



Dentifications of the Distribution of Treatment Effect with Duration Outcomes

David Koch, Steve Ofili and Abel Cudjoe
David Koch, Economic Development Department

Abstract:

The most critical factor in econometric estimations is parameter identification. Identification in econometric models formalizes prior assumptions and the data to information about a parameter of interest. However, there are two important features characterize duration data. The first one is that the data may be censored, and the second feature of duration data is that exogenous determinants of the event times characterizing the data may change during the event spell. The two characteristics lead some famous identification problems for the duration models. Following the recent literature in partial identification, we show the conditions when the duration models could be identified and provide several suggestions for the confidence bounds of partial identifications.

Keywords: Competing risk; Nonparametric; Confidence bounds.

JEL classification: C24; C41

1. Introduction:

There are several problems which researchers have to face, when one considers the treatment effect of a social experiment on duration outcomes. First, the randomization selection usually is not easy to satisfy. Even with randomized assignments, participants may not comply completely, which induces a selection bias. Second, the composition of survivors changes over time in different ways in the treatment and control groups, even though the selection is random, and this makes the analysis of individual hazard rate different from other treatment effect models. Third, the response to the social experiment may not be immediate, which induces the choice of an instrumental variable more restrictive. With later compliance, even though one can find a plausible instrumental variable which satisfies the condition of ex post exclusion restriction, it is almost impossible for researchers to non-parametrically identify the condition of ex ante exclusion restriction (Abbring and van den Berg, 2005), which means to test the efficiency of such an instrumental variable is almost intractable. To simplify the analysis, here we ignore the problem of later compliance and assume all participants in a social experiment will respond immediately to the assignment, mainly focus on the effect of self-selection.

To analyze the effect of self-selection, one of the classical methods is appealing to some instrumental variables. In the classical switching regime model, we usually require the number of variables in the decision function should be more than the outcome functions, since the identification of the effect on hazard rates would require additional information (see Heckman, Smith and Taber, 1998, for a discussion). However, to find an instrumental variable usually is not easy and in duration outcome model it is also almost impossible us to test its condition of exclusion restriction. As mentioned by Abbring and Berg, with duration outcome, even if we can find a valid instrumental variable, we still cannot non-parametrically identify the average treatment effects on individual hazard rates and have to use a semi-parametric structure. In this paper, we drop out the IV approach but also adopt the popular assumption of mixed proportional hazards (MPH). In the model, the hazard rate is written as a multiplicative function of observed explanatory variables x , the elapsed duration t , and a random term v representing unobserved explanatory variables. Specifically,

$$\theta(t | v, x) = \lambda(t)\phi(x)v \quad (1)$$

is the hazard rate of $t|x, v$. It is related to the distribution function $F(\cdot)$ of $t|x, v$ by $\theta(t|x, v) = -d \log(1 - F(t|x, v)) / dt$. The distribution of $t|x$ follows by integration of $F(t|x, v)$ w.r.t. the density of v . (See Lancaster, 1990 and Jenkins, 1995 for extensive surveys). The function $\lambda(t)$ is called the baseline hazard. Usually, $\phi(x)$ is specified as $\exp(x'\beta)$.

To analyze the treatment effect, Guo (2017) creatively introduce the potential outcome framework. Guo's seminal work assumes two different treatment outcomes. Our discussions heavily depend on Guo's results. With the sights from Shintani and Guo (2016), in this paper we make use of the potential outcome framework. In the potential outcome framework, the error terms in treatment equation and control equation are different but correlated. Without enforcing some additional conditions, it is impossible to identify the full joint distribution of them, since we cannot observe the potential outcomes jointly (Heckman 1990). However, regarding duration outcome, researchers usually assume that the error terms in treatment equation and control equation are the same, and the treatment effect differs from a parameter (e.g. Ham and LaLonde 1996; Abbring and van den Berg 2005). Obviously, this is a very strong assumption¹.

The identification of the hazard function of duration outcome has been studied by a lot of researchers. Elbers and Ridder (1982) consider a proportional hazard model. In their paper, they consider a single spell model and assume the hazard function to be $\theta(t|v, x) = \lambda(t)\phi(x)v$ where x can be interpreted as the observed and v as the unobserved heterogeneity in the sample. Since we can only identify the function $G(t, x) = 1 - \int_0^\infty e^{-m(t)\phi(x)v} f(v)dv$, with $m(t) = \int_0^t \lambda(u)du$, it is impossible for us to estimate $\lambda(t)$ and $\phi(x)$ directly. However, the authors prove that with some conditions, we can identify both $\lambda(t)$ and $\phi(x)$, and even the distribution of the unobserved heterogeneity $f(v)$. With self selection, the distribution function of the unobserved heterogeneity $f(v)$ will be distorted, since v is correlated with the decision rule, which means the participants with different v have different probabilities to be observed. In this paper, at first we will demonstrate that although $f(v)$ cannot be identified, but we still can identify $f(v_s | S)$.

The joint distribution of the error terms in the treatment outcome equation and control outcome equation has been studied by some econometricians (e.g. Cunha and Heckman, 2008; Wu and Fan, 2010; Park and Fan, 2012; van den Berg and Drepper, 2016; Reza and Rilstone, 2016), but for duration outcome it has been little studied. One reason may be the faintness of its identification, and the other reason may be that the estimation of the error terms still be ambiguous. Van den Berg (1997) studies the association measures for duration outcomes in bivariate hazard rate models, but Berg did not specify his analysis in the potential-outcome framework. In his paper, the unobserved heterogeneities from different duration variables may be dependent, so the duration variables may be dependent. As common assumptions, he firstly also assumes that the marginal hazard for both of two variables is the mixed proportional hazard, so the hazard rate $\theta_s(t_s | v_s, x) = \lambda_s(t_s)\phi_s(x)v_s$, $s = 1, 2$ for group 1 and 2 respectively. Then Berg specifies the baseline hazard as the so-called Weibull specification, i.e. $\lambda_i(t_i) = \alpha_i t_i^{\alpha_i - 1}$. As mentioned before, the dependence of different duration variables conditioning on observed variables comes from the unobserved heterogeneity terms v_s . With different joint distributions, the correlation of the two duration outcome variables will be different. Van den Berg then deduces the sharp bound for the correlation:

$$\frac{-1}{\sqrt{a_1 a_2} + \sqrt{(a_1 - 1)(a_2 - 1)}} < CORR(t_1, t_2 | x) < \frac{1}{\sqrt{a_1 a_2}} \quad (2)$$

Regardless of the values of $\theta_s(x)$, and also regardless of the shape of $G(v_1, v_2)$.

Here, we consider a social experiment. In the experiment, people are randomly selected into the treatment group, and the compliance will be immediate but incomplete. We assume the hazard rate functions of both the treatment group and the control group satisfy mixed proportional hazard assumption. The unobserved variables for treatment duration outcome

¹Heckman (1990) mentioned that some studies indicate that unobserved (by the econometricians) components of $v_1 - v_0$ contribute negligibly to the endogeneity of S , so the assumption of $E(v_1 - v_0 | S = 1, x) = 0$ can be held in some circumstance. We are not sure in duration outcome model, the assumption $v_1 = v_0$ plays how big role it does.

and control duration outcome are denoted by v_1 and v_2 , respectively. Intuitively, if one can estimate their marginal distribution and joint distribution, there are a lot of analyses will be quite straightforward, such as measurement inequality. The main purpose of this paper is trying to find a method to non-parametrically or semi-parametrically identify the distribution of the error terms v_1 and v_2 , and also their joint distribution.

The paper is divided into the following sections. The section two will present a model and give the identification of single duration outcome. Then, in section three we will talk about the estimation of the decision rule function. Section four is the core of the paper and will mainly focus on the estimation of the error term. Finally, we will compare this model to the competing risk model and give some directions of further research for both of them.

2. Setup of the Model and Identification of Single Duration Outcome

Following Abbring and Van den Berg (2005), in a social experiment the treatment is randomly assigned to agents, and the compliance is immediate but not complete. We are interested in the causal effect of the treatment on the duration spent in the state of interest (outcome). The framework I will use is the prevalent potential-outcome one. It defines causal effects in terms of potential outcomes or counterfactuals rather than in terms of the parameters of a regression model. Let $T_i(0)$ be the duration outcome without the treatment for individual i . $T_i(1)$ is the duration outcome with treatment. We assume that $\{T_i(s)\} := \{T_i(s) : s \in \{0, 1\}\}$ is a measurable stochastic process. Moreover, for ease of exposition we assume that each $T_i(s)$ is continuous, and denote the hazard rate of $T_i(s)$ for a specific individual by $\theta_{is}(t | x_i, s)$. The hazard rate for treatment state is as follow:

$$\theta_{i1}(t | x_i, s = 1) = \lambda_1'(t) \phi_1(x_i) v_{i1} \quad (3)$$

The hazard rate for control state:

$$\theta_{i0}(t | x_i, s = 0) = \lambda_0'(t) \phi_0(x_i) v_{i0} \quad (4)$$

where t denotes the time elapsed, x_i is a vector of explanatory variables which are observed by econometricians, such as education background, personal property and age in an unemployment training program, and v_{is} is a positive multiplicative disturbance, or unobservable heterogeneity. This specification of the hazard rate is so called mixed proportional hazard rate (MPH), which frequently has shown in duration outcome literature. The selection equation is as follow:

$$s_i^* = z_i' \gamma + \varepsilon_i \quad (5)$$

With $s_i := I(s_i^* > 0)$. As the classical potential-outcome setup, if there exists self selection bias, the error terms ε_i and v_{is} should not be independent.

To identify the functions $\lambda_s'(t)$ and $\phi_s(x_i)$, we follow Elbers and Ridder's approach (1982) and mainly focus the treatment group. The identification of the control group is quite similar and can use the same approach. Suppose the marginal distributions for v_{i1} and v_{i0} follow $f_{v_1}(v_1)$ and $f_{v_0}(v_0)$ respectively. The duration distribution for the observed treatment group is:

$$G_1(t, x_i | s = 1) = 1 - \int_0^t e^{-m_1(u) \phi_1(x_i) v_{i1}} f_{v_1}(v_{i1} | s = 1) dv_{i1} \quad (6)$$

with $m_1(t) = \int_0^t \lambda_1(u) du$. I make the following assumptions.

Assumption 1: The random variable v_{i1} is non-negative with distribution function $F_{v_1}(\cdot)$, and its expectation is $E(v_{i1}) = 1$.

Assumption 2: The function $m_1(\cdot)$, defined on $[0, \infty)$ can be written as the integral of a non-negative integrable function $\lambda_1(\cdot)$ which is defined on $[0, \infty)$, i.e.

$$m_1(t) = \int_0^t \lambda_1(u) du \quad t \geq 0 \quad (7)$$

Assumption 3: The set X is an open set in R^k . The function $\phi_1(\cdot)$ is defined on X and non-negative, differentiable and non-constant on X .

Theorem: Let $\{G_1(t, x_i); x_i \in X\}$ be a family of non-defective distribution functions which satisfy Equation (6). Then under the Assumptions 1, 2, 3 this representation is unique i.e. the functions $m_1(\cdot)$ and $\phi_1(\cdot)$ and the distribution of v_{i1} conditioning on selection $s = 1$ are uniquely determined ($\phi_1(\cdot)$ uniquely up to a constant).

With self selection, even if one assumes $E(v_{i1}) = 1$, $E(v_{i1} | s = 1) = 1$ may not hold. The cumulative distribution function one can identify directly becomes $\{G_1(t, x_i | s = 1); x_i \in X\}$. The process of the identification is as follows. Differentiating (6) with respect to t gives

$$g_1(t, x_i | s = 1) = \lambda_1(t) \phi_1(x_i) \int_0^\infty e^{-m_1(t)\phi_1(x_i)v_{i1}} f(v_{i1} | s = 1) dv_{i1}, \quad x \in X, t > 0. \quad (8)$$

Even without $E(v_{i1} | s = 1) = 1$, dividing by $g_1(t, x_c | s = 1)$ and let $t \rightarrow 0$, one still can find

$$\phi_1(x) = \phi_1(x_c) \lim_{t \rightarrow 0} \frac{g_1(t, x | s = 1)}{g_1(t, x_c | s = 1)} \quad (9)$$

Equation (9) determines $\phi_1(x)$ up to a constant. This result is quite intuitive. When the time close to the starting point, one can view the composition of the treatment group has not changed, so the hazard rate will equal to the survival probability approximately. Since v_{i1} is assumed to be positive, (6) can be written as

$$G_1(t, x | s = 1) = 1 - \zeta_{v_1} (m_1(t)\phi_1(x)) \quad t \geq 0 \quad (10)$$

Where ζ_{v_1} is the Laplace-transform of v_1 which is defined as

$$\zeta_{v_1}(y) = \int_0^\infty e^{-yv_1} dF(v_1 | s = 1) \quad y \geq 0 \quad (11)$$

Further, from equation (11), $z_1(t)$ can be written as

$$m_1(t) = \frac{H_1(1 - G_1(t, x | s = 1))}{\phi_1(x)} \quad t \geq 0, x \in X \quad (12)$$

where $H_1 = \zeta_{v_1}^{-1}$, the inverse of the Laplace-transform. Note that (12) implies that the right-hand side is a function of t alone and independent of x . Define $T_1 = 1 - G_1$, and take derivative of equation (12) with respect to x

$$\phi_1(x) H_1'(T_1(t, x | s = 1)) \frac{\partial(T_1(t, x | s = 1))}{\partial x} - H_1(T_1(t, x | s = 1)) \phi_1'(x) = 0 \quad t \geq 0 \quad (13)$$

By a change of variables

$$y = T_1(t, x | s = 1) \Leftrightarrow t = K_1(y, x) \quad (14)$$

with $t \geq 0$ and $0 \leq y < 1$, equation (13) can be written as

$$\phi_1(x) H_1'(y) \frac{\partial T_1(K_1(y, x), x)}{\partial x} - H_1(y) \phi_1'(x) = 0 \quad 0 \leq y < 1. \quad (15)$$

This differential equation can be solved for H . The solution is:

$$H_1(y) = c \exp\left\{\frac{d \ln \phi_1(x)}{dx} \int_{1/2}^s \frac{1}{\frac{\partial T_1}{\partial x}(K_1(u, x), x)} du\right\} \quad 0 \leq y < 1. \quad (16)$$

Next step we should prove the solution is unique. The process is the same as Elbers and Ridder's except we should make use of the condition of $f_{v_1}(v_1 | s = 1)$ fixed rather than the condition $E(v_s) = 1$. Here I skip this part. If one is interested in this part, please refer to Elbers and Ridder.

3. The Estimation of Decision Rules

Agents participate in the treatment program based on their selection. The selection indicator is a binary variable $s_i = I\{s_i^* \geq 0\}$, and s^* follows equation (5): $s_i^* = z_i' \gamma + \varepsilon_i$. The estimation of γ and the distribution of ε_i has been considered by several researchers (Manski 1988, Ichimura 1987, Powell, Stock and Stoker 1989, etc). Intuitively, one can easily estimate the propensity score conditioning on z_i . After estimating the propensity score, we can write the equation

$$\Pr(s = 1 | z_i) = E(s | z_i) = F_\varepsilon(z_i' \gamma) \quad (17)$$

Where F_ε is the assumed cumulative distribution function of $-\varepsilon$. As we can see, this is a typical single-index model. Through an estimation of a density-weighted average derivative of a general regression function, Powell, Stock and Stoker proposed an estimator $\hat{\gamma}$ of the coefficients γ and argue that it is root- N -consistent and asymptotically normal, even without assuming the distribution of the error terms. With known $\hat{\gamma}$, then one can trace out the cumulative density function by kernel estimation method.

4. The Marginal Distribution of the Error Terms

From section 2, it is possible for us to identify the functions $\phi_s(x)$ and $\lambda_s(t)$, and the conditional density function $f(v_s | s = 1)$, so one can identify the treatment effect for hazard rate individually. However, since the composition of the treated group changes over time, the distribution of error term v_{i1} may be interesting. Once the distribution of the error term is estimated, the quantile treatment effect or the measurement of inequality treatment can be analyzed straightforwardly. Since $f(v_s | s = 1)$ can be denoted as

$$f(v_s | s = 1) = \frac{f(v_s, s^* \geq -z' \gamma)}{\Pr(s^* \geq -z' \gamma)} \quad (18)$$

As we can see, if we take the limitation value of $\Pr(s^* \geq -z' \gamma)$ and consider the circumstance when it goes to very close to 1, the distribution function $f(v_s)$ can be traced out. Using similar approach by Heckman (1990), the marginal distributions of (v_0, v_1, ε) which can be identified are $f(v_0, \varepsilon)$, $f(v_1, \varepsilon)$, $f(v_0)$, $f(v_1)$ and $f(\varepsilon)$ ².

Since one cannot observed the residual pair (v_0, v_1) , it is impossible for us to estimate the joint distribution $f(v_0, v_1)$.

With known $\hat{f}(v_0, \varepsilon)$ and $\hat{f}(v_1, \varepsilon)$, to estimate $f(v_0, v_1)$, there are basically two directions. First, one may abandon point identification and estimation, and focus on bounding the parameters of interest (e.g. Wu and Fan (2008), Park and Fan (2008)). Second, one may impose a structure that is sufficiently specific to enable point identification and estimation of relevant treatment effects. For example, Cunha and Heckman (2006) used some semi-parametric structure and deduced the joint distribution of the residual terms.

²To interpret the estimation of $f(v_0, \varepsilon)$ and $f(v_1, \varepsilon)$, one can think in this way. Since the function $\theta_s(x)\lambda_s(t)$ can be estimated, the residual v_s can be estimated. Therefore, one can get a censored random residual pair (v_s, ε) , and their marginal distribution can be estimated using kernel estimation approach.

5. A Comparison with Competing Risk Model and Further Research

Most of the previous research starts from the hazard rate and mainly compare the difference of the hazard rates between treatment group and control group. However, one has to face the problem of changing composition in the corresponding groups if one uses the hazard rate approach. In this section, we will mainly use the probability approach.

Intuitively, the potential-outcome framework in duration also can be viewed as a competing risk one. Suppose there are variables which are observed by econometricians but fully observed by participants themselves. The participants compare the potential outcomes in different groups and rationally choose the best group to participate. Heckman and Honore (1989) studied the identification of the competing risks model. Through using proportional hazard model, Heckman and Honore argue that the joint distribution of failure times can be fully identified without invoking further assumption and the model will be fully identified. As the majority of the literature in duration outcome, the authors also assume the unobservable terms are the same for different duration outcomes. As mentioned before, this obviously is a very strong assumption.

In the model, each individual faces two competing treatment effects and chooses the best one between them. Associated with each treatment s , $s \in \{1, 2\}$ there is a potential duration outcome, T_s . The observed quantities are the duration to the minimum duration and the associated treatment assignment

$$(T, S) = \left\{ \min_s(T_s), \arg \min_s(T_s) \right\} \quad (19)$$

The assumptions of the hazard rate functions are the same as above. Following Heckman and Honore's approach, the dependence among potential duration is introduced in the following way. Suppose the survival functions for both treatments are:

$$S_s(t, x) = \int_0^\infty e^{-m_s(t)\phi_s(x)v_s} f_{v_s}(v_s) dv_s \quad (20)$$

with $m_s(t) = \int_0^t \lambda_s(u) du$. Since the potential duration outcomes are not independent, one can generate two dependent uniform distribution variables, U_1 and U_2 , with $S_s(t, x) = U_s \sim U(0,1)$. As mentioned in section 2, $\phi_s(x)$ can be identified from Equation (2). That is to say dependence between T_1 and T_2 can be introduced by assuming that U_1 and U_2 not necessarily independent. This implies that the joint survivor function of T_1 and T_2 conditional on X is

$$S(t_1, t_2 | x) = K \left[\int_0^\infty e^{-m_0(t)\phi_0(x)v_0} f_{v_0}(v_0) dv_0, \int_0^\infty e^{-m_1(t)\phi_1(x)v_1} f_{v_1}(v_1) dv_1 \right] \quad (21)$$

Although it is not quite the same as the model in Heckman and Honore, intuitively one still can conjecture the similar results by using the known probability functions

$$Q_1(t) = \Pr(T_1 > t, T_2 > T_1), \quad Q_2(t) = \Pr(T_2 > t, T_1 > T_2) \quad (22)$$

and the theorem 1 of Tsiatis (1975) that

$$Q_1' = \left[\frac{\partial S}{\partial t_1} \right]_{t_1=t_2=t}, \quad Q_2' = \left[\frac{\partial S}{\partial t_2} \right]_{t_1=t_2=t} \quad (23)$$

Conjecture: with regular assumptions, the joint distribution of v_0 and v_1 can be identified. The interpretation of this conjecture is that. First, one can identify the marginal distribution of v_s ; Second, with similar process one might identify the joint distribution of U_1 and U_2 , the function $K(\cdot, \cdot)$; Overall, one might identify the joint distribution.

From above analysis, the identification of the joint distribution of U_1 and U_2 might have some promising further research. To pursue this topic, there must be some topic to consider: what is the minimum condition for this identification; the exact processes and so on. As we mentioned in the previous footnote, for duration outcome model most of the literature mainly focus on the identification of the parameters the econometricians might be interested in. With identification, how to estimate them might need further research.

6. Conclusion

Partial identification in econometrics is an approach to conducting inference on parameters in econometric models that recognizes that identification is not an all-or-nothing concept and that models that do not point identify parameters of interest can, and typically do, contain valuable information about these parameters. This partial identification approach favors the principle that inference - and conclusions and actions - based on empirical models with fewer suspect assumptions are more robust, hence more sensible and believable. Stronger assumptions will lead to more information about a parameter, but less credible inferences can be conducted. Hazard-based models have been used extensively for several decades in biometrics and economics to examine issues such as life-expectancy after the onset of chronic diseases and the number of hours of failure of an economic program. In this paper, we consider partial identification issues in hazard-based models. Our results exhibit the conditions when the hazard-based models could be fully or partially identified. It would be interesting to develop some confidence bounds when the models are partially identified.

References

- [1] Abbring, J. and G. Berg (2005). "Social experiments and instrumental variables with duration outcomes." Free University Amsterdam, and Tinbergen Institute, Working Paper.
- [2] Berg, G. (1997). "Association measures for durations in bivariate hazard rate models." *Journal of Econometrics*, vol. 79, pp. 221-245.
- [3] Berg, G. and B. Drepper (2016). "Inference for shared-frailty survival models with left-truncated data." *Econometric Reviews*, vol. 35, pp. 1075-1098.
- [4] Cunha, F. and J. Heckman (2008). "A new framework for the analysis of inequality." *Macroeconomic Dynamics*, vol. 12, pp. 315-354.
- [5] Elbers, C. and G. Ridder (1982). "True and spurious duration dependence: the identifiability of the proportional hazard model." *Review of Economic Studies*, vol. 49, pp. 655-682.
- [6] Fan, Y. and S. Park (2012). "Confidence sets for the quantile of treatment effects in randomized experiments." *Journal of Econometrics*, vol. 167, pp. 330-340.
- [7] Fan, Y. and J. Wu (2010). "Partial identification of the distribution of treatment effects in switching regime models and its confidence sets." *Review of Economic Studies*, vol. 77, pp. 1002-1041.
- [8] Guo, Z. (2017). "Comparison of error correction models and first-difference models in CCAR deposits modeling." *Global Journal of Management and Business Research*, 2017, vol. 17, pp. 13-31;
- [9] Shintani, M. and Z. Guo (2016). "Improving the finite sample performance of autoregression estimators in dynamic factor models: A bootstrap approach", *Econometric Reviews*.
- [10] Ichimura, H. (1993). "Semiparametric least squares and weighted SLS estimation of single-index models." *Journal of Econometrics*, vol. 58, pp. 71-120.
- [11] Ham, J. and R. LaLonde (1996). "The effect of sample selection and initial conditions in duration models: evidence from experimental data on training." *Econometrica*, vol. 64, pp. 175-205.
- [12] Heckman, J. (1990). "Varieties of selection bias." *American Economic Review: Papers and Proceedings*, vol. 80, pp. 313-318.
- [13] Heckman, J., J. Smith and C. Taber (1998). "Accounting for dropouts in evaluations of social programs." *Review of Economics and Statistics*, vol. 80, pp. 1-14.
- [14] Honoré, B. and J. Heckman (1989). "The identifiability of the competing risks model." *Review of Economic Studies*, vol. 76, pp. 325-330.
- [15] Jenkins, S. (1995). "Easy estimation methods for discrete-time duration models." *Oxford Bulletin of Economics and Statistics*, vol. 57, pp. 129-136.
- [16] Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Econometric Society Monographs.
- [17] Manski, C. (1988). "Identification of binary response models." *Journal of the American Statistical Association*, vol. 83, pp. 729-738.
- [18] Powell, J., J. Stock and T. Stoker (1989). "Semiparametric estimation of index coefficients." *Econometrica*, vol. 57, pp. 1403-1430.

- [19] Reza, S. and P. Rilstone (2016). "Semiparametric efficiency bounds and efficient estimation of discrete duration models with unspecified hazard rate." *Econometric Reviews*, vol. 35, pp. 693-726.