W.E. UPJOHN INSTITUTE FOR EMPLOYMENT RESEARCH

Dissertation Awards

1-1-2001

Rank Estimation of Duration Models: Dissertation Summary

Govert E. Bijwaard
Free University, Amsterdam

Rank Estimation of Duration Models

Govert E. Bijwaard

MOTIVATION

Recently, social experiments have gained popularity as a method for evaluating social and labor market programs.[1] High-profile evaluations such as the National JTPA Study in the United States and the Self-Sufficiency Project in Canada have brought about real changes in the views and the actions of policymakers. Most of the literature relates to these two countries, in which there is a long-standing tradition of evaluating labor market programs. In this study, we also use data collected from an experiment in the United States. Indeed, in the United States, there is a requirement for public authorities to evaluate their programs. Few European countries have carried out rigorous evaluations; the Netherlands are not an exception. There, the most common method of "evaluation" consists of simply monitoring the labor market status and earnings of recipients for a brief period following their participation in the program. Although this kind of exercise provides useful information, it cannot answer the vital question as to whether the program fulfills its aim.

Ideally, the evaluator would like to know what the outcome would have been for a program participant if the person had not participated; the fundamental difficulty is that a person is never observed in both states. The observable target of estimation is typically the average effect, defined as the average difference between treated (in the program) and untreated outcomes across all persons in a population or in some subpopulation. In a non-experimental evaluation, statistical techniques are used to adjust the outcomes of individuals who choose not to participate in the program to resemble what the participants would have experienced had they not participated. By contrast, a random experiment directly produces the counterfactual of what would have happened to the participant had they not participated by forcing some potential participants not to participate.

In theory, data from a randomized experiment produce an unbiased estimate of the effect of an intervention or program on an outcome variable. A simple comparison of the average outcomes of the treatment group (consisting of those who participate in the program) and the control group (those who are excluded from participation) produces a consistent estimate of the impact of the program on its participants. In practice, a randomized experiment may suffer from the same problems that affect behavioral studies. In particular, the random assignment of the intervention is often compromised by noncompliance with the assigned intervention, i.e., members of the treatment sample may drop out of the program and members of the control group may participate. Noncompliance complicates the analysis of data from a randomized experiment in the same way as does nonresponse in (random) sample surveys and panel attrition in longitudinal studies. If the noncompliance is selective (i.e., is correlated with the outcome variable), then the difference of the average outcomes is a biased estimate of the average effect of the intervention.

Sample selectivity is a familiar problem for economists, and over the years a number of approaches have been suggested to reduce selectivity bias. Since Heckman's work, the dominant approach has been to model the selection process.[2] This is the natural approach if the selection process is of independent interest and the econometrician understands the process well enough to propose a reasonably accurate model. The first generation of these models required an assumption about the joint distribution of the response variable and the (latent) variable that determined participation in the program. In the second generation, this assumption is replaced by an elaborate model of the selection process under the assumption that an unbiased estimate of the intervention effect is obtained by comparing units with an (approximately) equal probability of participation.

Often—and the application considered in this thesis is a good example—there is not enough information to specify a model of the selection process.

Moreover, the available characteristics of the individuals, although significantly correlated with compliance, do not explain compliance well enough to enable a comparison between members of the treatment and control groups in a subsample with the same probability of compliance. Under these circumstances an approach that does not require a model of the selection process is preferable.

The method of instrumental variables (IV) gives an unbiased estimate of the intervention effect and does not require a model of participation. This method assumes that the treatment assignment results from a two-stage process, where in the first stage the sample is divided randomly in two (or more) groups, and in the second stage, units are free to decide whether to participate in the program or not. In the clinical literature, this experimental design is called the intention-to-treat (ITT) design.

Most of the evaluation literature has focused on static interventions, i.e., interventions that are administered at a particular point in time or in a particular time interval. If the outcome is a waiting time (e.g., the time until reemployment), the intervention can be dynamic; that is, it can be switched on and off over time. Examples are the unemployment insurance experiments 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. In general, the intervention may even depend on information that accumulates during the unemployment spell. With such a time-varying intervention, the effect of the intervention becomes dependent on the outcome.

A basic quantity, fundamental in duration analysis, is the hazard rate. The hazard rate is, roughly speaking, the probability of finding a job after some time spent in unemployment, given that the individual was still unemployed at that time. It has a direct relation to the density of a random variable. Economic models for durations, e.g., search models, often have direct implications for the hazard rate. One advantage of a hazard rate model is that incorporating time-varying interventions is fairly easy and natural.

Another reason to consider the effect of an intervention on the hazard rate is that duration data are usually censored. Censoring limits the observation period but is not a feature of the program. Hence, the estimated effect should be independent of the censoring time. Because the hazard rate is invariant to censoring, it is natural to relate the intervention to this quantity.

Two competing approaches to the estimation of the effect of a time-varying treatment on survival have been the (Mixed) Proportional Hazard ([M]PH) model and the Accelerated Failure Time (AFT) model. The MPH models have developed popularity among econometricians, and the AFT is commonly used by biostatisticians and medical statisticians. In the 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. There is a direct relation between hazard models (in particular the PH and MPH models) and transformation models (i.e., regression models in which the dependent variable is transformed). The simplest transformation model is the AFT model in which the dependent variable is the logarithm of the duration.

If all the covariates in the model are exogenous, many possible procedures to estimate the parameters of the AFT or MPH models exist. However, if some of the covariates are endogenous, all these methods fail and it is not straightforward how to generalize the standard IV methods to these nonlinear duration models. This thesis is one of the first attempts to provide instrumental methods for these models with possible endogenous covariates. It also gives a new solution to well-known inference problems of the MPH models. The estimation methods are all based on extensions of the (inverse of the) log-rank statistic.

In Chapter 2, we introduce a two-stage IV method for the MPH models to estimate the effect of a possible endogenous intervention on the hazard. This rank estimation method requires that the members of the control group are excluded from participation. The unobserved heterogeneity component of the MPH models is notoriously hard to estimate and empirically difficult to distinguish from the duration dependence, which can lead to misleading conclusions with regard to the regression parameters. In Chapter 3, we introduce the Generalized Accelerated Failure Time (GAFT) model, which is a generalization of both the AFT and the MPH models. We discuss a semi-parametric estimation procedure of the parameters of this model that is independent of the shape of the unobserved heterogeneity. In Chapter 3 we only consider exogenous interventions. The analysis of endogenous interventions in the GAFT is dealt with in Chapter 4. In that chapter, we develop an IV method for the GAFT models. First, I give a short overview of the data used in the applications.

REEMPLOYMENT BONUS DATA AND ITS IMPLICATIONS

Throughout this thesis, we illustrate the use of our proposed models and estimators by applying them to data from reemployment experiments conducted between mid 1984 and mid 1985 by the Illinois Department of Employment Security.[3] The experiment provides the opportunity to explore, within a controlled experimental setting, whether bonuses paid to unemployment insurance (UI) beneficiaries (treatment 1) or their employers (treatment 2) reduce the unemployment of beneficiaries relative to a randomly selected control group. The first treatment was intended to create an incentive for UI recipients to search more intensively for work and to become employed faster. The second treatment was designed to provide a marginal wage subsidy that might reduce the duration of insured unemployment.

Both treatments consisted of a $500 bonus payment, which was about four times the average weekly unemployment insurance benefit. In the experiment, newly unemployed claimants were randomly divided into three groups.[4]

- The members of the claimant bonus group were instructed that they 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. A total of 4,186 individuals were selected for this group. Of those, 3,527 (84 percent) agreed to participate.
- The employer bonus group was told that their next employer would qualify for a cash bonus of $500 if they, the claimants, found a job (of at least 30 hours) within 11 weeks and if they held that job for at least four months. Of the 3,963 selected for this group, 2,586 (65 percent) agreed to participate.
- The control group (i.e., all claimants not assigned to one of the other groups) consisted of 3,952 individuals.

The individuals assigned to the control group were excluded from participation in the experiment. In fact, they did not know that the experiment took place. The results of the experiment indicated a reduction in the mean number of weeks of insured unemployment. The average number of insured weeks of unemployment decreased about one and one-half weeks for all UI recipients assigned to the claimant bonus group and about half a week for those assigned to the employer bonus group (both relative to the control group). However, because the UI recipients are only observed while they receive UI benefits, for many (about 60 percent) unemployed we only know the time they spend on benefits and not the time until reemployment. Thus, the data is heavily censored at the maximum number of potential weeks of UI insurance, which is 26 weeks for all the individuals in the experiment.

If the bonus reduces the time spent in unemployment, it also reduces the probability that UI recipients will exhaust their UI benefits. Therefore, with censored data, taking averages can lead to wrong conclusions about the effect of the bonus on the duration of unemployment. In this thesis, we show that the logical means of accounting for this censoring problem is to relate the effect of the bonus to the hazard rate instead of to the average duration.

Another reason to consider the hazard rate is that the treatment itself varies over time. Because the period of bonus eligibility ends after 11 weeks, we expect that the effect of the bonus on the reemployment hazard disappears after that time. It is difficult to take such time-varying interventions—which are common in duration data—into account when taking averages. For all three groups the reemployment hazard is high at the start of the unemployment spell, decreases to a minimum after 16 to 20 weeks in unemployment, and rises again close to the benefit exhaustion time of 26 weeks. The hazards for both treatment groups are higher then the hazards for the control group.

However, we cannot draw strong conclusions from these hazard rates due to potential selection bias. About 15 percent of the claimant bonus group and 35 percent of the employer bonus group refused participation. The reason for this refusal is not known, and it is hard to think of an economic model for this decision. The refusal was not completely random, because it was significantly related to some characteristics of the participants, characteristics that are also important determinants of the reemployment hazard. Hence, we can not exclude the possibility that some unobserved variables affect both the compliance decision and the reemployment hazard.

The existence of heterogeneity will bias the aggregate treatment effect measure. If the specification of the hazard rate model is incomplete, then randomization at the start of the spell does not ensure an unbiased estimate of the treatment effect, if the intervention affects the hazard rate multiplicatively. To understand this, presuppose the existence of two types of unemployed where one type has higher hazard rates than the other. The unemployed with a large hazard rate leave the unemployed state quickly, and those with a small hazard rate typically have long unemployment spells. In the absence of the experiment, the sample of UI recipients still unemployed over time will have an increasing fraction of low-hazard-rate individuals. Therefore, the aggregate hazard would slope downward as more weight is placed on the low-hazard group. If the bonus leads to a proportional increase in the hazards of both types, the high-hazard types in the treatment group leave even faster than the high-hazard types in the control group, and this induces a correlation between the type and the intervention indicator. It is not difficult to see that the resulting bias in the intervention effect is toward zero. Note that there is no bias if the intervention has an additive effect on the hazard rate.

Thus, we need an estimator of the effect of the bonus on the reemployment hazard that is consistent if there is selective compliance to the assigned bonus regime, if the data is heavily censored and, finally, if the treatment varies over the duration of the unemployment spell. Chapters 2 and 4 address the estimation of the effect of the bonus for two related models. We propose two instrumental variable methods for duration models that adjust for the potential endogeneity of the choice to be eligible for the bonus. The first method explicitly uses the full compliance of the control group, while the second method allows for the possibility that some members of the control group are also eligible for the bonus. Chapter 3 only considers the data on the control group and discusses the issues concerning rank estimation of a model that generalizes two commonly applied duration models, without worrying about the endogeneity of any covariates in the model. In Chapter 4, we return to the whole dataset and extend the model of the preceding chapter to allow for selective compliance.

We do not attempt to estimate the effect of bonus-eligible individuals collecting the bonus or not. Before a bonus could be paid in this experiment, each member of the claimant and employer group had to make a number of decisions. Firstly, whether to agree to participate. Secondly, if a job was found within 11 weeks, whether to file the Notice of Hire. Finally, whether to file for a bonus if the job lasted at least four months. This thesis only examines the effect of the bonus on the first unemployment spell, corrected for bias due to nonparticipation.

Neither do we use the data on reemployment (quarterly) earnings. In principle, the wage data could be used to estimate a structural job search (see, e.g., Levine 1992; Davidson and Woodbury 1996). The earnings data from the Illinois experiment seem heavily blurred. For many reemployed individuals, the (quarterly) wage in their new job is unknown or too low to be true. Therefore, we focus on the effect of the bonus on the duration of insured unemployment using a reduced form model that does not incorporate the reemployment earnings data. We do, on the other hand, use the information on the pre-unemployment earnings.

An important limitation of the data is that we cannot ascertain the true labor status of a claimant when the UI benefits are exhausted. The administrative data are blind to the labor status of an individual unless he or she is employed in a UI-covered job or unemployed and receiving UI benefits. A significant fraction of individuals do not have reported positive post-UI benefit earnings. This is likely due to the same blindness of the data. In contradiction to job search theory, none of the earnings measures indicate a significant deviation of the earnings due to the bonus experiment.

If the bonus and the rules for collecting it would become widely known, both the worker and firm behavior may change in several dimensions. For example, many people who have a short spell of unemployment between jobs presently do not collect any UI benefits. The bonus could induce some of these individuals to apply for and receive some UI in order to collect the bonus. This is a classic example of a deadweight effect, in which individuals receive a bonus for behavior they would have displayed anyway. This general equilibrium effect would reduce the effect in the society of the bonus program relative to that estimated by the experiment (see Meyer 1995 for a discussion).

General equilibrium effects occur when the program affects individuals other than its participants. Other commonly mentioned general equilibrium effects are displacement effects (for example, if persons in the program take the job that otherwise would have been taken by other unemployed individuals); substitution effects (if the program induces the employers to substitute one group of workers for another group); and tax effects (whereby the taxes collected to finance the program distort the choices of both participants and nonparticipants). In the context of this thesis, we do not attempt to take these possible side effects into account and therefore we ignore their impact in the analyses.

THESIS OUTLINE

In Chapter 2, we propose an estimator that is consistent for the intervention effect if there is selective compliance to the intervention, the outcome variable is a censored duration, and the intervention varies over the duration. The duration model is the popular mixed proportional hazard (MPH) model, although we need not impose the restriction on the disturbance distribution that is implicit in the MPH model. The estimator is a generalization of the linear rank estimator of Tsiatis (1990) and Robins and Tsiatis (1991).

In particular, we allow for a nonconstant baseline hazard, which amounts to a transformation of the dependent duration in the regression representation of the MPH model. The estimator requires preliminary estimates of the baseline hazard. If there is compliance in the control group, then these preliminary estimates can be obtained from the control group sample. This is often the case in social experiments where the number of participants is a small fraction of the population and the members of the control group are not informed of the existence of the experimental program or can easily be excluded from participation. The preliminary estimates are substituted in the second-stage estimating equation of the intervention effect. This two-stage linear rank estimator is computationally attractive, because it avoids the choice of weighting functions for the estimation of the parameters of the baseline hazard. If the control group sample is also used in the second stage, the additional variability due to the preliminary estimates and the induced correlation between the preliminary estimates and the second-stage estimating equation complicates the computation of the asymptotic variance. Using the counting process representation of the first-stage score, we can obtain an estimable expression of this rather complicated variance.

The estimator is applied on the data from the Illinois unemployment bonus experiment. Our estimates show that compliance in the Illinois bonus experiment was indeed selective. We also investigate whether evidence of effect heterogeneity by income before unemployment and by the probability of benefit exhaustion is biased by selective compliance.

In Chapter 3, we consider a generalization of the accelerated failure time (AFT) models that includes the MPH models: that is, the generalized accelerated failure time (GAFT) model. In the GAFT model, the dependent variable, the time-to-event, is transformed, and this unknown transformation has to be estimated together with the regression parameters in the regression function and the error distribution. A major advantage of the GAFT model over the MPH model is that the unknown components of the GAFT model (i.e., the parameters of the regression function, the transformation, and the error distribution) are all estimable with the usual precision. The only disadvantage of the GAFT specification is that the estimated parameters cannot directly be related to the hazard of the time-to-event. We provide an alternative interpretation by relating the components of the GAFT model to the quantiles of the duration distribution.

We generalize the linear rank estimators that have been used to estimate the parameters of the AFT models to the GAFT models. The asymptotic properties of this linear rank estimator (LRE) are derived using counting process theory. We also address the semiparametric efficiency of this estimator. Simulation experiments indicate that the LRE performs well in samples of moderate size. An important side effect of the proposed methods in this chapter is that they provide a new solution to the notorious inference problems of the MPH models. Our semiparametric method leaves the distribution of the unobserved heterogeneity unspecified and gives consistent estimates of the other parameters.

The methods in Chapter 3 are only applicable for exogenous covariates, and therefore we exclusively focus on the control group in the Illinois reemployment experiment. This group consisted of individuals who were excluded from participation in the experiment. In fact, we are sure they did not know that the experiment took place. We estimate GAFT models for this data using linear rank estimators. Both the simulation experiments and the application show that incorrectly assuming an AFT model can lead to misleading conclusions about the regression coefficients.

The GAFT model can be applied to estimate the effect of a randomly assigned (time-varying) treatment on survival. Using the actual treatment as a normal covariate in the model will give biased results of the treatment effect if some of the individuals do not comply with their assigned treatment. The problem is that even if the intervention has no effect on the hazard, the treatment parameter 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. One could ignore the postrandomization compliance and rely on the analysis of the treatment assignment groups. This intention-to-treat estimator suffers from an error-in-variable bias.

In Chapter 4, we propose an instrumental variable method for GAFT models that adjusts for the possible endogeneity of the intervention without suffering the problems of the intention-to-treat method. We develop an estimation procedure that collapses to the linear rank estimator procedure for GAFT models without instrumenting. The GAFT models with instrumenting are a generalization of the rank-preserving structural failure time (RPSFT) models of Robins and Tsiatis (1991). Both models consider a transformation of the duration time to identify the treatment parameters, but the GAFT models allow for an extension of the transformation. The main drawback of the RPSFT models is that they assume a latent baseline duration time exists, representing the individual's survival time had the intervention always been withheld. This implies that if two individuals have identical durations and observed treatment and covariate histories, then they would have had identical durations had they never been treated. The GAFT models are not rank-preserving and therefore do not imply this strong noninteraction.

The two-stage linear rank (2SLR) estimator of Chapter 2 is related to the instrumental variable linear rank (IVLR) estimator proposed in Chapter 4. The 2SLR restricts the transformation to an MPH representation and requires preliminary estimates of the baseline hazard. The 2SLR is only applicable if there is full compliance in the control group, because only then are the preliminary estimates identified. The analysis in Chapter 4 does not impose the MPH assumption, nor does it require that the control group is totally excluded from treatment.

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. If the censoring is part of the study design or a consequence of administrative rules, we can often make the assumption that the (potential) censoring time is known at the start of the study. Then we can modify the GAFT transformation such that this modified transformation and the instruments are independent, and the IVLR estimator on this modified transformation leads to consistent estimators.

Again the data from the Illinois unemployment bonus experiment is used. In Chapter 4, we focus on the importance of extending the RPSFT models to the GAFT models for the reemployment data. Indeed, we find evidence that the RPSFT model is not the correct model for this application. Our results show that restriction the transformation to an AFT model leads to an overestimation of the effect of the bonus.

The Coase Theorem predicts that in a world of zero transaction costs, the bonuses paid to the claimant group and to the employer group would be equally efficient. Because they both imply the same amount of money, an employment contract would be established whenever the bonus was sufficient to enable a mutually advantageous bargain to be struck. Woodbury and Spiegelman (1987) concluded that only the claimant bonus significantly reduces unemployment. We cannot confirm the different treatment effects of the two bonuses. The selective decision to be eligible for the employer bonus seems to blur the effect of this treatment. Our results also indicate that the bonuses only influence the probability to find a job in the first 10 weeks, which is in line with the bonus eligibility period.

NOTES

This thesis consists of three separate articles. Each chapter is solely based on those articles. Every chapter can be read independently of the other chapters; this implies that there is some overlap between the chapters. Moreover, the introductory sections, the sections on the application on the Illinois data, and the appendices with asymptotic properties exhibit substantial repetition among the chapters. The common basis of the chapters is rank estimation in duration models.

This summary is from the author's doctoral dissertation at Free University of Amsterdam; his advisor was G. Ridder. Dr. Bijwaard is now at Erasmus University, Rotterdam.

1. The chapters by Heckman, LaLonde, and Smith (1999) and Angrist and Krueger (1999) in the most recent *Handbook of Labor Economics* discuss the tremendous flow of ideas on how to undertake evaluations of active labor market policies.
2. James Heckman won the 2000 Nobel Price for economics for his contributions to microeconometric theory on selection bias and evaluation of active labor market programs.
3. A complete description of the experiment and a summary of its results can be found in Woodbury and Spiegelman (1987).
4. The eligible population for either the claimant experiment or the employer experiment consisted of those who filed an initial claim for UI between July 29, 1984, and November 17, 1984, and who registered with one of the 22 job service offices in northern and central Illinois. Individuals had to be eligible for 26 weeks of UI benefits, had to be between ages 20 and 55, and had to have no nonmonetary eligible claims.

REFERENCES

Angrist, J.D., and A.B. Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics,* Volume 3A, O. Ashenfelter and D. Card, eds. Amsterdam: North-Holland, pp. 1277–1366.

Davidson, C., and S.A. Woodbury. 1996. *Unemployment Insurance and Unemployment: Implications of the Reemployment Bonus Experiments.* Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research.

Heckman, J.J., R. J. LaLonde, and J.A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics,* Volume 3A, O. Ashenfelter and D. Card, eds. Amsterdam: North-Holland, pp. 1866–2097.

Levine, P.B. 1992. "Testing Search Theory with Reemployment Bonus Experiments: Cross-Validation of Results from New Jersey and Illinois." Working paper, Wellesley College, Wellesley, Massachusetts.

Meyer, B.D. 1995. "Lessons from the U.S. Unemployment Insurance Experiments." *Journal of Economic Literature* 33(1): 91–131.

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(1): 354–372.

Woodbury, S.A., and R.G. Spiegelman. 1987. "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois." *American Economic Review* 77(4): 513–530.