Instrumental Variable Estimation for Duration Data

Govert E. Bijwaard*

Econometric Institute

Erasmus University Rotterdam

Econometric Institute Report EI 2007-14

forthcoming chapter in *Causal Analysis in Population Studies: Concepts,Methods, Applications* eds. Engelhardt, Kohler and Prskawetz, Springer-Verlag

*Erasmus University Econometric Institute, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; Phone: (+31) 10 40 81424; Fax: (+31) 10 40 89162; E-mail: `bijwaard@few.eur.nl`. This research is financially supported by the Netherlands Organization for Scientific Research (NWO) nr. 451-04-011. I am very grateful to Geert Ridder for his comments and suggestions. I also benefited from the comments of participants at the 2003 Extending the Tinbergen Heritage conference in Rotterdam, at the 2003 ESEM Meeting, at the 14^{th} EC^2 conference , at the Econometric Evaluation of Public Policies : Methods and Applications in Paris 2005 and at the Causal Analysis in Population Studies conference in Vienna 2006 for comments and feedback. I thank Steve Woodbury and the W. E. Upjohn Institute for making the data available to me.

Instrumental Variable Estimation for Duration Data

Abstract

In this article we focus on duration data with an endogenous variable for which an instrument is available. In duration analysis the covariates and/or the effect of the covariates may vary over time. Another complication of duration data is that they are usually heavy censored. The hazard rate is invariant to censoring. Therefore, a natural choice is to model the hazard rate instead of the mean.

We develop an Instrumental Variable estimation procedure for the Generalized Accelerated Failure Time (GAFT) model. The GAFT model is a duration data model that encompasses two competing approaches to such data; the (Mixed) Proportional Hazard (MPH) model and the Accelerated Failure Time (AFT) model. We discuss the large sample properties of this Instrumental Variable Linear Rank (IVLR) estimation based on counting process theory. We show that choosing the right weight function in the IVLR can improve its efficiency. We discuss the implementation of the estimator and apply it to the Illinois re-employment bonus experiment.

JEL classification: C21, C41, J64.
Key words: Endogenous Variable; Duration model; Censoring; Instrumental Variable.

1 Introduction

Social scientists have a long tradition of exploring the substantive implications of endogeneity in both methodological work and empirical work. Endogeneity is troublesome because it precludes the usual causal kinds of statements social scientists like to make. A canonical example is the evaluation of the effect of training programs of unemployment individuals on earnings and employment status. In general, the indicator for those who were trained is endogenous, because those individuals who choose to get training perceive the training as beneficial for earning or employment status. Other examples include the effect of union status and childbearing on labor market outcomes. All these problems have a treatment-control flavor. The notion that treatment status is endogenous reflects the fact that simple comparisons of treated and untreated individuals are unlikely to have a causal interpretation.

In recent years, social experiments have gained popularity as a method for evaluating social and labor market programs (see e.g. Meyer (1995), Heckman et al. (1999) and Angrist and Krueger (1999)). In experiments the assignment of individuals to the treatment can be manipulated. If assignment is random, the average impact of the treatment can be estimated. However, a randomized assignment may be compromised, if the individuals can refuse to participate, either by dropping out, if they are to receive the treatment, or by obtaining the treatment, if they are in the control group. If this non–compliance to the assigned treatment is correlated with the outcomes in the treatment or control regimes, the observed effect of the treatment is a biased estimate of the treatment effect. Thus, even with random assignment the actual treatment status can be endogenous.

Most of the evaluation literature has focused on static treatments, *i.e.* treatment that is administered at a particular point in time or in a particular time interval. If the outcome is a duration the treatment or its effect can be dynamic, *i.e.* it can be switched on and off over time. Examples are the unemployment insurance experiments (see Meyer (1995) for a survey) in which the unemployed receive a cash bonus if they find a job in a specified period. Another example is a temporary cut in unemployment benefits of unemployed individuals who do not expend sufficient effort to find a job (e.g. see Van den Berg et al. (2004) for The Netherlands and Ashenfelter et al. (2005) for the U.S.).

The problem of endogeneity in duration models is similar to other statistical models: when endogeneity is present the standard interpretations given by any statistical model generally do not hold. If the training is perceived beneficial those individuals who choose to get training differ ex ante from those who choose not to get training. Similarly, unemployed who choose to be eligible for a cash bonus if they find a job in time, differ both in observed and unobserved characteristics that may influence their job finding probability.

For linear models the problem of endogeneity can be solved if an instrument is available. The only requirement is that such an instrument affects the endogenous variable but is not correlated with the errors of the regression. We extend that notion to duration models that are inherently non-linear and propose an estimation technique. The basic idea behind this Instrumental Variable Linear Rank (IVLR) estimator is that on the correct transformed duration time the proportion of people with a particular value of the instrument remains the same.

Two competing approaches to the estimation of the effect of a, possible time–varying, covariate on duration has been the (Mixed) Proportional Hazard (MPH) model (for a recent survey see Van den Berg (2001)) and the Accelerated Failure Time (AFT) model (see a.o. Kalbfleisch and Prentice (2002), Brännäs (1992), and Klein and Moeschberger (1997)). In the Mixed Proportional Hazards (MPH) model the hazard is written as the product of the baseline hazard, a non-negative regression function, and a non-negative random variable that represents the covariates that are omitted from the regression function. Ridder (1990) introduced the Generalized Accelerated Failure Time (GAFT) model, a generalization of the AFT models that also includes the MPH model.

The GAFT model is based on transforming the duration and assuming some distribution for this transformed duration. The transformation is related to the integrated hazard of a PH model. The AFT model is obtained by restricting the transformation. The AFT does not restrict the distribution of the transformed duration, while the MPH model restricts this distribution to a mixture of exponentials. The regression coefficients in a GAFT model can be interpreted in terms of the effect of regressing on the quantiles of the distribution of the transformed duration for the reference individual. In an AFT model the relation between the q-th quantile of a individual with observed characteristics X and the q-th quantile of the reference individual is the acceleration factor. In a GAFT this acceleration factor is multiplied by the ratio of the 'duration dependence' at the two quantile durations.

For the population parameters of the GAFT model the distribution of the transformed durations is independent of the instrument. Thus, for a binary instrument, R, the proportion of the people with $R = 1$ remains the same over the survivors on the transformed duration time. Then the rank test statistic for the influence of this instrument on transformed duration should go to zero. The inverse of the rank test provides the estimation procedure we call the IVLR. The asymptotic properties of the estimator are derived using counting process theory.

The existence of endogenous covariates implies (possible) dependence between the transformed duration and the censoring time. This implies that the IVLR estimator, which exploits the independence between the transformed durations and the instruments, may give biased results. We can often make the assumption that the (potential) censoring time is known at the start of the study. In the re-employment bonus data, for example, we can only observed the unemployed while receiving UI benefits. In this case the potential censoring time is for all individuals at 26 weeks, the maximum duration of UI benefits in Illinois at the time of the experiment. With known (potential) censoring time we can modify the GAFT transformation by introducing additional censoring such that this modified transformation and the instruments become independent for the uncensored observations. Then, the IVLR estimator on this modified transformation leads to consistent estimators.

The IVLR estimation is based on a vector of mean restrictions on weight functions of the covariates, instrument and the transformed durations. Thus the IVLR is also related to GMM estimation. In GMM estimation it is feasible to get the most efficient GMM estimator in just two steps. In the first step directly observed weighting matrices lead to a consistent, but not necessary efficient estimator. From this consistent estimator we can consistently estimate the efficient weighting matrices. It is then possible to obtain an efficient estimate of the parameters involved in just two steps. A similar reasoning applies to the IVLR-estimator. In the first step we use simple weighting functions to obtain consistent estimates of the parameters of the GAFT model. From these parameters we can estimate the distribution of the transformed durations, which are needed to calculate the most efficient weighting functions. Then, in just one additional step the efficient IVLR is obtained.

For our empirical application we use data from the Illinois unemployment bonus experiment. These data have been analysed before with increasing sophistication by Woodbury and Spiegelman (1987), Meyer (1996) and Bijwaard and Ridder (2005). In this experiment a group of individuals who became unemployed during four months in 1984 were divided at random in three groups of about equal size: two bonus groups and a control group. The unemployed in the claimant bonus group qualified for a cash bonus if they found a job within 11 weeks and would hold this job for at least four months. In the employer bonus group, the bonus was paid to their employer. The members of the two bonus groups were asked whether they were prepared to participate in the experiment. About 15% of the claimant bonus and 35% of the employer bonus groups refused participation. It is very likely that the decision to be eligible for a bonus is related to the unemployment duration. This makes the participation indicator an endogenous variable in relation to the unemployment duration. Comparing our estimates with a standard ML-estimate shows that ignoring the endogeneity leads to underestimate the effect of the bonus on the re-employment probability. Assuming an AFT instead of a more general GAFT model overestimate the effect of the bonus.

The outline of the article is as follows. Section 2 introduces the GAFT model and its relation to the AFT and MPH models. We also discuss identification and interpretation of the GAFT model. Section 3 provides a discussion on endogenous variables in duration models and how such variables can be incorporated into a GAFT model. In Section 4 we explain the counting process framework for duration models and the GAFT model in particular. We also discuss the problems of endogenous censoring. Within the counting process framework, the asymptotic properties of our estimator, which is introduced in Section 5, can be derived using martingale theory. We prove the consistency and asymptotic normality of the estimator. We discuss the efficiency and the practical implementation of the IVLR. Section 6 discusses the empirical application of the IVLR estimator to the re-employment bonus experiment. We conclude with a summary and discuss possible avenues for further research in Section 7.

2 Duration Models

For many economic and demographic phenomena the timing of a transition from one state into another state is important. Examples include the time till re-employment of an unemployed individual, the time till marriage and the time till death. Two important features of such transition data are that some relevant characteristics of the individual may change over time, time-varying variables, and that, due to a limited observation window, we do not observe the completed duration for all individuals, right-censoring. In a duration model the timing of a particular event is modeled and it is straightforward to incorporate time-varying variables and allow for right-censoring.

The key variables in duration analysis are the duration till the next event, T, and the indicator of censoring, δ. The observed durations may be right–censored, i.e. we observe $\tilde{T} = \min(T, C)$ with C the censoring time. The possible time–varying covariates are given by the vector $X_i(t)$ where i refers to a member of the population. The path of the covariates are predetermined. Thus $\overline{X}(t) = \{X(s); 0 \leq s \leq t\}$ does not depend on future events.

Two competing approaches for the analysis of duration data has been the (Mixed) Proportional Hazard (MPH) model and the Accelerated Failure Time (AFT) model. The Mixed Proportional Hazards (MPH) model assumes that the hazard, the instantaneous probability of an event at duration t, given that no event occurred before t can be written as

$$\lambda(t|\overline{X}(t), V) = v\lambda_0(s; \alpha)e^{\beta' X(t)}$$

where $\lambda_0(t)$ represents the baseline hazard, that is, the duration dependence of the hazard common to all individuals. The covariates affect the intensity proportionally, $e^{\beta' X(s)}$ and the unobserved heterogeneity V is a random variable that captures variables not in X. For the population parameters, α_0 and β_0, the integrated hazard, $\Lambda(t) = \int_0^t \lambda_0(s;\alpha_0) e^{\beta_0' X(s)}\, ds$ times the unobserved heterogeneity is unit exponentially distributed, $\mathcal{E}(1)$ and therefore the MPH model is also equivalent to

$$\int_0^t \lambda_0(s;\alpha_0) e^{\beta_0' X(s)}\,\mathrm{d}s \sim \frac{\mathcal{E}(1)}{V} \tag{1}$$

Thus the distribution of integrated hazard is a mixture of exponential distributions.

The AFT model assumes that the survival function of an individual with covariate path $\overline{X}(t)$ is related to the survival function of the reference individual, the individual with $\overline{X}(t) = 0$, by

$$S\big(t|\overline{X}(t)\big) = S_0\Big(\int_0^t e^{\beta' X(s)}\,\mathrm{d}s\Big)$$

with $S_0(\cdot)$ is the survival function of the reference individual. If all covariantes are time-invariant then an AFT model implies that the distribution of T_X, the duration of an individual with covariate vector X, and the distribution of $e^{-\beta' X} T_0$ are the same. Thus X accelerates, $\beta < 0$, or decelerates, $\beta > 0$ the duration. This is equivalent with a linear regression model for the log-duration

$$\ln(T) = -\beta' X + \epsilon$$

where ϵ is a random variable with same distribution as $\ln(T_0)$.

2.1 The Generalized Accelerated Failure Time Model

A class of duration models that generalizes the AFT models in such a way that it also includes the MPH models is the Generalized Accelerated Failure Time (GAFT) model, introduced by Ridder (1990). The GAFT model is not specified by the distribution of the log-duration. Instead, we transform the duration, and assume that this transformed duration has some distribution. The transformation of the duration is related to the integrated hazard in a PH-model. The GAFT model is also related to the generalized regression model proposed by Han (1987).

The GAFT model assumes that the relation between the duration T and the covariates is specified as

$$\int_0^T \lambda(s;\alpha)e^{\beta' X(s)}\,\mathrm{d}s = U \tag{2}$$

where $\lambda(t;\alpha)$ is a non–negative function on $[0,\infty)$. The non–negative regression function $e^{\beta' X(s)}$ captures the effect of the covariates. A flexible functional form for λ is the piecewise constant function. The GAFT model is characterized by these functions and by the distribution of the non–negative random variable U. We denote the survivor function of U_0, the transformation in the population parameters α_0 and β_0 by $\overline{G}_0(u)$ and its hazard function by $\kappa_0(u)$, which do not depend on X. We assume that the distribution of U_0 is absolutely continuous. The semi–parametric estimators considered in this article avoid assumptions on the distribution of U_0.

As mentioned, the GAFT model contains as special cases the AFT, the PH and the MPH models. The AFT model restricts the transformation to $\lambda(t;\alpha)\equiv 1$, but leaves the distribution of U_0 unrestricted (with the exception of that U_0 should be non–negative, see e.g. Cox and Oakes (1984)). The (M)PH model restricts the distribution of U_0, but leaves the λ unrestricted (non–negative). The distribution of U_0 is an unit exponential distribution (PH) or a mixture of exponential distributions (MPH). A convenient assumption is that the unobserved heterogeneity has a gamma distribution with variance σ^2. Then, U_0 has a Burr distribution with density $g_0(u) = (1+u\sigma^2)^{-(1+1/\sigma^2)}$, survival function $\overline{G}_0(u) = (1+u\sigma^2)^{-1/\sigma^2}$, mean $1/(1-\sigma^2)$ and hazard function $(1+u\sigma^2)^{-1}$.

We denote the left–hand side of (2) by $h(T,\overline{X}(T);\theta)$ with $\theta = (\alpha',\beta')'$ the vector of parameters. The survivor function of T at t in the GAFT model for the population parameters is

$$\overline{F}(t \mid \overline{X}(t);\theta_0) = \overline{G}_0\Big(h(t,\overline{X}(t);\theta_0)\Big) \tag{3}$$

and the hazard of T at t is

$$\lambda_0(t;\alpha_0)e^{\beta_0' X(t)}\kappa_0\big(h(t,\overline{X}(t);\theta_0)\big). \tag{4}$$

We can interpret the model in terms of the effect of regressing on baseline quantiles, the quantiles for the reference individual. Let $t_q(\overline{X})$ be the q–th quantile of the population distribution of T with covariate history $\overline{X}$. Let t_q be the q–th quantile for the reference individual (with $X(t)$ identically equal to zero). Then from (3)

$$\overline{G}_0\Big(h\Big(t_q(\overline{X}),\overline{X}\big(t_q(\overline{X})\big);\theta_0\Big)\Big) = 1-q = \overline{G}_0\Big(h\big(t_q,0;\theta_0\big)\Big) \tag{5}$$

Hence we obtain a relation between $t_q(\overline{X})$ and t_q that is defined implicitly by

$$\int_0^{t_q(\overline{X})} \lambda(s;\alpha_0)e^{\beta_0' X(s)}\,\mathrm{d}s = \int_0^{t_q} \lambda(s;\alpha_0)\,ds \tag{6}$$

This implies that

$$\frac{d\,t_q(\overline{X})}{d\,t_q} = e^{-\beta_0' X\left(t_q(\overline{X})\right)}\frac{\lambda\big(t_q;\alpha_0\big)}{\lambda\big(t_q(\overline{X});\alpha_0\big)} \tag{7}$$

In the AFT model the duration of an individual with (time–constant) covariate X is distributed as $U_0 e^{-\beta_0' X}$. Thus, in the relation between the quantiles in the GAFT model the acceleration factor is multiplied by the ratio of the values of $\lambda(t)$ in the q–th quantile of the reference and the q–th quantile of the individual with covariate X.

In the MPH model we can interpret $\lambda(t)$ as the baseline hazard, i.e. the factor in the proportional hazard that captures the (duration) time variation in the hazard function. Thus, in the MPH model the ratio in (7) can be interpreted as the ratio of baseline hazards. The regression parameter, β, is the proportional change in the hazard rate due to a unit change in $X(t)$ for a unit with unobserved heterogeneity V.

2.2 Identification of the GAFT model

Assume that the regression function in the GAFT model is log–linear. Then, the model is characterized by the non–negative function $\lambda(t;\alpha)$ defined on $[0,\infty)$, the distribution of U_0 and the regression parameter β. Ridder (1990) has shown that if the covariates are time constant, all observationally equivalent GAFT models, i.e. models that give the same conditional distribution of T given X, have regression parameters $d\beta$, integrated transformation $c_1\Big(\int_0^t \lambda_0(s;\alpha_0)\,ds\Big)^{c_2}$ and U_0 distribution $G_0\Big(\big(\frac{u}{c_1}\big)^{1/c_2}\Big)$ for some constants $c_1, c_2 > 0$. The equivalent class follows from the fact that a GAFT model with time constant covariates can be expressed as a transformation model

$$\ln\Big(\int_0^T \lambda_0(s;\alpha_0)\,\mathrm{d}s\Big) \stackrel{d}{=} -\beta_0' X + \ln U,$$

and the constants c_1, c_2 correspond to addition of e^{c_1} to and division by c_2 of the left– and right–hand sides.

With time–varying covariates, the set of observationally equivalent GAFT models is generally smaller. In particular, the power transformation that gives an observationally equivalent model if the covariates are time constant, in general does not result in a GAFT model. As an example consider the GAFT model with time–varying regressors that differ between two groups. In group I

$$X(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq 1, \\ 0 & \text{if } t > 0. \end{cases}$$

and in group II, $X(t) = 0;\ t \geq 0$. Moreover $\lambda_0(t;\alpha) = \alpha t^{\alpha-1}$. With time constant regressors the parameter α is not identified. It can be shown that the observationally equivalent GAFT models have transformation $c_1 t^{\alpha}$ and U–distribution with survival $G_u\left(\frac{u}{c_1}\right)$. Hence, with time–varying covariates α is identified (and so is β).

We conclude that identification depends on whether the covariates are time constant or time–varying. If the covariates are time constant we can identify the transformation $h(T, \overline{X}(T); \theta_0)$ up to a power and β up to scale (with the power and the scale being equal). Moreover, if we fix the power we can identify $h(T, \overline{X}(T); \theta_0)^{c_2}$ up to scale and the distribution of U_0 up to the same scale parameter.

If the covariates are time–varying we can, except in special cases, identify $h(T, \overline{X}(T); \theta_0)$ and the distribution of U_0 up to a common scale parameter. Because we leave the distribution of U_0 unspecified in our estimation method, we can not use restrictions on U_0 to find the scale parameter. For that reason we normalize $h(T, \overline{X}(T); \theta_0)$ by setting $h(T, 0; \theta_0) = 1$ for some $t_0 > 0$. With time constant regressors we need the same normalisation, but in addition we need to set one regression coefficient equal to one. Of course, we could choose a class of transformations that is not closed under the power transformation. This amounts to identification by functional form.

Finally, we need a condition on the sample paths of X in the population. If we rewrite (2) as

$$\int_0^T e^{\ln \lambda(s;\alpha_0) + \beta_0' X(s)} \mathrm{d}s = U_0 \tag{8}$$

we require that

$$\Pr\Big(\ln\lambda(s;\alpha_0)+\beta_0'X(.)=0\Big)=0 \tag{9}$$

where the probability is computed over the distribution of X as a random function of t and 0 is the zero function. In other words, $\ln\lambda$ is not collinear with X.

3 Endogenous Covariates in Duration Models

It can rarely be defended that a study on unemployment durations includes all the relevant characteristics of the individuals looking for a job. For example, consider like in our application a study on the effect of a cash-bonus for finding a job within a certain period on the re-employment probability. Because such a bonus increases the reward of leaving unemployment it gives an incentive to search more intensively and therefore it increases the re–employment hazard. However, the search intensity of the unemployed individuals is usually not observed.

Suppose that the unemployed have to choose at the start of their unemployment spell whether they want to be eligible for a bonus. If they choose to be eligible they have to fill in some forms, notify their new employer and provide a proof that they held that new job for at least four months. Thus, joining the bonus program implies some administrative duties and cooperation with their new employer for the unemployed. This might refrain some individuals from joining the bonus program, as we see in our application. It is very likely that the unobserved motivation to return to work has an impact on both the decision to join the bonus program and the search intensity. This implies that the indicator of joining the bonus program is an endogenous variable for the analysis of the unemployed duration. Without adjusting for this (self)-selection standard duration analysis give biased results of the effect of the bonus on unemployment duration.

A way of adjusting for an endogenous variable is the conventional instrumental variable method that assumes instrument–error independence and an exclusion restriction. A familiar example of an instrumental variable is the treatment assignment–indicator of a randomly assigned treatment in which the actual treatment still depends on a decision by the agents (or on decisions made by those who execute the program). For instance, long–term unemployed can be randomly assigned to a training program, but for many programs they can still decide not to join, or the training manager can decide to withhold some training from some people. Then, the assignment indicator is an instrument for the actual indicator of training received.

The method of instrumental variables (IV) is widely used in econometrics. For illustration consider the simple linear model

$$Y = \beta' X + \gamma D + \epsilon$$

where Y is observed outcome, X is a vector of exogenous variables, D is an endogenous variable, and ϵ is a disturbance with mean 0. If D and ϵ are correlated OLS gives biased estimates of $\theta = (\beta, \gamma)$. The conventional IV method uses an instrument R that affects D but is uncorrelated with ϵ, like the assignment indicator in a random but compromised experiment. If we denote $Z = (X, R)$ and $\tilde{X} = (X, D)$ the IV estimator is

$$\hat{\theta}_{IV} = \big(Z'\tilde{X}\big)^{-1} Z'Y$$

Complications arise if the outcome variable of interest is a duration variable, like the unemployment duration. Models for duration data, as shown in section 2, are usually non-linear. In general the value of the endogenous variable may depend on information that accumulates during the evolution of the duration. The common approach to accommodate such time-varying variables is to relate them to the hazard rate. Another reason to consider the effect on the hazard rate is that duration data are usually (right)-censored, due to a limited observation window. The hazard rate is invariant to censoring and is therefore the natural choice for the analysis of duration data. Let $D(t)$ be the value of the endogenous variable at duration t. Then the GAFT model with endogenous variables becomes

$$\int_0^T \lambda(s;\alpha) e^{\beta' X(s) + \psi(s, D(s), \gamma)}\, \mathrm{d}s = U = h\big(T, \overline{X}(T), \overline{D}(T), \theta\big) \tag{10}$$

where $\psi(t, D(t), \gamma)$ captures the effect of the endogenous variable and $\theta = (\beta', \alpha', \gamma')'$. Without loss of generality we assume that D is a binary variable that only changes at prediscribed durations and that the effect of the endogenous variable may change over the duration. Then a flexible functional form for ψ is

$$\psi(t, D(t), \gamma) = \sum_{j=0}^{J} \gamma_j \cdot D_j \cdot I_j(t) \tag{11}$$

where $I_j(t) = I(t_j < t \leq t_{j+1})$ with $t_0 = 0$ and $t_{J+1} = \infty$.

If D is an exogenous variable, standard techniques for the analysis of survival time data can be used to estimate the γ's. For example, we can use a Mixed Proportional Hazards model and estimate γ using (semi–parametric) Maximum Likelihood procedures, depending on the assumptions we make about the distribution of the unobserved heterogeneity, V, and the baseline hazard. If the model is correctly specified the MLE yields a consistent estimate.

However, we will get biased estimation results for the parameters if the covariate D is endogenous. Consider, for example, a randomized trail with selective compliance to the assigned treatment. Let $R = 0, 1$ be the random assignment and $D = 0, 1$ the actual treatment. The problem is that even if D has no effect on the hazard the parameter vector γ may not have a causal interpretation, because those who comply with their assigned treatment differ in observed and unobserved characteristics from those who do not comply.

Since physical randomization implies that at time zero all attributes of the two treatment groups are (in expectation) identical, a commonly used solution to this problem, is to ignore the post–randomization compliance and rely on the analysis of the treatment assignment groups. This intention–to–treat (ITT) analysis replaces the actual treatment, D by the treatment assignment indicator, R in the estimation procedure. Further, if the model is correctly specified the estimated γ's effect will correspond to the overall effect that would be realized in the whole population, under the assumption that the compliance rate and the factors influencing compliance in the sample are identical to those that would occur in the whole population.

The major drawback of the intention–to–treat analysis is that the estimated effect is a mixture of the population effect and the effect on the compliance. Hence, if the treatment effectively raises the re–employment hazard, the intention–to–treat measure of this effect will diminish as non–compliance increases. Another disadvantage is that compliance is very likely to depend on the perceived effects of the treatment. If, for example, the unemployed know that being eligible for a re–employment bonus does not stigmatize them, they will be more prone to participate. Thus, when the pattern of compliance is a function of the perceived efficacy of the treatment the estimated intention–to–treat will not represent the overall effect of the treatment had it been adopted in the whole population.

3.1 GAFT model with endogenous interventions

We propose an Instrumental Variable method for duration models that adjusts for the possible endogeneity of the intervention, without suffering the problems of the intention–to–treat method. This is accomplished by using the GAFT model, defined in (10), that transforms the duration such that for the true, population, parameter the transformed duration is independent of the instrument.

The intuition behind the idea of transforming can be clarified by considering the simple example of an experiment with random assignment and selective compliance at the start of the study. If the treatment has no impact on the hazards the probability of observing an individual with $R = 1$ among the survivors at some duration t should be equal to the treatment assignment probability at the start, $\Pr(R = 1)$. If the treatment has an effect, say positive, on the hazard this does not hold, since the treated individuals find a job faster. However, on the transformed duration time-scale, using the true, population, parameters, $U_0 = h\big(T, \overline{X}(T), \overline{D}(T), \theta_0\big)$, in the GAFT model in (10) the proportion of the individuals in assignment group remains the same

$$\Pr\big(R = 1 \mid U_0 \geq u\big) = \Pr\big(R = 1 \mid T \geq 0\big), \tag{12}$$

In other words: the basic assumption underlying the IV estimator is that for the right GAFT model the distribution of the duration on the transformed time scale is independent of the instrument. This implies that the hazard of the *population* transformed duration is independent of the instrument. This independence only holds for the population parameters and therefore we can build an estimation procedure that exploits this conditional independence[1]. In section 5 we discuss this Instrumental Variable Linear Rank (IVLR) estimator.

The interpretation of the parameters of the GAFT-model is the same as defined in the section 2.1, (7). If, for example, the endogenous variable is time-invariant and has a fixed effect, thus $\psi(t, D, \gamma) = \gamma D$, then the relation between the q-th quantile of the transformed duration U_0 for $D = 0$, $t_q(0)$, and for $D = 1$, $t_q(1)$, is

$$\frac{\mathrm{d}\ t_q(1)}{\mathrm{d}\ t_q(0)} = e^{-\gamma_0' D} \exp\Big(-\beta_0'\Big(X(t_q(1)) - X(t_q(0))\Big)\Big) \frac{\lambda\big(t_q(0);\alpha\big)}{\lambda\big(t_q(1);\alpha\big)} \tag{13}$$

If $\gamma_0 > 0$ and assuming incorrectly an AFT model would overestimate the effect of the D if $\lambda\big(t_q(0);\alpha\big) < \lambda\big(t_q(1);\alpha\big)$ and underestimate the effect if $\lambda\big(t_q(0);\alpha\big) > \lambda\big(t_q(1);\alpha\big)$. The hazard or re–employment rate of unemployment durations often exhibits a spike just before the time that unemployment benefits are exhausted. This corresponds to a large increase in λ at that spike. For the bonus data this implies that, if we assume that all unemployed who find a job receive the bonus and that the bonus has a small positive effect on the job finding rate, $t_q(0) > t_q(1)$. If $t_q(0)$ is in the spike while $t_q(1)$ is not, then the left–hand side of (13) is greater than one and the AFT treatment effect at $t_q(1)$ is negative. Thus if there is (substantial) variation in λ, the AFT treatment effects will be biased.

The identification in the GAFT model with endogenous variables is the same as in the GAFT model with only exogenous variables, except that we need additional assumptions on the instrument. First, the instrument should only affect the duration through the endogenous variable and not directly. Second, the value of the instrument should influence the value of the endogenous variable in a non-trivial way. For example, if both the instrument and the endogenous variable are binary then $\Pr(D = 1|R = 1) > 0$ and $\Pr(D = 0|R = 0) > 0$.

[1]Here we only concentrate on a static binary instrument and a discrete, but possible time-varying according to a prescribed protocol, endogenous variable. It is not difficult to extend the analysis to more, discrete, levels of both the instrument and the endogenous variable and to have a sequential instruments.

4 Counting process interpretation

The density and the survival function of a duration T can be expressed as functions of the hazard rate. These expressions can be used to obtain a likelihood function. We use a different (but of course equivalent) representation of the relation between the hazard rate and the random duration. In particular, we use the framework of counting processes (see e.g. Andersen et al. (1993) and Klein and Moeschberger (1997)). The main advantage of this framework is that it allows us to express the duration distribution as a regression model with an error term that is a martingale difference. This simplifies the analysis of the estimator. The conditions for non selective observation can be precisely stated in this framework. The same is true for conditions on time–varying covariates.

The starting point is that the hazard of T is the intensity of the counting process $\{N(t); t \geq 0\}$ that counts the number of times that the event occurs during $[0, t]$. The counting process has a jump +1 at the time of occurrence of the event[2]. A jump occurs if and only if $\mathrm{d}N(t) = N(t) - N(t-) = 1$. For duration data, the event can only occur once. In many unemployment studies the individuals are only observed until re-employment. So, at most one jump is observed for any unit. To account for this we introduce the observation indicator $Y(t) = I(T \geq t)$ that is zero after re–employment. By specifying the intensity as the product of this observation indicator and the hazard rate we effectively limit the number of occurrences of the event to one. We assume that the observation indicator only depends on events up to time t. The observation process is assumed to have left–continuous sample paths. We define the history of the process up to time t by $H(t) = \{\overline{Y}(t), D, \overline{X}(t)\}$, where $\overline{Y}(t) = \{Y(s), 0 \leq s \leq t\}$. The history $H(t)$ only contains observable events.

Let V be some unobserved variables that both influence the endogenous variable and the duration. An example is the, usually, unobserved search intensity of unemployed looking for a job. We assume that V and $\overline{X}(t)$ are stochastically independent. Denote $H^V(t) = \{H(t), V\}$, the history that also includes the unobservables. As with dynamic

[2]The sample paths are assumed to be right-continuous.

regressors in time-series models, the time-varying $X(t)$ may depend on the dependent variable up to time t but not after time t (conditionally on V). Thus D only depends on $H^V(t)$ and $X(t)$ only on $H(t)$. In the counting process literature such a time-varying covariate is called predictable. We will use the econometric term predetermined.

If the conditional distributions of $N(t)$ given $H^V(t)$ or $H(t)$ are well-defined (see Andersen et al. (1993) for assumptions that ensure this) we can express the probability of an event in $(t - dt, t]$ as[3]

$$\Pr(\mathrm{d}N(t) = 1 \mid H^V(t)) = Y(t)\kappa\big(t \mid \overline{X}(t), D, V\big)\mathrm{d}t \tag{14}$$

with $\kappa(t \mid \cdot)$ is the hazard of T at t given $\overline{X}(t)$, D and V. By the Doob-Meier decomposition

$$\mathrm{d}N(t) = Y(t)\kappa\big(t \mid \overline{X}(t), D, V\big)\mathrm{d}t + \mathrm{d}M(t) \tag{15}$$

with $\{M(t); t \geq 0\}$ a (local square integrable) martingale. The conditional mean and variance of this martingale are

$$\mathrm{E}\big(\mathrm{d}M(t) \mid H(t)\big) = 0 \tag{16}$$

$$\mathrm{Var}\big(\mathrm{d}M(t) \mid H(t)\big) = Y(t)\kappa\big(t \mid \overline{X}(t), D, V\big)\mathrm{d}t \tag{17}$$

The (conditional on $H(t)$) mean and variance of the counting process are equal, so that the disturbances in equation (15) are heteroscedastic. The probability in equation (14) is zero, if the individual is not at risk.

A counting process can be considered as a sequence of Bernoulli experiments, because if $\mathrm{d}t$ is small equations (14) and (17) give the mean and variance of a Bernoulli random variable. The relation between the counting process and the sequence of Bernoulli experiments is given in equation (15), which can be considered as a regression model with an additive error that is a martingale difference. This equation resembles a time-series regression model. The Doob-Meier decomposition is the key to the derivation of the distribution of the estimator, because the asymptotic behavior of partial sums of martingales is well-known.

[3]Because the sample paths of $\{Y(t), X(t), t \geq 0\}$ are assumed to be left-continuous (as is the baseline hazard), we can substitute t for $t - dt$ in (14).

4.1 Transformed Counting process

The GAFT model transforms the observed duration T to a transformed duration U_0. The transformation involved a parameter vector $\theta_0 = (\beta_0', \gamma_0', \alpha_0')'$. We denote the transformation for parameter vectors $\theta \neq \theta_0$ by $U(\theta)$ with $U_0 = U(\theta_0)$. The distribution of $U(\theta)$ can also be represented by a (transformed) counting process $\{N^U(u); u \geq 0\}$. The relation between the original and transformed counting process, the observation indicator, and the time-varying exogenous covariates is

$$
\begin{aligned}
N^U(u;\theta) &= N\big(h^{-1}(u;\theta)\big) & \qquad Y^U(u;\theta) &= Y\big(h^{-1}(u;\theta)\big) \\
X^U(u;\theta) &= X\big(h^{-1}(u;\theta)\big) & \qquad I_k^U(u;\theta) &= I_k\big(h^{-1}(u;\theta)\big)
\end{aligned}
$$

with $h(T;\theta) = h\big(T, \overline{X}(T), \overline{D}(T);\theta\big)$, defined in (10), and $I_k(t) = I(t_k < t \leq t_{k+1})$. For $\theta = \theta_0$ we denote $h_0(T) = h(T;\theta_0)$. The corresponding history is $H^U(u;\theta) = \{\overline{Y}^U(u;\theta), \overline{X}^U(u;\theta), \overline{I}_k^U(u;\theta), D\}$. In the sequel we suppress θ and write $Y^U(u), N^U(u), \overline{X}^U(u), \overline{I}_k^U(u)$ and $H^U(u)$ for $\theta \neq \theta_0$ and $Y_0(u), N_0(u), \overline{X}_0(u), \overline{I}_{k0}(u)$ and $H_0(u)$ for $\theta = \theta_0$. The intensity of the transformed counting process with respect to history $H^U(u)$ is obtained by the innovation theorem (see Andersen et al. (1993), p. 80, 87)[4]

$$
\begin{aligned}
\Pr\big(\mathrm{d}N^U(u) = 1 \mid H^U(u)\big) = Y^U(u)\mathrm{E}\Bigg[&\frac{\lambda\big(h^{-1}(u;\theta);\alpha_0\big)}{\lambda\big(h^{-1}(u;\theta);\alpha\big)} e^{(\beta_0-\beta)'X^U(u)} \\
&\times \exp\Big(\sum_{k=1}^{K}(\gamma_{k0}-\gamma_k)I_k^U(u)D\Big)\kappa_0\Big(h_0\big(h^{-1}(u;\theta)\big)\Big)\Bigg| H^U(u)\Bigg]\mathrm{d}u \qquad (18)
\end{aligned}
$$

We implicitly integrate with respect to the distribution of the unobserved V conditional on $H^U(u)$. Note that these unobserved covariates are only introduced to ascertain the predictability of the endogenous covariate process. Although the distribution of those variables determines the distribution of U_0, the consistency of the IVLR is independent of that distribution. Unfortunately, even for the population parameters θ_0 the hazard of

[4] If $U = h(T)$ and κ_T is the hazard rate of the distribution of T, then the hazard rate of the distribution of U is

$$\kappa_U(u) = \kappa_T\big(h^{-1}(u)\big)\frac{1}{h'\big(h^{-1}(u)\big)}$$

U_0, $\kappa_0(u)$, still depends on the intervention path (through the correlation with V). If we condition on the history of the instruments instead of the actual endogenous covariates we do get the desired independence.

We must add the instrument R to the conditioning variables in (18) if we consider instrumenting the endogenous variable. Let the UR–history, $H^{UR}(u) = \{Y^U(s), X^U(s), R; 0 \leq s \leq u\}$, be the history on the transformed durations in which the endogenous variable D is replaced by the instrument. Then, another application of the innovation theorem gives the intensity of the transformed process on the UR–history

$$\Pr\big(\mathrm{d}N^U(u) = 1 \mid H^{UR}(u)\big) = Y^U(u)\mathrm{E}\Bigg[\frac{\lambda\big(h^{-1}(u;\theta);\alpha_0\big)}{\lambda\big(h^{-1}(u;\theta);\alpha\big)}e^{(\beta_0-\beta)'X^U(u)} \times \exp\Big(\sum_{k=1}^{K}(\gamma_{k0}-\gamma_k)I_k^U(u)D\Big)\kappa_0\Big(h_0\big(h^{-1}(u;\theta)\big)\Big)\Bigg|H^{UR}(u)\Bigg]\mathrm{d}u \quad (19)$$

which for the population parameters simplifies to $Y_0^U(u)\kappa_0(u)\mathrm{d}u$ with $H_0^{UR}(u) = H^{UR}(u;\theta_0)$. Note that (18) and (19) only differ in the history the intensities are conditioned on.

The intensity in (19) is independent of (the history) of X and R if we substitute the population parameter values, but not for other values of the parameters. This forms the basis for identification of the parameters. Independence of (the history) of X and R and the hazard rate of U_0 implies that the quantiles of the distribution of U_0 do not depend on X or R. By choosing C_0^U such that $\Pr(U_0 \leq C_0^U) = q$ we restrict the independence to the quantiles up to the q–th. For further reference we denote the intensity in (19) by $\kappa_i^U(u;\theta)$ such that

$$\Pr\big(\mathrm{d}N^U(u) = 1 \mid H^{UR}(u)\big) = Y^U(u)\kappa_i^U(u;\theta)\mathrm{d}u$$

which reduces to $\kappa_0(u)$ for the population parameters.

4.2 Censoring and endogenous covariates

A common feature of duration data is that some of the observations are censored. Assume the censoring time, C, is (potentially) known. For example, in the analysis of unemployment duration based on administrative data the duration is often only observed while the individual receives unemployment benefits. Usually, the maximum duration of receiving benefits is based on the labor market history of the individual and is recorded in the data. Then, the potential censoring time is known and the observed durations are $\tilde{T} = \min(T, C)$ and $\Delta = I(T \leq C)$, where Δ is one if T is observed.

One is tempted to define the censored transformed durations by the minimum of the transformed time till (potential) censoring and the transformed time till the event occurs, $\tilde{U}(\theta) = \min\big(h(T;\theta), h(C;\theta)\big) = h(\tilde{T};\theta)$. However, with endogenous covariates censoring makes some of the orthogonality conditions fail to hold. This can be illustrated by a simple example: Consider a fixed censoring time, all individuals have the same maximum duration of receiving benefits. Then for all individuals, irrespective of their value of the endogenous variable, censoring occurs at time C. Suppose the binary endogenous variable, D, and other covariates all be determined at the start of the study and have a constant effect on the hazard. Finally, we assume that except for the γ the effect of the endogenous variable, all parameters, β_0 and α_0, are known. Then, the transformation is

$$U_0 = e^{\gamma_0 D + \beta_0' X} \Lambda_0(T) \tag{20}$$

with $\Lambda_0(t) = \int_0^t \lambda(s, \alpha_0)\,\mathrm{d}s$. Hence, if $D = 0$ censoring in the transformed time occurs at $e^{\beta_0' X}\Lambda_0(C)$, but if $D = 1$ censoring occurs at $e^{\beta_0' X + \gamma_0}\Lambda_0(C)$. Thus, if $\gamma_0 > 0$, then all transformed durations in the interval $[e^{\beta_0' X}\Lambda_0(C), e^{\beta_0' X + \gamma_0}\Lambda_0(C)]$ have $D = 1$ (for $\gamma_0 < 0$ the boundaries are reversed). The hazard of U_0 on this interval clearly depends on D and hence on R. The independence of the hazard of U_0 and R only holds up to the lower bound of the interval. This implies that in the IVLR, which exploits this independence, the transformed durations that fall in the problematic interval have to be censored.

This can be generalized to the model with a time–varying coefficient of the endogenous variable and time–varying exogenous covariates. Assume the piecewise constant structure for the effect of the endogenous variable in (11). This implies that for $t_k < t \leq t_{k+1}$, the coefficient of $D = 1$ is e^{γ_k}. We define the transformed censoring time $C^U(\theta)$ (possibly depending on the observed history of other covariates) such that: (a) $T \geq C$ implies $h(T;\theta) \geq C^U(\theta)$ and (b) U_0 and R are independent on the interval bounded above by

$C^U(\theta)$.

Note that we either observe $T \le C$ and $\Delta = 1$, or $T > C$ and $\Delta = 0$. If some of the other covariates are also time–varying we have another identification problem, because these covariates are only observed up until $\tilde{T}$. The transformed censoring times (conditional on $T, C > t_k$) that take all these considerations into account are the sum of the transformed duration up to t_k, $h(t_k;\theta)$ and the censoring adjustment, *i.e*

$$C^U(\theta) = \begin{cases} \int_0^C \lambda(s;\alpha) e^{\beta' X(s)} P(s;\gamma)\, \mathrm{d}s & \text{if } T > C, \\ \int_0^T \lambda(s;\alpha) e^{\beta' X(s)} P(s;\gamma)\, \mathrm{d}s + \int_T^C \lambda(s;\alpha)\, ds & \text{if } T \le C. \end{cases} \tag{21}$$

where $P(s;\gamma) = I(s \ge t_k) \prod_{j=0}^{k} \min(e^{\gamma_j}, 1)$. From the last term on the right–hand side of (21) we see why we need to know C even for the uncensored observations. Otherwise we can not compute $C^U(\theta)$ for these observations. We can estimate the parameters of the Instrumental GAFT model from the following observed data

$$\tilde{U}(\theta) = \min\big(U(\theta), C^U(\theta)\big), \qquad \Delta^U(\theta) = I\big(U(\theta) < C^U(\theta)\big)$$

and $Y^U(u;\theta) = I\big(\tilde{U}(\theta) \ge u\big)$. Now $\tilde{U}(\theta_0)$ is independent of R for $\Delta^U(\theta_0) = 1$. Note that if, at least, one of the γ's is different from zero, we introduce extra censoring on the transformed durations, because then some units with $\Delta = 1$ have $\Delta^U(\theta) = 0$.

5 Instrumental Variable Linear Rank Estimation

In this section we introduce the Instrumental Variable Linear Rank (IVLR) estimator of the GAFT model. The estimator is based on independence of the transformed durations $\{\tilde{U}(\theta_0), \Delta^U(\theta_0)\}$ and R and the properties of the IVLR are derived using the transformed counting process $N^U(u)$. The IVLR is motivated by (19) and is defined on the possibly censored durations $\tilde{U}(\theta)$. For notational convenience we suppress the dependence on θ. Only when we want to emphasis this dependence we include it.

5.1 The IVLR estimator

For the population parameter vector θ_0 the hazard of U_0, $\kappa_0(u)$, is independent of the covariate and instrument history up to $h_0^{-1}(u)$. Because this is true only for $\theta = \theta_0$, we can use an estimate of (19) as an estimating equation. This independence can be used to construct test statistics close to the linear rank test (see Prentice (1978)). The IVLR also exploits this independence and is the estimation procedure derived from these rank tests.

The estimating equation that defines the IVLR estimator contains a left–continuous weight function W. The dimension of W is greater than or equal to the dimension of θ_0 which is p. The weight function may depend on $\tilde{U}_i(\theta) = \tilde{U}_i$ and $\overline{X}_i^U(u)$ and R. Typical examples are $W = \overline{X}^U(u)$, for the coefficient β of the exogenous variables and $W = R$, for a time-constant coefficient of the endogenous variable D. The variance of the IVLR estimator depends on W and in section 5.2 we discuss the optimal choice of this function. The IVLR estimator is defined by the estimating equations

$$S_n(\theta; W) = \sum_{i=1}^{n} \Delta_i^U \Big\{ W\big(\tilde{U}_i, \overline{X}_i^U(\tilde{U}_i), R_i; \theta\big)\big) - \overline{W}(\tilde{U}_i; \theta) \Big\} \tag{22}$$

where

$$\overline{W}(\tilde{U}_i; \theta) = \frac{\sum_{j=1}^{n} Y_j^U(\tilde{U}_i) W\Big(\tilde{U}_i, \overline{X}_j^U(\tilde{U}_i), R_j; \theta\Big)}{\sum_{j=1}^{n} Y_j^U(\tilde{U}_i)},$$

the average of the weight function evaluated at $\tilde{U}_i(\theta)$ among the individuals still at risk. Note that we use Δ_i^U instead of Δ_i to assure independence of the instruments and the transformed durations for all uncensored observations.

The interpretation of $S_n(\theta; W)$ is that it compares the weight function for a transformed duration that ends at $\tilde{U}_i(\theta)$ to the average of the weight functions at that time for those individuals that are still under observation. The suggestion is that the difference of the weight function for individual i and the average weight function for the individuals under observation is zero at the population parameters. Thus, the statistic $S_n(\theta; W)$ has mean zero at the population parameters and, therefore, we base our estimator on the roots of $S_n(\theta; W) = 0$. However, the estimating functions are discontinuous, piecewise constant, functions of θ and a solution may not exist. For that reason we define the Instrumental Linear Rank estimator (IVLR) $\hat{\theta}_n(W)$ as the minimizer of the quadratic form, *i.e.*

$$\hat{\theta}_n(W) = \inf\{\theta \mid S_n(\theta; W)' S_n(\theta; W)\} \tag{23}$$

The counting process interpretation of duration models allows for another, of course equivalent, formulation of the estimating equations in (22). The relevant counting measure, $N_i^U(u)$, can be seen as a discrete 'probability distribution' that assigns weight unity to uncensored transformed durations and is zero elsewhere. Then the estimating equations can be expressed as an integral with respect to that counting process

$$S_n(\theta; W) = \sum_{i=1}^{n} \int_0^{C_i^U} \Big\{ W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big\} \, \mathrm{d}N_i^{\tilde{U}}(u) \tag{24}$$

where C_i^U is the transformed censoring time defined in (21). To ensure weak consistency and asymptotic normality of the IVLR estimator we make the following assumptions. The random variable R is an instrument that is determined at the start. We restrict both the instrument, R, and the endogenous variable D, to be binary. The other assumptions can be found in the appendix.

If $S_n(\theta; W)$ were differentiable with respect to θ, then asymptotic normality can be proved using Taylor series expansion in a neighborhood of θ_0. Tsiatis (1990) showed that, if $S_n(\theta; W)$ is not differentiable, as in the current problem, we can still use a linear approximation of $n^{-1/2} S_n(\theta; W)$. Using this approximation and the asymptotic normality of $S_n(\theta_0; W)$, we can show that $\sqrt{n}(\hat{\theta}_n(W) - \theta_0)$ is asymptotically normal. Let $a(u;\theta_0)$ be the probability limit of the average weight function (see assumption A6), C_0 the transformed censoring time for $\theta = \theta_0$. Let $d_{i0}(u)$ the derivative of the hazard of $U(\theta)$ w.r.t. θ, i.e.

$$d_{i0}(u) = \frac{\partial \kappa_i^U(u;\theta)}{\partial \theta}\Big|_{\theta=\theta_0}$$

and $V(u,\theta)$ is the probability limit of

$$\frac{1}{n} \sum_{i=1}^{n} \Big[W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big] \times d_{i0}(u)' Y_i^U(u)$$

The asymptotic properties of the IVLR estimator are summarized in the following two theorems.

Theorem 1 (Consistency)**.**

If assumptions C1 to C7 hold $\hat{\theta}_n(W)$ converges in probability to θ_0.

Proof: See the appendix.

Theorem 2 (Asymptotic Normality)**.**

If assumptions C1 to C9 hold and $Q(W)$ *has full rank, then*

$$\sqrt{n}(\hat{\theta}_n(W) - \theta_0) \xrightarrow{d} N\big(0, Q^{-1}(W)\Omega(W)Q'^{-1}(W)\big) \tag{25}$$

where

$$\Omega(W) = \int_0^{C_0} a(u;\theta_0)\kappa_0(u)\, du \tag{26}$$

is the asymptotic variance of $n^{-1/2}S_n(\theta_0; W)$ *and,*

$$Q(W) = \int_0^{C_0} V(u, \theta_0)\, du \tag{27}$$

the limiting covariance matrix of the processes $W(u, \overline{X}_{i0}(u), R_i)$ *and* $d_{i0}(u)/\kappa_0(u)$*.*

Proof: See the appendix.

5.2 Efficiency of the IVLR estimator

Many different choices of the weight functions lead to consistent estimates of the parameters. By properly choosing the weight function the asymptotic variance of the IVLR can be minimized. Tsiatis (1990) has shown that for the AFT model with exogenous covariates weight functions proportional to $u\kappa_0'(u)/\kappa_0(u)X$ minimize the asymptotic variance of the estimated regression parameters. In general the distribution of U_0 is unknown. This distribution can, however, consistently be estimated from the implied U from any IVLR with a weight function not involving these functionals.

The IVLR estimation is based on a vector of mean restrictions on weight functions of the covariates, instrument and the transformed durations. GMM estimation is also based on moment conditions and in GMM estimation it is feasible to get the most efficient GMM estimator in just two steps. A similar reasoning applies to the IVLR-estimator. This justifies an adaptive construction of an efficient estimator. In the next section we address the practical implementation of an adaptive estimation procedure. First, we introduce the optimal weight function.

Theorem 3 (Optimal weight function in IVLR)**.**

The weight–function that gives the smallest asymptotic variance for $\hat{\theta}_n(W)$ *is*

$$W_{\text{opt}}(u, \overline{X}(u), \overline{R}(u)) \propto \left.\frac{\partial \ln \kappa^U(u;\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \frac{d_{i0}(u)}{\kappa_0(u)} \tag{28}$$

The asymptotic covariance matrix of the optimal IVLR estimator reduces to

$$\Omega^{-1}(W_{\text{opt}}) = Q^{-1}(W_{\text{opt}}). \tag{29}$$

Proof of theorem 3. From

$$\frac{1}{\sqrt{n}}\begin{pmatrix} S_n(\vartheta_0;W) \\ S_n(\vartheta_0;W_{\text{opt}})\end{pmatrix} \xrightarrow{D} N\left(0, \begin{pmatrix}\Omega(W) & Q(W)' \\ Q(W) & \Omega(W_{\text{opt}})\end{pmatrix}\right)$$

follows that the matrix

$$Z = \begin{pmatrix}\Omega(W) & Q(W)' \\ Q(W) & \Omega(W_{\text{opt}})\end{pmatrix}$$

is non–negative definite, the same is true for its inverse. In particular, the submatrices on the main diagonal of the inverse are non-negative definite. Hence the matrix

$$Q^{-1}(W)\Omega(W)Q'^{-1}(W) - \Omega^{-1}(W_{\text{opt}})$$

is a non-negative definite matrix. □

Consider, for example, a GAFT model with a piecewise constant λ function,

$$\lambda(t,\alpha) = \sum_{j=0}^{J} e^{\alpha_j} I(t_j < t \leq t_{j+1}) \tag{30}$$

with $t_0 = 0$ and $t_{L+1} = \infty$ and the hazard on the last interval is normalized to 1, $\alpha_L = 0$. Assume that the model has a constant coefficient for the endogenous variable then by

(28) the optimal weight functions are

$$W_{\text{opt},\beta} = X(u)\left[1 + u\frac{\kappa_0'(u)}{\kappa_0(u)}\right] \tag{31}$$

$$\begin{aligned} W_{\text{opt},\alpha_{\text{j}}} &= \left(1 + u\frac{\kappa_0'(u)}{\kappa_0(u)}\right) \cdot \Big(RI_j^1(u) + (1-R)I_j^0(u)\Big) + \\ &+ R\left[\big(1 + u\kappa_0(u)\big)\frac{f_0(u|1,R) - f_0(u)}{f_0(u)} + u\frac{f_0'(u|1,R) - f_0(u)}{f_0(u)}\right] I_j^1(u) + \\ &+ (1-R)\left[\big(1 + u\kappa_0(u)\big)\frac{f_0(u|0,R) - f_0(u)}{f_0(u)} + u\frac{f_0'(u|0,R) - f_0(u)}{f_0(u)}\right] I_j^0(u) \end{aligned} \tag{32}$$

$$\begin{aligned} W_{\text{opt},\gamma} &= R\left[1 + u\frac{\kappa_0'(u)}{\kappa_0(u)}\right] + \\ &+ R\left[\big(1 + u\kappa_0(u)\big)\frac{f_0(u|1,R) - f_0(u)}{f_0(u)} + u\frac{f_0'(u|1,R) - f_0(u)}{f_0(u)}\right] \end{aligned} \tag{33}$$

where $f_0(u|D,R)$ is the density of U_0 given D and R, $f_0'(\cdot)$ is the derivative of the density and $I_j^D(u) = I\big(m_j(X,D) < u \leq m_{j+1}(X,D)\big)$ for

$$m_j(X,D) = \int_0^{t_j} \lambda(s,\alpha)e^{\beta' X(s) + \gamma D}\,\mathrm{d}s$$

5.3 Estimation in practice

The statistic $S_n(\theta; W)$ is a multi–dimensional step–function. Therefore, the standard Newton–Raphson algorithm cannot be used to solve (23). One of the alternative methods for finding a zero of a non–differentiable function is the Powell-method. This method (see Press et al. (1986, §10.5) and Powell (1964)) is a multidimensional version of the Brent algorithm.

Related to the computation of optimal weight function is the estimation of the variance matrix for an arbitrary weight function.[5] The difficulty in estimating the covariance matrix lies in the calculation of the matrix $Q(W)$ and not in the calculation of the variance matrix

[5] Robins and Tsiatis (1991) suggested to use a numerical derivative of $n^{-1}S_n(\theta; W)$ that does not need an estimate of the optimal W–function to get $\hat{Q}(W)$. This numerical derivative is sensitive to the choice of the difference in θ. We found it hard to get stable results.

of the estimating equation. The latter can be consistently estimated by

$$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{C_i^U(\hat{\theta})}\Big[W\big(u,\overline{X}_i^U(u),R_i\big) - \overline{W}(u,\hat{\theta})\Big]\Big[W\big(u,\overline{X}_i^U(u),R_i\big) - \overline{W}(u,\hat{\theta})\Big]'\mathrm{d}\hat{N}_i^U(u) \tag{34}$$

where $\hat{N}_i^U(u)$ is the counting process of $U(\hat{\theta})$.

The optimal weight functions, the covariance matrix and the most efficient estimators are estimated in two steps. The first step consists of obtaining a consistent estimate of θ_0 using a weight function that does not depend on the distribution of U_0. For example, in a GAFT model with a piecewise constant λ and a time-invariant coefficient of the endogenous variable, the choice for the first-step weight functions could be $I_j(u), X(u)$ and R. The transformed durations for the estimated parameter values based on these first step weight functions are estimates of the unobserved population transformed durations. The second step concerns the estimation of the unknown distribution of U_0. Many different methods are available to get a reasonable estimate of an unknown distribution. We shall *not* apply the commonly used kernel based method. Although kernel–smoothed hazard rate estimators have been developed and adjusted to deal with the boundary problems inherent to hazard rates these methods can be difficult to implement due to the choice of the bandwidth. It is also unclear how the boundary corrections can be incorporated in the kernel estimates of the derivative of the hazard. We therefore choose to use a series approximation of the distribution.

Suppose the distribution of U_0 can be approximated arbitrary well using orthonormal polynomials. We base our approximation on Hermite polynomials using the exponential distribution as a weighting function:

$$g_0(u) = \frac{ae^{-au}}{\sum_{l=0}^{L} b_l^2}\Big[\sum_{l=0}^{L} b_l L_l(u)\Big]^2 \tag{35}$$

where

$$L_l(u) = \sum_{k=0}^{l}\binom{l}{k}\frac{(-au)^k}{k!} \tag{36}$$

are the Laguerre polynomials. The unknown parameters of this approximation are a and $b_0,\ldots,b_L$. If $b_l \equiv 0$ for all $l > 0$ the distribution of U_0 is exponential. Even for L as small as three (35) allows for many different shapes of $\kappa_0(u)$ and its derivative. Both can be derived analytically given the estimates of the parameters. The parameter estimators can be obtained from standard maximum likelihood procedures on the observed $(\tilde{U}_i(\hat{\theta}_n(W)), \Delta_i)$.

If a consistent but inefficient estimator $\hat{\theta}_n(W)$ of θ_0 is available and we have estimated the parameters of the polynomial approximation of the distribution of U_0 we can obtain an efficient estimator $\hat{\theta}_{\text{opt}}$ in just one extra step. From the linearization of the estimating equations, given in (41), we obtain an efficient estimator from

$$\hat{\theta}_{\text{opt}} = \hat{\theta}_n(W) - \hat{Q}(W)^{-1} S_n(\hat{\theta}_n(W); W_{\text{opt}})/n \tag{37}$$

This procedure is related to obtaining an efficient GMM estimator in two steps from a consistent, but possible, inefficient GMM estimator. It also possible to obtain the efficient estimator directly from minimizing the quadratic form. However, this involves again the minimization of a multi–dimensional step function.

6 Application to the Illinois Re-employment Bonus Experiment

Between mid–1984 and mid–1985, the Illinois Department of Employment Security conducted a controlled social experiment.[6] This experiment provides the opportunity to explore, within a controlled experimental setting, whether bonuses paid to Unemployment Insurance (UI) beneficiaries or their employers reduce the time spend in unemployment relative to a randomly selected control group. In the experiment, newly unemployed claimants were randomly divided into three groups: a *Claimant Bonus Group*, a *Employer Bonus Group* and, a *control group*. The members of both bonus groups were instructed that they (Claimant group) or their employer (Employer group) would qualify for a cash bonus of $500 if they found a job (of at least 30 hours) within 11 weeks and, if they held that job for at least four months. Each newly unemployed individual who was randomly

[6] A complete description of the experiment and a summary of its results can be found in Woodbury and Spiegelman (1987).

assigned to one of the two bonus groups had the possibility to refuse participation in the experiment.

Woodbury and Spiegelman (1987) concluded from a direct comparison of the control group and the two bonus groups that the claimant bonus group had a significantly smaller average unemployment duration. The average unemployment duration was also smaller for the employer bonus group, but the difference was not significantly different from zero. These results are confirmed in Table 1. Note that the response variable is insured weeks of unemployment. Because UI benefits end after 26 weeks, all unemployment durations are censored at 26 weeks. In Table 1 no allowance is made for censoring. In the table we distinguish between compliers, those who agreed to be eligible for a bonus if assigned to a bonus group, and non-compliers. We see that the claimant bonus only affects the compliers and that the average unemployment duration of the non-compliers and the control group are almost equal.

Table 1: Average unemployment durations:control group and (non-)compliers.

	Control Group	Claimant Bonus			Employer Bonus		
		All	Compl.	Non-compl.	All	Compl.	Non-compl.
Benefit weeks	18.33	16.96	16.74	18.18	17.65	17.62	17.72
	(0.20)	(0.20)	(0.22)	(0.50)	(0.21)	(0.26)	(0.35)
N	3952	4186	3527	659	3963	2586	1377

standard error of average in brackets.

About 15% of Claimant group and 35% of the employer group declined participation. The reason for this refusal is unknown. Bijwaard and Ridder (2005) showed that the participation rate is significantly related to some observed characteristics of the individuals that also influence that re–employment hazard. Hence, we cannot exclude the possibility of unmeasured variables that affect both the compliance decision and the re–employment hazard. Meyer (1996) analyzed the same data with a PH model with a piecewise constant baseline hazard. He used the randomization indicator instead of the actual bonus-group agreement indicator as an explanatory variable. Thus he used the ITT estimator. He found a significantly positive effect of the claimant bonus. However, as shown by Bijwaard and

Ridder (2005), the ITT has a downward bias.

We calculate the IVLR estimate of the effect of the claimant and employer bonus on the unemployment duration in a GAFT model and compare these estimates with the IVLR estimates of an AFT model, with ITT estimates in an MPH model and the ML estimates of an MPH model that ignores the endogeneity of the decision to participate in the bonus group. We consider the two interventions separately: thus Claimant Bonus group versus Control group and Employer Bonus group versus Control.

We shall consider two alternative specifications for the effect of the bonus on unemployment duration: (*i*) constant effect and, (*ii*) a change in the effect after 10 weeks, in line with the end of the eligibility period of the bonuses. Thus, the implied transformed durations are

$$U(\theta) = \int_0^T \lambda(s;\alpha) e^{\beta' X + (\gamma_1 I_1(s) + \gamma_2 I_2(s))D} \mathrm{d}s \tag{38}$$

with $I_1(t) = I(0 \leq t < 11)$ and $I_2(t)$ is its complement. Note that the covariates are all time–constant because the individual characteristics available in the data are all determined when the individuals register at the unemployment office. We include the following: the logarithm of the age (LNAGE), the logarithm of the pre–unemployment earnings (LNBPE), gender (MALE= 1), ethnicity (BLACK= 1), and the logarithm of the weekly amount of UI benefits plus dependence allowance (LNBEN). We employ two different specifications for $\lambda(t;\alpha_0)$: (*i*) AFT model, *i.e.* $\lambda(t;\alpha_0) \equiv 1$; and (*ii*) GAFT model with a piecewise constant λ on six intervals 0–2, 2–4, 4–6, 6–10, 10–25 and 25 and beyond.

For identification we need to set one of the parameters of the piecewise constant λ equal to one (or the log equal to zero). We let the base interval, the interval on which $\lambda = 1$, start on the last week before the end of the observation period, at 25 weeks. This allows us to capture the spike in the observed unemployment duration just before the UI eligibility period ends. The end of the UI eligibility period, at 26 weeks, is for all individuals the same and thus provides the potential censoring time.

For both the AFT and the GAFT specifications we estimate a first stage IVLR using the Powell-method and the one step optimal IVLR. The first stage IVLR uses the values of the covariates, X, (only for the GAFT-model) the interval indicators on the transformed duration, $I_j(u)$ and, the bonus group assignment indicator times the interval indicators on the transformed duration, $R \cdot I_1(u)$ and $R \cdot I_2(u)$, as the weight functions. From these first stage IVLR's the implied transformed duration are obtained. Then, we estimate the parameters of the polynomial approximation of the distribution of U conditional on R and D as mentioned in section 5.3. From these estimated parameters we calculate the hazard and its derivative of the transformed duration. These functions are then used as inputs to derive the optimal weight functions (see Theorem 3), which in turn are necessary to calculate the covariance matrix. We also calculate the 1-step efficient estimates with these optimal weight functions. In the case of a constant bonus effect, the optimal weight function are given in (31)–(33). When we assume that the effect of the bonus changes after 11 weeks the optimal weight function in (33) is more complicated and therefore not spelled out here.

The estimation results for the bonus effects are reported in Table 2. The results for the piecewise constant λ and for the regression coefficients in the AFT and GAFT models can be found in appendix B. A comparison of the results shows that AFT overestimates the effect and that both ML and ITT estimators underestimate the effect of the employer bonus. The results clearly indicate that the bonuses only influence the chances to find a job in the first ten weeks. This is in line with the bonus eligibility period: those who find a job after that period would not get the bonus. The effect of the Claimant Bonus increases from about 10% higher probability to find a job at every unemployment duration to about 15% higher probability to find a job in the first ten weeks (and no effect thereafter). The bonus for the Employer group raises the job finding probability with about 7% at every unemployment duration or with about 12% in the first ten weeks of unemployment.

In the GAFT (and AFT) model the effect of the bonus is defined in terms of the change in the quantiles, see (7) and (13). In an AFT model with a time-constant coefficient for the bonus this effect is constant and independent of the other covariates. In a GAFT model the λ function influences this effect directly and indirectly as the other covariates determine the quantiles. Using the distribution of U_0, already calculated to estimate the optimal IVLR and the variance-covariance matrix, we can derive the effect of the bonus in the GAFT depending on the quantile of the distribution. In Table 3 we present the effect of the bonus on the unemployment duration at the 80%, 60% and 40% survival for the reference individual and for a black individual, together with the AFT effect (first stage). Figure 1 till Figure 4 depict the change over the whole 90%-25% survival range of the effect of the bonus in the GAFT model.

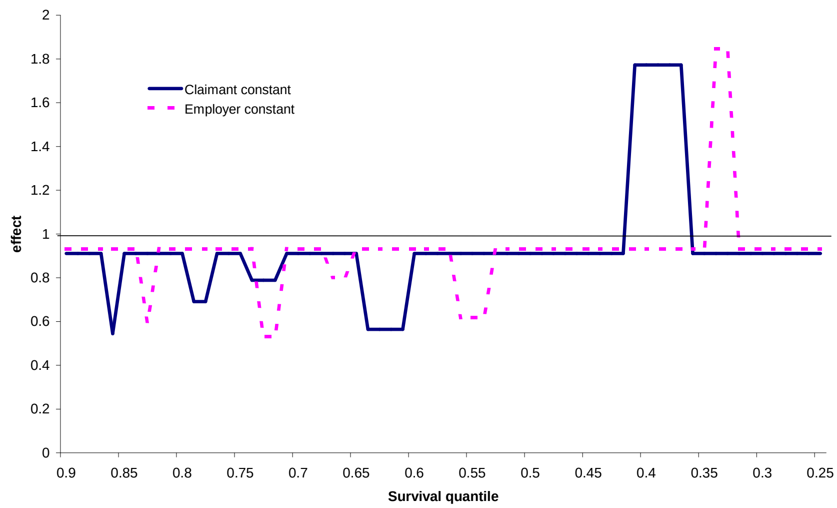

Figure 1: Effect of Bonus on quantiles of unemployment duration, $\frac{\mathrm{d}\ t_q(1)}{\mathrm{d}\ t_q(0)}$ (constant γ)

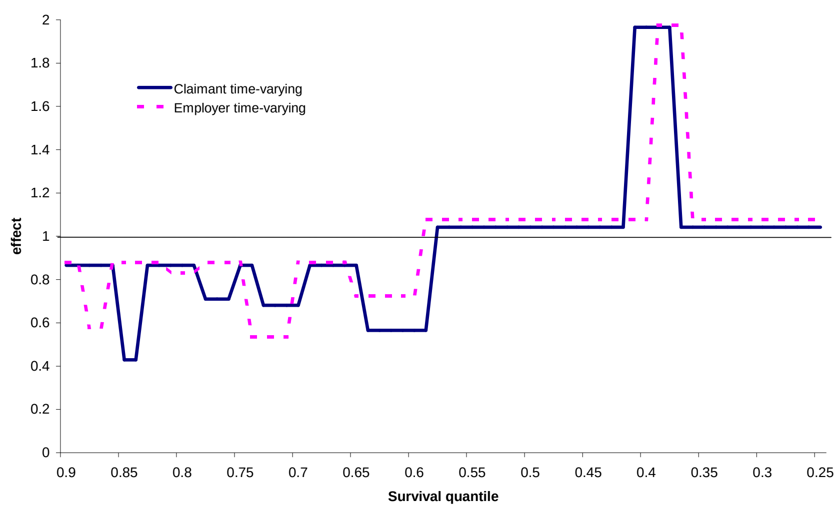

Figure 2: Effect of Bonus on quantiles of unemployment duration, $\frac{\mathrm{d}\ t_q(1)}{\mathrm{d}\ t_q(0)}$ (time-varying γ)

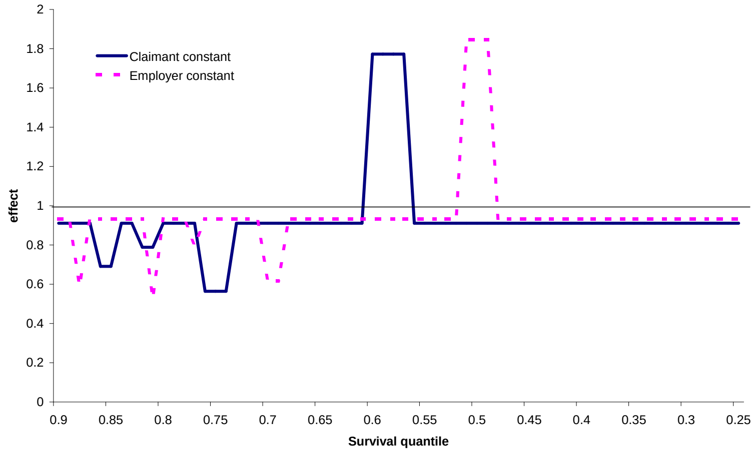

Figure 3: Effect of Bonus on quantiles of unemployment duration of BLACKS, $\frac{\mathrm{d}\ t_q(1)}{\mathrm{d}\ t_q(0)}$ (constant γ)

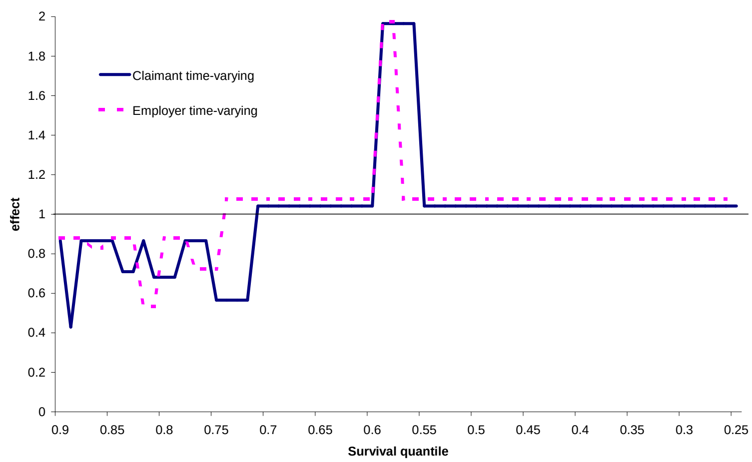

Figure 4: Effect of Bonus on quantiles of unemployment duration of BLACKS, $\frac{\mathrm{d}\ t_q(1)}{\mathrm{d}\ t_q(0)}$ (time-varying γ)

Table 2: Instrumental Variable Linear Rank estimates for the effect of the Bonus

Claimant group

	AFT	GAFT[a]	MLE	ITT
Constant effect				
First stage	0.1446	0.1024	-	-
	(0.0493)	(0.0523)	-	-
1–step optimal	0.1596	0.0932	0.1039	0.1117
	(0.0460)	(0.0380)	(0.0285)	(0.0303)
Time varying effect				
First stage				
0-10	0.2955	0.1433	-	-
	(0.0523)	(0.0907)	-	-
10+	-0.0720	0.0063	-	-
	(0.0608)	(0.0886)	-	-
1–step optimal				
0-10	0.3865	0.1439	0.1601	0.1516
	(0.0486)	(0.0578)	(0.0361)	(0.0378)
10+	-0.0437	-0.0411	-	-
	(0.0572)	(0.0850)	-	-

Employer group

	AFT	GAFT[a]	MLE	ITT
Constant effect				
First stage	0.1011	0.0721	-	-
	(0.0646)	(0.0470)	-	-
1–step optimal	0.1332	0.0696	0.0387	0.0516
	(0.0612)	(0.0425)	(0.0318)	(0.0307)
Time varying effect				
First stage				
0-10	0.2304	0.1103	-	-
	(0.0710)	(0.0736)	-	-
10+	-0.0783	-0.0048	-	-
	(0.0836)	(0.1253)	-	-
1–step optimal				
0-10	0.6334	0.1279	0.0881	0.0800
	(0.0674)	(0.0521)	(0.0402)	(0.0384)
10+	0.0330	-0.0747	-	-
	(0.0745)	(0.0882)	-	-

[a] GAFT piecewise constant intervals: 0–2, 2–4, 4–6, 6–10, 10–25, 25 →; *Notes:* Standard error in brackets.

Table 3: Effect of the Bonus on the length of unemployment duration

		Claimant		**Employer**	
		Constant	*Time-varying*	*Constant*	*Time-varying*
AFT	0-10	0.865	0.744	0.904	0.794
	10+	0.865	1.075	0.904	1.081
GAFT		*reference individual*			
80%	$t_q(0)$	3.9	3.7	2.8	4.3
	$t_q(1)$	3.5	2.9	2.5	3.7
	effect	**0.911**	**0.866**	**0.933**	**0.823**
60%	$t_q(0)$	12.8	12.6	8.9	12.7
	$t_q(1)$	10.4	9.4	7.8	10.0
	effect	**0.911**	**0.571**	**0.933**	**1.078**
40%	$t_q(0)$	25.7	25.7	20.7	24.3
	$t_q(1)$	22.8	23.1	18.3	22.5
	effect	**1.772**	**1.973**	**0.933**	**1.078**
GAFT		*black individual*			
80%	$t_q(0)$	7.5	6.8	4.8	8.1
	$t_q(1)$	6.5	5.3	4.1	6.4
	effect	**0.911**	**0.681**	**0.933**	**0.880**
60%	$t_q(0)$	25.3	24.4	18.44	24.22
	$t_q(1)$	22.1	21.0	16.2	22.5
	effect	**1.772**	**1.042**	**0.933**	**1.078**
40%	$t_q(0)$	35.6	35.1	30.7	34.2
	$t_q(1)$	32.9	33.8	28.90	33.9
	effect	**0.911**	**1.042**	**0.933**	**1.078**

Note that an effect smaller than one indicates that the bonus decreases the duration till re-employment and an effect bigger than one increases the duration. We see from the table (and more pronounced in Figure 1 and Figure 3) that even for a time-constant γ the effect of the bonus on the unemployment duration in the GAFT model changes with the duration. The huge spike in the effect at the survival quantile of 40% for the claimant group is because the re–employment rate exhibits a spike just before the time that unemployment benefits are exhausted, which is at 26 weeks. For the individuals in the control group the 40% survival time is just before 26 weeks, while in the claimant bonus group it is at 23 weeks. Thus the control group individuals are in the re-employment spike while the claimant bonus group are not. The interval boundaries of the other intervals of λ also cause, although not as pronounced, spikes. These spikes are downward because the λ is jumping to a lower level at these boundaries. The spikes are also visible in the effect of a time-varying coefficient of the bonus, see Figure 2 and Figure 4. Here, the change in γ at a duration of 10 weeks, after which the coefficient is negative, is reflected is a upward shift of the effect curve.

An indication that the AFT is not the right model is the difference between the first stage and one–step optimal estimators for the AFT model. For a correctly specified model both estimators are consistent and, therefore, do not differ much. In the GAFT model the first stage and one–step estimator are of the same magnitude. The estimated standard errors of the latter are, as expected, substantially lower in most situations.

Although the focus in this article is on the estimation of the effect of a possibly endogenous variable on the duration we also give a short discussion on the estimation results of the other parameters. These estimators can be found in the tables in appendix B. The regression parameters are overestimated (in absolute terms) if we assume an AFT model. These regression parameters hardly change from a model with constant bonus effect (Table 5) to a model with time–varying bonus effect (Table 6). The regression parameters for the Claimant data and the Employer data (both including the control group) are almost identical. Gender, MALE, is the exception; Gender has no significant influence on the re–employment probability in the Employer data. The shape of the estimated λ's indicate a U–shaped λ.

We end with a discussion on the selectivity in the bonus data. The compliance rate in the Claimant group, 85%, was much higher than the compliance rate in the employer group, 65%. Many individuals in the Employer group, apparently and contrary to our findings, did not perceive a bonus paid to their new employer beneficiary for their job search. Following Moffitt (1983) this partial compliance may be explained by a stigma effect. However, this is a tentative explanation because our analysis only adjust for (possible) selective compliance. It does not provide a model for the selection process. Thus, both an advantage and a drawback of our method is that we do not make any assumptions on the selection process and therefore cannot tell why individuals make such a selective decision.

7 Conclusion

In this article we proposed and implemented an instrumental variable estimation procedure for duration models. We show how the effect of an endogenous variable on the duration in a Generalized Accelerated Failure Time (GAFT) model can be estimated. The GAFT model is based on a transformation of the durations that encompasses both the Accelerated Failure Time (AFT), very popular in biostatistics, and the Mixed Proportional Hazards (MPH) model, very popular in econometrics. The interpretation of regression coefficients in the GAFT is in terms of shifting the quantiles of the distribution.

The Instrumental Variable Linear Rank (IVLR) estimation procedure is based on the inverse of an extended rank-test. It exploits the suggestion that for the population parameters the difference between the value of a weight function to the average of the weight functions for those individuals that are still under observation on the transformed duration goes to zero. This implies that for the population parameters the proportion of the individuals with a particular instrument value remains the same over the survivors on the transformed duration. The estimation procedure is related to quantile-regression, in particular Koenker and Bilias (2001) and Koenker and Geling (2001), and to Rank Preserving Structural Failure Time (RPSFT) Model estimator of Robins and Tsiatis (1991). The main difference with the RPSFT-model is that it assumes an AFT model while the IVLR allows for the more general GAFT-model, that includes duration dependence. The procedure is also related to the 2 stage Linear Rank (2SLR) estimation procedure of Bijwaard and Ridder (2005). The 2SLR assumes an MPH model and is a 2-steps procedure, while the IVLR is a one-step procedure.

We show that a counting process framework simplifies the derivation and interpretation of the IVLR. The counting process framework also enables us to derive the large sample properties of the IVLR. Because the IVLR is based on a vector of mean restrictions it is related to the well-known GMM estimation procedure. Similar to the application of GMM estimation choosing the right weight functions can improve the efficiency. However, again similar to the GMM, these optimal weight functions are not directly observable. Fortunately, an adaptive (or even 2 step) procedure can provide the efficient IVLR.

A difficulty with the proposed estimator is that it involves root-finding of a multi-dimensional step-function that is not differential. For this problem the Powell method provides a good algorithm to find the solution.

The empirical application shows that the ML and ITT estimates are downward biased due to endogeneity. Incorrectly assuming an AFT model can give misleading conclusions about the effects of a bonus on the re–employment hazard. In the Illinois bonus re-employment experiment many unemployed found a job just before their UI-benefits expires. This induces a spike in the re-employment hazard. In the GAFT, even with a constant regression coefficient, such a spike leads to an effect that changes over the quantiles. This has important implications for the evaluation of the effect of a possible endogenous variable on a duration.

Social experiments may provide instruments for an endogenous variable. With good instruments available the proposed method can be very useful in analyzing the effects of a possible endogenous variable on an inherently duration outcome. Examples in population studies include the effect of training programs on the unemployment duration, policies to increase the birth rate and migration policies.

There are several issues that need further research. First, the current approach to adjust for endogenous censoring implies loss of information and depends on the (unknown) parameters of the model. An important improvement would be to find a method to adjust for endogenous censoring that is parameter independent and minimizes the loss of information. Another related issue is that if the IVLR assumes that the censoring time is (potentially) known in advance. Further research on more general censoring patterns deserve attention. Second, in our empirical application we have, because of random assignment, a perfect assignment. Such an instrument is, however, not always available. Finding good instruments is therefore an important issue just as the influence of weak instruments on the properties of the estimator. A final issue for further research is the extension of the IVLR to recurrent duration data, like repeated unemployment spells.

References

Abbring, J. H. and G. J. van den Berg (2005). Social experiments and instrumental variables with duration outcomes. Tinbergen Institute, discussion paper, TI 2005–47.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer–Verlag.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Volume 3A*, Chapter 23, pp. 1277–1366. Amsterdam: North–Holland.

Ashenfelter, O., D. Ashmore, and O. Deschêne (2005). Do unemployment insurance recipients actively seek work? Evidence from randomized trials in four U.S. states. *Journal of Econometrics 125*, 53–75.

Bijwaard, G. E. and G. Ridder (2005). Correcting for selective compliance in a re–employment bonus experiment. *Journal of Econometrics 125*, 77–111.

Brännäs, K. (1992). *Econometrics of the Accelerated Duration Model*. Umeå: Solfjädern Offset AB.

Cox, D. R. and D. Oakes (1984). *Analysis of Survival Data*. London: Chapman and Hall.

Han, A. K. (1987). Non–parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics 35*, 303–316.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics, Volume 3A*, Chapter 31, pp. 1866–2097. Amsterdam: North–Holland.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data (second edition)*. John Wiley and Sons.

Klein, J. P. and M. L. Moeschberger (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer–Verlag.

Koenker, R. and Y. Bilias (2001). Quantile regression for duration data: A reappraisal of the Pennsylvania reemployment bonus experiments. *Empirical Economics 26*, 199–220.

Koenker, R. and O. Geling (2001). Reappraising medfly longevity: A quantile regression survival analysis. *Journal of the American Statistical Association 96*, 458–468.

Meyer, B. D. (1995). Lessons from the U.S. unemeployment insurance experiments. *Journal of Economic Literature 33*, 91–131.

Meyer, B. D. (1996). What have we learned from the Illinois reemployment bonus experiment? *Journal of Labor Economics 14*, 26–51.

Moffitt, R. (1983). An economic model of welfare stigma. *American Economic Review 73*, 1023–1035.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal 7*, 155–162.

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika 65*, 167–179.

Press, W. H., B. P. Flannert, S. A. Teukolsky, and W. T. Vetterling (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge UP.

Ridder, G. (1990). The non–parametric identification of generalized accelerated failure–time models. *Review of Economic Studies 57*, 167–182.

Robins, J. M. and A. A. Tsiatis (1991). Correcting for non–compliance in randomized trials using rank-preserving structural failure time models. *Communications in Statistics Part A: Theory and Methods 20*, 2609–2631.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics 18*, 354–372.

Van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple duration. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume V*, Chapter 55, pp. 3381–3460. Amsterdam: North–Holland.

Van den Berg, G. J., J. C. van Ours, and B. van der Klaauw (2004). Punitive sanctions and the transition rate from welfare to work. *Journal of Labor Economics 22*, 211–241.

Woodbury, S. A. and R. G. Spiegelman (1987). Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois. *American Economic Review 77*, 513–530.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics 21*, 76–99.

A Asymptotic properties of the IVLR

Assumptions:

C1: The covariate process $X(t)$ is predetermined, i.e. its distribution is independent of $\{H(s), s > t\}$. The sample paths of the covariate process are bounded and at least one of time–varying covariates is a continuous variable.

C2: The observation process $Y(t)$ is cadlag and $Y(t)$ is predetermined. Moreover,

$$\Pr\big(\mathrm{d}N(t) = 1 \mid Y(t) = 1, H(t)\big) = \Pr\big(\mathrm{d}N(t) = 0 \mid Y(t) = 0, H(t)\big)$$

C3: The population distribution of T given $\overline{X}$ and $\overline{D}$ satisfies

$$\int_0^T \lambda(s;\alpha_0) e^{\beta_0' X(s) + \psi(s,D;\gamma_0)}\, \mathrm{d}s = U_0$$

The absolutely continuous distribution of U_0 does not depend on $\overline{X}$ or $\overline{R}$. The p.d.f. of U_0 is bounded.

C4: The transformed observation process $Y^U(u) = I\big(\tilde{U}(\theta) \geq u\big)$ is cadlag and predetermined, with $\tilde{U}(\theta) = \min\big(U(\theta), C^U\big)$ and C^U defined in (21).

C5: The instrumental function W is bounded and left–continuous.

C6: The intensity of $U(\theta)$, $\kappa_i^U(u)$ given history $H^{UR}(u)$ in (19) can be linearized in a neighborhood of θ_0 as a function of θ, *i.e.* there exist $\mu(u)$ and $\epsilon > 0$ such that for $\|\theta - \theta_0\| < \epsilon$

$$\left|\kappa_i^U(u;\theta) - \kappa_0(u) - (\theta - \theta_0)' d_{i0}(u)\right| \leq \|\theta - \theta_0\|^2 \mu(u)$$

for $u \leq C_0 = C^U(\theta_0)$ with

$$d_{i0}(u) = \frac{\partial \kappa_i^U(u;\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}$$

C7: There exists a continuous function $a(u;\theta)$ of θ in a neighborhood B of θ_0 such that

$$\sup_{u \leq C_0} \sup_{\theta \in B} \left\|\overline{W}(u;\theta) - a(u;\theta)\right\| \xrightarrow{p} 0$$

where

$$\overline{W}(u;\theta) = \frac{\sum_{j=1}^n Y_j^U(u) W\big(u, \overline{X}_j^U(u), R_j\big)}{\sum_{j=1}^n Y_j^U(u)}$$

C8: There exists a continuous matrix function $A(u;\theta)$ of θ in a neighborhood B of θ_0 such that

$$\sup_{u \le C_0} \sup_{\theta \in B} \Bigg\| \frac{1}{n} \sum_{i=1}^n \Big[W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big] \\ \times \Big[W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big]' Y_i^U(u) - A(u;\theta) \Bigg\| \xrightarrow{p} 0$$

C9: There exists a continuous matrix-function $V(u;\theta)$ of θ in a neighborhood B of θ_0 such that

$$\sup_{u \le C_0} \sup_{\theta \in B} \Bigg\| \frac{1}{n} \sum_{i=1}^n \Big[W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big] \times d_{i0}(u)' Y_i^U(u) - V(u;\theta) \Bigg\| \xrightarrow{p} 0$$

The starting point is (24), which can, for θ in a small neighborhood of θ_0, be rewritten as

$$S_n(\theta;W) = \sum_{i=1}^n \int_0^{C_{i0}} \Big\{ W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big\} \,\mathrm{d}N_i^U(u) \\ + \sum_{i=1}^n \int_{C_i^U}^{C_{i0}} \Big\{ W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big\} \,\mathrm{d}N_i^U(u) \quad (39)$$

Substitution of the Doob–Meier composition in the first term on the right for N_i^U gives

$$S_n(\theta;W) = \sum_{i=1}^n \int_0^{C_{i0}} \Big\{ W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big\} \, dM_i^U(u) \\ + \sum_{i=1}^n \int_0^{C_{i0}} \Big\{ W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big\} \kappa_i^U(u) Y_i^U(u) \,\mathrm{d}du \quad (40)$$

We consider both terms separately. The first term is, for θ close to θ_0, close to $S_n(\theta_0;W)$ and for the second term we have

$$(\theta - \theta_0) \cdot \sum_{i=1}^n \int_0^{C_{i0}} \Big\{ W\big(u, \overline{X}_i^U(u), R_i\big) - \overline{W}(u;\theta) \Big\} \times \frac{\partial \kappa_i^U(u)'}{\partial \theta} Y_i^U(u) \,\mathrm{d}u + O_p\big(\|\theta - \theta_0\|^2\big)$$

Returning to (39) we note that the second term in this equation equals

$$\sum_{i=1}^{n}\left\{\left[W\Big(C_{i0},\overline{X}_{i0}(C_{i0}),R_i\Big)-\overline{W}\big(C_{i0};\theta_0\big)\right]\times\theta_0(C_{i0})Y_i(C_{i0})\right\}+O_p\big(\|\theta-\theta_0\|^2\big)$$

The term between brackets is the covariance between $\theta_0(C_{i0})$ and $W\big(C_{i0},\overline{X}_{i0}(C_{i0}),R_i\big)$ which is zero. Thus this whole term is zero for θ close to θ_0 and we have

$$S_n(\theta;W)\approx S_n(\theta_0;W)+n\int_0^{C_0}Z(u;\theta_0)\,du\cdot(\theta-\theta_0) \tag{41}$$

Hence, approximately for the IVLR estimator $\hat{\theta}_n(W)$

$$\sqrt{n}(\hat{\theta}_n(W)-\theta_0)=\left[\int_0^{C_0}Z(u;\theta_0)\,du\right]^{-1}\frac{1}{\sqrt{n}}S_n(\theta_0;W) \tag{42}$$

The proof of the consistency and asymptotic normality are both based upon the asymptotic linearity of $S_n(\theta;W)$ in the neighborhood of the true value θ_0. We follow the reasoning of Tsiatis (1990). Instead of a mean and variance condition, we have a mean and three covariance conditions. Let $\tilde{S}_n(\theta;W)$ be the right-hand side of (41). The following lemma shows that the linearization in (41) is uniformly close to the original estimating function

Lemma 1. *In neighbourhoods of $O(n^{-1/2})$ of θ_0*

$$n^{-1/2}\left\|\tilde{S}_n(\theta;W)-S_n(\theta;W)\right\|$$

converges uniformly to zero.

This lemma implies that $n^{-1/2}\tilde{S}_n(\theta;W)$ and $n^{-1/2}S_n(\theta;W)$ are asymptotically equivalent in a neighbourhood close to θ_0.

Proof: This can be proved in lines of Tsiatis (1990) Lemma (3.1) and (3.2) and theorem (3.2) and this is, because of the analogy, not repeated here.

Proof of theorem 1 and theorem 2. According to lemma 1 are $n^{-1/2}S_n(\theta;W)$ in a neighbourhood close to θ_0 asymptotically equivalent to $n^{-1/2}\tilde{S}_n(\theta;W)$. Then the estimates θ^* and $\hat{\theta}$, with $\tilde{S}_n(\theta^*;W)=0$, will also be asymptotically equivalent. Clearly, θ^* converges in probability to θ_0. Hence, if we show that $\sqrt{n}(\hat{\theta}-\theta^*)\xrightarrow{p}0$ then this would imply that

$\hat{\theta}$ also converges in probability to θ_0. Tsiatis (1990) argues that lemma 1 suffices to proof this. This proves theorem 1.

According to the Mann–Wald theorem convergence in probability implies convergence in distribution. We note that $\sqrt{n}(\theta^* - \theta_0) = n^{-1/2}Q^{-1}(W)S_n(\theta_0; W)$ clearly converges to a normal distribution with mean zero and variance matrix $Q^{-1}(W)\Omega(W)Q'^{-1}(W)$. This completes the proof of theorem 2. □

Remark. To establish detailed conditions on when $\tilde{S}_n(\theta; W)$ has a unique root is rather tedious; however Ying (1993) gave an excellent general treatment on rank estimation, which can also be used for the estimating equations in this article.

B Additional tables for the IVLR of reemployment bonus experiment

Table 4: Descriptive statistics for Control,Claimant Bonus and Employer Bonus group.

	Control Group	Claimant bonus	Employer bonus
White	0.632	0.651	0.647
Black	0.271	0.251	0.256
Other	0.097	0.099	0.097
Male	0.547	0.563	0.538
Age 20–29	0.425	0.436	0.424
Age 30–39	0.333	0.324	0.326
Age 40–49	0.179	0.185	0.187
Age 50–54	0.063	0.054	0.064
Weekly benefit			
-$51	0.088	0.085	0.084
$52–$90	0.201	0.212	0.217
$91–$120	0.169	0.176	0.179
$121–$160	0.190	0.196	0.181
$161–	0.353	0.331	0.339
Dependence allowance	0.323	0.345	0.332
Average pre–claim earnings	3188	3222	3215
Average age	33.0	32.9	33.1
Average weekly	119.9	118.8	118.5
N	3952	4186	3963

Table 5: Instrumental Variable Linear Rank estimates for the regression coefficients of the Illinois data (Constant Bonus Effect)

First stage	*Claimant*		*Employer*	
	AFT	GAFT[a]	AFT	GAFT[a]
LNAGE	-0.5718	-0.3424	-0.5219	-0.3379
	(0.0734)	(0.0897)	(0.0717)	(0.0699)
LNBPE	0.3528	0.2146	0.3188	0.2036
	(0.0510)	(0.0601)	(0.0512)	(0.0482)
BLACK	-0.6636	-0.3770	-0.6264	-0.3792
	(0.0526)	(0.0842)	(0.0510)	(0.0641)
MALE	0.1135	0.0663	0.0464	0.0295
	(0.0377)	(0.0330)	(0.0376)	(0.0305)
LNBEN	-0.5841	-0.3558	-0.6263	-0.4010
	(0.0867)	(0.1011)	(0.0871)	(0.0865)
One step Optimal				
	Claimant		*Employer*	
	AFT	GAFT[a]	AFT	GAFT[a]
LNAGE	-0.5204	-0.3612	-0.4733	-0.3110
	(0.0693)	(0.0653)	(0.0683)	(0.0603)
LNBPE	0.3537	0.2266	0.3133	0.1871
	(0.0473)	(0.0449)	(0.0483)	(0.0424)
BLACK	-0.6162	-0.3982	-0.5646	-0.3574
	(0.0509)	(0.0510)	(0.0495)	(0.0443)
MALE	0.1293	0.0691	0.0698	0.0227
	(0.0355)	(0.0303)	(0.0355)	(0.0303)
LNBEN	-0.5924	-0.3692	-0.6040	-0.3610
	(0.0813)	(0.0762)	(0.0826)	(0.0727)

[a] GAFT piecewise constant intervals: 0–2, 2–4, 4–6, 6–10, 10–25, 25 →; *Notes:* Standard error in brackets.

Table 6: Instrumental Variable Linear Rank estimates for the regression coefficients of the Illinois data (Time–varying Bonus effect)

First stage				
	Claimant		*Employer*	
	AFT	GAFT[a]	AFT	GAFT[a]
LNAGE	-0.5361	-0.3285	-0.5233	-0.3355
	(0.0693)	(0.0897)	(0.0706)	(0.0763)
LNBPE	0.3313	0.2139	0.3153	0.2029
	(0.0481)	(0.0617)	(0.0506)	(0.0530)
BLACK	-0.6086	-0.3665	-0.6268	-0.3771
	(0.0494)	(0.0861)	(0.0501)	(0.0740)
MALE	0.1036	0.0668	0.0461	0.0294
	(0.0352)	(0.0337)	(0.0371)	(0.0304)
LNBEN	-0.5470	-0.3564	-0.6187	-0.3989
	(0.0820)	(0.1043)	(0.0859)	(0.0959)
One step Optimal				
	Claimant		*Employer*	
	AFT	GAFT[a]	AFT	GAFT[a]
LNAGE	-0.4861	-0.3288	-0.4529	-0.3660
	(0.0653)	(0.0664)	(0.0675)	(0.0622)
BPE	0.3332	0.2061	0.3017	0.2236
	(0.0442)	(0.0455)	(0.0474)	(0.0434)
BLACK	-0.5644	-0.3615	-0.5286	-0.4189
	(0.0476)	(0.0533)	(0.0488)	(0.0480)
MALE	0.1176	0.0626	0.0622	0.0283
	(0.0332)	(0.0304)	(0.0349)	(0.0302)
LNBEN	-0.5501	-0.3343	-0.5813	-0.4284
	(0.0765)	(0.0770)	(0.0815)	(0.0752)

[a] GAFT piecewise constant intervals: 0–2, 2–4, 4–6, 6–10, 10–25, 25 →; *Notes:* Standard error in brackets.

Table 7: Estimated λ in GAFT model for the Bonus data

Claimant				
	Constant Bonus effect		*Time varying Bonus effect*	
interval	first	opt.	first	opt.
0–2	0.8098	0.7500	0.8625	0.9328
	(0.4638)	(0.2052)	(0.5262)	(0.2409)
2–4	0.3146	0.2348	0.3542	0.2309
	(0.3691)	(0.1462)	(0.4048)	(0.1799)
4–6	-0.0782	-0.0415	-0.0390	0.0318
	(0.2646)	(0.1220)	(0.3015)	(0.1552)
6–10	-0.2743	-0.1859	-0.2341	-0.2085
	(0.2392)	(0.1133)	(0.2807)	(0.1369)
10–25	-0.6868	-0.6655	-0.6077	-0.6345
	(0.1626)	(0.1006)	(0.1758)	(0.1261)
Employer				
	Constant Bonus effect		*Time varying Bonus effect*	
interval	first	opt.	first	opt.
0–2	0.7095	0.8929	0.7088	0.5647
	(0.3063)	(0.1450)	(0.4375)	(0.1716)
2–4	0.2540	0.4451	0.2542	0.1464
	(0.2134)	(0.0939)	(0.3344)	(0.1227)
4–6	-0.1217	-0.1178	-0.1195	0.0875
	(0.2008)	(0.0925)	(0.2330)	(0.1050)
6–10	-0.4552	-0.2707	-0.4526	-0.4098
	(0.1516)	(0.0751)	(0.2255)	(0.0975)
10–25	-0.7492	-0.6826	-0.7180	-0.6057
	(0.0971)	(0.0372)	(0.1015)	(0.0491)

Notes: Standard error in brackets.