.110***	-0.105***	-0.107***	-0.0951**	-0.0473
	[-0.161,-0.0596]	[-0.154,-0.0564]	[-0.153,-0.0604]	[-0.160,-0.0297]	[-0.153,0.0585]
age		-0.0161***	-0.0174***		0.261***
		[-0.0194,-0.0127]	[-0.0208,-0.0140]		[0.193,0.328]
married		0.223***	0.243***		0.183***
		[0.160,0.285]	[0.163,0.323]		[0.121,0.245]
onseam			-0.0349		
			[-0.0949,0.0251]		
education			0.000649		
			[-0.0103,0.0116]		
log total HH wealth			0.00644		
			[-0.0160,0.0289]		
discretization	none	none	none	$6 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$
N	4529	4529	4054	4529	4529

90% confidence intervals in brackets
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

We re-estimate the Cox model in Chetty (2008) based on the restricted sets of covariates used in our application. The sample is exactly the same as in his paper (e.g. the durations

over 50 weeks are censored). The variables "education" and "log HH wealth" are the same as the education and log total household wealth variables in Chetty's model (they are denoted "ed" and "l_hh_twlth" there, see Stata codes on his website). The "onseam" variable is a dummy indicating if the person is on the "seam" between the interviews (see Chetty (2008) for details). The results are in Table 4.

The point estimate of the elasticity of exit rate from unemployment w.r.t. UI benefit level in his model is -0.527 when he does not control for log total HH wealth (reported in the publication) and -0.514 when he does (obtained by us using his codes).[19] We see that our restricted set of covariates (columns (1) and (2)) give estimates (-0.504 and -0.52) very close to those reported in Chetty (2008). Moreover, including other variables in the model only slightly affects the estimate of the elasticity and none of these additional variables is statistically significant with low values of t statistics.

Next, in columns (4)-(5) we check how discretization affects our results. Although the grid for the regressors is quite coarse, the resulting point estimates and confidence intervals are very similar to respective estimates without discretization (columns (1) and (2)). In particular, the confidence intervals for the discretized model are not much wider than for the model with continuous covariates. This is reassuring. It confirms that discretization plays a minor role and that the wide bounds in our final result in Section 5 are not driven by the reduced variation in explanatory variables due to discretization but rather by relaxation of the independent censoring assumption.

C.2   Severance pay and unemployment duration: Chetty (2008)

Table 7 shows results from estimating a Cox model as in Chetty (2008). The only difference between his and our table is that we report estimates of additional coefficients, in particular the estimate of elasticity of unemployment exit rate w.r.t. UI benefit.

---

[19]In fact the note under Table 2 in his paper claims that the reported estimates control for total household wealth but they do not. Nevertheless, the estimates from both models are very similar.

Table 5: Replication of Chetty(2008), Table 4, column (2)

severance pay dummy	-0.233***
	(0.071)
age	-0.0191***
	(0.001)
marital status dummy	0.305***
	(0.046)
high school dropout	-0.308***
	(0.063)
college graduate	0.127**
	(0.053)
log UI benefit	**0.292*****
	(0.041)
N	2428

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Note: The model includes additional controls: state, occupation, industry dummies, 10 point log annual wage and tenure splines and time-varying effect of the severance pay.

C.3   Comparison with previous estimates

Table 6: Comparison with previous estimates

	assumptions			data	90% CI
	hazard	V	$V \perp C$		
column (6) of Table 3	Weibull	yes	no	SIPP 1985-2000	[-1.15, -0.22]
Meyer (1990)	nonparam.	yes	yes	UI records 1978-1983	[-1.22, -0.54]
Chetty (2008)	nonparam.	no	.	SIPP 1985-2000	[-0.97, -0.09]

Note: The estimates from Meyer (1990) are based on column 5 in Table V in his paper. The estimates from Chetty (2008) are based on column 1 in Table 2 in his paper.

C.4   Different values within the quantiles

In this section we repeat our main empirical exercise for datasets that use different discretizations than the one in the main text: within each quantile we set x_m to the minimal or

maximal value within the quantile instead of the mean value.[20] We use the Weibull model as this leads to faster computation. Comparing the results, given in Table 7, to those in Table 3 in the main text shows that the choice of the point within the quantile group does not affect the results in any dramatic manner. If anything, it leads to narrower bounds in most of the cases. Thus, our main conclusions would be preserved if we used these alternative discretizations.

Table 7: Confidence sets for the elasticity of unemployment exit rate with respect to unemployment benefit, 90% level

	(1)	(2)	(1)	(2)
	minimum within quantile		maximum within quantile	
log UI benefit	[-9.53, 4.02]	[-16.08, 3.58]	[-8.97, 0.12]	[-15.23, 3.33]
log annual wage	x	x	x	x
average unemployment rate	x	x	x	x
age		x		x
married		x		x
discretization	$6 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$	$6 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$
n	3986	3986	3986	3986
CI under independent censoring	[-0.84,-0.28]	[-1.16,-0.49]	[-0.73,-0.26]	[-0.95,-0.35]

Note: The row "discretization" gives the number of discrete values of the variables included in the model in the order they appear in the rows of the table, e.g. $6 \times 2 \times 3$ means 6 values of log UI benefit, 2 values of log annual wage, 3 values of unemployment rate. The number of bootstrap replications is 500 and $\kappa = 0.5$.

C.5 Different support points

In this section we investigate robustness of our empirical results to the choice of discretization of y. Again, for simplicity we use the Weibull model. Table 8 shows that the resulting bounds are similar when we use different support points for y. Note that more refined discretizations lead to more moment inequalities and longer computation (for example, for model in column (1) it takes 25 hours on 8 Intel Xeon skylake CPUs, running at 2.6GHz and 128GB of RAM, to compute the bounds with $y_s = 5, 10, \ldots, 50$ compared to 43 hours with

---

[20]For model (1) in Table 7 we also considered combining minimum value within quantile for the log benefit with maximum value within quantile for the unemployment rate and obtained similar results.

$y_s = 1, 3, \ldots, 49$). Thus, given our results in Table 8 in our empirical application we opt for using $y_s = 5, 10, \ldots, 50$.

Table 8: Confidence sets for the elasticity of unemployment exit rate with respect to unemployment benefit, 90% level

	(1)	(2)
log UI benefit, y=5:5:50	[-16.27, 3.4]	[-10.03, 6.82]
log UI benefit, y=2:2:48	[-16.35, 3.45]	[-14.47, 8.24]
log UI benefit, y=2:4:50	[-16.62, 3.43]	[-12.21, 7.43]
log UI benefit, y=1:5:46	[-16.12, 3.53]	[-10.96, 7.48]
log UI benefit, y=1:4:49	[-16.37, 3.53]	[-14.11, 7.62]
log UI benefit, y=1:2:49	[-16.15, 3.38]	[-16.83, 8.23]
log annual wage	x	x
average unemployment rate	x	x
age	x	
married	x	
discretization	$4 \times 2 \times 2 \times 2 \times 2$	$6 \times 2 \times 3$
n	3986	3986

Note: The row "discretization" gives the number of discrete values of the variables included in the model in the order they appear in the rows of the table, e.g. $6 \times 2 \times 3$ means 6 values of log UI benefit, 2 values of log annual wage, 3 values of unemployment rate. The number of bootstrap replications is 500 and $\kappa = 0.5$.

D Modified search model with liquidity contraints

In this section we show that, if one allows the cost of job search in Chetty's model to vary with the amount of assets and assume that marginal search cost is decreasing in asset holdings, then both the liquidity effect and the total effect of increasing UI benefits may have a positive sign. The intuition behind this modification is that the unemployed who are liquidity constrained may face higher marginal cost of search than wealthier individuals. For example, it may be more difficult or even impossible for them to search for a job in distant locations if they cannot afford to pay the transportation cost. Thus, the theory need not exclude positive values of ε_1.

Let s_t equal the probability of finding a job in the current period, A_t denote the current holding of assets and $v(\cdot), u(\cdot)$ denote flow consumption utilities if employed and unemployed,

44

respectively. Further, let w_t be the wage, b_t denote the unemployment benefit and τ a lump sum tax. Agents face a lower bound on assets L. Both the agent's discount rate and interest rate are zero. The cost of search effort is denoted by $\psi(s_t, A_t)$ where

$$\psi_s(s, A) > 0, \quad \psi_{ss}(s, A) > 0, \quad \psi_{sA}(s, A) < 0,$$

i.e. the cost function is increasing and convex in s and marginal cost of search effort is decreasing in the amount of asset holdings.

The value function for an individual who finds a job at the beginning of period t and holds assets A_t is:

$$V_t(A_t) = \max_{A_{t+1} \geq L} v(A_t - A_{t+1} + w_t - \tau) + V_{t+1}(A_{t+1}).$$

The value function for an individual who fails to find a job at the beginning of period t and remains unemployed is:

$$U_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + b_t) + J_{t+1}(A_{t+1}) \tag{7}$$

where

$$J_t(A_t) = \max_{0 \leq s_t \leq 1} s_t V_t(A_t) + (1 - s_t) U_t(A_t) - \psi(s_t, A_t).$$

The first order condition for optimal search choice is:

$$\psi_s(s_t, A_t) = V_t(A_t) - U_t(A_t). \tag{8}$$

Differentiating with respect to A_t we obtain a formula for the liquidity effect (the effect of

an unconditional cash grant on search intensity):

$$\frac{ds_t}{dA_t} = \frac{v'(c_t^e) - u'(c_t^u) - \psi_{sA}(s_t, A_t)}{\psi_{ss}(s_t, A_t)}$$

where c_t^e, c_t^u are consumption levels in the employed and the unemployed state, respectively. Therefore, even if $v'(c^e) - u'(c^u) \leq 0$ as in Chetty (2008), the liquidity effect may be positive if $-\psi_{sA}$ is sufficiently large. This shows that the liquidity effect cannot be signed in our extended model.

Furthermore, suppose that the unemployed is liquidity constrained, i.e. the constraint $A_{t+1} \geq L$ in (7) is binding. Now, if:

$$\psi_s(0, L) > V_{t+1}(L) - U_{t+1}(L)$$

the unemployed chooses zero search effort. An increase in the benefit level b_t relaxes the liquidity constraint $A_{t+1} \geq L$ and the unemployed can choose assets $A_{t+1} > L$. With this new level of assets it is possible that:

$$\psi_s(s_{t+1}^*, A_{t+1}) = V_{t+1}(A_{t+1}) - U_{t+1}(A_{t+1}), \quad s_{t+1}^* > 0$$

because the left hand side of (8) is decreasing in A_t (we need the right hand side to decrease slower than ψ_s). This shows that an increase in the UI benefit level may lead to an increase in the unemployment exit rate in our extended model.

E   Closed form expressions for the moment conditions in Section 2.2

The following results will be useful in the derivation below:

$$\int_{-\infty}^{w} e^{-e^{-s}(e^d+1)}e^{-s}ds = \frac{e^{-e^{-w}(e^d+1)}}{e^d + 1} \tag{9}$$

$$\int_{0}^{w} e^{-ps^2-s}ds = e^{\frac{1}{4p}}\sqrt{\frac{\pi}{p}}\left[\Phi\left(\sqrt{2p}\left(w + \frac{1}{2p}\right)\right) - \Phi\left(\frac{1}{\sqrt{2p}}\right)\right] \tag{10}$$

where Φ is the standard normal c.d.f. and the second equality holds for $w \geq 0$. These results follow from integration by parts.

Let's turn to the first moment inequality. We have:

$$P(Y > y|X = x) = P(\tilde{Y} > y, C > y|X = x) = P(\alpha \log \tilde{Y} > \alpha \log y, \alpha \log C > \alpha \log y|X = x) =$$

$$= P(-\log U < -\alpha \log y - \beta x - \log V, -\log V < c - \alpha \log y|X = x) =$$

$$= \int_{-\infty}^{\infty} P(-\log U < -\alpha \log y - \beta x + s, s < c - \alpha \log y|X = x, -\log V = s)e^{-e^{-s}}e^{-s}ds =$$

$$= \int_{-\infty}^{c-\alpha \log y} e^{-e^{-s}(e^{\alpha \log y+\beta x}+1)}e^{-s}ds = \frac{e^{-y^\alpha e^{-c}(y^\alpha e^{\beta x}+1)}}{y^\alpha e^{\beta x} + 1},$$

where the last equality follows from (9).

For the second moment condition we obtain:

$$E[D\mathbb{1}\{Y \le y\}|X = x] = P(\tilde{Y} \le C, \tilde{Y} \le y|X = x) =$$

$$= P(-\log U \ge -\beta x - c - 2\log V, -\log U \ge -\beta x - \log V - \alpha \log y|X = x) =$$

$$= \int_{-\infty}^{\infty} P(-\log U \ge \max\{-\beta x - c + 2s, -\beta x - \alpha \log y + s\}|X = x, -\log V = s)e^{-e^{-s}}e^{-s}ds =$$

$$= \int_{-\infty}^{\infty} 1 - \max\{e^{-e^{\beta x + c - 2s}}, e^{-e^{\beta x + \alpha \log y - s}}\}e^{-e^{-s}}e^{-s}ds =$$

$$= 1 - \int_{c-\alpha \log y}^{\infty} e^{-e^{\beta x + c - 2s}} e^{-e^{-s}} e^{-s}ds - \int_{-\infty}^{c-\alpha \log y} e^{-e^{\beta x + \alpha \log y - s}} e^{-e^{-s}} e^{-s}ds =$$

$$= 1 - e^{\frac{1}{4e^{\beta x + c}}} \sqrt{\frac{\pi}{e^{\beta x + c}}} \left[\Phi \left(\sqrt{2e^{\beta x + c}} \left(y^\alpha e^{-c} + \frac{1}{2e^{\beta x + c}} \right) \right) - \Phi \left(\frac{1}{\sqrt{2e^{\beta x + c}}} \right) \right]$$

$$- \frac{e^{-y^\alpha e^{-c}(y^\alpha e^{\beta x} + 1)}}{y^\alpha e^{\beta x} + 1},$$

where the last equality follows from integration by substitution, (10) and (9).