Estimating the non-compliance in employment experiments. An Application to the Illinois Bonus Experiment[*]

By Marcel Voia [†]

June 2005

Abstract

This paper develops an estimation methodology that applies to employment experiments with selection problems. In particular a counterfactual estimator is constructed for a non-compliance group. The estimation methodology can be used to answer policy questions that are relevant for employment experiments. Questions such as: How much would average unemployment duration decrease if a specific program were introduced generally? How much money would be saved if the program were introduced generally? What are the implications of the experiments for policy design? could be answered. To show the relevance of the new estimator an application to the Illinois Bonus Experiment is considered. In this particular case the estimator allows for the counterfactual assumption that all eligible individuals collect the bonus and thus, the bonus becomes available for the population at no cost. To construct the counterfactual duration, a mixed proportional hazard model that allows for unobserved heterogeneity and partial compliance is estimated. The distribution of unemployment durations for the non-treated individuals that would occur if they were treated is simulated. I find that allowing for partial compliance and unobserved heterogeneity, a generalization of the program to all eligible individuals reduces significantly the duration of unemployment. Also, the results suggest that other UI policies different than the regular ones could increase the monetary gains for the state UI office.

Keywords: Duration Models, Counterfactual, Treatment Effects
JEL Classification: C41, C24, J64

1 Introduction

This paper introduces a new estimator for duration models that can be applied to experiments with selection problems. The new estimator uses all the available data and it is an application of the integrated hazard principle as developed by Woutersen (2000). The

---

[*] I am very grateful to Tiemen Woutersen for all his comments and suggestions and I gratefully acknowledge stimulating suggestions from Brian McCall, Chris Robinson, Lance Lochner, Todd Stinebrickner and Audra Bowlus. All errors are mine.

[†] Correspondence addresses: Carleton University, Department of Economics, Social Science Centre, Ottawa, Ontario, Canada. Email: mvoia@connect.carleton.ca.

new estimator estimates the program effect under the assumption that an employment experiment is transformed into a policy, and thus, the non-compliers in a treatment setting become participants under the new policy. Therefore, the new estimator estimates the duration of unemployment for the non-compliers as if they were treated or in other words their counterfactual duration of unemployment.

In the biostatistics literature, Tsiatis (1990) and Robin and Tsiatis (1991) censor all their observations to deal with partial compliance. They hereby discard a large part of the data and cannot make statements about the hazard over the intervals that were censored away. To deal with partial compliance in the treatment group of the Illinois bonus experiment, Bijwaard and Ridder (2002) propose a two-stage linear rank ($2SLR$) estimator to mixed proportion hazard (MPH) models with exogenous and endogenous regressors. Their estimator is consistent if there is selective compliance in the treatment group of a randomized experiment and if the outcome variable is a censored duration. Their results show that compliance in the Illinois bonus experiment was selective, that is, the individuals from the treatment group that decline to participate are not doing it at random, but based on some characteristics, which are important to the re-employment hazard and thus, they found evidence of unobserved heterogeneity. They also found that the effect of the bonus is largest for individuals with a benefit exhaustion probability in the highest quartile and thus, the treatment effect is not the same for all individuals. Bijwaard and Ridder's estimator is inefficient in the sense that it only uses a fraction of the data.

The semiparametric estimator of Ridder and Woutersen (2002) uses all the data and is more efficient. Their estimator uses the method of moments to consistently estimate duration models with exogenous and endogenous regressors.

The estimation methodology introduced in this paper can be used to answer policy questions that are relevant for employment experiments such as: How much would average unemployment duration decrease if a specific program were introduced generally? How much money would be saved if the program were introduced generally? What are the implications of the experiments for policy design?

The above questions are answered using the Illinois bonus as an example of a new unemployment policy and by assuming that once the bonus is part of a new UI program it has a zero cost for the unemployed people (the cost of the experiment is removed).

Following the reasoning of two decades of econometric analysis of duration models, (see Lancaster (1990) and Van den Berg (2000) for overviews), I argue that one needs to control for unobserved heterogeneity (the difference in heterogeneity between the control and treatment groups due to the selection into treatment) to assess causality in a duration model. Therefore, to estimate the hazard rate a MPH model is considered. The MPH allows for a discontinuous baseline hazard $\lambda(t)$ or in other words allows for time-varying treatment effect. Given the full compliance in the control group, to estimate the baseline hazard for the non-compliers (non-participants), the moment conditions of Ridder and Woutersen are used. After the MPH model is estimated, the duration of unemployment for the non-compliers as if they were treated or in other words their counterfactual duration of unemployment is computed using the fact that the semi-integrated hazard (the integrated hazard divided by the unobserved heterogeneity) in the treatment group is equal with the one in the non-treated group. The counterfactual distribution is then estimated using a non-parametric bootstrap method.

My findings confirm the presence of unobserved heterogeneity in the Illinois bonus experiment. Thus, the results show a significant impact on the treatment effect when we consider that all individuals eligible for the bonus are considering the bonus in their labour decisions.

This paper is organized as follows. Section (2) shows how the integrated hazard principle is used to estimate simple hazard models and the model of the paper. Section (3) discusses counterfactuals in duration analysis and shows how to approximate the counterfactual distribution. Section (4) describes the data of this paper. Section (5) gives estimation results. Section (6) concludes.

2 The Integrated Hazard Principle

In this section, I show how to estimate simple hazard models. The estimator is using the identification results of Elbers and Ridder (1982) and the 'integrated hazard principle' of Woutersen (2000). The integrated hazard principle suggests to use those parameter values (or functions) as estimates for which the integrated hazard is an unit exponential variable. Let Z denote the integrated hazard

$$Z = \int_0^T \theta(t,x)dt \sim \varepsilon(1), \text{ where } \theta(t,x) \text{ is the hazard rate.} \tag{1}$$

Equation (1) can be used to estimate parameters of the hazard function. To illustrate how this is done, consider some simple problems.

Example 1. Let $t_1, ..., t_N$ be independent durations with constant hazard $\theta(t) = \lambda$ so the integrated hazard $Z_i = \lambda t_i$, with $i = 1 \dots N$. The integrated hazards are independent unit exponential: $Z_i \sim \varepsilon(1)$. Equating the sample analogue of the integrated hazard to one gives:

$$\frac{\sum_{i=1}^N \lambda t_i}{N} = 1.$$

This suggests the following estimator for λ,

$$\hat{\lambda} = \frac{N}{\sum_{i=1}^N t_i},$$

which is the maximum likelihood estimator.

Equation (1) can also be used for censored duration data. Suppose a duration is right censored if it lasts longer than c where c is exogenous. Let Z' denote the integrated hazard of this potentially censored observation and let the indicator d be zero if the observation is censored and one otherwise. Then the expectation of Z' equals the expectation of d, i.e.

$$EZ' = Ed \text{ where } d = 0, 1. \tag{2}$$

See Woutersen (2000) for a formal proof.
Equation (2) suggests to censor the durations at c and to estimate the hazard $g(\lambda) = \frac{\sum d_i}{N} - \frac{\sum z_i'}{N}$ is used.

Example 2. Let $T_1, ..., T_N$ be independent durations with the following hazard

$$\theta(t) = \begin{cases} \lambda_1 \text{ if } t \leq c \\ \lambda_2 \text{ if } t > c \end{cases}$$

The indicator d_i is one if $t_i \leq c$ and zero otherwise. We can write the realization of the integrated hazard of individual i as

$$z_i = d_i \lambda_1 t_i + (1 - d_i)\lambda_1 c + (1 - d_i)\lambda_2(t_i - c)$$

and the realization of the integrated hazard for the censored observation as

$$z_i' = d_i \lambda_1 t_i + (1 - d_i)\lambda_1 c.$$

Using the moment equations $g_1(\lambda) = \frac{\sum_i z_i}{N} - 1$ and $g_2(\lambda) = \frac{\sum_i d_i}{N} - \frac{\sum_i z_i'}{N}$ gives the following estimates

$$\begin{aligned} \hat{\lambda}_1 &= \frac{\sum_i d_i}{\sum_i \{d_i t_i + (1 - d_i)c\}} \\ \hat{\lambda}_2 &= \frac{\sum_i (1 - d_i)}{\sum_i (1 - d_i)(t_i - c)}. \end{aligned} \tag{3}$$

These estimates are, again, the maximum likelihood estimates. The maximum likelihood estimator of example 1 and 2 is consistent and efficient. One can therefore argue that the integrated hazard 'should' yield the same estimating function. This equivalence result holds for any number of intervals over which the hazard is supposed to be constant [1].

Example 3. Consider now a model with unobserved heterogeneity, and let the hazard have the following form $\theta_i = \eta_i \gamma^R$, where R is either 0 or 1, η_i is the realization of a mixing distribution, and γ defines the treatment. Suppose we observe N independent durations and the treatment is randomly assigned to half of the population ($R = 0$ for non treated individuals and $R = 1$ for the treated ones). The random assignment ensures independence of η_i and the treatment variable. Let the realizations denoted by t_i, $i = 1, ..., N$, and define the semi-integrated hazard as

$$s_i = \frac{1}{\eta_i} \int_0^{t_i} \theta_i du = \gamma^R.$$

An extension of example 1 shows how s_i can be used for estimation. Consider the following moment function.

$$g(\gamma) = \frac{1}{N} \sum_i (s_i^R - s_i^{1-R}).$$

This moment function is monotonic in γ and has expectation zero at $\gamma = \gamma_0$, the true value of γ.

$$Eg(\gamma) = (\frac{\gamma}{\gamma_0})E\frac{1}{\eta} - E\frac{1}{\eta} = (\frac{\gamma}{\gamma_0} - 1)E\frac{1}{\eta}.$$

Proof: See appendix 1.
Therefore, the treatment given by the parameter γ can be consistently estimated.

---

[1] If the number of intervals increases with N then integrated hazard estimator is equivalent to the non-parametric estimator of Kaplan and Meier (1958).

2.1 The Mixed Proportional Hazard (MPH) model

The MPH model is a generalization of the Proportional Hazard model introduced by Cox (1972) and was introduced by Lancaster (1979) and Manton et all (1981) . Let the individual durations be denoted by t_i, $i = 1, ..., N$, then, the hazard for this type of model has the following form

$$\theta_i\left(t_i\right) = \eta_i \phi(x_i)\lambda(t_i),$$

where η_i is the realization of a mixing distribution (unobserved heterogeneity) that captures variables that are not in x, $\phi(x_i)$ is a function of a time-constant regressor x_i and $\lambda(t_i)$ denotes the baseline hazard.

Bijwaard and Ridder (2002) show that compliance in the Illinois bonus experiment was selective thus, they found evidence of unobserved heterogeneity and they suggested a MPH model to estimate the program effect.

This result can be checked by looking at the correlation between an estimated unobserved heterogeneity η_i $(\widehat{\eta_i})$ and the observables. To do this, firstly the data is smoothed. Given the hazard $\theta(t) = \frac{f(t;x)}{1-F(t;x)} \Rightarrow$ the integrated hazard is $z = \int \theta(t;x)dt = -\ln(1 - F(t;x))$, where $F(t;x)$ is uniformly distributed between 0 and 1 and thus, $z \sim \varepsilon\left(1\right)$. Using the fact that the integrated hazard is unit exponential, the following estimator can be generated

$$\widehat{\eta_i} = \frac{1}{\widehat{s_i}},$$

where s_i is the semi-integrated hazard $\left(s_i = \frac{z_i}{\eta_i},\text{ with } z_i = \eta_i \phi(x_i)\Lambda\left(t_i\right)\right)$. An estimator for s_i is $\widehat{s_i} = \widehat{\Lambda}\left(t_i\right)$ with $\widehat{\Lambda}\left(t_i\right) = \int_0^{t_i} \widehat{\lambda}(t)dt$. Given that in the control group there is full compliance, the estimates for $\widehat{\lambda}(t)$ were obtained using the moment conditions of Ridder and Woutersen (2002) with data from this group.

Replacing η_i by $\widehat{\eta_i}$ in z_i and taking $\ln \widehat{\eta_i}$ gives

$$\ln \widehat{\eta_i} = -\ln \Lambda\left(t\right) - x_i\beta - \ln \varepsilon_i.$$

Regressing $\ln \widehat{\eta_i}$ on the non-participation group's observables helps identify if the unobserved heterogeneity is present in the experimental data. If there is correlation between observables and unobserved heterogeneity in the non-participation group, we have that the non-compliance in the experimental data was selective and thus, the MPH is a suitable model to analyze the duration of unemployment of this group.

2.1.1 Estimation of the MPH Model

Joint estimation of $\phi(x)$, $\lambda(t)$ and the distribution of η has to deal with one of the following complications. If a parametric mixing distribution is chosen, then, $\lambda(t)$ can be estimated by a parametric or nonparametric technique. However, Heckman and Singer (1985) show that the estimates are very sensitive to the choice of the mixing distribution (see appendix 2). The other option is to estimate the mixing distribution nonparametrically. As Horowitz (1999) shows, however, a deconvolution problem cannot be avoided and the rate of convergence would be very slow for all parameters (and functions). To avoid problems of joint estimation, I use moment functions whose expectation does not depend on the mixing distribution and were introduced by Ridder and Woutersen (2002). As 'building blocks' for these moment functions I use the indicator "*d*" of equation (2) and what it is called the "semi-integrated hazard" (s).

Next, I show how to estimate a MPH model with censored data.

The Model In example 2, data was censored at c and it was derived an estimator that was equivalent to maximum likelihood estimator. Given that a flexible base-line hazard is preferable, to estimate $\lambda(t)$ appropriate moment conditions are needed. In general, the durations can be artificially censored in order to create more moments. Suppose the duration of the treatment group is censored at c_0, and denote the semi-integrated hazard by s_i', then

$$Es_i' = E\frac{z_i'}{\eta_i} = \frac{\eta_i}{\eta_i}(1 - e^{-\eta_i s_{i\max}}) = (1 - e^{-\eta_i s_{i\max}})$$

where $s_{i\max} = \frac{1}{\eta_i}\int_0^c \theta_i dt$. Similarly, the indicator d_i has the following expectation

$$Ed_i = (1 - e^{-\eta_i s_{i\max}}).$$

Suppose the regressor x has two different values, x_0 and x_1 and let N_0 and N_1 denote the number of individuals with $x = x_0$ and $x = x_1$, respectively. All the observations with $x = x_0$ can be censored at c_0 and those with $x = x_1$ at c_1. The same normalized censoring time for the two spells is desired. Thus, to find c_1, the following moment function is used

$$g_1(c_1(\theta)) = \frac{\sum_{x=x_0} d_i}{N_0} - \frac{\sum_{x=x_1} d_i}{N_1}, \tag{4}$$

or, equivalently, $g_1(c_1(\theta)) = \bar{d}_0(c_0) - \bar{d}_1(c_1)$ yields a censoring point c_1 by solving

$$\frac{1}{N_0}\sum_{x=x_0} d_i - \frac{1}{N_1}\sum_{x=x_1} d_i = 0, \tag{5}$$

and $\bar{d}_0(c_0) < \bar{d}_1(c_1)$ implies $c_1 < c_0$. The expectation of this equation is satisfied if $s_{i\max}$ is the same for $x = x_0$ and $x = x_1$, i.e.

$$s_{i\max}(c_0, x_0) = s_{i\max}(c_1, x_1). \tag{6}$$

Suppose that the assumptions of Elbers and Ridder (1982) hold so that $\phi(x_0) \neq \phi(x_1)$ for time-invariant x. Then, $c_0 \neq c_1$ and equation (6) yields restrictions on the parameters. In fact, artificially censoring the data many times gives as many restrictions. Suppose that the regressor x is not relevant for the hazard rate so that $\phi(x_0) = \phi(x_1)$ and the identification condition of Elbers and Ridder fails. This touches upon what Elbers and Ridder consider the 'main contribution' of their paper, that time dependence and unobserved heterogeneity can be distinguished due to the variation of individual probabilities with the explanatory variable x.

Equation (6) sheds some light on the need of "variation of the individual probabilities with the explanatory variable x". Let $\phi(x_0) = \phi(x_1)$ then equation (6) holds for $c_0 = c_1$ so that

the function $\phi(x)$ is not identified through this restriction. This is in accordance with Elbers and Ridder since the MPH model is not identified in that case.
Note that

$$\begin{aligned} s_{0,\max} &= \frac{1}{\eta}\int_0^{c_0} \theta_i dt = \phi(x_0) * c_0 \\ s_{1,\max} &= \frac{1}{\eta}\int_0^{c_1} \theta_i dt = \phi(x_1) * c_1. \end{aligned}$$

This suggest the following estimator for $\frac{\phi(x_1)}{\phi(x_0)}$.

$$\widehat{\frac{\phi(x_1)}{\phi(x_0)}} = \frac{c_0}{c_1}. \tag{7}$$

To estimate the hazard rate, the following moment function that is a function of the parameters and c_1 is used

$$g_2(\theta, c_1(\theta)) = \frac{\sum_{x=x_0} d_i}{N_0}\frac{\sum_{x=x_1} s_i}{N_1} - \frac{\sum_{x=x_1} d_i}{N_1}\frac{\sum_{x=x_0} s_i}{N_0}. \tag{8}$$

Note that $\frac{\sum_{x=x_0} d_i}{N_0}$ is just the average of the indicators and not a function of the parameters. At the true value of the parameter of interest, θ_0, $\frac{\sum_{x=x_0} s_i}{N_0} = \frac{\sum_{x=x_0} \frac{1}{\eta}(1-e^{-\eta s_{\max}})}{N_0}$. The distribution of η_i does not depend on x_i and therefore $\frac{\sum_{x=x_0} s_i}{N_0} = \frac{\sum_{x=x_1} s_i}{N_1}$. This implies that the expectation of $g_2(\theta_0, c_1(\theta_0))$ is zero (see appendix 3). Note that both $g_1(c_1(\theta))$ and $g_2(\theta, c_1(\theta))$ are just functions of $\phi(x_0)$ and $\phi(x_1)$ since the durations were censored at (or before) c. Therefore, misspecification of the baseline hazard for $t > c$ has no effect on the consistency.

The next section shows how to estimate the counterfactual duration for the non-compliance group using the following identification condition

$$\frac{\sum_{x=x_0} s_i}{N_0} = \frac{\sum_{x=x_1} s_i}{N_1}. \tag{9}$$

3 Counterfactuals for non-compliers

3.1 Counterfactuals in duration analysis

The counterfactuals were introduced in econometric analysis to make inference of causal effects in treatments by analyzing causation in terms of possible worlds, which are of importance for empirical social scientists. Thus, counterfactuals characterizes a possible world with minimum deviation from the actual world. Three approaches are used in literature to model causation:

a) Structural Equation Model represented by the following type of equations

$$Y = X\beta + \varepsilon,$$

where β capture all causal connection between Y and X.

b) Potential Outcome Model (POM) of causation, which is used to explain the effect on the outcomes of some particular treatment (treatment group) relative to other particular treatment (control group), see Rubin (1977).

c) Directed Acyclic Graphs, which is a recent development and uses graphical approaches to causation, see Galles and Pearl (1998).

We will focus on counterfactual analysis using POM. The choice of the counterfactual has to be close enough to the data so that statistical methods provide empirical answers and the research questions are well defined.

Thus, one of the most important conditions necessary to have a counterfactual causal question is the relativity condition (the effect of one treatment is always relative to the effect of another treatment, in other words that causal inference about a treatment in the actual world is based on the counterfactual relation to what would have happened under exposure to a treatment in the alternative possible world). Rubin (1980, 1986) considered that the basic assumption necessary to assess potential outcomes and infer meaningful causal statements is the stable-unit-treatment-assumption, which states that the outcome of a treatment for a given unit is the same independent of the mechanism that is used to assign the treatment to that unit. He also argued that one of the most important causal parameters to be analyzed is the average causal effect of a treatment or the average treatment effect (of a treatment in the actual world versus a treatment in the possible world), and was defined as the difference between the expected value of the unit-level of outcomes in the actual world and possible world

$$ATE = E\left(Y_a\right) - E\left(Y_p\right).$$

To make the model applicable we need to have satisfied the independence assumption, which can be obtain by looking at a randomized experiment in which units are randomly assigned to different treatments. Thus, the initial population and the subpopulations in the treatments do not differ from each other on average. The Illinois bonus experiment satisfy all the above conditions and thus, a counterfactual analysis on the non-compliers in the treatment group can infer results about the applicability of the program to the overall population.

3.2 Counterfactuals for non-compliers

To get a consistent estimate for the counterfactual duration all data is considered (both the complete or uncensored duration spells and the incomplete or censored duration spells). Therefore, the counterfactuals are computed for both the uncensored spells and the censored spells.

3.2.1 Counterfactuals for complete spells

To compute the counterfactuals for the complete spells of unemployment, the non-compliers are hypothetically randomized into treatment. Now, consider the MPH model with hazard rate $\theta\left(t_i\right) = \eta_i \phi\left(x_i\right) \lambda\left(t_i\right)$,where η_i is the unobserved heterogeneity, $\lambda\left(t_i\right)$ is the baseline hazard, $\phi\left(x_i\right) = e^{\beta x_i}$. The integrated hazard z_i for the non-treated (non-complier) observation is $z_i = \eta_i e^{x_i \beta} \Lambda\left(t_i\right)$, and the semi-integrated hazard for the same observation is $s_i = e^{x_i \beta} \Lambda\left(t_i\right)$.

Suppose x_i^* is the counterfactual treatment, then using the identification assumption (9) , the counterfactual duration, t_i^* can be determined in the following way

$$\eta_i e^{x_i^* \beta} \Lambda\left(t_i^*\right) = z_i \Leftrightarrow e^{x_i^* \beta} \Lambda\left(t_i^*\right) = s_i,$$

$$\Rightarrow t_i^* = \Lambda^{-1}\left(\frac{s_i}{e^{x_i^* \beta}}\right) = \Lambda^{-1}\left(e^{\left(x_i - x_i^*\right)\beta} \Lambda\left(t_i\right)\right). \tag{10}$$

See a complete derivation of the counterfactual duration for the uncensored spells in the appendix 4.

An estimate of the counterfactual duration can be obtained by replacing $\lambda\left(t_i\right)$ with their estimates. Here we assume that $\lambda\left(t_i\right) = \left\{\lambda_j\right\}_{j=1,2,..,T}$ with $T =$ the number of weeks of UI benefits. To estimate the baseline hazard rates, the moment conditions developed by Ridder and Woutersen (2002) and data from the control (s_{0j}) and treatment (s_{1j}) groups are used:

$$g_j = s_{0j}(\lambda_j, c_i) - s_{1j}(\lambda_j, j), \ j = 1, 2, ..., T,$$

with g_j linear in λ_j. As an estimate for $e^{\left(x_i - x_i^*\right)\beta}$ use $\widehat{\frac{\phi(x_{i1})}{\phi(x_{i0})}} = c_j$.

Having estimates for the baseline hazard and the treatment effect, the counterfactuals are obtained using (10). To get the distribution of the counterfactual duration t_i^*, a non-parametric bootstrap[2] method is employed.

---

[2] The methodology is used to estimate the distribution of an estimator $\widehat{\theta}$ for a given parameter θ. The validity of the bootstrap follows from the fact that there exists a pivotal statistic, in this case

$$\sqrt{N}\frac{\text{counterfactual}\left(t_0\right) - E(\text{counterfactual}\left(t_0\right))}{\sigma} \rightarrow_d N\left(0, 1\right).$$

The employed methodology is nonparametric and is simulating bootstrap samples (pseudo-data). A single bootstrap sample is obtained by sampling randomly, with replacement, n observations from the original sample. Using the realizations $t_1, t_2, \ldots, t_n$ of the data ,construct the empirical distribution function, *edf* as $\widehat{F} := \frac{1}{n}\sum_{i=1}^{n} 1_{\{T_i < t\}}$,where $F\left(t\right)$ is the *cdf* of the probability model, that is $F = \Pr[T \leq t]$.Once $\widehat{F}$ is constructed, use it to generate new random variables from this distribution by inverting the *edf* and plugging uniform random variables into the resulting function. This is a bit problematic because the *edf* is not monotonic, and thereby is not easily invertible. In practice, a "generalized inverse" is constructed. Thus, use this method to generate at least $B = 1000$ new data sets and compute the statistic for some parameter θ on each pseudo-data set

$$\begin{aligned}
\widehat{\theta}^{(1)} &= \widehat{\theta}\left(T_1^{(1)}, T_2^{(1)}, \ldots, T_n^{(1)}\right) \\
\widehat{\theta}^{(2)} &= \widehat{\theta}\left(T_1^{(2)}, T_2^{(2)}, \ldots, T_n^{(2)}\right) \\
&\vdots \\
\widehat{\theta}^{(B)} &= \widehat{\theta}\left(T_1^{(B)}, T_2^{(B)}, \ldots, T_n^{(B)}\right),
\end{aligned}$$

which gives B new estimates of θ. This estimates are intimately related to the original $\widehat{\theta}$. Estimate the bias and the variance of the estimator using

$$\begin{aligned}
bias\left(\widehat{\theta}^{(*)}\right) &= \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{(b)} - \widehat{\theta} \\
Var\left(\widehat{\theta}^{(*)}\right) &= \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\theta}^{(b)} - \overline{\widehat{\theta}^{(*)}}\right)^2,
\end{aligned}$$

3.2.2 Counterfactuals for incomplete spells

Considering the period of UI benefits as T weeks, all the observations that exceed T weeks are censored at T. Assume that censored individuals are of two types: high hazard type $\eta_i = \eta_H$ with probability p and low hazard type η_L with probability $1-p$. To estimate η_H, η_L and p, the observations from the weeks $T-3$, $T-2$, $T-1$ and T are used to construct the following moment conditions:

$$\overline{F}(j-1) - \overline{F}(j) = E_\eta e^{-\eta\Lambda(j-1)} - E_\eta e^{-\eta\Lambda(j)},$$

with

$$E_\eta e^{-\eta\Lambda(j)} = pe^{-\eta_H\Lambda(j)} + (1-p)\, e^{-\eta_L\Lambda(j)}, \text{ where } j = T-2, T-1 \text{ and } T.$$

Then, the estimates η_H, p and η_L are used to estimate the survival probabilities for the censored observations. The model considers now T exogenously generated censoring points $(c_{0j},\ j = 1, ..., T)$ for the treatment group. Let $\lambda(t_i)$ be a piecewise constant function that allows for a different baseline hazard before and after the censoring point c_{0j}. For convenience, the individuals with $x_i = x_{i1}$ are considered from the control group and those with $x_i = x_{i0}$ from the treatment group. The control and treatment groups are censored endogenously at c_{1j} and c_{0j}. The censoring points are chosen in such a way that the survival probabilities in the two groups are equal. Thus, a sequence c_{0j} and c_{1j} is obtained by equating

$$\begin{aligned} \Pr(x_i = 1, j) &= \Pr(x_i = 0, c_{0j}) \\ \Pr(x_i = 0, j) &= \Pr(x_i = 1, c_{1j}). \end{aligned}$$

The following estimator for $\frac{\phi(x_{i1})}{\phi(x_{i0})}$ is used,

$$\left(\widehat{\frac{\phi(x_{i1})}{\phi(x_{i0})}}\right)_j = \frac{c_{0j}}{c_{1j}}, \text{ see } (7).$$

To find the counterfactual duration (t_i^*), hypothetically randomize into treatment the non-compliance group and then, use the identification assumption (9) to solve for t_i^*. Thus, the integrated hazard z_i for a non-complier without treatment is

$$z_i = \eta_i e^{x_i\beta}\Lambda(t_i) \ \Rightarrow s_i = e^{x_i\beta}\Lambda(t_i).$$

Suppose x_i^* is the counterfactual treatment for the non-complier, then the integrated hazard for the non-complier "treated" is

$$z_i^* = \eta_i e^{x_i^*\beta}\Lambda(t_i^*) \Rightarrow s_i^* = e^{x_i^*\beta}\Lambda(t_i^*).$$

Given the identification $s_i = s_i^*$,

$$\Rightarrow t_i^* = \Lambda^{-1}\left(\frac{s_i}{e^{x_i^*\beta}}\right) = \Lambda^{-1}\left(e^{(x_i - x_i^*)\beta}\Lambda(t_i)\right). \tag{11}$$

---

where the asterisk $^{(*)}$ holds for bootstrap quantities under the probability mechanism $\widehat{F}$ and $\overline{\widehat{\theta}}^{(*)} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^{(b)}$.

See a complete derivation of the counterfactual duration for the censored spells in the appendix 5.
The condition that $t_i^* < t_i$ is that $x_i - x_i^* < 0$. An estimate of the counterfactual duration can be obtained by replacing $\lambda_1, \lambda_2, \ldots, \lambda_T$, with their estimates. To estimate the baseline hazard rates, moment conditions developed by Ridder and Woutersen (2002) and data from the control (s_{0j}) and treatment (s_{1j}) groups are used:

$$g_j = s_{0j}(c_{0j}, \lambda_j) - s_{1j}(c_{1j}, \lambda_j), \ j = 1, \ldots, T.$$

After, replacing $e^{\left(x_{i1}-x_i^*\right)\beta}$ by $\left(\widehat{\frac{\phi(x_{i1})}{\phi(x_{i0})}}\right)_j = \frac{c_{0j}}{c_{1j}}$, the counterfactual duration for the non complier (t_i^*) can be computed using (11) and its variance using the bootstrap method.

Results for the Illinois bonus experiment are presented in Tables: 8, 9, 10, and 11 and show that the causal effect for the non-compliance group under the assumption that the non-compliers also get the treatment is significant and represents a 1.86 reduction in unemployment duration at an average duration of unemployment of 18.46 weeks.

4 Experimental Data

The data used to test the new methodology is from the Illinois bonus experiment. The Illinois bonus experiment is an interesting experiment about unemployment duration which was conducted by the Illinois Department of Employment Security. In particular, a randomized group of people who became unemployed was offered the chance to participate in a program in which unemployed people were offered US $500 bonus if they found a job within 11 weeks and kept it for at least four months. Some people declined to participate even if they received an offer to participate in the program. This group is called the non-compliance group.

The experiment randomly assigned those in the eligible population to one of three groups, which are identified by Meyer (1996) as control group, Claimant Experiment, and Employer Experiment. The goal of the experiment was to explore if the unemployment duration was reduced when a bonus was paid to Unemployment Insurance (UI) beneficiaries (treatment 1) or to their employers (treatment 2) relative to a randomly selected control group. I concentrate the attention on the Claimant Experiment (treatment 1), which consists of a random sample of new claimants for UI that received a $500 bonus if they found a job (of 30 hours or more per week) in less than 11 weeks after filing for UI and held that job for at least 4 months. The size of the bonus reflected a balancing of the experiments' budget constraint (a maximum of $750,000 in bonus payments) against an arbitrary judgement about how small a bonus could be to generate a response (5% of annual wage or 4 weeks of UI payments). The 11 week period was chosen to be approximately 40% of the potential duration of benefits in Illinois, which is 26 weeks. The minimum of 4 months employment was required to avoid fraudulent hire and to avoid payment of bonuses to seasonal workers and employers. Eligibility criteria excluded younger and older claimants in order to reduce the number of complicated factors such as: special programs for young people, and incentives to retire early for older workers which can influence the job-finding behavior of those enrolled in the experiment. Three variables can be used to control for the size of the sample: the number of sites (22), the length of the enrollment period, and the proportion of claimants selected at any given site. More sites permitted a shorter enrollment period, originally designed to be 13 weeks, ultimately 16. Thus, in the treatment 1 group, 4186 individuals were selected, 3527 (84%) (compliance group) agreed to participate in the experiment, while 659 (16%) (non-compliance group) did not agree. The individuals from the control group were excluded from participating in the experiment, they actually did not know that the experiment took place. The control group consisted of 3952 individuals.

5 Estimation Results

Regressing $\ln \widehat{\eta}_i$ on the non-participation group's observables {not exhausted benefits, (age-average(age), log(base period earnings)-average(log(BPE), dummy for race, dummy for sex), odd number of benefit weeks} helps identify if unobserved heterogeneity is present in the Illinois experimental data.

The results from Table 1 show a significant correlation between $\ln \widehat{\eta}_i$ and not exhausted benefits (NEB) and $\ln \widehat{\eta}_i$ and odd number of benefit weeks ($ONBW$).
Since the standard duration models assume that heterogeneity is uncorrelated with unobserved characteristics, and given that there is correlation between observables and unobserved heterogeneity in the non-participation group, we have that the non-compliance in the Illinois Bonus experiment was selective and thus, the MPH is a suitable model to analyze the duration of unemployment in this particular experiment.
where the dependent variable is $\ln \widehat{\eta}_i$, $Corr = Correlation\,(variable\,(X)\,, \ln \widehat{\eta}_i)$, NEB = Not exhausted benefits (first spell), ONBW =Odd number of benefit weeks, $\ln (wb\_d)$ = log(wkbenefit + depallow) with wkbenefit = weekly benefit amount, and depallow = weekly dependents' allowance.

5.1 Observables average comparison for control and non-compliance groups

A necessary condition for a counterfactual analysis is that the initial population and the subpopulations in the treatments do not differ from each other on average. The results are presented in Tables 2,3 and 4 for observables that show some difference in distribution between the control and non-compliance group. .
We observe that, even if there are slightly differences in average, these differences are not statistical significant and the conditions for a counterfactual analysis on non-compliers are satisfied. Also, we can observe that in the non-compliance group there are individuals with higher age and higher BPE.

5.2 Results for the causal treatment effect

Tables 5 and 6 show the same treatment effect and standard errors either by regressing the duration of unemployment on the randomized indicator for treatment or by computing the causal treatment effect and bootstrapping it to get standard errors.

Using IV regression with $\widehat{X}$ as an instrument for the treatment, gives the treatment effect on the treated. In this case, a slight increase in the $R-squared$ can be observed, and that the effect of the treatment effect on the treated is 1.8023 weeks, which is higher than the

average treatment effect on the overall treatment group (1.3391). This result shows that, if the program is applied to the overall population than smaller unemployment durations for the unemployed are obtained and, thus, the UI claims will be lower. These IV results are valid only if the treatment effect is the same for all individuals (see Heckman and Vytlacil (2000)). Given that individuals with shorter durations of unemployment have different treatment effects than the ones with longer durations of unemployment, the IV results did not give an accurate measure for the treatment effect of the treated. To overcome this problem, consider the non-compliance group and compute counterfactual durations (duration of unemployment for the non-treated as they were treated) obtaining the treatment effect on this group.
The following results are referring to the counterfactual duration for non-compliers.

5.3 Counterfactuals results when all available data is considered

Looking at the average duration of unemployment (Table 2) for non-compliers we observe that it is smaller (with 0.1336 weeks) than the average duration of unemployment for the control group (both groups being not treated). Also, the distribution of unemployment duration for non-compliers (fig.7), is different from the distribution of unemployment duration for control group (fig.5), in the sense that the duration of unemployment for non-compliers dominates the one from the control group. Thus, it seems that non-compliers have different incentives to find a job in the absence of a treatment (they are more willing to find a job quicker) and thus, if non-compliers are treated, they have a smaller duration of unemployment than the compliers in the treatment group. The counterfactual durations and they standard errors obtained using bootstrap method are presented in the Table 8, empirical hazards and standard errors for the non-compliance group are presented in Table 9.
From Fig 2, we observe that during the eligibility period, the hazard rate of the treated group is higher than the one in the control group, and there is no accounted difference after week 11.

5.3.1 Results for the causal treatment effect

Table 10 compares the average unemployment duration for non-compliers when they would get the treatment with the average unemployment duration in the other groups computed using all data.
Table 11 presents the average treatment effect when all data in non-compliance group is considered.
Observe that if we account for both the censored and uncensored observations, the average duration of unemployment for the counterfactuals is larger than the one obtained using IV. Looking at the causal effect, observe that the average unemployment duration decreased by about 1.86 weeks if the non-treated were treated, which implies that IV results (treatment effect on the treated) underestimate the treatment effect. The result shows that when we account for censored observations, the new estimator consistently estimates the treatment effect.

Table 12 presents the Program effects results (slope estimator for the treatment effect, $\widehat{\beta}$) for the Illinois Bonus Experiment using ML, ITT (Meyer's estimator), $2SLR$ (Bijwaard and Ridder's estimator), IH estimators with a flexible specification of the duration dependence. Results for Standard Errors for integrated hazard (IH) estimator are obtained using Bootstrap Method.

The IH estimator is obtained using the fact that when we account for all observations we have the following treatment

$$e^{\beta(x_{0i}-x_{1i})} = \frac{\phi\left(x_{0i}\right)}{\phi\left(x_{1i}\right)} = \frac{c_{1i}}{c_{0i}},$$

which gives

$$\widehat{\beta}_{IH} = \frac{1}{N}\sum_{i=1}^{N} \ln\left(\frac{c_{1i}}{c_{0i}}\right).$$

6 Conclusion

This paper develops an estimation methodology that applies to employment experiments with selection problems. In particular a counterfactual estimator is constructed for a non-compliance group. The estimation methodology uses the IH principle on a MPH model which allows for time-varying treatment effect (discontinuous $\phi(t)$). This estimator is used to evaluate the re-employment Illinois bonus experiment from a different perspective. Thus, the bonus becomes part of a new UI policy. In this particular case the estimator allows for the counterfactual assumption that all eligible individuals collect the bonus and thus, the bonus become available for the population at no cost. The distribution of unemployment durations for the non-treated individuals that would occur if they were treated is simulated. I find that allowing for partial compliance and unobserved heterogeneity, a generalization of the program to all eligible individuals reduces significantly the duration of unemployment. Also, the results suggest that other UI policies different than the regular ones could increase the monetary gains for the state UI office.

7 Appendix

Appendix 1 Let the hazard have the following form
$\theta_i = \eta_i \gamma^R$, with $\left\{ \begin{array}{c} R=0 \text{ for } \frac{N}{2} \text{ individuals} \\ R=1 \text{ for the other } \frac{N}{2} \text{ individuals} \end{array} \right.$ and where η_i is the realization of a mixing distribution. The random assignment ensures independence of η_i and X. Let the realizations denoted by t_i, $i = 1, ..., N$. The semi-integrated hazard has the following form

$$s_i = \frac{1}{\eta_i} \int_0^{t_i} \theta_i du = \gamma^R t_i.$$

Consider the following moment function.

$$g(\gamma) = \frac{1}{N} \sum_i (s_i^R - s_i^{1-R}) = \frac{1}{N} \sum_i \{(\gamma t_i)^R - t_i^{1-R}\}.$$

We have that $Et_i = \frac{1}{\gamma_0^R} E \frac{1}{\eta_i}$ and thus,

$$\begin{aligned} Eg(\gamma) &= \frac{1}{N} \left(\sum_{i,R=1} \{\gamma Et_i - 1\} + \sum_{i,R=0} \{1 - Et_i\} \right) \\ &= \frac{1}{N} \left(\sum_{i,R=1} \frac{\gamma}{\gamma_0} E \frac{1}{\eta_i} - \frac{N}{2} + \frac{N}{2} - \sum_{i,R=0} E \frac{1}{\eta_i} \right) \\ &= \frac{1}{N} \left(\frac{\gamma}{\gamma_0} \sum_i E \frac{1}{\eta_i} - \sum_i E \frac{1}{\eta_i} \right) \\ &= (\frac{\gamma}{\gamma_0}) E \frac{1}{\eta} - E \frac{1}{\eta} = (\frac{\gamma}{\gamma_0} - 1) E \frac{1}{\eta}. \end{aligned}$$

Appendix 2
In this appendix we are concerned with the estimation of two hazard models. The first model is the exponential model with a exponential mixing distribution.

$$\theta_i = \eta_i \text{ where } \eta_i \sim \text{Gamma}(\alpha, \beta). \tag{12}$$

This yields the following density function for t given η

$$f(t_i|\eta_i) = \eta_i e^{-\eta_i t_i}.$$

Let $M(\eta_i)$ denote the mixing distribution; the density of a observed realization has the following form.

$$g(t_i) = \int \eta_i e^{-\eta_i t_i} dM(\eta_i) = \int \frac{\beta^\alpha \eta_i^\alpha e^{-\eta_i(\beta + t_i)}}{\Gamma(\alpha)} d\eta_i = \frac{\alpha \beta^\alpha}{(\beta + t)^\alpha}. \tag{13}$$

However, the density of equation (13) could also be generated by the second model.

$$\theta = \frac{\alpha}{(\beta + t)}. \tag{14}$$

It follows from ER that we need a regressor to distinguish between the models of equation (12) and (14). Suppose this regressor has two values and that we normalize $\phi(x_0)$ to be one and denote $\phi(x_1)$ by γ. We first estimate γ.

$$\hat{\gamma} = \frac{\frac{1}{N_0}\sum_{x=x_0} s_i(\varepsilon)}{\frac{1}{N_1}\sum_{x=x_1} s_i(\varepsilon)}.$$

In large samples we have the following for equation (12).

$$\hat{\gamma} \approx \frac{\frac{1}{N_0}\sum_{x=x_0} Es_i}{\frac{1}{N_1}\sum_{x=x_1} Es_i} = \frac{\gamma\varepsilon}{\varepsilon}.$$

Appendix 3

Shows that the expectation of $g_2(\theta_0, c_1(\theta_0))$ is zero

$$\begin{aligned} Eg_2(\theta_0, c_1(\theta_0)) &= \frac{\sum_{x=x_0}(1-e^{-\eta s_{\max}})}{N_0}\frac{\sum_{x=x_1}\frac{1}{\eta}(1-e^{-\eta s_{\max}})}{N_1} \\ &\quad -\frac{\sum_{x=x_1}(1-e^{-\eta s_{\max}})}{N_1}\frac{\sum_{x=x_0}\frac{1}{\eta}(1-e^{-\eta s_{\max}})}{N_0} \\ &= 0. \end{aligned}$$

Appendix 4. Counterfactual duration for the uncensored spells

Consider that x_i is a time invariant regressor, with $\phi(x_i) = e^{x_i\beta} \neq \phi(x_i^*) = e^{x_i^*\beta}$. Let $\lambda(t_i)$ be a piecewise constant function that allows for a different baseline hazard before and after some artificially chosen censoring points $(1, 2, \ldots, T-1)$. The hazard rate for the a non-complier "treated" is defined as

$$\theta\left(t_i^*\right) = \begin{cases} \eta_i e^{x_i^*\beta}, & \text{if } t_i^* \leq 1 \\ \eta_i e^{x_i^*\beta}\lambda_j, & \text{if } j-1 < t_i^* \leq j,\ j = 2, \ldots, T \end{cases}.$$

Define the indicator d_j^*, such that

$$d_1^* = \begin{cases} 1 \text{ if } t_i^* \leq 1 \\ 0 \text{ otherwise} \end{cases}, \quad d_j^* = \begin{cases} 1 \text{ if } j-1 < t_i^* \leq j,\ j = 2, \ldots, T \\ 0 \text{ otherwise} \end{cases}.$$

Then the integrated hazard for the non-complier "treated" is

$$\begin{aligned} z_i^* &= \int_0^{t_i^*} \theta_i\left(t_i^*\right) dt = \eta_i e^{x_i^*\beta}\left(\int_0^{t_i^*} dt\right) d_1^* \\ &\quad + \eta_i e^{x_i^*\beta} \sum_{j=1}^{T-1}\left(1 + \sum_{s=2}^{j}\int_{s-1}^{s}\lambda_{s-1}dt + \int_j^{t_i^*}\lambda_j dt\right) d_{j+1}^*, \end{aligned}$$

with $\sum_{s=2}^{1}\int_{s-1}^{s}\lambda_{s-1}dt = 0$.

To identify the treatment function $(\phi(x_i))$, censoring points for the non-treated observation are needed, such that $s_{i\max}(j, x^*) = s_{i\max}(c_j, x)$, with $j = 1, ..., T$. To find c_j the following equality is used

$$\Pr(x_i = 0, j) = \Pr(x_i = 1, c_j),$$

with data from the treatment $(x_i = 0)$ and control $(x_i = 1)$ groups. Thus, given $c_j < j$, the hazard for the non-treated observation is defined as:

$$\theta(t_i) = \begin{cases} \eta_i e^{x_i\beta}, & \text{if } t_i \leq c_1 \\ \eta_i e^{x_i\beta}\lambda_{j-1}, \text{ if } c_{j-1} < t_i \leq c_j,\ j = 2, \ldots, T \\ \eta_i e^{x_i\beta}\lambda_T, \text{ if } c_T < t_i \leq T \end{cases},$$

and define d_j, such that

$$\begin{aligned} d_1 &= \begin{cases} 1 \text{ if } t_i \leq c_1 \\ 0 \text{ otherwise} \end{cases},\ d_j = \begin{cases} 1 \text{ if } c_{j-1} < t_i \leq c_j,\ j = 2, \ldots, T \\ 0 \text{ otherwise} \end{cases} \quad \text{and} \\ d_{T+1} &= \begin{cases} 1 \text{ if } c_T < t_i \leq T \\ 0 \text{ otherwise} \end{cases}. \end{aligned}$$

Then, the integrated hazard for the non-treated observation is defined as

$$\begin{aligned} z_i &= \int_0^{t_i} \theta_i(t_i)\, dt = \eta_i e^{x_i\beta} \left(\int_0^{t_i} dt \right) d_1 \\ &+ \eta_i e^{x_i\beta} \sum_{j=1}^{T} \left(c_1 + \sum_{s=2}^{j} \int_{c_{s-1}}^{c_s} \lambda_{s-1} dt + \int_{c_j}^{t_i} \lambda_j dt \right) d_{j+1}, \end{aligned}$$

with $\sum_{s=2}^{1} \int_{c_{s-1}}^{c_s} \lambda_{s-1} dt = 0$.

Given the identification 9 , $Es_i = Es_i^*$, the counterfactual duration, t_i^* can be determined by solving the following equation

$$\begin{aligned} & e^{x_i\beta} \left(\left(\int_0^{t_i} dt \right) d_1 + \sum_{j=1}^{T} \left(c_1 + \sum_{s=2}^{j} \int_{c_{s-1}}^{c_s} \lambda_{s-1} dt + \int_{c_j}^{t_i} \lambda_j dt \right) d_{j+1} \right) = \\ & e^{x_i^*\beta} \left(\left(\int_0^{t_i^*} dt \right) d_1^* + \sum_{j=1}^{T-1} \left(1 + \sum_{s=2}^{j} \int_{s-1}^{s} \lambda_{s-1} dt + \int_{j}^{t_i^*} \lambda_j dt \right) d_{j+1}^* \right) \Leftrightarrow \\ & t_i^* d_1^* + \sum_{j=1}^{T-1} \lambda_j (t_i^* - j)\, d_{j+1}^* = \\ & e^{(x_i - x_i^*)\beta} \left(t_i d_1 + \sum_{j=1}^{T} \left(c_1 + \sum_{s=2}^{j} (c_s - c_{s-1}) \lambda_{s-1} + ((t_i - c_j)\lambda_j) \right) d_{1,j+1} \right) \\ & - \sum_{j=1}^{25} \left(1 + \sum_{s=2}^{j} \lambda_{s-1} \right) d_{j+1}^* \Leftrightarrow \\ & t_i^* = \frac{e^{(x_i - x_i^*)\beta} \left(t_i d_1 + \sum_{j=1}^{T} \left(c_1 + \sum_{s=2}^{j} (c_s - c_{s-1}) \lambda_{s-1} + ((t_i - c_j)\lambda_j) \right) d_{1,j+1} \right) + \sum_{j=1}^{T-1} j \lambda_j d_{j+1}^* - \sum_{j=1}^{T-1} \left(1 + \sum_{s=2}^{j} \lambda_{s-1} \right) d_{j+1}^*}{d_1^* + \sum_{j=1}^{T-1} \lambda_j d_{j+1}^*}. \end{aligned}$$

Appendix 5. Counterfactual duration for the censored spells

The hazard for non-treated observation at each censoring point is defined as:

$$\theta(t_i) = \begin{cases} \eta_i e^{x_i\beta}, & \text{if } t_i \leq c_{11} \\ \eta_i e^{x_i\beta}\lambda_{j-1}, \text{ if } c_{1,j-1} < t_i \leq c_{1j},\ j = 2, \ldots, T \\ \eta_i e^{x_i\beta}\lambda_T, \text{ if } c_{1,T} < t_i \leq T \end{cases}$$

and define the indicators

$$\begin{aligned} d_1 &= \begin{cases} 1 \text{ if } t_i \leq c_{11} \\ 0 \text{ otherwise} \end{cases}, \; d_j = \begin{cases} 1 \text{ if } c_{1,j-1} < t_i \leq c_{1j},\; j = 2, \ldots, T \\ 0 \text{ otherwise} \end{cases} \\ d_{T+1} &= \begin{cases} 1 \text{ if } c_{1,T} < t_i \leq T \\ 0 \text{ otherwise} \end{cases}. \end{aligned}$$

Then the integrated hazard for the non-treated observation is defined as

$$\begin{aligned} z_i &= \int_0^{t_i} \theta_i (t_i)\, dt = \eta_i e^{x_i \beta} \left(\int_0^{t_i} dt \right) d_1 \\ &+ \eta_i e^{x_i \beta} \sum_{j=1}^{T} \left(c_{11} + \sum_{s=2}^{j} \int_{c_{1,s-1}}^{c_{1s}} \lambda_{s-1} dt + \int_{c_{1j}}^{t_i} \lambda_j dt \right) d_{j+1}, \end{aligned}$$

with $\sum_{s=2}^{1} \int_{c_{1,s-1}}^{c_{1s}} \lambda_{s-1} dt = 0$. Now, define the hazard for non-complier "treated" observation as

$$\theta (t_i^*) = \begin{cases} \eta_i e^{x_i^* \beta}, \quad \text{if } t_i^* \leq c_{01} \\ \eta_i e^{x_i^* \beta} \lambda_{j-1}, \text{ if } c_{0,j-1} < t_i^* \leq c_{0j},\; j = 2, \ldots, T \\ \eta_i e^{x_i^* \beta} \lambda_T, \text{ if } c_{0,T} < t_i^* \leq T \end{cases}$$

and the indicator

$$\begin{aligned} d_1^* &= \begin{cases} 1 \text{ if } t_i^* \leq c_{01} \\ 0 \text{ otherwise} \end{cases}, \; d_j^* = \begin{cases} 1 \text{ if } c_{0,j-1} < t_i^* \leq c_{0j},\; j = 2, \ldots, T \\ 0 \text{ otherwise} \end{cases} \\ d_{T+1}^* &= \begin{cases} 1 \text{ if } c_{0,T} < t_i^* \leq T \\ 0 \text{ otherwise} \end{cases}. \end{aligned}$$

The integrated hazard for the non-complier "treated" observation becomes

$$\begin{aligned} z_i^* &= \int_0^{t_i^*} \theta_i (t_i)\, dt = \eta_i e^{x_i^* \beta} \left(\int_0^{t_i^*} dt \right) d_1^* \\ &+ \eta_i e^{x_i^* \beta} \sum_{j=1}^{T} \left(c_{01} + \sum_{s=2}^{j} \int_{c_{0,s-1}}^{c_{0s}} \lambda_{s-1} dt + \int_{c_{0j}}^{t_i^*} \lambda_j dt \right) d_{j+1}^*, \end{aligned}$$

with $\sum_{s=2}^{1} \int_{c_{0,s-1}}^{c_{0s}} \lambda_{s-1} dt = 0$.
Given the identification (9), the counterfactual duration, t_i^* can be obtained by solving the following equation

$$\begin{aligned} & e^{x_i \beta} \left(t_i d_1 + \textstyle\sum_{j=1}^{T} \left(c_{11} + \sum_{s=2}^{j} \int_{c_{1,s-1}}^{c_{1s}} \lambda_{s-1} dt + \int_{c_{1j}}^{t_{i1}} \lambda_j dt \right) d_{1,j+1} \right) \\ & = e^{x_i^* \beta} \left(t_i^* d_1^* + \textstyle\sum_{j=1}^{T} \left(c_{01} + \sum_{s=2}^{j} \int_{c_{0,s-1}}^{c_{0s}} \lambda_{s-1} dt + \int_{c_{0j}}^{t_i^*} \lambda_j dt \right) d_{j+1}^* \right) \\ & \Leftrightarrow t_i^* d_1^* + \textstyle\sum_{j=1}^{T} \lambda_j (t_i^* - c_{0j})\, d_{j+1}^* = \\ & e^{\left(x_{i1} - x_i^*\right)\beta} \left(t_i d_1 + \textstyle\sum_{j=1}^{T} \left(c_{11} + \sum_{s=2}^{j} \int_{c_{1,s-1}}^{c_{1s}} \lambda_{s-1} dt + \int_{c_{1j}}^{t_{i1}} \lambda_j dt \right) d_{1,j+1} \right) \\ & - \textstyle\sum_{j=1}^{T} \left(c_{01} + \sum_{s=2}^{j} \int_{c_{0,s-1}}^{c_{0s}} \lambda_{s-1} dt \right) d_{j+1}^* \Leftrightarrow \\ & t_i^* = \frac{e^{\left(x_{i1} - x_i^*\right)\beta} \left(t_i d_1 + \sum_{j=1}^{T} \left(c_{11} + \sum_{s=2}^{j} (c_{1s} - c_{1,s-1}) \lambda_{s-1} + ((t_{i1} - c_{1j}) \lambda_j) \right) d_{1,j+1} \right)}{d_1^* + \sum_{j=1}^{T} \lambda_j d_{j+1}^*} \\ & + \frac{\sum_{j=1}^{T} \lambda_j c_{0j} d_{j+1}^* - \sum_{j=1}^{T} \left(c_{01} + \sum_{s=2}^{j} (c_{0s} - c_{0,s-1}) \lambda_{s-1} \right) d_{j+1}^*}{d_1^* + \sum_{j=1}^{T} \lambda_j d_{j+1}^*}. \end{aligned}$$

8 Bibliography

Bijwaard E.G. and Ridder, G. (2002): “Correcting for Selective Compliance in a Re-employment Bonus Experiment” *USC Center for Law, Economics & Organization*, Research Paper No. C01-15, forthcoming in: *Journal of Econometrics* (2003).

Chamberlain, G. (1985): “Heterogeneity, Omitted Variable Bias, and Duration Dependence,” in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer. Cambridge: Cambridge University Press.

Cochran, W.G. (1965), “The Planning of Observational Studies of Human Populations” (with discussion), Journal of the Royal Statistical Society Series A 128, 234-266.

Cox, David R. (1958), Planning of Experiments, New York: Wiley.

Dolton, P., and D. O’Neill (1996): “Unemployment Duration and the Restart Effect: Some Experimental Evidence,” *The Economic Journal*, 106, 387-400.

Elbers, C. and G. Ridder (1982): “True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model,” *Review of Economic Studies*, 49, 402-409.

Fisher, Ronald A. (1935), The Design of Experiments, Edinburgh: Oliver & Boyd.

Galles, David and Judea Pearl (1998), ”An Axiomatic Characterization of Causal Counterfactuals”, Foundations of Science 3, 151-182.

Glymour, Clark (1986), ”Statistics and Metaphysics”, Comment on ’Statistics and Causal Inference’ by P.W. Holland, Journal of the American Statistical Association 81, 964-966.

Gørgens, T. and J. L. Horowitz (1996): “Semiparametric Estimation of a Censored Regression Model with Unknown Transformation of a dependent variable,” *Journal of Econometrics*, 90, 155-191.

Hărdle, W. (1990): *Applied Nonparametric Regression.*: Cambridge University Press.

Heckman, James J. (2000), ”Causal Parameters and Policy Analysis in Economics: A Tntieth Century Retrospective”, Quarterly Journal of Economics 115, 45-97.

Heckman, J. J. (1991): “Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity,” *American Economic Review*, 81, 75-79.

Heckman, J. J., and G. J. Borjas (1980): “Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers for a Continuous Time Model of Heterogeneity and State Dependence,” *Economica*, 47, 247-283.

Heckman, J. J., and B. Singer (1982): “The Identification Problem in Econometric Models for Duration Data,” in *Advances in Econometrics,* ed. by W. Hildenbrand. New York: Cambridge University Press.

Heckman, J. J., and B. Singer (1984): “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data,” *Econometrica*, 52, 271-320.

Heckman, J. J., and E. J. Vytlacil (2000): “Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect,” NBER Technical Working Paper with number 259.

Honoré, B. E. (1990): “Simple Estimation of a Duration Model with Unobserved

Heterogeneity," *Econometrica*, 58, 453-473.

Honoré, B. E. (1993): "Identification Results for Duration Models with Multiple Spells," *Review of Economic Studies*, 60, 241-246.

Honoré, B. E. (1998): "A Note of the Rate of Convergence of Estimators of Mixtures of Weibulls," Working paper, Princeton University.

Honoré, B. E., and J. L. Powell (1998): "Pairwise Difference Estimators for Non-Linear Models," Working paper, Princeton University.

Horowitz, J. L. (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," Econometrica, 64, 103-107.

Horowitz, J. L. (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity" *Econometrica*, 67, 1001-1028.

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

Kluve, Jochen (2001): "On the Role of Counterfactuals in Inferring Causal Effects of Treatments" IZA Discussion Paper #354 (www.iza.org).

Lancaster, T. (1976): "Redundancy, Unemployment and Manpower Policy: a Comment," *Economic Journal*, 86, 335-338.

Meyer, D. (1995): Lessons from the U.S. Unemployment Insurance Experiments, Economics Literature, Vol.33, No.1, 91-131.

Meyer, D. (1996): "What Have We Learned from the Illinois Reemployment Bonus Experiment," Journal of Labor Economics, Vol.14, No.1, 26-51.

Neyman, Jerzy (1923 [1990]), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.", translated and edited by D.M. Sabrowska and T.P. Speed from the Polish original, which appeared in Roczniki Nauk Rolniczych Tom X (1923), 1-51 (Annals of Agriculture), Statistical Science 5, 465-472.

Ridder, G. and Woutersen, T. (2002), "Method of Moments Estimation of Duration Models with Exogenous and Endogenous Regressors," in preparation.

Ridder, G. and Woutersen, T. (2003), "The Singularity of the Information Matrix of the Mixed Proprtional Hazard Model, Econometrica, Vol. 71, No.5, 1579-1589.

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

Rubin, Donald B. (1977), "Assignment to Treatment Group on the Basis of a Covariate", Journal of Educational Statistics 2, 1-26.

Rubin, Donald B. (1980), "Comment" on 'Randomization Analysis of Experimental Data: The Fisher Randomization Test' by D. Basu, Journal of the American Statistical Association 75, 591-593.

Rubin, Donald B. (1986), "Which Ifs Have Causal Answers", Comment on 'Statistics and Causal Inference' by P.W. Holland, Journal of the American Statistical Association 81, 961-962.

Vytacil, E (2003): "Dummy Endogenous Variables in Nonseparable Models," in revision for Journal of Econometrics.

Woodbury, et al. (1987): "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois." American Economic Review 77, 513-30.

9 Tables and Figures

Table 1: Regression results for log of unobserved heterogeneity on observables (for claimant bonus experiment)

	Non	*Compliance*	*Group*
Variable	*Estimate*	*Standard Error*	*Correlation with* $\ln \widehat{\eta}_i$
Constant	-3.6628	0.0377	–
NEB	1.2145	0.0444	0.7586
$age - \overline{age}$	0.0019	0.0021	-0.0861
$\ln(BPE) - \overline{\ln(BPE)}$	-0.0861	0.0508	-0.0033
black	0.0271	0.0406	-0.0987
male	-0.0079	0.0377	0.0773
$\ln(wb\_d) - \overline{\ln(wb\_d)}$	0.0950	0.0830	0.0033
ONBW	-0.1501	0.0439	0.3121
$R - squared$	0.586		
# of individuals	658		

Table 2: Test :H0 :Average duration in control group is different than average duration in non-compliance group

	Results
H_0 :	0.1336
Standard errors	0.5126
$t - stat$	0.2607

Table 3: Test :H0 :Average(Age) in control group is different than Average(Age) in non-compliance group

	Results
H_0 :	-0.4098
Standard errors	0.5086
$t - stat$	-0.8057

Table 4: Test :H0 :Average(ln(BPE)) in control group is different than Average(ln(BPE)) in non-compliance group

	Results
H_0 :	-0.013
$Standard\ errors$	0.5379
$t - stat$	-0.025

Table 5: Regression results for causal treatment effect. Valid cases: 8134. Dependent variable: duration of unemployment

$Variable$	$Estimate$	$Standard$ $Error$	$t - value$	$Prob > \|t\|$
$Constant$	17.2816	0.1947	88.7369	0.000
$R = 0$ $(control\ group)$	1.3391	0.2794	4.7922	0.000

Table 6: Results for the causal treatment effect (treated versus control group)

	$Control\ Group$ $(R = 0)$	$Treated$ $All(R = 1)$	$Causal\ effect$ $ATE = E(Y_{nt}) - E(Y_t)$
$Benefit\ weks$	18.6207	17.28	1.3391
$Standard\ errors$	0.2028	0.2000	0.2739
$N(individuals)$	3951	4183	–

Table 7: Regression results for the treatment effect on the treated using IV Dependent variable: duration of unemployment

$Variable$	$Estimate$	$Standard$ $Error$	$t - value$	$Prob > \|t\|$
$Constant$	17.0565	0.2120	80.4336	0.000
$\widehat{X} = (R'R)^{-1} R'X$	1.8023	0.3286	5.4845	0.000

Table 8: Counterfactuals durations and standard errors for the non-compliance group obtained using estimated hazards

Weeks of unemployment	counterfactuals for 25 censoring points	Std. errors
1	0.7	0.1858
2	1.75	0.2018
3	2.65	0.2245
4	3.55	0.1816
5	4.45	0.1766
6	5.35	0.2423
7	6.25	0.2047
8	7.15	0.1821
9	8.05	0.1583
10	8.95	0.1749
11	9.85	0.1490
12	10.75	0.1670
13	11.65	0.1773
14	12.55	0.1847
15	13.45	0.1953
16	14.35	0.1808
17	15.25	0.1778
18	16.15	0.2138
19	17.05	0.2182
20	17.95	0.1887
21	18.85	0.1932
22	19.75	0.1687
23	20.65	0.1879
24	21.55	0.1776
25	22.45	0.1952
26	23.35	0.2145

Table 9: Hazard estimates and standard errors for the non-compliance group

weeks of unemployment	hazard rate for non-compliers	Standard errors	hazard rate for non-compliers treated	Standard errors
1	0.0784	0.0010	0.0830	0.0001
2	0.0522	0.0013	0.0666	0.0003
3	0.0275	0.0026	0.0366	0.0010
4	0.0548	0.0034	0.0602	0.0025
5	0.0252	0.0027	0.0302	0.0042
6	0.0436	0.0028	0.0461	0.0066
7	0.0247	0.0038	0.0263	0.0034
8	0.0388	0.0039	0.0506	0.0059
9	0.0209	0.0087	0.0259	0.0121
10	0.0375	0.0086	0.0402	0.0079
11	0.0210	0.0063	0.0331	0.0113
12	0.0301	0.0073	0.0311	0.0157
13	0.0193	0.0113	0.0250	0.0111
14	0.0345	0.0105	0.0385	0.0145
15	0.0183	0.0139	0.0205	0.0125
16	0.0260	0.0118	0.0317	0.0106
17	0.0173	0.0178	0.0201	0.0185
18	0.0281	0.0125	0.0320	0.0165
19	0.0167	0.0175	0.0202	0.0146
20	0.0317	0.0184	0.0319	0.0178
21	0.0205	0.0179	0.0160	0.0225
22	0.0420	0.0243	0.0332	0.0199
23	0.0172	0.0211	0.0211	0.0314
24	0.0377	0.0335	0.0366	0.0144
25	0.0259	0.0198	0.0301	0.0214
26	0.2434	0.0344	0.2166	0.0105

Table 10: Average unemployment durations and standard errors obtained using Bootstrap method (CG=control group)

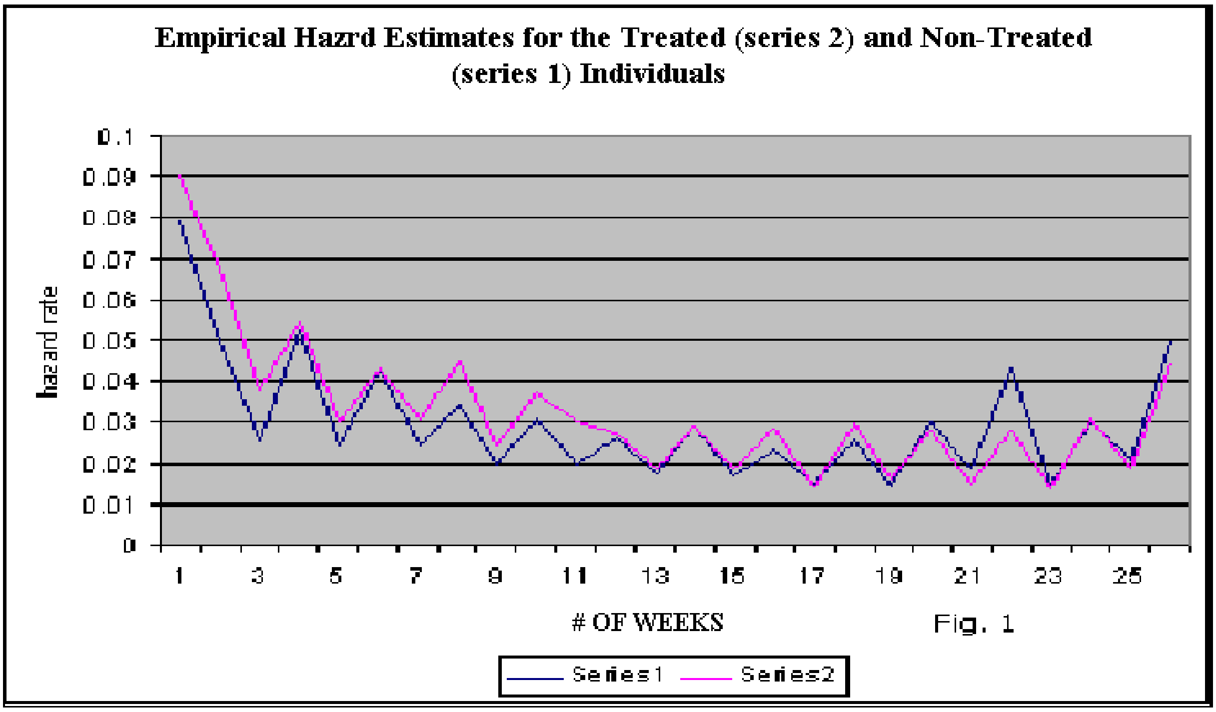
	CG		Treated		Counterfactuals
	(R=1)	All(R=0)	Compliance(X=1)	(X=0)	Noncompliance
Benefit weeks	18.6207	17.28	17.056	18.48	16.62
Std. errors	0.2028	0.2000	0.2103	0.4789	0.3877
N(individ.)	3951	4183	3525	658	658

Table 11: Results for the causal treatment effect for the non-compliers

	Noncompliance(X=0)	Counterfactuals Noncompliance	Causal effect ATE=E(Y_a)-E(Y_p)
Benefit weeks	18.48	16.62	1.86
Std. errors	0.4789	0.3877	0.5831
N(individ.)	658	658	658

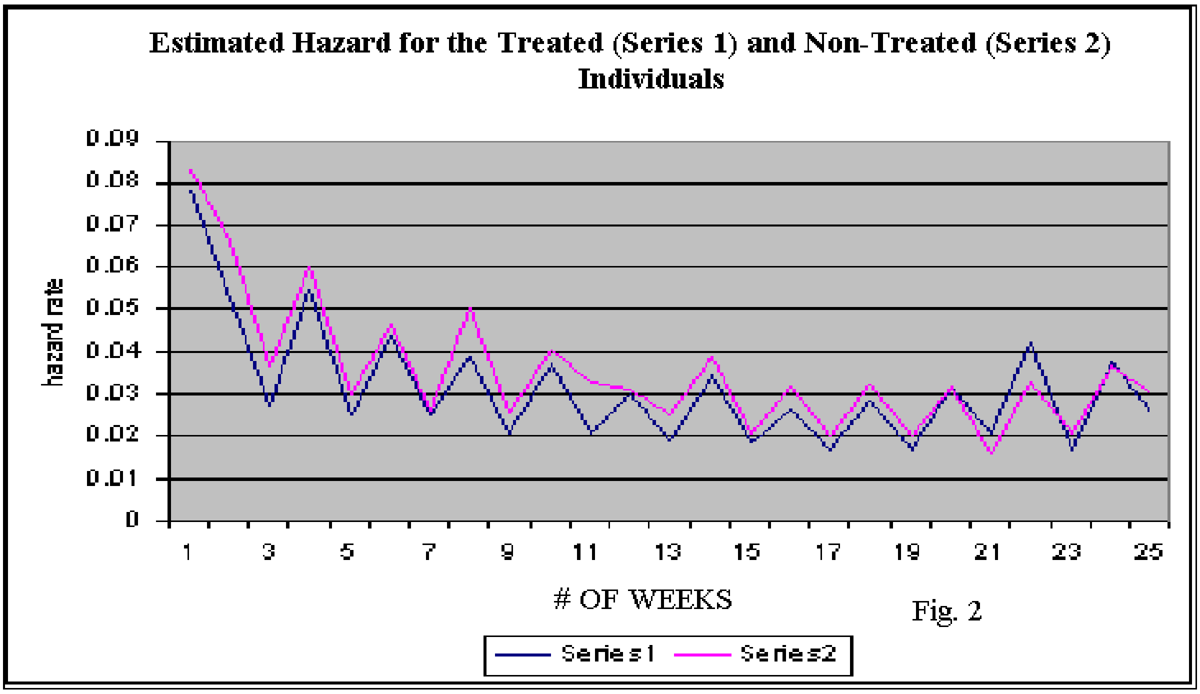
Table 12: Program effect (slope estimates for the treatment effect,beta) for the ML, ITT (Meyer's estimator), 2SLR (Bijwaard and Ridder's estimator) and IH (integrated hazard) estimators in the Illinois Bonus Experiment.

	ML	ITT	$2SLR$	IH
Claimant bonus	0.1601	0.1516	0.2283	0.2025
Standard Errors	0.0361	0.0378	0.0493	0.0287



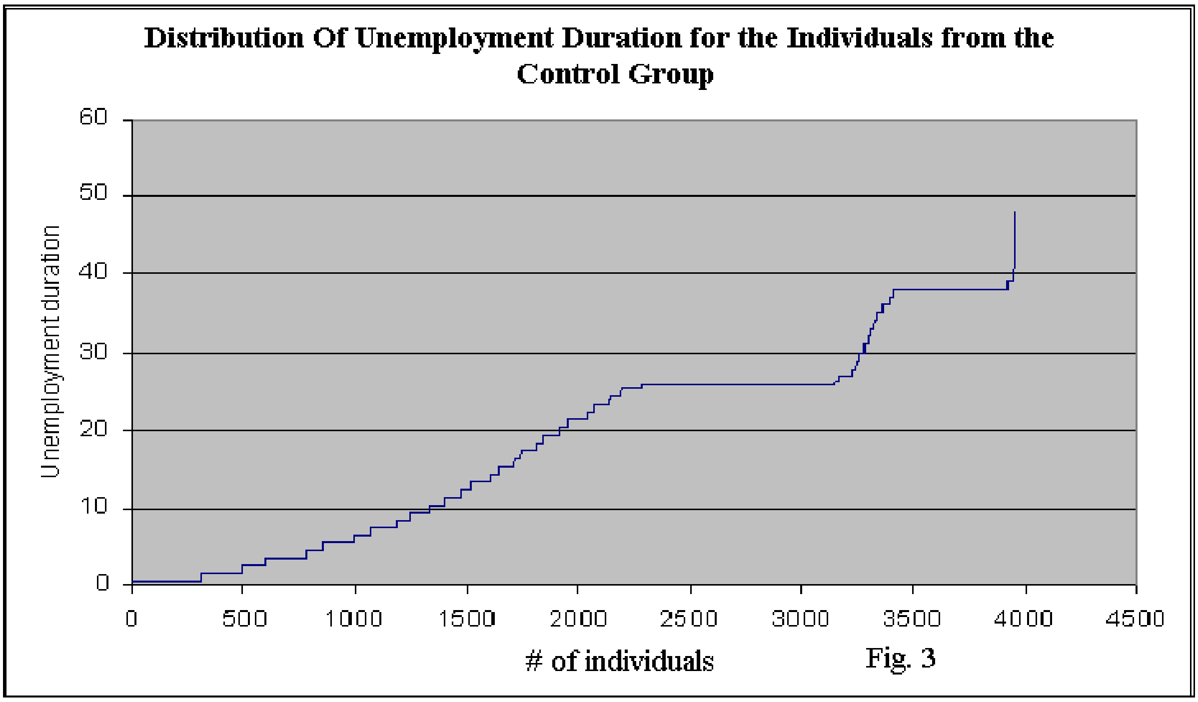
Empirical Hazrd Estimates for the Treated (series 2) and Non-Treated (series 1) Individuals

Fig. 1



Estimated Hazard for the Treated (Series 1) and Non-Treated (Series 2) Individuals

Fig. 2



Distribution Of Unemployment Duration for the Individuals from the Control Group

Fig. 3

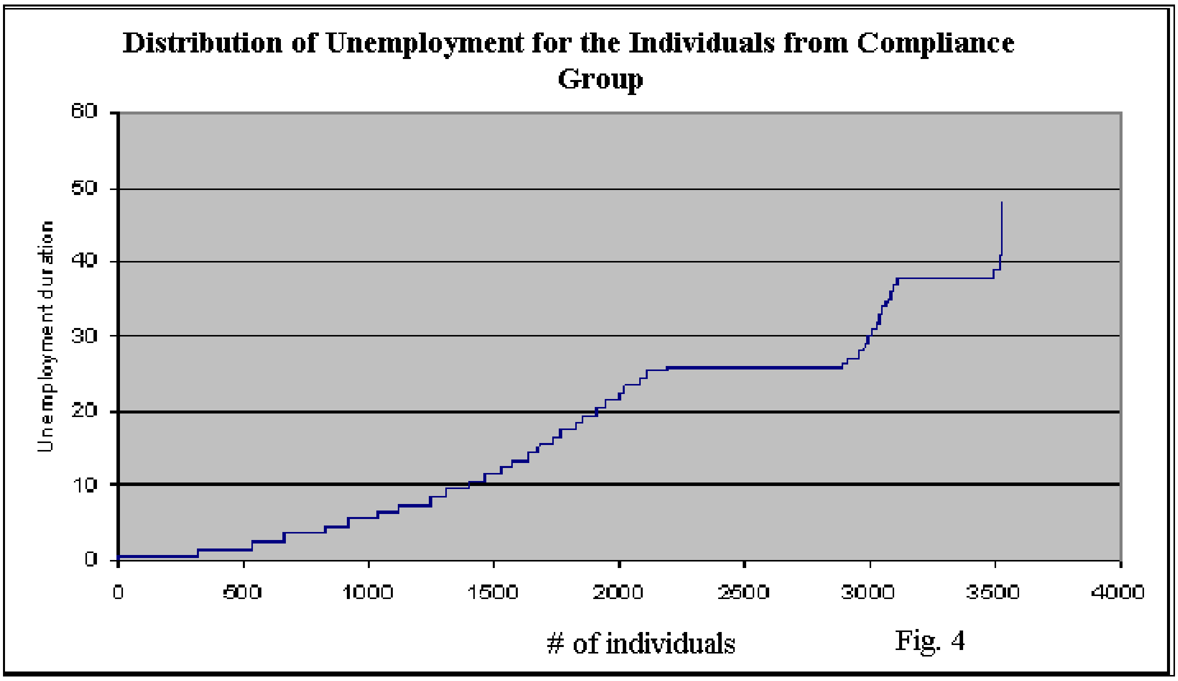
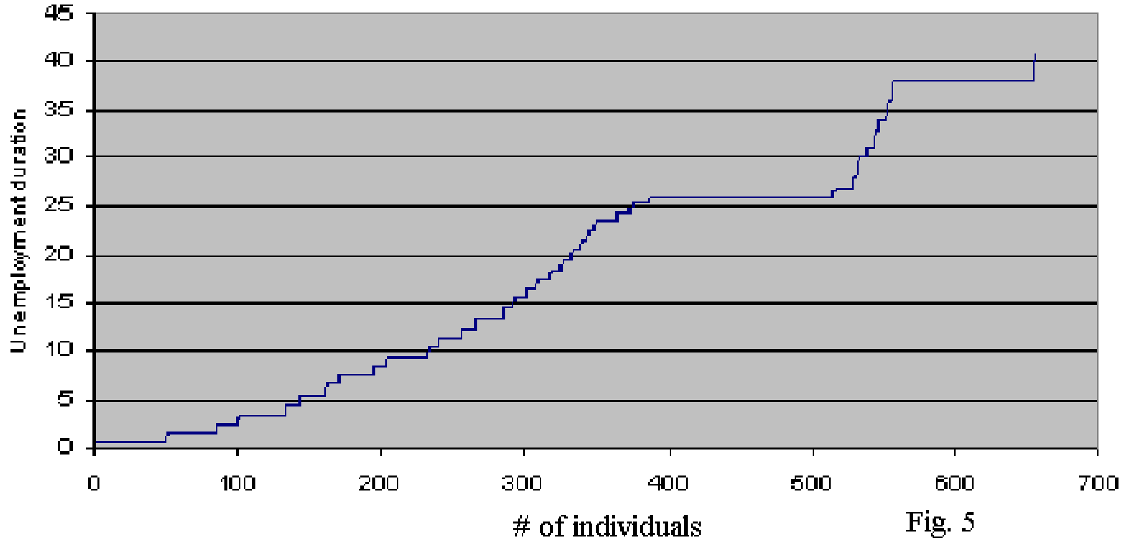

Distribution of Unemployment for the Individuals from Compliance Group
Unemployment duration
60
50
40
30
20
10
0
0
500
1000
1500
2000
2500
3000
3500
4000
# of individuals

Fig. 4



Distribution of Unemployment Duration for the Individuals from the Non-Compliance Group

Fig. 5

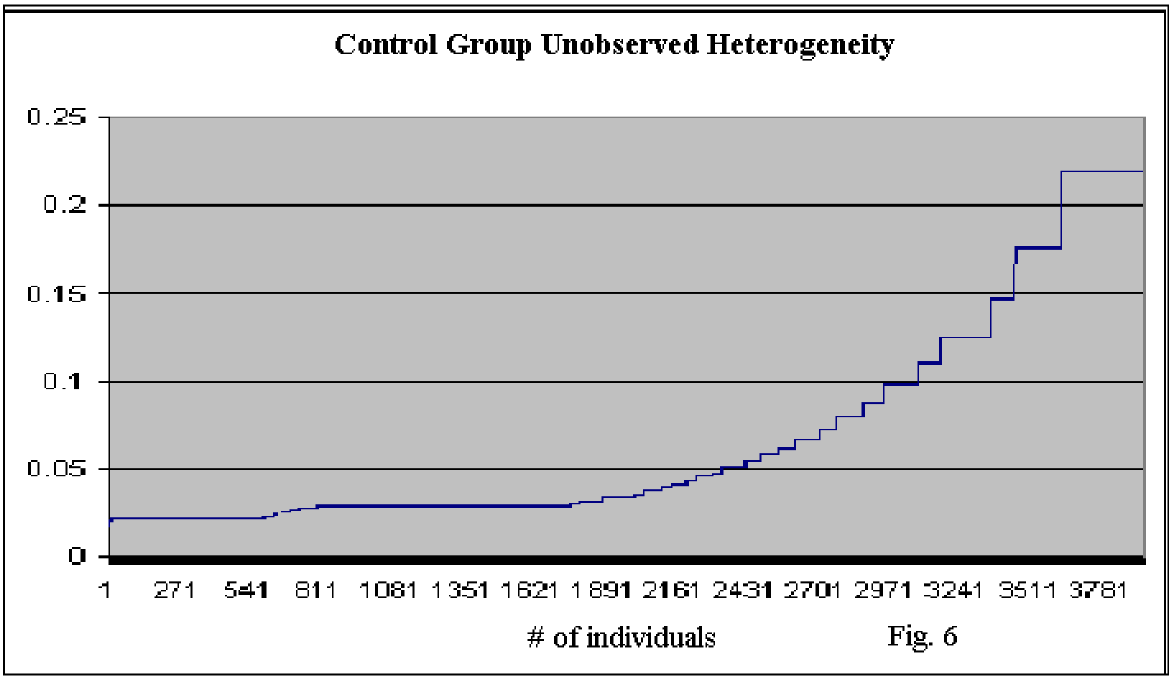

Control Group Unobserved Heterogeneity

Fig. 6

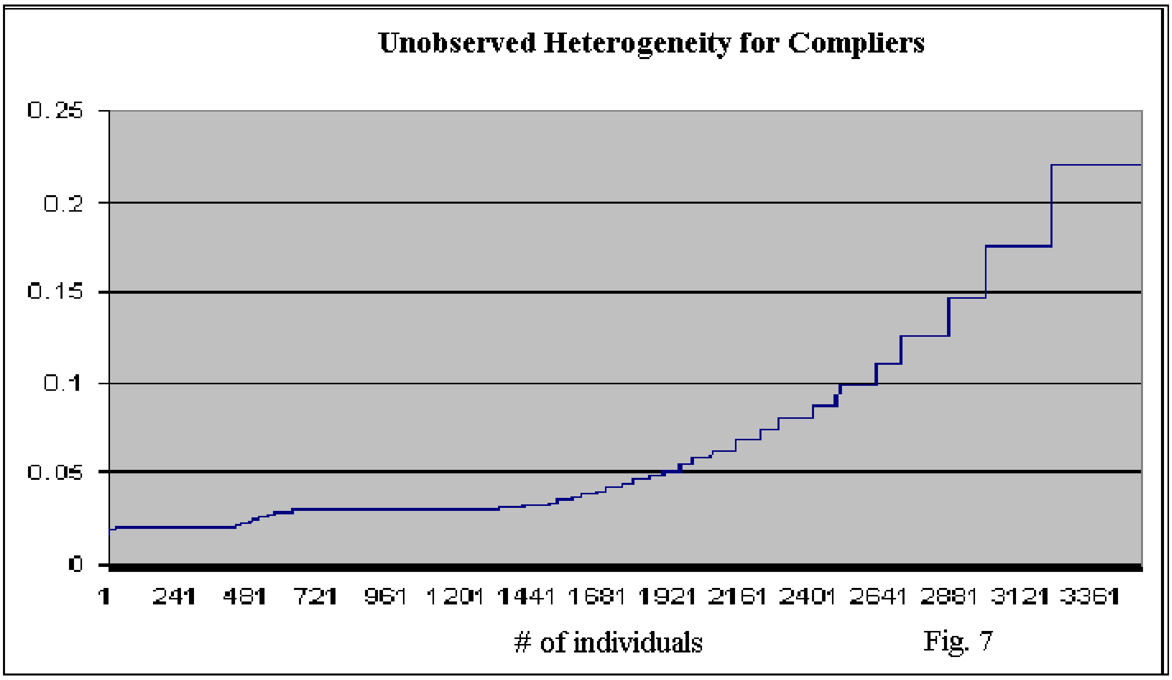

Unobserved Heterogeneity for Compliers
0.25
0.2
0.15
0.1
0.05
0
1
241
481
721
961
1201
1441
1681
1921
2161
2401
2641
2881
3121
3361
# of individuals

Fig. 7

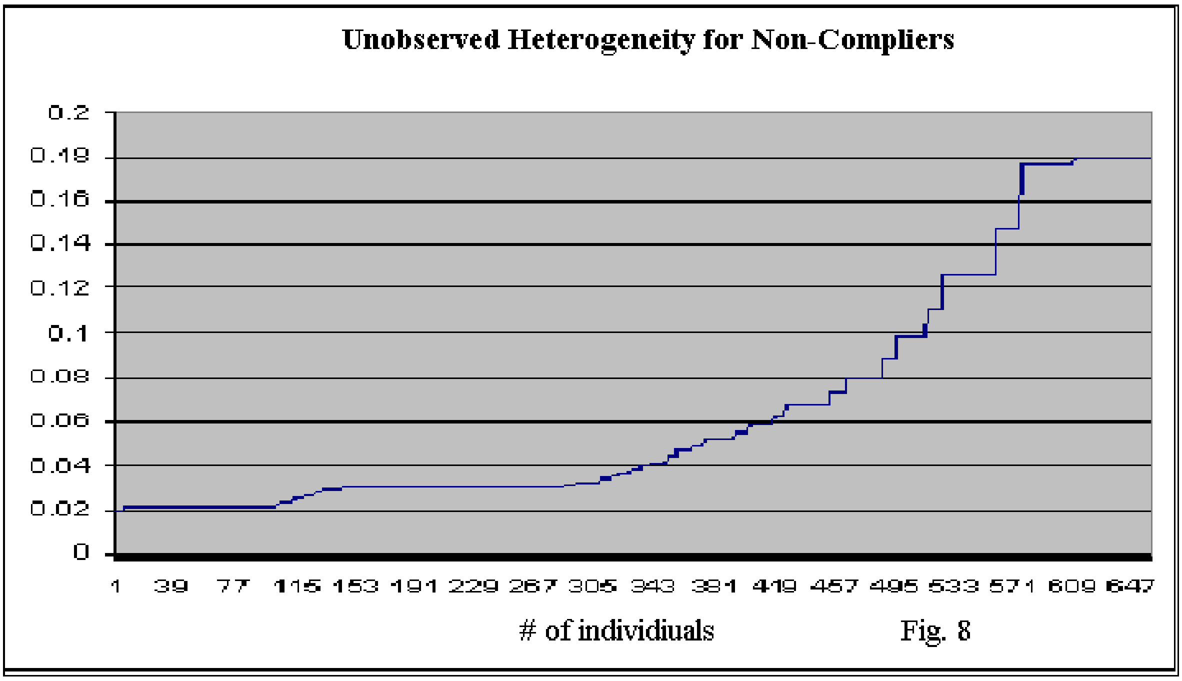

Unobserved Heterogeneity for Non-Compliers
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1
39
77
115
153
191
229
267
305
343
381
419
457
495
533
571
609
647
# of individiuals

Fig. 8

Estimating the noncompliance in employment experiments. An Application to the Illinois Bonus Experiment

Marcel Voia

Carleton University

CESG 22 Meeting

October 21, 2005

This paper:

- This paper introduces a new estimator for duration models that can be applied to experiments with selection problems.

- The new estimator estimates the program effect under the assumption that an employment experiment is transformed into a policy, and thus, the non-compliers in a treatment setting become participants under the new policy.

- It estimates the duration of unemployment for the non-compliers as if they were treated or in other words their counterfactual duration of unemployment.

- The new estimator is used to evaluate a social experiment (the Illinois bonus experiment) designed to test an alternative UI program with the purpose of reducing unemployment duration.

The Illinois bonus experiment

- The Illinois Department of Employment Security conducted the experiments at selected Job Search offices from mid-1984 to mid-1985. Eligible claimants were randomly assigned to the claimant experiment, employer experiment, or control group.

- Examine whether claimant or employer incentives reduce the cost or duration of unemployment.

Claimant experiment

- Out of 8000 individuals, 4186 individuals were randomly selected for treatment, 3527 (84 %), the compliance group, agreed to participate in the experiment, while 659 (16%), the non-compliance group, did not agree.

- Reemployment bonuses of 500 were available to the claimant if they found a job ($\geq$ 30 hours/week) in less than 11 weeks after filing for UI and held that job for at least 4 months.

How the new estimator is constructed?

1. Estimate a mixed proportion hazard (MPH) model (that allows for time-varying treatment effect and unspecified distribution for the unobserved heterogeneity) using the integrated hazard principle of Woutersen (2000), the identification results of Elbers and Ridder (1982) and the moment conditions of Ridder and Woutersen (2002),

2. Use the estimation results of the hazard rate to construct counterfactual durations for non-compliers

3. Simulate the distribution of unemployment durations for the non-treated individuals as they were treated using Bootstrap.

The estimator is used to answer the following policy questions:

1. How much would the average unemployment duration decrease if this program were introduced generally (make the bonus available to all eligible individuals at no cost)?

2. How much money would be saved if this program were introduced generally?

Literature

About duration analysis

- Lancaster (1990) and Van den Berg (2000) for overview.

- Tsiatis (1990) and Robin and Tsiatis (1991)
  - linear rank estimator
  - censor all their observations to deal with partial compliance.

- Woutersen (2000)
  - integrated hazard estimator, can deal with censoring
  - base inference on the whole dataset.

- Ridder and Woutersen (2002)
  - estimator for MPH models.
  - consistent if there is time-varying treatment and if the outcome variable is a censored duration.

About bonus experiment:

- Woodbury and Spiegelman (1987)

  - complete description of the Illinois Experiment and a summary of its results.

- Meyer (1995)

  - what can be learned about labor market and UI policy from four bonus experiments: Illinois, New Jersey, Pennsylvania and Washington.

- Meyer (1996) , Illinois bonus

  - examines predictions about the timing of exits from unemployment

  - tests labor supply and search theories of unemployment.

- Bijward and Ridder (2002), Illinois bonus experiment

  - develop a two-stage instrumental variable estimator

  - show that compliance in the bonus experiment was selective.

Integrated hazard principle

Simple example: Let $T_1, ..., T_N$ be independent durations with hazard $\theta(t) = \lambda$ so $Z = \lambda T$.

- **The integrated hazards are independent unit exponentials:** $Z = \lambda T \sim \varepsilon(1)$.

- **Equating the sample analogue of the integrated hazard to one gives:**

$$\frac{\sum_{i=1}^{N} \lambda t_i}{N} = 1.$$

- **This suggests an estimator for** λ,

$$\hat{\lambda} = (\frac{N}{\sum_{i=1}^{N} t_i}),$$

  which is the maximum likelihood estimator.

- **Equation**

$$Z = \lambda T \sim \varepsilon(1)$$

  can also be used for censored duration data.

Censored observations

- leaving out censored observations ⇒ bias estimate for the treatment effect

How to deal with censored observations?

- let Z' denote the integrated hazard of this potentially censored observation and let the indicator d be zero if the observation is censored and one otherwise.

  Equation ($EZ' = Ed$) suggests to censor the durations at c.

- Expectation of Z' equals the expectation of d, i.e.

  $$EZ' = Ed,$$

  where $d = \{0, 1\}$.

  Equation ($EZ' = Ed$) suggests to censor the durations at c.

- Use

  $$g(\lambda) = \frac{\sum d_i}{N} - \frac{\sum z'_i}{N}$$

  as a moment equation to estimate λ, and use $\hat{\lambda}$ to construct counterfactual durations for noncompliers.

Estimating the hazard when we have censored spells

- **Consider the following hazard model**

  $$\theta = \eta_i \phi(x) \lambda(t), with$$

  x **a time invariant regressor,** $\phi(x_0) \neq \phi(x_1)$, η_i **a realization of a mixing distribution.**

- **Consider individuals with** $x = x_0$ **as the treatment group and those with** $x = x_1$ **as the control group.**

- **Let** $\lambda(t)$ **be a piecewise constant function that allows for a different baseline hazard before and after** c.

- **Censor the treatment and control group at** c_0 **and** c_1

- **Choose the censoring points in such a way that the survival probabilities are equal.**

- **Get a sequence** c_{0j} **and** c_{1j}**,where** $j = 1, 2, \ldots, 25$, **by equating**

  $$\begin{aligned} \Pr(x = 1, j) &= \Pr(x = 0, c_{0j}) \\ \Pr(x = 0, j) &= \Pr(x = 1, c_{1j}). \end{aligned}$$

and using

$$\begin{aligned} 1-\Pr(x &= 0, c_{0j}) = e^{-c_{0j}\phi(x_0)} \\ 1-\Pr(x &= 1, c_{1j}) = e^{-c_{1j}\phi(x_1)} \end{aligned}$$

- **Estimate** $\frac{\phi(x_1)}{\phi(x_0)}$, **as** $\frac{\widehat{\phi(x_1)}}{\phi(x_0)} = \frac{c_0}{c_1}$.

- **Use moments of the form**

  $$g_j = s_{0j}(c_{0j}, \lambda_j) - s_{1j}(c_{1j}, \lambda_j), j = 1, 2, \ldots, 25,$$

  which are linear in parameters λ, **to estimate the hazard rates for the non-compliance group.**

- **Having estimates for the hazard compute the counterfactuals for the non-compliance group.**

- **Compute standard errors using bootstrap method.**

Counterfactual computation.

- **Define the semi-integrated hazard for the non-treated observation as $s_i = e^{x_1\beta}\Lambda\left(t_{i1}\right)$**

- **Consider x_i^*, the counterfactual treatment**
- **To find the counterfactual duration (t_i^*), hypothetically randomize into treatment the non-compliance group and then, use the identification assumption $\left(\frac{\sum_{x=x_0} s_i}{N_0} = \frac{\sum_{x=x_1} s_i}{N_1}\right)$ to solve for t_i^*. Thus, the integrated hazard z_i for a non-complier without treatment is**

  $$z_i = \eta_i e^{x_i\beta}\Lambda\left(t_i\right) \Rightarrow s_i = e^{x_i\beta}\Lambda\left(t_i\right).$$

  Suppose x_i^* is the counterfactual treatment for the non-complier, then the integrated hazard for the non-complier "treated" is

  $$z_i^* = \eta_i e^{x_i^*\beta}\Lambda\left(t_i^*\right) \Rightarrow s_i^* = e^{x_i^*\beta}\Lambda\left(t_i^*\right).$$

  Given the identification $s_i = s_i^*$,

  $$\Rightarrow t_i^* = \Lambda^{-1}\left(\frac{s_i}{e^{x_i^*\beta}}\right) = \Lambda^{-1}\left(e^{(x_i - x_i^*)\beta}\Lambda\left(t_i\right)\right).$$

Conclusion

- **I use a new estimator which can deal with partial compliance and estimates the duration of unemployment of the untreated observations as they were treated.**

- **I found that allowing for partial-compliance and unobserved heterogeneity, a generalization of the program to all eligible individuals will reduce significantly the duration of unemployment (the average unemployment duration decreased by about 1.86 weeks (Std. error = 0.5831) if the non-treated were treated.**

- **The results lead to more optimistic conclusions about monetary incentives than Meyer (1996).**

Tables

Table 1: Regression results for log of unobserved heterogeneity on observables (for claimant bonus experiment)

	Non	*Compliance*	*Group*
Variable	*Estimate*	*Standard Error*	*Correlation with* $\ln \widehat{\eta_i}$
Constant	−3.6628	0.0377	−
NEB	1.2145	0.0444	0.7586
$age - - - age$	0.0019	0.0021	−0.0861
$\ln(BPE) - - - - - - \ln(BPE)$	−0.0861	0.0508	−0.0033
black	0.0271	0.0406	−0.0987
male	−0.0079	0.0377	0.0773
$\ln(wb\_d) - - - - - - - \ln(wb\_d)$	0.0950	0.0830	0.0033
ONBW	−0.1501	0.0439	0.3121
$R - squared$	0.586		
# of individuals	658		

Table 2: Test :H0 :Average duration in control group is different than average duration in non-compliance group

	Results
H_0 :	0.1336
Standard errors	0.5126
$t - stat$	0.2607

Table 3: Test :H0 :Average(Age) in control group is different than Average(Age) in non-compliance group

	Results
H_0 :	−0.4098
Standard errors	0.5086
$t - stat$	−0.8057

Table 4: Test :H0 :Average(ln(BPE)) in control group is different than Average(ln(BPE)) in non-compliance group

	Results
H_0 :	-0.013
Standard errors	0.5379
$t - stat$	-0.025

Table 5: Regression results for causal treatment effect. Valid cases: 8134. Dependent variable: duration of unemployment

Variable	*Estimate*	*Standard Error*	$t - value$	$Prob > \|t\|$
Constant	17.2816	0.1947	88.7369	0.000
$R = 0$ *(controlgroup)*	1.3391	0.2794	4.7922	0.000

Table 6: Results for the causal treatment effect (treated versus control group)

	Control Group $(R = 0)$	*Treated* $All(R = 1)$	*Causal effect* $ATE = E(Y_{nt}) - E(Y_t)$
$\overline{Benefitweks}$	18.6207	17.28	1.3391
Standard errors	0.2028	0.2000	0.2739
$N(individuals)$	3951	4183	$-$

Table 7: Regression results for the treatment effect on the treated using IV Dependent variable: duration of unemployment

Variable	*Estimate*	*Standard Error*	$t - value$	$Prob > \|t\|$
Constant	17.0565	0.2120	80.4336	0.000
$\widehat{X} = (R'R)^{-1} R'X$	1.8023	0.3286	5.4845	0.000

Table 8: Counterfactuals durations and standard errors for the non-compliance group obtained using estimated hazards

Weeks of unemployment	counterfactuals for 25 censoring points	Std. errors
1	0.7	0.1858
2	1.75	0.2018
3	2.65	0.2245
4	3.55	0.1816
5	4.45	0.1766
6	5.35	0.2423
7	6.25	0.2047
8	7.15	0.1821
9	8.05	0.1583
10	8.95	0.1749
11	9.85	0.1490
12	10.75	0.1670
13	11.65	0.1773
14	12.55	0.1847
15	13.45	0.1953
16	14.35	0.1808
17	15.25	0.1778
18	16.15	0.2138
19	17.05	0.2182
20	17.95	0.1887
21	18.85	0.1932
22	19.75	0.1687
23	20.65	0.1879
24	21.55	0.1776
25	22.45	0.1952
26	23.35	0.2145

Table 9: Hazard estimates and standard errors for the non-compliance group

weeks of unemployment	hazard rate for non-compliers	Standard errors	hazard rate for non-compliers treated	Standard errors
1	0.0784	0.0010	0.0830	0.0001
2	0.0522	0.0013	0.0666	0.0003
3	0.0275	0.0026	0.0366	0.0010
4	0.0548	0.0034	0.0602	0.0025
5	0.0252	0.0027	0.0302	0.0042
6	0.0436	0.0028	0.0461	0.0066
7	0.0247	0.0038	0.0263	0.0034
8	0.0388	0.0039	0.0506	0.0059
9	0.0209	0.0087	0.0259	0.0121
10	0.0375	0.0086	0.0402	0.0079
11	0.0210	0.0063	0.0331	0.0113
12	0.0301	0.0073	0.0311	0.0157
13	0.0193	0.0113	0.0250	0.0111
14	0.0345	0.0105	0.0385	0.0145
15	0.0183	0.0139	0.0205	0.0125
16	0.0260	0.0118	0.0317	0.0106
17	0.0173	0.0178	0.0201	0.0185
18	0.0281	0.0125	0.0320	0.0165
19	0.0167	0.0175	0.0202	0.0146
20	0.0317	0.0184	0.0319	0.0178
21	0.0205	0.0179	0.0160	0.0225
22	0.0420	0.0243	0.0332	0.0199
23	0.0172	0.0211	0.0211	0.0314
24	0.0377	0.0335	0.0366	0.0144
25	0.0259	0.0198	0.0301	0.0214
26	0.2434	0.0344	0.2166	0.0105

Table 10: Average unemployment durations and standard errors obtained using Bootstrap method (CG=control group)

	CG	Treated			Counterfactuals
	(R=1)	All(R=0)	Compliance(X=1)	(X=0)	Noncompliance
$\overline{Benefitweeks}$	18.6207	17.28	17.056	18.48	16.62
Std. errors	0.2028	0.2000	0.2103	0.4789	0.3877
N(individ.)	3951	4183	3525	658	658

Table 11: Results for the causal treatment effect for the non-compliers

	Noncompliance(X=0)	Counterfactuals Noncompliance	Causal effect ATE=E(Y_a)-E(Y_p)
$\overline{Benefitweeks}$	18.48	16.62	1.86
Std. errors	0.4789	0.3877	0.5831
N(individ.)	658	658	658

Table 12: Program effect (slope estimates for the treatment effect,beta) for the ML, ITT (Meyer's estimator), 2SLR (Bijwaard and Ridder's estimator) and IH (integrated hazard) estimators in the Illinois Bonus Experiment.

	ML	ITT	$2SLR$	IH
Claimant bonus	0.1601	0.1516	0.2283	0.2025
Standard Errors	0.0361	0.0378	0.0493	0.0287