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

Arkadiusz Szydłowski
University of Leicester

June 22, 2017

Abstract

In economic duration analysis, it is routinely assumed that the process which led to censoring of the observed duration is independent of unobserved 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. We assume a parametric model for the duration of interest and leave the distribution of censoring unrestricted, allowing it to be correlated with both observed and unobserved characteristics. This leads to loss of point-identification. We provide a practical characterization of the identified set with moment inequalities and suggest methods for estimating this set. In particular, we propose a profiled procedure that allows us to build a confidence set for a subvector of the model parameters. We apply this approach to estimate the elasticity of exit rate from unemployment with respect to unemployment benefit and find that both positive and negative values of this elasticity are supported by the data. When combined with the welfare formula in Chetty (2008), these estimates do not permit us to put an upper bound on the size of the welfare

*I would like to thank Elie Tamer, Ivan Canay, Alex Torgovitsky for their guidance and support. I am grateful to Joel Horowitz for useful suggestions and discussions at the early stage of this project. I also benefited from comments by Agnieszka Szydłowska, Ryan Marsh, Shruti Sinha, Florian Hoffman, seminar participants at CERGE-EI and conversations with Tiago Pires. I would like to thank Raj Chetty for sharing his dataset and codes. This research used the ALICE High Performance Computing Facility at the University of Leicester and the Social Sciences Computing Cluster (SSCC) at Northwestern University.

change due to an increase in the unemployment benefit. We conclude that given the available data alone, one cannot credibly judge if the unemployment benefits in the US are close to the optimal level.

1 Introduction

1.1 Motivation and summary of the results

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 is 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 put 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 procedures to estimate confidence sets for the parameters of interest. In practice, as it is in our application, only a few elements of the parameter vector are of interest to the researcher. Thus, following Gandhi, Lu & Shi (2013) we propose a profiled procedure that allows us to build a confidence set for a subvector of the model parameters.

Secondly, we provide new set estimates of the elasticity of the exit rate from unemployment with respect to unemployment benefit from a model that does not restrict the distribution of censoring but uses 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 value of the elasticity of 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 Survey of Income and Program Participation (SIPP) and find that both positive and negative values are supported by the data. We show that positive values of this elasticity are encountered in existing studies and are not excluded on theoretical grounds.

When combined with the welfare formula, our 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. We also show that this formula is not well suited to deal with the wide range of plausible estimates for the elasticity of unemployment exit rate with respect to unemployment benefit. In particular, it does not apply to a significant portion of our confidence set. 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. Though, these results are in line with Chetty's results when interval estimates (incorporating standard errors) are used.

We take as a starting point the mixed proportional hazard model, in which 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.

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.[1]

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 & Bailey (2006) 30% of individuals in the initial sample left the survey by the final wave of the interviews. It is likely that unemployed leaving the survey differ from the remaining unemployed both with respect to observed and 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): This assumption 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. Those who do not exhaust their benefits may

[1]Further examples are given in Appendix A.

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. 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. For example, a person with little motivation to become employed is more likely to lose her benefits and, thus, be censored early. This would violate the assumption that censoring is independent of unobserved characteristics.[2]

These examples show that independent censoring implicitly entails strong economic assumptions that are rarely plausible. Thus, in this article we relax the assumption of independence between censoring and unobserved heterogeneity in the mixed proportional hazard model. We assume proportional hazard only for the exit from unemployment and leave the distribution of censoring unrestricted. We derive moment inequalities that characterize the identified set. In general, our moment inequalities allow for a nonparametric baseline hazard λ_0 and a nonparametric distribution of V. However, treating these components nonparametrically significantly complicates inference and computation. Thus, for practical reasons we maintain a parametric form for the baseline hazard and the distribution of unobserved heterogeneity. These functions are point-identified and can be estimated under independent censoring. Therefore, an applied researcher considering our approach faces a trade-off: he can either assume away dependent censoring and estimate the model semi-/non-parametrically or pose parametric (but possibly flexible) models for $(\lambda_0(\cdot), V)$ and allow for dependence between censoring and unobserved heterogeneity.

We provide a method for obtaining marginal confidence sets for subvectors of the param-

[2]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.

eter vector θ. Our confidence set is constructed by collecting candidate values that are not rejected by a bootstrap test. Our procedure has the benefit that if only a component of θ is of interest, as it is in our application, one only has to search over the space of values of this specific component, not all values of the whole vector θ. This can significantly simplify computation.

1.2 Related literature

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 & Lleras-Muney (2006)), which makes this approach unattractive to applied researchers.

Another possible choice is to assume mixed proportional hazards for both risks (exit from unemployment and censoring), see e.g. Van den Berg, Lindeboom & Ridder (1994). This restores point-identification under support conditions on the explanatory variables (cf. Heckman & Honoré (1989), Abbring & Van den Berg (2003)). However, justifying a parametric model for the censoring variable may be problematic.

Therefore, we take an intermediate approach. We assume proportional hazard only for the exit from unemployment and leave the distribution of the censoring variable unrestricted. The methods developed in this paper allow analysis of a competing risk model with many risks without the need to specify a model for all of the risks. For example, when studying causes of death in some population one may be interested only in the effect of covariates on one particular cause (e.g. cardiovascular disease). Our approach allows the researcher to focus on this specific cause without the need of posing a model for the other possible causes.

This makes the results more robust to misspecification.

Partial identification of a log-linear duration model with dependent censoring under median restriction was analyzed by Khan & Tamer (2009). However, they proceed with inference assuming that their condition for point-identification is satisfied. Khan, Ponomareva & Tamer (2016) investigate consequences of endogenous censoring in a panel data model. They suggest using a stochastic dominance test for inference but they discuss only inference on the whole parameter vector. Honoré & Lleras-Muney (2006) analyze a semiparametric competing risks model (accelerated failure time, AFT) with interdependent risks and a binary covariate. Thus, they relax the assumption on the support of covariates required for point-identification (Heckman & 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.

When it comes to inference, our moment inequalities can be viewed as stochastic dominance relationships. Additionally, they are indexed by a continuous parameter which resembles the situation in conditional moment inequalities models. Therefore, we draw from the literature on both of these topics (see Linton, Song & Whang (2010), Lee, Song & Whang (2013), Lee, Song & Whang (2015) and Andrews & Shi (2013)) and suggest a profiled test for inference on a subvector of parameters (θ_1): for each candidate θ_1 we minimize an appropriate criterion function over the remaining parameters. Gandhi et al. (2013) provide a general theorem for profiled inference in conditional moment inequality models. While their setup is different from ours, their results can still be applied. We translate our setup to their context and verify that their conditions are satisfied under our assumptions.

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 test has the right size 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 appendix.

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. We assume that the spells are observed from the beginning and are possibly right-censored. 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 & Ridder (1982), Heckman & Singer (1984), Horowitz (1999) and Ridder & 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 Abbring & Ridder (2015), Hausman & Woutersen (2014) for a review).

Let $\mathcal{X} \subset R^K$ denote the support of X_i. 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,* $\Lambda_0(0) = 0$, $\lim_{y \to \infty} \Lambda(y) = \infty$,

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

These assumptions correspond to Elbers & Ridder (1982). Under Assumptions 2.1(a)-(d) and no censoring (Λ_0, β, F_V) are identified. The same identification argument holds also for a model with completely random censoring and censoring correlated with observed covariates.

2.1 Identification with arbitrary censoring

Let us turn to the case where the censoring variable may be correlated with unobserved heterogeneity. Let $D_i = \mathbb{1}\{\tilde{Y}_i \leq C_i\}$ indicate observations that 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 $\Lambda_0(\cdot) = \Lambda(\cdot;\alpha), \alpha \in R^{d_\alpha}$. The parametric specifications need to satisfy Assumption 2.1. In the empirical application we will use a gamma distribution for the unobserved heterogeneity.[3] This is a common choice in the duration literature (see e.g. Meyer (1990), Nielsen, Gill, Andersen & Sørensen (1992)). We will also assume either a Weibull form for the hazard, $\Lambda(y,\alpha) = y^\alpha$, or use a step function $\Lambda(y,\alpha) = \sum_{l=1}^{L} \alpha_l \mathbb{1}\{y > c_l\}, \alpha_l \geq 0, l = 1,\ldots,L$, where $c_1,\ldots,c_L$ are known constants.

Denote $\theta = (\alpha,\beta,\gamma)$ and let Θ be a subset of R^{d_θ}, $d_\theta = d_\alpha + d_\gamma + K$. Define $F^x_{vc}(v,c|x)$ to be the joint cumulative distribution function of (V,C) conditional on $X = x$. Let $\mathcal{F} = \{F^x_{vc} : E[V] = 1, F_v(\cdot) = F_v(\cdot;\gamma), \gamma \in R^{d_\gamma}\}$ (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 $p_\theta(y,d,x;F^x_{vc})$ denotes the respective density generated from the censored parametric MPH model with dependence between V, C and X governed by F^x_{cv}. The identified set for the true value θ_{true} is the set of all θ's for which there exists a cumulative distribution function $F^x_{cv} \in \mathcal{F}$ such that the model probabilities are consistent with the true probabilities, i.e.

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

[3] We impose $E(V) = 1$ and parametrize the distribution by γ in the following way: $f_v(v) = \frac{v^{1/\gamma - 1} e^{-v/\gamma}}{\Gamma(1/\gamma)\gamma^{1/\gamma}}$, where Γ is the Gamma function.

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

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

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, Tamer & Torgovitsky (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^x_{vc} 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[4]: $Y_i \le \tilde{Y}_i \le D_i Y_i + (1 - D_i)\infty$, which implies:

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

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

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

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

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.

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

$$\mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta}; \gamma) - S(y|x) \geq 0; \quad \forall y \in [0, \infty), x \in \mathcal{X} \tag{4}$$

$$S^u(y|x) - \mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta}; \gamma) \geq 0; \quad \forall y \in [0, \infty), x \in \mathcal{X}, \tag{5}$$

and suppose Assumption 2.1 holds:

(a) Then, $\Theta_I \subseteq \Theta_0$.

(b) 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 $\Theta_0 = \{\theta_{true}\}$.*

Remark 1. *Part (a) follows from noticing that the moment inequalities 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). Thus, we can interpret our approach as an application of Peterson's bounds in the case when the model for one of the competing risks belongs to a MPH parametric family.*

Remark 2. Θ_0 *may be bigger than the identified set. The situation here is similar to Ciliberto & Tamer (2009) who use moment inequalities to estimate a static, discrete game. Though,* *their approach does not provide a sharp set, it is more practical then the other approaches that indeed give sharpness.*

For example, if censoring is independent of X and V but still random (and condition in part (b) does not hold), we can write:

$$S^u(y|x) = S(y|x) + P(\tilde{Y}_i > C_i, C_i \leq y | X_i = x).$$

and the second term in the above expression will be positive for $y > 0$. Thus, the inequalities will yield a set of parameter values, even though the model is point-identified under such random censoring (see Appendix B.3 for numerical examples). However, in view of the examples discussed in the Introduction, full independence will rarely occur in practice so lack of sharpness in this case is not a big concern.

Remark 3. *Part (b) provides sufficient conditions for point-identification. It 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 requirement 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.*

Remark 4. *Alternatively, the identified set can be characterized by the following moment inequalities:*

$$\begin{cases} \mathcal{L}_v(u;\gamma) - P(\log\Lambda(Y_i;\alpha) + x'\beta > \log u | X_i = x) & \geq 0; \\ 1 - E[D_i \mathbb{1}\{\log\Lambda(Y_i;\alpha) + x'\beta \leq \log u\} | X_i = x] - \mathcal{L}_v(u;\gamma) & \geq 0; \end{cases}$$
$$\forall u \in [0,\infty), x \in \mathcal{X}.$$

It can be shown that both characterizations are equivalent. However, the previous description *is more convenient when it comes to inference and application. In particular, it simplifies the technical arguments involved in developing the confidence set for the unknown parameters and it is computationally more attractive since the unknown parameters enter the moment inequalities through a continuous function.*

Remark 5. *Note that* (3) *provides alternative characterization of the MPH model.* $\mathcal{L}_v$ *is strictly decreasing, therefore the monotonicity of the Laplace transform can be used to obtain bounds on* β *up to scale. Using* (2) *and* (3) *we get that* $X_i'\beta \geq X_j'\beta$ *implies* $S^u(y|X_i) \geq S(y|X_j)$ *for all* y*. Hence, scale-free bounds for* β *can be obtained by minimizing an appropriate rank criterion function as in Khan et al. (2016). However, the scale of* β *is usually of interest in applications (and is identified with independent censoring). Thus, we prefer our approach which (partially) identifies the scale of* β*.*

For given y and x the inequalities in Lemma 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 (from now on, we will refer to Θ_0 as "the identified set") 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.

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. Since the moment inequalities are indexed by values of X, a reduction in the number of points of support of X will lead 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:

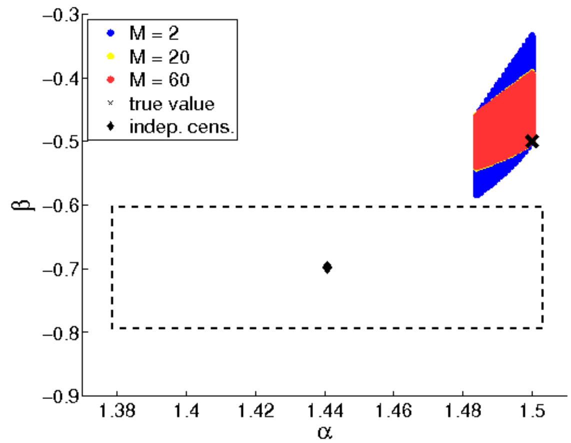
$$\alpha \log \tilde{Y} = -X\beta - \log V + \log U$$
$$\alpha \log C = c + \log V$$

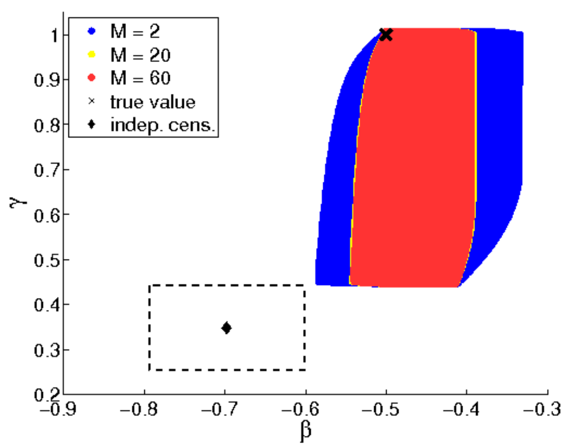
where V, U have the unit exponential distribution (which implies $\gamma = 1$) and are mutually independent. We set $\alpha = 1.5, \beta = -0.5, c = 2.7$ and impose an upper bound on the observed durations equal to 20 ($B_C = 20$ in the notation of Section 3). This guarantees a censoring rate around 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 model $\log C$ and $\log V$ are perfectly correlated (though, $corr(\log \tilde{Y}, \log C) = -0.71$), thus this specification corresponds to a somewhat extreme case. We investigate a model with imperfect correlation between censoring and unobserved heterogeneity in Appendix B.2.

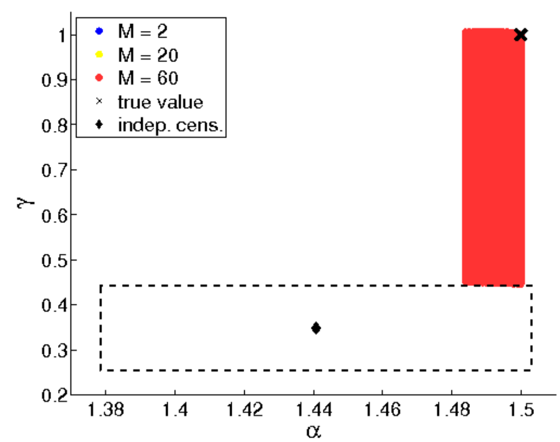
In this setup we can derive analytic expressions for the probabilities in (4)-(5) (see Appendix B.1). To check how the size of the identified set varies with the coarseness of the discretization of X we consider $x = 0, 1/M, 2/M, \ldots, 1 - 1/M$ for $M = 2, 20, 60$. Figure 1 portrays the identified set. To get a sense of how mistaken we can be when we assume 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.

Figure 1: Marginal identified sets

	$M = 2$	$M = 20$	$M = 60$
α	[1.48, 1.50]	[1.48, 1.50]	[1.48, 1.50]
β	[-0.58,-0.33]	[-0.54,-0.39]	[-0.54,-0.39]
γ	[0.45, 1.01]	[0.45, 1.01]	[0.45, 1.01]







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.

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 the example in Appendix B.2). Though, this may be specific to the examples considered.

Most importantly Figure 1 shows that the estimates obtained under the erroneous assumption of independent censoring may be far both from the identified set and the true value. Apart from the Weibull parameter α, the confidence intervals under independent censoring do not cover the true value nor they overlap with our identified set. As mentioned above, our example is somewhat extreme since $corr(\log C, \log V) = 1$. However, the true value is often outside the confidence intervals obtained under independent censoring also in the second model analyzed in Appendix B.2 in which censoring and unobserved heterogeneity are not perfectly correlated. This should serve as a warning sign that mistakenly assuming independent censoring may lead to invalid inference.

The identified set for β shrinks considerably when the number of x values goes from 2 to 20 but we see only minor tightening of the set when we increase M from 20 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 $x'\beta$ so that it takes around 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 (4)-(5). 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 one can also consider only a finite number of y values. In practice, the observations on Y are often recorded on a discrete scale, e.g. unemployment durations are recorded in weeks. In this case one can check if the moment inequalities are satisfied for the points of support of Y recorded in the data which further simplifies computation (if the cardinality of the support is relatively low). However, to keep the discussion general we do not assume that Y has discrete support.

In practice, the observation window is always finite which implies that C has a bounded support $[0, B_C]$ where $B_C > 0$ is known. To simplify notation we redefine y to be equal y/B_C, so that now $y \in [0, 1]$. Accordingly, we modify the integrated baseline hazard Λ such that its domain is $[0, 1]$.

3.1 Profiled ISD test

For a fixed value of x and θ the moment inequalities can be viewed as stochastic dominance relationships. For example, the first inequality (4) implies that the distribution of Y given $X = x$ is stochastically dominated by the distribution $1 - \mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta}; \gamma)$. This suggests that one can verify whether a candidate point belongs to the identified set by testing if the stochastic dominance condition holds for all $x_m \in \mathcal{X}$, thus we coin our procedure the profiled integrated stochastic dominance (ISD) test.

Let

$$\mu_m(\theta, y) = \begin{cases} \mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta_m}; \gamma) - S(y|x_m) & \text{for} \quad m = 1, .., M \\ S^u(y|x_{m-M}) - \mathcal{L}_v(\Lambda(y; \alpha)e^{x'\beta_{m-M}}; \gamma) & \text{for} \quad m = M + 1, .., 2M \end{cases}$$

and $\hat{\mu}_m(\theta, y)$ denote the estimator of $\mu_m(\theta, y)$, e.g.

$$\hat{\mu}_1(\theta, y) = \mathcal{L}_v(\Lambda(y;\alpha)e^{x_1'\beta};\gamma) - \frac{\sum_{i=1}^n \mathbb{1}\{Y_i > y, X_i = x_1\}}{\sum_{i=1}^n \mathbb{1}\{X_i = x_1\}}.$$

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, we provide a procedure that can be used to build a marginal confidence set for a subvector θ_1 of θ.[5] First, partition the parameter space $\Theta = \Theta_1 \times \Theta_{-1}$ and let:

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

Next, define the profiled statistic:

$$T_n(\theta_1) = \min_{\theta_{-1} \in \Theta_{-1}} \sum_{m=1}^{2M} \int_0^1 [\sqrt{n}\hat{\mu}_m(\theta, y)]_-^2 dw(y). \tag{6}$$

where $[\cdot]_- = \min\{\cdot, 0\}$ and $w(\cdot)$ is either the Lebesgue measure or a finite counting measure on [0,1]. In the former case the integral over y can be approximated using quadratures or Monte Carlo integration. In the latter case the integral amounts to a (weighted) sum over the support of Y_i in the dataset. As mentioned above, in practice durations are often recorded on a discrete grid (e.g. in our application $Y_i \in \{1, 2, \ldots, 50\}$) so instead of taking the integral we can check if the moment conditions are satisfied for the points on this grid.[6]

Alternative choice of the test statistic would involve taking the supremum over x_m and y instead of integrating (summing) over them. We prefer the integrated statistic for computational reasons. Note that the estimator of the survival function $\hat{S}(y|x_m)$ is a discontinuous function of x_m and y. Thus, $\hat{\mu}_m(\theta, y)$ may have multiple local maxima with respect to y

[5]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, Molinari & Stoye (2016).

[6]Note that we assume that $w(\cdot)$ does not depend on n. Alternatively, one could use a U-statistic type criterion function: $T_n^U(\theta) = \sum_{m=1}^{2M} \sum_{i=1}^{n} [\sqrt{n}\hat{\mu}_m^{-i}(\theta, Y_i)]_-^2$, where $\hat{\mu}_m^{-i}(\theta, Y_i)$ is an estimator of $\mu_m(\theta, Y_i)$ from a sample excluding the i-th observation. We leave this extension for further work.

since it is equal to the difference between $\hat{S}(y|x_m)$ and the smooth function $\mathcal{L}_v(\Lambda(y;\alpha)e^{x'\beta})$. Therefore, calculation of the supremum statistic would be prohibitively costly. Our form of the statistic is especially attractive when the integral over y turns into a sum as it is in our application.

The asymptotic distribution of this statistic will depend on the set of θ_{-1}'s that contains local deviations from Θ_0 ("limit graph" in the nomenclature of Chernozhukov, Hong & Tamer (2007)). This set cannot be consistently estimated, thus we need to resort to resampling methods to obtain the critical value. We suggest using a bootstrap procedure similar to Gandhi et al. (2013).[7] Although Bugni, Canay & Shi (2016) show that there are power benefits from combining this procedure with another resampling test, these benefits come at a high computational cost. Therefore, we forego this opportunity to improve on power of our test in order to keep the computational burden manageable.

We employ a nonparametric bootstrap, i.e. $\{Y_i^*, D_i^*, X_i^*\}_{i=1}^n$ is obtained by drawing from $\{Y_i, D_i, X_i\}_{i=1}^n$ with replacement. Let $\mu_m^*(\theta, y)$ be a bootstrap counterpart of $\hat{\mu}_m(\theta, y)$ and define $\bar{\mu}_m^*(y,\theta) = \mu_m^*(\theta, y) - \hat{\mu}_m(\theta, y)$. Define the bootstrap statistic:

$$T_{n,r}^*(\theta_1) = \min_{\theta_{-1}\in\Theta_{-1}} \sum_{m=1}^{2M} \int_0^1 [\sqrt{n}\bar{\mu}_m^*(\theta, y) + \sqrt{\kappa_n}\hat{\mu}_m(\theta, y)]_-^2 \, dw(y).$$

where:

Assumption 3.2. *$\kappa_n \to \infty$, $\kappa_n/n \to 0$.*

Denote $\mathcal{L}_v(\Lambda(y;\alpha)e^{x'_m\beta};\gamma)$ by $\mathcal{L}_m(\theta, y)$ and let the distance of a point a from a set A be given by:

$$d(a, A) = \inf_{\tilde{a}\in A} \|a - \tilde{a}\|.$$

[7]One can also use simulations to approximate the asymptotic distribution. We prefer the bootstrap procedure since its computational cost is the same as that of using simulations and Andrews & Soares (2010) show that the bootstrap procedure may have better power properties in finite samples (though bootstrap does not provide asymptotic refinements in this case).

where $\|\cdot\|$ is the Euclidean norm. Next define:

$$Q(\theta) = \sum_{m=1}^{2M} \int_0^1 [\mu_m(\theta, y)]_-^2 \, dw(y).$$

In order to show validity of the bootstrap procedure we introduce the following assumptions:

Assumption 3.3. *(a) Θ is compact, Θ_{-1} is convex for each $\theta_1 \in \Theta_1$ and Θ_0 is a strict subset of Θ.*

(b) $\mathcal{L}_m(\theta, y)$ is differentiable with respect to θ and the elements of the Hessian matrix $\frac{d^2\mathcal{L}_m(\theta,y)}{d\theta^2}$ are bounded above for all $(\theta, y) \in \Theta \times [0, 1]$ and all $m = 1, \ldots, M$.

(c) For any $\theta_1 \in \Theta_{0,1}$ there exist constants $C, c > 0$ such that:

$$Q(\theta) \geq C\{d(\theta, \Theta_0)^2 \wedge c\}$$

for all $\theta \in \Theta$ such that $\theta = (\theta_1, \theta_{-1})$.

Assumption 3.3(a) is needed to guarantee the existence of the minimum in the definition of the profiled statistic and its bootstrapped version. Assumption 3.3(b) allows us to analyze the asymptotic distribution of the statistic on the local parameter space using Taylor expansion arguments.[8] 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 $y = 0$. Therefore, a model with the Weibull hazard fails to satisfy Assumption 3.3(b). 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. In order to satisfy this assumption it is enough to redefine the model with the Weibull hazard such that the hazard takes a finite value near zero and has the Weibull shape further from zero.

[8] In fact Gandhi et al. (2013) require only some form of Hölder continuity of the gradient. However, in our case $\mathcal{L}_m(\theta, y)$ is smooth for all the interesting specifications of Λ and F_v, thus we prefer to state this condition in a stronger form.

Given the sampling scheme the data would not be able to distinguish this model from the standard Weibull hazard model.

Assumption 3.3(c) is a partial identification assumption and is used by Chernozhukov et al. (2007) and Gandhi et al. (2013), among others. It bounds below the rate at which the criterion function Q approaches the identified set - it prevents Q from being very flat in the neighborhood of Θ_0. This condition helps to derive the rate of convergence of the minimizer of Q to the identified set Θ_0 using the distance $d(\cdot,\cdot)$ (see Chernozhukov et al. (2007)). As noted by Kaido et al. (2016) Assumption 3.3(c) is violated when the identified set Θ_0 locally exhibits corners with extremely acute angles. Looking at all of our numerical examples (Section 2.2 and Appendix B.2) this does not seem to be the case in our model.

Define:

$$c_\tau^*(\theta_1) = \inf\left\{ t : \frac{1}{R}\sum_{r=1}^{R} \mathbb{1}\{T_{n,r}^*(\theta_1) \leq t\} \geq 1-\tau \right\},$$

where R is the number of bootstrap replications. The following theorem describes the asymptotic size of our bootstrap test:

Theorem 1. *Let $\theta_1 \in \Theta_{0,1}$ and $\{Y_i, D_i, X_i\}_{i=1}^n$ be an i.i.d. sample. Then under Assumptions 2.1-3.3 we have that as $n \to \infty$:*

$$\liminf_{n\to\infty} P(T_n(\theta_1) \leq c_\tau^*(\theta_1)) \geq 1-\tau.$$

if $c_\tau^*(\theta_1)$ is the continuity point of the asymptotic distribution of $T_n(\theta_1)$.

The theorem shows that the test with bootstrap critical value has a correct null rejection probability. This theorem is a consequence of Theorem 2(b) in Gandhi et al. (2013). In Appendix C.2 we verify that our assumptions are sufficient to apply their result. Theorem 1 can be strengthened to hold uniformly over all potential null distributions under additional assumptions as in Gandhi et al. (2013).

As argued by Chernozhukov et al. (2007), a normalized statistic:

$$\tilde{T}_n(\theta_1) = \min_{\theta_{-1}\in\Theta_{-1}} \sum_{m=1}^{2M} \int_0^1 [\sqrt{n}\hat{\mu}_m(\theta, y)]_-^2 \, dw(y) - \min_{\theta\in\Theta} \sum_{m=1}^{2M} \int_0^1 [\sqrt{n}\hat{\mu}_m(\theta, y)]_-^2 \, dw(y) \qquad (7)$$

may often yield improvement in power over the basic statistic if T_n does not attain the population value zero in finite sample. Similarly we can define the normalized bootstrap statistic:

$$\begin{aligned}\tilde{T}^*_{n,r}(\theta_1) = \min_{\theta_{-1}\in\Theta_{-1}} & \sum_{m=1}^{2M} \int_0^1 [\sqrt{n}\bar{\mu}^*_m(\theta, y) + \sqrt{\kappa_n}\hat{\mu}_m(\theta, y)]_-^2 \, dw(y) \\ & - \min_{\theta\in\Theta} \sum_{m=1}^{2M} \int_0^1 [\sqrt{n}\bar{\mu}^*_m(\theta, y) + \sqrt{\kappa_n}\hat{\mu}_m(\theta, y)]_-^2 \, dw(y).\end{aligned}$$

Arguments mirroring exactly the proof of Theorem 1 imply that such modified bootstrap procedure gives a correct null rejection probability. The computational issues involved in using the latter normalized test statistic are discussed in Appendix E.

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 $c_1 = 1.3, c_2 = c_3 = 2.5$. We impose an upper bound on the observed durations equal to 20. This guarantees a censoring rate around 22%. The covariate X_i is drawn from a discrete uniform distribution on $\mathcal{X} = \{-1, -29/30, -28/30, \ldots, 29/30\}$. The parameter values are chosen such that the designs resemble possibly closely the empirical application discussed in the next section.

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 similar to the one described in Section 2.2. The parameters are partially-identified in this setup, which is also the case in the third design. We report simulations for the profiled statistic, where β is the object of interest. In our SIPP dataset the unemployment durations are recorded in weeks from 1 to 50. Thus, instead of integrating over [0,1] in (6) we evaluate the statistic by taking a sum over this grid, i.e.

$$T_n(\theta) = \min_{\theta_{-1}\in\Theta_{-1}} \sum_{m=1}^{2M} \sum_{s=1}^{S_y} [\sqrt{n}\hat{\mu}_m(\theta, y_s)]_-^2,$$

where $y_s = 0.5, 1, 1.5, \ldots, 20$ are points of support of Y_i. We set $\kappa_n = n/(\kappa \log(n))$ and consider $\kappa = 0.5, 1, 1.5$. We consider different values of β to check if our test controls size correctly ($\beta = -0.5$, $\beta = -0.546$ and $\beta = -0.572$) and to examine power properties of our tests ($\beta = -0.7$). The results are reported in Table 1. We report results only for the normalized statistic given in (7) since simulations for the non-normalized statistic in (6) reveal that the latter leads to highly conservative inference (see Appendix F.1 for details), thus we recommend using the normalized test in applications.

Table 1: Results of Monte Carlo simulations, $M = 60$

	$n = 4000$						$n = 8000$					
	coverage				1 - power		coverage				1 - power	
	interior		boundary				interior		boundary			
	design 1 (point-identified)											
	$\beta = -0.5$		$\beta = -0.5$		$\beta = -0.7$		$\beta = -0.5$		$\beta = -0.5$		$\beta = -0.7$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	0.99	1	0.99	1	0.09	0.26	0.98	1	0.98	1	0	0.02
$\kappa = 1$	0.98	0.99	0.98	0.99	0.05	0.11	0.97	0.99	0.97	0.99	0	0
$\kappa = 1.5$	0.97	0.99	0.97	0.99	0.05	0.10	0.97	0.99	0.97	0.99	0	0
	design 2 (partially-identified)											
	$\beta = -0.5$		$\beta = -0.546$		$\beta = -0.7$		$\beta = -0.5$		$\beta = -0.546$		$\beta = -0.7$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	0.99	1	0.94	0.99	0.27	0.62	0.99	1	0.94	0.99	0.12	0.35
$\kappa = 1$	0.95	0.99	0.79	0.94	0.05	0.17	0.95	0.99	0.76	0.93	0.01	0.04
$\kappa = 1.5$	0.92	0.98	0.72	0.88	0.02	0.09	0.92	0.98	0.68	0.86	0	0.02
	design 3 (partially-identified)											
	$\beta = -0.5$		$\beta = -0.572$		$\beta = -0.7$		$\beta = -0.5$		$\beta = -0.572$		$\beta = -0.7$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	1	1	0.94	0.99	0.40	0.72	1	1	0.92	0.99	0.17	0.47
$\kappa = 1$	0.99	1	0.84	0.95	0.12	0.32	0.99	1	0.77	0.92	0.02	0.10
$\kappa = 1.5$	0.99	1	0.81	0.92	0.08	0.20	0.99	1	0.72	0.88	0.01	0.04

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).

For the lower end of our confidence set ($\beta = -0.546$ or $\beta = -0.572$) we get coverage fairly close to the nominal values only when $\kappa = 0.5$ with undercoverage 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 $\beta = -0.7$, which is outside the identified set, in at most 40% of the cases (for $\kappa = 0.5$).

We note that our test is quite conservative in the point-identified design. Thus, our approach comes at a cost compared to the standard approach when censoring is in fact independent (we compare the power of our approach and the standard approach in more detail in Appendix F.2). However, compared to the cost of misspecifying the censoring mechanism evidenced in Section 2.2 it seems worth incurring this power cost and using our approach instead of the standard one.

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	.	.	.

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 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.[9] This way we obtain an i.i.d. sample of single spells. The data consists 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).

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

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 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 Appendix G.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. Appendix G.1), 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:[10]

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

We also consider a second 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 $c_l = 1, 6, \ldots, 46$.

We pick $\kappa = 0.5$ for the bootstrap statistic. We also estimated the confidence sets with $\kappa = 1$ and obtained thinner sets, which confirms the results of MC simulations that the latter value will yield lower coverage probabilities. 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[11], 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.

Table 3 gives the results of our empirical study. For comparison we also report interval estimates from the model assuming independent censoring.[12] In the baseline specification (column (1)) we include log UI benefit, log annual wage and average unemployment rate as

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

[11] We consider alternative choice of values within quantiles in Appendix G.4.

[12] 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 Appendix G.1).

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

	Weibull hazard		Piecewise constant hazard	
	(1)	(2)	(3)	(4)
log UI benefit	[-1.61,0.6]	[-0.81,0.48]	[-2.49,1.26]	[-1.3,0.6]
log annual wage	x	x	x	x
avg. unemployment rate	x	x	x	x
age		x		x
married		x		x
censoring rate	0.26	0.26	0.26	0.26
discretization	$10 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$	$10 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$
n	3986	3986	3986	3986
CI under independent censoring	[-1.06,-0.49]	[-0.99,-0.4]	[-1.06,-0.49]	[-0.99,-0.4]

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. $10 \times 2 \times 3$ means 10 values of log UI benefit, 2 values of log annual wage, 3 values of unemployment rate. The number of bootstrap replications is 500 and $\kappa = 0.5$.

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 (2)). In Appendix G.3 we compare our results to the previous results available from Meyer (1990) and Chetty (2008).

The confidence set in column (1) contains the interval estimate under independent censoring, thus 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. For example, if the true value happens to be in the corner of the identified set as in Figure 1, then the confidence interval under independent censoring may often fail to cover this true value whereas our set will include it with approximately 90% confidence.

Interestingly, our confidence set in column (2) does not overlap completely with the set obtained under independent censoring, which shows that dependent censoring may be a concern here. Specifically, elasticities between -9.9% and -8.1% are not supported by the data even though they are plausible estimates under the (most likely) misspecified independent censoring model. This shows the advantage of our agnostic approach to modelling the censoring mechanism.

Looking at the results in columns (3) and (4) we find substantially wider bounds than in columns (1) and (2), which is natural given that piecewise constant hazard specification involves more parameters than the Weibull hazard with single parameter. What is re-assuring is that these bounds cover the respective values in columns (1) and (2), suggesting that, though parsimonious, the Weibull model is not misspecified.

All of our confidence sets are wide and include a range of positive elasticities. For example, based on the result in column (2) we cannot rule out that the 10% increase in the benefit would lead to a 4.8% increase in the hazard rate. 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 (this result is not reported in his paper, we obtained it using the Stata codes available on his website, see Appendix G.2).

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 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 Appendix H.

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:

$$\frac{dW}{db} = K_1 \left[\frac{f(\varepsilon_1, \varepsilon_2)}{1 - f(\varepsilon_1, \varepsilon_2)} + \frac{\varepsilon_1}{K_2} \right] \tag{8}$$

where:

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

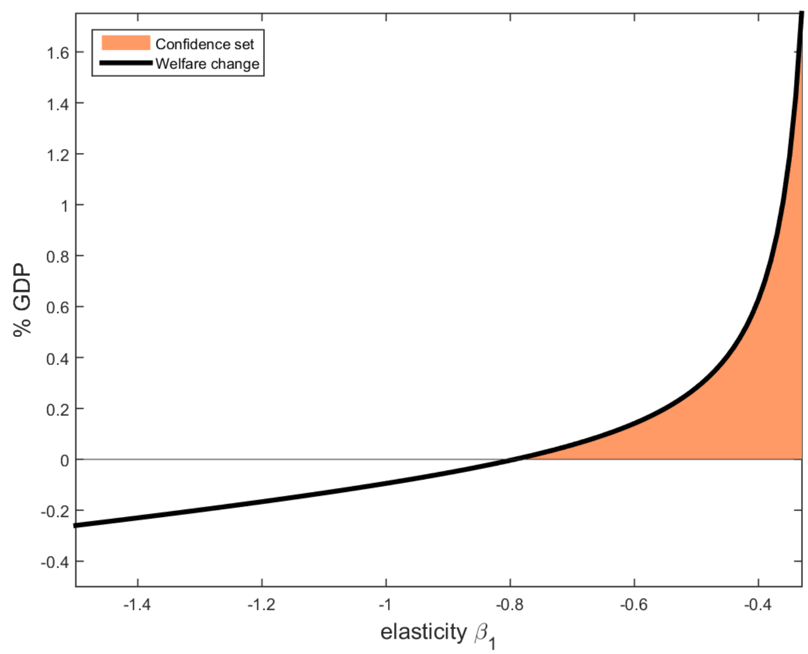
ε_1 is the elasticity of the unemployment exit rate with respect to the benefit level and ε_2 is the percentage change in unemployment exit rate associated with receipt of severance pay. K_1, K_2, K_3 are positive constants that are calibrated from the macro data.

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 the liquidity effect is negative and its value exceeds the total effect, i.e. the above welfare formula is valid only when:

$$0 \leq f(\varepsilon_1, \varepsilon_2) \leq 1. \tag{9}$$

With the above welfare formula at hand we can now assess what are the policy implications of our set estimate of ε_1. For example, a big estimated welfare gain would suggest that a jump in the benefit level would be desirable. 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: Yearly welfare change from a 1$ increase in weekly UI benefit as a percent of GDP



Note: The confidence set contains the range of values in our confidence set from column (2) in Table 3 that are consistent with condition (9).

Figure 2 portrays the results. We use the estimates from column (2) of Table 3 since this specification provides a better control of observed heterogeneity than the model in column (1) (note that using bounds from column (4) would lead to similar conclusions). The black curve plots the welfare change as a function of the unemployment exit rate elasticity. The shaded area corresponds to our confidence set. Unfortunately, condition (9) is not satisfied for all values of ε_1 in our confidence set[13], thus in the figure we only include a relevant region where condition (9) holds.

Even in this restricted region we cannot put any upper bound on the welfare effect. 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 change from increasing the benefits is in the range $[-0.01, \infty]$% GDP. This implies that the current level of benefits may be optimal or a jump in the benefit level would be desirable.

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.[14]

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. 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.

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

[13] The welfare formula in (8) is discontinuous at ε_1 for which $f(\varepsilon_1, \varepsilon_2) = 1$ and approaches $-\infty$ to the right of this point.

[14] 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.

behavior that fail in many applications of interest. We show how to proceed with inference without this assumption. Our model does not restrict the distribution of censoring and partially identifies the parameters of interest.

Our procedure is computationally intensive because it involves optimization and bootstrapping. If one is not willing to estimate the whole confidence set, one can use our test as a check of validity of the assumption of independent censoring. Namely, after estimating the model under the assumption of independent censoring we can run the test to check if our point or interval estimate would belong to the confidence set in a model without imposing this assumption. To the best of our knowledge, this is the first test of this kind in the literature. One can model unemployment duration and censoring using a competing risk MPH model and then test if they are independent using a likelihood ratio test. Van den Berg et al. (1994) implement this method using a competing risks model with Weibull hazards and two-point distributions for unobserved heterogeneity. They test if the unobserved heterogeneity in unemployment duration is correlated with unobserved heterogeneity in censoring. However, our test is more general since it does not impose a specific model on the distribution of censoring and does not restrict unobserved heterogeneity to have a two-point distribution.

We applied our test 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. This calls for the need to obtain alternative formulas and richer datasets to sharpen the discourse about unemployment insurance policies.

References

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

Abbring, J. H. & 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. & Shi, X. (2013), 'Inference based on conditional moment inequalities', *Econometrica* **81**(2), 609–666.

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

Bugni, F. A., Canay, I. A. & Shi, X. (2016), 'Inference for subvectors and other functions of partially identified parameters in moment inequality models', *Quantitative Economics* .

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

Chernozhukov, V., Hong, H. & Tamer, E. (2007), 'Estimation and confidence regions for parameter sets in econometric models', *Econometrica* **75**(5), 1243–1284.

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

Ciliberto, F. & Tamer, E. (2009), 'Market structure and multiple equilibria in airline markets', *Econometrica* **77**(6), 1791–1828.

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

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

Gandhi, A., Lu, Z. & Shi, X. (2013), Estimating demand for differentiated products with error in market shares. Working paper.

Hausman, J. A. & Woutersen, T. (2014), 'Estimating a semi-parametric duration model without specifying heterogeneity', *Journal of Econometrics* **178**, 114 – 131.

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

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

Honoré, B. E. & 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), pp. 1001–1028.

Kaido, H., Molinari, F. & Stoye, J. (2016), Confidence intervals for projections of partially identified parameters. Working paper.

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

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

Lee, S., Song, K. & Whang, Y.-J. (2013), 'Testing functional inequalities', *Journal of Econometrics* **172**(1), 14 – 32.

Lee, S., Song, K. & Whang, Y.-J. (2015), Testing for a General Class of Functional Inequalities. CEMMAP Working Paper.

Linton, O., Song, K. & Whang, Y.-J. (2010), 'An improved bootstrap test of stochastic dominance', *Journal of Econometrics* **154**(2), 186–202.

Meyer, B. D. (1990), 'Unemployment insurance and unemployment spells', *Econometrica* **58**(4), 757–82.

Nielsen, G. G., Gill, R. D., Andersen, P. K. & Sørensen, T. I. A. (1992), 'A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models', *Scandinavian Journal of Statistics* **19**(1).

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 of the United States of America* **73**(1), pp. 11–13.

Ridder, G. & Woutersen, T. M. (2003), 'The singularity of the information matrix of the mixed proportional hazard model', *Econometrica* **71**(5), 1579–1589.

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

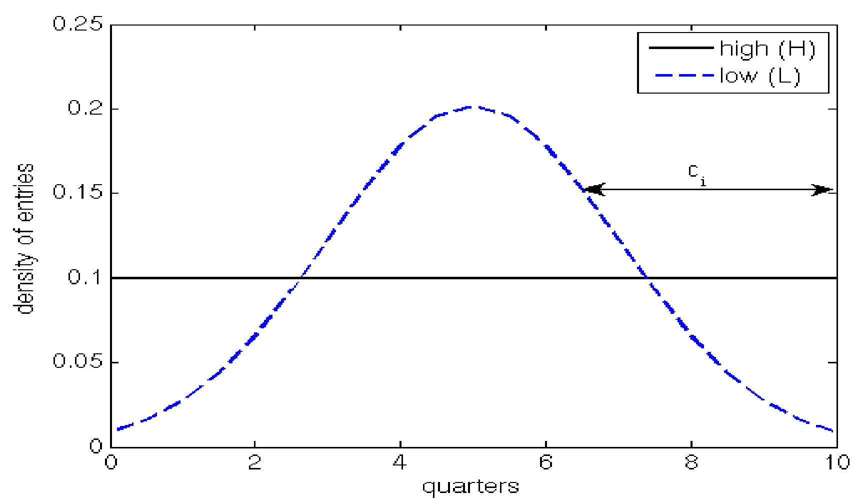
Van den Berg, G. J., Lindeboom, M. & Ridder, G. (1994), 'Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behaviour', *Journal of Applied Econometrics* **9**(4), 421–35.

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

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 & 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).

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.

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 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 Shape of the identified set

B.1 Closed form expressions for the moment conditions in Section 2.2

The following results will be useful in the derivation below:

$$\int_{-\infty}^{w} e^{-e^{-s}(e^d+1)} e^{-s} ds = \frac{e^{-e^{-w}(e^d+1)}}{e^d+1} \tag{10}$$

$$\int_{0}^{w} e^{-ps^2-s} ds = e^{\frac{1}{4p}} \sqrt{\frac{\pi}{p}} \left[\Phi\left(\sqrt{2p} \left(w + \frac{1}{2p} \right) \right) - \Phi\left(\frac{1}{\sqrt{2p}} \right) \right] \tag{11}$$

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:

$$\begin{aligned}
&P(Y > y|X = x) = P(\tilde{Y} > y, C > y|X = x) = P(\alpha \log \tilde{Y} > \alpha \log y, \alpha \log C > \alpha \log y|X = x) = \\
&\quad = P(-\log U < -\alpha \log y - \beta x - \log V, -\log V < c - \alpha \log y|X = x) = \\
&\quad = \int_{-\infty}^{\infty} P(-\log U < -\alpha \log y - \beta x + s, s < c - \alpha \log y|X = x, -\log V = s) e^{-e^{-s}} e^{-s} ds = \\
&\quad = \int_{-\infty}^{c-\alpha \log y} e^{-e^{-s}(e^{\alpha \log y + \beta x} + 1)} e^{-s} ds = \frac{e^{-y^{\alpha} e^{-c}(y^{\alpha} e^{\beta x} + 1)}}{y^{\alpha} e^{\beta x} + 1},
\end{aligned}$$

where the last equality follows from (10).

For the second moment condition we obtain:

$$\begin{aligned}
&E[D\mathbb{1}\{Y \leq y\}|X = x] = P(\tilde{Y} \leq C, \tilde{Y} \leq y|X = x) = \\
&\quad = P(-\log U \geq -\beta x - c - 2\log V, -\log U \geq -\beta x - \log V - \alpha \log y|X = x) = \\
&\quad = \int_{-\infty}^{\infty} P(-\log U \geq \max\{-\beta x - c + 2s, -\beta x - \alpha \log y + s\}|X = x, -\log V = s) e^{-e^{-s}} e^{-s} ds = \\
&\quad = \int_{-\infty}^{\infty} 1 - \max\{e^{-e^{\beta x + c - 2s}}, e^{-e^{\beta x + \alpha \log y - s}}\} e^{-e^{-s}} e^{-s} ds = \\
&\quad = 1 - \int_{c-\alpha \log y}^{\infty} e^{-e^{\beta x + c - 2s}} e^{-e^{-s}} e^{-s} ds - \int_{-\infty}^{c-\alpha \log y} e^{-e^{\beta x + \alpha \log y - s}} e^{-e^{-s}} e^{-s} ds = \\
&\quad = 1 - e^{\frac{1}{4e^{\beta x + c}}} \sqrt{\frac{\pi}{e^{\beta x + c}}} \left[\Phi\left(\sqrt{2e^{\beta x + c}} \left(y^{\alpha} e^{-c} + \frac{1}{2e^{\beta x + c}} \right) \right) - \Phi\left(\frac{1}{\sqrt{2e^{\beta x + c}}} \right) \right] \\
&\qquad - \frac{e^{-y^{\alpha} e^{-c}(y^{\alpha} e^{\beta x} + 1)}}{y^{\alpha} e^{\beta x} + 1},
\end{aligned}$$

where the last equality follows from integration by substitution, (11) and (10).

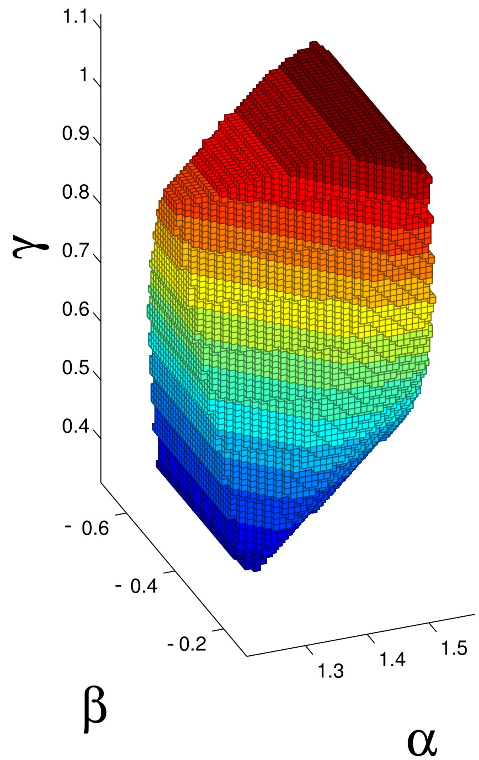
B.2 Shape of the identified set - alternative specification

We analyze a MPH competing risks model of the form:

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

$$\alpha_c \log C = c + \log V + \log U_c$$

Figure 4: Shape of the identified set (M=20)



where V, U, U_c have the unit exponential distribution (which implies $\gamma = 1$) and are mutually independent. We set $\alpha = 1.5, \beta = -0.5, \alpha_c = 1, c = 3.5$ and impose an upper bound $B_C = 20$. This implies a censoring rate around 22%.

In this setup the expressions for the probabilities in (4)-(5) are given by:

$$
\begin{aligned}
P(Y_i \geq y|X_i = x) &= P(\tilde{Y}_i \geq y, \tilde{Y}_i \leq C_i|X_i = x) + P(\tilde{Y}_i > C_i, C_i \geq y|X_i = x) = \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big[\max\Big\{e^{-e^{-s+\alpha\log y+\beta x}} - e^{-e^{-(1+\frac{\alpha}{\alpha_c}\rho)s-\frac{\alpha}{\alpha_c}(s_c-c)+\beta x}}, 0\Big\} \\
&\quad + \max\Big\{e^{-e^{\rho s+\alpha_c\log y-c}} - e^{-e^{-c+(\frac{\alpha_c}{\alpha}+\rho)s-\frac{\alpha_c}{\alpha}(s_c+\beta x)}}, 0\Big\}\Big] e^{-e^{-s}-e^{-s_c}} e^{-s-s_c} ds ds_c
\end{aligned}
$$

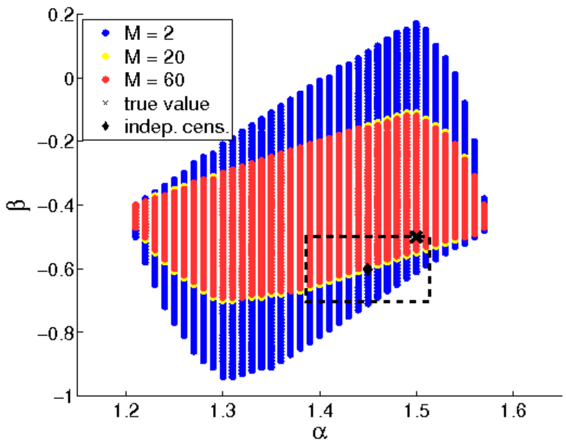
$$
\begin{aligned}
&1 - E[D_i \mathbb{1}\{Y_i < y\}|X_i = x] = \\
&\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1 - e^{-e^{-\max\left\{s-\alpha\log y-\beta x, (1+\frac{\alpha}{\alpha_c}\rho)s+\frac{\alpha}{\alpha_c}(s_c-c)-\beta x\right\}}}\right) e^{-e^{-s}-e^{-s_c}} e^{-s-s_c} ds ds_c.
\end{aligned}
$$

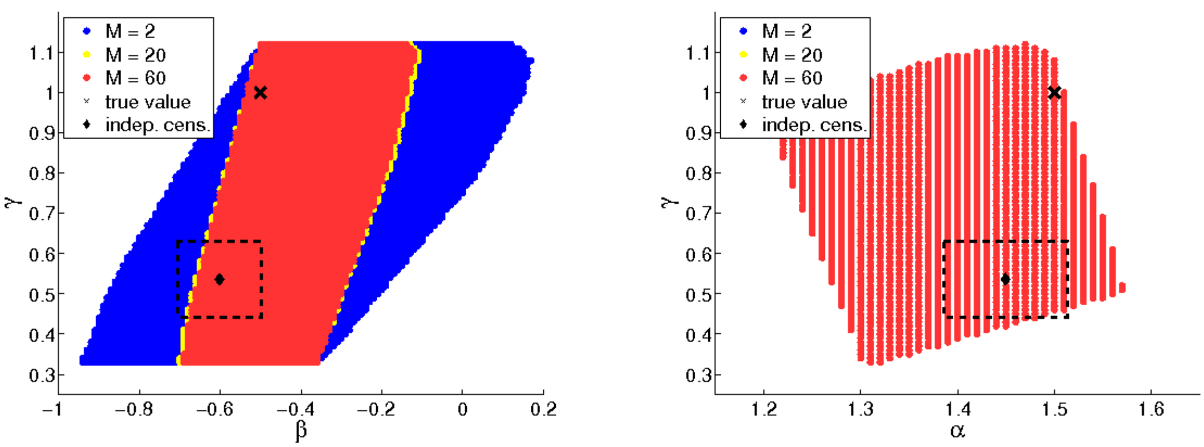
We use numerical integration to evaluate the expressions and employ grid search with step

0.01 to find the identified set. We include the candidate value in the set if it satisfies the inequalities for $y = 0, 0.4, 0.8, 1.2, \ldots, 20$ within integration tolerance 1e-5. As in Section 2.2 we consider $x = 0, 1/M, 2/M, \ldots, 1 - 1/M$ for $M = 2, M = 20$ or $M = 60$. Figures 4-5 portray the identified set.

Figure 5: Marginal identified sets

	$M = 2$	$M = 20$	$M = 60$
α	[1.21, 1.57]	[1.21, 1.57]	[1.21, 1.57]
β	[-0.94,0.17]	[-0.7,-0.11]	[-0.69,-0.12]
γ	[0.33, 1.12]	[0.33, 1.12]	[0.33, 1.12]





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.

Similar to the example in Section 2.2 the identified set turns out to be convex. The identified set and the confidence interval that we would get if we disregard dependent censoring overlap here. The median confidence interval under independent censoring contains the true α but does not cover the true γ. The true value of β lies on the boundary of this interval. This means that if we assume independent censoring, we may often fail to cover the true values of β and γ with the resulting confidence interval. This confirms that imposing an erroneous assumption about dependence between censoring and durations may lead to invalid inference. Moreover, as in Section 2.2 we conclude that discretization of covariates should not lead to significant widening of the confidence set if the number of support points is reasonably large (here $M \geq 20$).

B.3 Identified set under random censoring

As discussed in Remark 2 our moment inequalities do not provide a sharp characterization of the identified set when censoring is fully random, i.e. independent both of observed and unobserved characteristics. In order to visualize how much identifying power is lost in this case, we look at two numerical examples. Unemployment duration and censoring are generated from the model similar to the one in Section 2.2:

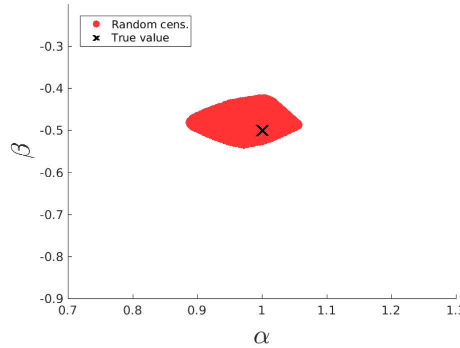
$$\alpha \log \tilde{Y} = -X\beta - \log V + \log U$$
$$\alpha \log C = c + \log \varepsilon$$

where $\beta = -0.5; V, U \sim \mathit{Exponential}(1); X \sim U[-1, 1]$ and in design RC1: $\alpha = 1, c = 0, \varepsilon \sim U[0, 20]$, in design RC2: $\alpha = 1.5, c = 1.9, \log \varepsilon \sim \mathit{Logistic}$.

Figures 6 and 7 show the calculated marginal sets. First note that the MPH model is point-identified under the assumption of random censoring whereas our moment inequalities only partially identify the model parameters and the resulting identification regions may be quite large (especially for design RC2). Thus, our method should not be used when the researcher has all the reasons to believe that data has been censored completely randomly since it would lead to unnecessarily conservative inference. Nonetheless, random censoring happens rarely in practice so lack of sharpness in this case is not of major concern. Recall that

Figure 6: Marginal identified sets, RC1

	$M = 60$
α	[0.89, 1.06]
β	[-0.54, -0.42]
γ	[0.78, 1.1]



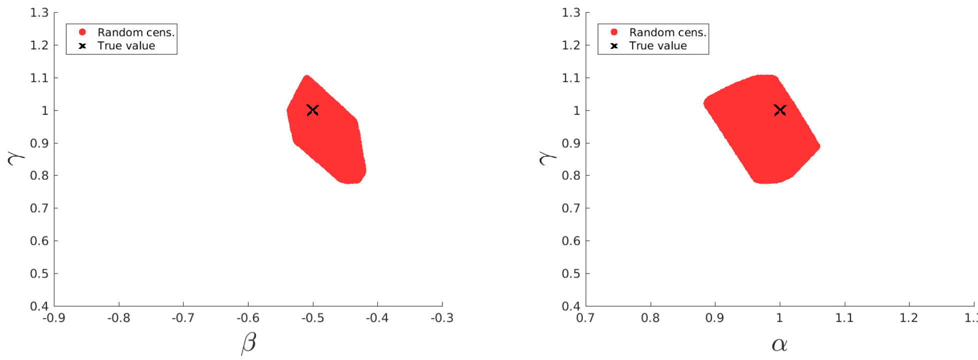
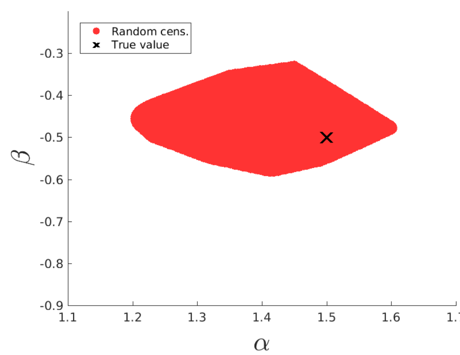
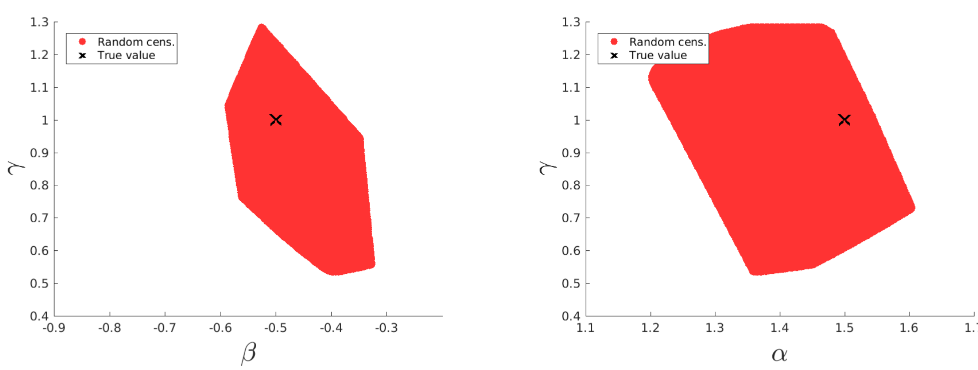


Figure 7: Marginal identified sets, RC2

	$M = 60$
α	[1.2, 1.6]
β	[-0.59, -0.33]
γ	[0.54, 1.29]





if censoring depends only on observed covariates or is random with support bounded from below (and integrated baseline hazard is an analytic function), our moment inequalities point-identify the model parameters, thus they provide sharp characterization of the identified set.

C Proofs

C.1 Proof of Lemma 1

Proof. (a) Note that:

$$S(y|x) = P(\tilde{Y} > y, \tilde{Y} \leq C|X = x) + P(C > y, \tilde{Y} > C|X = x)$$
$$S^u(y|x) = P(\tilde{Y} > y, \tilde{Y} \leq C|X = x) + P(\tilde{Y} > C|X = x)$$

Thus, the bounds $S(y|x) \leq \tilde{S}(y|x) \leq S^u(y|x)$ correspond to the bounds in Peterson (1976), Theorem 1, part 2. He shows that these bounds are sharp. Since his bounds are nonparametric and we impose a parametric model on $\tilde{S}(y|x)$, we have $\Theta_I \subseteq \Theta_0$.

(b) Note that the width of Peterson's bounds is $P(C \leq y, D = 0|X = x)$. Thus, under our assumption bounds 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 $y \in [0, \infty)$ for any $x \in \mathcal{X}_{ID}$. Thus, we can apply standard identification arguments for the MPH model (e.g. as in Elbers & Ridder (1982)) to show that $\theta = (\alpha, \beta, \gamma)$ is point-identified.

□

C.2 Proof of Theorem 1

Let $l_\infty(X)$ be the set of uniformly bounded, real functions on X and $\rightsquigarrow$ denote weak convergence as defined in Van der Vaart & Wellner (1996). We need the following lemma:

Lemma C.1. *Define* $S(y) = [S_1(y) \dots S_{2M}(y)]'$, $\hat{S}(y) = [\hat{S}_1(y) \dots \hat{S}_{2M}(y)]'$ *where:*

$$S_m(y) = \begin{cases} P(Y > y|X = x_m) & m = 1, \dots, M \\ E[D\mathbb{1}\{Y \leq y\}|X = x_{m-M}] & m = M+1, \dots, 2M \end{cases}$$

and:

$$\hat{S}_m(y) = \begin{cases} \frac{\sum_{i=1}^n \mathbb{1}\{Y_i > y X_i = x_m\}}{\sum_{i=1}^n \mathbb{1}\{X_i = x_m\}} & m = 1, \ldots, M \\ \frac{\sum_{i=1}^n D_i \mathbb{1}\{Y_i \leq y, X_i = x_{m-M}\}}{\sum_{i=1}^n \mathbb{1}\{X_i = x_{m-M}\}} & m = M+1, \ldots, 2M. \end{cases}$$

Then:

$$\sqrt{n}(\hat{S}(y) - S(y)) \rightsquigarrow G(y),$$

where $G(y) = [G_1(y) \ldots G_{2M}]'$ is a tight Gaussian process on $l_\infty^{2M}([0,1])$ with covariance kernel:

$$R_{m_1,m_2}(y_1, y_2) =$$

$$= \begin{cases} \frac{cov(\mathbb{1}\{Y_i > y_1, X_i = x_m\} - p_m(y_1)\mathbb{1}\{X_i = x_m\}, \mathbb{1}\{Y_i > y_2, X_i = x_m\} - p_m(y_2)\mathbb{1}\{X_i = x_m\})}{P(X = x_m)^2} & \\ & \text{for} \quad m_1 = m_2 = m \leq M \\ \frac{cov(D_i \mathbb{1}\{Y_i \leq y_1, X_i = x_{m-M}\} - p_m(y_1)\mathbb{1}\{X_i = x_{m-M}\}, D_i \mathbb{1}\{Y_i \leq y_2, X_i = x_{m-M}\} - p_m(y_2)\mathbb{1}\{X_i = x_{m-M}\})}{P(X = x_{m-M})^2} & \\ & \text{for} \quad M < m_1 = m_2 = m \\ \frac{cov(\mathbb{1}\{Y_i > y_1, X_i = x_{m_1}\} - p_{m_1}(y_1)\mathbb{1}\{X_i = x_{m_1}\}, D_i \mathbb{1}\{Y_i \leq y_2, X_i = x_{m_2 - M}\} - p_{m_2}(y_2)\mathbb{1}\{X_i = x_{m_2 - M}\})}{P(X = x_{m_1})^2} & \\ & \text{for} \quad m_2 = M + m_1 \\ 0 & \text{otherwise}, \end{cases}$$

where $p_m(\cdot) = P(Y > \cdot | X = x_m)$ for $m = 1, \ldots, M$ and $p_m(\cdot) = P(D = 1, Y \leq \cdot | X = x_{m-M})$ for $m = M+1, \ldots, 2M$.

Proof. We have:

$$\sqrt{n}(\hat{S}_m(y) - S_m(y)) =$$

$$= \begin{cases} \frac{1}{\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i = x_m\}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}\{Y_i > y, X_i = x_m\} - p_m(y)\mathbb{1}\{X_i = x_m\}\right) & m \leq M \\ \frac{1}{\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i = x_{m-M}\}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n D_i \mathbb{1}\{Y_i \leq y, X_i = x_{m-M}\} - p_m(y)\mathbb{1}\{X_i = x_{m-M}\}\right) & m > M. \end{cases}$$

Note that $D_i \mathbb{1}\{Y_i \leq y, X_i = x_{m-M}\} = \mathbb{1}\{Y_i \leq C_i, Y_i \leq y, X_i = x_{m-M}\}$. Let $f_{1,m}^y : \mathcal{X} \times [0,1] \mapsto$

$[-1,1]$, $f^y_{2,m} : \mathcal{X} \times [0,1]^2 \mapsto [-1,1]$ and:

$$\mathcal{F}_{1,m} = \{f^y_{1,m} : f^y_{1,m}(z_1, z_2) = \mathbb{1}\{z_2 > y\}\mathbb{1}\{z_1 = x_m\} - p_m(z_2)\mathbb{1}\{z_1 = x_m\}; y \in [0,1]\}$$

and

$$\mathcal{F}_{2,m} = \{f^y_{2,m} : f^y_{2,m}(z_1, z_2, z_3) = \mathbb{1}\{z_2 \leq z_3\}\mathbb{1}\{z_2 \leq y\}\mathbb{1}\{z_1 = x_m\} - p_m(z_2)\mathbb{1}\{z_1 = x_m\}; y \in [0,1]\}.$$

The classes $\mathcal{F}_{1,m}$ are Donsker classes of functions for each m. This can be shown as follows. The class $\mathbb{1}\{\cdot \leq y\}$ is Donsker by Example 2.5.4 in Van der Vaart & Wellner (1996). This implies also that $\mathbb{1}\{\cdot > y\}$ is Donsker. The functions $p_m(\cdot), \mathbb{1}\{\cdot = x_m\}$ and $\mathbb{1}\{\cdot \leq \cdot\}$ are not indexed by y and are uniformly bounded. Thus, the claim follows by Example 2.10.10 in Van der Vaart & Wellner (1996).

Therefore, the processes

$$\hat{v}_{1,m}(y) = 1/\sqrt{n} \sum_{i=1}^{n} f^y_{1,m}(X_i, Y_i)$$

and

$$\hat{v}_{2,m}(y) = 1/\sqrt{n} \sum_{i=1}^{n} f^y_{2,m}(X_i, Y_i, D_i)$$

are stochastically equicontinuous, which implies that the vector-valued process $\hat{v}(y) = [\hat{v}_{1,1} \ldots \hat{v}_{2,2M}]$ is stochastically equicontinuous. Since $E[\mathbb{1}\{Y_i > y, X_i = x_m\} - p_m(y)\mathbb{1}\{X_i = x_m\}] = 0$ and $E[D_i\mathbb{1}\{Y_i \leq y, X_i = x_{m-M}\} - p_m(y)\mathbb{1}\{X_i = x_{m-M}\}] = 0$, the central limit theorem and Cramer-Wold device imply fi-di convergence of $\hat{v}(y)$. Moreover, Assumption 3.1 implies:

$$\frac{1}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = x_m\}} \to^p \frac{1}{P(X = x_m)}.$$

Therefore, by Slutsky's lemma (cf. Example 1.4.7 in Van der Vaart & Wellner (1996)) we have that $\hat{G}(y) = \sqrt{n}(\hat{S}(y) - S(y))$ converges weakly to $G(y)$ in $l^{2M}_{\infty}([0,1])$, where G has a covariance kernel given in the statement of the lemma. □

Proof of Theorem 1. The statement of the theorem follows from Theorem D.1 in the next section, proved in Gandhi et al. (2013) (GLS). First, let's translate their setup to our environment. Let $\|\cdot\|$ be the Euclidean norm. Our profiled statistic and bootstrap statistic can be written as follows:

$$T_n(\theta_1) = n \min_{\theta_{-1}\in\Theta_{-1}} \int_0^1 \|[\hat{\mu}(\theta,y)]_-\|^2 \, dw(y),$$

$$T^*_{n,r}(\theta_1) = \min_{\theta_{-1}\in\Theta_{-1}} \int_0^1 \left\|[\sqrt{n}(\mu^*(\theta,y) - \hat{\mu}(\theta,y)) + \sqrt{\kappa_n}\hat{\mu}(\theta,y)]_-\right\|^2 dw(y).$$

Comparing them to (12) and (13) we observe that γ in GLS corresponds to our θ_1, $\Gamma_{-1}(\gamma)$ to Θ_{-1}, g to y, $d\mu(g)$ to $dw(y)$, $\bar{\rho}_n(\theta,g)$ to $\hat{\mu}(\theta,y)$ and $\bar{\rho}^*_n(\theta,g)$ to $\mu^*(\theta,y)$. Our arguments assume that the integral (or the sum) over y is evaluated precisely thus $\mathcal{G}_n$ in GLS corresponds simply to $[0,1]$ in our case and their Assumption D.4 is irrelevant in our setup.

Now we verify that their conditions are satisfied.

Assumption D.1 In our case $\Gamma(\theta) = \theta_1$, thus this assumption is implied by our Assumption 3.3(a).

Assumption D.2(a) It is assumed in the statement of Theorem 1 that the observations are i.i.d.

Assumption D.2(b) Note that in GLS $\bar{\rho}(\theta,g)$ is just a sample mean of $\rho(W_i,\theta,g)$. In this case the Donsker property of the class $\{\rho(\cdot,\theta,g) : (\theta,g) \in \Gamma(\gamma)^{-1} \times \bar{\mathcal{G}}\}$ means that the empirical process $\sqrt{n}(\bar{\rho}_n(\theta,g) - \rho(\theta,g))$ converges weakly to some tight Gaussian process and implies that this empirical process is stochastically equicontinuous with respect to Euclidean metric. It also implies (e.g. by Theorem 3.6.1 in Van der Vaart & Wellner (1996)) that the bootstrap empirical process $\sqrt{n}(\bar{\rho}^*_n(\theta,g) - \bar{\rho}_n(\theta,g))$ converges weakly to the same Gaussian process and is stochastically equicontinuous w.p. 1. This is all that's needed for the proof of Theorem D.1 in GLS. It is not necessary that $\bar{\rho}(\theta,g)$ has the form of the sample mean.

Thus, in order to verify Assumption D.2(b) we need to show that the vector-valued empirical process $\hat{v}(y) = \sqrt{n}(\hat{\mu}(\theta,y) - \mu(\theta,y))$and the bootstrap process $\sqrt{n}(\mu^*(\theta,y) - \hat{\mu}(\theta,y))$ converge weakly to a tight Gaussian process (w.p.1 for the latter). The first part follows from Lemma C.1 since $\hat{v}(y) = \sqrt{n}(\hat{S}(y) - S(y))$. The weak convergence of the bootstrap process follows from the same arguments as in Lemma C.1 and Theorem 3.6.1 in Van der Vaart & Wellner (1996).

Assumption D.2(c) We need to show that the gradient of μ w.r.t. θ exists and is Hölder continuous. Since every differentiable function is Lipschitz, and thus Hölder, if it has a bounded derivative, our Assumption 3.3(b) implies that this assumption is satisfied with $\delta_1 = 1$.

Assumption D.3 This is satisfied with $\delta_2 = 2$ by Assumption 3.3(c).

$\square$

D Notation and results in Gandhi, Lu & Shi (2013)

GLS prove a general theorem for inference on a subvector of parameter vector in conditional moment inequality models. Here we present a simplified version of their results that corresponds closely to our setup. In particular we do not discuss uniformity over the underlying distribution. Secondly, we consider only specific form of the criterion function in the profiled test statistic that does not allow studentization of the estimators of moment conditions. As a result, we can simplify or drop some of the assumptions in their paper.

GLS analyze the following moment inequality model:

$$E[\rho(W_i, \theta_0, g)] \geq 0, \qquad \forall g \in \mathcal{G},$$

where $\rho(\cdot,\cdot,\cdot)$ is a vector-valued function of size k, W_i is a vector of random variables, $\theta_0 \in \Theta_0 \subset \Theta$ is a finite-dimensional vector and $\mathcal{G}$ is a collection of indicator functions over properly defined sets (see GLS for details). The parameter of interest, γ_0, is a function of θ_0. They are related through a mapping Γ:

$$\gamma_0 \in \Gamma(\theta_0).$$

For example, Γ may return the first component of θ.

In principle, the inequalities will only partially identify γ_0 so we need a procedure to estimate confidence sets for this parameter. Let $\Gamma^{-1}(\gamma) = \{\theta \in \Theta : \gamma \in \Gamma(\theta)\}$ and define:

$$\bar{\rho}_n(\theta, g) = \frac{1}{n}\sum_{i=1}^{n} \rho(W_i, \theta, g).$$

Let $\|\cdot\|$ denote the Euclidean norm. GLS propose to build the confidence set by inverting the following test statistic:

$$\hat{T}_n(\gamma) = n \min_{\theta \in \Theta : \theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}_n} \|[\bar{\rho}_n(\theta, g)]_-\|^2 \, d\mu(g), \tag{12}$$

where $\mathcal{G}_n$ is a truncated/simulated version of $\mathcal{G}$, $\mu(\cdot)$ is a probability measure on $\mathcal{G}$.

The critical value for the test can be obtained by bootstrapping the following modified statistic:

$$T_n^*(\gamma) = \min_{\theta \in \Theta : \theta \in \Gamma^{-1}(\gamma)} \int_{\mathcal{G}_n} \left\|[\sqrt{n}(\bar{\rho}_n^*(\theta, g) - \bar{\rho}_n(\theta, g)) + \sqrt{\kappa_n}\bar{\rho}_n(\theta, g)]_-\right\|^2 d\mu(g), \tag{13}$$

where κ_n is a slackness sequence and $\bar{\rho}_n^*(\theta, g)$ is the equivalent of $\bar{\rho}_n(\theta, g)$ calculated on the bootstrap sample. Let $c_{bt}^*(\gamma, p)$ be the p-th quantile of the bootstrap distribution. Then the critical value for the test is given by:

$$c_n^{bt}(\gamma, p) = c_{bt}^*(\gamma, p + \eta^*) + \eta^*,$$

where $\eta^* > 0$ is an infinitesimal number. If one is willing to assume that the asymptotic distribution of $\hat{T}_n(\gamma)$ doesn't have a mass point at zero of size greater or equal than p, then one can set $\eta^* = 0$.

Now we list the assumptions needed for the main result in GLS.

Assumption D.1. *(a) Θ is compact, (b) Γ is upper hemi-continuous, (c) $\Gamma^{-1}(\gamma)$ is either convex or empty for any $\gamma \in R^{d_\gamma}$, and (d) Θ_0 is a strict subset of Θ.*

Let $\bar{\mathcal{G}} = \mathcal{G} \cup \{1\}$, $\rho(\theta, g) = E[\rho(W_i, \theta, g)]$ and $G(\theta, g)$ denote the derivative of $\rho(\theta, g)$ with respect to θ.

Assumption D.2. *For all $\gamma \in R^{d_\gamma}$ such that $\Gamma^{-1}(\gamma) \cap \Theta \neq \emptyset$:*

(a) $\{\rho(W_i, \theta, g)\}_{i=1}^n$ is an i.i.d. sample for any $(\theta, g) \in \Theta \times \bar{\mathcal{G}}$;

(b) the class of functions $\{\rho(\cdot, \theta, g) : (\theta, g) \in \Gamma(\gamma)^{-1} \times \bar{\mathcal{G}}\}$ is Donsker;

(c) $\rho(\theta, g)$ is differentiable w.r.t. $\theta \in \Theta$ and there exist constants C and $\delta_1 > 0$ such that, for any

$(\theta^{(1)}, \theta^{(2)})$:

$$\sup_{g \in \bar{\mathcal{G}}} \left\| G(\theta^{(1)}, g) - G(\theta^{(2)}, g) \right\| \leq C \left\| \theta^{(1)} - \theta^{(2)} \right\|^{\delta_1};$$

Let Γ_0 denote the identified set for γ. Define:

$$Q(\theta) = \int_{\mathcal{G}} \| [\rho(\theta, g)]_- \|^2 \, d\mu(g).$$

and $\Theta_0(\gamma) = \{\theta \in \Theta : \quad Q(\theta) = 0 \quad \& \quad \gamma \in \Gamma(\theta)\}$.

Assumption D.3. *For any $\gamma \in \Gamma_0$ there exist $C, c > 0$ and $2 \leq \delta_2 < 2(\delta_1 + 1)$ such that:*

$$Q(\theta) \geq C(d(\theta, \Theta_0(\gamma))^{\delta_2} \wedge c)$$

for all $\theta \in \Gamma(\gamma)^{-1}$.

Assumption D.4. *(a) $\mathcal{G}_n \uparrow \mathcal{G}$ as $n \to \infty$;*

(b) $\lim_{n\to\infty} \sup_{\theta \in \Gamma(\gamma)^{-1}} \int_{\mathcal{G}/\mathcal{G}_n} \| [\sqrt{n}\rho(\theta, g)]_- \|^2 \, d\mu(g) = 0$ for all $\gamma \in \Gamma_0$.

Note that we do not need Assumption C.6 in GLS since their Lemma C.1(a) verifies that it is satisfied in the above setup. Moreover, we add D.1(d) which is implicitly assumed in GLS (for example, it is required for step 4 in their proof of Theorem E.1).

We are now ready to state the main theorem in GLS:

Theorem D.1. *Suppose that Assumptions D.1-D.4 hold, then:*

$$\liminf_{n \to \infty} P(\hat{T}_n(\gamma) \leq c_n^{bt}(\gamma, p)) \geq p$$

for all $\gamma \in \Gamma_0$.

This theorem corresponds to Theorem 2(b) in GLS.

E Computation

To facilitate discussion define:

$$Q_n(\theta) = \sum_{m=1}^{2M} \sum_{s=1}^{S_y} [\sqrt{n}\hat{\mu}_m(\theta, y_s)]_-^2$$
$$Q_n^*(\theta) = \sum_{m=1}^{2M} \sum_{s=1}^{S_y} [\sqrt{n}\bar{\mu}_m^*(\theta, y_s) + \sqrt{\kappa_n}\hat{\mu}_m(\theta, y_s)]_-^2$$

Since $\hat{\mu}_m(\theta, y_s), m = 1, 2, \ldots, 2M$ are differentiable in θ, $Q_n(\theta)$ and $Q_n^*(\theta)$ are differentiable everywhere besides points at which $\hat{\mu}_m(\theta, y_s) = 0$ for some m. Therefore, we can use gradient based methods to find minimum of $Q_n(\theta)$ and $Q_n^*(\theta)$ over θ_{-1} or θ. We use the MATLAB *fmincon* function with supplied gradient to find the minima involved in calculating our test statistic.

The function $Q_n(\theta)$ may not be globally convex. This problem does not arise in simulations since we can start the optimization procedure at the true parameter value, which in every simulated sample will be close to the identified set, and Q_n is convex in the relevant region. In application we start our optimization procedures at several points to look for the global optimum. In a few bootstrap samples we get a negative value of $\tilde{T}_{n,r}^*(\theta_1)$. In such cases we reset $\min_\theta Q_n^*(\theta)$ to the value we get from $\min_{\theta_{-1}} Q_n^*(\theta)$ and recalculate the critical value (if necessary, we iterate this procedure until the fraction of bootstrap samples with negative values is zero). Since we fix the bootstrap samples for testing different values of θ_1, the tests for different values of θ_1 give us multiple runs of the local optimization $\min_\theta Q_n^*(\theta)$ with different starting values (though, only in the θ_1 dimension).

Moreover, note that since Q_n and Q_n^* are continuous in θ_1 and the minimum with respect to θ_{-1} is taken over a compact set (Assumption 3.3(a)), the test statistic $\tilde{T}_n(\theta_1)$ as well as the critical value obtained from bootstrapping $\tilde{T}_{n,r}^*(\theta_1)$ should be continuous in θ_1 (given that we use the same bootstrap samples for each θ_1). Thus, if any discontinuities appear, this would suggest that the optimization procedures find a local optimum not a global one so the number of the starting points should be increased. In our application starting the minimization at only two points worked well in practice.

The average time to compute the test for a single candidate value was around 2-3 hours. We used cluster computing so the total time needed to obtain a confidence set was around 4-6 hours (first we used a coarse grid and then refined the grid around the edges of the set found in the first step).

F Monte Carlo simulations: additional results

F.1 Non-normalized statistic

We compare the performance of non-normalized and normalized statistics given in (6) and (7). Coverage probabilities in Table 4 show that the non-normalized test is highly conservative. All the values considered, including the ones outside the identified set, are contained in the confidence set with probability one. In fact, simulations for $\beta = -1$ (Table 5) show that for $n = 4000$ this value is included in the confidence set almost with certainty. One requires a sample as large as $n = 12000$ for this test to have good power.

We would expect the test to be conservative for the first design since it poses a worst-case scenario. In this setup all the inequalities are binding or very close to binding for all y values. As shown by Linton et al. (2010), the stochastic dominance test will not have power against local alternatives where the inequalities are close to binding uniformly over y. Moreover, we have a large number of inequalities in our design (120 in total), which further exacerbates power properties of our test.

On the other hand, design 2 does not suffer from these difficulties since only a small subset of inequalities is binding at the boundary of the identified set. The overall poor performance of the non-normalized test may stem from the moment inequalities being violated around $y = 0$ due to the sampling error in estimating the survival functions around this point. This may lead to non-zero value of the test statistic in finite sample even for β's in the identified

Table 4: Results of Monte Carlo simulations, $n = 4000, M = 60$

	non-normalized						normalized					
	coverage				1 - power		coverage				1 - power	
	interior		boundary				interior		boundary			
	design 1 (point-identified)											
	$\beta = -0.5$		$\beta = -0.5$		$\beta = -0.7$		$\beta = -0.5$		$\beta = -0.5$		$\beta = -0.7$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	1	1	1	1	1	1	0.99	1	0.99	1	0.09	0.26
$\kappa = 1$	1	1	1	1	1	1	0.98	0.99	0.98	0.99	0.05	0.11
$\kappa = 1.5$	1	1	1	1	1	1	0.97	0.99	0.97	0.99	0.05	0.10
	design 2 (partially-identified)											
	$\beta = -0.5$		$\beta = -0.546$		$\beta = -0.7$		$\beta = -0.5$		$\beta = -0.546$		$\beta = -0.7$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	1	1	1	1	1	1	0.99	1	0.94	0.99	0.27	0.62
$\kappa = 1$	1	1	1	1	1	1	0.95	0.99	0.79	0.94	0.05	0.17
$\kappa = 1.5$	1	1	1	1	1	1	0.92	0.98	0.72	0.88	0.02	0.09
	design 3 (partially-identified)											
	$\beta = -0.5$		$\beta = -0.572$		$\beta = -0.7$		$\beta = -0.5$		$\beta = -0.572$		$\beta = -0.7$	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
$\kappa = 0.5$	1	1	1	1	1	1	1	1	0.94	0.99	0.40	0.72
$\kappa = 1$	1	1	1	1	1	1	0.99	1	0.84	0.95	0.12	0.32
$\kappa = 1.5$	1	1	1	1	1	1	0.99	1	0.81	0.92	0.08	0.20

Table 5: False coverage probabilities ($\beta = -1$) for various n, $M = 60$

	design 2 (partially-identified)					
	90% level			95% level		
	$n = 4000$	$n = 8000$	$n = 12000$	$n = 4000$	$n = 8000$	$n = 12000$
$\kappa = 0.5$	0.99	0.06	0	1	0.17	0
$\kappa = 1$	1	0.05	0	1	0.15	0
$\kappa = 1.5$	1	0.20	0	1	0.45	0

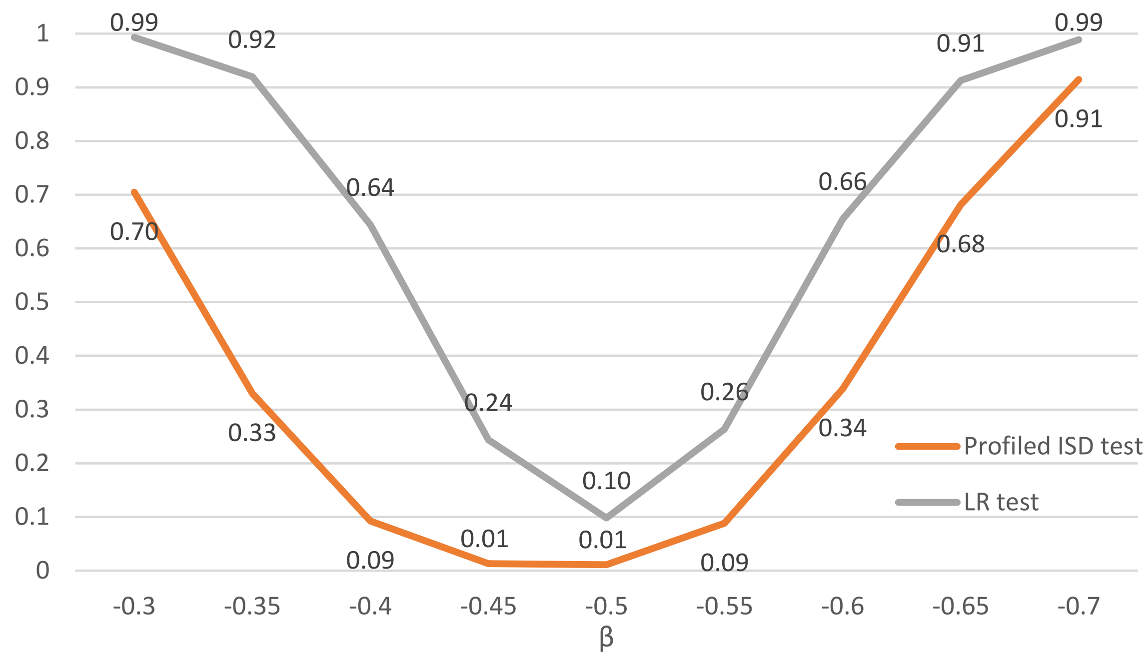
Note: 1000 Monte Carlo simulations, 500 bootstrap replications. All simulations were performed for the non-normalized statistic in (6).

set. We interpret the results for the non-normalized test as a manifestation of this problem.

F.2 Power loss under point-identification

As mentioned in Section 4 our approach will lead to conservative inference when censoring is in fact independent (i.e. possibly correlated with X but not with V). In order to get some insight about the power cost of using our method compared to the standard maximum likelihood approach with independent censoring, we run additional Monte Carlo simulations

Figure 8: Power curves for profiled ISD test (arbitrary censoring) and likelihood-ratio test (independent censoring), design 1.



Note: The orange line shows power of the profiled ISD test using our moment inequality approach (with discrete X). The grey line shows the power of the likelihood ratio test using standard maximum likelihood estimates from the MPH model with Weibull hazard and Gamma distributed unobserved heterogeneity (without discretizing the covariates). We run 2000 simulations with $n = 4000$.

and compare the power curves of the standard method and our method around the true parameter value ($\beta = -0.5$) for design 1 (point-identified). Figure 8 shows the results. We can see that the power cost of using our method versus the standard method is significant, especially around the true value $\beta = -0.5$, but the differences start to narrow quickly for values ± 0.2 from the true value. Overall, we think this cost is still small compared to the cost of mistakenly assuming independent censoring, as portrayed in Section 2.2 (cf. Figure 1).

Table 6: Cox model estimates

	(1)	(2)	(3)	(4)	(5)
log UI benefit	-0.504**	-0.520**	-0.455**	-0.487**	-0.410**
	[-0.795,-0.214]	[-0.795,-0.246]	[-0.727,-0.183]	[-0.791,-0.182]	[-0.726,-0.0930]
log annual wage	0.0522**	0.103***	0.0948***	0.0865**	0.113***
	[0.0153,0.0892]	[0.0663,0.139]	[0.0480,0.142]	[0.0297,0.143]	[0.0605,0.166]
av. unempl. rate	-0.110***	-0.105***	-0.107***	-0.0987**	-0.0473
	[-0.161,-0.0596]	[-0.154,-0.0564]	[-0.153,-0.0604]	[-0.165,-0.0324]	[-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	$10 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$
N	4529	4529	4054	4529	4529

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

G Application: additional results

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

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 6.

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).[15] 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.

[15] 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.

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

Table 7: 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.

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.

G.3 Comparison with previous estimates

Table 8: Comparison with previous estimates

	assumptions			data	90% CI
	hazard	V	$V \perp C$		
column (2) of Table 3	Weibull	yes	no	SIPP 1985-2000	[-0.81, 0.48]
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.

G.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. Comparing the results, given in Table 9, 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.

H 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

Table 9: Confidence sets for the elasticity of unemployment exit rate with respect to unemployment benefit, 90% level

	(1) minimum within quantile	(2) minimum within quantile	(1) maximum within quantile	(2) maximum within quantile
log UI benefit	[-1.15,0.36]	[-0.5,0.37]	[-1.75,0.5]	[-0.7,0.42]
log annual wage	x	x	x	x
av. unemployment rate	x	x	x	x
age		x		x
married		x		x
discretization	$10 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$	$10 \times 2 \times 3$	$4 \times 2 \times 2 \times 2 \times 2$
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. $10 \times 2 \times 3$ means 10 values of log UI benefit, 2 values of log annual wage, 3 values of unemployment rate. The number of bootstrap replications is 500 and $\kappa = 0.5$.

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, 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:

$$U_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + b_t) + J_{t+1}(A_{t+1}) \tag{14}$$

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:

$$\psi_s(s_t, A_t) = V_t(A_t) - U_t(A_t). \tag{15}$$

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 (14) 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 (15) 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.