Report
1/2004

Causality and Selection in Labour Market Transitions. Dessertation for the Dr.Polit Degree

Tao Zhang

Stiftelsen Frischsenteret for samfunnsøkonomisk forskning
Ragnar Frisch Centre for Economic Research

Report 1/2004

Causality and Selection in Labour Market Transitions Dissertation for the Dr.Polit Degree

Tao Zhang

Abstract: The dissertation contributes to the micro-economic and -econometric literatures of labour market transitions. By utilising the Norwegian official administrative register data on the entire unemployment population from 1989 to 2000, the candidate for the doctorate has investigated some of the most important causal relations pertaining to the transitions from unemployment to employment. The selection problem associated particularly with empirical analysis of labour market data is also the focal point of this thesis. The candidate has presented some of the methodological contributions to non-parametric identification of causality, and decomposing causal relations and selection mechanisms.

Keywords: Causality, selection, treatment effect, mixed proportional hazard rate model, unobserved heterogeneity, non-parametric estimation, Monte Carlo study.

Contact: tao.zhang@frisch.uio.no, www.frisch.uio.no, phone + 47 22 95 88 16

ISBN 82-7988-043-7
ISSN 1501-9721

Table of Contents

Acknowledgement

This thesis derives from my research work at the Ragnar Frisch Centre for Economic Research. I started working as a research assistant at Frisch Centre in early 1998, and had the fortune to pursue a Dr. Polit. (Ph.D) degree at the Department of Economics, University of Oslo, from 1999, working mostly on labour market transitions and unemployment duration analysis. Studying at the institute that was founded by Ragnar Frisch and associated with the names of great economists such as Trygve Haavelmo and Leif Johansen, and working at the research institute named after the great Frisch, is a honourable experience, which I believe has great impact on my academic perspectives.

I own debts to many people who have helped me during different stages of my doctoral study. First, I would like to thank my supervisors Harald E. Goldstein and Steinar Holden at the Department of Economics, University of Oslo for insightful and invaluable comments and suggestions on the topics and approaches of my thesis. My gratitude is also towards the generous financial support of the Frisch Centre and particularly to the Research Director Erik S. Hernæs, for providing me the opportunity to work with various projects, which lays the ground for my thesis.

I am deeply grateful for the kindness of Knut Røed at the Frisch Centre, whom I have the fortune to learn from and work with during my entire doctoral study period. It is him who not only have inspired the topics of my dissertation, but also read through the entire thesis and provided valuable critiques, comments and suggestions. I have the pleasure to include a joint work with him in my thesis. Without his contribution, this thesis would not be possible.

I would also like to acknowledge here my gratitude towards Oddbjørn Raaum at the Frisch Centre and Hege Torp at the Institute for Social Research for co-authorship of my second paper in this thesis. They provided me in-depth views of causal theories and empirical approaches of evaluating causal effects. They also reviewed the fourth chapter of my thesis and have offered constructive critiques and helpful suggestions for improvement.

Special thanks are given to Simen Gaure at the High Performance Computing Centre, University of Oslo, for the ingenious programming assistance, without which the estimations in this thesis would have taken years to finish.

The stimulating environment and teamwork at the Frisch Centre encourage and facilitate the creation of my dissertation. I have greatly benefited from the discussions with the highly competent colleagues at the Frisch Centre and the seminar participants at the Research Seminar at the Department of Economics, University of Oslo. Individual acknowledgements are given in each paper.

To my parents and my sister who have patiently waited me to fulfil my degree, I thank them for their support and affection.

Last, but certainly not least, I am indebted to my beloved Xiaolin, for her endearment, encouragement and constant inspiration during my thesis work.

Needless to say, the faults remaining are entirely my own.

30th July 2003
Oslo

Chapter 1 Causality and Selection in Labour Market Transitions: Introduction and Summary

Causality and Selection in Labour Market Transitions: Introduction and Summary*

1. Introduction

Unemployment has always been a central topic in labour economics. Most importantly, unemployment involves a large loss for society in general, in the form of reduced output, as well as a loss of welfare for the individuals that are affected. In addition, unemployment is important because of its strong impact on the wage and price setting at macro level, which in turn influence the general equilibrium of an open economy, and on the political concerns of stability of the society. To reduce involuntary unemployment is always a central policy goal.

To combat the unemployment, one must first acquire the understanding of the causes of unemployment and how the causes may interact, so that counter measures can be developed to increase the prospects of employment. Further, knowledge of dynamic aspects of labour market is a key for causal inferences of unemployment. The labour market dynamics are characterised by constant movements of individuals from one state to another, such as transitions from unemployment to employment, from open unemployment to partly employed, entering and withdrawing from labour force, job changes etc. The knowledge on the mechanisms that lie behind all these transitions, particularly the transitions from unemployment, is of great importance for understanding the causes of unemployment, and for policy designs aimed at preventing rising unemployment.

* Thanks to Harald Goldstein, Steinar Holden, Knut Røed and Oddbjørn Raaum for helpful comments and suggestions for improvement and language corrections.

This thesis aims to contribute to the understanding of the driving forces behind the movements in the labour market and in particular to the causal mechanisms that influence the transitions from unemployment to employment. By utilising recently available Norwegian official register-based data and newly developed computational techniques, I provide new insights on the identification of causal parameters with respect to the probability of (re)employment, and on the methods of dealing with selection bias. The key parameters addressed in this thesis are the causal effects on the escape rate from unemployment to employment of (i) unemployment spell durations, (ii) economic incentives and (iii) labour market programme participations.

Trying to identify and understand the causes and effects within labour market transitions involves several challenging aspects.

1. Due to the nature of being a discipline of social science, the data available to economists are almost entirely observational. Contrary to established methods in natural science that identify causality through experiments, the possibility for economists to conduct experimental like research is very limited. Since observational data usually cannot be easily manipulated, this implies that to assess the causal effects, counterfactual thinking is of essential importance. Also one of the most common problems associated with the empirical economic studies using observational data is the selectivity of data due to unobserved population heterogeneity.
2. When the economic theories do not provide explicit guidance on the underlying causality, it is desirable to model the causal parameters in a flexible way, so as to avoid misspecification of the arbitrarily chosen functional form. Semiparametric and non-parametric methods thrive in the econometric literatures. But the identification of causality with flexible functional forms in the presence of selection is still a challenging issue.
3. From a philosophical point of view, the notion of causality is much debated, and it is fair to say that there is no general consensus on the conceptual meaning of causality. Different views of causality give rise to different approaches in identifying causes and causal effects. This not only implies that to derive causal relations from empirical observations one has to be conscious on which causal theory is relevant,

but also means that given the evidence from the observed data, one has to rely on the relevant causal theory applied in the analysis to interpret causal effects.

This thesis is mainly investigating the operational aspects of causal theories and approaches. By exploring and uncovering some particular cases of cause-effect relationships in labour market transitions, I wish to increase the understanding of the causal mechanisms that affect the labour market dynamics. Also, the challenge of selection that plague studies on observational data is addressed in the thesis. In this chapter, I present a brief discussion of causal theories that I consider to be relevant for the empirical research of unemployment durations. The later chapters are dealing with some particular cases of causal inferences. The remainder of this chapter is organized as follows: Section 2 gives a brief account of two causal theories that in my view are particularly suitable for causal analysis of labour market transitions, which I believe also have general relevance to economic researches. Section 3 discusses the problem of selection and the data-based identification of causality with the presence of selectivity bias. I also provide an overview of empirical data that are used throughout the entire thesis. Section 4 offers a synopsis of the other chapters in the thesis.

2. Causal views and causal approaches in analysis on the labour market transitions

Causality refers to the "way of knowing" that one thing causes another. It is the explanation of cause-effect relationships among variables. Through establishing the causal relationships, we obtain a deep understanding of a given phenomenon, and with the acquiring of cause-effect knowledge we are able to predict the future outcome given the proper circumstances and conditions from past experience that we derive such causal knowledge.

Causality is perhaps the most debated notion in the philosophical literature, and can be traced back to Aristotle. Earlier philosophers were concentrating on the conceptual issues of causality, while later philosophers were more concerned with operational aspects. Hume (1740, 1748) is the most known philosopher to postulate a wholly empirical definition of causality. In his view, he emphasizes three elements that can be verified through observation. According to Hume, "X causes Y" if (1) Temporal ordering: X precedes Y in time; (2) Contiguity: X and Y are contiguous in space and time; (3) Constant conjunction: X and Y always co-occur (or not occur). Modern advances in probability theories and statistics have provided new insights both in the conceptual aspect and in the operational fields in many applied researches. Among those, Hill (1971) provides a set of seven general criteria for assessing the extent to which available evidence supports a causal relation. The full review and discussion of philosophical as well as empirical aspects of causality are beyond the ambition and scope of this thesis. In the following, I will give a brief account focusing on the operational side of causality and the causal approaches in econometrics that I view as particularly relevant for analysing labour market dynamics. For a general survey of causality, see e.g. Pearl (2000).

Probabilistic causality

One of the first lessons one learns in statistical courses is that statistical association is not the same as causation. The fact that the events *C* and *E* are jointly observed with positive probability does not imply that *C* is a cause for *E*, or conversely. One reason for this is that statistical association is symmetric. To qualify a statistical association to be causation, the relationship must be asymmetric. One of the qualification for the asymmetry between *C* and *E* is the temporal ordering: if *C* were to be a cause for *E*, *C* must precede *E* in time. Another qualification is that the association cannot be explained by the dependence on features other than *C*. The formal notion of causality via association is the probabilistic causality, initiated by Good (1961-62), and further pursued by Suppes (1970), Cartwright (1979), Eell (1991) and Salmon (1998). A formal account of probabilistic causality can be found in e.g. Suppes (1970).

A simple setting for probabilistic causal model can be viewed as following (e.g. Cox 1992): given a relevant background context *K*, event *C* is said to be a *prima facie cause* (in terms of Suppes 1970) of *E* if

$$P(E \mid C) > P(E \mid \text{not C})$$

This implies that to qualify *C* as a cause for *E*, it requires that occurrence of *C* must increase the probability for occurrence of *E*. If *C* is a prima facie cause for *E*, let *B* be another variable or collection of variables, if

$$P(E \mid C \text{ and } B) = P(E \mid \text{not } C, \text{ and } B)$$

then C is said to be a *spurious cause* because the association of C and E is screened off by existence of B. If C is a prima facie cause and not a spurious cause, then C is said to be a *genuine cause* of E, given the background context K.

In my view, the probabilistic causality theory has particular relevance in applied economic research for several reasons. First, the probabilistic causality has the appealing virtue that it adopts the notion of indeterminism and statistical determinism used in modern physics. As Russell (1913) argued, philosopher's concept of causation involving, as it does, the law of universal determinism that every event has a cause and the associated concept of causation as a relation between events, is "otiose"; and in modern science is replaced by the concept of causal laws understood in terms of functional relations, where these causal laws are not necessarily deterministic. Modern physics suggests that the lawful regularities that were highly regarded by Laplacian philosophers are indeed statistical at bottom, and the causal patterns may be probabilistic ones. If this represents the actual structure of the world, then many events, also social, will have to be viewed as probabilistic outcomes of stochastic processes.

Second, the probabilistic causality theory provides operational feasibility to deduct general causal relation. This is achieved by injecting causal hypothesis into observed association and by restricting attention to subject-matter variables and temporal ordering to establish the asymmetry of causal relations. It offers a mechanism to qualify a potential variable to have a causal meaning among all variables. In e.g. regression analysis, the potential cause variables are modelled as exogenous variables to the regression equation while the caused variables are modelled as endogenous. To say that x_C is a cause for response variable y_E requires that the introducing of other exogenous variable x_B does not diminish the statistical significance of x_C in the regression equation.

Third, to view causal relations in terms of probability, the probabilistic causality theory offers a view of general causal law, a population level causal relation. Whether a particular outcome of such causal relation occurs or not is of less importance. Economic theories are general theories by nature, in the sense that the descriptions and explanations of economic phenomena and common interactions among economic variables are almost entirely at population level, while the behaviours of units, such as an individual or a firm, are subject to both structural determination of general economic theories and random elements. A particular observation is often regarded as the outcome of a stochastic process that is governed by the structural causal laws and random disturbances. From this point of view, probabilistic causality is suitable for economics as a causal approach to derive general economic theories.

Fourth, the probabilistic causality theory recognises that causal interpretation has to be confined within the relevant context. The background context *K* copes with the relevant conditions and circumstances, also possibly the unobserved variables, which might have impacts on the *C* and *E*. Thus the causal inference is not universal, but is conditional on the background context *K*. The control of relevant background context is an open issue and subject to concrete case-specific settings. As for the economics, it is generally understood by economists that causal statements must also be confined within the relevant context. In the sensitivity analysis of policy parameters in economics, the common practice is to fix all other variables and manipulate the potential causal variable to expose if such manipulation has any significant impact, so that causal effect of such variable is conditional on other fixed variables. In regression analysis, exogeneity of x_C and possibly x_B is understood to be conditional relative to the equation system. x_C and x_B might well follow their own stochastic processes. But such processes are understood or assumed to have no causal impact on the equation system in analysis. The "other things equal" or ceteris paribus clause is central to causal interpretation and reflects the conditional nature of causal thinking in economics.

In later chapters, I provide some empirical studies of causality in labour market transitions. Chapter 2 gives an analysis of the importance of economic incentives for the transition probability to job. The causal hypothesis is that the higher the unemployment replacement ratio, the lower the probability of finding a job would be. Chapter 4 considers the effect of participation in active labour market programmes on the likelihood of a successful transition to ordinary employment. Both chapters illustrate the operational aspects of probabilistic causality theory. The main econometric method used in these chapters is the hazard rate model, which I find to be suitable for causal analysis of labour market transitions within the context of probabilistic causality. Chapter 5 focus on methodological aspects of the hazard rate models and through Monte Carlo experiments it offers some insights on the properties and estimations of such models non-parametrically.

Applications of the duration analysis and hazard rate model have flourished in the empirical literatures (see van den Berg (2001) for a recent exposure). The most popular duration model is perhaps the (mixed) proportional hazard rate (Cox (1972), Lancaster (1985)). The model expresses the transition probability to a destination state as a function of observed and unobserved explanatory variables and the elapsed duration spent in the current state. In the context of causal inference, this model is often used to describe the causal influences of the explanatory variables on such transition probability. For a general inquiry of duration analysis, see e.g. van den Berg (2001).

This thesis mainly employs the mixed proportional hazard rate model for exploration of causality in labour market dynamics. In my view, the mixed proportional hazard rate model (MPH) is particularly suitable for causal inferences on labour market transitions. This can be viewed in several ways: First, the MPH model follows the general settings of probabilistic causality by modelling the causal effect through changes on the transition probabilities between states. Thus it has clear meanings with respect to causes and effects. Second, MPH has such flexibility of modelling underlying economic variables of interests in causal inference. Potential cause variables can be incorporated within the MPH as covariates and their effects can be easily estimated with standard econometric methods e.g. the maximum likelihood. As van den Berg (2001) states: “Part of the attractiveness of the (M)PH model stems from the fact that it is difficult to think of a more parsimonious specification of the hazard that includes all single major determinants of it”. Third, the MPH model has the ability to cope with problems pertained to the observational data such as selection, censoring and measurement errors. In addition, MPH offers flexible ways of modelling the causal parameters and do not restrict to particular functional form assumptions. The properties of semiparametric and non-parametric approaches in duration analysis have been thoroughly studied in the literatures and readily implemented in applied researches.

Experimental causality

Modern statisticians believe that well designed random experiments would be powerful tools in investigation of causation and causal inference. Rubin (1974,1977, 1978, 1980) has originated his account of causality from experimental settings. Further development and application of Rubin's theory on measuring causal effects through experiments can be found in e.g. Holland and Rubin (1980), Holland and Rubin (1983), and Heckman and Smith (1995) and Lalonde (1986).

The experimental view of causality maintains that no causal inference without experiments is possible. A much stronger statement is "no causation without manipulation" (Holland, 1986). In this causal modelling work, cause and treatment are interchangeable concepts. To outline the basic ideas of causation, suppose that causal inference is conducted on a population of units ($u \in U$), which are subjects of study. Assume two causes of which *t* is the cause, or treatment applied upon unit *u*, and *c* is control treatment (or equivalently non-treatment). The key notion here is that at some time or in some time interval, unit *u* is exposed to *t* or *c*, and each one of these two elements could have been assigned to the same unit *u*. Thus for unit *u*, either *t* or *c* would be applied, but never both. Let $Y_t(u)$ denote outcome (response) of treatment (or cause), while $Y_c(u)$ denote outcome of control (non-treatment). Then the effect of cause *t relative to* cause *c* to be defined as

$$Y_t(u) - Y_c(u)$$

within the context of this experiment. Since for the same unit *u*, either *t* or *c* is received, but never both, it is impossible to derive the causal inference based on observation of both $Y_t(u)$ and $Y_c(u)$. Rubin calls this to be the fundamental problem of causal inference (Holland (1986)).

Nevertheless, the statistical solution to this is to evaluate average causal effect *T* of *t* (*relative to c*) over the population *U*, which is the expected treatment effect over *U*

$$E(Y_t(u) - Y_c(u)) = T$$

This is equivalent to

$$E(Y_t(u)) - E(Y_c(u)) = T$$

given that $Y_t(u)$ and $Y_c(u)$ are independent. Therefore, in practical applications, one only needs to calculate average response of treatment $E(Y_t(u))$ and average response for non-treatment (control) $E(Y_c(u))$. The algebraic difference is then in fact the causal effect.

The key notion behind this statistical modelling is that, for the same unit *u*, treatment effect and non-treatment effect cannot be observed at the same time, but observation on different units would contribute in probability to identify the counterfactual outcomes to the treatment (or non-treatment). This is however, based on some assumptions lying within the context of an "ideal" experiment: (a) homogenous unit assumption: units subject to causal analysis are homogenous with respect to all possible and relevant aspects. In scientific laboratory work, this is done by carefully prepared subject units so that they are "identical" in every conceivable way. (b) assignment of treatment is conducted by randomisation. For each unit *u*, exposure to treatment *t* (or *c*) is independent of exposure to treatment *c* (or *t*). Thus the response $Y_t(u)$ and $Y_c(u)$ is statistically independent. Let *S* denote exposure to *t* or *c*, then the observed data is (Y_s, S). Under randomisation, we have

$$E(Y_t \mid S = t) = E(Y_t)$$

and

$$E(Y_c \mid S = c) = E(Y_c)$$

Hence

$$T = E(Y_t \mid S = t) - E(Y_c \mid S = c)$$

is the average treatment effect and is well defined in statistical sense.

Experimental causality has also seen vast applications in economics, particularly in the programme evaluation literature. However, due to the fact that most of empirical data available for economic research is observational, the experiment settings in economics are usually quasi-experiment with limited room for manipulation. Some economists pursue the possibility of the natural experiment through exogenous policy changes that affect some by not the whole population, as the closest to the laboratory experiments. A reference is Lalonde (1986).

In an experimental setting, cause and effect have clear interpretations . However, it is not without difficulties in non-controlled observational studies. Particularly, as the subject in social science is often a human being, it is sometime impossible to conduct experiments that fully satisfy the experimental settings. This is because firstly, it is impossible to manipulate human beings so that the target population could be homogenous. Thus unexpected results could arise due to uncontrolled heterogeneous characteristics that might have influence on the outcome. Secondly, treatment sometime cannot be assigned randomly in observational study, either because it is improbable, unethical or prohibitive by cost. Thirdly, note that a key feature of experimental causality is counterfactual thinking. To assess the effect of a cause, one must be able to deduce the counterfactual effect of non-cause. As explained above, effect of a given treatment *t* is the difference of outcome $Y_t(u)$ of this treatment and the counterfactual outcome $Y_c(u)$ that the same unit would have had it be subject to treatment *c*. Although statistically this counterfactual effect can be meaningfully defined and obtained, as Glymour (1986) points out, counterfactual conditions could be logically false: Unit *u* could have outcome other than $Y_c(u)$ had it been exposed to treatment *c*.

Within the context of quasi-experiment, methodological development has facilitated the causal inferences. A tool that has proven useful in many applications is matching, see e.g. Heckman, Ichimura and Todd (1998). Matching is based on the assumption of conditional independence between the treatment assignment and the outcome. By pairing units in the treated group with the similar units within the controlled group, matching can circumvent some of the difficulties that pertain to the studies on the observational data. The merit of such matching technique is that by construction of an analysing sample of matched "identical twins", one can control for observed heterogeneity, so that the effect in question is not driven by the differences of the distributions of the observed heterogeneity across the treatment and controlled groups. This reduces the risk of selection bias due to observed heterogeneity. In addition, matching enables researchers to assess another important aspect of experimental causality, namely counterfactual thinking. Since the matched "identical twin" resembles the treated as much as possible, by evaluating the same quantities of $Y_t(u)$ and $Y_c(u)$ on the treated and controlled groups, it is feasible to acquire counterfactual effect that is central to experimental causal inference.

Chapter 3 offers an application of matching to a study of the effects of Norwegian labour market training programmes. The hypothesis that participation in labour market training programmes contributes to post-programme success measured by increases of yearly labour earnings is tested on multiple matched cohorts over 6 years. In particular, that chapter matches the treated and the controlled through propensity scores that acquired through estimations on the multinomial choice models.

3. Selection, labour market transition, and data based identification of causality

Selection is a frequently encountered problem in empirical economic research. It is a less accurate but informative term that implies that the sample used in analysis might be selective such that it has different properties than the target population upon which inference is to be made. A classic example is the work of Heckman (1979) on the female labour supply, where the hours worked are only observed for those who participate in the labour force. Failing to control for selection would lead to biased estimators on the causal parameters in question. Existence of and effect due to the selection are acknowledged by most empirical economists, and quite a few statistical and econometric methods are developed to deal with selection bias, see e.g. Heckman, Ichimura, Smith and Todd (1998) for a detailed discussion.

Selection can arise from two possible sources: sampling practice and unobserved population heterogeneity. Survey based data sampling is especially vulnerable to the selection bias if the sampling's design does not take explicit account for which subpopulation has been chosen for the study. Even the register-based data sampling is not free from the selectivity due to e.g. updating practice. For example, the Norwegian administrative unemployment registers are updated based on the response cards submitted by unemployed workers each week. For those with unemployment benefit entitlements, registration with the Public Employment Services is mandatory. For those without benefit entitlements, they would have weak incentives to register. Thus the registered durations of unemployment spells based on submitted response cards might be overrepresented by those with the unemployment benefits. Another source of selection is due to unobserved population heterogeneity. This is often referred as self-selection. Loosely put, this could mean that individual's endowments, motivations, incentives and will of determination, or other underlying attributes that cannot be observed by researchers, make the individual more inclined towards certain choices, decisions or actions. The observed outcomes of potential causal factors can thus be affected by such individual's self-selection.

In short, it is likely that the data researchers work with is selective by nature. Ignoring or failing to control for selection would inevitably lead to false or invalid inferences on the causes and effects. A typical example is the duration dependence in labour market transition analysis. Empirically it is often observed that the probability of finding a job is decreasing with the length of the unemployment spell. A possible causal statement might suggest the discouraged worker effect or the stigmatisation due to long-term unemployment. But Heckman and Singer (1985, proposition 1, pp. 53) have demonstrated that "Uncontrolled unobservables bias estimated hazards toward negative duration dependence." Therefore a valid causal statement on duration dependence requires that one control for such possible biases due to unobserved heterogeneity.

Self-selection is particularly relevant in the evaluation of treatment effects of active labour market programmes. If an individual has higher motivation and/or expectations towards the programmes, he or she might be more eager to participate and benefit more from the participation. Also, if the individual has higher ability and preferable employment characteristics, he or she might be more successful even without the treatment. If this were the case, the effects of programmes would be over-estimated, unless such self-selections are accounted for.

Controlling for the impact of selection bias on estimates of causal variables can be done in various ways. For selection due to sampling practice, better design of data collection and better control for sampling practices are the effective ways. However, control for selection due to unobserved variables or self-selections requires more thorough consideration and accumulation of causal knowledge. In the probabilistic causality context, control for selection can be carried out by conditioning the causal inference on the possible distribution of unobserved heterogeneity. This is equivalent to regarding the unobserved heterogeneity as part of the background context K. In hazard rate models, this is done by conditioning the inference of the causal parameters of the hazard rate model on the often unknown distribution of the unobserved heterogeneity. In the experimental causality, testing method on conditional independence of causal effects on the treatment assignment has been proposed and carefully studied (Rosenbaum and Rubin 1983, Rosenbaum 1984). Within the evaluation literature, econometric methods have been developed to tackle the selection problems. Popular approaches include matching, differences-in-differences, index-sufficient, etc. (see Heckman, Ichimura, Smith and Todd (1998)). In particular, matching by propensity scores are showed to be an effective way of including unobserved variables as part of the causal explanation.

Recently, large administrative register data have become available for research purposes, providing a welcome opportunity for non-parametric identification of causality in labour market transitions. At the Ragnar Frisch Centre for Economic Research, we have over several years built a complete unemployment register for the Norwegian unemployment population. In the present form, "The Frisch Centre Database" covers the whole Norwegian population aged 16-69 and contains information about labour market status during the 1989-2002 period (some status, such as unemployment can be updated almost continuously). Each individual record comprises demographic information (age, gender, country of birth, marital status etc), education attainment, current-stage income, income history and labour market experiences. With the access to large-scale detailed individual information for 12 years, we are not only able to identify many mechanisms that would remain unexplored with survey data, but also able to conduct flexible reduced-form estimations without rigid assumptions on the functional distributions on the causal parameters for the labour market transitions. Røed and Raaum (2003) have provided a comprehensive account for the potentials of administrative register data in empirical researches.

Observation of lagged explanatory variables can provide a valuable source for identification of unobserved heterogeneity that is central to the control of selection bias. At the more general level, this idea can be related to Leibniz[1] (1686, quoted from the translation in Loemker, 1969, p. 500): ... "*for since this command in the past no longer exists at present, it can accomplish nothing unless it has left some subsistent effect behind, which has lasted* *and operated until now, and whoever thinks otherwise renounces any distinct explanation of things, if I am any judge, for if that which is remote in time and space can operate here and now without any intermediary, anything can be said to follow from anything else with equal right*". The quotation of Leibniz can be interpreted to imply the following: the past event itself cannot have direct influence upon the present event, other than through the influence in the past upon the current variables that generate causal effect in the present. This insight could further motivate the empirical data-based identification of the unobserved heterogeneity using an easily acquirable lagged explanatory variable in the form of calendar time variations of exit rates from unemployment: the past calendar time (in the form of time-varying covariates) of unemployment spell is an important instrument in the identification of the unobserved heterogeneity, provided that the past labour market conditions experienced earlier in the spell do not have causal effects on the current transition probability, given current state of labour market conditions. Within the context of hazard rate models, this could be elaborated further that conditional on all current values of observed explanatory variables and given that the unobserved heterogeneity does not vary over time, any dependence between the current hazard rate and past (lagged) values of explanatory variables must reflect the influence of the unobserved heterogeneity on the hazard rates during the elapsed spell. The intuition behind the proposition of using lagged explanatory variables to identify unobserved heterogeneity can be thought of as following: Consider two individuals that are identical in every observed aspect and have the same length of unemployment spell. The only observed difference between them is the calendar time at which they enter into unemployment. Given the assumption of proportional hazard, these two should experience the same hazard rate if they have the same value of unobserved heterogeneities. But if one experiences unemployment during a slump period when "everyone" is hit by the unemployment risk while the other starts unemployment in a boom time when job opportunity is good and the overall outflow rate is high, it is intuitively plausible that the individual being unemployed in the boom time should have a better job opportunity and shorter duration than that of the "identical twin" in the slump time. The fact that they have the same spell length can then only be accredited to the unobserved differences between them, in addition to pure chance element. It is likely that the one unemployed in the boom time have more unfavourable personal characteristics than the one

[1] Thanks to Knut Røed for making me aware of this quotation and inspiring the related idea.

in the slump time with the same spell length. This is to say that, the calendar time at which unemployment spells take places and undergo is a source of hazard rate variation, ceteris paribus, that contains information about the expected value of unobserved heterogeneity. Formal proof of identification of hazard rate model with unobserved heterogeneity by utilising time-varying covariates can be found in McCall (1994) and Brinch (2000). Access to large administrative register data with complete observations of unemployment history provides the ground for such data-based identifications on unobserved heterogeneity. Chapter 2 and 4 are applications utilise this identification source, and Chapter 5 provides some statistical insights and evidences of such data-based identification.

In addition, the longitudinal data can reflect possible outcomes due to policy changes and policy practices that are exogenous from the individual's point of view. An example is the acquirement of independent variation in the explanatory variables that otherwise are observationally correlated. In Chapter 2, we explore the subtle feature of Norwegian unemployment benefit system that uses calendar year income as basis for computation of unemployment compensation. The arbitrary administrative regulation provides in this case the necessary variation on the replacement ratio that is independent of previous income and labour market experience, as well as spell durations. This kind of independent variation is not available in survey data or cohort data with limited coverage.

The longitudinal nature of unemployment register data also invites new thoughts on the treatment evaluation. The traditional approaches are mostly of static nature. Typically the assignment of treatment occurs at one point of time, and the effect of the treatment is evaluated at a later instance. Within the context of evaluation of labour market programmes, information about for how long a person is in unemployment before the treatment is undertaken, and how fast a person obtains a job after a treatment, are typically ignored. The newly evolved time-to-event approach is based on the idea that the duration until event provides valuable information that may work as an identification source for selection effects. The hazard rate framework has proved to be a well-suited method in treatment evaluation (Abbring and van den Berg 2003). By modelling the selection into treatment through a competing risks hazard rate model, it can cope with the randomness of the treatment assignment so that the outcomes of treatment are statistically independent to the probability of receiving treatment. The self-selection due to unobserved characteristics can be captured by mixing the distribution of unobserved heterogeneity within a proportional hazard rate framework as well. In addition, hazard rate models open the possibility that the causal effects of treatment can be time-varying, thus offering the opportunity to evaluate the causal parameters in a dynamic way.

Large administrative register data also facilitate quasi-experimental studies, especially offers the promising opportunity for matching techniques. This can be viewed in three aspects: 1. an observationally identical person for using as a match for the treated to assess counterfactual outcome is always available; 2. with large administrative data, it is possible to acquire homogeneous samples with respect to pre-programme labour market experience, opportunity sets, as well as personal characteristics *prior to* the actual matching. This would certainly increase the validity of conditional independence assumption. 3. register data allows a maximum degree of flexibility to evaluate causal effects across individual heterogeneity as well as other explanatory variables such as business cycle conditions.

The old saying within the economic community "Good data helps a lot" is perhaps a proper summary for this section. Though practioner economists begin to be aware of the potential of data-based identification, we have not yet seen many applications. The later chapters provide some empirical evidences with respect to data-based non-parametric identification of causality in labour market transitions. In particular, they also show the applicability of the data-based identification and the data-based control of unobserved heterogeneity that gives rise of selection biases.

4. Synopsis

Does unemployment compensation affects unemployment duration? (with Knut Røed)

This paper addresses the causal relationships between economic incentives embedded in the unemployment compensation system and the transition probability out of unemployment. The overriding problem associated with analysis of effects of unemployment benefit system is the lack of independent variation in the replacement ratios. It is typical that the benefit entitlement is correlated with the previous incomes, which in turn are correlated with the unobserved individual characteristics. The views on whether the economic incentives have significant impact on the employment probability and on the feasibility of identification of such effects are mixed within the whole literatures of unemployment compensations.

In the standard search theory, an unemployed worker chooses the search intensity and the degree of job selectivity so as to maximise discounted expected utility. According to the reservation wage model, unemployed jobseekers accept only offers that exceed a reservation wage level. If the search involves cost (search effort and/or opportunity cost of lost leisure), the optimal search model predicts that the individual would accept the wage offer when the marginal search cost equates the marginal utility gain of acceptance. As a special case, the exit rate from unemployment is homogenous of degree 0 with respect to expected wage and benefit level, thus only the replacement ratio (the benefit level relative to the expected wage) would affect the transition probability from unemployment to job. Since the replacement ratio is likely to be strongly correlated with unobserved individual characteristics, spell length, as well as business cycle conditions, it is very difficult to acquire necessary independent variations of replacement ratio that is required for a sensible identification of causal effects of unemployment compensation on the escape rate.

We utilise two unique features of the Norwegian unemployment insurance system that provide the required independent variations on the replacement ratios. First, for the new entrants to unemployment that have less than two years of full-time employment prior to the entrance, the unemployment benefit is calculated on the basis of previous calendar year's income, which means for a given income level over the last twelve months, the more it is concentrates within the last calendar year, the higher the benefit would be. The second source arrives from the general indexation rules applying on the spells starting in May-December by the administrative regulation. For new entrants to unemployment in May-December period, the base income is index-adjusted according to adjustment factor applying to pension system before the benefit is calculated, while for spells starting in January-April there is no such indexation. These two subtle features give rise to variation in the replacement ratios that we consider to be conditionally independent of the unobserved characteristics at the individual level.

We base our analysis on the Norwegian unemployment population from 1991 to 1999 and develop a mixed hazard rate model with non-parametric specifications on the duration dependence and the unobserved heterogeneity. The data is carefully grouped according to the source of variation of incentive variables. We also model the calendar time effects and business cycle conditions using flexible non-parametric specification.

Our findings suggest that there are disincentive effects associated with the unemployment benefit system. Generous unemployment compensation has a negative impact on the transition probability to job. The disincentive effects do not seem to be sensitive toward business cycle and spell lengths. There is evidence of heterogeneous effects of marginal changes of unemployment benefit with respect to individuals' characteristics. The threat of benefit termination has a substantial positive effect on the exit rate from unemployment in the months just prior to benefit exhaustion. Our findings also offer some policy implications for the design and implementation of the unemployment benefit system.

Business cycles and impact of labour market training programmes (with Oddbjørn Raaum and Hege Torp)

Active labour market programmes have been used extensively to combat rising unemployment during the past decades in Norway. The causal impact of programme participation on employment and labour market success is of considerable importance from a policy point of view. In this paper we explore the possible causal relationship between programme participation and post-programme labour earnings. In particular, we investigate to what extent the impact of participation in the labour market training programmes depends on the business cycle and the labour market conditions, looking for possible cyclical patterns in the treatment effect.

As noted earlier, with observational data, the assignment of treatment cannot be viewed as entirely random. As in the evaluation of labour market training programmes, the administrative selection (admission) and the self-selection into the participation are the main sources of non-randomness of the treatment assignment. In addition, an individual is observed receiving either treatment or non-treatment, but never both. To assess the treatment effect (differentials of the effects of treatment and the effects of non-treatment) involves counterfactual thinking. Hence the challenge is multi-dimensional.

In this paper, we utilise the Norwegian administrative register data on unemployment registers, earning and taxation data from 1991 to 1997 to evaluate treatment effects of labour market training programmes on post-programme labour earnings for multiple cohorts starting the participation at different stages of business cycles, and follow their earnings over long post-programme period to assess cyclical patterns of treatment effects. Each cohort is constructed by matching those in the treatment group (participants in the training programmes) with those in the control group (non-participants) based on the propensity scores. In particular, we estimate the probabilities of both participation in training programmes and other possible labour market transitions through a multinomial logit model, and match the treatment groups with the non-treated control groups through the propensity score matching. We conduct matching separately for men and women, for unemployed with and without unemployment benefit entitlement. The post-programme labour earning differences for the treated and the non-treated are evaluated for matched samples. Estimated average treatment effects are also evaluated under different business cycles conditions. Two possible business cycle indicators are used: one is characterised by gender-specific macro-level unemployment rates, the other is a gender-specific county-level empirical job opportunity indicator estimated through a hazard rate model.

Using multiple cohorts of participants, we are able to estimate first, second and third year effects under different labour market conditions, controlling for fixed regional effects. We find evidence that the labour market training programmes have significant positive impact on post-programme earnings. The impact is persistent, so that the positive impact of training programmes on earnings remains even after several years. We also find that the impact of training programmes is significantly procyclical, i.e. the effects are strongest when the labour market conditions are favourable. This insight is based on meta analysis of the large number of group- and cohort-specific training effects based on two business cycle indicators. Training effects are positively correlated with job opportunities measured by both indicators. The influence from labour market conditions, i.e. business cycles, on estimated programme effects can also be useful when assessing and explaining differences in effects across time, regions and even countries. Our results are also useful for policy making as the optimal timing and volume of active labour market programmes must take into account that individual effects are likely to vary over the business cycle.

As in most non-experimental studies, the estimated training effects can be driven by selection on unobservables rather than a causal impact on post-training outcomes. In our case, the institutional setting does not provide any clearcut indication. As for most programmes targeted at unemployed, the recruitment to labour market training programme is a mixture of self-selection and administrative decisions. Previous studies of selection processes suggest that there is, if any, a positive selection to labour market training programme, i.e. participants have observed – and possibly also unobserved - characteristics assumed to correlate positively with employability. In this study pre-training earnings records are available. When looking at whether individual earnings are correlated with future labour market training programme participation, the null hypothesis of no correlation is not rejected for any of the groups (for which pre-training earnings are available). Since pre-training earnings are not significantly different for participants and non-participants, we gain more confidence in the identification of causal effects.

Identifying treatment effects of active labour market programmes for Norwegian adults

This paper aims to contribute to the understanding of the dynamics of causal effects of the active labour market programmes. In the massive econometrics literature on evaluation of the effects of active labour market programmes, the approaches have largely been of static nature. It is typically assumed that at one point of time, the participation in programmes is characterised by binomial or multinomial choice models. At a later stage, the outcome of participation is evaluated by some suitable measures such as employment status, labour earnings, job length etc. This kind of static evaluation requires strong assumptions on the independence between assignment probability of treatment and causal effect of such treatment. Also these static evaluation measures have limited capability in dealing with selection biases due to unobserved population heterogeneity.

The duration model framework offers the possibility to tackle the selection bias on the probability of treatment assignment in a novel way. The idea is that the dynamic processes towards the participation in programmes and towards the outcomes of such programmes convey valuable information on the selections due to unobserved heterogeneity. This could reflect on the probability of receiving treatment and the probability of the outcome. By modelling the probability of receiving treatment in the form of hazard rate, it provides the necessary randomness for programme participation, which ensures that the determinants for treatment assignment are stochastic at the individual level. By controlling for unobserved heterogeneity, the hazard rate approach has the capability to minimise the impact of administrative selection and self-selection biases on the treatment effects of programme participation. Also a feature embedded in the observational duration data has been utilised to facilitate the control for unobserved heterogeneity, namely the time-varying calendar time as covariates that reflect the selections on the unobserved personal characteristics at earlier stage of the spell.

The data used in this study is carefully prepared such that individuals in analysing sample are homogenous in terms of labour market preferences and prospects. By utilising rich Norwegian administrative register data of unemployment for the period of 1990-2000, the paper has evaluated three types of major active labour market programmes for Norwegian prime-aged unemployed jobseekers. The labour market training programmes, the public employment programmes and the wage subsidy programmes are evaluated separately and simultaneously. We have opened the possibility that the causal effects of programme participations are different during the actual participation and during the post-programme periods. By allowing both the during-programme and post-programme effects to be time-varying and modelled non-parametrically, we can explore the dynamics of treatment effects on the employment probabilities. Another feature of this paper is the assessment of heterogeneous treatment effects with respect to individual's characteristics such as age, gender, and educational attainment. Also, the diverse programmes are evaluated in the conjunction with business cycle and general labour market conditions to uncover possible cyclical patterns and trends.

The econometric model used is a non-parametric mixed proportional competing risks hazard rate model with dynamically assigned risk sets. The intended effects of the participation in the active labour market programmes are measured by the changes on the transition probability to ordinary employment. The sample is restricted to the prime-aged jobseekers with unemployment benefit, to avoid possible selection due to short-term economic incentives of participation. Several interesting findings have emerged:

1. There is evidence of selections into different programmes with respect to individual characteristics. Employment programmes have been dominated by low-qualified unemployed workers, while wage subsidy programmes are targeted to jobseekers with favourable employment qualifications.
2. The participation in programmes has different effects on transition probabilities to employment at different stages of the unemployment spell: during the participation, the transition probability is low compared to non-participation; while after-programme effects are significantly positive and participants have on average higher employment probability than non-participants.
3. Heterogeneous effects of different programmes with respect to individual characteristics. The training programmes and wage subsidy programmes have their intended effects of enhancing job opportunities, while employment programmes do not seem to have strong impact on post-programme job probability. Women seem to benefit more from participation than men. Also the younger jobseekers and those with high education attainment benefit more from programmes participation than others in general.
4. There is evidence that the effects of training programmes and wage subsidy programmes are procyclical, and that the positive effects are persistent over time.
5. Some stylised analysis based on simulation show that the overall treatment effects of the active labour market programmes are significantly positive in terms of reduced total exposure of unemployment.

A Monte Carlo study on non-parametric estimation of duration models with unobserved heterogeneity

This paper aims to provide insights on the non-parametric identification and estimation of mixed proportional hazard rate models through Monte Carlo experiments. Though the advantages of non-parametric estimations have been realised by several authors in years, e.g. Heckman and Singer (1984), Horowitz (1999), Baker and Melino (2000), the computational challenge associated with non-parametric specifications has been a major obstacle for further assessment of properties of estimators. With recently acquired high-performance computational power, we hope to bring some new understanding on the front of non-parametric estimation of duration model.

The Monte Carlo method provides a suitable laboratory-like framework for study the problem. It also gives rise the opportunity to test the proposition of using time-varying lagged explanatory variables as an additional source of identification on unobserved heterogeneity. The data are simulated to resemble the real observational data that are familiar to empirical researchers. We consider a variety of model combinations for single risk hazard rate models with respect to the functional duration dependence and the unobserved heterogeneity distribution. The constant and the negative duration dependence assumptions are considered in detail; the familiar Gamma distributed unobserved heterogeneity and the discrete distributed unobserved heterogeneity are thoroughly investigated. We also consider a spectrum of calendar time variations to explore the effect of these time-varying covariates on the identification and estimation of mixed proportional hazard rate models. All model terms are estimated non-parametrically to avoid arbitrary functional form assumptions. Sampling properties are also addressed.

Our findings indicate that the totally non-parametric specified duration models can be successfully estimated. The non-parametric specifications can approximate the parametric distributions reasonably well. Our proposition on using time-varying calendar variations as additional identification source for the unobserved heterogeneity proves to be reasonably successful. The findings suggest that the inclusion of such calendar time with large variations would improve the identifiability of model considerably.

We also find the use of maximum penalised likelihood to be an important control for convergence. The pure maximum likelihood method has the tendency to over parameterise the mixing unobserved heterogeneity distribution by finding more points of support than necessary to characterise the distribution. Consequently, the structure parameters and the duration dependences are estimated with positive biases. We find that when sample sizes are small and the calendar time variations are limited, maximum penalised likelihood in the form of information criteria would improve the qualities of estimators. When sample size increases, maximum likelihood and maximum penalised likelihood converge to each other.

The Monte Carlo experiments are also extended to bivariate competing risks models. Our results show that the estimation methods on single risk models can be easily applied to competing risks models. Furthermore, the results suggest that complex bivariate competing risks models can be estimated with non-parametric specifications on all model terms. We find positive evidence that calendar time variation contributes to the control for unobserved population heterogeneity, hence reduces the potential bias on structure parameters and duration dependence. The behaviours of estimators for the unobserved heterogeneity are not clear at this moment, which is an issue that invites further exploration. Also the asymptotic properties of non-parametric maximum likelihood estimators remain a challenge for future research.

References

Abbring, J. H. and van den Berg, G. J., (2003) "The Non-Parametric Identification of Treatment Effects in Duration Models" *Forthcoming in Econometrica.*

Baker, M. and Melino A., (2000), "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study" *Journal of Econometrics* 96, 357-393.

Brinch, C. (2000) "Identification of Structural Duration Dependence and Unobserved Heterogeneity with Time-varying Covariates" Memorandum No. 20, 2000, Department of Economics, University of Oslo.

Cartwright, N., (1979), "Causal Laws and Effective Strategies", Nous 13:419-37.

Cox, D.R., (1972) "Regression Models and Life-tables (with discussion)" *Journal of Royal Statistics Society*, 34, 187-202.

Cox, D.R., (1992) "Causality: Some Statistical Aspects" *Journal of Royal Statistics Society*, A, 155, part 2, pp. 291-301.

Eell, E. (1991) *Probabilistic Causality*. Cambridge University Press.

Glymour, C. (1986) "Statistics and Metaphysics", Comment on "Statistics and Causal Inference" by Holland, P.W., *Journal of the American Statistical Association*, vol 81, issue 396, 964-966.

Good, I. J. (1961-62) " A Causal Calculus I-II" British Journal for the Philosophy of Science 11: 305-318; 12: 43-51; Errata and Corrigenda, 13:88.

Heckman, J.J., (1979) "Sample Selection Bias as a Specification Error (with an Application to the Estimation of Labor Supply Functions)", *Econometrica* 47(1) 153-161.

Heckman, J. J., Ichimura, H., Smith, J. and Todd, P., (1998) "Characterizing selection bias using experimental data." *Econometrica,* 66, pp.1017-1089.

Heckman, J. J., Ichimura, H. and Todd, P. (1998), "Matching as an econometric evaluation estimator." *Review of Economic Studies,* 65, pp. 261-294.

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

Heckman, J.J. and Singer, B., (1985) "Econometric Duration Analysis", *Journal of Econometrics*, 24, 63-132, 1984.

Heckman, J. J. and Smith, J. (1995) "Assessing the Case for Social Experiments (in Symposia: Social Experiments)" *The Journal of Economic Perspectives*, Vol. 9, No. 2. pp. 85-110.

Hill, A.B. (1971) *Principles of Medical Statistics*, 9th ed. New York: Oxford University Press.

Holland, P.W., (1986) "Statistics and Causal Inference", *Journal of the American Statistical Association*, vol 81, issue 396, 945-960.

Holland, P. W. and Rubin, D. B. (1980) "Causal Inference in Prospective and Retrospective Studies," address given at the Jerome Cornfield Memorial Session of the American Statistical Association Annual Meeting, August 1980.

Holland, P. W. and Rubin, D. B. (1983), "On Lord's Paradox," in Principals of Modern Psychological Measurement, eds. H. Wainer and S. Messick, Hillsdale, NJ: Lawrence Erlbaum.

Horowitz, J.L. (1999) "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity." *Econometrica*, 67-5, pp. 1001-1028.

Hume, David, (1740), *A Treatise of Human Nature*.

Hume, David, (1748), *An Inquiry Concerning Human Understanding*.

LaLonde, R.J. (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *The American Economic Review*, Vol. 76, No. 4., pp. 604-620.

Lancaster, T. (1985), "Generalised residuals and heterogeneous duration models, With applications to the Weibull model" *Journal of Econometrics*, Vol. 28(1), 155-169.

Loemker, L.E. (ed.) (1969) Leibniz: Philosophical Papers and Letters. Dordrecht.

McCall, B. P., (1994) "Identifying State Dependence in Duration Models", *Proceedings of the Business and Economics Section American Statistical Association*, 14-17.

Pearl, J. (2000), *Causality: models, reasoning and inference*. Cambridge University Press.

Rosenbaum, P.R. (1984) "From association to causation in observational studies." *Journal of American Statistical Association*, 79, 41-48.

Rosenbaum, P.R. and Rubin, D. B. (1983) "The central role of the propensity score in observational studies for causal effect." *Biometrika*, 70, 41-55.

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

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

Rubin, D.B., (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6,34-58.

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

Russell, B., (1913), "On the Notion of Cause". *Proceedings of the Aristotelian Society* 13: 1-26.

Røed, K., and Raaum, O. (2003) "Administrative Registers – Unexplored Reservoirs of Scientific Knowledge?" *The Economic Journal*, Vol. 113 Issue 488 258-281.

Salmon, W., (1998), *Causality and Explanation*, Oxford University Press.

Suppes, P. C., (1970), *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.

van den Berg, G. J. (2001), "Duration Models: Specification, Identification, and Multiple Durations." *Handbook of Econometrics*, Vol. 5, Chapter 55, Elsevier.

Chapter 2: Does Unemployment Compensation Affect Unemployment Duration?*

By Knut Røed and Tao Zhang

* This paper has been published in *Economic Journal*, vol.113, No.484, 190-206.

Does Unemployment Compensation Affect Unemployment Duration?

By Knut Røed and Tao Zhang[*]

Abstract

We use a flexible hazard rate model with unrestricted spell duration and calendar time effects to analyse a dataset including all Norwegian unemployment spells during the 1990's. The dataset provides a unique access to conditionally independent variation in unemployment compensation. We find that a marginal increase in compensation reduces the escape rate from unemployment significantly, irrespective of business cycle conditions and spell duration. The escape rate rises sharply in the months just prior to benefit exhaustion. While men are more responsive than women with respect to marginal changes in compensation, women are most responsive with respect to benefit exhaustion.

Keywords: Unemployment spells, business cycles, unemployment compensation, non-parametric duration analysis.

JEL Classification: C41, J64.

[*] The Ragnar Frisch Centre for Economic Research, Oslo. We wish to thank the Research Council of Norway for financial support and Christian Brinch, Harald Goldstein, Costas Meghir, Espen Moen, Steinar Strøm, Rolf Aaberge and three anonymous referees for helpful comments.

Chapter 3: Business Cycle and Impact of Labour Market Programmes.*

By Oddbjørn Raaum, Hege Torp and Tao Zhang

* This paper has been published in Memoranda 14/2002, Department of Economics, University of Oslo, ISSN: 0801-1117

Business cycles and the impact of labour market programmes*

Oddbjørn Raaum

The Ragnar Frisch Centre for Economic Research, University of Oslo

Hege Torp

Institute for Social Research, Oslo

Tao Zhang

The Ragnar Frisch Centre for Economic Research, University of Oslo

Oslo, July 22nd, 2002

Abstract

By comparing mean outcomes for a large number of matched samples of participants and non-participants we estimate individual earnings effect of the Norwegian labour market training programme (LMT) targeted at unemployed adults in the years 1991-1996. The average training effect on the trained is positive, even after three years. The training effect is positively correlated with post-training job opportunities in the (local) labour market, when job opportunities are measured by time-varying human capital adjusted national and region-specific exit rates from unemployment.

Keywords: Causal effects, matching estimators, training unemployed, impact on annual earnings.

JEL classification: C14, J64, and J68.

* Acknowledgements: The research is funded by the Research Council of Norway (grants no 124583/510 and 124613/510), at the Ragnar Frisch Centre for Economic Research and the Institute for Social Research. Comments from Karl-Gustav Löfgren, Ragnar Nymoen, Knut Røed, participants at the Labour Market Workshop, October 2001, arranged by Norwegian Research Council, and at the Annual Meeting of Norwegian Economists, January 2002, are gratefully acknowledged.

1. Introduction

To combat high and persistent unemployment and promote labour force participation, *active labour market programmes* (ALMPs) may provide a better alternative than income support for the unemployed workers. Through skill upgrading of the unemployed, ALMPs may improve the match between vacancies and unemployed and thus reduce wage-inflation, increase employment and decrease unemployment. During the 1990's annual public expenditures for ALMPs targeted at unemployed exceeded 1 per cent of GDP in many European countries, and the average annual participant inflow in these programmes was more than 5 per cent of the labour force. The Nordic countries are top ranked on this list, although the decline in unemployment has reduced Norwegian expenditures significantly in recent years.[2] The focus on ALMP in the Norway is also illustrated by a high share of active expenditures relative to total unemployment expenditures, see Calmfors, Forslund and Hemström (2001).

While the relative volume of such programmes is significantly lower, the public focus on *programme efficiency* seems to be stronger in the United States than in Europe. The international literature on evaluation of labour market programmes is extensive and growing, both in the United States and Europe, including the Nordic countries. A large part of this literature focuses on post-programme outcomes at the individual level, measured by employment probabilities, employment duration, or annual earnings. The major challenge for such evaluations is to report unbiased estimates of the causal impact of the treatment. In line with most microeconomic evaluation studies, this paper deals with individual effects only and more precisely, the average effect for the participants. A positive impact on the labour market success for participating individuals is a necessary, but probably not sufficient to achieve the overall macroeconomic goals: reduced wage inflation, increased employment and reduced unemployment. The net impact at the macro level also depends on any dead-weight loss, substitution and displacements effects, see e.g. Calmfors (1994) and Heckman, LaLonde and Smith (1999).

Very few, if any, of the programme evaluations question whether the individual programme effect depends on the state of the (local) labour market during the post-programme period. When no employers open new jobs, or if they do not open vacancies in response to voluntary quits, any improvement of skills through labour market programs will not help unemployed back to work. This extreme case is unrealistic, even during a slump, but it illustrates the possibility that job opportunities available in the post-programme labour market may affect the individual effects. On the other hand, when competition for vacant jobs is intensive, unemployed persons who upgrade their skills through ALMP may improve their job prospects significantly, compared to a situation where firms face labour shortages and hires "whoever" comes along. Consequently, programme effects may vary systematically over the *business cycles* at the national or local level, but the direction needs to be studied empirically. If business cycles matter, it may explain why short and long run effects differ simply because the macroeconomic, or even the local labour market conditions, change over time. Insight into the influence of business cycles on individual programme effects is highly relevant to policy-makers who aims at choosing an optimal timing and volume of ALMP. This knowledge can also turn out to be useful when assessing and comparing the estimated impact of various programmes across time and regions or even between countries.

In this paper we use non-experimental Norwegian data covering the years 1991-1997 to study the importance of business cycles on short term and medium term individual effects of a labour market training programme, the *LMT programme*. This is the largest labour market programme in Norway targeted at unemployed adults, offering classroom training in a large number of subjects, mainly vocational, but also some general subjects. The courses last typically 5 - 20 weeks. Similar programmes are found in many other countries.

In order to identify any impact of post-programme labour market conditions on individual programme effects, data containing labour market variation are needed. The typical evaluation considers one cohort, or a limited number of cohorts, of participants. Such studies have to rely on spatial (or regional) variation in labour market conditions since a given post-programme period is only observed under one set of macroeconomic conditions.[3] Thus, if the programme operates at a limited number of geographical locations, the impact of labour market conditions on the programme effects will not be identified.

[2] This observation is based on figures published by OECD in Employment Outlook 1990-2001 and is also shown in Calmfors, Forslund and Hemström (2001).
[3] "Time since programme" will be (perfectly) correlated with "calendar time effects".

Unlike most evaluation studies, we are able to disentangle the impact of post-training business cycles from the importance of the time span between the training and the post-training period. Using several cohorts of participants, we can estimate first, second and third year effects under different labour market conditions, even controlling for fixed regional effects.

It is well known that non-experimental evaluation methods may provide *biased estimates* of the impact of programmes. The conventional evaluation bias comprises the bias due to selection on unobservables as well as bias due to non-overlapping supports of the explanatory variables in the treatment sample and the comparison sample (mismatching) and different distributions of these variables within the two samples (misweighting). Evaluation methods based on *matching techniques* may reduce the conventional measure of bias - as far as selection on observables is concerned. Such estimators are increasingly being used in evaluation studies; in this paper we apply such estimators as well.

A commonly used conditioning set is the probability of being in the participant group versus in the comparison group. Provided that the outcome is independent of participation conditional on this probability the matching estimator is unbiased. This is the *conditional independence assumption*, CIA. Participation in a specific programme is, however, not the outcome of a simple binary choice, or of a selection process with only two mutually exclusive outcomes. First, the target group is often offered alternative programmes. Second, those who participate in programmes do also have more than one option – not only unemployment, but possibly also options as employment, education, retirement etc. Thus, matching the samples to be compared on *all* these propensities would make the CIA more plausible. In this paper we use probability scores matching estimators to assess the impact of a single programme by comparing participants (the treatment group) with unemployed non-participants (the comparison or no-treatment group). The two samples are matched by probabilities of (a) taking part in the programme to be evaluated, (b) taking part in other alternative programmes, and (c) leaving the unemployment register, all as alternatives to staying unemployed. To make each group of participants homogenous we conduct separate analyses for those starting LMT at about the same time, i.e. in winter or in autumn each year. This gives us 12 cohorts of participants (2 x 6 years). The data are cut by gender and unemployment benefit entitlement giving a total of 48 subsamples for which we estimate separate training effects.

Our analyses show that the impact of LMT on annual earnings is *positive.* With few exceptions, the effects are statistically and economically significant. The positive training effects persist. Even after three years, earnings of participants with recent work experience, i.e. those who receive unemployment benefits before the start of the training spell, are significantly higher than among the non-participants. Among participants without recent work experience, i.e those without UB entitlement, not all effects are statistically significant.

A meta-analysis of the large number of group- and cohort-specific training effects confirms that the average training effect on the trained do vary over the business cycle. Participants gain more when job opportunities in the post-training period are favourable. The effects are significantly lower when the national, or the local, labour market is characterised by a high unemployment rate and few transitions from unemployment to jobs.

The paper is organised as follows. Section 2 reviews some previous studies relevant for our study. Section 3 discusses briefly the evaluation problem, and section 4 presents the matching procedure. Section 5 presents the programme to be evaluated and the Norwegian labour market during the period covered by the study.

Section 6 presents the design of the study and the data. Data contain participants in LMT and non-participants during the period 1991-1996. The matching procedure and the outcomes of the matching for each of the 48 subsamples are presented in section 7. Section 8 presents the results of the effect evaluation, both first year effects for all the 48 subsamples, and later years effects as far as data on the outcome are available. In this section we also present a test for unobserved heterogeneity based on pre-training annual earnings.

Section 9 contains meta analyses of the estimated training effects for each the subsamples, focusing on how the impact of LMT on annual earnings correlates with job opportunities (the business cycle) at the national as well as the regional level. Section 10 concludes.

2. Previous studies

The international literature on evaluation of labour market programmes is extensive and growing, both in the United States and in Europe. A large part of this literature focuses on post-programme outcomes at the individual level. Barnow (1987), LaLonde (1995), Fay

(1996) and Heckman, LaLonde and Smith (1999) review much of the empirical results and the methodological discussions. No consensus about the impact of the active labour market programmes on individual success has emerged from the large number of evaluations in recent years. The content and the organisation of the programmes, the target groups and recruitment procedures as well as the economic environment at the time of the evaluation differ across the studies. There is also a large variety in evaluation design and estimating techniques. Thus there is no surprise that the results diverge. The mixed results may also reflect a lack of suitable data as well as robust estimation methods.

The general impression is that some, but not all programmes do have the intended impact, at least for some of the participants and in the short run. Large scale, low cost programmes perform not as good as more costly programmes, targeted at smaller groups of unemployed (OECD 1993, Martin 1998). Activation strategies especially targeted at people receiving unemployment benefits, encouraging them to intensify job search with later obligation to participate in programmes, have also shown evidence of increased job motivation and increased transitions to employment (Martin and Grubb 2001).

Nordic studies

Recently quite a number of studies in the Nordic countries have been published. Sweden has a long tradition with ALMPs, and the labour market training programme has been evaluated several times. This programme, AMU in Swedish, is quite similar to the Norwegian programme evaluated in the present study, the LMT programme, or AMO in Norwegian.

The early Swedish evaluations report positive impacts on employment probabilities and earnings. Axelsson (1992) evaluates the labour market training programme by comparing annual earnings (before taxes) for a sample of participants and a sample of unemployed non-participants in 1981. The analyses are based on non-experimental data within the framework of a fixed effect log-linear earnings model as well as by difference in differences. The overall impact is estimated to be positive and significant, and the second year effect turns out to be larger than the first year effect: about 9,000 and 7,000 Swedish kroner respectively.

Evaluations using data from the late 1980'ies onwards show, however, insignificant and even negative impacts of the labour market training programme. Many of the Swedish studies focus on impact for young participants, either in special programmes for youth or in ordinary training and employment programmes. The results for this group are also mixed, and in general not very positive: Ackum (1991) finds mostly insignificant effects for young participants, Korpi (1994) presents both significantly positive and insignificant effects, while Regner (1997) reports negative impacts of training programmes. Larsson (2000) evaluates two programmes for youth (20-24 years) and concludes that (in the short run, after one year) both seem to have negative impacts on employment and annual earnings (after two years the impacts are insignificant) – whereas the impacts on transition to ordinary education are mostly insignificant.

In a recent study, Sianesi (2002) applies a multiple-treatment matching framework to evaluate the differential performance of six main types of Swedish labour market programmes. This study covers 30,600 adults 25-54 who became unemployed for the first time during 1994 and who were eligible for unemployment benefits. The sample is followed until the end of November 1999, i.e. a post-training period of maximum 5 years. The differential performance of the six programmes – and the non-treatment state (waiting longer in open unemployment and searching for a job) - is assessed in relation to employment rates over time and the probability to be in a compensated unemployment spell. On average people who is in a programme (any programme) at a given moment subsequently enjoy higher employment rates than if they had postponed participation. Secondly, the best programme is clearly *employment subsidies*, not surprisingly as this is an arrangement based on a job promise by the programme employer – after completion of the programme. The employment probability is 40 percentage points higher about 7 months after entering the programme - compared with waiting. The impact decreases over time and is about 20 percentage points 60 months after entering the programme.

One of the six programmes evaluated is *labour market training*. Compared with waiting, participation in LMT is found to have a positive, significant effect on employment, increasing from about 5 percentage points 12 months after entering the programme to almost 20 percentage points 60 months after entering. Compared with the other five programmes LMT is the least effective when it comes to employment rates over time. Further details on the Swedish experience can be found in surveys by Björklund (1990), Zetterberg (1996), Ackum Agell & Lundin (2001) and Calmfors, Forslund and Hemström (2001).

Also in Denmark and Finland there are programmes similar to the Norwegian LMT programme. Jensen et al. (1993) evaluate the Danish LMT, which offers somewhat shorter courses (2-5 weeks), mainly targeted at employed, but open for unemployed as well. Effects on subsequent wage level and unemployment are analysed by fixed effects models. Wage effects are found to be small and insignificant in most cases. When it comes to effects on unemployment, the estimated models predict that participants with substantial pre-training unemployment will experience a decrease in post-training unemployment.

Westergaard-Nielsen (1993) evaluates the same programme for a different period and within a different framework. This study shows that training gives an overall positive impact on the wage level, significant for men – also for those with some unemployment experience - and insignificant for women. When it comes to subsequent unemployment, participation in LMT gives a small overall reduction for men, not for women. As Jensen et al. (1993), Westergaard-Nielsen (1993) finds that this is the case also for those with pre-training unemployment experience. However, for those with substantial unemployment experience, Westergaard-Nielsen (1993) finds that post-training unemployment increases. The Danish Ministry of Labour, AM (2000) and Westergaard-Nielsen (2001) recently evaluate effects of the Danish employability enhancement programmes. While AM (2000) is rather optimistic with respect to the individual effects, Westergaard-Nielsen (2001) is more sceptical when it comes to the efficiency of the active labour market policy in Denmark

Evaluation studies of ALMPs in Norway typically report more positive results than the evaluations in the other Nordic countries. For labour market training, evaluations indicate positive impacts on employment probabilities, see Torp (1994) and Aakvik (1998), while Raaum and Torp (2002) find positive annual earnings effects.

A business cycle perspective on programme effects

There are numerous reasons for why training can affect future earnings of the LMT participants. First, successful training helps trainees to accumulate human capital that is relevant to potential employers. Increased human capital may have a positive effect on wages as well as the probability of employment. However, if training increases the reservation wage, this may have the opposite effect on the employment probability. Secondly, as training represents a meaningful activity to most participants, it may help to prevent social isolation and mental problems during a period of non-employment. This may in turn enhance job search efficiency and reduce the probability that unemployed workers drop out of the labour force. Thirdly, LMT may represent a signal about unobserved characteristics like motivation and effort, which correlates with productivity. Potential employers may consider a personal unemployment record, which include LMT to be better than a record with only open unemployment. This “signalling effect” of LMT is crucially dependent on the reputation of the programme. Programmes associated with long-term or low-qualified unemployed may give a negative signal to employers. Finally, training has an alternative cost as time available to ordinary job search activities is reduced. Various empirical studies show that labour programme participants have very low transition rates to ordinary employment during the programme period; see e.g. Røed and Zhang (1999).

Turning to the impact of labour market conditions, the location in the business cycle may influence active labour market programmes in different ways. First, the composition of the eligible population, typically unemployed adults, may change with respect to observed and unobserved characteristics as both demand and supply of labour change. Secondly, the recruitment process may change. This applies to both self-selection (who wants to participate?) and the administrative selection, reflecting changing priorities in the implementation of labour market policy. Finally, the state of the local or national labour market and the demand for labour in the post-training period may affect the impact of training on individual outcomes, e.g. earnings.

The purpose of the present study is to make identical evaluations of a programme at various points in time over a business cycle. The LMT programme is well suited for studying how the state of the labour market, i.e. business cycles, affects the impact of ALMP. First, it has a fairly long record and it has been operated at a significant volume every year in the period of interest - even when unemployment was as low as 3 per cent of the labour force. Second, the eligibility criteria are quite simple and have mainly been the same in the whole period. Participants have to be unemployed and to register at the local PES (public employment service), they have to be 19 years or older (our study includes only persons 25-50 years) and employable, i.e. not vocationally disabled and ready to take a job. Thus, even if the mix of courses has changed over the business cycle, e.g. more general training during the slump and training more targeted at specific need in the market during the boom, the evaluated programme is essentially the same throughout the first half of the 1990s.

Assume an unbiased estimator for the average treatment effect of the treated is identified. Any variation in the estimated effect over the business cycle is then a mix of changes in the composition of the treatment group (assuming effect heterogeneity) and variation in demand for labour across post-programme periods.

In the present study, to overcome some of this possible composition effects, we specify four different treatment groups: men and women, entitled and not entitled for unemployment benefits. It turns out that the observable composition of each treatment group is rather stable over the period. Thus, interpreting variations in the estimated effect over the evaluation period we focus on the last point, i.e. changes in the demand for labour.

Our hypothesis is that the effects of ALMPs are more positive (less negative) during a boom than during a slump. When employment is increasing employers have to recruit from outside the market: young entrants, re-entrants and unemployed. Among unemployed we believe employers will prefer job applicants with some programme experience compared with other unemployed with the same characteristics. Decreasing employment and increasing unemployment means a low turnover rate and very few job openings. Thus even for the best qualified among the unemployed the employment probability is low during a slump.

This kind of business cycle impact is probably stronger for effects of programmes emphasising *quick-job-entry*, as intensified employment service and job search training. It is probably less strong for *human capital development* programmes focusing on basic skills and vocational training. As we only evaluate one programme we are not able to test this hypothesis.

3. The evaluation problem

There are various concepts of causal effects – even for a specific and well-defined treatment and for a given outcome. First, the treatment in question needs to be contrasted with *an alternative* treatment or to non-treatment. Second, we have to specify for *whom* we evaluate the impact, whether it is the average effect for a specific group or the whole distribution of effects.

Denote Y_1 as the given outcome at the relevant point in time conditional on the specific treatment of interest, and denote Y_0 as the outcome conditional on non-treatment, or the alternative treatment. Defining the impact as the difference between these two, we get (Y_1 - Y_0.) Thus the causal impact of the treatment does not only rely on the specification of the treatment to be evaluated. The definition of the non-treatment status is just as important.

For each person only one outcome is observed. Thus whether we want to estimate the expected impact for *any* potential participant, for those *not* participating, or for those who *do* participate, we need to estimate or simulate the counterfactual outcome.

Assume we have cross-sectional data. Let D = 1 for those in the treatment group and let D=0 for those in the non-treatment group. Let X be a vector of observed characteristics. Assume the outcome Y depends on X and D, as well as an unobserved error term U:

(1a) D= 1: $Y_1 = a_1X + U_1$

(1b) D= 0: $Y_0 = a_0X + U_0$

The most common evaluation parameter of interest is the mean impact for participants or *average (expected) treatment effect for the treated* (ATET).[4] The ATET is the expected difference between Y_1 and Y_0, conditional on D=1, given by

(2) $\Delta(X) = E(Y_1 - Y_0 \mid X, D=1) = E(Y_1 \mid X, D=1) - E(Y_0 \mid X, D=1)$

To identify this parameter we have to predict Y_0, because this is not observable for D=1. Given model (1) the effect $\Delta(X)$ defined by (2) is a mix of structural effects $\{ a_1 X - a_0 X \}$ and error terms $E(U_1 - U_0 \mid X, D=1)$.

There are many methods of constructing the unobserved counterfactual $E(Y_0 \mid X, D=1)$. One common method is to use the outcomes of non-participants (or participants in the alternative treatment) as a proxy, i.e. $E(Y_0 \mid X, D=0)$. However, comparing participants and non-participants for instance in a standard regression analyses, i.e. comparing the expectations $E(Y_1|X, D=1)$ and $E(Y_0|X, D=0)$, we may get a biased estimate of $\Delta(X)$. This *selection bias* is given by

(3) $B(X) = E(Y_0 \mid X, D=1) - E(Y_0 \mid X, D=0)$

[4] Other parameters of interest are for instance the average (expected) treatment effect for a person drawn randomly from the eligible population or the expected effect for a person drawn randomly from the combined sample of participants and non-participants. In addition it is of interest to assess the whole distribution of effects: What fraction of the participants benefits from the treatment, and what is the effect for those in the left-hand-side tail of the outcome distribution?

B(X) is rigorously defined only for values of X common to D=1 and D=0. Conditional on this X the bias rigorously defined is due to genuine differences in the distributions of the error terms (unobserved differences).

The *conventional evaluation bias* (LaLonde 1986) defined by $B = E(Y_0|D=1) - E(Y_0|D=0)$ is analogous to selection bias B(X) given by (3) but does not condition on X. Heckman, Ichimura, Smith and Todd (1998) show that the conventional evaluation bias comprises the selection bias rigorously defined as well as bias due to non-overlapping supports of X in the two samples (mismatching) and different distributions of X within the two samples (misweighting). Heckman, Ichimura and Todd (1997) demonstrate that, in the Job Training Partnership Act (JTPA) study, bias due to selection on unobservables is empirically less important than selection due to lack of matching on X for the samples of participants and non-participants.

In the jungle of complicated econometric evaluation models, it is important to keep in mind one of the fundamentals in empirical research; "Good data help a lot".[5] From assessments of evaluation strategies on US data, there seems to be a consensus that some features are of particular importance. Heckman, Ichimura and Todd (1997) summarise these as follows:

(I) Participants and controls have the same distributions of unobserved attributes. (II) Participants and controls have the same distributions of observed attributes. (III) The same questionnaire is administrated to both groups, so outcomes and characteristics are measured in the same way. (IV) Participants and controls are placed in a common economic environment.

In the present study of LMT, we estimate the ATET where the treatment and the non-treatment groups are sampled from the same populations. All persons are fulltime unemployed, registered at the local branch of PES, at the same time, i.e. taking care of (IV). Information on all groups is collected in the same way and from the same sources without sample attrition (administrative registers), i.e. fulfilling (III). The matching procedure described in the next section takes care of feature (II).

[5] This has indeed been stressed by e.g. Heckman and his colleagues in numerous contributions over the last ten years.

4. Matching

The logic of matching is to re-establish some of the features characterising experimental data when we actually use non-experimental data. By matching we construct samples of participants and non-participants to ensure that they meet certain conditions related to independence between the outcome (or the effect to be evaluated) and treatment status. The brief presentation to follow leans heavily on Heckman, Ichimura and Todd (1997, 1998).

Assume that the outcomes (Y_0, Y_1) and the treatment status D are statistical independent conditional on X. (This X-vector may be the same or another than the X-vector in the outcome model.) Thus

(4) $(Y_0, Y_1) \perp\!\!\!\perp D \mid X$

This is equivalent to $\text{Prob}(D=1 \mid Y_0, Y_1, X) = \text{Prob}(D=1 \mid X)$, which rules out the Roy model of self-selection. In addition, assume that

(5) $0 < P(X) = \text{Prob}(D=1 \mid X) < 1$

By (5) we exclude cases of $P(X)=1$ and $P(X)=0$, i.e. persons with X-values that ensure they will always or never receive treatment. Such persons are not possible to match with persons from the other group. According to Rosenbaum and Rubin (1983) condition (4) is the *ignorability condition* for D, while together with (5) it constitutes the *strong* ignorability condition.

Conditions (4) and (5) are, however, stronger than what is necessary to estimate ATET. To identify $E(Y_0 \mid X, D=1)$ it is sufficient to assume

(4') $Y_0 \perp\!\!\!\perp D \mid X$

(5') $P(X) < 1$

(4') is called the conditional independence assumption (CIA). This does not rule out the dependence of D and Y_1. To get an unbiased estimate of ATET it is sufficient with the even weaker assumption: $E(Y_0 \mid X, D=1) = E(Y_0 \mid X, D=0)$

Assume the X-variables that meet the conditions (4') and (5') are identified. Thus by *matching* the two subsamples on these variables we eliminate the bias in the $\Delta(X)$ estimator given by (2), but only the bias due to observables.[6] Provided that the CIA holds, we have B(X) = 0 for the matched samples. If CIA does not hold, other estimation methods may eliminate selection on unobservables. Difference-in-differences will for instance eliminate selection on person specific, time-invariant unobservables.[7] When the number of matching variables (observed variables that may affect the relation between participation status and outcome) is very large, multivariate matching on explanatory variables is hard to handle.

Rosenbaum and Rubin (1983) show that if CIA holds, matching the two samples on the *propensity score* P(X) is sufficient to secure unbiased estimates. They show that (for random variables D, Y and X) when Y_0 is independent of D conditional on X, Y_0 is also independent D, conditional on P(X)= Prob(D=1|X).

If the propensity score is smaller than one, then $E(Y_0| D=1, P(X)) = E(Y_0| D=0, P(X))$. Thus, if P(X) is known or if it can be parametrically (or semi-parametrically) estimated, we may match the two samples on the univariate propensity score.

The propensity score matching methods are further developed by Heckman, Ichimura and Todd (1997, 1998), see also Heckman, Ichimura, Smith and Todd (1998), Imbens (2000) and Lechner (2001a). Empirical implementations of the various estimators are found in some of the same papers as well as in Deheija and Wahba (1998, 1999), Brodaty, Crepon and Fougere (2001), Smith and Todd (2002), Larsson (2000) and Lechner (2001b).

Although increasingly popular, the propensity score matching technique is not necessarily an easy way to obtain non-biased estimates using non-experimental data. For instance, Smith and Todd (2002) find little support for claims by e.g. Deheija and Wahba (1998, 1999), about the effectiveness of these estimators as a method for controlling for selectivity bias. They find that various cross-sectional matching estimators are highly sensitive to the choice of sub-sample and to the variables used to estimate the propensity scores. Smith and Todd (2002) also find that difference-in-differences matching estimators

[6] Matching here means pairing each programme participant with one (or several) non-participant, selected from the population of non-participants (without or with replacement). The pairs are constructed on the bases of identity or similarity in the X variables. The mean impact of the treatment on treated is then estimated by the mean differences in the outcomes of the matched samples.
[7] We do not estimate difference-in-differences for two reasons. First, it turns out that pre-training earnings differentials between participants and non-participants are very small and statistically insignificant. Second, pre-training earnings are not observed for all cohorts in the data.

may perform better. As an explanation they point at possible problems with the data, for instance that the features (III) and (IV) mentioned above, are not achieved.

Of special interest for our study is the extension of the method from a conventional two-state framework to allow for the case with *multiple mutually exclusive states*, developed by Imbens (2000) and Lechner (2001a). Lechner (2001a) presents a *matching protocol*, suggesting a specific algorithm – with some variants - in four steps for estimating the treatment effects. As pointed by Lechner (2000a) this algorithm does not give asymptotically efficient estimators, because the trade-off between bias and variance is not addressed (the algorithm minimises the bias). See Hirano, Imbens and Ridder (2000) for a discussion of efficiency of estimators based on propensity score matching. More sophisticated and computer intensive matching estimators - that also control for *un*observables - are discussed by Heckman, Ichimura and Todd (1998).

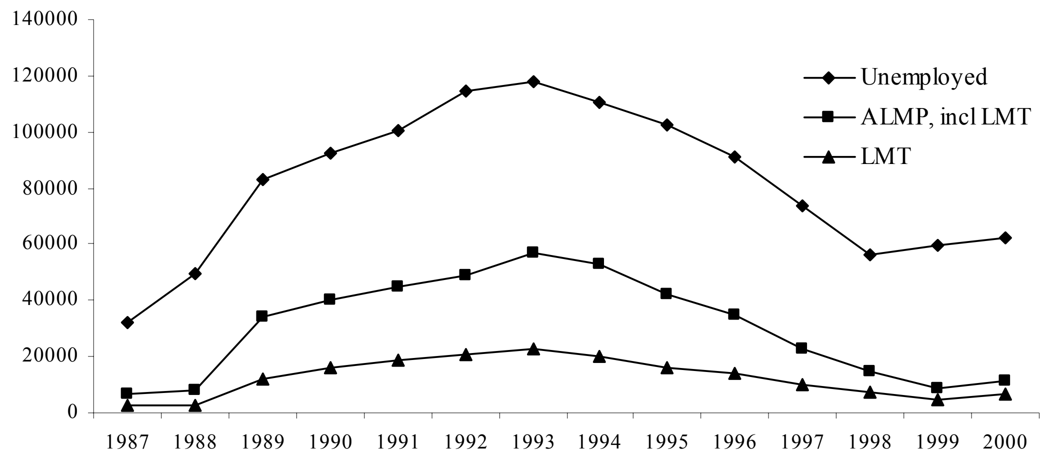
5. Labour market policies and the business cycle in Norway during the 1990s

During the period covered by our data, 1990-97, unemployment has fluctuated as illustrated in Figure 1. In 1990 unemployment was relatively high by Norwegian standards and increasing with a peak in 1993. The unemployment rate peaked in 1993 at 5.5 per cent, increasing from 1.5 in 1987 and sliding back to 3.3 per cent in 1997 and 2.4 in 1998.[8]

The average number of persons involved in ALMPs increased from 7,000 in 1987 to 57,000 in 1993, showing how ALMPs are used to dampen the labour market effects of business cycles. During the bottom of the slow-down, 2.5 per cent of the labour force participated in these programmes. As most ALMPs last for less than half a year, the total number of persons participating in programmes *during* one year is about twice the participation rate at a point in time. From 1993 to 1997, the average participation in ALMPs decreased from 57,000 to 23,000. In 1999 the number of participants was as low as 8,000.

[8] Open unemployment are persons registered as fulltime unemployed and searching for a job, source Directorate of Labour. These figures deviate from the statistics published by OECD, which are based on the Norwegian Labour Force Survey.

Figure 1. Unemployment, participants in active labour market programmes (ALMP) and training (LMT). Persons, annual average. 1987-2000.



The *Labour Market Training* programme is by far the largest programme, covering about 40 per cent of all ALMP-participants. The aim of LMT is to maintain and improve the skills of the unemployed and thereby to enhance their employability. The programme is organised as off-the job courses, mainly targeted at unemployed adults. Moreover, a substantial number of people (re-)enter the labour market via the training programme.

In the first training year of this study, 1991, the average number of participants was 19,000. Then it increased to 23,000 in 1993, before it started to decrease: in 1996 the average number of participants was 14,000, in 1999 only 4,500.

The programme is funded by the central government and organised by the local employment service under the supervision of the Directorate of Labour and the Ministry of Labour. The courses are provided by the employment service, often in co-operation with other public and private institutions. Vocational training is dominant and a wide range of subjects and crafts are covered. Most of the courses are short, from 5 to 20 weeks. In some cases there are basic courses and follow-up courses within the same subject, with a total duration of one year (or even more). LMT is available for all job seekers and participation is voluntary.[9] Unemployed persons who refuse to accept offers of training may lose their unemployment insurance benefit. This sanction is, however, rarely carried out.

The courses are free of charge. All participants are entitled to a training allowance, but recipients of unemployment benefits (UB) may opt to collect their benefits. Participants eligible to unemployment benefits typically keep these as the training allowance is lower. UB compensates about 62.4 per cent of previous earning, while the allowance is flat rated. Economic incentives to participate in LMT are driven by the training allowance, but also related to the eligibility and exhaustion of unemployment benefits. Time spent in LMT and the allowances collected *do not* qualify the participants for unemployment insurance.[10] But as unemployed not eligible for UB receive training allowances, they sure have economic incentives to take part in LMT.

The capacity of most courses is limited. The rate of rationing at each course depends on the number of qualified applicants related to the capacity of the course. Thus the recruitment to LMT is partly a self-selection process and partly an administrative selection process. Røed, Torp, Tuveng and Zhang (2000) have studied this recruitment process. Based on register data as well as interviews with the administrative staff at local branches of PES they point at a possible trend towards *positive selection to LMT*, i.e. participants have observed – and may be also unobserved - characteristics assumed to correlate positively with employability. They find it difficult to draw any conclusions on how this selection changes over the business cycle. It seems, however, that the positive selection is weaker when unemployment is low (as in 1997-1999) than when unemployment is high (as in 1992-1994). On the other hand the staff at PES reports that during the slump the capacity of LMT was sufficient to offer training to "everyone". During the boom the administrative staff had to be more selective. Courses directed at expressed needs of labour among employers were given priority, as were unemployed expected to be able to fill manifest vacancies.

Selection to LMT and the variation in selection over the business cycle may be captured by *observable* characteristics of the participants and the non-participants.

[9] For some courses applicants have to qualify through education, previous vocational training or work experience to be eligible.
[10] According to the Norwegian system it is necessary to have earnings from an ordinary job to qualify for unemployment insurance benefits. Until 1997 earnings received during a temporary employment programme (but not in a training programme as LMT), counted as qualification for future unemployment benefits.

Matching estimators of ATET will then be unbiased and any variation in the estimates over the business cycle will reflect that training effects do depend on labour market conditions in the post-programme period. However, if the difference is tied to *un*observables it will cause biased estimates of the effects, and this bias may change over the business cycle. Assume the positive selection of participants to LMT is weaker during the boom than during the slump, as indicated by Røed et al. (2000). Then we would expect the estimated effects to be less upward biased in the boom than in the slump. Thus the influence of a change in the selection bias will be to partly disguise any positive association between training effects and post-training job opportunities for the unemployed.

6. Data and design of study

The data are drawn from a large Frisch Centre database containing individual level information from numerous administrative registers, delivered by Statistics Norway. We select individuals from all entrants (and re-entrants) in the public unemployment register during December 1990 – July 1996. Our data from this register contain monthly observations of unemployment, labour market programme participation by type and unemployment benefit entitlement. From this population we select 12 cohorts of LMT participants, two cohorts every year from 1991 to 1996.

We use annual labour earnings 1992-1997 measured in Norwegian 1997 kroner to estimate the impact of the programme. A large number of individual pre-training characteristics are available. The comprehensive and detailed data sources constitute a solid basis for evaluating the effects of the LMT program throughout the first half of the 1990s. The data enable us to study the extent to which training effects vary with the state of the labour market, i.e. job opportunities, and how training effects evolve as post-training time prolongs. In this section we describe the data which then is used to model the selection into training and the creation of comparison groups of non-participants (section 7) as well as estimating the training effects (section 8).

Participants and non-participants

LMT courses typically start in August or September and then there is another wave of courses starting in January and February. The composition of training courses does not differ substantially between the autumn and winter seasons. As the majority of courses last for 5-20 weeks, most courses in the autumn are completed by the end of the year, but in some cases continuation courses start early next year. Most winter courses end before the summer, while some continue after the summer holiday. Since the post-programme success of the training is measured by annual earnings, and the time passed after having completed the training may affect the impact on earnings, it is preferable to analyse the impact of autumn and winter courses separately. We then have 12 (training year*season) cohorts, where each cohort is split into four groups by gender and unemployment benefit entitlement. We restrict ourselves to participants aged 25-50, since selection into other programmes and education, as well as labour market behaviour in general, are different for teen-agers and young adults. The upper age limit set is to avoid transitions out of the labour force due to early retirement or disability pension which become increasingly important as we include unemployed in their fifties and sixties.

The participants and non-participants are defined by the same procedure across cohort groups. The population of potential LMT participants consists of all fulltime unemployed persons registered at the end of December and July, for the winter and autumn cohorts respectively.[11] Then we consider the register status two months later, i.e. at the end of February and September, respectively. LMT participants constitute the treatment group. In order to define a suitable comparison group we divide non-participants into three groups according to their status in the register; still unemployed (U), participating in another labour market programme (PROG), or having left the unemployment register (OUT). Those who leave enter jobs or exit from the labour force, but we cannot distinguish between the two transitions.

The comparison group is selected among those still unemployed.[12] From these individuals we select non-participants who are "observationally equivalent" to the participants, as far as pre-training characteristics are concerned. The logic behind this matching, how it is implemented and the results of procedure are described in following section.

[11] By this sample restriction we exclude LMT participants who enter training directly from outside the register.
[12] Lechner (2001b) estimates the impact of four different programmes on employment (relative to non-participation), measured by number of days employed during limited a post-programme period (per cent), by data from the Swiss canton of Zurich. The paper presents and compares different estimators of the causal impact. It is shown that effect based on a comparison of a treatment group to an aggregated comparison group of individuals has no meaningful causal interpretation, while pair-wise effects give clear-cut causal effects.

In Table 1 we report the sample sizes of the different cohorts, by group. The sum columns two to four constitutes the populations at risk, defined as the members of unemployment stock two months before and each column shows how they are distributed according to LMT, PROG and OUT transitions. The U-group is those still unemployed. Cohort W91 consists of all fulltime unemployment in December 1990 and their status at the end of February 1991, cohort A92 consists of all fulltime unemployment in July 1991 and their status at the end of September 1991, and so on.

The samples of participants vary between 600 and 2,500 individuals. Unemployed without unemployment benefits (No UB) are more likely to enter training. Among those with UB, men and women are equally likely to participate in training. For those without UB, more women than men enter the training programme.

Table 1. Sample sizes and transitions from fulltime unemployment, 1991-1996.

Males, UB					
	No trans.	Transition to			
Cohort	U	OUT	PROG	LMT	LMT rate
W91	12427	4794	704	1238	0.065
A91	16079	7856	1295	2185	0.080
W92	19699	5806	989	1558	0.056
A92	19144	8643	1612	2491	0.078
W93	21442	5374	1362	1900	0.063
A93	18157	8842	2379	2206	0.070
W94	18882	5322	1559	1566	0.057
A94	14250	8269	1981	1777	0.068
W95	14594	4664	966	862	0.041
A95	12034	6887	1402	1484	0.068
W96	11445	4161	889	836	0.048
A96	10325	5781	941	1086	0.060

Females, UB					
	No trans.	Transition to			
Cohort	U	OUT	PROG	LMT	LMT rate
W91	8054	4328	545	907	0.066
A91	10636	7303	766	1964	0.095
W92	11298	3850	509	988	0.059
A92	13368	8538	1151	2110	0.084
W93	12831	4075	906	963	0.051
A93	13014	8872	1739	1752	0.069
W94	12141	3717	1022	853	0.048
A94	11714	8502	1816	1893	0.079
W95	11055	3815	744	695	0.043
A95	11365	8312	1404	1618	0.071
W96	9631	3951	725	711	0.047
A96	10495	7704	1192	1351	0.065

Males, No UB					
	No trans.	Transition to			
Cohort	U	OUT	PROG	LMT	LMT rate
W91	3818	2307	317	638	0.090
A91	4688	3068	512	1055	0.113
W92	5422	2310	463	733	0.082
A92	5427	3384	574	1132	0.108
W93	6065	2461	385	807	0.083
A93	6828	3562	740	1127	0.092
W94	7441	2954	543	844	0.072
A94	7183	3990	884	1237	0.093
W95	7933	2940	492	634	0.053
A95	7227	4071	763	1186	0.090
W96	6679	3355	467	685	0.061
A96	6902	3713	654	982	0.080

Females, No UB					
	No trans.	Transition to			
Cohort	U	OUT	PROG	LMT	LMT rate
W91	2315	1684	237	595	0.123
A91	3396	2671	495	1540	0.190
W92	3429	1792	345	760	0.120
A92	3901	2860	518	1705	0.190
W93	3928	1923	339	776	0.111
A93	4756	3171	782	1561	0.152
W94	4628	2349	451	735	0.090
A94	5112	3428	949	1914	0.168
W95	5988	2325	393	734	0.078
A95	6060	3755	874	1942	0.154
W96	5050	2892	465	857	0.093
A96	6192	4019	835	1762	0.138

W9j = Winter 199j , A9j = Autumn 199j, j = 1,..6.

The transition probabilities are estimated for each of the 48 sub-samples as functions of a large number of individual characteristics. From the unemployment register we collect information on pre-training labour market program participation, unemployment record, previous occupation and unemployment benefit entitlement (UB). In addition we have register information on age, gender, material status, number of children, educational attainment, work experience measured by yearly pension points (proportional to earnings), immigrant status and school enrolment during the previous six months. Fixed county of residence effects, measured at the time when training starts, is used to control for variations in local labour market conditions and supply of labour market programmes. All these variables are used to model the transitions from unemployment, including enrolment into LMT, see section 7. More information on the individual characteristics is given in the Appendix.

Earnings profiles of participants

According to the Norwegian labour market authorities, the main objective of LMT is to increase the ability of unemployed to get permanent jobs. Even if employment is the overall goal of the programme, several arguments favour the use of *post-training earnings* to measure programme effects. The first argument is *relevance*. Post-training earnings in year t (Y_t) can be decomposed into days of employment (e_t), average hours per day employed (h_t) and average wages per hour (w_t), which gives: $Y_t = e_t h_t w_t$. Here e_t measures how quickly the person enters employment as well as the stability of the job. h_t depends on opportunities, qualifications and preferences of the individual. Part-time unemployment is common among LMT applicants, indicating that many are rationed with respect to working hours. If the training effect on earnings is due to longer daily working hours, this should be considered as a success in line with (re-)employment. The hourly wage reflects productivity and the quality of the employment match. If LMT contributes to more productive employees and a better matching, these effects are obviously socially beneficial. As the Norwegian wage structure is fairly compressed; see e.g. Barth and Zweimüller (1994), earnings mainly reflect the duration of employment. If there is a positive effect of training on earnings we do not expect wage increments to be an important explanation. Finally, cost-benefit comparisons also favour earnings as a measure to evaluate the programme effect.

In line with the objective of LMT, earnings should include wages as well as income from self-employment. Transfers, unemployment benefits, social support and training allowances ought to be excluded. Our data on earnings are collected from public tax- and wage-registers. Unemployment benefits are subtracted, implying that our earnings are very close to income from work, including earnings from self-employment.[13]

Since Ashenfelter (1978), studies of programme effects have been concerned about the earnings dynamics. Participants typically experience that earnings drop prior to the training period and gradually increase during the post-programme period. Figures 2 and 3 illustrate the mean earnings profiles of the 1995 cohorts, by gender and unemployment benefit status.

The "Ashenfelter-dip" is clearly experienced by those with unemployment benefits. The earnings profiles of the groups without benefits are clearly very different, illustrating that LMT is a potential stepping stone in the process of (re-)entering the labour market. One might also suspect that financial incentives (i.e. training allowances) make it economically wise to spend time on LMT during this process, even if the effects on future labour market prospects from this investment are minor. Anyhow, Figures 2 and 3 clearly motivate our split by gender and unemployment benefit eligibility when we estimate earnings effects of LMT.

Short and medium run effects

The earnings data presently available cover the years 1992-1997. Consequently, we estimate first year (short run) effects for all cohorts, while second and third year effects can only be estimated for the first ten and eight cohorts, respectively. The Norwegian business cycle turned some time during 1993, which means that the variation in job opportunities is somewhat limited when we consider the effects beyond three years. In a companion paper, Raaum, Torp and Zhang (2002b) we compare individual long run effects and direct programme costs of LMT, focusing on participants in 1992 and 1993.

[13] The available measure of income, "Income Qualifying for Pension" (PI), includes unemployment benefits and wage earnings from various labour market programmes. Training allowances are not included. We adjust the PI for unemployment benefits, but earnings from participation in other labour market programme than training are difficult to sort out and are therefore included.

Figure 2. Earnings profiles. Participants Winter 1995. By gender and unemployment benefit.

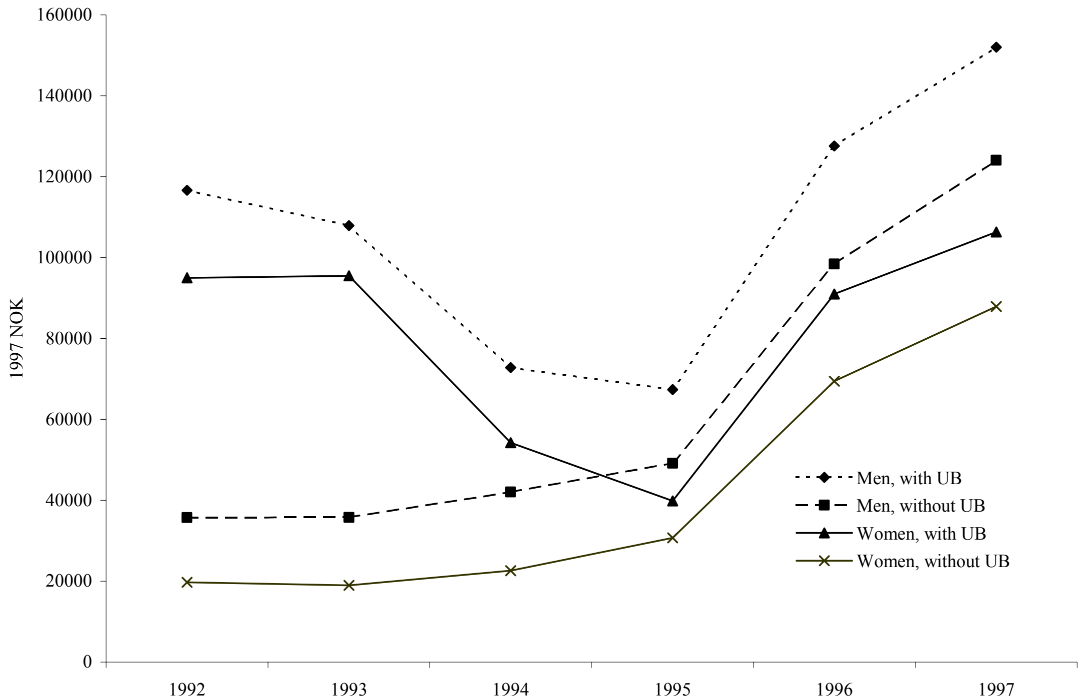
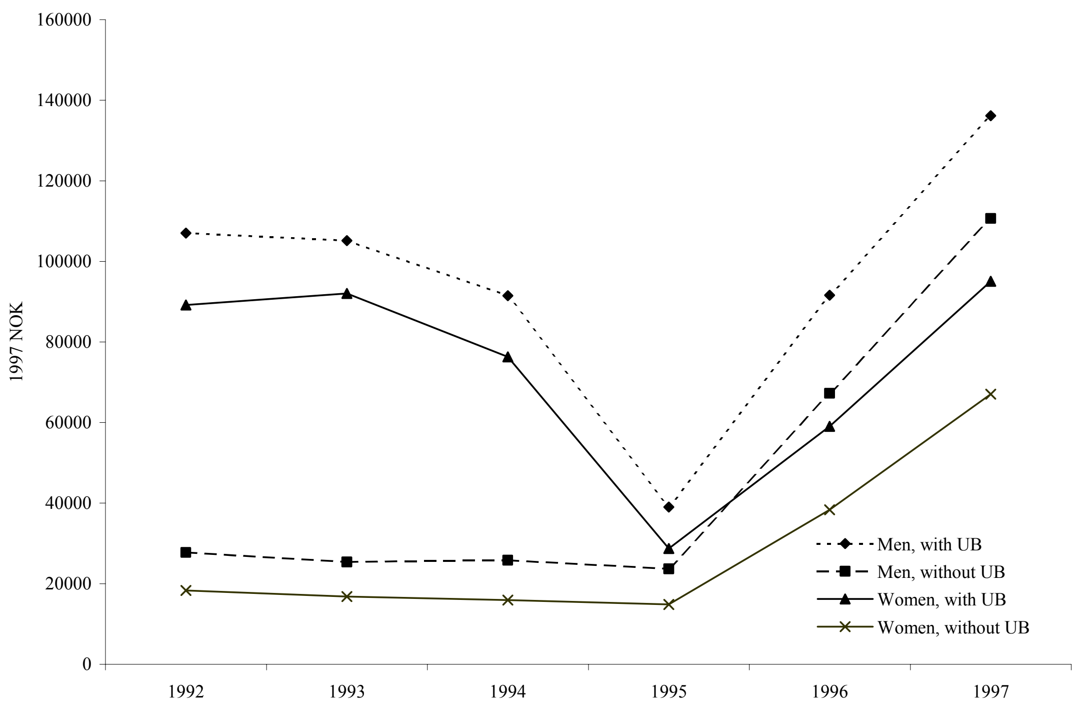


Figure 3. Earnings profiles. Participants Autumn 1995. By gender and unemployment benefit.



7. Selection on observables and matching

This section describes how the comparison groups of non-participants are established and used to simulate the counterfactual outcome of LMT participants, i.e. $E(Y_0 \mid X, D=1)$. There are various matching techniques and estimators used in the evaluation literature. In this study we apply a variant of traditional pair-wise nearest-neighbour-matching in the case of a multinomial choice model, inspired by the *matching protocol* suggested by Lechner (2001a).

We start out with the population of all fulltime unemployed, registered at time *t,* who are eligible for the LMT programme. Each member faces several options, here specified as four mutually exclusive states at time *t+dt*: to remain unemployed (U), to take part in the programme to be evaluated (LMT), to take part in another programme (PROG), or to leave the unemployment register (OUT). We use a multinomial logit model to estimate the probabilities of each state at time *t+dt*, as functions of a large number of individual characteristics at time *t*. *dt* is approximately two months. As we have 12 cohorts split by gender and unemployment benefit status, we estimate and predict probabilities for 48 different samples.

The estimated parameters from the multinomial logit model are used to predict the probabilities of LMT, PROG and OUT for each individual in the subsample of participants (LMT=1) as well as for all those still potential participants when the programme starts, i.e. unemployed non-participants (U=1).

To eliminate as much as possible of the potential selection bias, we select unemployed non-participants with the same predicted structure of transition probabilities as those in the treatment group. The first step in the matching procedure is to exclude observations outside the common support, i.e. we exclude observations from the sample of participants with estimated probabilities that are larger than the maximum value of the same probabilities in the comparison group. Similar we exclude observations from the LMT-sample with estimated probabilities that are smaller than the minimum value of the same probabilities in the comparison group. Then we use the same procedure to exclude observations in the comparison group with estimated probabilities outside the range of the probabilities in the LMT-sample.[14] This defines the common support samples.

Next we take each observation from the sample of participants and search through the comparison group to find the closest match based on the three estimated probabilities. In this process we use the *Mahalanobis metric* as a measure of distance with the inverse covariance matrix from the original gross sample as weights; see Rubin (1979). Since we keep the matched comparison group member (i.e. matching with replacement), a single observation may be used several times. In our case, however**,** a limited number of non-participants are used more than once. Across cohorts and groups, 5 to 12 per cent are used twice, up to 3 per cent are used three times as control while up to 2 per cent are used three times or more.

What explains participation?

In a separate working paper, Raaum, Torp and Zhang (2002a), we report the estimates of the multinomial logit model for selected subsamples, women and men, with and without unemployment benefits for some cohorts. The observables used to estimate the propensity scores are defined in the Appendix of this paper. The estimations show that various explanatory variables have some influence on the transitions from unemployment to the three other states. The partial impact of most variables differs, however, across subsamples.

When it comes to the relative probability of LMT, there are some robust patterns. First of all, the probability is higher for those who participated in LMT the previous quarter as well, ceteris paribus. This parameter is significantly positive for most subsamples. Next, those with a fairly long unemployment record, i.e. 11 months or more, are less probable to participate in LMT (relative to stay in U). We also find that for a majority of the subsamples the relative probability of LMT is larger for immigrants than others. This may mirror the fact that LMT includes special courses target at this group. The partial impact of age seems to be negligible (ceteris paribus), even if those aged 46-50 years are less apt to participate in LMT for some subsamples. Education is somewhat more important, as low education (10

[14] We compare one probability at the time: First we accept all observations from the comparison group with estimated values Prob(LMT=1| X) within the range of estimated values of Prob(LMT=1|X) for the participant group. Next we accept all observations from the treatment group with estimated values Prob(LMT=1| X)

years or less) and unknown education is negatively correlated with the relative probability of LMT. We also find substantial regional differences. For many of the subsamples the relative probability of LMT is larger in the northern counties of Norway (Finmark, Troms, and Nordland) than in the southern and central parts. This illustrates the importance of comparing participants and non-participants from the same location if we are to eliminate the misweighting on observables.

Matching results

We assess the success of our matching procedure in two ways. First, we compare the distributions of the predicted probabilities among (i) the participants, (ii) all potential unemployed non-participants and (iii) the unemployed non-participants picked by the matching procedure. Next we compare mean predicted probabilities and average pre-training characteristics among participants and the matched non-participants.

As illustrated in Table 1 the number of observations in the group of unemployed non-participants is much larger than the number of observations in the group of participants. This holds for all 48 subsamples. This simplifies the matching. The common support criteria (based on the predicted probabilities) leaves out rather few observations. Across cohorts and groups, less than 5 per cent of the unemployed non-participants (on average about 1.5 per cent) and less than 2 per cent of the participants (on average about 0.5 per cent) are excluded because they do not meet this common support criterion, see Raaum, Torp and Zhang (2002a) for details.

Figures 4 and 5 present plots of the predicted probabilities of Prob(LMT=1), Prob(OUT=1) and Prob(PROG=1) for the two 1991 cohorts, respectively.[15] In each panel there are three lines. The line marked with *squares* is for all unemployed non-participants. For the predicted values of Prob (LMT=1) this line is (in general) to the left of the two other lines.

within the range of estimated values of Prob(LMT=1|X) for the comparison group. Then we proceed with similar comparisons of estimated values Prob(OUT=1|X) and Prob(PROG=1|X) for both samples.

[15] Plots are estimates of *Epanevhnikov Kernel* densities on predicted probabilities P_i(I=LMT,OUT). Bandwidth is estimated by $h=0.9m/(n^{1/5})$, where m=min(sqrt(variance(p_i), interquartilerange(p_i)). The densities are estimated with *STATA,* see “Reference Manual, [R] kdensity” (2001), *Stata Statistical Software, Release 7.0*, StataCorp.

The thicker left-side tail indicates more people with a low probability of LMT=1. When it comes to the predicted values of Prob(OUT=1) the difference between the three lines is not as large (and the line marked with squares is often to the right, indicating more people with a high probability of OUT=1).

The two other lines are for the matched samples, non-participants marked with *triangles* and participants marked with *circles.* As can be seen these two lines are very close both for predicted values of Prob(LMT=1) and the predicted values of Prob(OUT=1). The closer the lines, the more successful is the matching with respect to the propensity scores.

Figure 5 presents similar panels for the subsamples of Autumn 1991 cohort and reveals that patterns are quite stable across cohorts.[16] For all subsamples the matching seems pretty successful, see Raaum, Torp and Zhang (2002a) where similar plots for all cohorts are included.

The success of the matching procedure can also be assessed by studying the differences in mean propensity scores and X-variables included in the multinomial choice model. In Appendix we present mean predicted propensity scores as well as mean values of X-variables *after matching* for selected cohorts, in Tables A1 and A2, respectively. Generally the predicted probabilities are very close. Comparing the mean values of the predicted probability of taking part in LMT (LMT=1) for participants and non-participants we typically find differences less than 0.5 per cent, Table A1. Similar and small differences are found for predicted probabilities of PROG=1 and OUT=1, see Raaum, Torp and Zhang (2002a) for the other cohorts. As expected from the similarity in predicted probabilities, mean values of pre-training observables are very similar for participants and matched non-participants, see Table A2 in Appendix.

[16] Similar illustrations for the other cohorts as well as for the predicted values of Prob(PROG=1) and Prob(U=1) can be found in Raaum, Torp and Zhang (2002a).

Figure 4. Predicted probability distributions Prob(LMT=1), Prob(OUT=1) and Prob(PROG=1). Cohort winter 1991.

—o— LMT participants ----Δ--- Matched non-participants ......□....... All unemployed

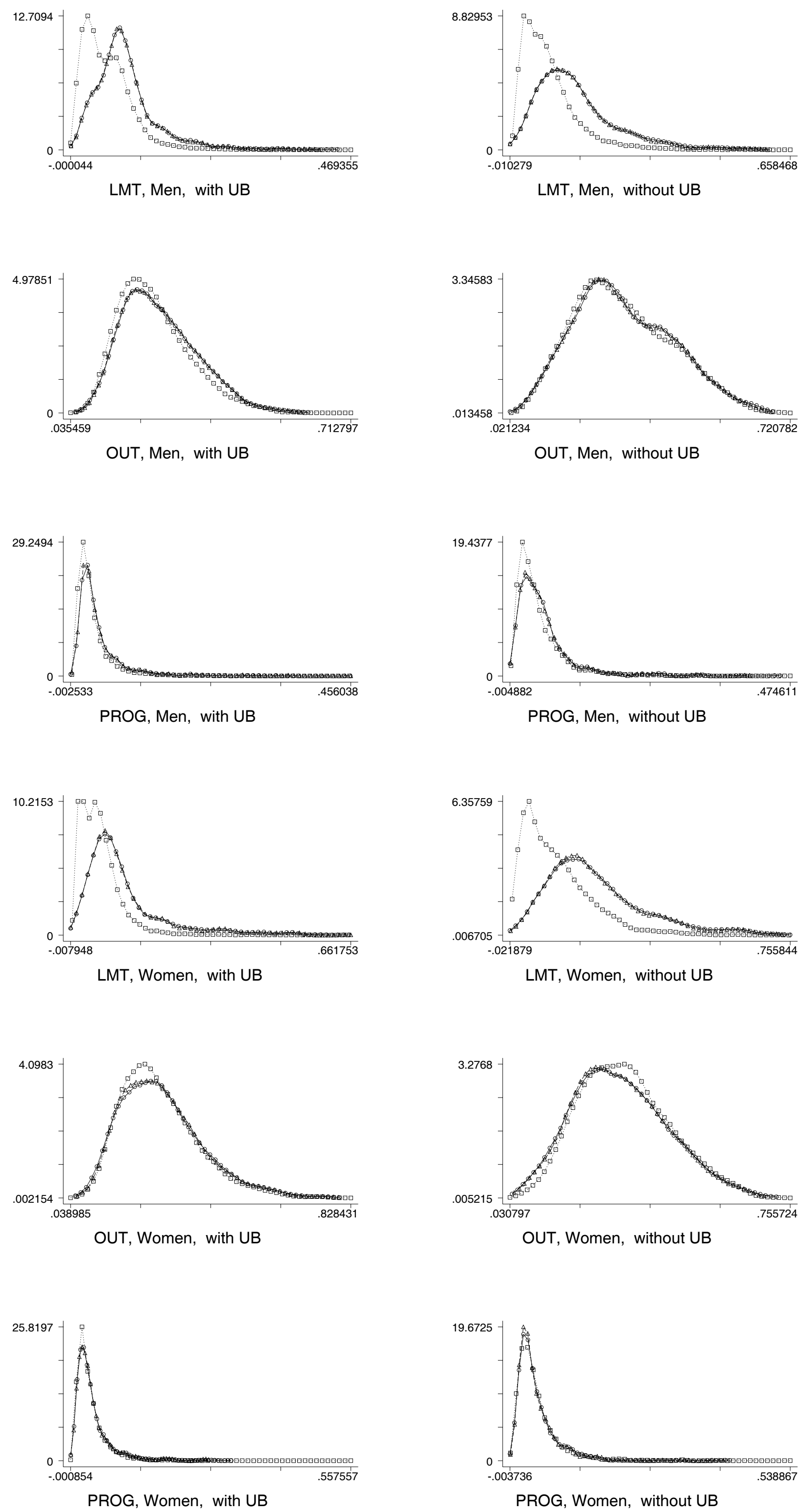
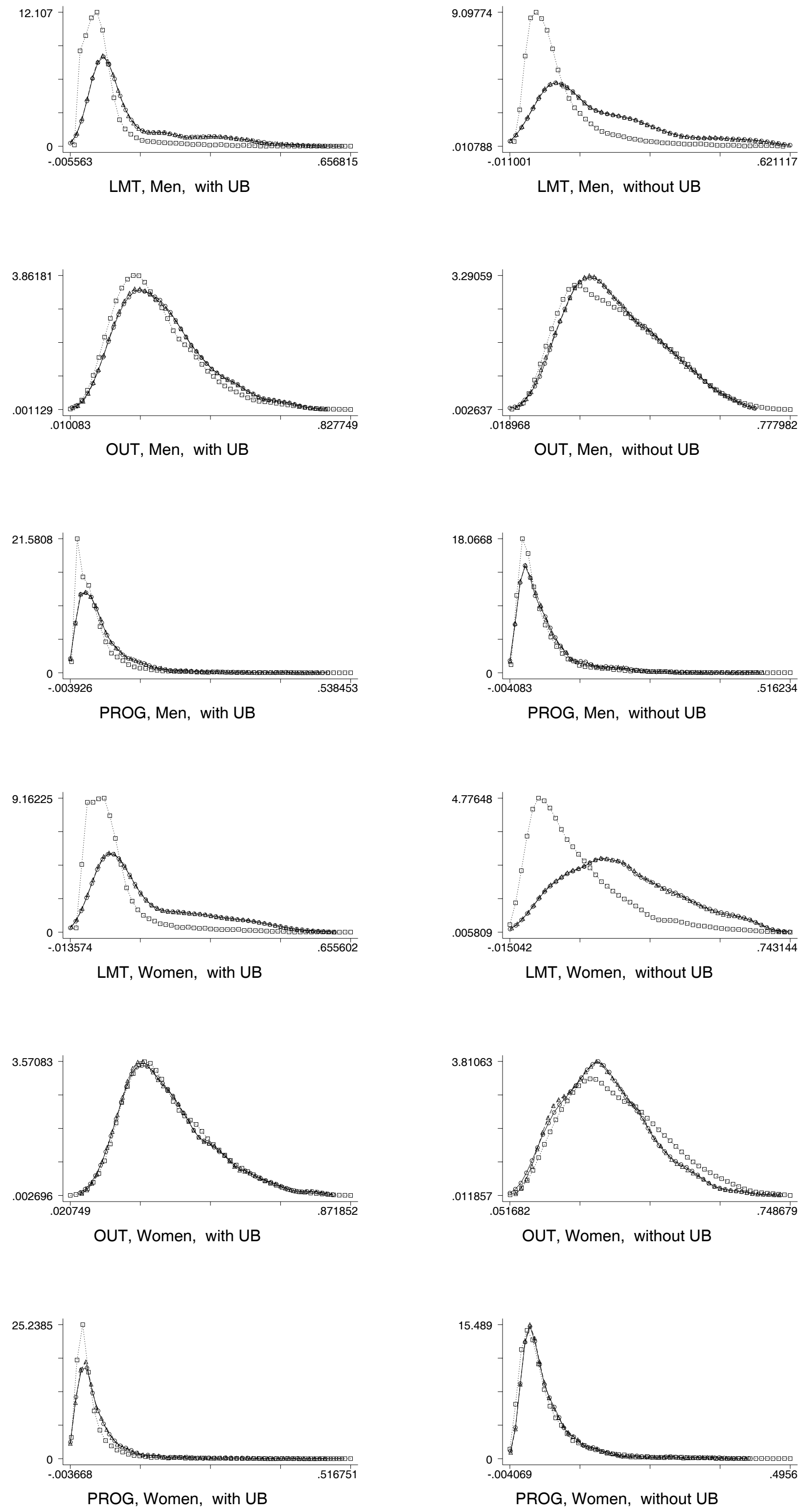


Figure 5. Predicted probability distributions Prob(LMT=1), Prob(OUT=1) and Prob(PROG=1). Cohort autumn 1991.



Cross over and substitution

Participation is defined according to training status by the end of February (Winter) and September (Autumn). The majority of courses start in the beginning of the term, i.e. January/February and August/September. Non-participants are not excluded by administrative procedures, nor by our matching procedure, to start training later on, either in the same term (January –June, July-December) or in the next.

Table 2. Participation in LMT. Cross over and substitution. Fraction by gender and unemployment benefit entitlement. Average across all cohorts.

	Cross over	Substitution ("delay")	
Period:	Same term	Next term	Two terms later
Male, UB			
Participants	1	0.4842	0.2164
Non-participants	0.0717	0.1207	0.1024
Female, UB			
Participants	1	0.5207	0.2214
Non-participants	0.0588	0.1089	0.1002
Male, No UB			
Participants	1	0.5186	0.2607
Non-participants	0.0781	0.1466	0.1313
Female, No UB			
Participants	1	0.6067	0.2965
Non-participants	0.0928	0.1576	0.1422

If members of the comparison group start in LMT the same term (6 months), they are characterised as cross-overs. If they enrol during the two following terms, we label it substitution. In Table 2, we first report the average fraction of cross-overs and find that about 6 to 9 per cent of the non-participants do enrol in training during (i.e. later in) the training term. Second, between 10 and 15 percent of the non-participants turn up as participants during the following terms. Participants, however, are much more likely to be enrolled in the following two terms. This is partly due to courses with long durations that stretch into the next term. There in no strong indication of inter-temporal substitution in the sense that participation is delayed for a substantial fraction of the non-participants.

8. Training effects

The individual effects are estimated by group (i.e. gender, cohort and unemployment benefit entitlement) for each of the three post-training years.[17] All effects are *average training effect on the trained*, simply defined as the mean earnings of the participants minus the mean of the matched unemployed non-participants, as explained in section 3. The main results are presented in Table 3, where we have aggregated the cohort-specific training effects.

Table 3. Average effects of training on annual earnings (NOK 1997). Average over cohorts.

	First year effect		Second year effect		Third year effect	
Season	Winter	Autumn	Winter	Autumn	Winter	Autumn
Male, UB						
Average effect	11,120	-3,755	14,052	8,127	13,517	9,142
Std.error	3,826	2,618	3,825	2,992	3,767	3,085
# positive effects	5	0	5	3	4	4
# negative effects	0	4	0	0	0	0
# insignificant effects	1	2	0	2	0	0
Female, UB						
Average effect	11,966	-6,316	17,113	8,762	20,215	14,107
Std.error	3,379	1,931	3,631	2,300	3,742	2,425
# positive effects	5	0	5	4	4	4
# negative effects	0	1	0	0	0	0
# insignificant effects	1	5	0	1	0	0
Male, no UB						
Average effect	8,851	-34	11,940	10,134	10,935	11,587
Std.error	4,780	3,252	5,199	3,893	5,472	4,242
# positive effects	2	0	4	3	2	2
# negative effects	0	1	0	0	0	0
# insignificant effects	4	5	1	2	2	2
Female, no UB						
Average effect	7,931	-9,645	10,102	-1,533	10,574	2,675
Std.error	3,224	1,713	3,594	2,077	3,611	2,228
# positive effects	2	0	3	0	2	1
# negative effects	0	5	0	1	0	0
# insignificant effects	4	1	2	4	2	3
Training years	*1991-1996*	*1991-1996*	*1991-1995*	*1991-1995*	*1991-1994*	*1991-1994*

Note: Average effects are weighted by the number of participants in each cohort.

[17] In total, 120 (=48+40+32) effect estimates. All are presented in the Appendix, Tables A3-A5.

The weighted average training effect, with the corresponding standard error, is reported together with the number of statistically significant positive and negative effects.

The effects differ by UB status and season. Consider first the group who collected unemployment benefits at the time of enrolment (two first panels). The training effects are positive and significant, in economic as well as in statistical terms. The only exception is the first year effects for Autumn courses. These negative effects are likely to reflect the short time span between the training and the outcome periods, amplified by the continuation of a training period into the next calendar year, among autumn course participants. In practice, what we call the post-training period actually include periods on training for more than one third of the participants.[18]

For winter courses, annual effects vary between 11, 000 and 22,000 NOK. The effect is positive for most groups and outcome years. The effect of autumn courses varies between –6,300 and 17,100 NOK. Ignoring the (negative) first year effect for autumn courses, 44 of 48 effects are significantly positive. The effects of winter and autumn courses tend to converge as we expand the distance between the training and the post-training period. We find relatively small differences between men and women. If any, women seem to gain more than men.

The results are more mixed for the participants without unemployment benefits. These individuals are typically in the process of entering the labour market and the training effects are less favourable. The effects differ by season and gender. Winter courses have positive effects, both for men and women, but they are not always significant in statistical terms. Half of the estimated effects (15 out of 30) are significantly positively, while the rest are not different from zero. There is no obvious gender difference for the winter courses. The first year effects of the autumn courses are again negative, but males without UB entitlement have similar training effects as males with UB. For women without UB who started training in August-September, however, the training effect is close to zero.

In total, Table 3 shows that the LMT programme raises the earnings of the participants. Assessing the effects three years after training, significantly positive effects are found for 23 of the 32 groups (cohort*gender*UB entitlement) and the average effect on annual earnings is more than 10 000 NOK.

[18] In Raaum and Torp (2001) we study training starting in August-September 1991, but define 1993 as the first outcome year.

The estimated training effects are significant and some may find them suspiciously large, given that the typical participant spent 4-7 months on training during the training period. We are not extremely successful in modelling the selection into training and critical observers may argue that many of the unobserved characteristics which determine participation is likely to be correlated with earnings potential, violating the CIA. From pre-training earnings records we can gain more confidence in the consistency of our estimates. By looking at whether individual earnings are correlated with *future* LMT participation, we test the joint null hypothesis that CIA holds for post-training earnings *and* the over-identifying restriction that pre-training earnings are uncorrelated with unobserved characteristics determining participation, see Heckman and Hotz (1989).[19] Of course, this test has no power with respect to an alternative null where only post-training earnings are correlated with training status via unobserved characteristics.

We cannot perform the pre-training test on all cohorts since earnings are not observed in our data before 1992. Hence, we can only test from cohort five (winter 1993) onwards. In Appendix, Table A6, we report estimates of the earnings differential between participants and (matched) non-participants in Y_{k-s} where k is the training year and s varies between one and four. The results are clear. For none of the cohorts, groups or pre-training years we are able to reject the null. Pre-training earnings are not (significantly) different for participants and non-participants. Even the point estimates are generally low and we find positive as well as negative pre-training differentials. There are, however, some indications that post-training earnings of participants are somewhat higher among 1995-96 cohorts of males receiving UB.

Ideally one would like to have an internal comparison group of rejected applicant to measure the counterfactual outcome for participants, see e.g. Raaum and Torp (2002). In our previous study of training effect we argue that; "Our data indicate that training programmes attract applicants with better employment prospects than the average unemployed. This kind of self-selection, e.g. on post-training variables, is hard to identify and correct for". However, the magnitude of the bias is not very large. Moreover, our previous recommendation follows from a study with a stock-sampled comparison group without matching. We believe our previous warnings about external comparison groups do not necessarily undermine the strategy in this paper.

[19] The matching means that we do not have to worry about observables.

Finally, it is worth noticing the considerable heterogeneity in training effects across groups and outcome years. Evaluation studies typically study a limited number of cohorts and outcome periods. Our results illustrate the potential problem of low external validity, at least in studies of Norwegian data.

9. Are training effects higher when job opportunities are favourable?

With the average training effects at hand, we are able to test the hypothesis that individual programme effects are higher when labour market conditions are favourable. This section offers a meta-analysis of the large number of estimated training effects, exploiting variations in job opportunities *during the post-training periods* between men and women, across time and regions. First, we consider the training effects reported in the previous section and investigate whether these effects correlate with job opportunity indicators at the national level. Second, we disaggregate by geographical region and test whether county-specific training effects vary systematically with local labour market conditions, across and within counties. Although the main focus is on the association between training effects and the business cycle, we also investigate whether there are systematic differences in training effects between men and women or by unemployment benefit entitlement.

9.1 Macro Variation in Job Opportunities

We use two indicators of how job opportunities vary over time and by gender. First, we consider the average annual unemployment rates from the Labour Force Surveys (LFS) for those aged 25-54, by gender. However, the unemployment stock is likely to be a noisy measure of variation in job opportunities over time and across groups, because it is heavily influenced by the *inflow* into unemployment. The gender-specific outflow rate from unemployment to employment is likely to be a better measure of job opportunities for the unemployed. This motivates our second indicator, which is a human capital adjusted unemployment outflow rate calculated for the 1990's by means of data covering all unemployment spells in Norway.[20] The yearly indicator is equal to the average monthly exit

[20] The cohorts used to estimate the LMT effects are extracts from this complete data base. Further details on the data and the duration model used to estimate the outflow rates are given in the Appendix.

rates for prime aged unemployed receiving unemployment benefits, evaluated at mean value of observables like age, schooling, marital status and unemployment duration. While the outflow rate follows the LFS unemployment rate over time, the two indicators differ systematically when job opportunities are compared across gender, see Figure A1 in Appendix. The LFS unemployment rate indicates a more favourable labour market for women than for men, up to 1996, while the outflow rates (at any given year) show that job opportunities among unemployed men are considerably better than for unemployed women.[21] Tables 4 and 5 describe how the cohort- and group-specific training effects reported in section 8 correlate with the two job opportunity indicators (JOI's). We also include season, gender and UB entitlement dummies.

In the preferred specification, the gender differential differs by UB entitlement. As noted in section 8, the first year effect is negative for the Autumn courses but significantly positive for Winter courses, see second column Table 4. The average second year effect is positive for both seasons, see Table 5. No significant difference is found between males with and without UB entitlement. While men and women with unemployment benefits have about the same effect of training, there is a significant gender difference in favour of men for participants without UB entitlement.

Including the LFS unemployment rate, we find that a one percentage point increase in unemployment is associated with a reduction in the first year effect of about 6,000 NOK and somewhat less for the second year, see column four in Tables 4 and 5. The gender differential is amplified because women, according to the LFS unemployment rate, met more favourable labour market conditions.

While the estimate for the stock indicator can be interpreted directly, the impact of the outflow measure needs further explanations. The average monthly exit rate is about 0.06. If the outflow rate increases by 0.01, the estimated first year effect increases by around 3,500 NOK. Comparing the 10^{th} and the 90^{th} percentile in the observed outflow distribution, the predicted first year training effect difference is around 17,500 NOK. The impact of job prospects on training effects is somewhat weaker for the second year, but it remains significant, see Table 5, column six.

[21] This is consistent with a lower inflow to unemployment among women, see Brinch (2000) for Labour Force Survey (LFS) evidence. Actually, a comparison of LFS-based outflows from unemployment to employment, see Brinch (2000), reveals that rates are higher for men (Figure V.3.3) than for females (Figure V.9.3) during the first half of the 1990's.

Table 4. The Impact of Job Opportunities on First year Training Effects.

Estimated OLS parameters (standard errors).

	Without Job Opportunity indicator		**Job Opportunity indicators**			
			Unemployment rate (%)		Unemployment outflow rate	
Unemployment rate (%)			-5,872 (1,021)	-5,728 (991)		
Unemployment outflow					361,612 (61,350)	357,181 (58,657)
Winter	16,102 (1,614)	16,071 (1,573)	15,980 (1,227)	15,953 (1,188)	15,967 (1,215)	15,940 (1,161)
No UB	-1,950 (1,427)	1,731 (2,460)	-2,501 (1,089)	536 (1,868)	-2,551 (1,078)	833 (1,820)
Female	-3,745 (1,518)	-1,449 (1,946)	-9,500 (1,527)	-7,473 (1,802)	1,635 (1,462)	3,674 (1,664)
Female*no UB		-5,414 (2,982)		-4,447 (2,258)		-4,965 (2,200)
Constant	-3,061 (1,441)	-4,470 (1,608)	863 (1,293)	-391 (1,405)	-4,777 (1,126)	-6,048 (1,214)
R^2	0.7228	0.7426	0.8434	0.8566	0.8467	0.8633
# observations	48	48	48	48	48	48

Note: Dependent variable is the average training effect = mean earnings differential between participants and non-participants; by gender, UB entitlement, season and training year. OLS weighted by the inverse of the standard error of the estimated training effect. Reference group is men, with UB on Autumn courses. The unemployment outflow is measured as deviations from the mean.

Table 5. The Impact of Job Opportunities on Second Year Training Effects. Estimated OLS parameters (standard errors)

	Without Job Opportunity indicator		**Job Opportunity indicators**			
			Unemployment rate (%)		Unemployment outflow rate	
Unemployment rate (%)			-4,340 (1,416)	-4,175 (1,259)		
Unemployment outflow					245,941 (99,380)	242,913 (88,473)
Winter	7,707 (1,851)	7,669 (1,671)	7,614 (1,668)	7,581 (1,482)	7,707 (1,732)	7,670 (1,542)
No UB	-6,292 (1,726)	187 (2,695)	-6,559 (1,557)	-331 (2,382)	-6,559 (1,618)	-136 (2,477)
Female	-2,306 (1,797)	1,620 (2,104)	-5,565 (1,937)	-1,674 (2,105)	1,943 (2,403)	5,781 (2,458)
Female*no UB		-9,780 (3,311)		-9,387 (2,922)		-9,691 (3,039)
Constant	9,623 (1,706)	7,309 (1,735)	1,1687 (1,678)	9,388 (1,653)	7,233 (1,866)	4,969 (1,807)
R^2	0.5013	0.6009	0.6068	0.6984	0.6009	0.6733
# observations	40	40	40	40	40	40

Note: See Table 4.

We find a statistically as well as economically significant association between post-training job opportunities and average training effects. The gender difference in training effects is, however, sensitive to the type of JOI used. While the outflow specification attributes a significant part of the lower female training effect to less favourable job opportunities for women, the gender difference is amplified when the unemployment rate is used to proxy job prospects. Actually, among participants with UB, women gain more from training than men when we control for differences in job opportunities by means of the outflow indicator, see column six in Tables 4 and 5.

9.2 Job Opportunities Between and Within Counties

The JOI based on the adjusted outflow from unemployment is also calculated at the county level, see Appendix, Table A9 for details. Correspondingly, the estimated training effects are split into 19 county-specific effects, by gender, unemployment benefit entitlement and training cohort, using participants and non-participants from the same county.[22] We regress the group specific training effect in each county on the corresponding county level JOI, separately by outcome years. Table 6 reports estimates with and without county dummies. The inclusion of county fixed effects can be motivated by differences in training programmes and labour market characteristics across regions as well by time-invariant measurement error in the county-specific JOI.

The differences between men and women, winter and autumn courses as well as between those entitled to unemployment benefit and those without are very similar to those reported in Table 4 and 5. The significantly positive effect of the JOI shows that the training effects are higher in local labour markets with more favourable employment opportunities. Consequently, using variation in job opportunities across and within regions confirm our findings based on macro-level indicators.

The impact of job opportunities on training effects is economically important. The difference between the 10th and the 90th percentile in the county-specific JOI distribution is about 0.05. Thus, the corresponding predicted training effect differential is found by dividing the coefficient in the first row of Table 6 by twenty.

Across outcome years, we find that the effect differential is about 10,000 to 15,000 NOK when comparing the 10th and the 90th percentile in the county-specific JOI distribution. Whether we use macro or regional variation in job opportunities, we find a similar association between the state of the labour market and the impact of labour market training. We consider this as strong evidence for the case that training effects do depend on how outcome years are located in the business cycle.[23]

[22] Note, however, that the matching is not performed at the county level.
[23] We have also included the average outflow JOI during the last 12 months before training among the controls in Table 4,5 and 6, and the estimates of the other variables, like the JOI during outcome years, are basically unchanged. No systematic pattern is found for the impact of the pre-training JOI.

Table 6. The Impact of Job Opportunities on First, Second and Third Year Training Effects. County-level effects. Estimated parameters (standard errors).

	First year effect		Second year effect		Third year effect	
Unemployment outflow JOI	213,440 (47,566)	303,270 (44,841)	190,722 (63,709)	216,203 (65,489)	239,013 (81,189)	269,572 (93,147)
Winter	16,613 (1,215)	15,254 (1,070)	7,070 (1,432)	6,260 (1,354)	4,796 (1577)	4,368 (1,544)
No UB	2,267 (1,867)	210 (1,642)	-48 (2,278)	-1,377 (2,148)	-48 (2,518)	-947 (2,459)
Female	2,045 (1,634)	3,183 (1,456)	4,998 (2,104)	5,462 (2,045)	10,028 (2,590)	10,516 (2,748)
Female*no UB	-6,051 (2,262)	-5,018 (1,982)	-10,943 (2,793)	-9,983 (2,627)	-11,548 (3,125)	-10,706 (3,044)
Constant	-6,423 (1,216)	5,050 (1,780)	5,432 (1,546)	15,844 (2,402)	5,594 (1,967)	12,721 (3,036)
Fixed effects	None	County Dummies	None	County Dummies	None	County Dummies
R^2	0.2222	0.4176	0.1147	0.2399	0.0934	0.1714
# observations	874	874	722	722	589	589

Note: Dependent variable is the county level training effect = average earnings differential between participants and non-participants; by gender, UB entitlement and season. OLS weighted by the inverse of the standard error of the estimated training effect. Reference group is men, with UB on Autumn courses in the Oslo region.

10. Conclusions

By comparing mean outcomes for matched samples of participants and non-participants we evaluate the Norwegian labour market training programme (LMT) targeted at unemployed adults. We estimate the average earnings effects of training on the trained, using individuals participating in LMT drawn from all entrants (and re-entrants) in the Norwegian public unemployment register during December 1990 – July 1996. As we evaluate average effects only, we construct fairly homogenous groups of participants with separate analyses for those starting training at about the same time, i.e. in winter or in autumn each year. This gives us 12 cohorts of participants over the six year period. The samples are also separated by gender and unemployment benefit entitlement. This gives a total of 48 subsamples of participants and their partners of unemployed non-participants.

Matched samples of unemployed non-participants are used to simulate the counterfactual outcomes of the participants. The matching procedure selects unemployed non-participants with the closest set of predicted probabilities from a multinomial choice model where training, participation in other programmes, exit from the unemployment register or remaining unemployed are alternative outcomes.

Unlike most evaluation studies that typically study one, or a limited number of cohorts of participants, we are able to disentangle the impact of business cycles on training effects from the importance of the time span between the training and the post-training period. Since a single cohort study will have perfect correlation between “time since programme” and “calendar period”, any impact of post-programme labour market conditions can only be identified by means of spatial (or regional) variation. Using several cohorts of participants, we are able to estimate first, second and third year effects under different labour market conditions, even controlling for fixed regional effects.

Two main conclusions can be drawn from this study. First, the training programme has, on average, a positive impact on post-training annual earnings for those who participate. Second, the training effect is larger when job opportunities in the post-programme period are favourable. When job opportunities are bleak, participants gain less in terms of post-training earnings. This insight follows from the *meta analysis* of the large number of group- and cohort-specific training effects where two indicators of post-training job opportunities are used. The first indicator is the national average of annual unemployment rates by gender from the Labour Force Surveys. The second indicator is based on all Norwegian unemployment spells in the 1990's and is a measure of annual average of outflow from unemployment, based on human capital adjusted monthly exit rates. We argue that the outflow indicator is a more precise measure of post-training labour market opportunities. Training effects are positively correlated with job opportunities measured by both indicators.

As in most non-experimental studies, the estimated training effects can be driven by selection on unobservables rather than a causal impact on post-training outcomes. In our case, the institutional setting does not provide any clearcut indication. As for most programmes targeted at unemployed, the recruitment to LMT is a mixture of self-selection and administrative decisions. Previous studies of selections process suggest that there is, if any, a positive selection to LMT, i.e. participants have observed – and possibly also unobserved - characteristics assumed to correlate positively with employability. In this study pre-training earnings records are available. When looking at whether individual earnings are correlated with *future* LMT participation, the null hypothesis of no correlation is not rejected. Since pre-training earnings are not significantly different for participants and non- participants, we gain more confidence in the consistency of our estimates.

Information on how labour market conditions affect estimated programme effects are useful when assessing and explaining differences in effects across time, regions and even countries. For example, in the case of Sweden, programme effects have changed systematically over the business cycle. The programme effects are more negative (or less positive) when evaluations are based on post-programme outcomes during the area of high unemployment in the early 1990s, compared to studies using data for the 1980s or towards the end of the 1990s. Our results are also useful for policy making as the optimal timing and volume of ALMP must take into account that individual effects are likely to vary over the business cycle.

References

Ackum, Susanne (1991), "Youth Unemployment, Labour Market Programmes and Subsequent Earnings." *Scandinavian Journal of Economics* 93, pp. 531-453.

Ackum Agell, Susanne and Martin Lundin (2001), "Erfarenheter av svensk arbetsmarknadspolitikk." *Ekonomisk debatt,* 29, pp 239-250.

AM (2000), Effects of Danish Employability Enhancement programmes. Copenhagen: Danish Ministry of Labour.

Aakvik, Arild (1998), "Assessing the effects of labour market training in Norway" in *Five essays on the microeconometric evaluation of job training programs.* Dissertation in Economics No 15. University of Bergen.

Aakvik, Arild (2001), "Bounding a matching estimator: the case of a Norwegian training program**."** *Oxford Bulletin of Economics and Statistics*, Vol 63, No 1, 115-

Aakvik, Arild, James J. Heckman and Edward J. Vytlacil (2000), "Treatment effects for discrete outcomes when responses to treatment vary among observationally identical persons: an application to Norwegian vocational rehabilitation programs." *NBER Technical Working Paper No 262.* Cambridge, MA.

Barnow, B. S. (1987), "The impact of CETA programs on Earnings: A review of the Literature." *Journal of Human Resources* XXII, pp. 157-193.

Björklund, Anders (1990), "Evaluation of Swedish Labor Market Policy." *Finnish Economic Paper* 3. pp. 3-13.

Brinch, Christian (2000), "Arbeidstilbud i vedvarende gode tider", Rapport 7/2000. Oslo: Ragnar Frisch Centre for Economic Research.

Brodaty, Thomas, Bruno Crepon and Denis Fougere (2001), "Using marching estimators to evaluate alternative youth employment programs: Evidence from France, 1986-1988." In M. Lechner and F. Pfeiffer (eds) *Econometric evaluation of Labour market Policies*. Heidelberg: Physica.

Calmfors, Lars (1994), *Active labour market policies and unemployment: a framework for the analysis of crucial design features.* Economic studies 22, pp.7-49. Paris: OECD.

Calmfors, Lars, Anders Forslund and Maria Hemström (2001), *Does active labour market policy work? Lessons from the Swedish experiences*. Working paper. Uppsala: IFAU – office of labour market policy evaluation.

Couch, Kenneth (1992), "New evidence on long term effects of employment and training programs." *Journal of Labor Economics*, 10, pp 380-388.

Deheija, Rajeev and Sadek Wahba (1998), "Propensity score marching methods for nonexperimental causal studies." *NBER Working Paper No 6829.* Cambridge, MA.

Deheija, Rajeev and Sadek Wahba (1999), "Causal effects in nonexperimental studies: Reevaluating the evaluations of training programs." *Journal of the American Statistical Association,* 94, pp 1053-1062.

Fay, Robert G. (1996), "Enhancing the effectiveness of active labour market policies: evidence from programme evaluations in OECD countries." *Labour market and social policy occasional papers no 18*. Paris: OECD.

Friedlander, Daniel and Gary Burtless (1995), *Five years after: The long term effects of Welfare-to-work programs.* New York: Russell Sage.

Fraker, Thomas and Rebecca Maynard (1987), "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs." *The Journal of Human Resources* XXII, pp. 194-227.

Heckman, James J. and Joseph Hotz (1989), "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association* 84, pp. 862-874.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith and Petra Todd (1998), "Characterizing selection bias using experimental data." *Econometrica,* 66, pp.1017-1089.

Heckman, James J., Hidehiko Ichimura and Petra Todd (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme." *Review of Economic Studies,* 64, pp. 605-654.

Heckman, James J., Hidehiko Ichimura and Petra Todd (1998), "Matching as an econometric evaluation estimator." *Review of Economic Studies,* 65, pp. 261-294.

Heckman, James. J., and Jeffrey Smith (1995), "Assessing the Case for Social Experiments." *Journal of Economic Perspectives* 9, Spring 1995, 85-100.

Heckman, James . J., Robert LaLonde and Jeffrey Smith (1999), "The Economics and Econometrics of Active Labor Market Programs." In O. Ashenfelter and D. Cards, (eds.), *Handbook of Labor Economics, Volume III,* Amsterdam: North-Holland Elsevier.

Hirano, Keisuke, Guido W. Imbens and Geert Ridder (2000), "Efficient estimation of average treatment effects using the estimated propensity score." *NBER Technical Working Paper No 251.* Cambridge, MA.

Imbens, Guido W. (2000), "The role of the propensity score in estimation dose-response functions." *Biometrica* 87:706-710.

Jensen, Peter, Peder J. Pedersen, Nina Smith, Niels Westergård-Nielsen (1993), "The effects of labour market training on wages and unemployment: Some Danish results." In Bunzel et al. (eds.), *Panel Data and Labour Market Dynamics.* Amsterdam: North-Holland Elsevier.

LaLonde, Robert (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76, pp. 604-620.

LaLonde, Robert (1995), "The Promise of Public-Sponsored Training Programs." *Journal of Economic Perspectives* 9, pp. 149-168.

Larsson, Laura (2000), Evaluation of Swedish youth labour market programmes*. Working Paper 2000:1. Uppsala: The Office of Labour Market Policy Evaluation.*

Lechner, Michael (2001a), "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption." In M. Lechner and F. Pfeiffer (eds) *Econometric evaluation of Labour market Policies,* 43-58. Heidelberg: Physica.

Lechner, Michael (2001b), "Programme heterogeneity and propensity score matching: An application to the evaluation of active labour market policies." *Discussion Paper*. Swiss Institute for international economics and applied economic research. University of St.Gallen.

Martin, John P. (1998), What works among active labour market policies: Evidence from OECD countries experiences. *OECD Labour market and social policies. Occasional papers no 35*. Paris: OECD.

Martin, John P., and David Grubb (2001), What works for whom: a review of OECD countries' experiences within active labour market policies. Working paper. Paris: OECD.

Meyer, B. D. (1990) "Unemployment Insurance and Unemployment Spells". *Econometrica*, Vol. 58, No. 4, 757-782.

OECD (1993), *Employment outlook*. Paris: OECD.

Prentice, R. L. and Gloeckler, L. A. (1978) "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data" *Biometrics*, Vol. 34, 57-67.

Regnér, Håkan. (1997), *Training at the job and training for a new job: two Swedish studies,* Dissertation Series, Swedish Institute for Social Research, Stockholm.

Regnér, Håkan. (2002), "A nonexperimental evaluation of training programs for the unemployed in Sweden." *Labour Economics,* 9 (2002).

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The central role of the propensity score in observational studies of causal effects." *Biometrica*, 70:41-55.

Rubin, Donald B. (1979), "Using multivariate matched sampling and regression adjustment to control bias in observational studies." *Journal of American Statistical Association,* 74:318-328.

Røed, Knut and Tao Zhang (1999), "Unemployment Duration in a Non-Stationary Macroeconomic Environment", Memorandum No 14, Department of Economics, University of Oslo

Røed, Knut, Hege Torp, Irene Tuveng and Tao Zhang (2000), "Hvem vil og hvem får delta? Analyser av rekruttering og utvelgelse av deltakere til arbeidsmarkedstiltak i Norge på 1990-tallet", Report 4/2000. Oslo: Ragnar Frisch Centre for Economic Research.

Røed, Knut. (2001) Hvor stramt er arbeidsmarkedet. Et forslag til en ny konjunkturindikator (How tight is the labour market? A proposal for a new business cycle indicator)" *Økonomisk forum*, No. 3, 25-32 (2001)

Røed, Knut and Tao Zhang (2002) "Does Unemployment Compensation Affect Unemployment Duration?" *Economic Journal*, Vol. 113, issue 484, 190-206.

Raaum, Oddbjørn and Hege Torp (2002), "Labour market training in Norway – effect on earnings." *Labour Economics* 9 (2002), 207-247.

Raaum, Oddbjørn, Hege Torp and Tao Zhang (2002a) "Labour Market Training Participation in Norway 1991-1996: Construction of treatment and non-treatment groups by matching on observables", Frisch Centre Working paper 2/2002.

Raaum, Oddbjørn, Hege Torp and Tao Zhang (2002b), "Do individual programme effects exceed the costs? Norwegian evidence on long run effects of labour market training." Memorandum 15/2002, Department of Economics, University of Oslo.

Sianesi, Barbara (2002), "An evaluation of the Swedish system of active labour market programmes in the 1990s" Working Paper W02/01. London: The Institute for Fiscal Studies.

Smith, J., Todd, P., (2002). "Is matching the answer to LaLonde's critique of nonexperimental methods?" *Journal of Econometrics*, forthcoming.

Torp, Hege (1994), "The impact of training on employment: Assessing a Norwegian labour market programme." *Scandinavian Journal of Economics*, 96:531-550.

Westergård-Nielsen, Niels (1993), "Effects of training. A fixed effect model." In *Measuring Labour Market Measures* (proceedings from the Danish Presidency Conference). Ministry of Labour, Denmark.

Westergaard-Nielsen, Niels (2001), *Danish labour Market Policy: Is it worth it?* Working paper 01-10. Aarhus: Centre for Labour Market and Social Research.

Appendix

1. Definition of variables

Dependent variable

(a) Annual earnings defined by tax registers, including wages, self-employment income and sickleave benefits measured in 1997 NOK. (Unemployment benefits and other transfers are not included).

Explanatory variables

(a) Married (dummy)

(b) Level of education: Educ1 - Educ6
6 dummies: le 9 years, 10 ys, 11-12 ys (reference), 13-16 ys, ge 17 ys, and unknown

(c) County of residence: 19 dummies
one for each county in Norway, county of Oslo as reference

(d) Age: Age1 – Age5
5 dummies: 25-30 ys, 31-35 ys, 36-40 ys (reference), 41-45 ys, 46-50 ys

(e) Immigrant from outside OECD: Immigrant (dummy)

(f) Unemployment history: Month-1 – Months-1923,
i.e. number of months of unemployment before *t*: 16 dummies: 0 (reference), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,13-15, 16-18, 19-23 months

(g) Earnings history: Earnings1-Earnings2123,
i.e. number of years of annual earnings above B.a. before the year *T* (B.a. = Basic amount in the Norwegian Social Insurance Scheme, annually regulated, about Euro 5-6,000 in the period of interest). Earnings history serves as an indicator of aggregated experience. 13 dummies: 0 (reference), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11-15, 16-20, and 21-23 years of income above B.a. (23 is maximum since for the first sample (T= 1991) as the scheme was established in 1967, when the Norwegian Social Insurance Scheme was established)

(h) LMT history: LMT1 – LMT8,
i.e. number of quarters participated in LMT during last 23 months before *t:* 8 dummies, one for each of the latest 8 quarters LMT8 means last quarter, LMT7 means the second last quarter .... LMT1 means two years ago

(i) Programme history: PROG1 – PROG8,
i.e. number of quarters participated in other programmes than LMT during last 23 months before *t:* 8 dummies, on for each of the latest 8 quarters PROG8 means last quarter, PROG7 means the second last quarter .... PROG1 means two years ago

(j) Occupational background: Occup1-Occup6
categories (based on ISCO): (1) technical, physical science, humanistic and artistic work (teachers, nurses, doctors, technicians etc), (2) administrative executive work, clerical work and sales work, (3) agriculture, forestry, fishing and related work, (4) manufacturing work, mining, quarrying, building and construction work (reference), (5) service work, transport and communication, (6) unknown

(k) Children in household: Kid1 – Kid3
4 dummies: 0 (reference), 1, 2 or 3 children and more, below the age of 18

(l) Previous annual earnings: AnnEarn1 – AnnEarn3
3 continuously distributed variables for last year before *t,* second last year, and next to second last year: annual earnings measured by B.a.

(m) Left ordinary education just before *t;* LeftEduc (dummy),
only available for cohorts x-12

(n) Left ordinary high level education just before *t;* LeftEducHigh
(dummy, two highest levels of education) , only available for cohorts x-12.

2. Matching results

Table A1. Means of predicted probabilities for matched samples. Selected cohorts.

Subsample	Winter 1991 Partic.	Winter 1991 Non-part	Autumn 1991 Partic.	Autumn 1991 Non-part	Winter 1995 Partic.	Winter 1995 Non-part	Autumn 1995 Partic.	Autumn 1995 Non-part
Women without UB								
Pred prob of LMT	0.2177	0.2158	0.2905	0.2879	0.1366	0.1357	0.2643	0.2632
Pred prob of PROG	0.0480	0.0472	0.0627	0.0623	0.0472	0.0472	0.0684	0.0682
Pred prob of OUT	0.3172	0.3164	0.2914	0.2917	0.2266	0.2263	0.2351	0.2352
Women with UB								
Pred prob of LMT	0.1195	0.1182	0.1687	0.1677	0.0724	0.0720	0.1191	0.1186
Pred prob of PROG	0.0431	0.0427	0.0468	0.0465	0.0485	0.0484	0.0662	0.0661
Pred prob of OUT	0.2959	0.2960	0.3165	0.3160	0.2377	0.2373	0.3213	0.3211
Men without UB								
Pred prob of LMT	0.1493	0.1480	0.1911	0.1902	0.0865	0.0856	0.1735	0.1724
Pred prob of PROG	0.0525	0.0520	0.0594	0.0592	0.0469	0.0468	0.0649	0.0647
Pred prob of OUT	0.3067	0.3056	0.3047	0.3033	0.2331	0.2330	0.2562	0.2560
Men with UB								
Pred prob of LMT	0.0931	0.0928	0.1398	0.1391	0.0628	0.0621	0.1057	0.1053
Pred prob of PROG	0.0447	0.0444	0.0582	0.0578	0.0509	0.0509	0.0686	0.0685
Pred prob of OUT	0.2518	0.2515	0.2774	0.2769	0.2233	0.2223	0.2801	0.2804

Table A2. Descriptive statistics. Participants and matched non-participants. Selected cohorts.

(a) Participants and matched non-participants <u>*with*</u> *unemployment benefits*

	Men Winter 1991		Men Winter 1995		Women Winter 1991		Women Winter 1995	
	Particip.	Non-part.	Particip.	Non-part.	Particip.	Non-part.	Particip.	Non-part.
Age (years)	33.848	33.823	34.352	34.055	35.441	35.570	34.703	34.913
Married	0.485	0.498	0.545	0.574	0.215	0.206	0.338	0.341
Number of children	0.936	0.879	0.845	0.833	1.158	1.177	1.185	1.204
Low education (9 years or less)	0.086	0.076	0.146	0.131	0.054	0.072	0.143	0.146
Low medium educ (10 years)	0.400	0.423	0.355	0.335	0.469	0.464	0.391	0.393
Med educ (11 to 12 years)	0.345	0.342	0.345	0.357	0.347	0.341	0.323	0.322
High educ 1 (13 to 16 years)	0.139	0.124	0.102	0.111	0.108	0.111	0.103	0.105
High educ 2 (17 years +)	0.020	0.020	0.016	0.019	0.018	0.011	0.010	0.007
Ostfold	0.058	0.047	0.053	0.071	0.076	0.068	0.079	0.082
Akershus	0.070	0.069	0.091	0.092	0.089	0.076	0.103	0.133
Oslo	0.090	0.076	0.133	0.115	0.076	0.068	0.153	0.146
Hedmark	0.035	0.046	0.029	0.034	0.035	0.024	0.036	0.027
Oppland	0.048	0.032	0.037	0.034	0.047	0.061	0.052	0.043
Buskerud	0.045	0.052	0.080	0.073	0.038	0.032	0.066	0.061
Vestfold	0.049	0.045	0.044	0.043	0.068	0.058	0.045	0.043
Telemark	0.026	0.028	0.024	0.024	0.025	0.017	0.020	0.032
A_Agder	0.027	0.029	0.026	0.027	0.016	0.021	0.022	0.029
V_Agder	0.056	0.050	0.037	0.035	0.042	0.054	0.032	0.025
Rogaland	0.096	0.104	0.078	0.072	0.128	0.139	0.049	0.059
Hordaland	0.139	0.157	0.084	0.092	0.142	0.122	0.097	0.094
Sogn	0.015	0.012	0.008	0.008	0.006	0.012	0.007	0.006
Moere	0.063	0.067	0.049	0.044	0.056	0.055	0.052	0.035
S_Trond	0.038	0.034	0.051	0.051	0.054	0.064	0.040	0.051
N_Trond	0.022	0.022	0.050	0.055	0.021	0.028	0.051	0.056
Nordland	0.071	0.071	0.060	0.072	0.050	0.056	0.055	0.051
Troms	0.025	0.028	0.035	0.033	0.019	0.023	0.025	0.022
Finnmark	0.027	0.032	0.031	0.028	0.013	0.022	0.017	0.007
Work experience (years)	12.058	11.969	11.816	11.697	9.162	8.964	9.463	9.343
Aggregated pension points	2.861	2.843	2.618	2.570	1.802	1.764	1.806	1.844
Prevevios open unempl (months)	7.882	7.934	9.476	9.591	7.586	7.701	8.049	8.156
Indic. Of last year's income	3.946	3.879	3.671	3.573	2.820	2.784	2.896	2.896
Indic. Of second last year's income	4.367	4.367	3.645	3.522	2.995	2.974	2.744	2.682
Indic of third last year's income	4.350	4.346	3.512	3.419	2.775	2.700	2.546	2.546
Occup 1	0.165	0.135	0.114	0.114	0.186	0.165	0.245	0.261
Occup 2	0.113	0.109	0.135	0.146	0.493	0.501	0.421	0.433
Occup 3	0.026	0.024	0.029	0.021	0.017	0.017	0.013	0.013
Occup 5	0.517	0.532	0.484	0.483	0.085	0.093	0.077	0.079
Occup 6	0.169	0.182	0.205	0.204	0.202	0.209	0.228	0.204
LMT1	0.033	0.027	0.094	0.094	0.021	0.017	0.068	0.077
LMT2	0.036	0.031	0.089	0.080	0.030	0.024	0.061	0.068
LMT3	0.047	0.040	0.078	0.073	0.050	0.052	0.066	0.079
LMT4	0.057	0.047	0.077	0.074	0.063	0.073	0.074	0.075
LMT5	0.087	0.079	0.143	0.131	0.094	0.106	0.137	0.120
LMT6	0.078	0.073	0.152	0.133	0.113	0.114	0.144	0.139
LMT7	0.070	0.062	0.145	0.140	0.111	0.099	0.165	0.150
LMT8	0.063	0.058	0.138	0.146	0.117	0.115	0.196	0.193
PROG1	0.051	0.045	0.132	0.130	0.060	0.060	0.104	0.114
PROG2	0.091	0.086	0.147	0.156	0.054	0.059	0.113	0.123
PROG3	0.088	0.092	0.153	0.179	0.054	0.058	0.113	0.156
PROG4	0.088	0.084	0.165	0.180	0.060	0.056	0.118	0.149
PROG5	0.073	0.068	0.109	0.121	0.075	0.080	0.095	0.134
PROG6	0.088	0.091	0.113	0.131	0.069	0.080	0.097	0.115
PROG7	0.062	0.059	0.075	0.062	0.031	0.039	0.075	0.088
PROG8	0.028	0.034	0.034	0.017	0.016	0.016	0.049	0.064
Just left education			0.144	0.144			0.180	0.180

(b) Participants and matched non-participants <u>without</u> unemployment benefits

	Men Winter 1991		Men Winter 1995		Women Winter 1991		Women Winter 1995	
	Particip.	Non-part.	Particip.	Non-part.	Particip.	Non-part.	Particip.	Non-part.
Age (years)	32.748	32.786	32.989	32.765	34.764	33.978	34.022	34.715
Married	0.502	0.521	0.552	0.557	0.165	0.162	0.267	0.253
Number of children	0.907	0.984	0.637	0.618	1.280	1.355	1.459	1.443
Low education (9 years or less)	0.076	0.066	0.110	0.110	0.039	0.035	0.137	0.125
Low medium educ (10 years)	0.400	0.430	0.286	0.248	0.487	0.493	0.307	0.330
Med educ (11 to 12 years)	0.295	0.265	0.263	0.256	0.313	0.313	0.285	0.262
High educ 1 (13 to 16 years)	0.175	0.169	0.139	0.166	0.126	0.125	0.101	0.106
High educ 2 (17 years +)	0.033	0.054	0.014	0.016	0.020	0.022	0.022	0.016
Ostfold	0.071	0.072	0.058	0.060	0.089	0.074	0.069	0.067
Akershus	0.055	0.055	0.104	0.109	0.071	0.083	0.147	0.151
Oslo	0.150	0.126	0.222	0.243	0.072	0.066	0.169	0.212
Hedmark	0.027	0.014	0.028	0.035	0.030	0.025	0.036	0.034
Oppland	0.028	0.032	0.021	0.024	0.027	0.030	0.029	0.033
Buskerud	0.032	0.039	0.076	0.074	0.037	0.025	0.032	0.033
Vestfold	0.060	0.090	0.062	0.047	0.057	0.044	0.066	0.066
Telemark	0.027	0.028	0.032	0.032	0.025	0.022	0.032	0.026
A_Agder	0.025	0.022	0.017	0.024	0.017	0.014	0.011	0.012
V_Agder	0.049	0.058	0.035	0.017	0.046	0.044	0.037	0.041
Rogaland	0.107	0.087	0.057	0.060	0.160	0.194	0.073	0.055
Hordaland	0.150	0.142	0.087	0.088	0.128	0.128	0.107	0.092
Sogn	0.013	0.016	0.003	0.002	0.005	0.007	0.008	0.008
Moere	0.035	0.033	0.038	0.038	0.047	0.054	0.045	0.052
S_Trond	0.060	0.068	0.035	0.027	0.077	0.091	0.044	0.040
N_Trond	0.014	0.019	0.024	0.022	0.030	0.024	0.037	0.029
Nordland	0.046	0.044	0.041	0.049	0.049	0.051	0.038	0.027
Troms	0.028	0.033	0.046	0.035	0.010	0.008	0.014	0.010
Finnmark	0.025	0.022	0.016	0.016	0.022	0.019	0.010	0.012
Work experience (years)	7.836	8.066	6.120	5.533	5.931	5.823	4.481	4.895
Aggregated pension points	1.721	1.832	1.324	1.187	1.143	1.141	0.971	1.007
Prevevios open unempl (months)	7.148	7.594	7.462	7.323	5.596	5.572	5.223	5.295
Indic. Of last year's income	1.791	1.907	1.063	1.038	0.847	0.909	0.507	0.528
Indic. Of second last year's income	2.080	2.188	1.114	1.067	0.982	1.066	0.516	0.565
Indic of third last year's income	2.277	2.357	1.120	1.103	1.112	1.203	0.570	0.601
Occup 1	0.140	0.173	0.117	0.126	0.182	0.162	0.214	0.227
Occup 2	0.087	0.090	0.117	0.112	0.347	0.355	0.292	0.312
Occup 3	0.044	0.041	0.036	0.027	0.012	0.019	0.003	0.006
Occup 5	0.419	0.397	0.314	0.314	0.074	0.067	0.081	0.062
Occup 6	0.113	0.107	0.164	0.144	0.172	0.168	0.164	0.169
LMT1	0.038	0.032	0.115	0.099	0.030	0.030	0.099	0.097
LMT2	0.039	0.035	0.112	0.095	0.046	0.049	0.092	0.095
LMT3	0.061	0.057	0.122	0.104	0.072	0.084	0.101	0.106
LMT4	0.088	0.071	0.120	0.106	0.089	0.106	0.110	0.114
LMT5	0.120	0.106	0.174	0.147	0.126	0.133	0.132	0.115
LMT6	0.096	0.088	0.156	0.137	0.111	0.123	0.126	0.129
LMT7	0.069	0.065	0.131	0.110	0.123	0.108	0.190	0.199
LMT8	0.085	0.080	0.112	0.103	0.148	0.131	0.216	0.222
PROG1	0.052	0.033	0.055	0.062	0.052	0.049	0.032	0.029
PROG2	0.077	0.066	0.062	0.066	0.049	0.052	0.029	0.030
PROG3	0.074	0.079	0.080	0.073	0.054	0.052	0.034	0.032
PROG4	0.101	0.102	0.088	0.077	0.056	0.052	0.055	0.055
PROG5	0.082	0.071	0.080	0.069	0.057	0.057	0.059	0.069
PROG6	0.093	0.104	0.112	0.098	0.069	0.064	0.077	0.077
PROG7	0.074	0.066	0.085	0.077	0.037	0.030	0.051	0.051
PROG8	0.044	0.028	0.047	0.039	0.025	0.030	0.045	0.040
Just left education			0.188	0.181			0.226	0.240

3. Detailed training effects. By group and outcome year.

Table A3. First year effects. By season, gender, unemployment benefit entitlement and training year.

Training year	1991	1992	1993	1994	1995	1996
Winter courses						
Male, UB	11414	8509	7014	13979	8139	22649
	[3784.07]	[3210.01]	[3084.97]	[3332.31]	[5491.92]	[5044.47]
Male, no UB	9110	38	568	7994	13231	24862
	[4825.44]	[4553.98]	[4388.8]	[4282.51]	[5361.35]	[5389.45]
Female, UB	9614	14603	12445	15513	6648	11589
	[3145.15]	[2859.55]	[3100.11]	[3342.19]	[3951.31]	[4068.22]
Female, no UB	3111	5708	10570	6231	7188	12968
	[3620.43]	[3213.62]	[3416.26]	[3381.53]	[3764.92]	[3526.58]
Autumn courses						
Male, UB	-7875	-6185	-4810	-7000	5853	4436
	[2302.32]	[2051.02]	[2346.41]	[2740.42]	[3099.72]	[3742.11]
Male, no UB	-7040	-3128	-2679	3227	4766	4063
	[3192.91]	[3028.56]	[3194.79]	[2935.77]	[3252.95]	[3931.44]
Female, UB	-12735	-6831	-6349	-3801	-9296	3943
	[1748.03]	[1645.94]	[1957.65]	[1917.35]	[2092.67]	[2338.86]
Female, no UB	-14941	-10501	-8320	-11028	-9883	-3593
	[1861.4]	[1796.42]	[1964.72]	[1731.21]	[1781.69]	[2015.6]

Table A4. Second year effects. By season, gender, unemployment benefit entitlement and training year.

Training year	1991	1992	1993	1994	1995	1996
Winter courses						
Male, UB	13606	11761	13968	17113	13457	na
	[4114.45]	[3515.27]	[3226.13]	[3639.87]	[5217.43]	
Male, no UB	14837	-2417	10133	17321	20710	na
	[5316.2]	[4934.32]	[4857.87]	[4925.34]	[6076.64]	
Female, UB	14250	18970	16355	22009	13244	na
	[3412.31]	[3216.56]	[3455.25]	[3767.5]	[4439.52]	
Female, no UB	-420	3130	18332	16038	11249	na
	[3979.53]	[3619.47]	[3749.19]	[4034.93]	[4321.21]	
Autumn courses						
Male, UB	3380	8845	8499	5296	16739	na
	[2765.92]	[2542.49]	[2790.15]	[3277.77]	[3826.45]	
Male, no UB	436	5365	8592	16799	17686	na
	[3886.01]	[3770.06]	[3911.96]	[3800.26]	[4088.12]	
Female, UB	3251	11269	10985	12378	5553	na
	[2132.76]	[2088.17]	[2349.76]	[2363.02]	[2608.43]	
Female, no UB	-5517	-993	2712	-2412	-1402	na
	[2382.13]	[2212.91]	[2399.72]	[2232.31]	[2325.83]	

Table A5. Third year effects. By season, gender, unemployment benefit entitlement and training year.

Training year	1991	1992	1993	1994	1995	1996
Winter courses						
Male, UB	8746	14661	11608	18469	na	na
	[4219.21]	[3568.96]	[3482.13]	[3910.1]		
Male, no UB	15325	-4253	10711	21011	na	na
	[5598.28]	[5236.74]	[5711.5]	[5340.2]		
Female, UB	18099	21003	18466	23525	na	na
	[3569.69]	[3384.2]	[3694.2]	[4323.92]		
Female, no UB	6535	2833	15841	16308	na	na
	[4275.83]	[3778.01]	[3966.32]	[4422.68]		
Autumn courses						
Male, UB	7145	9913	10707	8574	na	na
	[2970.39]	[2703.24]	[3095.34]	[3664.46]		
Male, no UB	5112	6516	14058	19403	na	na
	[4168.54]	[3992.43]	[4404.87]	[4369.46]		
Female, UB	10569	13460	18419	14507	na	na
	[2288.66]	[2234.07]	[2581.22]	[2609.56]		
Female, no UB	-1217	3921	6996	1159	na	na
	[2617.54]	[2402.72]	[2652.1]	[2559.16]		

Table A6. Fourth year effects. By season, gender, unemployment benefit entitlement and training year.

Training year	1991	1992	1993	1994	1995	1996
Winter courses						
Male, UB	15206	16376	7853	na	na	na
	[4258.43]	[3777.54]	[3794.94]			
Male, no UB	17276	7030	12517	na	na	na
	[5735]	[5660.09]	[5773.17]			
Female, UB	17252	18617	18981	na	na	na
	[3625.11]	[3573.27]	[3923.41]			
Female, no UB	9258	-84	18848	na	na	na
	[4350.85]	[4101.17]	[4284.37]			
Autumn courses						
Male, UB	8156	9183	19332	na	na	na
	[3077.35]	[2948.19]	[3337.35]			
Male, no UB	6006	11313	9122	na	na	na
	[4422.26]	[4552.15]	[4852.38]			
Female, UB	8839	13409	21641	na	na	na
	[2421.07]	[2401.03]	[2780.09]			
Female, no UB	1261	8360	7923	na	na	na
	[2628.39]	[2564.03]	[3001.56]			

4. Pre-training tests

Table A7. Earnings differentials between participants and non-participants in the pre-training years. By pre-training year, season, gender, unemployment benefit entitlement and training years.

Earnings one year before training (s=1)

Season	Winter				Autumn			
Training year	1993	1994	1995	1996	1993	1994	1995	1996
Male, UB	493	1827	5732	5885	-180	694	4675	7358
	[2587.64]	[3456.96]	[4691.82]	[5282.02]	[2771.96]	[3290.66]	[3721.59]	[4315.91]
Male, no UB	-744	-2786	-905	4028	-445	734	-378	-5395
	[3993.16]	[4077.56]	[3836.11]	[4094.48]	[2873.85]	[2869.86]	[2568.46]	[3185.19]
Female, UB	3472	-4463	5854	-167	503	2777	-1144	368
	[2782.59]	[3663.14]	[3860.23]	[4024.09]	[2177.88]	[2298.91]	[2527.83]	[2824.17]
Female, no UB	-3966	986	-1738	1387	-2480	-1878	-1271	1857
	[2450.31]	[2351.49]	[2481.96]	[2272.99]	[1755.63]	[1396.73]	[1359.33]	[1449.55]

Earnings two years before training (s=2)

Season	Winter			Autumn		
Training year	1994	1995	1996	1994	1995	1996
Male, UB	1706	2715	6615	-1913	3201	3734
	[2631.18]	[4364.53]	[5016.06]	[2934.93]	[3671.25]	[4349.99]
Male, no UB	-4083	218	3870	-3289	-1545	-5683
	[3930.46]	[3875.63]	[3851.32]	[2379.97]	[2494.84]	[3283.28]
Female, UB	4450	5183	6482	-85	3190	2337
	[2873.13]	[3790.57]	[3952.12]	[2085]	[2474.36]	[2758.5]
Female, no UB	990	-482	626	-2510	-1575	-697
	[2327.5]	[2353.48]	[2181.42]	[1271.63]	[1332.25]	[1407.47]

Earnings three and four years before training (s=3,4)

	s=3				*s=4*	
Season	Winter		Autumn		Winter	Autumn
Training year	1995	1996	1995	1996	1996	1995
Male, UB	-1727	885	-336	2569	3519	152
	[3761.84]	[4838.95]	[3275.88]	[3972.94]	[3573.9]	[3612.27]
Male, no UB	978	3369	1570	-3798	4155	-5802
	[4242.59]	[4627.81]	[2333.71]	[3203.64]	[4095.59]	[3059.75]
Female, UB	-3041	2167	-2240	4670	524	1904
	[3090.73]	[3737.01]	[2250.82]	[2600.38]	[2834.53]	[2376.02]
Female, no UB	-3516	1122	-383	-1231	-3073	-392
	[2669.96]	[2142.33]	[1203.68]	[1403.5]	[2189.47]	[1312.85]

5. Job Opportunity Indicators

While the LFS unemployment rate is the official unemployment rate produced by Statistics Norway, the alternative indicator is estimated from individual unemployment spells in the Frisch Centre Data base covering *all* registered spells throughout the 1990's. Note that the our alternative JOI is constructed by means of all spells and not only those covered by the study of training effects.

Outflow from unemployment

The computation of our job opportunity indicator is based on a hazard rate model with the exit probability from unemployment as the dependent variable. This approach has been motivated by Røed (2001) and Røed and Zhang (2002). It resorts on the idea of a proportional hazard model in that the hazard rate is proportional in factors depending on calendar time, spell duration and (time varying) explanatory variables. Let *d* denote actual duration of unemployment spell, *t* denote calendar time, λ_d be parameters of duration *d*, and σ_t be parameters associated with calendar time. Note here we make no distributional assumptions on both duration and calendar time. Therefore λ_d and σ_t are estimated non-parametrically. The monthly probabilities of individual exits from unemployment are then parameterised as follows:

$$h(t,d,x_t) = 1 - \exp(-\exp(\sigma_t + x_t^{'}\beta + \lambda_d))$$

where $\exp(\sigma_t + x_t^{'}\beta + \lambda_d)$ is interpreted as the integral taken over an underlying continues time hazard rate for the time interval corresponding to spell duration month number *d* (Prentice and Gloeckler (1978), Meyer (1990)), hence the parameters can be interpreted in terms of the underlying hazard rate. The vector of explanatory variables, x_t , includes a total number of 43 covariates capturing age, gender, educational attainment, immigrant status, and dummy for part time employment. Taking the complete administrative unemployment register from 1990 to 1999, we estimate this model on prime age individuals from age 25 to 59, which are entitled to unemployment benefit. The total number of monthly observations used for the estimation of equation is 10,182,300. Table A8 provides summary statistics of the estimation sample.

Job Opportunity Indicators in Table 4, 5 and 6

To calculate the job opportunity indicator used in section 9, we estimate predicted monthly exit probabilities for representative mean individuals for the period of 1992 to 1997, separately for men and women. The representative individual is constructed by taking mean value of all explanatory variables $\overline{x}_t$ for both men and women. The estimated monthly transition probabilities are then

$$\hat{h}(t,\overline{d},\overline{x}_t) = 1 - \exp(-\exp(\hat{\sigma}_t + \overline{x}_t'\hat{\beta} + \hat{\lambda}_{\overline{d}}))$$

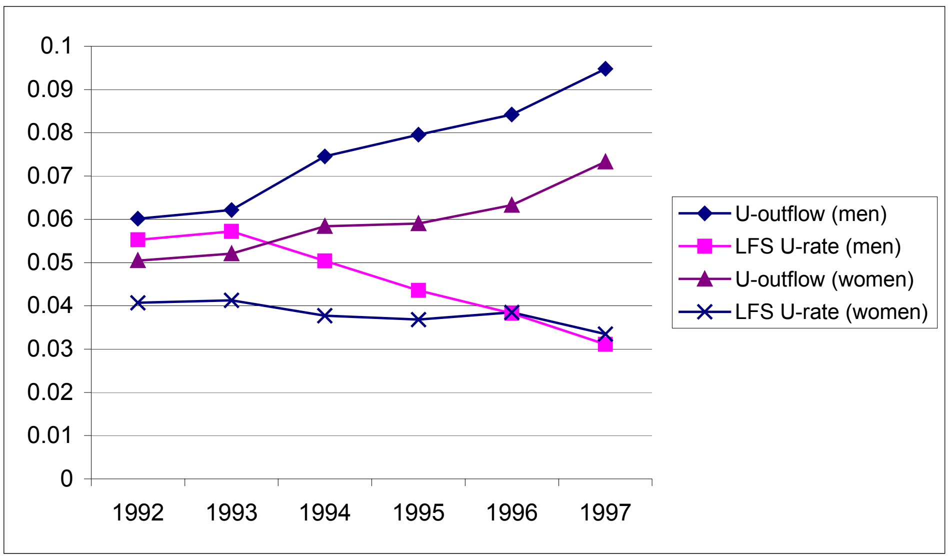
We then take the yearly average of these as proxies for the aggregated job opportunity indicators for each of analysing year. Røed and Zhang (2002) provides detailed discussion on properties of this indicator and extension to the case of mixed proportional hazard model.

Table A8. Summary statistics of estimation sample

	Men	*Women*
Number of monthly observations	4 741 250	5 441 050
Mean transition	0.082	0.066
Mean duration	10.82	11.52
Mean age	36.78	37.35
Education attainment		
Primary school (<= 9 years)	0.17	0.18
1 year Secondary school (10 year)	0.30	0.37
Vocational school (11-12 years)	0.23	0.13
Secondary school (12 years) ref	0.11	0.19
3 years college engineer education (13-15 years)	0.02	0.00
Lower college/university (13-15 years)	0.06	0.06
Higher university (> 15 years)	0.07	0.05
Unknown	0.06	0.04
Part-time employment	0.22	0.42
Married	0.33	0.55
Non-OECD immigrant	0.12	0.07

Figure A1 compares the (estimated) national monthly unemployment outflow rate and the standard unemployment rate from the Labour Force Survey (LFS), by gender for the period 1992 to 1997. Røed (2001) has showed that unemployment rate (as well as the aggregate outflow rate from unemployment) lags behind the estimated job opportunity indicator, which reflects a systematic sorting effect on unemployed over the business cycle. The figure also depicts that the two JOI's tell different stories about the gender differential in job opportunities. Although men have experienced higher unemployment rates than women, the unemployed men are more likely to leave unemployment and enter jobs, than women. We find the job opportunity indicator based on estimated transition probabilities to be a better measure of the how the job opportunities of unemployed job seekers vary over the business cycle.

Figure A1. Unemployment outflow and the LFS unemployment rate.



Job Opportunity Indicator in Table 6

The model where the business cycle effect is identified by within county variations in outflow rates uses the county level variable for men and women as displayed in Table A9.

Table A9. County level (monthly) outflow rates. By gender for 1992-1997.

Men		***1992***	***1993***	***1994***	***1995***	***1996***	***1997***
	Østfold	0.052	0.058	0.075	0.084	0.082	0.098
	Akershus	0.053	0.055	0.077	0.084	0.089	0.103
	Oslo	0.046	0.050	0.064	0.072	0.075	0.083
	Hedmark	0.052	0.058	0.079	0.085	0.088	0.094
	Oppland	0.058	0.062	0.070	0.082	0.089	0.098
	Buskerud	0.053	0.057	0.075	0.087	0.092	0.105
	Vestfold	0.060	0.061	0.079	0.089	0.089	0.109
	Telemark	0.054	0.061	0.074	0.081	0.084	0.093
	Aust-Agder	0.072	0.073	0.086	0.089	0.098	0.114
	Vest-Agder	0.073	0.066	0.085	0.086	0.100	0.101
	Rogaland	0.082	0.076	0.075	0.076	0.088	0.100
	Hordaland	0.052	0.066	0.073	0.076	0.083	0.091
	Sogn og Fjordane	0.086	0.088	0.091	0.099	0.107	0.115
	Møre og Romsdal	0.080	0.073	0.090	0.100	0.100	0.119
	Sør-Trøndlag	0.058	0.056	0.069	0.079	0.084	0.093
	Nord-Trøndlag	0.068	0.062	0.070	0.072	0.080	0.083
	Nordland	0.067	0.064	0.071	0.076	0.079	0.094
	Troms	0.069	0.069	0.078	0.078	0.086	0.099
	Finnmark	0.084	0.085	0.077	0.071	0.064	0.073
	Mean	*0.064*	*0.065*	*0.077*	*0.082*	*0.087*	*0.098*
	St.dev	*0.013*	*0.010*	*0.007*	*0.008*	*0.010*	*0.011*
Women		***1992***	***1993***	***1994***	***1995***	***1996***	***1997***
	Østfold	0.040	0.045	0.053	0.056	0.059	0.067
	Akershus	0.049	0.052	0.060	0.066	0.070	0.083
	Oslo	0.046	0.049	0.057	0.063	0.067	0.074
	Hedmark	0.045	0.050	0.053	0.054	0.058	0.068
	Oppland	0.050	0.053	0.053	0.057	0.062	0.070
	Buskerud	0.049	0.049	0.055	0.062	0.068	0.081
	Vestfold	0.047	0.046	0.062	0.058	0.061	0.077
	Telemark	0.047	0.048	0.056	0.054	0.059	0.073
	Aust-Agder	0.055	0.058	0.058	0.060	0.068	0.083
	Vest-Agder	0.055	0.052	0.062	0.060	0.068	0.078
	Rogaland	0.058	0.058	0.063	0.059	0.063	0.075
	Hordaland	0.044	0.051	0.055	0.058	0.062	0.071
	Sogn og Fjordane	0.070	0.066	0.077	0.070	0.080	0.086
	Møre og Romsdal	0.051	0.052	0.062	0.062	0.071	0.074
	Sør-Trøndlag	0.047	0.047	0.057	0.059	0.060	0.069
	Nord-Trøndlag	0.047	0.049	0.053	0.051	0.053	0.059
	Nordland	0.054	0.052	0.055	0.053	0.056	0.067
	Troms	0.056	0.055	0.062	0.057	0.062	0.073
	Finnmark	0.083	0.087	0.076	0.068	0.063	0.068
	Mean	*0.052*	*0.054*	*0.059*	*0.059*	*0.064*	*0.073*
	Std.dev.	*0.010*	*0.009*	*0.007*	*0.005*	*0.006*	*0.007*

Chapter 4: Identifying treatment effects of active labour market programmes for Norwegian adults

By Tao Zhang

Identifying treatment effects of active labour market programmes for Norwegian adults

By Tao Zhang

The Ragnar Frisch Centre for Economic Research

Abstract

We investigate treatment effects of active labour market programmes for Norwegian adults for the 1990 to 2000 period. Three types of active labour market programmes are evaluated within a competing risks hazard rate model. Non-parametric specifications on both duration dependence and unobserved heterogeneities are used. By utilising rich administrative data, we find that active labour market programmes do have intended effects on enhancing the transition probability to employment *after* the completion of programmes participation, but *during* the participation, the transition probability is low relative to that for non-participants. There is some evidence of heterogeneity of treatment effects with respect to observed individual characteristics, and effects for training programmes and wage subsidy programmes are pro-cyclical and more favourable at boom time. The positive treatment effects of labour market programmes are long lasting, at the same time diminish gradually over time when individuals remain in unemployment after completion of programmes. The net impact of active labour market programmes in terms of reduced total amount of unemployment exposure is estimated to be about 6.42%.

Keywords: labour market programmes, treatment effects, competing risk, non-parametric estimation.
JEL Classification: C41, J24, J64.

* Thanks to Knut Røed, Oddbjørn Raaum, Hege Torp and Steinar Holden, and participants at Frisch Seminar at the Ragnar Frisch Centre for helpful comments and suggestions for improvement.

1. Introduction

Active labour market programmes have been widely applied in attempts to combat rising unemployment during the past decades. In most OECD countries, active labour market programmes have been used extensively when the economy has been at slump. In continental Europe, particularly in the Nordic countries, labour market programmes have a long tradition and major status in government policy consideration.

Evidence on the impact of active labour market programmes displays a mixed picture. In the US, no clear conclusion has been made on the effects of programmes in terms of enhanced employment opportunity and job perspective. In Europe, some encouraging results regarding the success of programmes have been seen lately. But equally many empirical studies have showed no or even negative effects. For a general survey, see e.g. Heckman et al (1999) and Fay (1996). Some recent evaluation literature on the European active labour market policies can be found in e.g. Fertig et al (2002) and Gerfin and Lechner (2002). It is notable that quite a few studies have showed that the Swedish model, which is associated with the most ambitious active labour market policy of all, has failed to produce convincing evidence of favourable treatment effects, see e.g. Ackum et al (2001) and Calmfors et al (2001).

Recent studies on the evaluation of Norwegian active labour market programmes have, however, produced some encouraging results. Raaum et al. (2002) have found that with income as a measure for post programmes success, labour market training programmes have a significant positive treatment effect. They also find that the effect is strongly influenced by business cycle conditions and is strongest when job opportunities are favourable. Røed and Raaum (2003) have evaluated the total effects of participation in active labour market programmes and also found positive impacts on the transition probability from unemployment to employment. Aakvik et al (2000) have provided some positive evidences on the Norwegian rehabilitation training programmes.

In this paper, we evaluate the Norwegian active labour market programmes within the model framework of unemployment duration analysis. We model the assignment of treatment in terms of hazard rates, and evaluate the resulting changes of transition probability to job *during* and *after* programmes participation as a measure for success. We apply a 5-state competing risks model, non-parametric specification for both duration dependence and unobserved heterogeneity. The econometric approach used in this paper starts out from the dependent competing risks treatment evaluation framework provided by Røed and Raaum (2003). However, we extend their model in three directions: First, while Røed and Raaum (2003) view the "treatment" as a single one-dimensional state, we model the selection into and the causal effects resulting from programme participation separately for three different types of programmes. Second, while Røed and Raaum (2003) treat the duration of the programmes as exogenous, we model the duration of the programmes as part of the competing risks structure. And finally, while Røed and Raaum (2003) model the treatment effects as constant at individual level (except from a "dying out" effect after completion), we model the treatment effects as varying freely from month to month, both through the participation phase and afterwards. Our findings suggest that in general, the basket of active labour market programmes has a significant positive impact on the probability of employment.

The difficulty that lies beneath all studies on the labour market programmes effects is that in the studies on observational data, assignment of treatment is unlikely to be totally random due to the population heterogeneity. In evaluation of treatment effects, the fundamental problem is thus the unobserved heterogeneity (Heckman et al 1999). If the treatment is assigned to a subpopulation with systematically different characteristics than the population as a whole, we would inevitably have a sample selection problem. Heckman (1979) was the first to analyse the impact of such selectivity bias due to population heterogeneity. It is well conceived that participants who receive treatment may have some unobserved characteristics that researchers cannot assess, e.g. that they are more inclined to participate and more responsive towards the treatment. If the assignment of treatment is not random, the outcome of such treatment can be driven by the same factors that influence the probability of receiving treatment itself. Fail to controll the unobserved heterogeneity in the form of self-selection into the treatment, the effect of treatment is obviously estimated with bias.

The ideal way to avoid such self-selection bias (also administrative selections) is to randomise the assignment of treatment to identically assembled sample such that the outcome is not conditional on factors that influence the assignment. This is a common practice in experimental studies such as in medicine and biology. Rubin has done an extensive research on the causal effect of treatment within the experimental settings (e.g. Rubin (1974), Holland (1986)), and maintained that causal effect can only be identified through experiments (Holland (1986), also see Lalonde (1986)).

Some methods have been developed to minimise the risk of selection bias in analysing non-experimental data. Matching techniques have been applied widely in evaluating treatment effects (Heckman et al. (1997)). A central assumption behind the matching technique is the conditional independence assumption, which means that conditional on observed heterogeneity, the effect of treatment does not depend on the assignment probability of such treatment. Under this assumption, by matching treated with untreated who has "similar" observed characteristics, the estimated effect is unbiased. Nevertheless, in practice, it is not always possible to ensure that observed characteristics catch up all the population heterogeneity. In addition, the conditional independence assumption is a rather strong assumption and justification is not without difficulty in practice.

Another method in tackling independence of treatment assignment is the exclusion assumption. This means that by utilising some instrumental variables that enter the determination of treatment assignment but do not affect the outcome, the probability of assignment and effect of the treatment are not perfectly correlated. It is however difficult in practice to find such instrumental variables and justification of independence is often questionable. Some other methods such as difference-in-difference have been developed in the evaluation literature. A comprehensive reference is Heckman et al. (1999).

Most of such methods have a style of "binary-choice and binary-effect". This means that the assignment of treatment is modelled by a binary (or multivariate) choice model and the effect of treatment is modelled in a similar way. They are mostly static evaluation practices and the time until treatment and the time until outcome are usually ignored. Abbring and van den Berg (2003b) have argued that it is precisely the time until treatment and time until outcome that convey important information in capturing the selection into the treatment and selection towards outcome. They suggest that duration modelling framework is suitable in treatment evaluation, and prove that the treatment effect is non-parametrically identified within the context of a mixed proportional hazard rate model. Richardson and van den Berg (2002), Lalive et al. (2001), and Røed and Raaum (2003) are recent applications in evaluating treatment effects using duration model framework. In this paper, we adopt identification results of Abbring and van den Berg (2003b), also identifications based on time-varying covariates suggested by McCall (1994) and Brinch (2000). We propose here a dynamic analysis of treatment evaluation within the context of duration model with unobserved heterogeneity.

The rest of this paper is organised as follows: Section 2 gives a brief introduction to the institutional settings of Norwegian active labour market programmes, also describes the data at hand and estimation strategies. Section 3 gives an account of econometric theory and modelling of the treatment evaluation problem. Focus is given to the identification of treatment effects in competing risks model and advocate the use of duration modelling. Section 4 presents main results. We will show the estimation of determinants of programmes participation, also present estimated treatment effect in a dynamic setting. Section 5 concludes and offers some policy implications.

2. Norwegian labour market programmes and data used in this analysis

Norwegian active labour market programmes have been important policy tools to combat rising unemployment for many years and have been applied extensively during the past decades. One of the stated goals of active labour market programmes is to increase the employability of the participants. In addition to its primary intention, the active labour market programmes also have a number of welfare implications. By admitting unemployed workers to programmes with some form of allowance or economic compensation, it may prevent poverty and avoid individuals from dropping out of labour market, and maintain their social network.

Active labour market policy also serves as an incentive scheme particularly for unemployment benefit claimants. For most of the period covered by the analysis in this paper, unemployment benefit claimants were entitled to benefit for a maximum duration of 186 weeks (about 47 months), but with a possible cut-off period of 13 weeks after the first

80-week period (18-20 months). If the unemployed fails to meet certain criteria for active job search after the exhaustion of the first benefit period, the benefit is cut-off for a quarantine period. Although strict enforcement rules of cut-off were rarely applied, benefit claimants have often been required to participate in some programmes in order to maintain the benefit entitlement during or after the quarantine period.

The scope and volume of active labour market programmes are adjusted according to the overall unemployment situation. When in the slump time, a wide range of programmes are offered to unemployed, while in boom time the programmes are scaled down. All programmes evaluated here are offered and organised by the public employment services – some of them in cooperation with other agents, both private and public.

Since the one of the primary goals of active labour market programmes is to increase the employability of unemployed persons, it is natural in this paper to define the success of programmes by the enhancement of the employment probability. We use the term *treatment* to denote the participation in the active labour market programmes. The *causal effects* of programmes are measured by the changes in transition probability from unemployment to employment for programme participants. For a systematic study of programme effects, we classify the Norwegian active labour market programmes into 3 groups[24]:

1. Labour market training programmes. This group mainly consists of formal training courses offered by the public employment services. It is mainly a qualification scheme. By participating in courses in different areas including general and occupational specific trainings, the participants are able to improve their individual qualifications for either their existing occupations, or the new careers. Also, some special courses are offered to immigrants in order to improve their language skills. The duration of these programmes varies, but most of them last for 1-5 months. Some courses are preparatory, leading to some more advanced courses next term. Thus we may observe periods of programme participation last for up to 10 months. Admission usually takes place in the start of spring and autumn seasons. But in reality, exceptions are often made to suit the individual's particular needs. Training

programmes constitute almost half of active labour market programmes in Norway. They are in principle open for all unemployed jobseekers, and no specific qualification is required for participation.

2. Temporary employment in public sectors (including voluntary sector). This group of programmes is targeted at the long-term unemployed, or those who are particularly "hard-to-employ". Duration of programmes is normally up to one year. By offering temporary placement in public sectors, the programmes aim to prevent unemployed from dropping out of the labour force. The employment programmes were offered in large scale during the last slump period in early 1990's, but reduced dramatically as labour market conditions became favourable in late 1990's. From the year 2000, the employment programmes have no longer been offered.

3. Wage subsidy, stand-in jobs, courses in active job search, etc. The wage subsidy programmes work with private establishments to employ jobseekers, while part of the wages are subsidised by governmental employment offices. Stand-in jobs are on temporary basis, but are normal employments by nature. Courses in active search are aimed to provide information about job market and personal adjustments to fit in, etc. We group these programmes together under the name "wage subsidy programmes". This group of programmes is aimed to assist ready-to-work jobseekers by enhancing their competitiveness in the job market, hence the participants are usually more qualified than those in other programmes.

The motivation for us to focus on these 3 groups of programmes and evaluate effects of participation in these programmes separately is this: it is obvious from the design of ordinary active labour market programmes that the targeted participants of different programmes are different. The wage subsidy programmes are targeted to those ready-to-employ jobseekers with higher qualification, while the employment programmes are targeted long-term unemployed who have the most difficulties in the job market. Thus admission to different programmes is selective both in terms of individual characteristics, and in terms of administrative admission requirement (administrative selection), see e.g.

[24] For a description of Norwegian active labour market programmes, see e.g. Torp (1995).

Røed et al. (2000) for a detailed exposure. Secondly, the treatment effects of different labour market programmes are likely different due to different selection mechanisms into the programmes. If evaluating various programmes aggregately, the total effects would be driven by shares and compositions of participants of different programmes.

The data we use in this analysis is from a wide range of official administrative registers collected at the Ragnar Frisch Centre for Economic Research. They include unemployment registers for the whole Norwegian unemployment population from 1989 to 2000, combined with detailed demographic information for the Norwegian population collected in 1993-1997, and detailed labour market experiences from 1967 to 1999.

In this analysis we focus on the core of the labour force, i.e. adult male and female job seekers, aged 25-50, not temporarily laid off, who have been full time employed for at least 12 months prior to entering the registers as unemployed. All of them are entitled to unemployment benefits (i.e. they are entitled to about 62.4 per cent compensation of previous before-tax earnings up to a ceiling for a period about 47 months). The reason for putting these restrictions on the sample is to have a pretty homogenous analysing population when it comes to preferences and labour market options. For unemployed members of the core of the labour force, gainful employment is supposed to be the preferred and dominant transition. For very young and senior unemployed, other options as well - such as education and retirement - are possibly both preferred and available. Restricting the analysis to the core of the labour force enables us to identify the effects of the evaluated programmes on the employment for participants who are motivated for returning to employment. At the same time we are essentially excluding out-of -labour-force as a possible transition state. The purpose of restricting our attention to those entitled to unemployment benefits is due to registration phenomena. Without unemployment benefit, the incentive to register with public employment service is weak; therefore the unemployment spells are likely not correctly measured for non-benefit-claimants. To avoid the contamination of data due to incomplete unemployment registration, we censor the spell once the jobseeker loses his or her unemployment benefit.

Our observational window is set from January 1990 to December 2000. Since employment programmes ceased to exist from year 2000, and there were very few participants already from 1998, we censor the employment programmes from January 1998. For each month, we record the status of unemployment, along with observed individual characteristics. All observed individual characteristics are in principle time-varying if their values change during the spell.

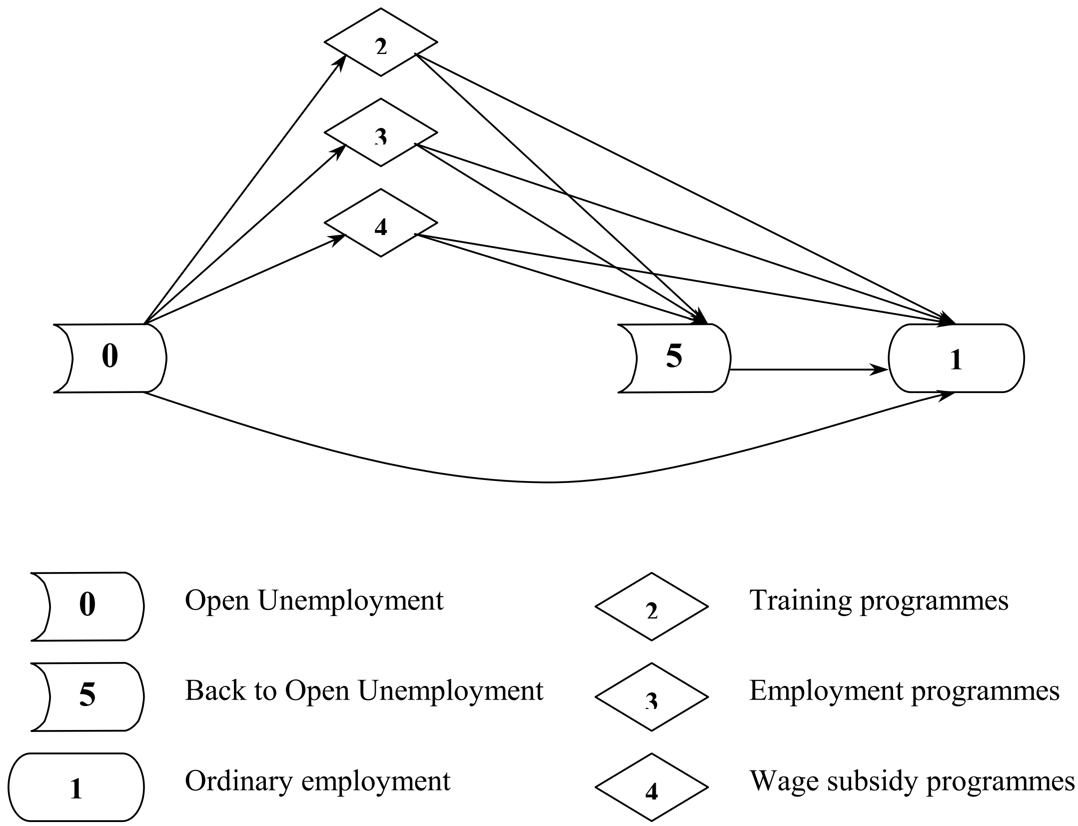
Each individual enters our analysing data as a new entrant to open unemployment. We define 4 possible transitions from the entrance to open unemployed: to ordinary employment and to 3 labour market programmes defined above. The duration of the unemployment spell is defined in the following way: the spell starts as open unemployment with 4 possible transitions (after January 1998, there are only three possible transitions: ordinary employment, labour market training programmes and wage subsidy programmes). We follow the spell until there is a consecutive two months of registered part-time job observed, or a consecutive three months absence from the unemployment register[25]. We define it as a successful transition to employment. If a transition to any of the 3 programmes has occurred, we then followed the spell further until a termination (defined below) is observed. When the individual is participating in labour market programmes, we only allow two possible transitions: back to open unemployment or to ordinary employment. If individual completes a labour market programmes and returns to open unemployment with unemployment benefit, we only allow one possible transition: to ordinary employment.

To better illustrate the dynamics of transitions between unemployment, labour market programmes and ordinary employment, Figure 1 provides a flowchart for these complex processes. The arrow lines indicate the directions of possible transitions an individual can take from the current state he or she is occupying. Let j be the origin states and k be the destination states for transitions. The individual enters open unemployment as state $j=0$ and facing 4 possible transitions indicated by four arrow lines ($k=1,2,3,4$). Once a transition to one of the programmes is made, e.g. in Figure 1, if the individual is at state $j=1$, the only possible transition is $k=5$ (back to open unemployment after completing the programmes)

[25] The part-time jobs are registered in the unemployment register. However, we do not have the complete employment register for the whole analysing period. Therefore, we have to rely on such criteria to define transition to employment.

or to state $k=1$ (ordinary employment). If the individual is at state $j=5$, then the only possible transition is to job ($k=1$), indicated by the single arrow line.

Figure 1: Flowchart of dynamic transitions between unemployment, active labour market programmes and ordinary employment.



Spells are terminated either due to transition to ordinary employment, or censored. Spells are censored after 36 months of the total length, as the observed frequencies of transitions are too low for a precise estimation after 36 months. Also we restrict our attention to only one treatment at a time. This means that if an individual has taken a transition from one programme to another, we censor the spell accordingly at the transition month. This is because we are interested in evaluating each treatment effect on the transition probability to job alone. If we allow multiple programmes participation, the estimated effects may be the joint effects of multiple programmes. Our censoring scheme allows us to isolate the pure effect on transition probability to job from each and one programme alone, and avoid complicated joint effects of multiple treatments.

After careful preparation, we have to our disposal 115,557 individuals with 126,034 spells. The total number of monthly observations in our estimation sample is 664,250. Table 1 gives a summarizing view of analysing data. A quick glance of Table 1 reveals that there are more men (57%) than women in our sample, probably as a result of the previously full-time work requirement. Average age is around 35. We find that those with low educational attainment (less than high school) are in majority with 78%. Average time spent in unemployment is approximately 5.27 months. For those eventually participating in some labour market programmes, time spent before participation is on average 7.53 months. Mean duration of programmes is 3.45 months. Those remaining unemployed after participation would have on average 4.41 more months before a possible job transition (or being censored). Of all 126,034 spells, 93,528 have a successful transition to ordinary employment. Among these, 6,624 make the transition with the assistance of participations in the labour market programmes.

3. Econometric model and identification of treatment effect.

We consider a mixed proportional hazard rate model with k competing destination states of transitions from origin state j over a continuous time τ . The transition specific hazard rate can be defined by

$$(1)\ \theta_{jk}\left(\tau \mid \mathbf{X}_{jk\tau}, v_k\right) = \lim_{\Delta t \to 0} \frac{P(\tau \leq T_{jk} \leq \tau + \Delta\tau, K = k \mid T_{jk} \geq \tau, \mathbf{X}_{jk\tau}, v_k)}{\Delta\tau} = \lambda_{jk}(\tau) \cdot \phi(\mathbf{X}_{jk\tau}) \cdot v_k$$

here $\lambda_{jk}(\tau)$ is the underlying duration baseline associated with transition k; $\phi(\mathbf{X}_{jk\tau})$ is the structure term of covariates affecting the transition specific hazard rate. Note the subscript τ and k, which indicate that the effects of covariates can be transition specific and time-variant. v_k is meant to capture the unobserved transition specific heterogeneity with unknown distribution. We assume v_k is constant throughout the spell duration. In applied research, we often encounter discrete time units, which might be due to the observational or data sampling practice. In Norwegian official registers for unemployment, the available data is commonly updated at the end of each calendar month, which implies that

Table 1: Statistics of estimation data.

# of individuals	115,557	
# of spells	126,034	
# of monthly observations	664,250	
Means*	**mean**	**std**
Gender (1=male)	0.5744	0.4944
Age (years)	34.9793	7.3265
Married (1=yes)	0.4499	0.4975
Having children under 18 years (1=yes)	0.5682	0.4953
Non-OECD immigrants (1=yes)	0.0529	0.2238
Immigrants with Norwegian citizenship (1=yes)	0.0266	0.1609
Having relevant experience for intended job	0.9274	0.2595
Having relevant training for intended job	0.8107	0.3917
County of residence		
Akershus, Hedmark, Oppland, Buskerud	0.2358	0.4245
Vestfold, Telemark, Aust-Agder, Vest-Agder	0.1803	0.3845
Rogaland, Hordaland	0.1333	0.3399
Sogn og Fjoldane, Møre og Romsdal, Sør-Trondlag, Nord-Trondlag	0.1507	0.3578
Nordland, Troms, Finnmark.	0.1077	0.3101
Education attainment (percent)		
up to 9 years	0.7812	0.4134
10 years	0.1568	0.3636
11-12 years	0.0616	0.2405
13-16 years	0.0002	0.0138
17 or more years	0.0002	0.0146
Occupational Background		
Technical, physical science, humanistic and artistic	0.2178	0.4127
Administrative executive work, clerical work and sales work	0.2979	0.4574
Agriculture, forestry, fishing and related work	0.0177	0.1319
Manufacturing work, mining, quarrying, building and construction work (reference)	0.2767	0.4474
Service work, transport and communication	0.1817	0.3856
Unspecified	0.0081	0.0898
Transitions (# of spells)	**#**	
To Job	93,528	
Have not participated in Active labour market programmes	86,904	
Have participated in Active labour market programmes	6,624	
To Training programmes	10,732	
To Employment programmes	2,106	
To Wage subsidy programme	7,533	
Back to open unemployment after participation in programmes	6,978	
Censored	26,538	
Durations (months)**	**mean**	**std**
Average total spell durations	5.2704	5.7098
Average durations until programmes	7.5313	6.0819
Average durations after programmes	4.4110	4.2507
Average durations of programmes	3.4599	2.8795

Note: * means are calculated on spell basis. ** means are calculated within relevant groups.

the smallest reliable time unit is month. In the case of some discreteness, a convenient assumption is that the competing hazard rates are constant within each time unit. This is probably an innocent assumption, provided that the time unit is small. In our case, let *d* be discrete time unit (e.g. month, $d = 1, 2, \ldots$), we can for example define the integrated hazard rate within interval $[d - 1, d]$ as (for transition from *j* to *k*):

$$\int_{d-1}^{d} \theta_{jk}(u \mid \mathbf{X}_{jk\tau}, v_k) du = \int_{d-1}^{d} \lambda_{jk}(u) \cdot \phi(\mathbf{X}_{jk\tau}) \cdot v_k du$$

As in the dynamic transition process depicted in Figure 1, we define origin and destination states as:

$j,k = 0$ open unemployment
$j,k = 1$ ordinary employment
$j,k = 2$ training programmes
$j,k = 3$ employment programmes
$j,k = 4$ wage subsidy programmes
$j,k = 5$ back to open unemployment

In our context of programmes evaluation, we start by model a four-state competing risk model from origin state of open unemployment, with mixing unknown distributions for unobserved heterogeneity for each state. We follow the spell until a transition to either of the four states has occurred. Note however that, once a transition to one of the labour market programmes has taken place, the possible transitions from labour market programmes are restricted to job and back to open unemployment again. Here we do not allow cross programmes transitions. While the individual is participating in the programme, the competing risks are reduced to only two possible transitions, *k=1* and *k=5* (back to open unemployment), with origin state be *j=2,3,4,* (participating in programmes). If the individual has finished participation in labour market programmes and nevertheless still remains unemployed, we define the origin state be *j=5* (have participated, but still unemployed after the participation), and the only possible transition is reduced to job (*k=1*).

We also introduce an important model term: the calendar variation, as time-varying dummy variables σ_t, to denote the calendar months *t* at which each individual is at the risk set. These are meant to capture the aggregate labour market conditions and business cycle and seasonal effects on the transition probabilities out of unemployment. There are totally 132 such dummies. The estimation of calendar dummies itself can be proven of great interest. Røed and Zhang (2003) have showed that the predicted hazard rate (from estimation of a single risk model) on each calendar dummies can have the convenient interpretation as the cyclical variation in the transition. Røed (2002) gives a detailed account on the properties of such estimates and possible business cycle interpretation of these.

We are only interested in treatment effects on job probability in this paper. Therefore we model the treatment effects only in the hazard rate for job transition. The total treatment effects are modelled by

(2) $\Delta_{kt} = \delta_{0k}(\mathbf{D}_{0kt} + \mathbf{x}_{\mathbf{0}kt} + b_t) \cdot I_0 + \delta_{1k}(\mathbf{D}_{1kt} + \mathbf{x}_{1kt} + b_t) \cdot I_1,\ \ \Delta_{kt}{=}0\ \mathit{if}\ \ k \neq 1.$

The (exponential of) Δ_{kt} is the aggregated treatment effect in calendar month t which affects hazard rate proportionally. Note that Δ_{kt} by definition only affects transitions to job. In order to fully assess the treatment effects of labour market programmes, we have made some decomposition of aggregated effects Δ_{kt}. First, we let effects vary *during* the participation and *after* the participation. δ_{0k} is the effect of labour market programmes while under the participation; δ_{1k} is the effect after the participation. We denote accordingly *while-treatment effects* and *after-treatment effects* respectively. I_0 and I_1 are indexing functions, which indicate if the individual is currently under participation and if the individual has participated in the labour market programmes earlier in the same spell, respectively.

Second, we allow the treatment effects to vary over time. In (2), $\mathbf{D}_{0kt}$ is a set of dummies to indicate 1,2, … months after the *start* of treatment, while $\mathbf{D}_{1kt}$ is a set of dummies to indicate 1,2, … months after the *completion* of treatment. By interacting treatment effects with these two sets of dummies, we can then fully examine the time pattern of treatment effects on an individual's transition probabilities, both while under treatment and after treatment.

We introduce further heterogeneous treatment effects for various demographic observables by letting while-treatment effects and after-treatment effects to be dependent on observed covariates such as gender, age and education, by introducing interactive terms. In (2)

$\mathbf{x}_{0kt}$ and $\mathbf{x}_{1kt}$ are vectors of individual characteristics that we wish to interact with treatment effects.

It is also conceivable that treatment effects may vary with respect to job opportunities and labour market conditions. Raaum et al (2002) find that the impact of labour market programmes varies over the business cycle. At least for training programmes, a strong pro-cyclical tendency has been found. Therefore it is of importance to look into the treatment effects within the context of business cycle. We also wish to investigate here how the while-treatment effects and after-treatment effects of programmesparticipation would be affected by business cycle conditions. To facilitate that, we let the treatment effects be interacted with business cycle indicators b_t. The business cycle indicator is taken from Gaure and Røed (2003). It is a vector of smoothed calendar time parameters for each calendar month from 1989 to 2002, estimated from a comprehensive hazard rate models for transitions from unemployment to employment. Similar usage of estimated outflow rates as business cycle indicators can be found in e.g. Raaum et al. (2002) and Røed and Zhang (2003). For a more detailed discussion of this set of business cycle indicators, see Gaure and Røed (2003).

Let $\lambda_{jkd} = \log(\int_{d-1}^{d} \lambda_{jk}(u)du)$, $\phi(\mathbf{X}_{jkt}) = \exp(\mathbf{X}_{jkt}'\beta_k)$, $\mu_k = \log(v_k)$, the integrated hazard rate for interval $[d-1, d]$ is thus

$$\varphi_{jk}(d,t,\mathbf{X}_{jkt},\mu_k) = \exp(\lambda_{jkd} + \mathbf{X}_{jkt}'\beta_{jk} + \sigma_{kt} + \Delta_{jkt} + \mu_k)$$

and the monthly transition probability from origin *j* to destination *k* is given by

$$(3)\ \ h_{jk}(d,t,\mathbf{X}_{jkt},\mu_k) = \left[1-\exp(-\sum_k \varphi_{jk}(d,t,\mathbf{X}_{jkt},\mu_k))\right] \cdot \frac{\varphi_{jk}(d,t,\mathbf{X}_{jkt},\mu_k)}{\sum_k \varphi_{jk}(d,t,\mathbf{X}_{jkt},\mu_k)}$$

for relevant *j* and *k*.

One important feature of our model framework in (3) is the dynamic definition of risk sets. Depending on which transition is realised, the subsequent risk sets an individual occupies are endogenously defined. Once a transition to a labour market programme has occurred, a new possible transition (back to open unemployment) is added, and the individual finds himself within another risk sets. In our model framework, we open for interdependence of different transitions by allowing unobserved heterogeneity μ_k to be dependent across transitions, i.e. $\operatorname{cov}(\mu_k, \mu_m) \neq 0$ for $k \neq m$.

Equation (3) has a familiar form of complementary loglog model. It has the advantage of flexibility, which we find is very suitable for non-parametric duration analysis. There is an obvious advantage to adopt non-parametric specification because in reality, the economic theories and observational data do not provide convincing arguments towards using any particular parametric functional form specifications. Arbitrary chosen functional form would involve the risk of misspecification, particularly for unobserved heterogeneity. Heckman and Singer (1984) have warned about the danger of "overparameterising" the unobserved heterogeneity, and suggested the approach of specifying a discrete distribution with support of unknown number of points to non-parametrically estimate the unknown mixing distribution. They have proved consistency of such non-parametric maximum likelihood estimators. Baker and Melino (2000) provide Monte Carlo evidence for single risk models. Their conclusions are more in favour of semi-parametric specification, where the unobserved heterogeneity is modelled by discrete mass points, while the duration dependence is modelled by some parametric family. Zhang (2003) has done an extensive Monte Carlo study on models where both duration dependence and unobserved heterogeneity are specified non-parametrically. He found that non-parametric specified duration dependence and unobserved heterogeneity can be consistently estimated. Further more, even when the true underlying distribution is parametric, non-parametric estimation can still produce reasonable approximations. He also showed the evidence of consistent estimation for competing risks model with bivariate normal distributed unobserved heterogeneity. In our model, we have 5 distinct competing states and 5 mixing distributions for unobserved heterogeneity. We find it difficult to apply any parametric functional form on all duration dependences, not to mention combinations for unobserved heterogeneity. Therefore the non-parametric specification is especially suitable in our context. We have chosen to model the duration dependence and unobserved heterogeneity non-parametrically by using step functions for λ_{jkd} and μ_k in equation (3).

The calendar time of commencing and elapsed spell duration can function as additional identification source for unobserved heterogeneity. The intuition behind this is as follows:

In applied study, it is typical that local or macro economic environments will have effects on the transitions from unemployment to work. Consider two individuals that are identical in every observed aspect and have the same length of elapsed unemployment spell. Given the assumption of proportional hazards, these two should experience the same hazard rate if they have the same value of unobserved heterogeneities. But if one experiences unemployment during a slump period when "everyone" is hit by the unemployment risk while the other starts unemployment in a boom time when job opportunity is good and the overall outflow rate is high, it is intuitively plausible that the individual being unemployed at the boom time should have a better job opportunity and shorter duration than that of the "identical twin" in the slump time. The fact that they have the exact length of spell can then only be accredited to the unobserved differences between them (plus random factor). It is therefore likely that the one unemployed in boom time might have somewhat unfavourable personal characteristics that comparing to the one in the slump time with the same spell length, which implies lower chance of getting employed, even though the observed characteristics are identical. The same argument can apply on the transitions to different labour market programmes as well. Given the same length of elapsed spells, the different hazard rates for transition to one type of labour market programmes of two otherwise identical individuals must reflect different unobserved characteristics associated with the programme transition. This is to say that, time of the unemployment spell taking places and undergoing is the only source of hazard rate variation, cerates paribus. Therefore by including control for such exogenous variations of calendar time within the hazard rate formulation, the identifiability of unobserved heterogeneity should be greatly improved.

Identification of such competing risks model has been a focal point in the hazard rate model literature, see Heckman and Honoré (1989), McCall (1997) and Abbring and van den Berg (2003a). The general review on identification of competing risks duration model with time-varying covariates can be found in e.g. van den Berg (2001). Abbring and van den Berg (2003a) have proved that under proportionality and some regularity assumptions, the dependent competing risks model is non-parametrically identified. McCall (1996, 1997) has showed some identification results when models possess time-varying covariates. Brinch (2000) has proved that with time-varying covariates, the proportionality assumption can be relaxed, and the mixed hazard model is identified non-parametrically. Zhang (2003) provides Monte Carlo evidences both for single risk and competing risk models showing the advantages of including time-varying calendar variations as identification sources for unobserved heterogeneity.

Abbring and Van den Berg (2003b) have discussed and proved that under some regularity conditions, the treatment effect is identified non-parametrically within the duration model framework. The *timing-to-event approach* is suitable for treatment effect estimation in several aspects: 1. *Randomness in treatment assignment*: Though in practice the determinants that affect the assignment of treatment are never fully known, they are modelled in the form of competing risks hazard rates, therefore whether an individual is receiving the treatment is characterised by a transition probability, which by definition of probability itself ensures the randomness in assignment. 2. *Selection problems:* as elaborated earlier, it is the unobserved population heterogeneity that produces selection biases on the evaluation of treatment effects. Control of the observed heterogeneity can minimise the impact of selection on treatment effect, but never fully eliminate the source. In the mixed hazard rate model, not only the observed heterogeneity is fully modelled, but also the unobserved heterogeneity is taken into account by mixing its distribution with the hazard rate. Moreover, by allowing the unobserved heterogeneity associated with different transitions to be correlated, the selection is captured by the correlation coefficients of unobserved heterogeneity across transitions. By ensuring randomness in treatment assignment and controlling for selection bias due to unobserved population heterogeneity, the causal effect of treatment can be successfully revealed in hazard rate model framework, see Abbring and van den Berg (2003b).

The model is estimated with maximum likelihood method. Due to the complexity of transition processes, we find it convenient to divide the total duration of a spell into 3 segments: duration before possible transitions to labour market programmes, duration while participating, and duration of post-programme period. Let

$\kappa_t = 0$ if $j = 0, k = 1,2,3,4$

$\kappa_t = 1$ if $j = 2,3,4, k = 1,5$

$\kappa_t = 2$ if $j = 5, k = 1$

Define $d = \sum_{\kappa_t} d_{\kappa_t}$, where d_{κ_t} is the duration associated with each spell segment according to κ_t .

We use spells, rather than individuals as the basic unit for unobserved heterogeneity as well as likelihood formulation. This implies that we have ignored the information provided by multiple spells of the same individual. Although repeated spells are valuable sources for identification of unobserved heterogeneity (Honoré (1993)), we find several reasons to use spells after all. First, it is not likely that the unobserved characteristics for an individual would remain constant across spells, especially since we have conditioned the entrance to our data on the 12 months absence rule. Thus treating repeated spells from the same individual as independent spells in our context is more reasonable. Second, persons with repeated spells are not likely to be representative. This is because the length of the second spell is inversely related to the length of the first spell, given the observational window. This might have imposed some possible selection problem that persons with multiple spells are likely to have shorter earlier spells. Also from the statistics showed in Table 1, there are relatively few individuals with repeated spells. Therefore we feel using spell as unit for unobserved heterogeneity is justified.

We model the unobserved heterogeneity in the form of a discrete distribution with w different mass points. Let p_w be the probability of a particular combination of unobserved variables, $\sum_w p_w = 1$. Also note that since calendar time and spell duration effects are varying from month to month, we have to divide each spell into many one-month long subspells, which sum up to original spell length. This is a known technique in dealing with time-varying covariates.

Let $Z_{\kappa t}$ be an dummy indicator variable that:

$Z_{0t} = 1$ if $\kappa_t = 0$, $Z_{0t} = 0$ otherwise
$Z_{1t} = 1$ if $\kappa_t = 1$, $Z_{1t} = 0$ otherwise
$Z_{2t} = 1$ if $\kappa_t = 2$, $Z_{2t} = 0$ otherwise

This implies that $Z_{z+1} = 1 \Rightarrow Z_z = 1$, for $z = 0,1$. Further let i be spell id, the segmental likelihood for duration d_{κ_t} of spell i is then given by

$$(4)\quad L_{i\kappa_t} = \left[\prod_k h_{jk}(d_i, t_i, \mathbf{X}_{ijkt}, \mu_{ik})\right]^{y_{ijkt}} \cdot \prod_{l=1}^{d_{\kappa_t} - y_{ijkt}} \left[\exp(-\sum_k \varphi_{jk}(l, t_i, \mathbf{X}_{ijkt}, \mu_{ik}))\right]^{1-y_{ijkt}}$$
$$for\ \kappa_t = 0,1,2,\ i = 1,2,...,N$$

Here y_{ijkt} is the censoring indicator, which takes value 1 if transition from *j* to *k* is realised, 0 if the spell is censored.

The likelihood for a complete spell is thus

$$L_i = \sum_w p_w \cdot \prod_{\kappa_t} L_{i\kappa_t}{}^{z_{\kappa_t}}$$

and finally the total likelihood function for the whole sample can be easily acquired as

$$(5)\ \ L = \prod_{i=1}^{N} L_i = \prod_{i=1}^{N} \sum_w p_w \cdot \prod_{\kappa_t} L_{i\kappa_t}{}^{z_{\kappa_t}}$$

The randomness of treatment assignment incorporated in the hazard rate model also implies that any deterministic mechanism of assignment cannot be revealed to the intended treated *prior to* treatment. But in reality, it might happen that the individual has certain expectations regarding the probability of being treated and accordingly adjust his optimal strategy either to increase the probability of receiving treatment, or to avoid the treatment. For example, the Norwegian social security system in principal requires a quarantine period when the first benefit period (18-20 months) is exhausted. In order to maintain economic support after this period, participation in some labour market programmes might be required by authorities. In this case, the unemployed might intensify his search activity prior to the benefit cut-off time to avoid possible impending programmes participation. Equally possible, knowing that programme participation is highly probable, the unemployed might reduce his effort in job search accordingly to await forthcoming programmes. In either case, ignoring this anticipation effect of treatment would result in biased estimate on treatment effects, since treatment also affects the behaviour of non-participants. In practice, it is very difficult to find suitable proxies for such anticipating effect of treatment, but anticipation effects that are systematically related to spell duration will be captured by duration baseline hazard rates.

The totally non-parametric specification of our model is very ambitious to estimate. We have adopted an "implicit dummy" approach to effectively reduce the computational cost on multiplications of large amount of dummy variables; see Gaure and Røed (2003). We apply a maximum likelihood approach in estimation, starting by no unobserved heterogeneity and add one point of support to the vector of unobserved heterogeneity at each iteration, until the overall likelihood cannot be improved further. The maximization routine is hard-coded in Fortran 90 with MPI implementation for parallel processing[26].

As Zhang's (2003) Monte Carlo results suggest, the optimal number of points found for the unobserved heterogeneity distribution is sensitive with respect to maximising routine and search directions, and it is advisable to adopt some information criteria to penalise the excessive points found for the discrete mixing distribution when the sample size is small. However, the maximum penalised likelihood estimators converge to pure maximum likelihood estimators when the sample size is sufficiently large. We also adopt methods in Zhang (2003) on maximum penalised likelihood to check if our results are sensitive with respect to number of points found. In our analysis, we have at our disposal about 120,000 individuals, and our results appear to be robust with respect to the selection of maximum likelihood and maximum penalised likelihood stopping rules.

4. Results

Due to non-parametric specifications, there are totally over 1,000 model parameters. To outline our main findings, we organise the presentations of results as following: we first report the duration baselines for transitions from open unemployment to ordinary employment and labour market programmes by plotting estimated baseline hazard rates. We then look in to the deterministic factors of selections into each transition by examining the estimated coefficients for covariates. Secondly, we report estimated treatment effects on the transition probability to employment, both time-varying effects and heterogeneous treatment effects with respect to individual characteristics. Thirdly, we present stylised figures to further illustrate effects of labour market programmes over unemployment spells. We also offer a measure for the total effects of active labour market programmes through simulation.

[31] We are fortunate to have Senior Analyst Simen Gaure at the University Information Technology Centre at University of Oslo to help us programme the estimation routine. All estimations are done on HP Superdome at High Performance Computing Centre, University of Oslo.

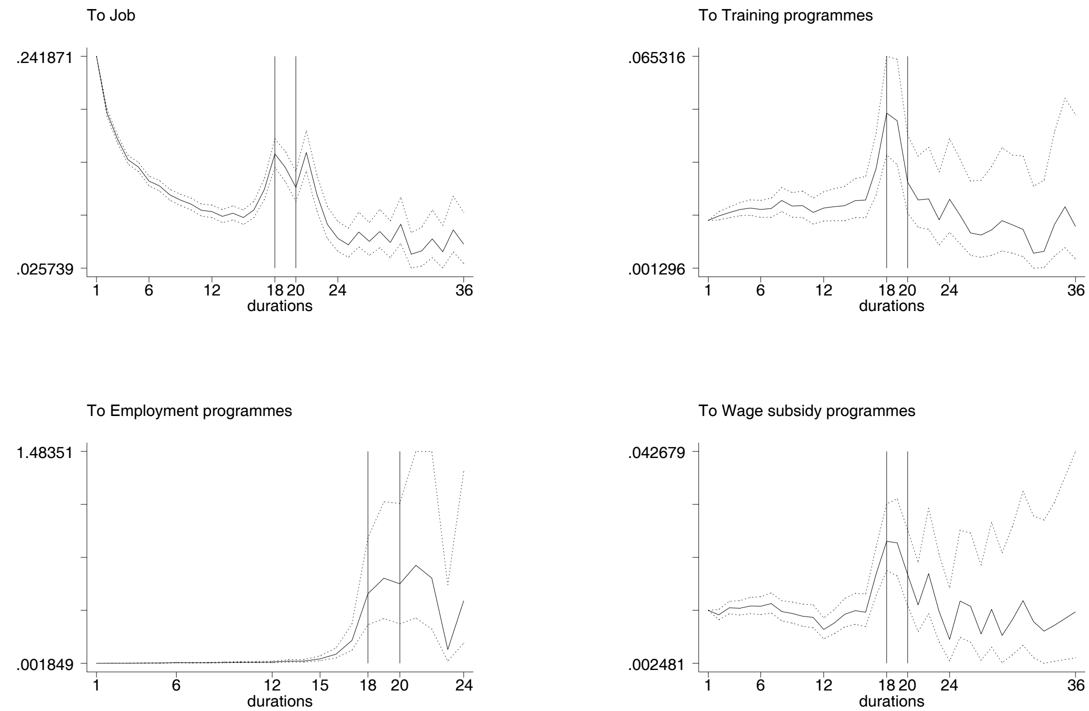
Selection into programmes

We plot estimates of the baseline hazard rates for transition to job and transitions to labour market programmes in Figure 2, together with 95% confidence intervals. The plotted curves are exponential of estimated coefficients and are normalised to the observed empirical hazard rates for the first month of spells. Note that they are estimates to the baseline hazard rates λ_{jkd} and do not have a direct interpretation of transition probabilities. For the baseline hazard rate to job, we find significant negative duration dependence even after controlling for unobserved heterogeneity. The hazard rate drops by half just after about 5 months. This suggests strong discouraged-worker effects and stigmatisation effect on transition to work. This agrees with several earlier studies on Norwegian labour market dynamics, see e.g. Røed and Zhang (2000), that even after control for unobserved heterogeneity, there is still strong negative duration dependence for the baseline hazard to job.

An interesting finding in the baseline for transition to job is that, at approximately 18 months of spell length, the hazard rate rises sharply from 0.11 up to 0.15 and remains this level until it drops down to 0.10 again after 21 months. The 18 months corresponds to about 80 weeks of first unemployment benefit entitlement period according to the Norwegian regulation, after which the benefit may be cut-off and there is a quarantine period before a possible renewal can take place. This seems to have a significant impact on the hazard rate out of the unemployment, as Røed and Zhang (2003) pointed out. Here the sharp rise of the hazard rate can to some extent be interpreted as an anticipation effect of impending labour market programmes. Given the knowledge of possible benefit sanction, the "threat" of participation in programmes seems to have considerably increased the hazard rate to employment.

The baseline hazard rates for transitions to training programmes and wage subsidy programmes seem to have no particular duration dependence at the beginning of spells. However, it is also interesting to observe the sharp rise of the hazard rates around 18 months for transitions to labour market programmes. Before the 18 months cut-off point, the baseline is almost flat. Around the benefit exhaustion time, the hazard rate estimates to both training programmes and wage subsidy programmes rise sharply by almost 100%

Figure 2: Duration baseline hazard rates with 95% confidence intervals.

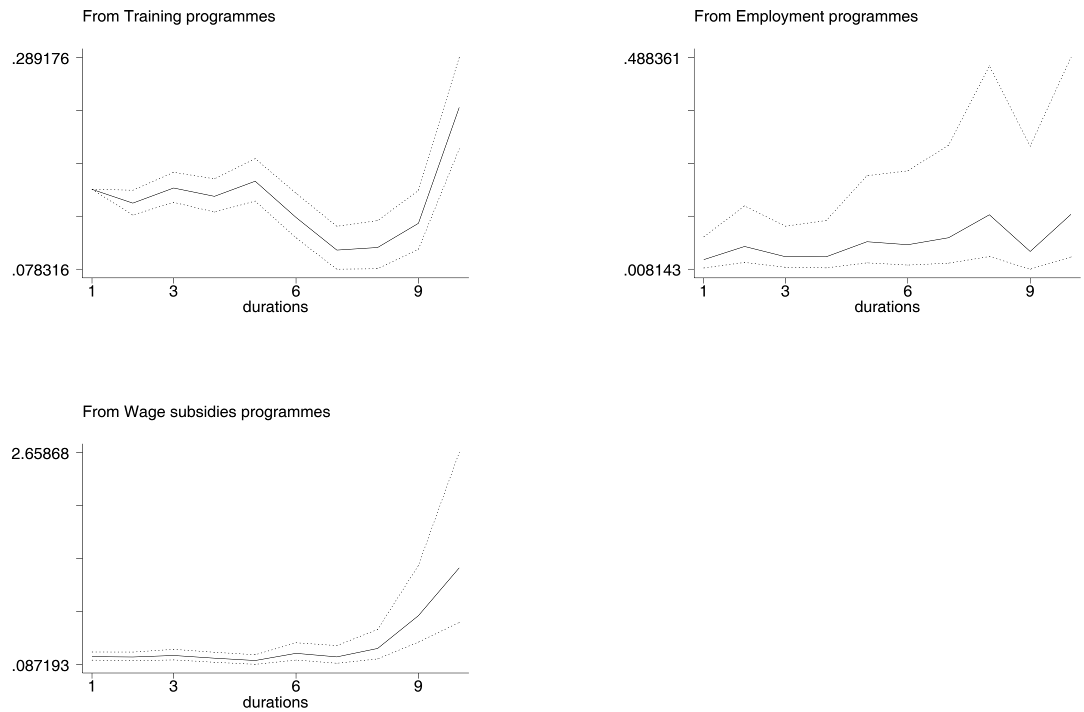


Note: Baseline hazard rates are normalised to the observed empirical hazard rates at the first month of spell. Duration for employment programmes is censored after 24 months due to lack of observations.

(from 0.05 to 0.08 for training programme, and from 0.002 to 0.005 for wage subsidy programme). This implies strong evidence of selection into programmes driven by the benefit exhaustion, which can be both self-selection (economic incentive of acquiring programmes allowances when benefits are exhausted), and administrative selection (priority of admission is given to those with exhausted benefits). The hazard rate approach here thus is able to take account of such self-selection into programmes via duration baselines.

The baseline hazard rate for transition to employment programmes is very low at the beginning of spell (around 0.0017) and quite flat until 15 months, where it rises sharply. This is possible evidence that employment programmes is designed for long-term unemployed and admission to programmes only occurs after certain length of duration. Due to lack of observations for longer spells for this transition, the baseline is censored after 24 months.

Figure 3: Duration baseline hazard rates for transitions back to open unemployment from participation in programmes, with 95% confidence intervals.



Note: all the baseline hazard rates are normalised to the observed empirical hazard rate of the first month participating in training programmes.

We also plot the baseline hazard rates for the transitions back to open unemployment while participating labour market programmes in Figure 3. We do not find significant evidence that the time spent in participation affects the hazard rate back to open unemployment. However, some positive duration dependence can be observed, especially for the wage subsidy programme. Explanation can be that the longer an individual stays in the programme participation, the higher the probability for back to open unemployment is, and it seems that the participation in programmes has a somewhat delayed effect which does not guarantee an immediate success.

Table 2 reports some of the important covariates estimations from the competing risks model. We restrict our attention to individual characteristics and their influences on hazard rates. The first column is the estimations of covariates for transitions from open unemployment to job. The second to fourth columns are estimations of transitions from open unemployment to labour market programmes. The last column is the estimations on the transition from programmes back to open unemployment (all three programmes combined). As to the determinants of transition to employment, female has a slightly better chance for job comparing to male unemployed (about 5.56%[27]); married adults seem to be more eager in finding employment, may be due to family responsibility; having younger dependents may reduce the probability of transitions to employment, possibly because taking care of youngsters reduces search intensity. Immigrants from non-OECD countries have difficulty in finding jobs comparing to natives, but once immigrants have acquired citizenship, the chance for an employment would increase by 7.32%. This is probably due to the fact that citizenship requires certain length of years staying in the country. Adaptation to language and culture, and basic knowledge of society would certainly contribute to success in the labour market. We find that younger people have better job prospects than elderly jobseekers. An individual's qualification measured in years of educational attainment plays an important role in employment opportunity. Compared to high school educated, with only primary school education would reduce the job probability by 15.1%. Also having relevant job training and experience helps to find employment quickly. All these findings are in concord with earlier studies on labour market dynamics, such as in Røed and Zhang (2000, 2003).

[32] The percentage changes are simply calculated by $\frac{e_1 - e_0}{e_0}$, where $e_1 = \exp(\hat{e})$, $e_0 = \exp(\hat{e}_0)$. $\hat{e}$ is the estimator, $\hat{e}_0$ is the reference.

Table 2: Estimated coefficients for competing risks hazard rate model.

	To job		To training programmes		To employment programmes		To wage subsidy programmes		Back to open unemployment	
	Est.	Std.	Est.	Std.	Est.	Std.	Est.	Std.	Est.	Std.
Gender (1=male)	-0.0581	0.0082	-0.0570	0.0254	-0.3683	0.0776	0.5038	0.0290	-0.0940	0.0322
Married (1=yes)	0.1965	0.0081	0.0250	0.0250	-0.0028	0.0730	0.1876	0.0288	-0.0898	0.0309
Having children under 18 years (1=yes)	-0.2050	0.0080	-0.0630	0.0248	-0.1318	0.0719	-0.0308	0.0285	0.0649	0.0304
Non-OECD immigrants (1=yes)	-0.5531	0.0255	0.1750	0.0643	-1.0815	0.2324	-1.1733	0.1126	0.0107	0.0901
Immigrants with Norwegian citizenship (1=yes)	0.0706	0.0333	-0.0647	0.0824	-0.4136	0.3280	0.4137	0.1386	0.0466	0.1143
Having relevant experience for intended job (1=yes)	0.1568	0.0144	-0.2753	0.0409	-0.2830	0.1236	0.0605	0.0506	-0.1077	0.0535
Having relevant training for intended job (1=yes)	0.0963	0.0095	0.1635	0.0291	-0.1723	0.0792	0.0772	0.0329	0.0293	0.0366
Age (ref: 36-40)										
25-30	0.1126	0.0109	-0.1006	0.0335	0.0907	0.0942	-0.0343	0.0388	-0.1127	0.0436
31-35	0.0199	0.0112	-0.0560	0.0342	0.0172	0.1005	-0.0852	0.0397	-0.0058	0.0443
41-45	-0.0493	0.0124	0.0303	0.0371	-0.0593	0.1091	-0.0561	0.0430	0.0289	0.0470
46-50	-0.1808	0.0139	-0.0886	0.0405	-0.2626	0.1214	-0.1512	0.0461	0.1804	0.0499
Educational Attainment (ref: 11-12 years)										
up to 9 years	-0.1637	0.0166	0.4057	0.0610	0.6613	0.2059	0.0969	0.0635	0.2795	0.0881
10 years	-0.0293	0.0173	0.3030	0.0635	0.4181	0.2125	0.2219	0.0659	0.0881	0.0917
13-16 years	0.2249	0.2480	*	*	*	*	-0.2779	0.9478	*	*
17 or more years	0.1577	0.2491	0.3014	0.9539	*	*	-0.6862	0.9896	*	*
County of residence (ref: Oslo)										
Akershus, Hedmark, Oppland, Buskerud	0.1522	0.0112	0.2742	0.0341	0.6406	0.1080	0.3755	0.0400	0.1370	0.0427
Vestfold, Telemark, Aust-Agder, Vest-Agder	0.1767	0.0130	0.2305	0.0395	0.8152	0.1260	0.4783	0.0447	0.0053	0.0513
Rogaland, Hordaland	0.1184	0.0120	0.3106	0.0368	0.3194	0.1207	0.0119	0.0453	0.0309	0.0466
Sogn og Fjoldane, Møre og Romsdal, Sør-Trondlag, Nord-Trondlag	0.2855	0.0126	0.1019	0.0419	1.6872	0.1182	0.2142	0.0473	0.1056	0.0535
Nordland, Troms, Finnmark.	0.3428	0.0142	0.1763	0.0477	2.0421	0.1262	0.5260	0.0492	-0.2057	0.0631
Occupational background (ref: Unspecified)										
Technical, physical science, humanistic and artistic	-0.0741	0.0400	-0.2500	0.1278	-0.4106	0.3526	0.2377	0.1706	-0.3279	0.1877
Administrative executive work, clerical work and sales	-0.2941	0.0400	-0.0936	0.1272	-0.9358	0.3523	0.3388	0.1698	-0.1664	0.1866
Agriculture, forestry, fishing and related work	0.0177	0.0477	-0.3798	0.1564	0.1892	0.3926	0.1040	0.1954	-0.1347	0.2233
Manufacturing, mining, quarrying, building, construction	-0.1360	0.0401	-0.1105	0.1272	-0.4538	0.3508	0.1082	0.1703	-0.0443	0.1868
Service work, transport and communication	-0.1320	0.0402	-0.3080	0.1282	-0.8773	0.3544	0.0700	0.1712	-0.0422	0.1875

Note: * indicates that this variable is omitted in estimation due to the fact that there is no observation for this variable in this particular transition.

As to the determinants of transition to labour market programmes, we find that participation in particular type of programmes is highly selective with respect to individual characteristics. For employment programmes, people with low qualification and low job market competitiveness have a large probability to participate. Compared with high school graduates, those with only primary educational attainment have 93.7% larger chance for participation in employment programmes. Females seem to have a large tendency to participate. Also there is strong regional variation in participating in the employment programme. The northern counties are over represented (in terms of probability of participation).

Participations in the training programmes and the wage subsidy programmes do not seem to have the same strong pattern of selection as observed for employment programmes. Men seem to have better chance to participate in the wage subsidy programmes, as well as married adults. Immigrants that have acquired Norwegian citizenship seem to have a larger chance to participate than non-OECD immigrants without citizenship. For training programme, those with lower education than senior high school (11-12 years) have strong tendency to participate. There are strong regional variations in terms of participation probabilities as well.

Noticeably from Table 2, estimated coefficients for the covariates associated with transition back to unemployment after participating in labour market programmes seem in general to have negative signs comparing to estimates from the job transition. This implies that those return to open unemployment after participation on average have lower job prospects. It is plausible that a sorting mechanism "selects" unemployed individuals out of the unemployment pool according to qualification and employability. Those with highest qualification get job first and leave unemployment quickest; those who need assistance in job search leave after participation in labour market programmes; those with lower qualification would eventually return to unemployment even after participation. Hence, it is evident that evaluation of labour market programmes on these different groups must take into account that those treated are in general a selected group. This aspect can be uncovered by duration model framework as demonstrated here, but usual static evaluation methods do not have the mechanism to explore this.

Our maximisation routine returns 5 points of support for the unobserved heterogeneity distribution for all transitions. Table 3 provides the calculated first and second order moments (in exponential form) of the unknown mixing distributions for the unobserved heterogeneity. We are hesitant to give an interpretation of estimated mass points, such as “ability” or “motivation”, as some authors suggest. Zhang (2003) has showed, it is often not possible to retrieve exact number of points for unobserved heterogeneity distribution, even when the true distribution is discrete with known number of support points. Rather, we suggest that emphasis should lie on the correct control for the unobserved heterogeneity such that the other model parameters of interest can be consistently estimated, see Zhang (2003) for a detailed discussion of this aspect. We find that estimators for model’s structure parameters do not have any significant numerical differences after finding of 4 points, neither do the estimated moments for mixing distribution.

Table 3 also reports the estimated correlations coefficients for the unobserved heterogeneities across diverse transitions. The correlations coefficients could have an interpretation of selections on the unobservables between different transitions (according to Abbring and Van den Berg(2003b)). Loosely put, a positive correaltion between two transitions means those with higher probability for transition to one state would also have a somewhat higher probability taking another transition; and vice versa. We find that there is a slight positive selection between job and training transition, also a somewhat positive selection has been found between job and wage subsidy programme. It seems also that different programmes are substitutes as all correlations coefficients between programmes are negative. Not surprising is the strong negative selection between job and back to open unemployment when the individual is participating in the programmes. This implies a negative selection between job transition and transition back to unemployment. Those with preferable employment prospects would leave unemployment earlier, possibly with the assistance of programme participation. Since we do not have the uncertainty measures for these estimators of correlations coefficients (they are not directly estimated but calculated from the estimated mass points distribution), we hesitate to draw any firm conclusion with respect to this and apply great caution on the interpretation.

Table 3: Estimations on the moments of the unobserved heterogeneity distributions and correlations coefficients for the unobserved heterogeneity between transitions.

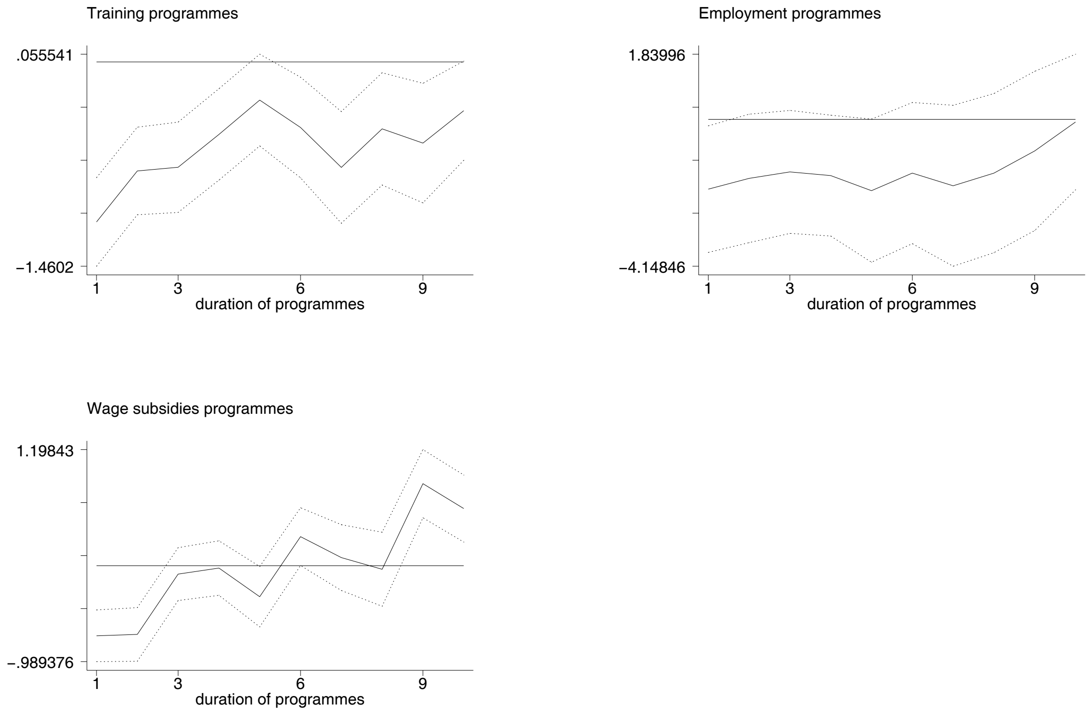
	Expectation	Variance
Transition to job	0.2505	4.72E-03
Transition to training programmes	0.0270	9.56E-04
Transition to employment programmes	0.0010	4.08E-06
Transition to wage subsidy programmes	0.0021	4.16E-06
Back to open unemployment	0.1293	8.20E-03
Correlations coefficients of unobserved heterogeneities between transitions		
before transitions to programmes		
job and training programmes		0.0648
job and employment programmes		0.0455
job and wage subsidy programmes		0.1136
training programmes and employment programmes		-0.2370
training programmes and wage subsidy programmes		-0.5925
employment programmes and wage subsidy programmes		-0.3532
after transitions to programmes		
job and back to open unemployment		-0.8019

Note: 1. maximisation returns 5 mass points of support for unobserved heterogeneity. 2. expectations and variances are calculated with exponential transformations. 3. correlations coefficients for unobserved heterogeneities are calculated based on the estimates for the mass points distributions.

Treatment effects

To assess the dynamics of treatment effects over time, we estimate the while-treatment effects and after-treatment effects of participation in each of labour market programmes with two step-functions ($\mathbf{D}_{0kt}$ and $\mathbf{D}_{1kt}$ in equation (3)). To facilitate the interpretations of the positive and negative sides of the effects, we plot directly the estimated coefficients for while-treatment effects and after-treatment effects over time in Figure 4 and 5, together with 95% confidence intervals. These coefficients of effects are estimated relative to that of a female middle-aged non-participant, under average labour market conditions and all other covariates taking mean values, which is indicated by the solid horizontal line with value zero in each figure. The formal comparisons of these effects on transition probabilities should be calculated by inputting these coefficients into the competing risks hazard rate formulations as showed in (3). Here we for the expository purpose report the effects of these estimates on the integrated monthly hazard rates $\varphi_{jk}(d,t,\mathbf{X}_{jkt},\mu_k)$.

Figure 4: While-treatment effects for transitions to employment.



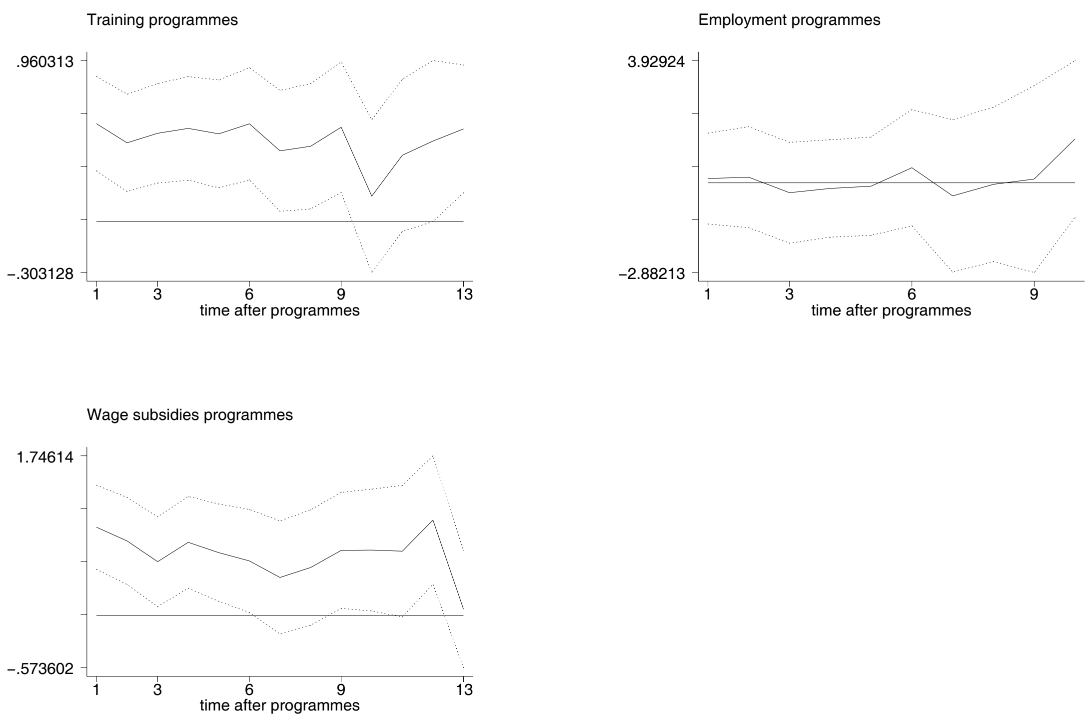
Note: 1. effects are measured relative to a middle-aged female non-participant with medium education, under average labour market conditions. 2. horizontal lines indicate zero effects (reference).

It is remarkable that while the treatment is undergoing, i.e. while participating in labour market programmes, we find a strong negative effect for the training programme relative to the reference, on average about 45.6%[28] reduction on the hazard rate to employment while participating in the training programmes. The negative effect is even stronger for employment programme (76.8%), while the wage subsidies programmes seem not have a significant impact on hazard rate to job on average (0.7%). We also observe that although the while-treatment effects are mostly negative, they increase over time. This holds for all types of labour market programmes. The increase is strongest for the wage subsidy programmes, just after 6 months of participation in the programmes, the effect on the hazard rate to employment is already positive, and remains growing. Equally increasing effect can

[34] Average effects are calculated based on estimators of time-varying treatment effects, relative to a middle-aged female non-participant with medium education, under average labour market conditions. Full set of estimators is available upon request.

be found with the training programmes as well. Even though the effect is negative during the entire training period, the estimated coefficients have increased sharply from -1.144 to -0.518 within 4 months. Somewhat increasing effect for the employment programmes can be observed as well.

Figure 5: After-treatment effects for transitions to employment.



Note: 1. effects are measured relative to a middle-aged female non-participant with medium education, under average labour market conditions. 2. horizontal lines indicate zero effects (reference).

The after-treatment effects are more encouraging. Positive effects are observed both for the training programmes and the wage subsidy programmes. Those that have been to the training programmes have achieved a 59.6% average increase of the hazard rate to employment. For wage subsidy group, the effect is even higher at 86.9%[29]. All these effects are significant (viewed from confidence intervals in Figure 5). The positive effects are not temporary, but lasting for a long post treatment period. However, the after-treatment effects

[35] See footnote 34.

do decline gradually with time. The effects are strongest immediately after completion of programme, and decline gradually as individual still remain unemployed. For participants in employment programmes, the average after-treatment effect is 19.2%, but not significantly different from zero (Figure 5).

A possible explanation for the negative effects of the training programmes while the programmes are undergoing, is that participation in such programmes possibly reduces search intensity. It might be the case that participants wish to take advantage of the training opportunity to enhance their qualifications and human capitals. Such enhancement needs certain amount of time to accumulate. Once the programmes are completed, the job probability is increased significantly and the ex post effects of the training programmes are significantly positive. This is in accordance with the findings of Raaum et al. (2002).

The negative while-treatment effects of the employment programmes can be thought of as lock-in effects. Since the employment programmes are targeted at long-term unemployed to prevent them from dropping out of the labour force, they do not have the immediate goal to systematically improve the qualifications of low-skilled jobseekers. Instead, they have a kind of safety net feature by providing temporary employment with pay. Therefore the participants might as well lack of motivation in search for normal employment.

Contrary to the employment programmes participants, the wage subsidy programmes participants are generally more qualified and ready for employment. The programmes have already positive effects on the transition probability even when the programmes are undergoing. Immediately after the completion of the wage subsidy programmes, a significant increase on the hazard rate to ordinary employment can be observed.

Table 4 reports the estimated heterogeneous while-treatment and after-treatment effects with respect to selected individual characteristics as well as with business cycle conditions. For the while-treatment effects, our first observation is that there is not much difference across individual characteristics. However some of the after-treatment effects vary across individual characteristics. For the training programmes, women seem to benefit more from participation with a 15.2% higher effect than men. Similar findings have also been observed in Raaum et al. (2002). It holds for the wage subsidy programmes as well, where females have an even higher advantage to males with an increase of hazard rate to job as much as 21.2%. For the employment programmes, men seem to have a stronger effect than women, but this difference is not significant. Younger jobseekers benefit strongest from employment programme. Perhaps the most significant observation is that low education seems to have a negative impact on the effects of the employment programmes. As for the wage subsidy programmes, the impact of participation is stronger for women than for men. For the while-treatment effects, the training programmes display a significant pro-cyclical pattern. This implies that the effect of participating in a training programme is larger if the labour market condition is favourable. A similar pattern is observed for the after-treatment effects for the training programmes with some significance. In Raaum et al. (2002), they also find the pro-cyclical patterns of labour market training programmes. The intuition behind this finding can be thought of as follows: when the job market is unfavourable, job vacancies are scarce. Therefore it might be of little importance whether one has participated in the labour market programmes or not, since there are not many jobs to fill in anyway. When the labour market condition is good, those who have participated in labour market programmes might signal more positive qualifications than those who have not. Thus, the participation in the programmes has stronger impact on the employment probability when at the boom time.

Stylised analysis

To illustrate the dynamic effects of participation in labour market programmes on the transition probabilities to ordinary employment, we conduct a highly stylised analysis that resembles the matching study. The idea is that by keeping all other covariates that affect hazard rate fixed, we are able to isolate the causal effects of participation in labour market programmes by comparing the predicted hazard rates with and without the presence of programme participation.

We construct a representative unemployed jobseeker with all individual characteristics taking mean values of the estimation sample. We also fix calendar months and business cycle indicators to sample references. By using estimators for job transition, we predict hazard rates over a 36-months period using equation (3) for non-participants.

Table 4: Heterogeneous treatment effects.

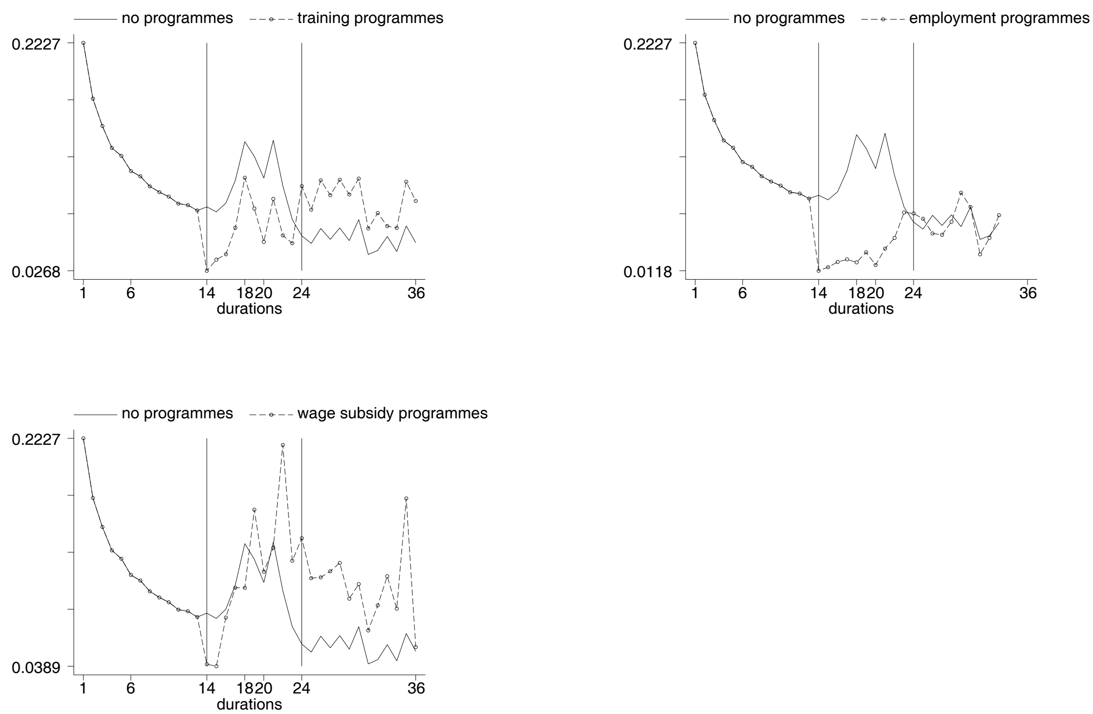
While-treatment effects	Est.	Std.
Training programmes		
× business cycle indicator	0.6693	0.1572
× low education (up to 9 years)	-0.0515	0.0672
× high education (more than 12 years)	1.3085	0.8386
× male	0.0321	0.0535
× younger jobseeker (age ≤ 30)	0.0288	0.0577
× elder jobseeker (age > 45)	0.0687	0.0805
Public employment programmes		
× business cycle indicator	-0.2382	0.9980
× low education (up to 9 years)	-0.0951	0.3479
× male	0.3122	0.2309
× younger jobseeker (age ≤ 30)	0.1499	0.2420
× elder jobseeker (age > 45)	-0.0950	0.4064
Wage subsidy programmes		
× business cycle indicator	-0.1466	0.1515
× low education (up to 9 years)	0.2870	0.0591
× male	-0.0318	0.0508
× younger jobseeker (age ≤ 30)	0.0097	0.0578
× elder jobseeker (age > 45)	-0.0403	0.0722

After-treatment effects	Est.	Std.
Training programmes		
× business cycle indicator	0.2760	0.1286
× low education (up to 9 years)	0.1056	0.0644
× male	-0.1621	0.0455
× younger jobseeker (age ≤ 30)	0.0704	0.0532
× elder jobseeker (age > 45)	-0.1260	0.0638
Public employment programmes		
× business cycle indicator	-1.9588	1.7547
× low education (up to 9 years)	-1.2637	0.6912
× male	0.4156	0.4134
× younger jobseeker (age ≤ 30)	0.9581	0.4551
× elder jobseeker (age > 45)	0.6820	0.5027
Wage subsidy programmes		
× business cycle indicator	0.0688	0.2559
× low education (up to 9 years)	-0.1627	0.1037
× male	-0.2309	0.0859
× younger jobseeker (age ≤ 30)	-0.0328	0.1081
× elder jobseeker (age > 45)	-0.0957	0.1108

Note: the reference is middle-aged female (31-45 years) with above 9 years educational attainment under the average labour market conditions.

Assume that at the start of 14th month[30], the "artificial jobseeker" participates in a labour market programme that takes 10 months to finish. We add while-treatment effect estimators to the hazard rate formulation and predict the "while-treatment hazard rate". After completing of programme, we follow the spell further until 13 months and calculate "after-treatment hazard rate" by including after-treatment effect estimators into hazard rate formulation. We predict such representative hazard rates for all three groups of labour market programmes that we evaluate.

Figure 6: Predicted treatment effects on transition probabilities to employment.



Note: vertical lines indicate the start and the end of programmes.

Figure 6 depict the stylised figures on how participation in labour market programmes affects the hazard rate to employment. We observe that immediately after starting a

[30] Because the total length of spell in estimation sample is 36 month, and we have 10 estimators for the programme duration and 13 estimators for the post-programme duration, therefore the pre-programme spell duration is set to 13 month.

programme, the hazard rate drops significantly. While participating in a training programme, the hazard rate is lower than that of non-participation, but gradually catches up over the duration of participation. After the completion of the training programme, the hazard rate for after-treatment period rises sharply above that of non-participation, though it again decreases gradually as the spell lengthens. For the wage subsidy programme, the effect of increasing the hazard rate comes much earlier. After only 3 months of participation, the hazard rate due to participation is already higher than that of non-participation. The hazard rate remains higher as well and lies above that of non-participation after the participation is finished. For the employment programme, we observe the decrease of the hazard rate during the participation, but the after-treatment hazard rate is almost the same as that of non-participation.

The above figures give some visual illustration of the impacts of active labour market programmes on the hazard rates to job. Since the treatment effects are mostly negative during the participation, and positive after the participation, it is desirable to derive a measure for the total impact of the active labour market programmes on the spell length. However, the prediction of expected spell duration with programme participations cannot be solved analytically, since we do not have the knowledge of future development of labour market conditions, as well as the covariate processes that have influences on the hazard rates. We provide here an approach based on simulation to offer an assessable measure of the total impact of the treatment effects.

The idea here is to first simulate a counterfactual situation that no programmes have any effects on the hazard rate to job[31]. Based on the estimation sample, we predict the expected spell durations in our competing risks model. We take one individual and record his/her observed characteristics at the first month of the unemployment spell, as well as at the calendar time at which the spell starts. Then by utilising the complete estimates for the transition to job and labour market programmes (coefficients of covariates, baseline hazard

[31] There are however the possible anticipation effects of programmes remaining. We have found that the existence of programmes could possibly affect the behaviour of individuals even for non-participants. Because we do not have any estimators for such anticipation effects, we cannot predict the spell durations excluding such anticipations. Since the anticipation effects are (in part) captured by the baseline hazard rates, the predicted spell durations based on those baseline hazard rate estimates are compatible to those in the real data.

rates, estimates for the calendar time effects and the averages for the unobserved heterogeneity), we predict the progression of each spell. For the sake of simplicity, we fix the individual characteristics throughout the spell. The previous censoring scheme is applied here as well such that the spell is censored after 36 months, or if the spell has exceeded the observation window (from Jan. 1990 to Dec. 2000). The dynamic processes depicted in Figure 1 are followed in the simulation. Repeating this process for all spells, we get a sample for the unemployment within the counterfactual state of no programmes effects. The total amount of unemployment months is then measured.

We next consider the situation where only one of the programmes has effects corresponding to our point-estimates, while the others have zero effects[32]. Interaction terms of treatment effects with individual characteristics and business cycle indicators are also added to the hazard rates. After simulation of the spells for this single programme effects situation, the total amount of unemployment months, compared with that from no programme effects, gives us a measure of marginal impact from one particular programme. We conduct this simulation separately for all three active labour market programmes.

Last, in the similar manner we predict the complete competing risks model, including all three programmes' effects evaluated earlier. Again, individual characteristics are fixed. By incorporating the time-varying while-treatment effects and after-treatment effects to the hazard rates, we predict a sample of unemployment spells when there are three types of active labour market programmes that have effects on the hazard rates. This simulation provides a sample that bears the satisfactory similarity to that of the estimation sample in terms of distribution of spell lengths. The total amount of unemployment months are then measured and used to compare with that from the counterfactual situation of no programme effects to assess the total impact of active labour market programmes in terms of the changes in the total amount of unemployment.

[38] In the simulation, we censor the programme duration after 10 months, and post programme duration after 13 months, respectively, to resemble the same censoring practice in the estimation earlier.

We conduct the above simulation routines 100 times to get the average total impact of active labour market programmes with uncertainty measures. Table 5 reports the results from this highly stylised exercise.

Table 5: Total impacts of the active labour market programmes.

	Total amount of unemployment months of all spells	
	mean	**std**
No programme effects	819533.00	2206.44
With effects from the training programmes	791289.32	2422.91
Changes due to the training programmes	**-28243.68**	**3277.02**
Causal effects of the training programmes	-3.45 %	
With effects from the employment programmes	824070.26	2404.52
Changes due to the employment programmes	4537.26	3263.45
Causal effects of the employment programmes	0.55 %	
With effects from the wage subsidy programmes	790612.94	2352.19
Changes due to the wage subsidy programmes	**-28920.06**	**3225.08**
Causal effects of the wage subsidy programmes	-3.53 %	
With effects from all three programmes	766933.90	2208.92
Changes due to all three programmes	**-52599.10**	**3122.13**
Causal effects of all three programmes	-6.42 %	

Note: 1. bald-faced fonts indicate significant estimators. 2. mean and standard errors are calculated across 100 simulation trials.

The means and standard errors are calculated across 100 simulations. The impacts of programmes are measured as reduced total amount of unemployment months, and the percentage changes could have the interpretation as the causal effects of the programmes. We see that both training programmes and wage subsidy programmes have positive effects in terms of reduced total unemployment. The causal effect of the training programmes alone is about 3.45%, while for the wage subsidy programmes is about 3.53%. The employment programmes do not seem to have significant effect on reducing the total unemployment.

When viewing all three programmes together, the total impact of active labour market programmes is about 6.42% reduction of total unemployment and the effect is significant[33].

5. Conclusions

By estimating treatment effects of Norwegian labour market programmes on transition probabilities to employment, we evaluate causal effects of participation in the active labour market programmes for Norwegian prime-aged unemployed workers. The estimation is carried out by applying non-parametric competing risks hazard rate model.

We find significant impacts of participations in active labour market programmes on the transition probabilities to ordinary employment. Both training programmes and wage subsidy programmes have significant positive effects on employment probabilities *after* the completion of programmes. There is some evidence that these two groups of active labour market programmes have their intended effects on enhancing job opportunities and function as effective tools in combating the unemployment. However, *during* the participation of programmes, the transition probabilities are low comparing to non-participants. This can be due to the nature of programmes participation (reduced search intensity during participation). The employment programmes on the other hand do not display strong causal effects on the transition probabilities after the programmes have finished. During the programmes period, the transition probabilities are significantly lower than that of non-participants. There is limited evidence on the heterogeneous treatment effects with respect to the individual characteristics. Women seem to benefit more after participating in the training programmes and the wage subsidy programmes. The younger jobseekers benefit more from the employment programmes.

[33] Recall that in the estimation data, we also censor the spell once a transition from one programme to another programme has occurred. Also if there are repeated participations in the same or different programmes, the spell is censored as well. Ideally, we should also include such options as possible transitions and censor the spells accordingly in the simulation. But since we do not have the estimates for cross-programme transitions and repeated participations, in our simulation, such cross-programmes or repeated participations are not modelled. Although in estimation such censoring does not impose bias on the estimators, this innocuous practice in simulation might have the consequences on the predicted spell durations. Thus the total effects of programmes in terms of reduced amount of unemployment might be overestimated.

There is some evidence of selection into different programmes with respect to individual characteristics. This may be due to that the different programmes are targeted on the different population of participants. The employment programmes are targeted on long-term unemployed to prevent them from dropping out of labour force, while the wage subsidy programmes offer qualified jobseekers a final assistance in finding employment. The evaluation of effects across different programmes must take account for the differences of the intended treated.

The effects of labour market programmes change over time and business cycle conditions. Effects of both training programmes and wage subsidy programmes have a pro-cyclical pattern, which means the effects are stronger the better the labour market conditions are. Also we find that the treatment effects change over time spent during participation and time spent after participation. During the programmes participation, the effects of programmes grow with elapsed the programme duration. There is evidence that treatment effects need time to build up. The after-treatment effects are significantly positive both for training programmes and wage subsidy programmes. The effects are strongest when participants have just finished the programmes, and persistent over the spell length for participants remaining unemployed.

The total impacts of all three active labour market programmes are measured in terms of reduced total unemployment volumes by simulations. We find a significant effect of 6.42% reduction of the total amount of unemployment months due to the active labour market programmes. However, we interpret these results with caution, because the simulation method used here might not be suitable (see footnote 39). The case of evaluation of treatment effects due to cross-programme transitions and multiple participations is remaining for future research.

By studying various types of programmes over time within the duration model framework, we hope to provide some insights on the causal effects of Norwegian active labour market programmes and the dynamics of these effects. Nevertheless, the social gains of the active labour market programmes must be evaluated in the conjunction with the costs of programmes, both in terms of individuals' opportunity cost during the participation, and the administrative cost of providing these programmes. A cost-benefit analysis might be a nature continuation of this study. The policy implications of this study should be focused on the dynamic side of programmes effects. Given the evidence of heterogeneity of treatment effects both over intended treated and over unemployment duration and business cycle, it is of importance for policy makers to design active labour market programmes tailored to the different needs across the different unemployment population, and to adjust the scope and volume accordingly at different stages of business cycles.

References

Aakvik, Arild, James J. Heckman and Edward J. Vytlacil (2000), "Treatment effects for discrete outcomes when responses to treatment vary among observationally identical persons: an application to Norwegian vocational rehabilitation programs." *NBER Technical Working Paper No 262.* Cambridge, MA.

Abbring, Jaap H. and Gerard J. van den Berg (2003a) "The Identifiability of the Mixed Proportional Hazards Competing Risks Model" Forthcoming in *J. Royal Statistical Society: Series B*.

Abbring, Jaap H. and Gerard J. van den Berg (2003b) "The Non-Parametric Identification of Treatment Effects in Duration Models" *Forthcoming in Econometrica.*

Ackum Agell, Susanne and Martin Lundin (2001), "Erfarenheter av svensk arbetsmarknadspolitikk." *Ekonomisk debatt,* 29, pp 239-250.

Baker, Michael and Angelo Melino (2000), "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study" *Journal of Econometrics* 96, 357-393.

Brinch, Christian (2000) "Identification of Structural Duration Dependence and Unobserved Heterogeneity with Time-varying Covariates" Memorandum No. 20, 2000, Department of Economics, University of Oslo.

Calmfors, Lars, Anders Forslund and Maria Hemström (2001), *Does active labour market policy work? Lessons from the Swedish experiences*. Working paper. Uppsala: IFAU – office of labour market policy evaluation.

Fay, Robert G. (1996): "Enhancing the Effectiveness of Active Labour Market Policies: Evidence from Programme Evaluations in OECD countries", *Labour Market and Social Policy Occasional Papers*, #18, OECD.

Fertig, Michael, Christoph M. Schmidt, and Hilmar Schneider (2002) "Active Labor Market Policy in Germany - Is There a Successful Policy Strategy?" *IZA Discussion Papers* 576, Institute for the Study of Labor (IZA).

Gaure, Simen and Knut Røed (2003), “ How Tight is the Labour Market? A Micro-Based Macro Indicator” *Memoranda,* Department of Economics, University of Oslo, Norway 09/2003.

Gerfin, Michael and Michael Lechner, (2002), “Microeconometric Evaluation of the Active Labour Market Policy in Switzerland”, *The Economic Journal*, 112 (482), 854-893.

Heckman, James. J. (1979), “Sample Selection Bias as a Specification Error (with an Application to the Estimation of Labor Supply Functions)”, *Econometrica* 47(1) 153-161.

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

Heckman, James and Burton Singer (1985), “Social science duration analysis” in Heckman and Singer (ed.) *Longitudinal analysis of labor market data*, Cambridge University Press, 39-110.

Heckman, James and Bo Honoré (1989) “The Identifiability of the Competing Risks Model” *Biometrika* Vol. 76, No. 2. pp. 325-330.

Heckman, James J., Hidehiko Ichimura and Petra Todd (1997), “Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme.” *Review of Economic Studies,* 64, pp. 605-654.

Heckman, James J., Robert J. Lalonde and Jeffrey Smith (1999), “The economics and econometrics of active labor market programs.” In O. Ashenfelter and D. Card (Eds.) *Handbook of Labor Economics*, Vol. 3a, North-Holland.

Holland, P.W. (1986), “Statistics and Causal Inference”, *Journal of the American Statistical Association*, vol 81, issue 396, 945-960.

Honoré, Bo (1993), “Identification Results for Duration Models with Multiple Spells” *The Review of Economic Studies*, Vol. 60, No. 1. (Jan., 1993), pp. 241-246.

Lalive, Rafael, Jan C. van Ours and Josef Zweimueller (2001) ”The Impact of Active Labor Market Programs on the Duration of Unemployment, *IEW Working Paper No. 51*, University of Zurich.

LaLonde, Robert J. (1986) “Evaluating the Econometric Evaluations of Training Programs with Experimental Data”, *The American Economic Review*, Vol. 76, No. 4., pp. 604-620.

McCall, Brian P. (1994), “Identifying State Dependence in Durations Models,” *American Statistical Association 1994, Proceedings of the Business and Economics Section*, 14-17.

McCall, Brian P. (1996), "The Identifiability of the Mixed Proportional Hazards Model with Time-Varying Coefficients," *Econometric Theory*, 12, 733-738.

McCall, Brian P. (1997), “The Non-Parametric Identifiability of a Competing Risks Model with Regressors” mimeo, University of Minnesota.

Raaum, Oddbjørn, Hege Torp and Tao Zhang (2002), “Business cycles and the impact of labour market programmes”, Memoranda, 14/2002, University of Oslo.

Richardson, Katarina and Gerard J. van den Berg (2002) “The effect of vocational employment training on the individual transition rate from unemployment to work” *IFAU working paper*, 2002:8.

Rubin, D.B. (1974), “Estimating Causal Effects of Treatments in Randomised and Nonrandomised Studies,” *Journal of Educational Psychology*, 66, 688-701.

Røed, Knut and Tao Zhang (2000), “Labour Market Transitions and Economic Incentives.” *Memoranda,* Department of Economics, University of Oslo, Norway 15/2000.

Røed, Knut and Oddbjørn Raaum (2003) “The Effect of Program Participation on the Transition Rate from Unemployment to Employment.” *Memoranda,* Department of Economics, University of Oslo, Norway 13/2003.

Røed, Knut and Tao Zhang (2003), “Does Unemployment Compensation Affect Unemployment Duration?” *Economic Journal*, Vol. 113, issue 484, 190-206.

Torp, Hege (1995), “Labour market Institutions and Development. Norway Baseline Study.” Oslo: Institute for Social Research, Report 1995:11.

van den Berg, Gerard J. (2001), Duration Models: Specification, Identification, and Multiple Durations. *Handbook of Econometrics*, Vol. 5, Chapter 55, Elsevier.

Zhang, Tao (2003) “A Monte Carlo Study on Non-parametric Estimation of Duration Model with Unobserved Heterogeneity” Working paper.

Chapter 5: A Monte Carlo study on non-parametric estimation of duration models with unobserved heterogeneity

By Tao Zhang

A Monte Carlo study on non-parametric estimation of duration models with unobserved heterogeneity

By Tao Zhang
The Ragnar Frisch Centre for Economic Research

Abstract

We conduct extensive Monte Carlo experiments on non-parametric estimations of duration models with unknown duration dependence and unknown mixing distribution for unobserved heterogeneity. We propose a full non-parametric maximum likelihood approach, based on time-varying lagged explanatory covariates from observational data. By utilising this data-based identification source, we find that both duration dependence and unobserved heterogeneity can be reliably estimated. Our Monte Carlo evidences show that variation in time-varying lagged explanatory variables contributes to the identification of both duration dependence and unobserved heterogeneity, especially when sample sizes are limited. For limited sample sizes, maximum penalised likelihood with information criteria seems to produce more accurate estimators than pure maximum likelihood. Our approach can be easily extended to multivariate competing risks model with dependent unobserved heterogeneities.

Keywords: duration dependence, unobserved heterogeneity, non-parametric estimation, Monte Carlo study, time-varying covariates.
JEL Classification: C14, C15, C41

*Thanks to Knut Røed and Harald Goldstein for helpful comments and suggestions for improvement.

1. Introduction

The hazard rate model has seen many applications in applied econometric analysis, especially in unemployment duration studies. The aim of unemployment duration analysis is typically to study how the variables of economic interests, such as economic incentives, affect the transition probabilities to employment. Often, the uncontrolled population heterogeneity casts bias on estimation of causal parameters of interest, e.g. Lancaster (1979) showed that uncontrolled unobserved heterogeneity biases estimators of structure parameters towards zero. Heckman and Singer (1985, p.53) also prove that uncontrolled heterogeneity biases estimated hazard rates towards negative duration dependence. One of the most important challenges in unemployment duration analysis is hence whether the distribution of unobserved heterogeneity can be identified and estimated consistently from observational data, so that the bias arising from uncontrolled heterogeneity on parameters of economic interests can be eliminated. Conventional mainstream analysis adopts parametric specifications for both duration dependence and unobserved heterogeneity. However if economic theories do not provide explicit guidance, there is a risk of misspecification with parametric modelling. Several authors have suggested that flexible specifications on duration dependence and/or unobserved heterogeneity are superior compared to parametric specifications. See van den Berg (2001) for a recent survey. Some semi-parametric approaches have also been suggested, e.g. Horowitz (1999).

In this paper, we are exploring the identification and estimation feasibility of non-parametric maximum likelihood estimation (NPMLE) of duration models, particularly when data exhibits some discreteness. Two distinguishing features are represented in our analysis: by utilising newly available high performance computing techniques, we are able to overcome the computational barrier encountered in the earlier studies and estimate the non-parametrically specified hazard models in large scale and variety. This provides unique opportunity to assess the properties of non-parametric maximum likelihood estimators. We also utilise a unique feature of observational data that has become available with the access to administrative registers for research purpose, namely the time-varying explanatory variables embedded in the exogenous calendar time variation. We will show that time-varying explanatory variables have great value in facilitating identification and estimation of both duration dependence and unobserved heterogeneity.

There are two important sources of misspecification bias arising in duration models: misspecification of duration dependence and misspecification of the distribution of unobserved heterogeneity. In applied research, one often observes negative duration dependence. It seems in this case plausible to specify the duration dependence with a popular functional form that displays a monotonous relationship between the elapsed spell length and the transition probability. However, an observed declining hazard rate is not necessarily a causal consequence of spell length, but rather spurious duration dependence due to uncontrolled population heterogeneity, see e.g. Heckman and Singer (1985). To control the unobserved heterogeneity, many empirical studies have adopted a mixture distribution approach by assuming a parametric specification for the unobserved heterogeneity. Heckman and Singer (1984) have demonstrated that assuming parametric distribution for unobserved heterogeneity without sufficient economic evidence may lead to an "overparameterising" of the duration models. Such misspecification has some time posed great difficulty in estimation and inference of structure parameters of interest, as pointed out by Kiefer (1988). Due to the complexity of duration models, causal inference is often clouded by the uncontrolled unobserved heterogeneity and misspecification of the distributions for such unobserved heterogeneity.

Flexible specifications of duration dependence and unobserved heterogeneity seem to be a natural remedy to misspecification. With the evolvement of non-parametric approaches, many researchers turn to more flexible ways of modelling the duration models. Due to the complexity of non-parametric modelling, compromises are often made though to make estimation and inference feasible. The semi-parametric approach has been popular for many years; often the duration dependence is modelled non-parametrically, by a step function or spline approximation, so that no particular functional form is assumed. But in most of the semi-parametric studies a Gamma mixture model is used to account for unobserved heterogeneity and inference about structure parameters is conditioned on this distribution. Lancaster (1979) was the first to adopt a Gamma distribution for control of unobserved heterogeneity. Heckman and Singer (1984) argued that estimation on structure parameters is sensitive to the specification of the mixing distribution. They were the first to introduce the non-parametric specification for unobserved heterogeneity distribution, together with parametric duration dependence. Though in theory it is applicable to specify both the duration dependence and distribution of unobserved heterogeneity totally non-parametrically, the computational complexity involved seems to be a major obstacle. Very few previous successful implementations on non-parametric specification of both duration dependence and unobserved heterogeneity have been seen; among those is Røed and Zhang (2003).

We explore in this paper the identification results based on time-varying covariates from McCall (1994) and Brinch (2000). We estimate the mixed proportional hazard model with a set of unique time-varying covariates, namely calendar time variables that represents pure time changes in the hazard rates, e.g. business and seasonal cycles. With extensive Monte Carlo experiments, we provide empirical evidence that these explanatory variables are important additional identification sources. Our results show that the time-varying explanatory variables contribute to non-parametric identification and estimation of hazard models with mixing distribution of unobserved heterogeneity, and that sufficient variation in time-varying explanatory variables is a key to robust identification.

The rest of this paper is organized as follows: section 2 gives a brief discussion of the econometric approach and presents the non-parametric modelling of both duration dependence and unobserved heterogeneity. Identification of such models is discussed. Section 3 presents the main structure of experimental settings, the data generating process, and the computational strategies. Section 4 presents the main results for the single risk models. Special focus is given to how much the introduction of time-varying explanatory variables can contribute to estimation of unobserved heterogeneity non-parametrically, and how well the non-parametric approach can recover the structural parameters as well as the underlying duration dependence. Discussion of model selections with information criteria is included. Section 5 offers a summarising discussion of the estimation results for model components. Some implications of our findings are elaborated. We also provide some measures for the overall fit. Section 6 extends the method to dependent competing risks models where the unobserved heterogeneities from two competing states are assumed to be bivariate normal distributed. Section 7 gives concluding remarks.

2. Econometric model

In applied unemployment research the actual duration data that researchers are facing result from a combination of joint effects of several factors, such as spell duration, business cycle, seasonal and regional variations of labour market conditions etc. In many empirical studies of unemployment duration, the available data possess a discreteness feature. It could be due to the observational practice, such as in official unemployment registers, where updating of unemployment status happens at certain interval points of time, e.g. days, weeks or months (Røed and Zhang (2003)). It is also the case for interview based data sampling, in that retrospective data sampling repeats at certain time intervals. Another reason might be that the true transition does occur at discrete time, e.g. if completing an unemployment programme is mandatory for participants, transition will only occur after the programme has been finished. The estimation must take the discreteness into account. Røed and Zhang (2002) have showed that time-aggregation bias could result from disregarding the discrete data pattern. All these factors require that an econometric model should be carefully tailored to cope with these elements.

Let the duration of an unemployment spell be a stochastic variable T and its realization be τ. The formal definition of hazard rate (in a single risk case) is the probability of leaving original state within the small interval $(\tau, \tau + \Delta\tau)$, given that transition has not occurred prior to τ, conditional on other observed factors $\mathbf{X}$ and unobserved heterogeneity v that might have influence on transition probability.

(1) $\theta\left(\tau \mid \mathbf{X}, v\right) = \lim_{\Delta\tau \to 0} \dfrac{P(\tau \le T \le \tau + \Delta\tau \mid T \ge \tau, \mathbf{X}, v)}{\Delta\tau}$.

The most popular formulation of hazard rate is due to Cox (1972). The hazard rate is proportional such that,

(2) $\theta(\tau) = \lambda(\tau) \cdot \exp(\mathbf{X}'\beta) \cdot v$

where $\lambda(\tau)$ is called the baseline hazard rate, $\exp(\mathbf{X}'\beta)$ is the effect of covariates that influence the hazard rate proportionally, and v is meant to capture unobserved heterogeneity across individuals. Such hazard rate model is well known with the name Mixed Proportional Hazard rate model (MPH), where v usually has an unknown distribution.

It is often for the computational simplicity assumed that the spell duration is measured in continuous time. In that case one often assumes a continuous function form for hazard rate, e.g. a Weibull specification. When data possesses discreteness and when the discreteness is of importance, one needs a specification of hazard rate that takes account for that. Kiefer (1989), Prentice and Gloeckler (1978) have proposed a grouped hazard model when data is observed with some interval $\Delta\tau$. For the sake of simplicity, we can normalise the interval be $\Delta\tau = 1$ without loss of generality. The conditional survival function within interval $[d-1, d]$ (the subject survives until d given that no transition has occurred prior d-1, d=1,2,…) can be derived as $\exp(-\int_{d-1}^{d}\theta(\tau)d\tau)$. Thus the probability that transition taken place within interval $[d-1, d]$, given that no transition occurred before *d-1*, is then

$$(3) \qquad h(d) = 1 - \exp(-\int_{d-1}^{d}\theta(\tau)d\tau)$$

Here we use d as indicator of grouped hazard within interval $[d-1, d]$ and τ as underlying continuous time. In empirical applications, one often specifies the covariates and unobserved heterogeneity in exponential forms as well as the integral part of equation (3). Using equation (2), we can rewrite (3) to

$$(4) \qquad h(d, x, v) = 1 - \exp\left[-\exp\left(\log(\int_{d-1}^{d}\lambda(\tau)d\tau) + \mathbf{X}'\beta + \log(v)\right)\right]$$

where we assume for the time being that $\mathbf{X}$ and v do not vary over time. Equation (4) specifies the grouped hazard rate for interval $[d-1, d]$ in single risk case to be of a complementary log-log form.

The unobserved heterogeneity v has an unknown distribution. A popular approach is to adopt a Gamma distribution due to its computational advantage (Lancaster (1985)), but no particular justification has been advanced until recently in Abbring and van den Berg(2001). They have showed that a large class of distribution families converges to the Gamma distribution asymptotically, and in some cases the convergence is rather rapid. Heckman and Singer (1984) have introduced a non-parametric approach and showed that the support of unobserved heterogeneity can be specified by a set of mass points. They prove that the non-parametric maximum likelihood estimators are consistent.

The complementary log-log form of hazard rate in equation (4) has its great advantage of flexibility. Both the duration dependence and unobserved heterogeneity can be specified non-parametrically. By applying a step function to the duration it is possible to approximate a large class of parametric hazard rate family. Also, Heckman and Singer (1984) have showed that non-parametric specification of unobserved heterogeneity can approximate parametric distribution reasonably well. We believe it is also the most empirically applicable model that fits the true observational data. To specify non-parametrically the baseline, we can use e.g. a set of dummies λ_d to characterize the continuous baseline $\log(\int_{d-1}^{d} \lambda(\tau) d\tau)$. Define $\mu = \log(v)$. Equation (4) can be further elaborated to

(5) $$h(d,x,v) = 1 - \exp\left[-\exp\left(\lambda_d + \mathbf{X}'\beta + \mu\right)\right]$$

Let L_i denote the likelihood for the individual i[34]. If the spell is censored, we only observed that the spell lasts until d_i. The likelihood is then represented by the overall survival function up to d_i:

$$\prod_{s=1}^{d_i} \left(1 - h_i(s,x,v)\right), \text{ where } s = 1,2,3,\ldots,$$

If a spell with duration d_i is not censored, the contribution of this spell to the likelihood consists of two parts: the overall survival function up to d_i-1; and for the last time interval, $h_i(d)$. Let y_i be the censoring indicator, of which, $y_i = 1$ if the spell is not censored and $y_i = 0$ if it is censored. The overall likelihood for individual i with spell duration d_i is then given by:

(6) $$L_i = \left(h_i(d_i, x_i, v)\right)^{y_i} \cdot \prod_{s=1}^{d_i - y_i} \left(1 - h_i(s, x_i, v)\right)^{1-y_i}$$

With discrete distributed unobserved heterogeneity, assume the unobserved heterogeneity v has a support of N mass points, with probabilities $P_j, j = 1,\ldots,N$ and satisfies that $\sum_j P_j = 1$. the likelihood of an individual i with observed duration d_i is thus

[34] Here we in effect ignore the repeated spells from individuals, so each individual only contributes one spell. See paragraph 2 on page 183 for the motivation for this.

(7) $$L_i = \sum_j P_j \left[\left(h_i(d_i, x_i, v_j) \right)^{y_i} \cdot \prod_{s=1}^{d_i - y_i} \left(1 - h_i(s, x_i, v_j) \right)^{1-y_i} \right], \ \sum_j P_j = 1$$

The overall likelihood function for population of all individuals is then

(8) $$L = \prod_i \sum_j p_j \left[\left(h_i(d_i, x_i, v_j) \right)^{y_i} \cdot \prod_{s=1}^{d_i - y_i} \left(1 - h_i(s, x_i, v_j) \right)^{1-y_i} \right], \ \sum_j P_j = 1$$

and the log likelihood l is easily acquired by

(9) $$l = \sum_i \log \left\{ \sum_j p_j \left[\left(h_i(d_i, x_i, v_j) \right)^{y_i} \cdot \prod_{s=1}^{d_i - y_i} \left(1 - h_i(s, x_i, v_j) \right)^{1-y_i} \right] \right\}, \ \sum_j P_j = 1$$

Note that with the non-parametric specification of unobserved heterogeneity, the overall log likelihood is not additive, which imposes great computational challenge.

This likelihood specification has the advantage that it not only can cope with the censoring problem, but also easily allow time-varying covariates in an unrestricted form. Further more, it does not actually require a proportionality assumption. By e.g. interacting duration with covariates of interest, one can investigate how these affect the hazard rates at different phases of the unemployment spells.

A few serious attempts have showed that within the context of reduced form duration analysis, the mixed proportional hazard rate model is non-parametrically identified in the sense that given observations of (d_i, x_i, y_i), it is possible to derive the unique mapping from the data to the parameters of hazard rate model, within the general setting such as in equation (2), see e.g. van den Berg(2001) for a recent exposure. One of the earliest contributions to identification of duration dependence and unobserved heterogeneity is due to Elbers and Ridder (1982). They show that at least within the family of proportional hazard model, the duration dependence and unobserved heterogeneity are identified. Their result is generally based on parametric identification. Heckman and Singer (1984) utilise Kiefer and Wolfowiz (1956) and Lindsey (1983) results on identification of mixture distribution and propose a non-parametric specification of unobserved heterogeneity (as formulated above) and prove the identifiability in a non-censored Weibull model. Their work could be considered a milestone in non-parametric estimation of unobserved heterogeneity within the duration models. They find that structure parameters of hazard rate model can be well estimated by non-parametric specification of the unknown distribution of unobserved heterogeneity. But due to its complexity, empirical implementation of non-parametric estimation has rarely been successful. Baker and Melino (2000) conduct an extensive Monte Carlo study on the Heckman and Singer approach and show that non-parametric specification of both duration dependence and unobserved heterogeneity tends to produce biased estimators on structure parameters. This bias is somewhat due to over compensation or correction for the dispersion of unobserved heterogeneity and the estimators are bias away from zero. They suggest hence that the use of some penalised likelihood method will produce more accurate estimators.

Another identification source is by utilising repeated spells. Honoré (1983) provides identification results based on multiple spell cases, also see van den Berg(2001) for a survey. Roughly speaking, the idea is to adopt a fixed-effect approach similar to the ordinary panel data analysis and estimate the joint densities of multiple spells. This, however, imposes some difficulties in empirical application. First, the assumption that unobserved heterogeneity v is constant for repeated spells is rather strong. It is more likely that v can vary from spell to spell. Suppose we think that v represents individual's motivation in job search. It is conceivable that earlier unemployment experience would have demoralising effect, and hence the motivation for job search in late spells would be lower. Another problem might be that the number of available repeated spells can be strongly influenced by the observational window. The probability of having a second spell within a given time period is inversely related to the length of the first spell. The shorter the analysing period, the fewer repeat spells are available. Also, the uncensored part of the second spell is proportionally inverse to the length of the first spell, i.e. the longer the first spell is the shorter the uncensored part of the second one could be, given the fixed observational window.

McCall (1994) explores another identification source and proves that by including time-varying covariates, mixed proportional hazard model can be non-parametrically identified. Brinch (2000) extends the results of McCall and proves that it applies even without proportionality assumption. As long as there is sufficient variation in the covariates over time, combined with variation across observations, the mixed hazard model can be non-parametrically identified.

We utilise the ideas of McCall (1994) and Brinch (2000) and explore some unique feature of observational data that is often accessible in today's microeconometric research. The idea is to improve the identifiability by including an extra set of time-varying covariates that are exogenous to individuals as control for population heterogeneity. In applied studies, it is typical that local or macro economic environments will have effects on transitions from unemployment to work. Consider two individuals that are identical in every observed aspect and have experienced the same length of unemployment. The only observed difference between them is the calendar time at which they enter the unemployment. Given the assumption of proportional hazard, these two should experience the same hazard rate if they have the same value of unobserved heterogeneities. But if one experiences unemployment during a slump period when "everyone" is hit by the unemployment risk, while the other starts unemployment in a boom time when job opportunity is good and the overall outflow rate is high, it is intuitively plausible that the individual being unemployed in the boom time should have a better job opportunity and shorter duration than that of the "identical twin" in the slump time. The fact that they have the same spell length can then only be accredited to the unobserved differences between them, in addition to pure chance element. It is likely that the one unemployed in the boom time have more unfavourable personal characteristics than the one in the slump time with same spell length. This is to say that, the calendar time at which unemployment spells take places and undergo is a source of hazard rate variation, ceteris paribus, that contains information about the expected value of unobserved heterogeneity. Therefore by including such exogenous variation within the hazard rate formulation, the identifiability of unobserved heterogeneity should be improved. In time-series literature, this type of covariates is often named as lagged explanatory variables. We use this term to denote these calendar time covariates. Brinch (2000) provides a proof of identification of mixed hazard model based on time-varying covariates. He shows that even without proportionality assumption, variation of covariates over time is sufficient in identifying duration dependence, controlled for unobserved heterogeneity. We adopt his identification results and argue that the lagged explanatory variables in form of calendar time variation are unique time-varying covariates that contribute to the identification of unobserved heterogeneity.

One key assumption to facilitate the argument above is that the causal impact of any factors on the transition probability only occurs in current period of time, while their influences in earlier periods only have affected the selection of persons who have reached the current period. As Elster (1976 p.373) elegantly put: "the past itself cannot have influence upon the present over and above the influence that is mediated by the traces left by the past in the present". Loosely expressed, this implies that *conditional on all current values of the explanatory variables, any dependence between the current hazard rate and past (lagged) values of explanatory variable must reflect the influence of unobserved heterogeneity.* We could think that there is a sorting mechanism in labour market that "selects" unemployed out of unemployment within every period of time. Those who have favourable labour market characteristics would be selected first and those remaining are the sorted-out groups that have "unfavourable" employment characteristics. Thus the past unemployment history only reflects this selection mechanism, while the causal impact of any variables of interests will only affect the transition probability of the current period. By including control for this sorting mechanism, we have then an additional source of identification of unobserved heterogeneity.

The empirical evidence of this sorting mechanism is demonstrated in Røed and Zhang (2000). A critical prerequisite for utilising calendar time for identification purposes is of course that it is not perfectly correlated with spell durations. This utilisation also implies that multiple cohorts that start at different calendar time are required. With a single cohort that starts at one point of time, there is no variation in calendar time conditional on the duration, therefore it is impossible to identify unobserved heterogeneity without resting on the proportionality assumption. We have seen many studies using single cohort or limited number of cohorts due to limited access of data. With the increasing availability and variety of large administrative register data, particularly in the Nordic countries, researchers begin to be aware of the potentials that these data can provide.

In line with the argument above, we can further decompose the covariates into two groups: usual covariates such as individual observed heterogeneity, demographic characteristics, etc; and calendar time effects σ_t. By using a set of dummies, we can also estimate the calendar time effects non-parametrically. The formal hazard rate model used in our Monte Carlo investigations is thus (suppressing the subscript *i* for individual):

(5*) $$h(d,t,x,v) = 1 - \exp\left[-\exp\left(\lambda_d + \sigma_t + \mathbf{X}'\beta + \mu\right)\right]$$

where *t* represents the calendar time.

3. Design of study

Data Generating Process (DGP)

The main hazard rate model in simulation and estimation is that of equation (5*), namely a grouped hazard complementary log-log model. We choose time unit to be of integer length. To facilitate the comparison with real empirical work, we denote the time unit to be month, and scale the hazard rates such that they resemble monthly exit rates from unemployment. In the following, all time units are in terms of months.

We have experimented with different sizes of simulated samples and found that to maintain reasonably precise estimation yet manageable computational cost, sample size of at least 5,000 individuals is preferred. We also simulate samples of 10,000 and 50,000 individuals to explore the large sample property.

We only consider the case of one time-invariant covariate in $\mathbf{X}$ and define it to be dummy that has probability of 0.6 for $x = 1$. To simplify the interpretation, the coefficient β is set to 1. In empirical researches, **X** usually is the structural covariate that has the interpretation as a causal variable, e.g. it can represent economic incentives or treatment. Thus correct estimation of β is important for any causal inferences derived from the model.

For unobserved heterogeneity, we have considered a variety of distributions, both parametric and non-parametric. It is important at this point to make it clear which model term our simulation is made. In the MPH formulation, the unobserved heterogeneity is captured by a multiplicative term v (equation (2)), while in our complementary log-log formulation on grouped hazard, the estimation is actually conducted on $\mu = \log(v)$ (see e.g. equation (5*) above). In our experiments, we simulate the distribution of v directly, and transform to model term $\mu = \log(v)$. For each individual, we make a draw of v from a pre-decided distribution and use the logarithm of the simulated value additively into the complementary log-log hazard model. In order to make comparison across parametric and non-parametric distributions of unobserved heterogeneity, we simulate v such that they all have the unit mean. In the following, we simulate the unobserved heterogeneity from a discrete mass point distribution of 3 points with mean 1 and variance 0.6475, and a Gamma distribution with the same mean and variance as discrete mass points distribution.[35] Table 1 gives a brief overview of simulated distributions of unobserved heterogeneity.

We then simulate artificial datasets for each combination of sample sizes and distribution types of unobserved heterogeneity. There are 6 samples (3 sample sizes, 2 distributional types for unobserved heterogeneity), with fixed distribution of observed heterogeneity **X**.

The observational window is set to be 24 months long and we simulate 24 monthly calendar time covariates to indicate the calendar time effects. An important feature we wish to study is how the size of the variation of these calendar time covariates affect the identification and estimation of the model, so we simulate a set of 4 different cases of calendar parameters drawn from a $N(0,\sigma^2)$ distribution (see Table 1 for details).

Yet another important model term needs to be simulated, namely the baseline hazard rate. To be focused on the point and maintain manageable computational cost, we in the following will concentrate on constant hazard and negative hazard models. We use the widely applied Weibull distribution to represent the negative dependence baseline with shape parameter 0.9 and scale parameter 0.1,

$\theta_{weibull}(\tau) = \lambda^{\alpha} \cdot \alpha\tau^{\alpha-1}, \quad \lambda = 0.10,\ \alpha = 0.90.$

Since the Weibull hazard is continuous in time, and our model is based on discrete grouped hazard with time unit 1 month, some discretising is needed. We simply calculate the definite integral

$\int_{d-1}^{d} \theta_{weibull}(\tau)d\tau = \int_{d-1}^{d} \lambda^{\alpha} \cdot \alpha\tau^{\alpha-1}d\tau = \lambda^{\alpha} \cdot \tau^{\alpha}\Big|_{d-1}^{d},\ \lambda = 0.10, \alpha = 0.90.$

The first month grouped hazard rate is thus 0.1259. For the sake of comparison, constant duration dependence is given by an exponential distribution baseline with parameter *log(*0.1259*)=-2.0723* such that the hazard rate is approximately equal to that of the first month of Weibull.

[35] We have experimented with other parametric distributions of v such as lognormal, as well as discrete mass point distributions with 2 points, 4 points, 7 points of support. They all have unit means, but variances differ from each other. Based on our experiments and consideration on computational cost, we choose 3 mass points distribution and Gamma distribution in our formal Monte Carlo studies.

We are then able to simulate unemployment spells following equation (5*). Taking one of those simulated 6 individual samples and one set from 4 simulated calendar time samples, we first randomly choose a start month from 1-24, and calculate up to 12 monthly hazard rates with inputted covariates, unobserved heterogeneity terms, baseline and calendar time effects, from the start month and onwards. Then for each month, we simulate actual transition from a uniform distribution. If the drawing from this distribution does not exceeded the calculated hazard rate, a termination of the spell has established, and we set the transition indicator *y* equal to 1. If at the end of observation window, i.e. month 24, there is still no transition, the spell is then censored and *y* takes the value 0. If on the other hand the spell length has reached 12 months and still no transition, the spell is censored as well. We conduct this routine of spells simulation for all combinations of duration dependences and distributions of unobserved heterogeneity, as well as different calendar parameters variations and sample sizes. There are totally 48 model combinations (3 sample sizes, 2 duration dependences, 2 types of distributions for unobserved heterogeneity and 4 cases of calendar time variations). We then repeat the sampling process 100 times to get 4,800 samples.

Since in the model the calendar time terms function as time-varying covariates, in estimation, it is necessary to split each spell into many subspells, each has duration of *one month* and total number of subspells should be equal to the total length of original spell. Each splitted one-month spell has been defined as censored except the last one, which retains its original censoring status. This is a known technique in dealing with time-varying covariates. This episode splitting operation results in data sets with monthly observations ranging from 25,000 up to over 300,000.

Table 1: Data Generating Process (DGP)

Duration dependence	*Distribution*	*Scale factor*	*First month hazard rate*
No duration dependence	Exponential, $\lambda = -2.0723$	-2.0723	0.1259
Negative duration dependence	Weibull $\lambda = 0.10, \quad \alpha = 0.90$	-2.0723	0.1259
Observed Heterogeneity, X	Pr(x=1)=0.6, Pr(x=0)=0.4		
Calendar time variation	$N(0,\sigma^2)$, $\sigma^2 = 0,\ 0.001,\ 0.1,\ 1$		

Unobserved heterogeneity			*Mean*	*Variance*
Gamma			1	0.6475
Discrete	*Support Points*	*Probability*	*Mean*	*Variance*
3 points	1.80	0.50	1	0.6475
	0.30	0.30		
	0.05	0.20		

Computational strategies

Heckman and Singer (1984) propose a three-step algorithm in determining the number of mass points for unobserved heterogeneity: they start with one point of support, maximise the loglikelihood to achieve start value for search; in step 2, they scan a grid of admissible intervals of potential support for mass points conditional on estimated parameters in step 1 and acquire the interval which gives the largest Gateaux derivatives. If the Gateaux derivative is non-positive everywhere within the interval, stop. Otherwise, estimate the model with 2 points. Proceed to step 3 by evaluating the Gateaux derivative conditional on estimated parameters in step 2 and repeat the procedure until Gateaux derivative is negative or zero. They find that the EM algorithm usually provides a satisfactory convergence. Baker and Melino (2000) use a similar approach but instead of Gateaux derivatives in step 2 and step 3, they use the more familiar Kuhn-Tucker multiplier and maximise loglikelihood function under constraint $\sum_j P_j = 1$.

The choice of algorithm used by Heckman and Singer (1984) as well as Baker and Melino (2000) is most likely due to computational tangibility, in that at each iteration, the gradient searching direction is (hopefully) optimised. However, it might be the case, as we experienced, that the search interval might lead to a local maximum. By restricting search direction by such interval, it is difficult to switch to the "correct" path once the search direction is already leading to an inferior maximum. Also Baker and Melino (2000) find it is often the case that the optimal solution to Kuhn-Tucker is found at the corner of the search interval with negligible probability. This is unfavourable with respect to computational cost. In addition, Heckman and Singer as well as Baker and Melino algorithms are only for single risk case. We originate our programming with the consideration to apply on competing risks models as well, and it has proven to be quite cumbersome to evaluate Gateaux derivatives in multiple dimensions. Therefore we choose a more direct approach: we start with 1 point of support and add one additional point at each iteration until likelihood cannot be improved numerically. At each iteration, we first carry out a few line searches with BFGS method to acquire search direction that makes increment of likelihood largest with added point. The initial value for search is taken from the previous iteration, except that the distribution of mass points is randomly chosen (scrambled). After an optimal search direction is found, we switch to Newton-Raphson method for functional maximisation. It proves that in most of cases our approach seems to perform well.

In the construction of simulated hazard, the duration baseline is normalised to the first month by the scale factor; the calendar months are normalised to month 13. Hence the model is estimated with a constant term c.

(5**) $h(d,v) = 1 - \exp\left[-\exp\left(c + \lambda'_d + \sigma'_t + \mathbf{X}'\,'\beta + \mu\right)\right]$

where $(\lambda'_d, \sigma'_t, \mathbf{X}')$ are all normalised to their respective references. In the case of no unobserved heterogeneity, the exponential of constant term c is thus the true duration baseline hazard rate of the first duration month, i.e. $\exp(c)$ with the mean calendar variation for a person with $x = 0$. With the presence of unobserved heterogeneity, the constant is actually the sum of c and μ, i.e. we do not obtain directly estimates for μ. Thus in estimation and post estimation inferences, we evaluate the estimated sum $(c + \mu)$ in (5**).

The probability P_j is formulated with a logistic formulation

$$P_j = \frac{\exp(\gamma_k)}{1 + \sum_k \exp(\gamma_k)} \text{ for } k = 2,3,\ldots,N \text{ and } P_j = \frac{1}{1 + \sum_k \exp(\gamma_k)} \text{ for j} = 1$$

to ensure the probabilities lie within [0,1]. However, this also means that the probability of an additional point can never be exactly zero, which implies that additional points may be included even though the probability for this point is extremely small, and increment of likelihood is numerically insignificant. Therefore we choose from time to time an ad hoc criterion to stop the search for further points when distribution of current estimated mass points involves some very small probabilities. The threshold for small probability in most cases is set to 10^{-4}.

In maximising the finite mixing distribution characterized by (5*), the maximised log likelihood might raise the problem of selection of optimal number of points for the mixing distribution. In our case, it might be that the number of points found are more than necessary for a good fit of the observational data. Leroux (1992) suggests that a procedure that penalises overfitting might be preferable to pure maximum likelihood, and proposes a solution that he labels the maximum-penalised-likelihood. Huh and Sickles (1994) have showed that the maximum penalised likelihood estimators are consistent in duration models with unobserved heterogeneity, provided the mixing distribution can be characterised by a finite number of points of support. The general form for a maximum-penalised-likelihood is (Leroux (1992))

$$l_n(\hat{\mu}_m) - a_{mn}$$

where $l_n(\hat{\mu}_m)$ is maximised loglikelihood with estimator $\hat{\mu}_m$ and a_{mn} is the penalty term, m is the number of components in finite mixing distribution, while n is number of observations. Baker and Melino (2000) propose to use either Bayesian Information Criterion (BIC, Schwarz (1978)) or Hannan-Quinn Information Criterion (HQIC, Hannan and Quinn (1979)) to penalise the additional spurious point that introduces "overparameterisation" bias. The BIC is defined with $a_{mn} = (1/2)\log(n)\dim(\hat{\mu}_m)$, while HQIC is defined with $a_{mn} = \log(\log(n))\dim(\hat{\mu}_m)$, where $\dim(\hat{\mu}_m)$ is the dimension of mixing distribution, which is equal to number of independent parameters. We consider these two information criteria in our analysis. In addition, we also include Akaike (1973) information criterion (AIC) based on Kullback's symmetric divergence (1968), as alternative definitions for penalty term. A variant of AIC can be defined as $AIC = l_n(\hat{\mu}_m) - \dim(\hat{\mu}_m)$. Thus in evaluation of convergence and optimal dimension of mass points, we apply both pure maximum likelihood criterion and maximum penalised likelihood with 3 information criteria to determine the optimal model choice with respect to number of support points found.

It is known that maximisation of non-parametrically specified likelihood is extremely cumbersome (Baker and Melino (2000), footnote 12). In this paper we solve this problem by using an approach which we call “implicit dummy” technique. This technique efficiently reduce computational cost on redundant multiplications of zero value dummies that are due to non-parametric specification, and hence remarkably improves the speed of the maximisation. The maximisation routine is hard-coded in Fortran 90 with MPI implementation for parallel processing[36]. All estimations are carried out on a HP Superdome (44 PA8600 CPUs prior to July 2003) running HP-UX with HP’s Fortran 90 compiler. Compiler-native Lapack and BLAS have been used. A typical run for a sample size of 50,000 individuals, up to 50 parameters, with 4 CPUs utilised, takes approximately 40-50 minutes in real time.

We emphasize at this stage that maximisation is extremely difficult in the region around potential maximum, as already pointed out by Heckman and Singer (1984). The likelihood function is not globally concave, and our experience suggests that the likelihood is quite flat around the potential maximum and has a “wash-board” like texture with plenty of local maxima. We need to distinguish two types of local maxima: sets of equivalent maxima and qualitatively different maxima. By equivalent maxima we mean that given the random search direction, our iteration can end up in a set of numerically equivalent maxima, characterised by approximately the same estimators on the coefficients and moments of the distribution of unobserved heterogeneity, as well as likelihood function value. Thus convergence to any of these maxima can be regarded as convergence to the global maximum. The qualitatively different maxima refer to the fact that by altering search direction, convergence might be reached at another maximum that is significantly different both in terms of likelihood function value and estimators of the parameters than we otherwise might find. This is a more serious problem.

To ensure that the global maximum is located, for each model, we find it necessary to repeat each estimation multiple times with randomly chosen starting values and

[36] We are fortunate to have Senior Analyst Simen Gaure at the University Information Technology Centre at University of Oslo to help us programme the estimation routine. All estimations are done on HP Superdome at High Performance Computing Centre, University of Oslo.

randomly chosen search direction in each linear search. It turns out that our method in most cases is robust regarding the starting values and ends up approximately the same maximised likelihood. However, since currently no explicit guidance is available on determination of global maximum when likelihood function is non-concave globally, we interpret our results with caution. Nevertheless in most cases we are reasonably confident that the global maximum is found.

4. Results

We conduct extensively non-parametric maximum likelihood estimation on all simulated data. To be concise about our results, we focus on two representative models: the constant duration dependence (non duration dependence) hazard rate model with 3 mass points discrete distribution for the unobserved heterogeneity, and the negative duration dependence (Weibull hazard rate model) with a parametric mixture distribution for the unobserved heterogeneity characterised by the Gamma distribution[37]. To investigate our proposition that time-varying covariates in the form of calendar dummies improves the identifiability, we look into the cases that calendar time variations are generated with variances being set as 0, 0.001, 0.1 and 1. The main results are organised as following: We first report maximisation of log likelihood and iteration process for the selected models. Special attention is given on the choice of models in terms of estimated number of mass points. We also look into the estimation on the structure parameter β and how the estimator changes over iterations. Second, we report the distribution of estimators on β and distributional properties of estimators for different model settings using kernel densities of estimated β through 100 repetitions. Third, we report the estimated duration dependence with respect to support points found by plotting estimated baseline hazard rates. Also we report the measure of average weighted squared errors for estimators on duration dependence parameters. Fourth, we will comment the estimation of time-varying calendar time parameters by reporting average weighted squared errors for estimators as well. We also look into the consequences of ignoring such time-varying calendar variations in estimation. Last, we

[37] We have also looked into models of constant hazard with Gamma mixture and Weibull hazard with 3 points discrete mixture distributions. The findings from these models are virtually the same as those we present in this section. The full sets of all results are available upon request.

will compare estimated moments of mass point distributions of unobserved heterogeneity to see how well the estimates can approximate the true distribution moments. We follow the results of model components estimation by some discussions of the implications of our findings and a measure for overall fit in next section. For all results we will consider a variety of specifications for both maximum likelihood and maximum penalised likelihood, as well as sampling properties and effects of variation of calendar times.

1. Convergence and choice of optimal model dimension

We first report the maximum number of support points found by maximum likelihood method and maximum penalised likelihood in the form of information criteria, in the 100 trials. Table 2 reports the maximum points found for samples with 5,000 individuals, for constant duration dependence and negative duration dependence models. An immediate observation is that, when the true mixing distribution for unobserved heterogeneity is generated with discrete distribution, the pure maximum likelihood method tends to find more points than used to generate the data. For example, in the first panel of Table 2, when the unobserved heterogeneity in DGP is discretely distributed with 3 mass points, the maximum likelihood method tends to find number of points ranging 3 to 6, while AIC, BIC and HQIC in most cases are able to find correct number of points. Similar pattern can be found for loglikelihood method when the true unobserved heterogeneity distribution is Gamma. The optimal number of points found by loglikelihood is ranging 3-6, while AIC and HQIC find optimal number of support points to be 2 and 3. BIC is quite conservative with respect to added points when the unobserved heterogeneity is Gamma distributed, and in most cases fails to find more than 1 point of support.

A second observation is that when the variation of calendar time parameters increases, the number of trials that found excessive points is somewhat reduced. This can be seen from the first panel for the discrete distribution case. When the variance of calendar variation is zero, there are 24 estimations in which the loglikelihood criterion results in 6 or more points of support for the unobserved heterogeneity distribution, while when variance is 1, only 9 out of 100 estimations return 6 or more points. The pattern is not clear for Gamma distributed unobserved heterogeneity.

Table 2: Maximum number of support points found.

Constant hazard 3 points generated unobserved heterogeneity, 5,000 obs.

	Var(month)	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
loglikelihood	0	23	0	13	22	18	14	10
	0.001	5	0	18	20	29	15	13
	0.1	8	0	26	30	26	9	1
	1	18	0	14	32	27	7	2
AIC	0	23	0	63	13	1		
	0.001	5	0	69	18	8		
	0.1	8	0	84	7	1		
	1	18	0	73	9	0		
BIC	0	23	0	77	0	0		
	0.001	6	1	93	0	0		
	0.1	9	24	67	0	0		
	1	18	17	65	0	0		
HQIC	0	23	0	77	0	0		
	0.001	5	0	94	1	0		
	0.1	8	3	89	0	0		
	1	18	1	80	1	0		

Weibull hazard, Gamma distributed unobserved heterogeneity, 5,000 obs.

	Var(month)	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
loglikelihood	0	18	1	13	34	21	9	4
	0.001	4	0	29	28	25	12	2
	0.1	13	2	18	31	26	8	2
	1	12	0	22	37	22	5	2
AIC	0	18	14	62	5	1		
	0.001	9	15	67	7	2		
	0.1	18	40	39	2	1		
	1	12	50	32	6	0		
BIC	0	95	4	1	0	0		
	0.001	96	2	2	0	0		
	0.1	93	7	0	0	0		
	1	63	37	0	0	0		
HQIC	0	28	41	31	0	0		
	0.001	35	39	26	0	0		
	0.1	62	31	7	0	0		
	1	17	70	13	0	0		

Table 3 reports maximum number of support points found for different sample sizes. Again, the loglikelihood methods have the tendency to find excessive points regardless of sample sizes. For discrete distributed unobserved heterogeneity most information criteria methods returns 3 or 4 points of support regardless the sample size. However, when the mixing distribution is generated by Gamma, BIC and HQIC does not seem to be able to find more than 3 points of support, but increasing sample sizes do enable the BIC and HQIC to find more than 1 point of support.

Table 3: Maximum number of support points found across sample sizes.

Constant hazard 3 points generated unobserved heterogeneity, var(month)=0.1

	Obs	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
Log likelihood	5000	8	0	26	30	26	9	1
	10000	20	0	7	29	27	13	4
	50000	1	0	15	35	26	19	4
AIC	5000	8	0	84	7	1		
	10000	20	0	72	5	3		
	50000	1	0	85	11	3		
BIC	5000	9	24	67	0	0		
	10000	20	25	55	0	0		
	50000	1	0	99	0	0		
HQIC	5000	8	3	89	0	0		
	10000	20	1	78	1	0		
	50000	1	0	97	2	0		

Weibull hazard, Gamma distributed unobserved heterogeneity, var(month)=0.1

	Obs	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
Log likelihood	5000	13	2	18	31	26	8	2
	10000	13	1	12	39	28	7	0
	50000	2	0	3	44	36	12	3
AIC	5000	13	2	18	31	26		
	10000	14	42	39	5	0		
	50000	2	7	74	15	2		
BIC	5000	93	7	0	0	0		
	10000	69	30	1	0	0		
	50000	2	94	4	0	0		
HQIC	5000	93	7	0	0	0		
	10000	19	72	9	0	0		
	50000	2	49	49	0	0		

By looking into some of the typical iteration processes from estimations, we will show more clear pictures of convergence and impact of number of support points found on estimation of structure parameters. In Table 4-1 to 4-2, we report some typical iteration processes and convergences of loglikelihood for small sample (5,000 observations) models with and without duration dependence, together with non-parametrically and parametrically generated unobserved heterogeneity distributions. To produce these tables, for each selected combination of duration dependence, unobserved heterogeneity and calendar variation, we *arbitrarily* choose one result from 100 trials that returns more than one point of support. The loglikelihood for each iteration is reported, as well as the penalised loglikelihood by various information criteria. The bold faced values indicate the optimal choice of points according to each criterion. The estimated structure parameter $\hat{\beta}$ serves in this case as a benchmark to evaluate how well each criterion

Table 4-1 Constant hazard, 3 points distributed unobserved heterogeneity, 5,000 individuals.

model	obs	points	# parameter	loglikelihood	AIC	BIC	HQIC	$\hat{\beta}$	std
Constant Hazard	31094	1	1	-10225.7765	-10226.7765	-10230.9489	-10228.1130	0.5545	0.0359
Var(σ)=0	31094	2	3	-10214.3447	-10217.3447	-10229.8619	-10221.3541	0.7456	0.0790
(2)	31094	3	5	**-10191.6759**	**-10196.6759**	**-10217.5378**	**-10203.3583**	0.9606	0.0631
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	25851	1	1	-8829.0631	-8830.0631	-8834.1432	-8831.3816	0.6137	0.0386
Var(σ)=0.001	25851	2	3	-8809.3181	-8812.3181	-8824.5583	-8816.2735	0.8965	0.0783
	25851	3	5	-8781.8830	**-8786.8830**	**-8807.2833**	**-8793.4754**	1.0921	0.0654
(1)	25851	4	7	**-8780.5567**	-8787.5567	-8816.1171	-8796.7860	1.0947	0.0760
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	28555	1	1	-8339.5278	-8340.5278	-8344.6592	-8341.8563	0.7155	0.0419
Var(σ)=0.1	28555	2	3	-8339.2958	-8342.2958	-8354.6899	-8346.2813	0.7152	0.0505
	28555	3	5	-8318.4729	-8323.4729	**-8344.1298**	**-8330.1155**	1.0920	0.0885
	28555	4	7	-8316.1579	**-8323.1579**	-8352.0775	-8332.4575	1.0946	0.1164
	28555	5	9	**-8315.6447**	-8324.6447	-8361.8270	-8336.6014	1.2439	0.1416
	28555	6	11	-8316.1672	-8327.1672	-8372.6122	-8341.7809	1.1120	0.1359
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	17469	1	1	-7180.5639	-7181.5639	-7185.4480	-7182.8430	0.5355	0.0344
Var(σ)=1	17469	2	3	-7100.5928	-7103.5928	-7115.2451	-7107.4302	0.7529	0.0499
	17469	3	5	-7057.5060	-7062.5060	-7081.9264	-7068.9016	0.9568	0.0575
	17469	4	7	-7048.0878	-7055.0878	-7082.2764	-7064.0417	1.0392	0.0656
(2)	17469	5	9	**-7035.1158**	**-7044.1158**	**-7079.0727**	**-7055.6280**	1.1845	0.0728

Note: 1. Number of observation listed in table is number of monthly observation is estimation data. 2. Var(σ) is variance of calendar month in simulation.
3. Number of parameters is free parameters associated with unobserved heterogeneity. (1) indicates iteration terminates when approximate zero probability on added point is encountered. (2) indicates numerical difficulty prevents further search of mass points.

Table 4-2 Weibull hazard, Gamma distributed unobserved heterogeneity, 5,000 individuals.

model	obs	points	# parameter	loglikelihood	AIC	BIC	HQIC	$\hat{\beta}$	std
Weibull Hazard	29757	1	1	-10622.5733	-10623.5733	**-10627.7237**	-10624.9055	0.7361	0.0357
Var(σ)=0	29757	2	3	-10616.8160	-10619.8160	-10632.2672	**-10623.8126**	0.9027	0.0822
	29757	3	5	-10614.0579	**-10619.0579**	-10639.8100	-10625.7191	0.8952	0.0687
(2)	29757	4	7	**-10613.5947**	-10620.5947	-10649.6476	-10629.9203	1.0212	0.1559
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	24278	1	1	-9128.2983	-9129.2983	-9133.3470	-9130.6106	0.7420	0.0382
Var(σ)=0.001	24278	2	3	-9125.9219	-9128.9219	-9141.0679	-9132.8588	0.7432	0.0466
	24278	3	5	-9106.6811	-9111.6811	**-9131.9245**	**-9118.2425**	1.0675	0.0691
	24278	4	7	-9105.1306	-9112.1306	-9140.4712	-9121.3165	1.4290	0.1201
	24278	5	9	**-9099.1835**	**-9108.1835**	-9144.6214	-9119.9939	1.6184	0.1438
	24278	6	11	-9099.2332	-9110.2332	-9154.7685	-9124.6682	1.5805	0.4039
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	27702	1	1	-8580.3997	-8581.3997	**-8585.5143**	**-8582.7249**	0.7870	0.0420
Var(σ)=0.1	27702	2	3	-8580.0429	-8583.0429	-8595.3868	-8587.0187	0.7876	0.0508
(2)	27702	3	5	**-8571.3220**	**-8576.3220**	-8596.8951	-8582.9482	0.9668	0.0833
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	28163	1	1	-8164.4505	-8165.4505	**-8169.5734**	-8166.7774	0.7892	0.0407
Var(σ)=1	28163	2	3	-8154.3023	-8157.3023	-8169.6710	**-8161.2829**	0.9797	0.0695
	28163	3	5	-8151.9303	**-8156.9303**	-8177.5447	-8163.5647	0.9852	0.0653
	28163	4	7	-8150.1599	-8157.1599	-8186.0201	-8166.4480	0.9959	0.0689
(1)	28163	5	9	**-8148.7837**	-8157.7837	-8194.8896	-8169.7255	1.0325	0.0705

Note: 1. Number of observation listed in table is number of monthly observation is estimation data. 2. Var(σ) is variance of calendar month in simulation.
3. Number of parameters is free parameters associated with unobserved heterogeneity. (1) indicates iteration terminates when approximate zero probability on added point is encountered. (2) indicates numerical difficulty prevents further search of mass points.

performs. Also in each table, we consider results from which the calendar dummies have variances ranging from 0 to 1.

At first iteration, the number of mass points is simply 1, which means there is no control for unobserved heterogeneity. We observe immediately that the estimated $\hat{\beta}$ is significantly biased towards zero. For example in Table 3-1, when no variation of calendar dummies (or the single cohort that starts at the same calendar time), without control of unobserved heterogeneity, the estimated $\hat{\beta}$ is 0.5545, which has a bias as large as 45%. This confirms the well-known fact that uncontrolled unobserved heterogeneity produces non-negligible biased estimates towards zero. At iteration two, we add 1 point of support to the distribution of unobserved heterogeneity. This also means the free parameters associated with the distribution of unobserved heterogeneity is 3 (2 for mass points and 1 for probability). At this stage, by examining estimates from all models, we see no significant improvement on estimation of structure parameter. When we have 3 mass points, in almost all models the estimate on $\hat{\beta}$ is very close to 1. But it seems to be the case that the likelihood can be improved further by adding additional points. It is observed immediately that when there are 4 points (Table 4-1, constant hazard, $\text{var}(\sigma) = 0.1$), the estimate of $\hat{\beta}$ is quite larger than the true value 1, and with additional points being added, the $\hat{\beta}$ displays stronger positive bias. We continue the iteration until the likelihood deteriorates[38]. For example in constant hazard with calendar dummies' variance set to 0.1, the maximum likelihood criterion would conclude that maximum is reached when there are 5 points of support found. However, if we adopt some form of information criterion, we would find that the optimal choice of number of points is reached at 3 points (BIC and HQIC). And the $\hat{\beta}$ is very close to the true value 1 at 3 points.

Another important finding is that, when sample size is relative small (e.g. 5,000 individuals), we find evidence that applying some kind of information criterion to penalize excessive points is more favourable than pure likelihood criterion, especially when the true distribution of unobserved heterogeneity is characterized by 3 points of support. We find that AIC seems more in line with likelihood, and BIC and HQIC are more conservative with adding extra points. On the other hand, BIC and HQIC seem to have the tendency to underestimate $\hat{\beta}$. A possible explanation can be attributed to the definition of information criteria: since BIC and HQIC depend not only on the number of free parameters, but also on the sample size. Given the number of free parameters, the size of penalty is solely decided by number of observations. In small sample cases, it seems that the increment of loglikelihood value from iteration to iteration is small relative to the penalty term, and the BIC and HQIC often "overcorrect" the excessiveness of loglikelihood and return estimates below the true value. Since AIC does not involve sample size, it seems to be the most balanced choice among all. Although in most cases the information criteria give roughly the same (or statistically equivalent) estimates, we find evidence in favour of using AIC as a suitable measure for model choice.

Increasing sample sizes does show improvement of estimator for $\hat{\beta}$, even though the number of points found exceed the true mixing distribution when it is generated with 3 points. Appendix Table A1-1 to A1-2 report evidence of estimation on samples generated with 10,000 individuals. Appendix Table A2-1 to A2-2 report some results of reestimated the same models on even larger sample of 50,000 individuals. For models with no duration dependence and discretely generated unobserved heterogeneity, our findings on small sample become more obvious. We find again that maximum likelihood estimator tends to find excessive points for distribution of unobserved heterogeneity and overestimate the structure parameter. This also holds for Weibull baseline model with Gamma distributed unobserved heterogeneity (Table A1-2). We find that differences between information criteria become smaller when sample size is larger. When sample size is sufficiently large (Table A2-1, A2-2), we find that all criteria on model choice give almost equivalent results. The estimates on $\hat{\beta}$ are almost the same whichever criterion is chosen. And the estimates are very close to the true value with very good precision. This can be seen for both models. Also, variation of calendar dummies seems to be less important, though including calendar dummies as an

[38] This could purely be due to the numerical precision phenomenon since we can always set values of additional parameters equal to 0. On the other hand, this could also imply that the search has switched to another (inferior) local maximum.

additional source of identification contributes to the precision of estimation, which can be recognized by examining standard errors of $\hat{\beta}$ for Table A2-1 to A2-2.

The above crude examinations of some typical iterations and maximising processes across sample sizes and variations of time-varying covariates for calendar time provide some intuitions on how the maximisation process recovers the true structure parameter. It suggests that the likelihood criterion seems to find excessive number of points and "overparameterise" the dispersion of unobserved heterogeneity. Consequently, to compensate this "overestimation", the structure parameters are upwards biased. There are evidences that information criteria produce better estimators when sample sizes are small, while variations of time-varying calendar covariates help the estimation on the structure parameters.

2. Estimation on β

To further illustrate the relationship between excessive number of points returned by loglikelihood and biases produced by such excessive points on structural estimator, we calculate mean deviation between estimators for β acquired by loglikelihood in 100 trials, and the true value 1 in DGP. We also look into different cases of calendar time variations and how they affect the mean deviations. Figure 1 and 2 are plots of mean deviations for models of constant hazard with 3 points distributed unobserved heterogeneity and Weibull hazard with Gamma distributed mixture in DGP.

It is clear that in both figures, when estimation fails to find more than 1 point of support for unobserved heterogeneity (which means no control for unobserved heterogeneity), the estimates on β are biased towards zero (mean deviations from true value 1 in DGP are all negative). With more points found by loglikelihood criterion, the mean deviations turn to be positive and increase with the number of mass points. When the unobserved heterogeneity is generated from 3 points distribution, and when the loglikelihood finds the correct number of points, the mean deviation is the smallest (Figure 1). For Weibull model with Gamma distributed unobserved heterogeneity in DGP, 4 points seem to give the smallest bias, although 3 and 5 mass points perform relative well too.

Figure 1: Mean deviations of estimated $\hat{\beta}$ from true value 1 by maximum number of mass points found by maximum loglikelihood. Constant hazard, 3 points mixture in DGP. var represents calendar variation in DGP. Obs=5,000.

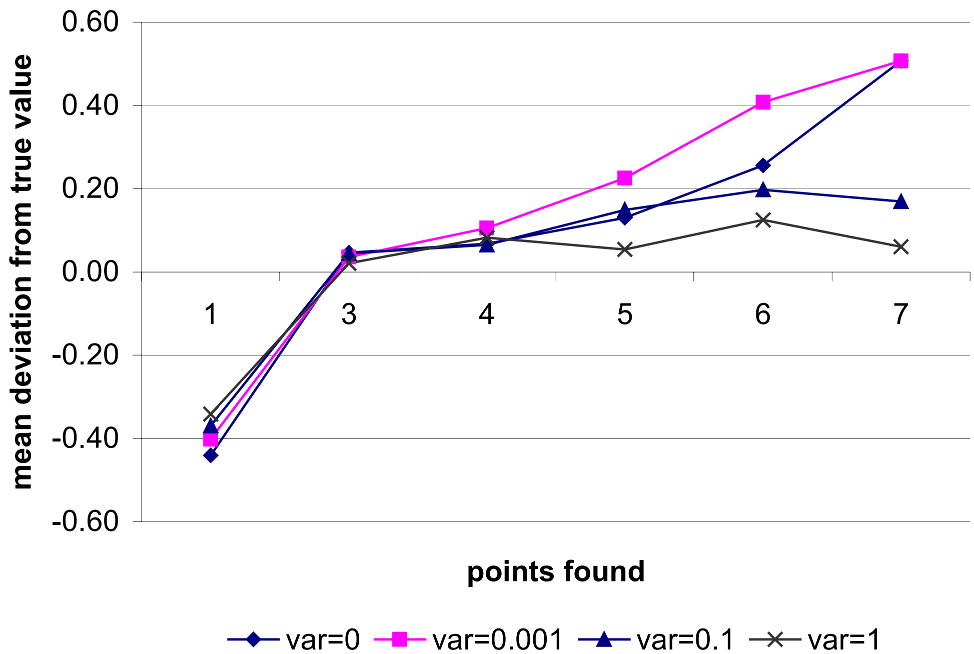
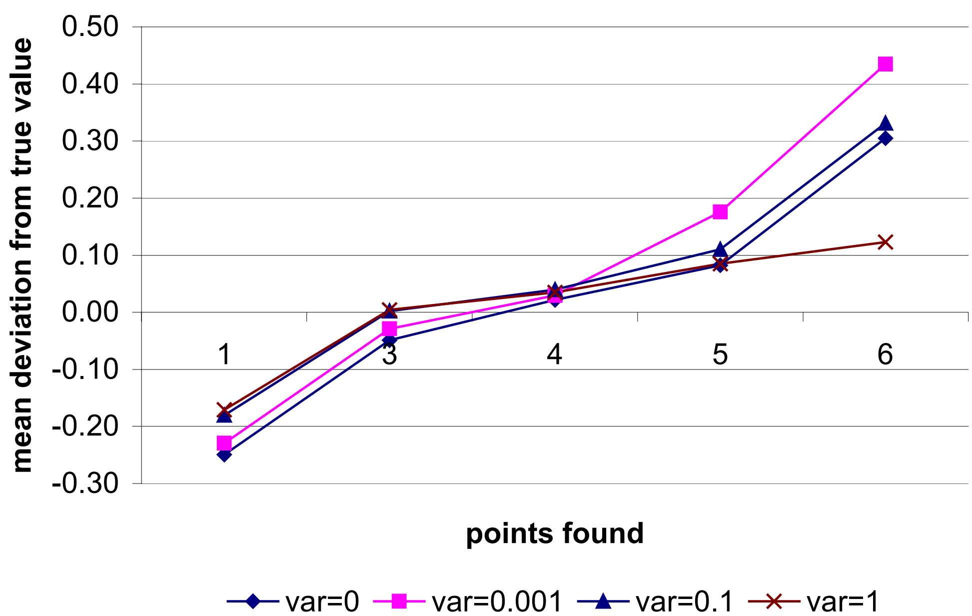


Figure 2: Mean deviations of estimated $\hat{\beta}$ from true value 1 by maximum number of mass points found by maximum loglikelihood. Weibull hazard, Gamma mixture in DGP. var represents calendar variation in DGP. Obs=5,000.



It is also notable that the mean deviations seem to be smaller when calendar variation is larger. From both Figure 1 and Figure 2, we observe that when the calendar variation is sufficiently large (variance is 1), the biases measured by mean deviations are smaller and roughly below 0.1, compared to small or none calendar time variations. In constant hazard case with large calendar time variation, even when the loglikelihood finds 7 or more points, the biases are moderate compared with that in small calendar variation cases. This suggests also that the calendar time variations seem to contribute to the reduction of estimation biases on structure parameters, at least in the case when loglikelihood criterion returns excessive number of points for unobserved heterogeneity distribution.

We will turn to the details of the distributional properties of non-parametric maximum likelihood estimators and maximum penalised likelihood estimators for structure parameter β. Table 5 reports estimated structure parameter $\hat{\beta}$, for constant hazard with 3 points support and Weibull hazard with Gamma distributed unobserved heterogeneity. Means and standard deviations are calculated across trials that find more than 1 point of support for the unobserved heterogeneity. An encouraging observation is that for most of the estimations, the structure parameter $\hat{\beta}$ is very well estimated, the means are very close to the true value 1 in DGP[39].

Several observations can be made: First, the log likelihood criterion has the tendency to overestimate the structure parameter $\hat{\beta}$ when sample sizes are small, particularly when there is no or little calendar time variation in hazard rates. It seems that the data is less informative for a successful recovery of structure parameter when there is no or little calendar variation. In this case, it helps for the estimation when some form of information criterion is used to penalise the excessive mass points found by log likelihood. We find that for constant hazard with 3 points discrete unobserved heterogeneity (Table 5 upper panel), both BIC and HQIC perform well. AIC is more in

[39] Table A3 in Appendix reports number of trials among each 100 repetitions that the 95% confidence intervals of estimators cover the true value 1. For almost all estimations, over 70 per cent trials produce the confidence intervals that cover the true value. Among model selection criteria, there does not seem to be much difference, except BIC in Weibull hazard with Gamma distributed unobserved heterogeneity. For small samples, BIC is extremely poor comparing to other criteria. As sample sizes increase and with large calendar variations, BIC performs much better. While in Constant hazard 3 points unobserved heterogeneity models, all criteria seem to be equally successful.

line with loglikelihood. While in the case of Weibull hazard with Gamma unobserved heterogeneity (Table 5 lower panel)[40], all model selection criteria give roughly the same means for estimated $\hat{\beta}$.

Table 5: Estimated means and standard errors of $\hat{\beta}$.

Constant hazard, 3 points unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	1.1803	0.2491	1.0691	0.1237	1.0275	0.0558	1.0275	0.0558
	0.001	1.2387	0.3022	1.1238	0.1853	1.0446	0.0665	1.0474	0.0733
	0.1	1.0967	0.1106	1.0621	0.0914	1.0315	0.0688	1.0458	0.0620
	1	1.0645	0.0746	1.0456	0.0739	1.0275	0.0768	1.0406	0.0688
10000	0	1.0654	0.0886	1.0260	0.0513	1.0159	0.0418	1.0191	0.0427
	0.001	1.0914	0.2091	1.0248	0.0567	1.0156	0.0548	1.0192	0.0476
	0.1	1.0601	0.0851	1.0300	0.0726	0.9997	0.0680	1.0185	0.0637
	1	1.0336	0.0608	1.0226	0.0547	1.0092	0.0623	1.0186	0.0542
50000	0	1.0310	0.0455	1.0141	0.0380	0.9931	0.0178	0.9976	0.0269
	0.001	1.0334	0.0742	1.0055	0.0323	0.9930	0.0192	0.9941	0.0211
	0.1	1.0188	0.0342	1.0078	0.0266	1.0031	0.0213	1.0045	0.0247
	1	1.0136	0.0191	1.0052	0.0188	1.0013	0.0171	1.0013	0.0171

Weibull hazard, Gamma distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	1.0919	0.2147	0.9825	0.0779	1.0301	0.0657	0.9669	0.0576
	0.001	1.1300	0.3224	1.0205	0.1530	1.0853	0.0279	0.9923	0.0673
	0.1	1.0869	0.1510	1.0297	0.0836	1.1552	0.0579	1.0655	0.0694
	1	1.0475	0.0877	1.0231	0.0722	1.0508	0.0514	1.0206	0.0603
10000	0	1.0917	0.2431	0.9967	0.1012	0.9361	0.0473	0.9455	0.0541
	0.001	1.0909	0.2436	0.9845	0.0944	0.9492	0.0381	0.9579	0.0611
	0.1	1.0081	0.0786	0.9824	0.0581	1.0160	0.0396	0.9834	0.0472
	1	0.9970	0.0579	0.9849	0.0531	0.9766	0.0440	0.9765	0.0447
50000	0	1.0165	0.0550	0.9892	0.0374	0.9704	0.0210	0.9724	0.0269
	0.001	1.0377	0.1784	0.9956	0.0635	0.9694	0.0231	0.9735	0.0290
	0.1	1.0105	0.0391	0.9981	0.0402	0.9934	0.0249	0.9905	0.0276
	1	1.0107	0.0246	1.0051	0.0254	0.9922	0.0213	0.9981	0.0236

Note: 1. means are calculated among estimations that successfully found more than 1 points of support for unobserved heterogeneity. 2. var(month) is the variance of calendar month variation in DGP.

Second, there is strong evidence that given the sample size, increase of calendar variation would considerably increase the quality of estimation on the structure parameter. This is particularly the case for small samples. For instance, in the constant hazard model with 5,000 individual observations, the standard deviation for $\hat{\beta}$ from

[40] Since means for BIC are calculated from a handful estimations that return more than 1 point (referring to Table 3), we should not put too much weight on these results.

loglikelihood estimation reduces from 0.2491 when no calendar variation to 0.0746 when variance of calendar variation is 1. Similar observations can be found for other maximum penalised estimators.

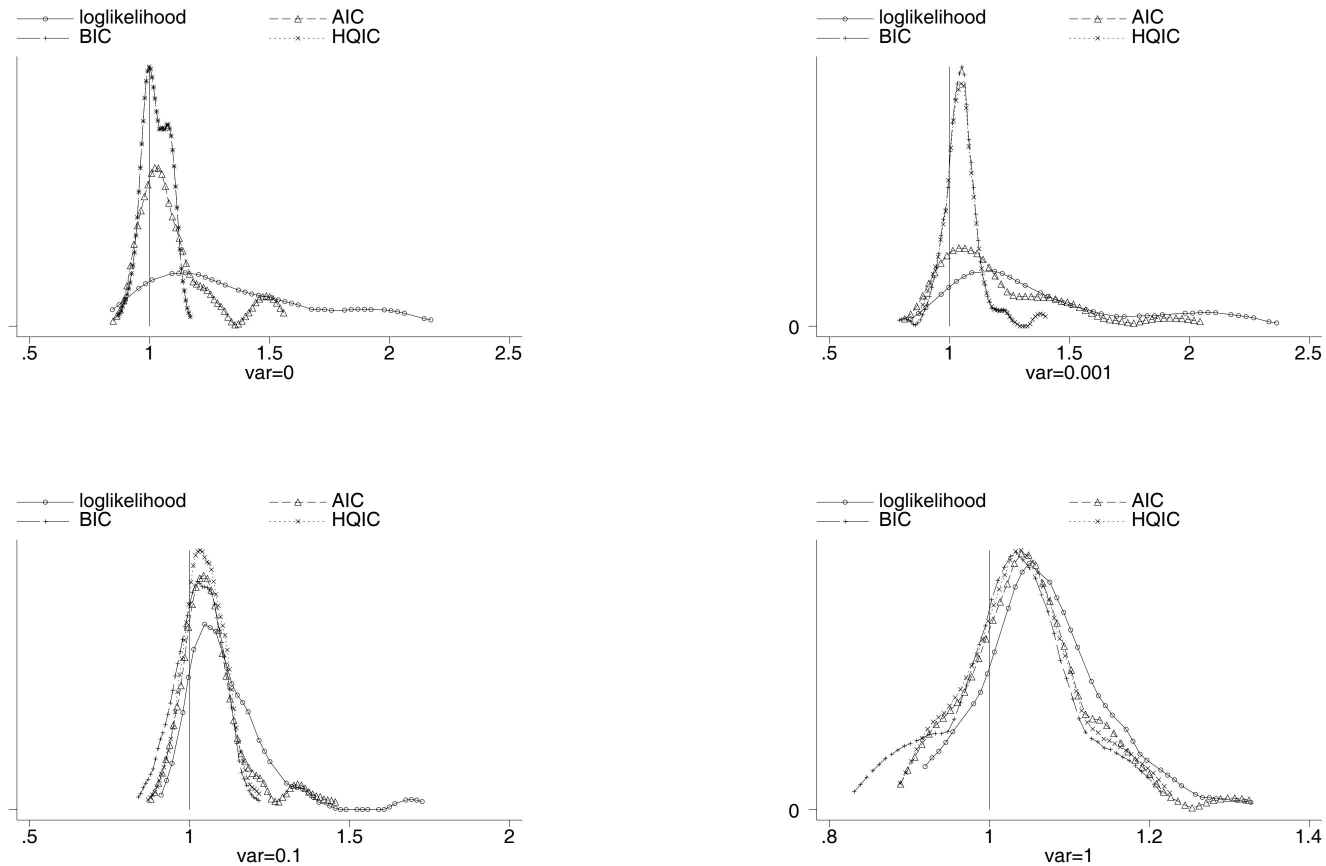
Third, sample size matters. Large sample size improves the identifiability of the model. This can be seen from increased accuracy of means and reduced standard deviations when sample size increases. For given calendar variation, the standard errors for loglikelihood estimators reduce in line with factor of $\sqrt{N}$.

To facilitate the presentation of our findings, we plot the kernel densities for estimated $\hat{\beta}$ for samples with 5,000 individual observations[41]. Figure 3 and 4 depict the kernel densities for $\hat{\beta}$, by calendar variations for maximum likelihood and maximum penalised likelihood estimators.

It is clear from the figures that when there are no or little calendar variations, the distribution of loglikelihood estimator (as well as AIC) has a wide dispersion and heavy tail, while BIC and HQIC are more concentrated around the true value 1. This confirms the finding above that maximum likelihood criterion can impose positive bias on estimation of structure parameter. But as calendar variation increases, it is more likely that estimators from both maximum likelihood and maximum penalised likelihood have the similar distributions.

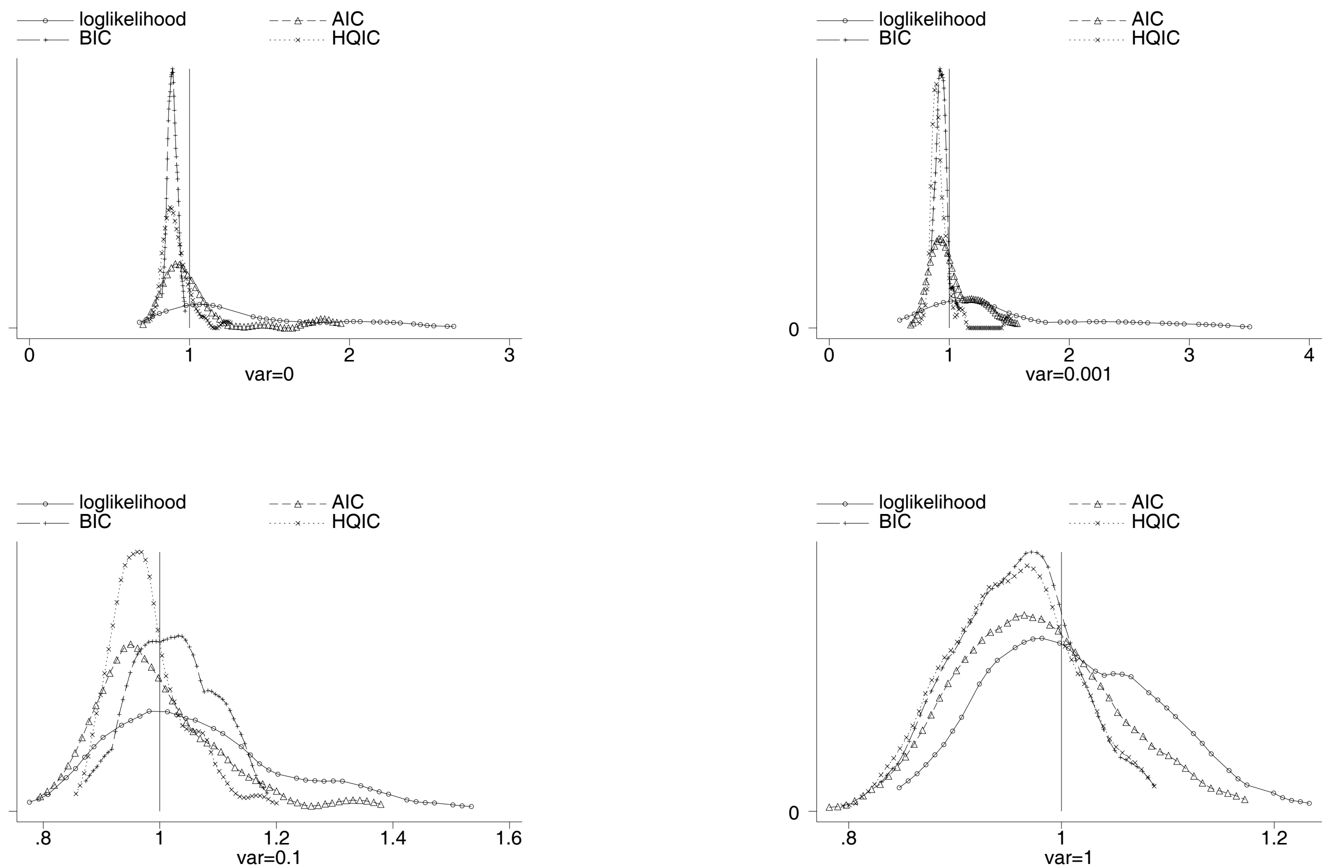
[41] Plots are estimates of *Epanevhnikov Kernel* densities on $\hat{\beta}$ across successful estimations that return more that one point of support. The densities are estimated with *STATA*. Bandwidth is estimated by $h=0.9m/(n^{1/5})$, where m=min(sqrt(variance($\hat{\beta}$), interquartilerange($\hat{\beta}$)). n is the number of values of $\hat{\beta}$ that we estimate kernel densities on. We use the default value n=50 for all kernel density estimations. See "Reference Manual, [R] kdensity" (2001), *Stata Statistical Software, Release 7.0*, StataCorp.

Figure 3: Kernel densities of estimated β. Constant hazard, 3 points mixture, 5,000 individuals.



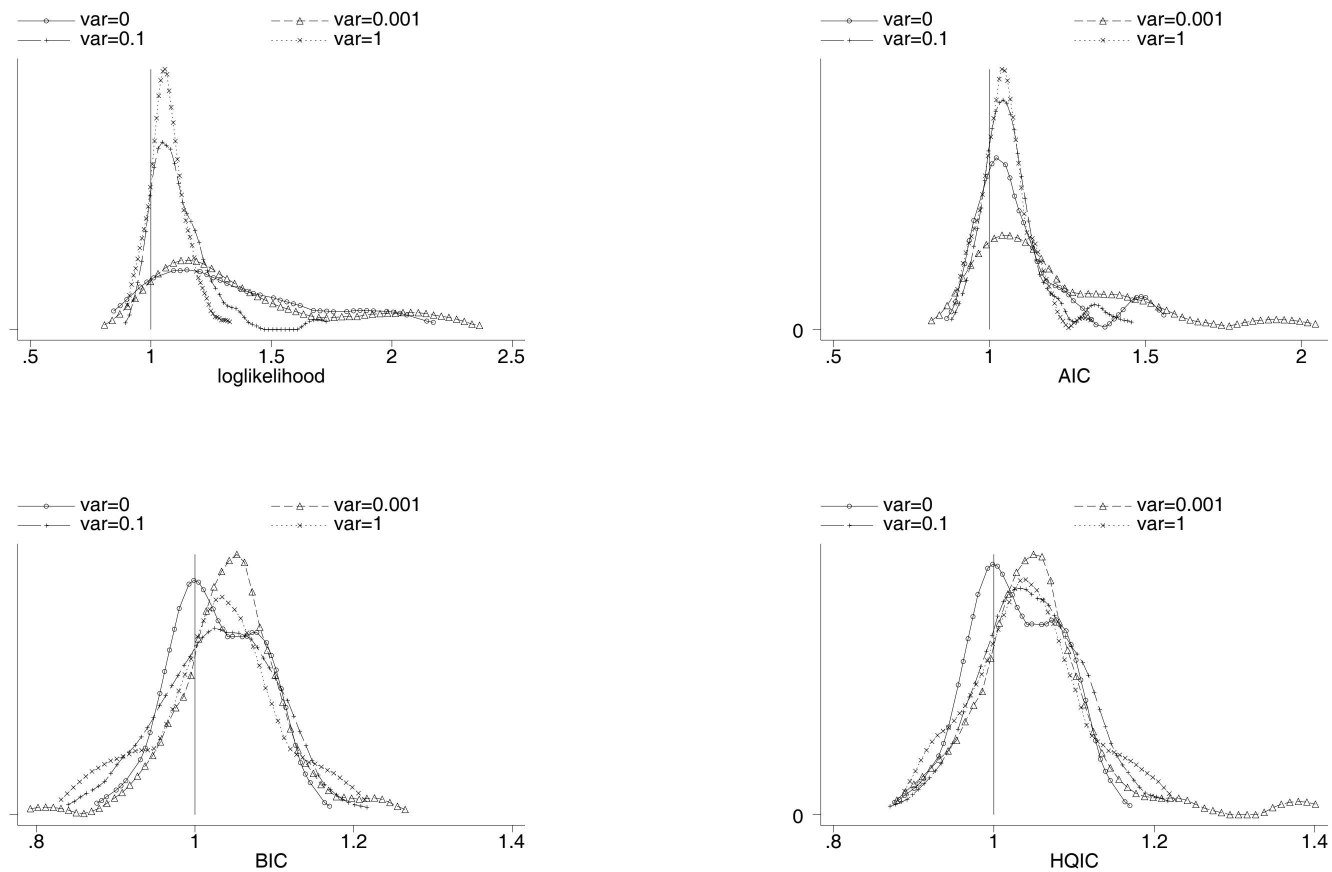
Constant hazard, 3 points, 5000 obs

Figure 4: Kernel densities of estimated β. Weibull hazard, Gamma mixture, 5,000 individuals.



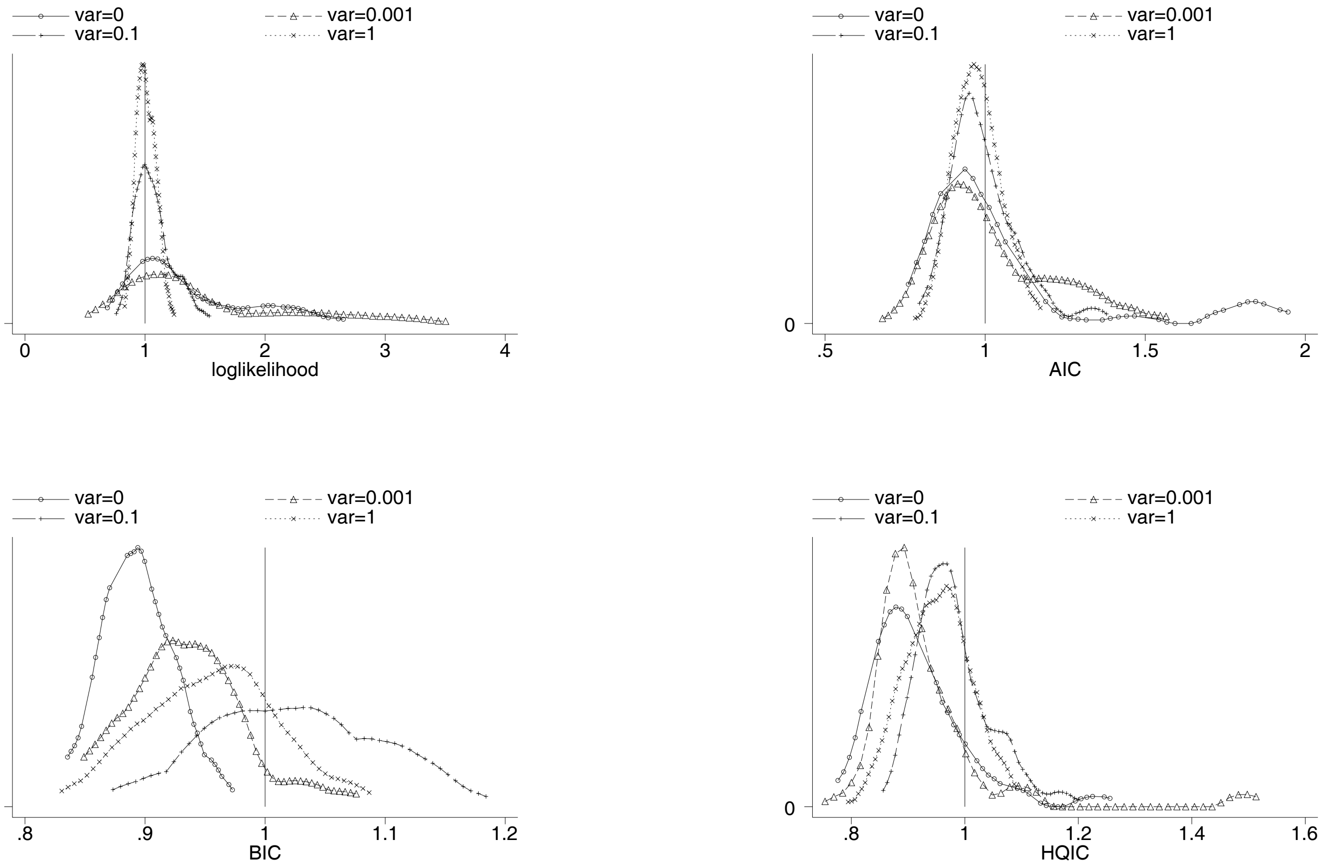
Weibull hazard, Gamma, 5000 obs

Figure 5: Kernel densities of estimated β by calendar variations. Constant hazard, 3 points mixture, 5,000 individuals.



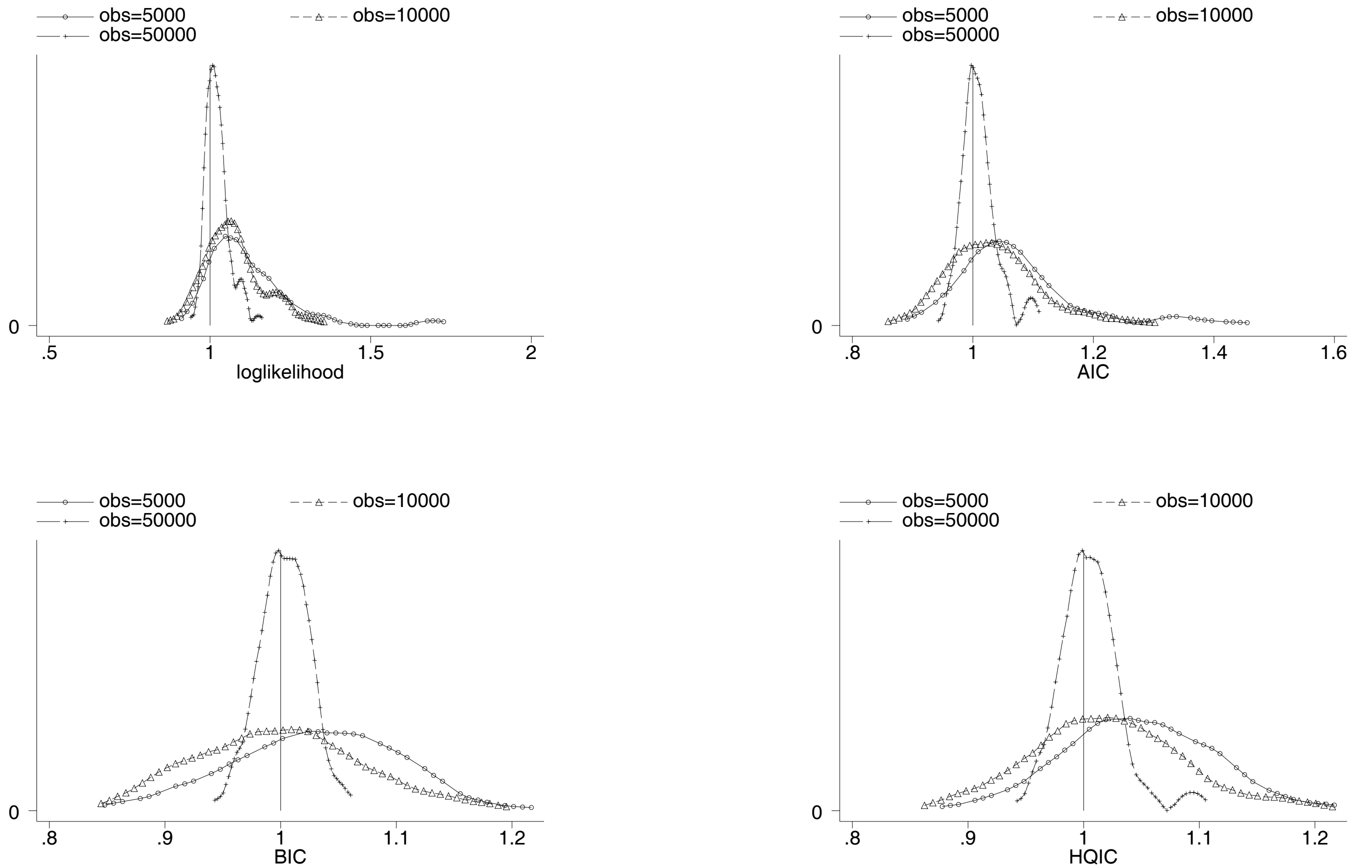
Constant hazard, 3 points, 5,000 obs

Figure 6: Kernel densities of estimated β by calendar variations. Weibull hazard, Gamma mixture, 5,000 individuals.



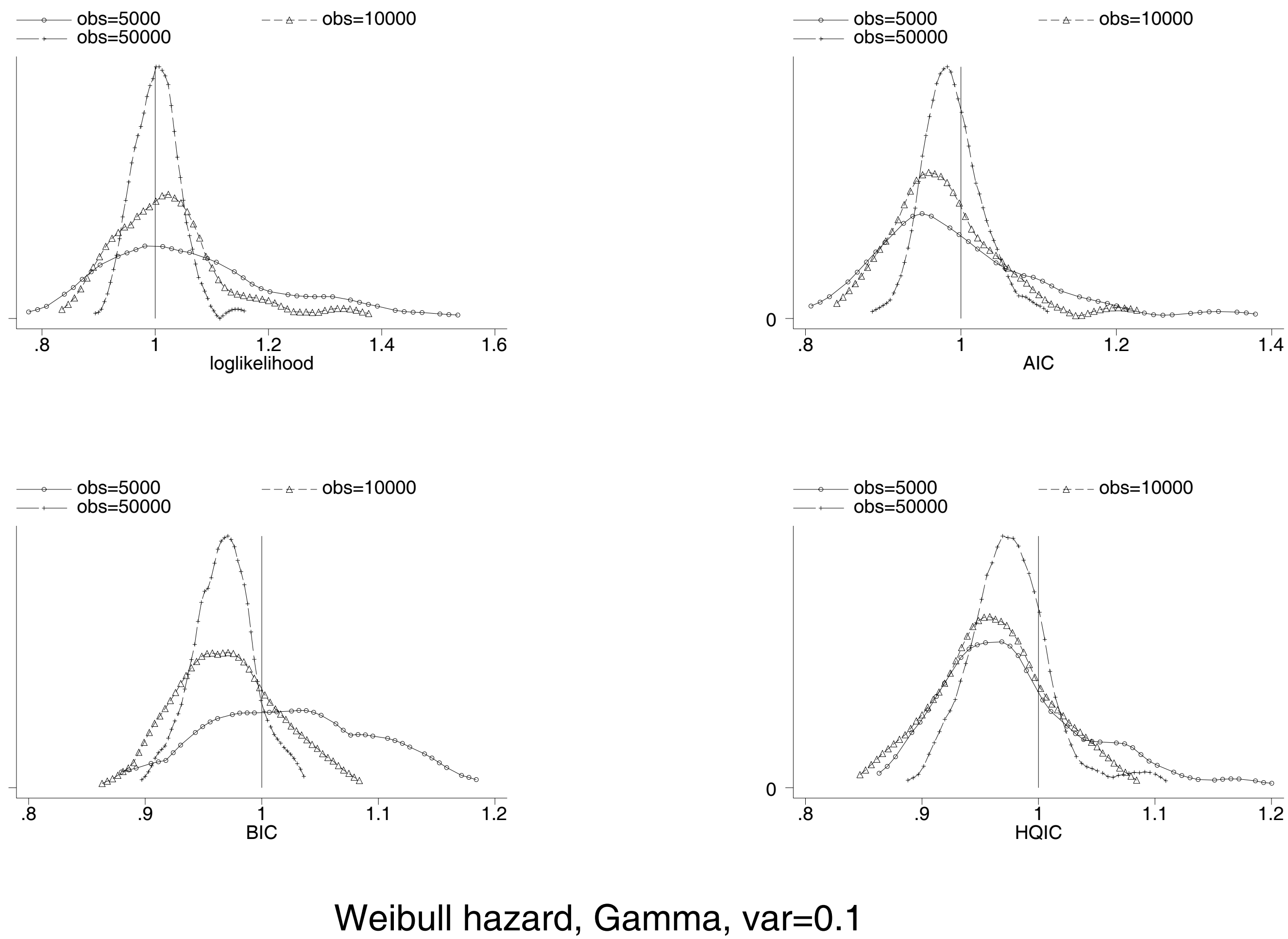
Weibull hazard, Gamma, 5,000 obs

Figure 7: Kernel densities of estimated β by sample sizes. Constant hazard, 3 points mixture, var(month)=0.1.



Constant hazard, 3 points, var=0.1

Figure 8: Kernel densities of estimated β by sample sizes. Weibull hazard, Gamma mixture, var(month)=0.1.



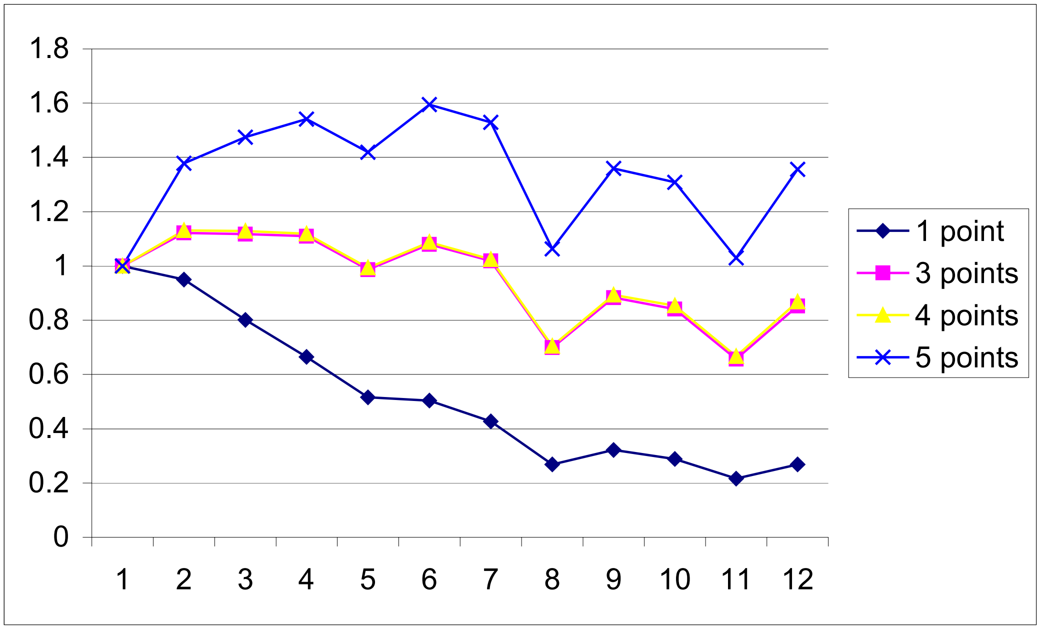
Figures 5 and 6 look into the effect of calendar variation on the estimated $\hat{\beta}$. One could come to the conclusion from figures that BIC and HQIC are less sensitive towards calendar variations than loglikelihood. At least for constant hazard model with 3 points distributed unobserved heterogeneity, kernel densities for BIC and HQIC do not vary much across calendar variations. On the other hand, maximum likelihood method seems to be much sensitive towards calendar variations. With large variance of calendar time, kernel density for loglikelihood estimator is more concentrated on the true value.

Figure 7 and 8 display kernel densities of $\hat{\beta}$ across different sample sizes, with fixed calendar variation being 0.1. Not surprisingly, the larger the sample is, the more concentrated the $\hat{\beta}$ around the true value. For large samples with 50,000 individuals, the distributions for estimated $\hat{\beta}$ have a familiar bell-shape. There is evidence that maximum loglikelihood and maximum penalised likelihood converge to each other.

3. Duration Dependence

In our non-parametric estimation settings, the duration baselines are represented by a set of 12 dummies. As the iteration processes indicate in Table 4-1 and 4-2, the estimators on $\hat{\beta}$ are sensitive with respect to how many points of support found for the unobserved heterogeneity distribution. As more points added to the support of mixing distribution, the estimators move away from zero. It turns out that the duration baseline hazard has the same response with respect to the points found for the mixing distribution. Since uncontrolled unobserved heterogeneity would produce negative duration dependence, it is intuitive that excessive control would produce positive duration dependence. This can be seen from Figure 9.

Figure 9: Duration baseline for constant hazard with 3 points mixing distribution in DGP, estimated by maximum likelihood, 5,000 individuals, var(month)=0.1.



Note: duration baselines are estimated from the same estimation in Table 4-1. Each line represents estimated baseline hazard with respective number of support points found for the unobserved heterogeneity distribution. All baselines are normalised to the first month.

In Figure 9, when only 1 point of support for the unobserved heterogeneity (no control for unobserved heterogeneity), the baseline displays a negative duration dependence. By referring to Table 4-1, we can see that the best estimator for $\hat{\beta}$ is found at 3 points (BIC and HQIC) or 4 points (AIC). The baseline associated with the best $\hat{\beta}$ estimator is almost flat, as seen in Figure 9. But the optimal number of points for support found by likelihood criterion is 5, which not only causes a positive bias on $\hat{\beta}$ (Table 4-1), but also a somewhat positive duration dependence for baseline hazard (Figure 9).

Figure 9 is just an illustration of the possible consequences of number of support points found by maximisation on the estimation of duration dependences. However, to assess overall performance of non-parametric estimation on duration dependence, we would need a single measure for overall biases on the estimators. We report in Table 6 average weighted squared errors for duration baseline estimators for the constant hazard model with 3 points unobserved heterogeneity, and the Weibull model with Gamma unobserved heterogeneity. Though this might be somewhat ad hoc, these average weighted squared errors do provide an intuitive overall measure of goodness of fit for the duration baseline estimates. The squared errors are calculated as squared differences

Table 6: Average Weighted Squared Errors for duration baseline estimators.

Constant hazard, 3 points unobserved heterogeneity

# obs	var(month)	Loglikelihood mean	Loglikelihood std.	AIC mean	AIC std.	BIC mean	BIC std.	HQIC mean	HQIC std.
5000	0	0.3715	0.8991	0.0817	0.2176	0.0216	0.0208	0.0216	0.0208
	0.001	0.5465	1.2508	0.1852	0.5174	0.0276	0.0392	0.0324	0.0606
	0.1	0.0685	0.2298	0.0403	0.1030	0.0288	0.0236	0.0218	0.0178
	1	0.0234	0.0365	0.0223	0.0356	0.0252	0.0223	0.0190	0.0160
10000	0	0.0623	0.1168	0.0161	0.0303	0.0098	0.0152	0.0107	0.0194
	0.001	0.2410	1.2267	0.0213	0.0602	0.0172	0.0214	0.0140	0.0137
	0.1	0.0615	0.1352	0.0352	0.1046	0.0201	0.0159	0.0170	0.0203
	1	0.0181	0.0219	0.0135	0.0175	0.0156	0.0144	0.0121	0.0114
50000	0	0.0149	0.0431	0.0094	0.0207	0.0041	0.0037	0.0059	0.0088
	0.001	0.0368	0.1483	0.0079	0.0211	0.0049	0.0044	0.0051	0.0047
	0.1	0.0070	0.0147	0.0039	0.0073	0.0027	0.0024	0.0035	0.0064
	1	0.0023	0.0020	0.0020	0.0016	0.0019	0.0013	0.0019	0.0013

Weibull hazard, Gamma distributed unobserved heterogeneity

# obs	var(month)	Loglikelihood mean	Loglikelihood std.	AIC mean	AIC std.	BIC mean	BIC std.	HQIC mean	HQIC std.
5000	0	0.2563	0.8084	0.0383	0.0643	0.0118	0.0061	0.0251	0.0160
	0.001	0.4398	1.5500	0.1177	0.5884	0.0186	0.0046	0.0287	0.0185
	0.1	0.1245	0.2989	0.0320	0.0829	0.0421	0.0245	0.0206	0.0182
	1	0.0297	0.0306	0.0197	0.0174	0.0136	0.0092	0.0144	0.0100
10000	0	0.3033	1.1682	0.0586	0.1130	0.0197	0.0104	0.0244	0.0204
	0.001	0.3056	0.8927	0.0468	0.0990	0.0193	0.0106	0.0272	0.0403
	0.1	0.0315	0.0382	0.0168	0.0178	0.0092	0.0070	0.0105	0.0089
	1	0.0126	0.0140	0.0108	0.0096	0.0079	0.0058	0.0086	0.0062
50000	0	0.0167	0.0613	0.0093	0.0093	0.0096	0.0070	0.0102	0.0077
	0.001	0.1203	0.9850	0.0216	0.0904	0.0099	0.0077	0.0105	0.0097
	0.1	0.0074	0.0283	0.0079	0.0274	0.0028	0.0032	0.0039	0.0047
	1	0.0024	0.0036	0.0027	0.0036	0.0028	0.0019	0.0026	0.0022

Note: 1. means are calculated among estimations that successfully found more than 1 point of support for unobserved heterogeneity. 2. var(month) is the variance of calendar time variation in DGP. 3. Weighted squared errors are calculated by $\frac{1}{w}(\hat{\lambda}-\lambda)^2$, where w is the weight that is inversely proportional to the estimated standard error for $\hat{\lambda}$.

between estimators and the true value in DGP. Each squared difference is weighted by the standard error of the estimator, such that larger standard error gives a smaller weight. The average is taken over trials that successfully return more than 1 point of support. Several observations can be made: Firstly, the average weighted squared errors are relatively small, for most models they are below 10%. We interpret this as a sign of relatively good fit. Secondly, there is also evidence that given the sample size, with increased calendar variations the averaged weighted errors for baseline estimators decrease considerably. This is particularly visible for log likelihood estimators. For example, for the small sample of 5,000, constant hazard model with discrete mixture, when no calendar variations present, the average weighted squared error for baseline is 0.3715. As the calendar variation being 1, the average weighted squared error is reduced to 0.0234. Similar pattern can be observed for Weibull model with Gamma mixture as well. Thirdly, large sample sizes increase the estimation precision by reducing the average weighted squared errors, as expected.

In Appendix Figures A3-A5, we provide some plots of confidence intervals associated with the estimated baselines. They are just some illustrative figures from the same results that produce Table 4-1, 4-2, A2-1 A2-2. They give some informative views on how the estimation of duration dependences are affected by sample sizes and calendar variations embedded in the data.

4. Unobserved heterogeneity

Recall that the model is estimated with a constant term, therefore when there is no control for the unobserved heterogeneity, the constant represents first month hazard rate for an individual with $x=0$ in the reference calendar month. The unobserved model term μ is an additive term to the constant such that in estimation, the estimated constant is the sum of μ and parameter for a representative individual's hazard rate. In our simulation, we predetermined the constant to be $log(0.1259)=-2.07233$ (see above) and rescale the unobserved heterogeneity term accordingly. But in reality, this constant is never known. Therefore all estimated discrete points in models are sum of both genuine constants and the chosen (log) points of support.

In non-parametric specification of the mixing distribution of unobserved heterogeneity, we evidently approximate an unknown distribution with a set of discrete mass points. We find it natural in our case to compare estimated moments to those in the true DGP to assess the quality of identification of the mixing distribution. For the convenience of interpretation, from estimators for points and their associated probabilities, we calculated (in exponential form) first and second moments[42]. These facilitate the comparison with the true moments used in DGP.

Table 7: Estimated means and standard errors of the first moment (expectation) of the unobserved heterogeneity distribution, exponential form.

Constant hazard, 3 points unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	var(month)	mean	std.	mean	std.	mean	std.	mean	std.
5000	0	0.1283	0.0170	0.1239	0.0084	0.1241	0.0075	0.1241	0.0075
	0.001	0.1326	0.0248	0.1272	0.0182	0.1257	0.0141	0.1258	0.0140
	0.1	0.1251	0.0151	0.1235	0.0130	0.1237	0.0133	0.1231	0.0126
	1	0.1230	0.0145	0.1201	0.0125	0.1216	0.0126	0.1205	0.0120
10000	0	0.1260	0.0105	0.1235	0.0062	0.1245	0.0055	0.1243	0.0053
	0.001	0.1365	0.0919	0.1242	0.0137	0.1240	0.0111	0.1237	0.0110
	0.1	0.1270	0.0128	0.1246	0.0096	0.1251	0.0091	0.1246	0.0091
	1	0.1337	0.0517	0.1254	0.0145	0.1251	0.0092	0.1244	0.0090
50000	0	0.1259	0.0065	0.1251	0.0034	0.1253	0.0028	0.1252	0.0029
	0.001	0.1270	0.0060	0.1261	0.0043	0.1261	0.0039	0.1260	0.0040
	0.1	0.1258	0.0049	0.1257	0.0046	0.1256	0.0042	0.1256	0.0042
	1	0.1261	0.0052	0.1257	0.0043	0.1257	0.0041	0.1257	0.0041

Weibull hazard, Gamma distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	var(month)	mean	std.	mean	std.	mean	std.	mean	std.
5000	0	0.1292	0.0198	0.1266	0.0144	0.1125	0.0092	0.1258	0.0073
	0.001	0.1334	0.0298	0.1273	0.0154	0.1242	0.0245	0.1262	0.0127
	0.1	0.1377	0.0508	0.1271	0.0151	0.1157	0.0044	0.1217	0.0097
	1	0.1305	0.0268	0.1244	0.0122	0.1231	0.0115	0.1249	0.0119
10000	0	0.1441	0.0796	0.1294	0.0086	0.1299	0.0057	0.1293	0.0069
	0.001	0.1343	0.0196	0.1287	0.0115	0.1294	0.0088	0.1282	0.0096
	0.1	0.1339	0.0251	0.1288	0.0101	0.1238	0.0080	0.1281	0.0089
	1	0.1312	0.0103	0.1287	0.0078	0.1301	0.0079	0.1293	0.0075
50000	0	0.1262	0.0054	0.1252	0.0033	0.1250	0.0027	0.1251	0.0027
	0.001	0.1262	0.0084	0.1247	0.0053	0.1241	0.0047	0.1239	0.0046
	0.1	0.1263	0.0067	0.1263	0.0097	0.1251	0.0043	0.1249	0.0043
	1	0.1265	0.0052	0.1256	0.0045	0.1257	0.0040	0.1250	0.0041

Note: 1. means are calculated among estimations that successfully found more than 1 point of support for unobserved heterogeneity. 2. var(month) is the variance of calendar month variation in DGP. 3. the true expectation in DGP is 0.125893.

[42] Recall that in MPH model (equation 2), v is the term for unobserved heterogeneity, and $E(v)=1$, $var(v)=0.6475$ (Table 1). Define $y=log(v)+c$, where c is the genuine constant term (-2.07233). y is then the point of support we acquire from estimation. In DGP, the first moment for y is (in exponential form) simply $E(\exp(y)) = E(\exp(\log(v)+c)) = E(v)\exp(c) = 0.1259$; the second moment of y is then

$$E(\exp(y)^2) = \mathrm{var}(\exp(y)) + (E(\exp(y)))^2 = (\exp(c))^2 \left[\mathrm{var}(v)+1\right] = 0.0261$$

Table 7 provides the summarised results for the first moment of distribution of the unobserved heterogeneity for selected models. We find the high agreement between the estimated means and the true value in the DGP. In quite a few cases, the differences for expectations are less than 0.01. For the simulated parametric Gamma distributions of unobserved heterogeneity, the estimators acquired using by pure loglikelihood approach seem to be a little upwards biased. This is again probably due to the fact that loglikelihood finds more points for the support of the unobserved heterogeneity. There is not much difference with respect to which information criterion is used. The second moments are also well estimated as showed in Table 8. Except a few cases with loglikelihood, all estimations return the estimated second moments that are very close to the true one in DGP, with very good precision in terms of standard errors.

Variation of calendar dummies does not seem to have strong impact on estimations of the unobserved heterogeneity. There is no firm relationship between the variation of calendar covariates and estimation quality from Table 7 and 8. But for the second moment, when sample sizes are limited, large dispersions for this estimator have been seen from loglikelihood estimators. When sample sizes are sufficiently large (50,000), all model selection criteria return reasonably good first and second moments estimators for the unobserved heterogeneity.

To further assess the properties of non-parametric estimators on unobserved heterogeneity, we also provide plots for the kernel densities of estimated means (first moment) of unobserved heterogeneity distribution in appendix. Figures A6 and A7 display the kernel densities of estimated means from various model selection criteria for sample size of 5,000. It is clear from the figures that we find again loglikelihood estimators have a large dispersion of distribution and long tail in the distribution. This is in accordance with the finding in earlier section. Figure A8 and A9 in appendix depict the effects of calendar variations on the estimated first moment of unobserved heterogeneity. Contrary to kernel densities for structure parameter estimators, it seems that the less calendar variation, the more concentrated the density on estimated means of unobserved heterogeneity is. As we plot kernel densities of estimated means for unobserved heterogeneities across sample sizes in Figures A10 and A11, we find large sample sizes do increase the precision of estimators. The distribution of estimated means is more concentrated on the true value in DGP when sample size is 50,000, at least for discrete generated unobserved heterogeneity.

Table 8: Estimated means and standard errors of the second moment of the unobserved heterogeneity distribution.

Constant hazard, 3 points unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	0.0427	0.0364	0.0270	0.0098	0.0243	0.0032	0.0243	0.0032
	0.001	0.0534	0.0564	0.0337	0.0235	0.0259	0.0064	0.0261	0.0064
	0.1	0.0326	0.0228	0.0261	0.0085	0.0257	0.0061	0.0245	0.0050
	1	0.0319	0.0251	0.0236	0.0067	0.0247	0.0061	0.0234	0.0047
10000	0	0.0339	0.0235	0.0257	0.0027	0.0260	0.0026	0.0259	0.0026
	0.001	0.1577	1.1199	0.0277	0.0160	0.0261	0.0054	0.0259	0.0053
	0.1	0.0363	0.0238	0.0279	0.0088	0.0271	0.0043	0.0264	0.0053
	1	0.1986	1.1431	0.0313	0.0445	0.0268	0.0045	0.0260	0.0040
50000	0	0.0302	0.0173	0.0266	0.0033	0.0251	0.0012	0.0254	0.0017
	0.001	0.0308	0.0150	0.0264	0.0037	0.0254	0.0016	0.0254	0.0016
	0.1	0.0275	0.0050	0.0261	0.0036	0.0255	0.0018	0.0256	0.0018
	1	0.0276	0.0066	0.0260	0.0023	0.0256	0.0018	0.0256	0.0018

Weibull hazard, Gamma distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	0.0410	0.0558	0.0257	0.0335	0.0198	0.0044	0.0225	0.0034
	0.001	0.0504	0.0903	0.0266	0.0224	0.0247	0.0091	0.0236	0.0056
	0.1	0.1381	0.7704	0.0292	0.0301	0.0245	0.0018	0.0246	0.0042
	1	0.0636	0.2013	0.0249	0.0078	0.0257	0.0049	0.0249	0.0052
10000	0	0.5237	4.3580	0.0275	0.0143	0.0254	0.0027	0.0233	0.0034
	0.001	0.0465	0.0593	0.0258	0.0137	0.0253	0.0040	0.0232	0.0048
	0.1	0.0601	0.1904	0.0260	0.0101	0.0256	0.0038	0.0256	0.0037
	1	0.0307	0.0129	0.0259	0.0065	0.0270	0.0033	0.0261	0.0035
50000	0	0.0278	0.0122	0.0235	0.0047	0.0214	0.0013	0.0216	0.0021
	0.001	0.0302	0.0217	0.0243	0.0095	0.0211	0.0020	0.0213	0.0028
	0.1	0.0290	0.0212	0.0310	0.0543	0.0248	0.0017	0.0234	0.0021
	1	0.0273	0.0053	0.0250	0.0035	0.0247	0.0022	0.0236	0.0024

Note: 1. means are calculated among estimations that successfully found more than 1 point of support for unobserved heterogeneity. 2. var(month) is the variance of calendar month variation in DGP. 3. the true second moment in DGP is (rescaled) 0.026111.

5. Calendar variations

The calendar variations enter the hazard rate models as the time-varying covariates, and in our estimations, they are modelled by a set of dummies with reference to month 13. We also present average weighted squared errors as those for duration baseline estimates as a measure for estimation quality. Table 9 displays the results from trials that find more than 1 point of support for the mixing distribution, for constant hazard with 3 points mixture and Weibull hazard with Gamma mixture. The average weighted errors are small and in most cases below 0.05. We interpret this as evidence for a good fit. Also note that the errors using information criteria estimators are considerably smaller than that from using pure maximum loglikelihood. This is especially the case for small samples. Variations of calendar time covariates certainly contribute the accuracy of estimators. When sample sizes increase, all estimators have little or negligible average weighted squared errors.

Table 9: Average Weighted Squared Errors for calendar variation estimators.

Constant hazard, 3 points unobserved heterogeneity

		Loglikelihood		**AIC**		**BIC**		**HQIC**	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	-	-	-	-	-	-	-	-
	0.001	0.5465	1.2508	0.0168	0.0102	0.0159	0.0100	0.0161	0.0101
	0.1	0.0685	0.2298	0.0174	0.0098	0.0173	0.0089	0.0171	0.0087
	1	0.0234	0.0365	0.0225	0.0121	0.0226	0.0121	0.0223	0.0120
10000	0	-	-	-	-	-	-	-	-
	0.001	0.2410	1.2267	0.0101	0.0081	0.0100	0.0080	0.0100	0.0080
	0.1	0.0615	0.1352	0.0085	0.0060	0.0086	0.0062	0.0085	0.0061
	1	0.0181	0.0219	0.0099	0.0048	0.0099	0.0047	0.0098	0.0047
50000	0	-	-	-	-	-	-	-	-
	0.001	0.0368	0.1483	0.0016	0.0012	0.0015	0.0012	0.0015	0.0012
	0.1	0.0070	0.0147	0.0017	0.0009	0.0017	0.0009	0.0017	0.0009
	1	0.0023	0.0020	0.0021	0.0009	0.0021	0.0009	0.0021	0.0009

Weibull hazard, Gamma distributed unobserved heterogeneity

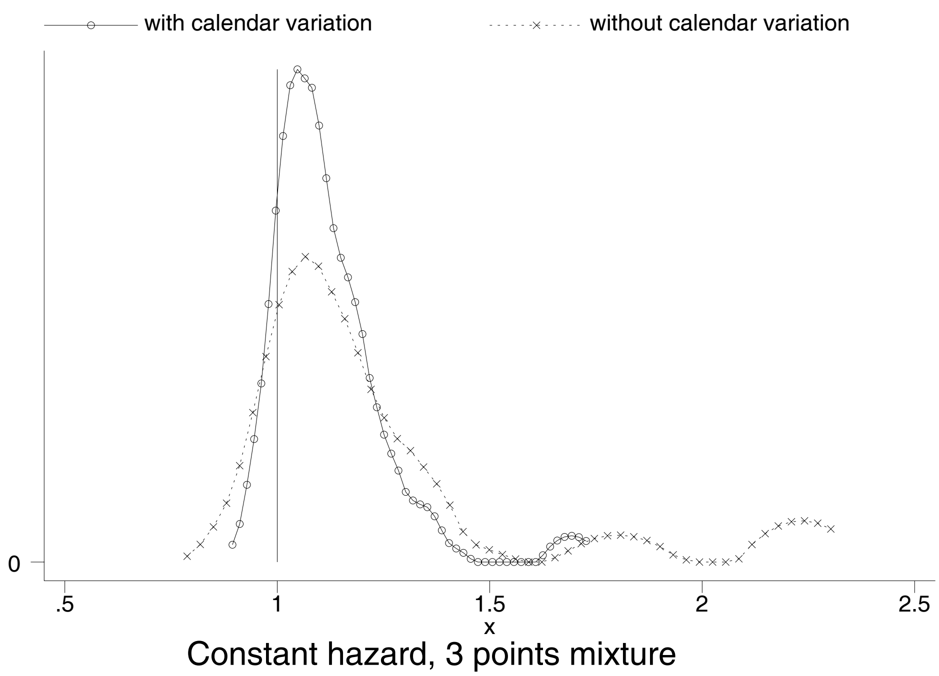
		Loglikelihood		**AIC**		**BIC**		**HQIC**	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	-	-	-	-	-	-	-	-
	0.001	0.4398	1.5500	0.0151	0.0116	0.0231	0.0078	0.0156	0.0117
	0.1	0.1245	0.2989	0.0144	0.0063	0.0129	0.0039	0.0135	0.0049
	1	0.0297	0.0306	0.0193	0.0092	0.0191	0.0090	0.0191	0.0092
10000	0	-	-	-	-	-	-	-	-
	0.001	0.3056	0.8927	0.0071	0.0043	0.0066	0.0040	0.0070	0.0043
	0.1	0.0315	0.0382	0.0074	0.0034	0.0080	0.0044	0.0075	0.0035
	1	0.0126	0.0140	0.0102	0.0045	0.0102	0.0046	0.0103	0.0045
50000	0	-	-	-	-	-	-	-	-
	0.001	0.1203	0.9850	0.0016	0.0012	0.0015	0.0012	0.0015	0.0012
	0.1	0.0074	0.0283	0.0016	0.0009	0.0016	0.0009	0.0016	0.0008
	1	0.0024	0.0036	0.0019	0.0008	0.0019	0.0008	0.0019	0.0008

Note: 1. means are calculated among estimations that successfully found more than 1 point of support for unobserved heterogeneity. 2. var(month) is the variance of calendar time variation in DGP. 3. Weighted squared errors are calculated by $\frac{1}{w}(\hat{\lambda}-\lambda)^2$, where w is the weight that is inversely proportional to the estimated standard error for $\hat{\lambda}$.

The estimated hazard rates for each calendar time dummy conditional on observed covariates and unobserved heterogeneity have also particular empirical interpretations.

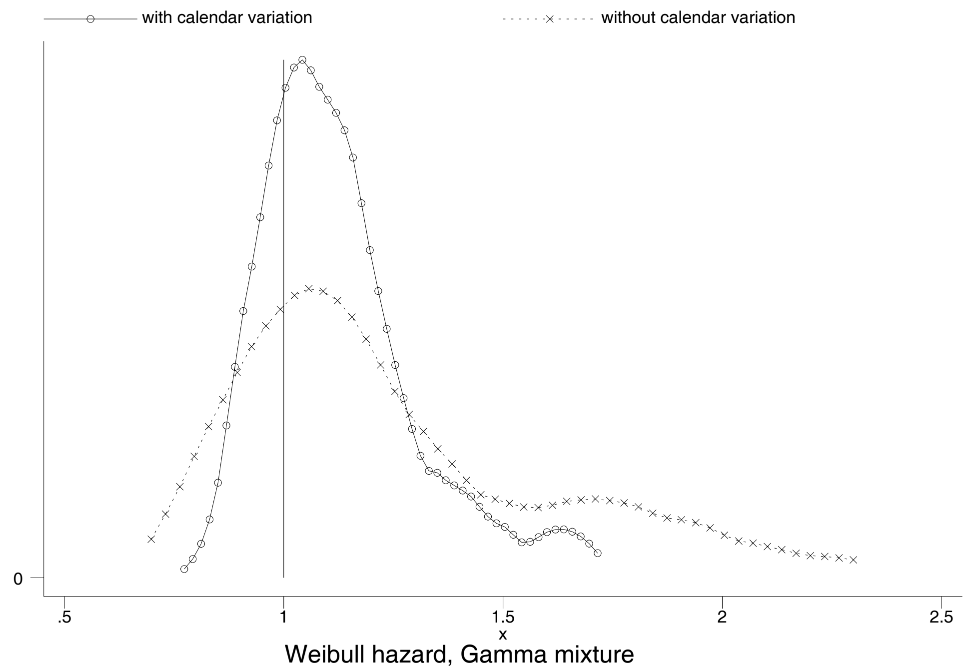
These estimated monthly hazard rates characterise the business and seasonal cycle conditions, as Gaure and Røed (2003) point out. As suggested earlier, the information on labour market conditions during the elapsed time of active spells would contribute to the identification of unobserved personal characteristics. Consequently, ignoring such information would probably result in ineffective control for bias on the structural parameters due to unobserved heterogeneity.

Figure 10-1: Kernel densities of estimated β with and without calendar variations. Constant hazard, 3 points mixture, 5,000 individuals, var(month)=0.1 in DGP.



To see the consequences of ignoring calendar time variations, we plot the kernel densities of estimated β from maximum loglikelihood estimations with and without covariates of calendar variations in Figures 10-1 and 10-2. As clearly seen from the figures, distributions of estimated β when the calendar variations are ignored have a wider dispersion and heavier tails. With calendar variations, estimators are more concentrated on the true value 1. This pattern is seen for both constant hazard and Weibull hazard models, which we regard as additional evidence for our proposition on the roll of calendar time variations in control of unobserved heterogeneity.

Figure 10-2: Kernel densities of estimated β with and without calendar variations. Weibull hazard, Gamma mixture, 5,000 individuals, var(month)=0.1 in DGP.



5. Discussions

Our findings from the Monte Carlo studies on the non-parametric estimation of single risk duration models with unobserved heterogeneity can be summarised as following: Firstly, the mixed proportional hazard rate model can be reasonably well estimated with non-parametric specifications on both duration dependence and distribution of unobserved heterogeneity. In most of the model estimations, the recovery of the true model parameters is rather satisfactory. This can be viewed from e.g. Table 5[43]. Secondly, there is evidence that inclusion of time-varying covariates, e.g. in the form of calendar time variations can considerably increase the identifiability of the model components. We have seen that inclusion of large calendar variations has contributed to the estimations on both duration dependence and the structural parameter. Thirdly,

[43] In Appendix tables A4-A8, we provide results for constant hazard model with Gamma distributed unobserved heterogeneity, as well as Weibull hazard model with 3 points discrete mixing distribution. We report statistics for estimated structural parameter, as well as estimated first and second moments for the unobserved heterogeneity mixing distributions. They show the same patterns as we have seen in Table 5, 7 and 8. Our conclusions are thus robust with respect to mixed proportional hazard rate models with different combinations of duration dependences and unobserved heterogeneity distributions.

when sample sizes are small, it is observed that pure maximum likelihood estimators tend to overestimate the absolute sizes of the structure parameters as well as the dispersions of the distribution of unobserved heterogeneity. It is sensible in this case to adopt some form of information criteria to penalise excessive points found by likelihood. We find some evidence in favour of Akaike's Information Criterion, but in some cases the Bayesian Information Criterion and the Hannan-Quinn Information Criterion seem to perform better. When sample sizes are sufficiently large, maximum likelihood and maximum penalised likelihood tends to converge to each other.

Our results show that in non-parametric estimation of the hazard rate model, the number of support points included in the unobserved heterogeneity distribution seems to have a substantial impact on the estimators of other model components. Less points means failing to sufficiently control for the unobserved heterogeneity; on the other hand more points than "necessary" means an excessive control, which we have showed could produce disturbance on the estimation of the structural model parameters as well. Thus the key task in such non-parametric estimation is to find the optimal number of support points for the unknown mixing distribution so that the impact of uncontrolled unobserved heterogeneity on other model components can be eliminated as much as possible.

The fact that pure maximum likelihood tends to find excessive points of support of the mixing unobserved heterogeneity distribution might be an indication of the flatness of the loglikelihood function around the potential maximum. In some cases, it seems that even though the change for likelihood function value from iteration to iteration is minimal, there is still room for an extra point with extremely small probability to barely increase the likelihood function value. These points presumably lie at the tails of distribution and do not provide significant information in identification of the distribution. Nevertheless, such excessive points have showed to produce distortions on estimation of other model components. Since uncontrolled unobserved heterogeneity bias the duration dependence downwards and structure parameters towards zero, it is not surprising that excessive control would bias the estimators away from zero. Our results above show that at least for small samples, it seems to be the case that maximum likelihood has the tendency to produce such positive bias on the structure parameters. This is in accordance with the findings of Baker and Melino (2000). They find that Heckman and Singer's non-parametric maximum likelihood approach produces quite large bias on estimators of structure term, which not only diminishes very gradually with sample sizes, but also has the direction away from zero. This positive bias seems less dramatic in our case.

Maximum penalised likelihood operates as a safeguard against excessive control on unobserved heterogeneity. Given the maximum fit of the data, the choice of pure maximum likelihood versus maximum penalised likelihood is essentially to find a balance point between maximal overall fit and reliable recovery of duration dependence and structure parameters. Our results have showed that for small samples, it is of particular importance to control the estimations with information criteria such as AIC, BIC and HQIC. Baker and Melino (2000) find that HQIC performs well, and BIC is virtually not different from HQIC. We find that BIC and HQIC seem to put too much weight on the sample sizes and have the restrictive tendency for allowing an additional point. In our cases, AIC seems to be a balanced choice between pure maximum likelihood and BIC and HQIC. Our finding confirms the suggestion of Huh and Sickles (1994) that the maximum likelihood estimators and maximum penalised estimators converge to each other when sample size is large.

The utilisation of time-varying calendar variation shows to be a novel approach in facilitating the identification of unobserved heterogeneity. Although the mixed proportional hazard rate model is identified even without the time-varying calendar variation, the inclusion of such calendar variations has showed to increase the identifiability of structural model parameters. The unobserved heterogeneity represents in the traditional econometric sense the omitted regressors. Without taking account of the calendar time when the spell starts and undergoes, the calendar variations are implicitly included in the unobserved heterogeneity terms as omitted regressors. This implies further that by modelling explicitly the calendar time variations, we in effect have controlled a substantial part of the unobserved heterogeneity, and the larger the calendar variations are, the less the uncontrolled population heterogeneity is. Our Monte Carlo results have showed the improvement of the estimations through calendar time variations.

We characterise the calendar time variation as a data-based identification source. The potential of such data based identification has not yet seen many applications. This is of particular empirical relevance, because in applied research, the calendar variation is easily acquirable. Our results indicate that when data quantity is small, it is less informative for a precise estimation of model components solely based on the proportional assumption. Therefore inclusion of calendar variation as an additional source for identification of unobserved heterogeneity can be helpful for empirical inference based on duration data.

Another implication of our results on approximation of unknown distribution of unobserved heterogeneity can be thought of as following: since the unobserved heterogeneity is a nuisance parameter, it is of less importance that the exact number and location of mass points can be retrieved. Rather, the emphasis should lie on the correct control for this nuisance parameter's distribution so that bias on other parameters of interests can be minimised. Heckman and Singer (1984, pp. 309) have argued that "… Imposing a false, but very flexible, mixing distribution may not cause much bias in the estimates of the structural coefficients." In our models, e.g. when loglikelihood gives a 5 mass points finite distribution which involves 9 independent parameters to characterise the mixture, it should provide a more accurate approximation than the usual 2 parameters parametric distributions such as Gamma. Our results show that at least the first and second moments of unknown mixing distribution can be well estimated by non-parametric maximum likelihood. This we believe has more relevance than estimators of mass point location and probabilities themselves. Some previous empirical attempts in estimation of hazard rate model with non-parametric specification of unobserved heterogeneity (e.g. Richardson and van den Berg(2002), Lalive et al (2002)) typically assume a two-points mass points distribution with associated probabilities and estimate the model taking these two points as parameters. They also interpret estimators of these two points as values of two types of individuals that differ in e.g. productivity. Our finding suggests that it is generally not sensible to interpret any estimates as such, not to mention that the two points "parametric assumption" may not produce sufficient approximation for the true mixture. Instead, we suggest that the most important objective of minimising the spurious duration dependence and biases on structure parameters of real interests can be achieved by using a non-parametric approach.

We have showed that our main model components of mixed proportional hazard model: the duration dependence, the structure parameter and the mixing distribution of unobserved heterogeneity, as well as the time-varying calendar variations, can, in most of the cases, be estimated with negligible bias. Since asymptotic properties of the non-parametric maximum likelihood estimators are unclear, it is difficult to apply known statistic tests for overall performances of our models. Therefore we choose a somewhat direct approach to assess the overall performance of non-parametric estimation. This is simply done by in-sample and out-of-sample prediction of distributions for spell duration. To be concise of presentation, we only report results for Weibull hazard with Gamma distributed unobserved heterogeneity, and the calendar variation is fixed to be 0.1. Sample size is fixed at 5,000 individuals.

The in-sample prediction is done in the following way: For each sample used in estimation, we only keep the distribution of X and calendar time when each spell starts. We then for each individual calculate the predicted spell duration according to equation (5*), but using estimated parameters (from the estimation on this sample) instead of true parameters used in DGP. Both maximum likelihood estimators and maximum penalised likelihood estimators for baseline hazard rates, structure parameter β, and estimators on calendar variations are used (for each estimation, we have acquired 4 sets of estimators). The unobserved heterogeneities are simulated from the estimated distribution. For 100 samples of estimation data, we thus acquire 400 new predicted samples. The out-of-sample prediction is done similarly: we first simulate a fresh set of 100 samples using the same DGP as before. Then by only keeping the distribution of X and start calendar time of each spell, using the same estimators as we acquired from previous estimations and used in in-sample predictions, we have made predictions of spell durations the same way as in in-sample prediction. We have then 100 fresh samples and 400 samples from the out-of-sample prediction.

Table 10 reports the cumulative distributions of durations from the estimation data, and from the in-sample and the out-of-sample predictions. We observe that the cumulative distributions of spell duration from resimulated data using maximum likelihood estimators fits the original data very well. The same is also true for the predictions made from maximum penalised likelihood estimators. The cumulative frequencies are virtually the same for both maximum likelihood and maximum penalised likelihood. Even for fresh sample prediction, the agreement is very high. We regard this as a strong evidence of overall goodness of fit.

Table 10: Cumulative frequencies of spell lengths for fitted Weibull hazard model with Gamma distributed unobserved heterogeneity. Obs=5,000, var(month)=0.1.

Duration	Estimation data	Loglikelihood	AIC	BIC	HQIC
1	19.43	20.57	19.35	19.53	19.37
2	32.47	33.72	32.36	32.53	32.38
3	42.25	43.43	42.24	42.42	42.23
4	50.13	51.17	50.09	50.23	50.13
5	56.66	57.58	56.69	56.74	56.67
6	62.22	63.04	62.25	62.32	62.21
7	67.05	67.76	67.15	67.20	67.07
8	71.18	71.80	71.31	71.37	71.22
9	74.61	75.12	74.71	74.73	74.58
10	77.74	78.13	77.80	77.83	77.64
11	80.53	80.88	80.57	80.65	80.44
12	100	100	100	100	100

Duration	Fresh data	loglikelihood	AIC	BIC	HQIC
1	19.39	20.58	19.35	19.40	19.44
2	32.40	33.76	32.41	32.46	32.47
3	42.22	43.50	42.27	42.29	42.26
4	50.03	51.23	50.21	50.13	50.14
5	56.58	57.64	56.72	56.61	56.64
6	62.19	63.07	62.31	62.26	62.25
7	67.10	67.80	67.24	67.17	67.17
8	71.19	71.81	71.33	71.31	71.31
9	74.64	75.14	74.70	74.68	74.65
10	77.76	78.17	77.78	77.79	77.75
11	80.54	80.87	80.59	80.58	80.57
12	100	100	100	100	100

Note: 1. duration is measured in month. 2. numbers are cumulative percentage of frequencies. 3. numbers in first panel are calculated based on estimators acquired from diverse model selection criteria, for all estimation samples. 4. numbers in second panel are calculated based on fresh-generated samples.

6. Competing risks model

We now briefly turn our attention to the more complex competing risks model. Identification of duration baselines and unobserved heterogeneity has proven to be more challenging in competing risks models. In this section, we extend our model specification for single risk model of mixed proportional hazard rate for grouped hazard to a two-state competing risks model and apply the non-parametric specification for both duration dependence and unobserved heterogeneity.

Identification of the competing risks model has also been a focal point in hazard rate model literature, for example Heckman and Honoré (1989), McCall (1997) and Abbring and van den Berg (2003), to name a few. If the unobserved heterogeneity terms involved in the e.g. two competing transitions are independent, it is straightforward to estimate the competing risks model as two independent single risk models, provided that the issue of discrete durations is disregarded. However in general, there is no justification that these two competing risks are independent. Therefore we will have a dependent competing risks case in that the underlying unobserved variables for each competing state are correlated. Abbring and van den Berg(2003) have proved that under proportionality and some regularitory assumptions, the dependent competing risks model is non-parametrically identified. Here we also invoke our earlier results that the inclusion of time-varying explanatory variables may contribute to the identification. The argument for this is similar to that of single risk case: given the assumption that the unobserved heterogeneity does not change over the spell length, the lagged explanatory variables represent the variations of unobserved heterogeneity in the earlier stage of the spell, so that the effect of the unobserved heterogeneity on current stage hazard rate can be captured by these. Other variables only have causal impacts on the transition rates in current stage.

We consider two possible transitions from origin state 0, and denote these two states be 1 and 2. In practice, we can regard the spell to be unemployment, and transitions can be thought of as e.g. either to job or to labour market programmes. Let θ_1 and θ_2 denote underlying hazard rates associated with transitions 1 and 2, which satisfy proportionality assumptions. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ denote the respective observed heterogeneities. It is possible that $\mathbf{X}_1$ and $\mathbf{X}_2$ have different components. Further, let v_1 and v_2 be the unobserved heterogeneities associated to transitions 1 and 2 respectively. The overall survival function for spell within interval $[d-1,d]$ (probability that no transition has occurred during $[d-1,d]$) is that

$$\exp\left(-\sum_k \int_{d-1}^{d} \theta_k(\tau)d\tau\right), \text{ for } k=1,2.$$

By using the same non-parametric specification for both duration baseline and unobserved heterogeneity, the state-specific transition probability can be written as:

$$h_k(d,t,x,\mu) = \left(1-\exp\left(-\sum_k \exp(\lambda_{dk}+\sigma_{kt}+\mathbf{X}_k'\beta_k+\mu_k)\right)\right) \times \frac{\exp(\lambda_{dk}+\sigma_{kt}+\mathbf{X}_k'\beta_k+\mu_k)}{\sum_k \exp(\lambda_{dk}+\sigma_{kt}+\mathbf{X}_k'\beta_k+\mu_k)} \tag{10}$$

for $k = 1,2$. The overall likelihood function can be then specified similarly as in equations (8) and (9). For individual *i*, the individual likelihood for transition *k* (*k=1,2*) is given by,

$$(11)\ L_{ik} = \left(h_{ik}(d_{ik}, t_i, x_{ik}, \mu_k)\right)^{y_{ikt}} \cdot \prod_{s=1}^{d_{ik}-y_{ikt}} \left(1 - h_{ik}(s, t_i, x_{ik}, \mu_k)\right)^{1-y_{ikt}}$$

where y_{ikt} is the censoring indicator which equals to 1 if a transition to *k* is realised, and zero otherwise. The overall likelihood is then given by,

$$(12)\ L = \prod_{i=1}^{N} \sum_{l=1}^{W} q_l \prod_{j=1}^{k} L_{ij} \mid \mu_l, \quad \sum_{l=1}^{W} q_l = 1$$

$\mu_l = (\mu_{l1}, \mu_{l2})$ is the vector of unobserved heterogeneities associated with transitions 1 and 2. Here we assume that unobserved variables have a discrete distribution with *W* different mass points, q_l is the probability of a particular combination of unobserved variables.

The Data Generating Process (DGP) is done similarly as in single risk case in section 3. We simulate a two-state mixed proportional hazard model, only consider the case of no duration dependence. The baseline hazard rates for transition 1 and 2 are simply 0.1259 and 0.0629. There is only one time-invarying dummy covariate in each hazard rate with 0.6 probability for x=1, and coefficients 1 and 0.5 respectively for transitions 1 and 2.

The calendar time variations are simulated from $N(0, \sigma^2)$. We consider the combination of three calendar time variations: no calendar time; a small variation case that the variances of the calendar time for transition 1 and 2 are 0.1 and 0.05; a large variation case with variances 1 and 0.5.

To simulate the dependence between unobserved heterogeneity terms associated with two competing hazard rates, we choose without lost of generality to simulate μ_j directly instead of simulating v_j and taking logarithm afterwards. For the sake of simplicity, we simulate a bivariate normal distributed μ_1 and μ_2. To do that, we first simulate independently two variables μ_1 and u from standard normal distribution **N**(0,1). μ_2 is then defined by

$\mu_2 = a\mu_1 + u,\ \ E(\mu_2) = 0,\ \ Var(\mu_2) = a^2 + 1,$ for a suitable constant a.

The covariance and correlations coefficient between μ_1 and μ_2 can then be derived in terms of a:

$Cov(\mu_1, \mu_2) = a, \ \ \rho = \frac{a}{\sqrt{a^2+1}}$.

by choosing $a = 1$, we have then a bivariate normal distributed μ_1 and μ_2 with $(\mu_1, \mu_2) \sim N(0,0,1,2,0.70)$.

We consider two sample sizes with 10,000 and 50,000 individuals. All in all we have 6 models (2 sample sizes, 3 calendar variations), and with 100 repetitions, we have 600 samples for estimation.

Most of our previous findings still hold for competing risks case. For expository reasons, we only report estimated means and standard errors for structure parameters β_1 and β_2. We also put our focus on how calendar variation affects identification of β_1 and β_2. Table 11 reports means and standard errors of estimated β_1 and β_2 for sample sizes 10,000 and 50,000, across 100 trials. It is encouraging to see that even for small samples with no time-varying calendar variation, the maximum likelihood still give reliable estimator for transition 1. When the calendar variation increases, the precisions of estimators are largely improved. However, maximum penalised likelihood estimators seem to be overly cautious in competing risks case, especially BIC estimators display a strong negative bias. For transition 2, we find that the quality of estimations is not as good as transition 1. It is not surprising since in DGP we deliberately fix the calendar time variation for transition 2 to be half of that of in transition 1. Lack of or low calendar time variation seems to be the reason for less accurate identification of structure parameters for transition 2. When the calendar time variation is at its largest, β_2 for transition 2 can nevertheless be reasonably well estimated by loglikelihood, AIC, HQIC, but not BIC. It seems to be advisable to avoid using BIC in competing risks cases.

Increased sample sizes certainly improve the precisions of estimators. The second panel in Table 11 reports results for estimations on samples of 50,000 individuals. The results again confirm our proposition that inclusion of time-varying covariates in the form of calendar time variation increases identifiability of structure terms of the model. When the variation is large, even BIC can reproduce the structure parameters reasonably well.

Also with large calendar variation, dispersion of estimators is reduced accordingly, given sample sizes. In Appendix Figures A12-1 and A12-2, we plot the kernel densities of estimated β_1 and β_2, across degree of calendar time variations for samples of 10,000 individuals.

Table 11: Estimated means and standard errors of $\hat{\beta}_1$ and $\hat{\beta}_2$.

		loglikelihood		AIC		BIC		HQIC	
Transition 1, $\beta_1 = 1$		**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
obs=10000	var=0	1.0057	0.1528	0.9130	0.0800	0.9624	0.0566	0.9013	0.0519
	var=0.1	1.0499	0.1374	0.9503	0.1085	0.9189	0.0810	0.8984	0.0540
	var=1	1.0343	0.0624	1.0077	0.0599	0.8586	0.0564	0.9528	0.0673
Transition 2, $\beta_2 = 0.5$									
obs=10000	var=0	0.4104	0.1896	0.3497	0.0926	0.1260	0.0479	0.2785	0.1016
	var=0.05	0.4934	0.1831	0.3993	0.1177	0.2905	0.1131	0.3142	0.0849
	var=0.5	0.5080	0.0779	0.4824	0.0738	0.3559	0.0565	0.4300	0.0840

		loglikelihood		AIC		BIC		HQIC	
Transition 1, $\beta_1 = 1$		**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
obs=50000	var=0	0.9431	0.0966	0.8930	0.0651	0.8522	0.0243	0.8581	0.0303
	var=0.1	0.9845	0.0585	0.9476	0.0572	0.8624	0.0219	0.8889	0.0421
	var=1	0.9889	0.0281	0.9826	0.0280	0.9399	0.0359	0.9672	0.0269
Transition 2, $\beta_2 = 0.5$									
obs=50000	var=0	0.4148	0.0924	0.3736	0.0711	0.3179	0.0457	0.3332	0.0345
	var=0.05	0.4690	0.0786	0.4315	0.0705	0.3223	0.0410	0.3697	0.0549
	var=0.5	0.4799	0.0295	0.4715	0.0295	0.4308	0.0415	0.4575	0.0270

Note: var=0 means the calendar time variation is 0 (none). var=0.1 means the variance for calendar time variation is 0.1, etc.

We see from the plots that the larger the calendar time variation is, the more concentrated the kernel densities are on the true parameter values. This holds for both estimators of β_1 and β_2. Note also that although BIC estimates β_2 with negative bias, the larger the calendar time variation is, the smaller the bias is. In any case, there is some evidence that time-varying calendar time variation improves identification of structure parameters in competing risks cases.

The distribution of bivariate normally distributed unobserved heterogeneity is however not very well estimated comparing to those in the single risk cases. Table 12 lists estimated means and standard errors for μ_1 and μ_2, comparing with those in DGP. The point estimators are somewhat less accurate. We also plot the distribution of estimated first moments for transitions 1 and 2 in Appendix. A surprising finding is that larger calendar time variation does not seem to help in estimation of unobserved heterogeneity. From Appendix Figures A13-1 and A13-2, we find that the best results are found with moderate calendar variations,(var=0.1 for transition 1 and var=0.05 for transition2). This phenomenon is also observed with large sample experiments (not showed here). We do not have an explanation for this at the moment, but it would certainly remain for future research.

Table 12: Estimated means and standard errors of $E(\hat{\mu}_1)$ and $E(\hat{\mu}_2)$.

		loglikelihood		AIC		BIC		HQIC	
		mean	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
Transition 1, $\mu_1 = 0.2093$									
obs=10000	var=0	0.1822	0.0320	0.1746	0.0116	0.1596	0.0049	0.1721	0.0086
	var=0.1	0.2472	0.0806	0.2217	0.0503	0.2026	0.0157	0.2110	0.0162
	var=1	0.1814	0.0708	0.1576	0.0286	0.1492	0.0144	0.1481	0.0216
Transition 2, $\mu_2 = 0.1659$									
obs=10000	var=0	0.1469	0.0392	0.1309	0.0119	0.1348	0.0060	0.1300	0.0063
	var=0.05	0.2179	0.0952	0.1777	0.0329	0.1619	0.0259	0.1681	0.0176
	var=0.5	0.1245	0.0974	0.1016	0.0281	0.0844	0.0127	0.0923	0.0185

		loglikelihood		AIC		BIC		HQIC	
		mean	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
Transition 1, $\mu_1 = 0.2093$									
obs=50000	var=0	0.1954	0.0219	0.1860	0.0093	0.1817	0.0040	0.1817	0.0045
	var=0.1	0.2696	0.0824	0.2449	0.0341	0.2215	0.0081	0.2268	0.0143
	var=1	0.2148	0.1156	0.1622	0.0200	0.1541	0.0082	0.1582	0.0173
Transition 2, $\mu_2 = 0.1659$									
obs=50000	var=0	0.1452	0.0246	0.1339	0.0088	0.1294	0.0032	0.1292	0.0035
	var=0.05	0.2288	0.0788	0.1970	0.0344	0.1695	0.0090	0.1776	0.0169
	var=0.5	0.1292	0.0995	0.1022	0.0175	0.0946	0.0074	0.0973	0.0135

Note: var=0 means the calendar time variation is 0 (none). var=0.1 means the variance for calendar time variation is 0.1, etc.

7. Conclusions

We have conducted extensive Monte Carlo experiments on non-parametric estimation of mixed proportional hazard rate models. The hazard rate is modelled with a complementary log-log formulation such that it has the flexibility to cope with arbitrary functional form of underlying hazard rate. We also simulate both parametrically and non-parametrically the duration dependence and unobserved heterogeneity. By utilising newly available computational power, we are able to estimate the mixed proportional hazard model with totally non-parametric fashion.

In addition to established identification results, we utilise the calendar time variation in hazard rates that is not perfectly correlated to spell durations, as an additional source in identification of unobserved heterogeneity. The intuition behind this is the idea that the history of the elapsed spells (in terms of previous hazard rates) could provide valuable information about population heterogeneity. By comparing estimation results from models with and without calendar time variation, we find that inclusion of calendar time variation as lagged explanatory variables improves identifiability of model parameters.

In most of our experiments, models with non-parametric specifications of both duration dependence and unobserved heterogeneity can be well estimated. This includes all model terms: duration dependence, distribution of unobserved heterogeneity (in terms of first and second moments), and covariates. We've also conducted some limited experiments on bivariate competing risks model. Our Monte Carlo results show that the conclusions on single risk models can be extended to competing risks models. Again, we find positive evidence that calendar time variation contributes to control the population heterogeneity, hence minimise the potential bias on structure parameters and duration dependence.

The non-parametric control of unobserved heterogeneity shows to be successful in most of our simulated analysis. The unknown mixing distribution, being finite discrete distributions or parametric family distributions, can be approximated by discrete masse points distribution. We find at least the first and the second moments of unknown distribution can be estimated with negligible biases, especially for single risk model cases. Our results advocate the application of Heckman and Singer's non-parametric approach in estimation of mixed proportional hazard model. However, we find that even though the data is generated with discrete distribution, our estimation in general does not return the same number of points. Rather, our estimation returns the correct moments of such discrete distribution. Therefore, we do not find the support for interpretation of such estimated supports, which we have seen in several empirical applications.

We find the sample size matters for the optimal choice of model. When sample size is small, or the variation of calendar time covariates is small or none, the maximum likelihood tends to overparameterise the mixing distribution by finding excessive mass points. This in turn produces bias away from zero on structure parameters and positive duration dependence bias. Our finding is in concord with that of Baker and Melino (2000), but much less dramatic. When sample size is increased and/or variation of calendar time is sufficiently large, the bias diminishes rapidly.

In the case of small samples, our findings suggest the use of maximum penalised likelihood. We have evaluated several popular information criteria in penalising the excessive points and find that Schwarz's Bayesian Information Criterion seems to be conservative to additional points and tend to underestimate the structural parameters, while Akaike's Information Criterion seems to be most balanced one between maximum likelihood and maximum penalised likelihood with other information criteria such as Hannan-Quinn Information criterion. We find in most cases that AIC is the recommended choice for maximum penalised likelihood. Nevertheless, when sample size is sufficiently large, maximum likelihood and maximum penalised likelihood converge to each other and choice of information criteria is of less importance.

Our findings have particular empirical relevance, because our simulation setting is based on the properties of observational data and sampling practice. With more accessible register-base data and advances of computational capacity, utilisation of data in non-parametric estimation of mixed hazard rate model can become a common practice in applied labour research.

It is important to emphasize that totally non-parametric specification of mixed proportional hazard model inevitably introduces significantly large amount of parameters in estimation, hence the computational burden sometimes seems insurmountable and the use of such flexible modelling might seem unattractive from a cost-benefit point of view. Since the likelihood function is not globally concave, ad hoc methods are needed to judge the maximum. To date, it still remains a challenge to find an effective way of determining global maximum in non-concave likelihood optimisation. Also, the asymptotic properties of non-parametric maximum likelihood estimators remain to be explored in further research. The non-parametric estimation and properties of such estimators for dependent competing risks model are also challenging subjects for future investigations.

References

Abbring, Jaap H. and Gerard J. van den Berg(2003) "The Non-Parametric Identification of the Mixed Proportional Hazards Competing Risks Model" Forthcoming *in Journal of Statistical Society, Series B*.

Abbring, Jaap H. and Gerard J. van den Berg(2001) "The Unobserved Heterogeneity Distribution in Duration Analysis" Working paper, 2001.

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle", in: B.N. Petrov and F. Csaki, eds. 2nd *International Symposium on Information Theory* (Akademia Kiado, Budapest) pp. 267-281.

Baker, Michael and Angelo Melino (2000), "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study" *Journal of Econometrics* 96, 357-393.

Brinch, Christian (2000) "Identification of Structural Duration Dependence and Unobserved Heterogeneity with Time-varying Covariates" Memorandum No. 20, 2000, Department of Economics, University of Oslo.

Cox, D.R. (1972) "Regression Models and Life-tables (with discussion)" *Journal of Royal Statistics Society*, 34, 187-202.

Elbers, Chris and Geert Ridder (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model" *Review of Economic Studies* (1982) XLIX, 403-409.

Elster, Jon (1976) "A Note on Hysteresis in the Social Sciences" Synthese, Vol. 33 (1976), 371-391.

Gaure, Simen and Knut Røed (2003), " How Tight is the Labour Market? A Micro-Based Macro Indicator" *Memoranda,* Department of Economics, University of Oslo, Norway 09/2003.

Hannan, E.J., and Quinn, B.G. (1979), "The determination of the order of an autoregression", *Journal of the Royal Statistical Society* B 41, 190-195.

Heckman, James and Bo Honoré (1989) "The Identifiability of the Competing Risks Model" *Biometrika* Vol. 76, No. 2. pp. 325-330.

Heckman, James and Burton Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data" *Econometrica*, Vol. 52, 271-320.

Heckman, James and Burton Singer (1985), “Social science duration analysis” in Heckman and Singer (ed.) *Longitudinal analysis of labor market data*, Cambridge University Press, 39-110.

Honoré, Bo E. “Identification Results for Duration Models with Multiple Spells” *Review of Economic Studies*, Vol. 60, 241-246.

Horowitz, Joel L. (1999) “Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity.” Econometrica, 67-5, pp. 1001-1028.

Huh, Keun and Robin, C. Sickles (1994) “ Estimation of the Duration Model by Nonparametric Maximum Likelihood, Maximum Penalised Likelihood, and Probability Simulators” *The Review of Economics and Statistics*, Vol. 76 (4), pp. 683-694.

Kiefer, Nicolas M., (1988) “Economic duration data and hazard functions” *Journal of Economic Literature*, Vol. XXVI, pp. 646-679.

Kiefer, Nicolas M., (1989) “Econometric Methods for Grouped Duration Data” in *Panel Data and Labor Market Studies*, (J. Hartog, G. Ridder, and J. Teeuwes, eds.), Amsterdam: Elsevier, 1989.

Kullback, S. (1968) *Information Theory and Statistics*, Dover, New York.

Lalive, Rafael, Jan C. van Ours and Josef Zweimueller (2001) ”The Impact of Active Labor Market Programs on the Duration of Unemployment, *IEW Working Paper No. 51*, University of Zurich.

Lancaster, Tony (1979), “Econometric Methods for the Duration of Unemployment” *Econometrica*, Vol. 47, No. 4, 939-956.

Lancaster, Tony (1985), “Generalised residuals and heterogeneous duration models, With applications to the Weibull model” *Journal of Econometrics*, Vol. 28(1), 155-169.

Lindsay, Bruce G. (1983) “The Geometry of Mixture Likelihoods: A General Theory” *The Annals of Statistics*, Vol. 11, 86-94.

Loemker, Leroy E. (ed.) *Leibniz: Philosophical Papers and Letters*. Dordrecht (1969).

McCall, Brian P. (1994), “Identifying State Dependence in Durations Models,” *American Statistical Association 1994, Proceedings of the Business and Economics Section*, 14-17.

McCall, Brian P. 1996, "The Identifiability of the Mixed Proportional Hazards Model with Time-Varying Coefficients," *Econometric Theory*, 12, 733-738.

McCall, Brian P. 1997, “The Non-Parametric Identifiability of a Competing Risks Model with Regressors” mimeo, University of Minnesota.

Prentice, Richard L. and Lynn A. Gloeckler (1978), “Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data” *Biometrics*, Vol. 34, 57-67.

Richardson, Katarina and Gerard J. van den Berg(2002) “The effect of vocational employment training on the individual transition rate from unemployment to work” *IFAU working paper*, 2002:8.

Røed, Knut and Tao Zhang (2000), “Labour Market Transitions and Economic Incentives.” *Memoranda,* Department of Economics, University of Oslo, Norway 15/2000

Røed, Knut and Tao Zhang (2002), “A Note on the Weibull Distribution and Time Aggregation Bias” *Applied Economic Letters*, 2002,9, 469-472.

Røed, Knut and Tao Zhang (2003), “Does Unemployment Compensation Affect Unemployment Duration?” *Economic Journal*, Vol. 113, issue 484, 190-206.

Schwarz, G. (1978), “Estimation the dimension of a model”, *Annals of Statistics* 6, 461-464.

van den Berg, Gerard J. (2001), “Duration Models: Specification, Identification, and Multiple Durations.” *Handbook of Econometrics*, Vol. 5, Chapter 55, Elsevier.

Appendix

Table A1-1 Constant hazard, 3 mass points distributed unobserved heterogeneity, 10,000 individuals.

model	obs	points	# parameter	loglikelihood	AIC	BIC	HQIC	$\hat{\beta}$	std
Constant Hazard	62177	1	1	-20962.0016	-20963.0016	-20967.5204	-20964.4029	0.5765	0.0253
Var(σ)=0	62177	2	3	-20925.3318	-20928.3318	-20941.8884	-20932.5357	0.8630	0.0468
	62177	3	5	-20896.3318	**-20901.3318**	**-20923.9261**	**-20908.3384**	1.0470	0.0469
	62177	4	7	**-20896.0052**	-20903.0052	-20934.6373	-20912.8144	1.0451	0.0557
	62177	5	9	-20896.0100	-20905.0100	-20945.6798	-20917.6218	1.0449	0.0845
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	51493	1	1	-17933.8436	-17934.8436	-17939.2682	-17936.2277	0.6192	0.0273
Var(σ)=0.001	51493	2	3	-17904.9660	-17907.9660	-17921.2398	-17912.1183	0.9027	0.0516
	51493	3	5	-17887.7413	**-17892.7413**	**-17914.8643**	**-17899.6617**	1.0458	0.0551
(1)	51493	4	7	**-17887.3836**	-17894.3836	-17925.3558	-17904.0723	1.0304	0.0598
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	54873	1	1	-18005.4370	-18006.4370	-18010.8933	-18007.8269	0.6408	0.0279
Var(σ)=0.1	54873	2	3	-17980.0437	-17983.0437	-17996.4128	-17987.2135	0.9104	0.0560
	54873	3	5	-17966.5309	**-17971.5309**	**-17993.8129**	**-17978.4806**	1.0120	0.0576
(2)	54873	4	7	**-17965.7898**	-17972.7898	-18003.9845	-17982.5193	1.0092	0.0614
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	31875	1	1	-14188.2544	-14189.2544	-14193.4392	-14190.5933	0.6899	0.0245
Var(σ)=1	31875	2	3	-14052.9035	-14055.9035	-14068.4579	-14059.9201	0.9046	0.0343
	31875	3	5	**-14031.3681**	**-14036.3681**	**-14057.2920**	**-14043.0624**	1.0013	0.0387
	31875	4	7	-14032.7800	-14039.7800	-14069.0736	-14049.1522	1.0364	0.0457

Note: 1. Number of observation listed in table is number of monthly observation is estimation data. 2. Var(σ) is variance of calendar month in simulation. 3. Number of parameters is free parameters associated with unobserved heterogeneity. (1) indicates iteration terminates when near zero probability on added point is encountered. (2) indicates numerical difficulty prevents further search of mass points.

Table A1-2 Weibull hazard, Gamma distributed unobserved heterogeneity, 10,000 individuals.

model	obs	points	# parameter	loglikelihood	AIC	BIC	HQIC	$\hat{\beta}$	std
Weibull Hazard	60763	1	1	-21145.3042	-21146.3042	-21150.8116	-21147.7034	0.7482	0.0254
Var(σ)=0	60763	2	3	-21128.8475	-21131.8475	**-21145.3696**	-21136.0452	0.9633	0.0553
	60763	3	5	-21122.2988	**-21127.2988**	-21149.8356	**-21134.2950**	1.0448	0.0588
	60763	4	7	-21121.5678	-21128.5678	-21160.1194	-21138.3625	1.0539	0.0847
(1)	60763	5	9	**-21120.6688**	-21129.6688	-21170.2351	-21142.2619	1.0561	0.0741
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	49577	1	1	-18399.4842	-18400.4842	**-18404.8898**	-18401.8648	0.7663	0.0270
Var(σ)=0.001	49577	2	3	-18398.5146	-18401.5146	-18414.7315	-18405.6564	0.7659	0.0329
	49577	3	5	-18378.6182	**-18383.6182**	-18405.6464	**-18390.5211**	0.9704	0.0513
	49577	4	7	**-18377.4539**	-18384.4539	-18415.2934	-18394.1181	0.9890	0.0727
	49577	5	9	-18377.6288	-18386.6288	-18426.2796	-18399.0541	0.9878	0.0993
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	57146	1	1	-17161.4105	-17162.4105	**-17166.8871**	-17163.8041	0.7713	0.0299
Var(σ)=0.1	57146	2	3	-17154.7222	**-17157.7222**	-17171.1523	**-17161.9032**	0.9483	0.0605
	57146	3	5	-17154.1197	-17159.1197	-17181.5031	-17166.0879	0.9432	0.0626
	57146	4	7	-17153.7483	-17160.7483	-17192.0850	-17170.5038	0.9500	0.0742
(2)	57146	5	9	**-17153.6975**	-17162.6975	-17202.9876	-17175.2403	0.9502	0.0765
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	57081	1	1	-16196.1120	-16197.1120	-16201.5881	-16198.5055	0.8301	0.0292
Var(σ)=1	57081	2	3	-16178.1004	-16181.1004	**-16194.5288**	**-16185.2811**	1.0208	0.0510
	57081	3	5	-16174.0864	**-16179.0864**	-16201.4669	-16186.0541	1.0424	0.0518
	57081	4	7	-16173.7278	-16180.7278	-16212.0606	-16190.4826	1.0474	0.0553
	57081	5	9	-16173.4978	-16182.4978	-16222.7828	-16195.0397	1.0514	0.0558
	57081	6	11	-16173.4186	-16184.4186	-16233.6558	-16199.7475	1.0697	0.0575
(2)	57081	7	13	**-16173.2179**	-16186.2179	-16244.4074	-16204.3339	1.0804	0.0587

Note: 1. Number of observation listed in table is number of monthly observation is estimation data. 2. Var(σ) is variance of calendar month in simulation. 3. Number of parameters is free parameters associated with unobserved heterogeneity. (1) indicates iteration terminates when near zero probability on added point is encountered. (2) indicates numerical difficulty prevents further search of mass points.

Table A2-1 Constant hazard, 3 mass points distributed unobserved heterogeneity, 50,000 individuals.

model	obs	*points*	# parameter	loglikelihood	AIC	BIC	HQIC	$\hat{\beta}$	std
Constant Hazard	315443	1	1	-102425.1077	-102426.1077	-102431.4386	-102427.6463	0.5433	0.0114
Var(σ)=0	315443	2	3	-102244.7582	-102247.7582	-102263.7508	-102252.3739	0.8130	0.0224
	315443	3	5	-102043.6526	-102048.6526	**-102075.3070**	**-102056.3455**	1.0075	0.0206
	315443	4	7	**-102040.5558**	**-102047.5558**	-102084.8718	-102058.3259	1.0357	0.0278
	315443	5	9	-102040.8163	-102049.8163	-102097.7941	-102063.6636	1.0376	0.0358
model	**obs**	***points***	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	261663	1	1	-88542.7267	-88543.7267	-88548.9641	-88545.2504	0.5820	0.0123
Var(σ)=0.001	261663	2	3	-88387.7595	-88390.7595	-88406.4717	-88395.3306	0.8449	0.0247
	261663	3	5	-88222.4987	**-88227.4987**	**-88253.6858**	**-88235.1173**	0.9872	0.0212
	261663	4	7	-88221.5047	-88228.5047	-88265.1666	-88239.1707	1.0062	0.0318
	261663	5	9	**-88220.4275**	-88229.4275	-88276.5642	-88243.1409	1.0115	0.0276
	261663	6	11	-88220.4458	-88231.4458	-88289.0573	-88248.2066	1.0130	0.0411
model	**obs**	***points***	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	242683	1	1	-87156.5071	-87157.5071	-87162.7069	-87159.0248	0.5566	0.0119
Var(σ)=0.1	242683	2	3	-86960.4634	-86963.4634	-86979.0627	-86968.0164	0.8087	0.0212
	242683	3	5	-86770.4751	**-86775.4751**	**-86801.4738**	**-86783.0633**	0.9927	0.0202
	242683	4	7	-86769.1550	-86776.1550	-86812.5533	-86786.7786	1.0218	0.0286
	242683	5	9	**-86767.2713**	-86776.2713	-86823.0691	-86789.9302	1.0046	0.0239
	242683	6	11	-86768.4883	-86779.4883	-86836.6856	-86796.1825	1.0645	0.0454
model	**obs**	***points***	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Constant Hazard	176483	1	1	-72982.9780	-72983.9780	-72989.0185	-72985.4697	0.5229	0.0109
Var(σ)=1	176483	2	3	-72312.9198	-72315.9198	-72331.0413	-72320.3947	0.7142	0.0154
	176483	3	5	-71977.8526	-71982.8526	-72008.0550	-71990.3108	0.8910	0.0179
	176483	4	7	-71909.3361	-71916.3361	-71951.6195	-71926.7775	0.9691	0.0207
	176483	5	9	**-71876.6380**	**-71885.6380**	**-71931.0024**	**-71899.0627**	1.0175	0.0218
	176483	6	11	-71876.6416	-71887.6416	-71943.0870	-71904.0495	1.0175	0.0239

Note: 1. Number of observation listed in table is number of monthly observation is estimation data. 2. Var(σ) is variance of calendar month in simulation.
3. Number of parameters is free parameters associated with unobserved heterogeneity. (1) indicates iteration terminates when approximate zero probability on added point is encountered. (2) indicates numerical difficulty prevents further search of mass points.

Table A2-2 Weibull hazard, Gamma distributed unobserved heterogeneity, 50,000 individuals.

model	obs	points	# parameter	loglikelihood	AIC	BIC	HQIC	$\hat{\beta}$	std
Weibull Hazard	307303	1	1	-106369.3795	-106370.3795	-106375.6973	-106371.9160	0.7307	0.0113
Var(σ)=0	307303	2	3	-106312.1446	-106315.1446	-106331.0980	-106319.7541	0.9082	0.0259
	307303	3	5	-106293.7916	-106298.7916	**-106325.3806**	**-106306.4742**	0.9293	0.0238
	307303	4	7	**-106290.8566**	**-106297.8566**	-106335.0812	-106308.6123	0.9863	0.0411
	307303	5	9	-106291.1861	-106300.1861	-106348.0463	-106314.0148	0.9730	0.0557
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	248757	1	1	-91153.4145	-91154.4145	-91159.6266	-91155.9341	0.7584	0.0121
Var(σ)=0.001	248757	2	3	-91099.3846	-91102.3846	-91118.0209	-91106.9435	0.9437	0.0262
	248757	3	5	-91085.7817	**-91090.7817**	**-91116.8423**	**-91098.3800**	0.9721	0.0270
	248757	4	7	-91084.8411	-91091.8411	-91128.3259	-91102.4786	0.9976	0.0400
	248757	5	9	-91084.5795	-91093.5795	-91140.4886	-91107.2564	0.9938	0.0484
(2)	248757	6	11	**-91083.7393**	-91094.7393	-91152.0725	-91111.4554	1.0016	0.0460
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	287995	1	1	-85369.6122	-85370.6122	-85375.8976	-85372.1436	0.7900	0.0134
Var(σ)=0.1	287995	2	3	-85333.3239	-85336.3239	**-85352.1799**	-85340.9180	0.9543	0.0270
	287995	3	5	-85327.0783	**-85332.0783**	-85358.5050	**-85339.7351**	0.9384	0.0258
	287995	4	7	-85326.9290	-85333.9290	-85370.9264	-85344.6486	0.9399	0.0327
	287995	5	9	**-85326.7030**	-85335.7030	-85383.2712	-85349.4853	0.9491	0.0336
(1)	287995	6	11	-85329.2171	-85340.2171	-85398.3559	-85357.0622	0.9022	0.0215
model	**obs**	**points**	**# parameter**	**loglikelihood**	**AIC**	**BIC**	**HQIC**	$\hat{\beta}$	**std**
Weibull Hazard	288414	1	1	-81110.7014	-81111.7014	-81116.9874	-81113.2329	0.8044	0.0130
Var(σ)=1	288414	2	3	-81021.5975	-81024.5975	**-81040.4557**	-81029.1920	0.9957	0.0227
	288414	3	5	-81015.3996	**-81020.3996**	-81046.8300	**-81028.0570**	1.0029	0.0229
(2)	288414	4	7	**-81015.0915**	-81022.0915	-81059.0940	-81032.8119	1.0093	0.0245

Note: 1. Number of observation listed in table is number of monthly observation is estimation data. 2. Var(σ) is variance of calendar month in simulation.
3. Number of parameters is free parameters associated with unobserved heterogeneity. (1) indicates iteration terminates when approximate zero probability on added point is encountered. (2) indicates numerical difficulty prevents further search of mass points. ,

Table A3: Number of trials that confidence intervals for estimated $\hat{\beta}$ cover the true parameter value 1 used in DGP.

Constant hazard, 3 points unobserved heterogeneity

# obs	var(month)	Loglikelihood	AIC	BIC	HQIC
5000	0	66	67	75	75
	0.001	74	71	84	85
	0.1	84	85	88	88
	1	73	74	73	74
10000	0	80	77	79	79
	0.001	79	77	76	78
	0.1	71	70	72	72
	1	74	75	71	75
50000	0	90	86	94	91
	0.001	95	94	93	93
	0.1	90	92	94	92
	1	97	97	98	98

Weibull hazard, Gamma distributed unobserved heterogeneity

# obs	var(month)	Loglikelihood	AIC	BIC	HQIC
5000	0	76	79	5	72
	0.001	83	84	4	62
	0.1	79	77	5	35
	1	78	84	37	81
10000	0	78	70	54	71
	0.001	69	68	48	70
	0.1	81	79	30	77
	1	76	76	73	77
50000	0	96	84	76	75
	0.001	94	82	78	80
	0.1	92	85	95	90
	1	91	91	96	94

Note: the number of trials that the estimated confidence intervals cover the true value 1 in DGP is calculated based on the estimations that return more than 1 point of support for the unobserved heterogeneity distribution.

Table A4: Maximum number of support points found.

Constant hazard Gamma distributed unobserved heterogeneity, 5,000 obs.

	Var(month)	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
loglikelihood	0	14	0	16	24	21	19	6
	0.001	0	0	23	22	32	17	6
	0.1	3	0	7	36	28	20	6
	1	19	1	11	32	26	7	4
AIC	0	14	5	71	9	1		
	0.001	1	11	70	13	5		
	0.1	5	26	55	13	1		
	1	19	27	45	8	1		
BIC	0	92	2	6	0	0		
	0.001	95	2	3	0	0		
	0.1	89	11	0	0	0		
	1	64	36	0	0	0		
HQIC	0	22	20	58	0	0		
	0.001	25	22	53	0	0		
	0.1	32	46	22	0	0		
	1	22	58	20	0	0		

Weibull hazard, 3 points distributed unobserved heterogeneity, 5,000 obs.

	Var(month)	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
loglikelihood	0	18	0	13	18	32	11	8
	0.001	5	0	15	23	42	11	4
	0.1	12	0	18	23	31	10	6
	1	16	0	17	33	28	6	0
AIC	0	18	0	77	4	1		
	0.001	5	0	84	10	1		
	0.1	12	7	77	3	1		
	1	16	2	75	7	0		
BIC	0	18	4	78	0	0		
	0.001	7	9	84	0	0		
	0.1	21	57	22	0	0		
	1	16	49	35	0	0		
HQIC	0	18	0	82	0	0		
	0.001	5	1	93	1	0		
	0.1	12	21	67	0	0		
	1	16	10	72	2	0		

Table A5: Maximum number of points found across sample sizes.

Constant hazard Gamma distributed unobserved heterogeneity, var(month)=0.1

	Obs	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
Log likelihood	5000	3	0	7	36	28	20	6
	10000	19	0	9	34	24	11	3
	50000	1	0	2	37	42	15	3
AIC	5000	5	26	55	13	1		
	10000	19	17	52	10	2		
	50000	1	0	72	25	2		
BIC	5000	89	11	0	0	0		
	10000	48	51	1	0	0		
	50000	1	36	63	0	0		
HQIC	5000	32	46	22	0	0		
	10000	21	49	29	0	1		
	50000	1	0	92	7	0		

Weibull hazard, 3 points distributed unobserved heterogeneity, var(month)=0.1

	Obs	1 point	2 points	3points	4 points	5 points	6 points	7 or more points
Log likelihood	5000	12	0	18	23	31	10	6
	10000	19	0	14	35	25	7	0
	50000	3	0	21	41	22	10	3
AIC	5000	12	7	77	3	1		
	10000	19	2	72	6	1		
	50000	3	0	86	10	1		
BIC	5000	21	57	22	0	0		
	10000	19	54	27	0	0		
	50000	3	0	97	0	0		
HQIC	5000	12	21	67	0	0		
	10000	19	13	68	0	0		
	50000	3	0	97	0	0		

Table A6: Estimated means and standard errors of $\hat{\beta}$.

Constant hazard, Gamma distributed unobserved heterogeneity

# obs	var(month)	Loglikelihood mean	Loglikelihood std.	AIC mean	AIC std.	BIC mean	BIC std.	HQIC mean	HQIC std.
5000	0	1.0969	0.2250	0.9879	0.1117	1.0082	0.0683	0.9529	0.0631
	0.001	1.1328	0.2994	1.0108	0.1807	1.0439	0.0685	0.9681	0.0722
	0.1	1.1128	0.1671	1.0261	0.1010	1.0606	0.0411	1.0192	0.0643
	1	1.0324	0.0794	1.0022	0.0740	0.9972	0.0468	0.9837	0.0619
10000	0	1.0314	0.1040	0.9890	0.0851	0.9295	0.0545	0.9514	0.0472
	0.001	1.0716	0.2036	1.0057	0.1770	0.9409	0.0519	0.9534	0.0594
	0.1	1.0306	0.0865	0.9888	0.0769	0.9717	0.0377	0.9698	0.0604
	1	1.0036	0.0512	0.9840	0.0456	0.9533	0.0344	0.9684	0.0429
50000	0	1.0128	0.0514	0.9909	0.0465	0.9545	0.0216	0.9633	0.0352
	0.001	1.0441	0.1824	0.9936	0.0665	0.9562	0.0204	0.9685	0.0395
	0.1	1.0134	0.0315	0.9991	0.0320	0.9780	0.0209	0.9890	0.0258
	1	1.0056	0.0249	1.0001	0.0264	0.9856	0.0217	0.9911	0.0238

Weibull hazard, 3 points distributed unobserved heterogeneity

# obs	var(month)	Loglikelihood mean	Loglikelihood std.	AIC mean	AIC std.	BIC mean	BIC std.	HQIC mean	HQIC std.
5000	0	1.1871	0.3021	1.0574	0.0904	1.0413	0.0660	1.0462	0.0589
	0.001	1.1722	0.2589	1.0593	0.0926	1.0401	0.0684	1.0453	0.0664
	0.1	1.0958	0.1117	1.0418	0.0776	1.0350	0.0631	1.0348	0.0726
	1	1.0893	0.0732	1.0731	0.0666	1.0468	0.0656	1.0650	0.0648
10000	0	1.0896	0.1499	1.0418	0.0860	1.0246	0.0508	1.0299	0.0636
	0.001	1.1064	0.1961	1.0360	0.0833	1.0050	0.0622	1.0176	0.0500
	0.1	1.0388	0.0954	1.0212	0.0840	1.0013	0.0608	1.0098	0.0637
	1	1.0238	0.0510	1.0148	0.0501	0.9972	0.0530	1.0100	0.0521
50000	0	1.0340	0.0713	1.0047	0.0253	0.9981	0.0179	0.9991	0.0205
	0.001	1.0269	0.0548	1.0086	0.0374	0.9975	0.0239	1.0002	0.0256
	0.1	1.0167	0.0304	1.0106	0.0291	1.0066	0.0256	1.0066	0.0256
	1	1.0069	0.0229	1.0025	0.0222	0.9998	0.0197	0.9999	0.0197

Note: 1. means are calculated among estimations that successfully found more than 1 points of support for unobserved heterogeneity. 2. var(month) is the variance of calendar month variation in DGP.

Table A7: Estimated means and standard errors of the first moment for the unobserved heterogeneity distribution.

Constant hazard, Gamma distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	0.1313	0.0207	0.1266	0.0088	0.1207	0.0045	0.1269	0.0085
	0.001	0.1430	0.0871	0.1287	0.0119	0.1188	0.0097	0.1278	0.0124
	0.1	0.1342	0.0242	0.1266	0.0131	0.1180	0.0089	0.1249	0.0113
	1	0.1294	0.0209	0.1355	0.0904	0.1236	0.0110	0.1252	0.0125
10000	0	0.1346	0.0456	0.1277	0.0068	0.1291	0.0066	0.1277	0.0060
	0.001	0.1373	0.0335	0.1307	0.0190	0.1284	0.0097	0.1277	0.0098
	0.1	0.1297	0.0140	0.1275	0.0083	0.1281	0.0088	0.1282	0.0085
	1	0.1303	0.0130	0.1284	0.0111	0.1298	0.0092	0.1286	0.0097
50000	0	0.1258	0.0050	0.1251	0.0033	0.1249	0.0024	0.1249	0.0028
	0.001	0.1298	0.0296	0.1251	0.0055	0.1246	0.0039	0.1247	0.0039
	0.1	0.1255	0.0044	0.1251	0.0040	0.1252	0.0037	0.1249	0.0038
	1	0.1253	0.0046	0.1249	0.0046	0.1242	0.0041	0.1242	0.0041

Weibull hazard, 3 points distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	0.1304	0.0280	0.1226	0.0088	0.1230	0.0090	0.1227	0.0087
	0.001	0.1265	0.0206	0.1219	0.0142	0.1221	0.0143	0.1220	0.0142
	0.1	0.1289	0.0205	0.1251	0.0146	0.1239	0.0143	0.1249	0.0143
	1	0.1241	0.0187	0.1219	0.0156	0.1233	0.0153	0.1223	0.0161
10000	0	0.1243	0.0077	0.1236	0.0056	0.1237	0.0056	0.1237	0.0055
	0.001	0.1274	0.0177	0.1235	0.0108	0.1245	0.0117	0.1236	0.0108
	0.1	0.1253	0.0105	0.1246	0.0090	0.1256	0.0093	0.1250	0.0092
	1	0.1283	0.0148	0.1266	0.0121	0.1266	0.0106	0.1259	0.0108
50000	0	0.1275	0.0100	0.1258	0.0031	0.1256	0.0022	0.1256	0.0022
	0.001	0.1267	0.0057	0.1257	0.0049	0.1257	0.0048	0.1257	0.0049
	0.1	0.1260	0.0055	0.1258	0.0053	0.1258	0.0052	0.1258	0.0052
	1	0.1260	0.0043	0.1257	0.0041	0.1258	0.0039	0.1257	0.0039

Note: 1. means are calculated among estimations that successfully found more than 1 points of support for unobserved heterogeneity. 2. var(month) is the variance of calendar month variation in DGP. 3. the true first moment in DGP is (rescaled) 0.125893.

Table A8: Estimated means and standard errors of the second moment for unobserved heterogeneity distribution.

Constant hazard, Gamma distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	0.0446	0.0663	0.0237	0.0117	0.0197	0.0021	0.0215	0.0039
	0.001	0.2027	1.5527	0.0274	0.0198	0.0187	0.0031	0.0222	0.0048
	0.1	0.0606	0.0967	0.0274	0.0158	0.0237	0.0039	0.0241	0.0048
	1	0.0387	0.0513	5.0320	45.0592	0.0244	0.0046	0.0235	0.0052
10000	0	0.1174	0.7808	0.0251	0.0088	0.0224	0.0032	0.0217	0.0044
	0.001	0.0639	0.1549	0.0334	0.0478	0.0230	0.0041	0.0220	0.0039
	0.1	0.0367	0.0349	0.0254	0.0079	0.0261	0.0040	0.0247	0.0047
	1	0.0328	0.0289	0.0264	0.0150	0.0260	0.0039	0.0248	0.0047
50000	0	0.0274	0.0109	0.0240	0.0046	0.0204	0.0009	0.0212	0.0026
	0.001	0.0702	0.3879	0.0245	0.0116	0.0203	0.0013	0.0216	0.0039
	0.1	0.0268	0.0060	0.0243	0.0048	0.0228	0.0022	0.0228	0.0034
	1	0.0261	0.0075	0.0253	0.0090	0.0219	0.0019	0.0225	0.0031

Weibull hazard, 3 points distributed unobserved heterogeneity

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	0.0519	0.0869	0.0250	0.0060	0.0246	0.0042	0.0243	0.0038
	0.001	0.0415	0.0483	0.0250	0.0069	0.0246	0.0063	0.0242	0.0060
	0.1	0.0431	0.0531	0.0263	0.0074	0.0285	0.0071	0.0262	0.0070
	1	0.0343	0.0532	0.0258	0.0098	0.0278	0.0072	0.0259	0.0099
10000	0	0.0321	0.0169	0.0267	0.0051	0.0258	0.0029	0.0260	0.0034
	0.001	0.0403	0.0380	0.0273	0.0080	0.0267	0.0057	0.0259	0.0047
	0.1	0.0298	0.0108	0.0268	0.0060	0.0288	0.0048	0.0263	0.0042
	1	0.0337	0.0297	0.0282	0.0105	0.0283	0.0058	0.0266	0.0049
50000	0	0.0329	0.0277	0.0263	0.0042	0.0255	0.0009	0.0256	0.0012
	0.001	0.0295	0.0091	0.0266	0.0043	0.0255	0.0020	0.0258	0.0027
	0.1	0.0276	0.0048	0.0266	0.0035	0.0260	0.0022	0.0260	0.0022
	1	0.0270	0.0037	0.0261	0.0029	0.0257	0.0017	0.0257	0.0018

Note: 1. means are calculated among estimations that successfully found more than 1 points of support for unobserved heterogeneity. 2. var(month) is the variance of calendar month variation in DGP. 3. the true second moment in DGP is (rescaled) 0.026111 .

Figure A1: Kernel densities of estimated β by sample sizes. Constant hazard, 3 points mixture, var(month)=0.1.

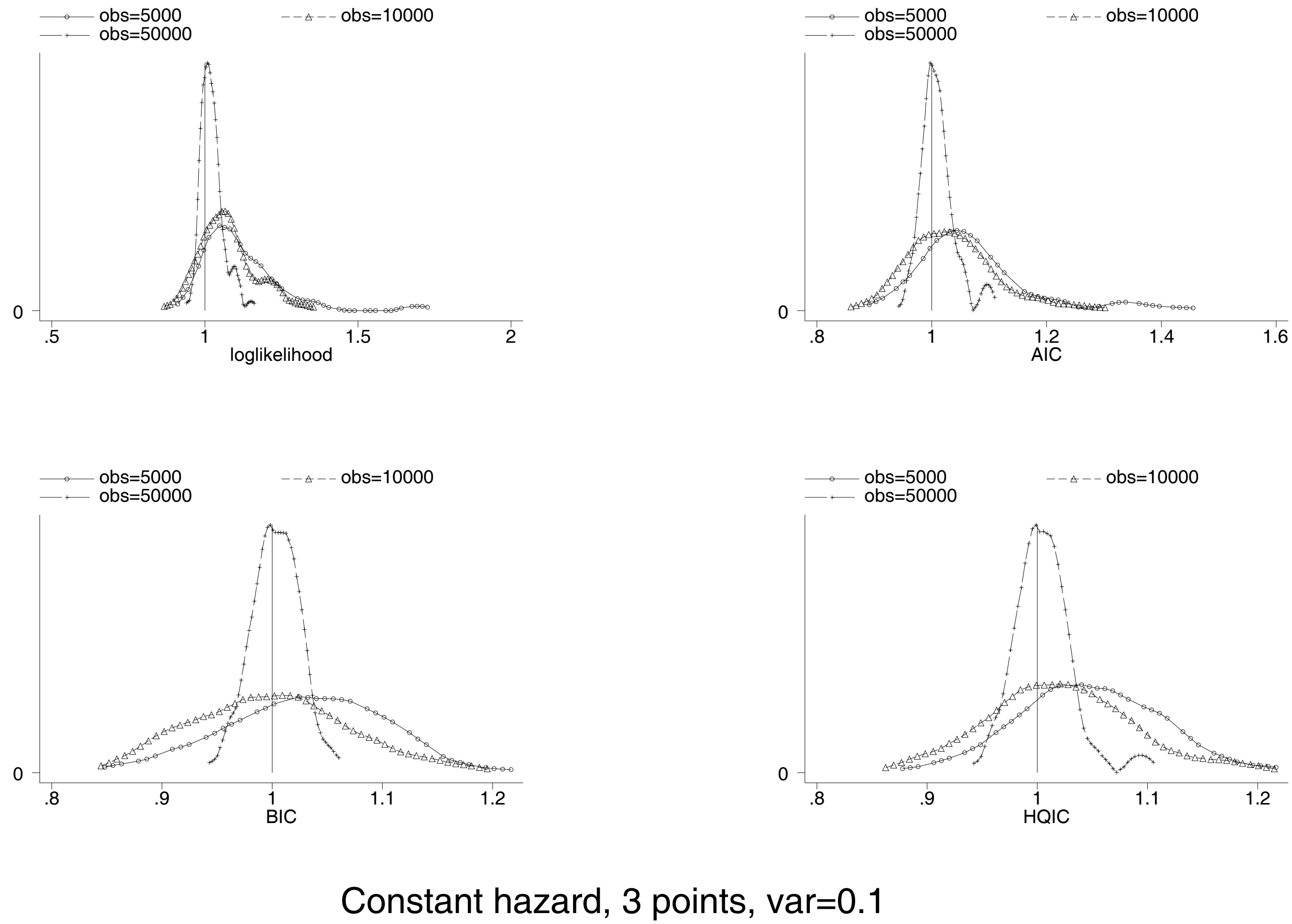


Figure A2: Kernel densities of estimated β by sample sizes. Weibull hazard, Gamma mixture, var(month)=0.1.

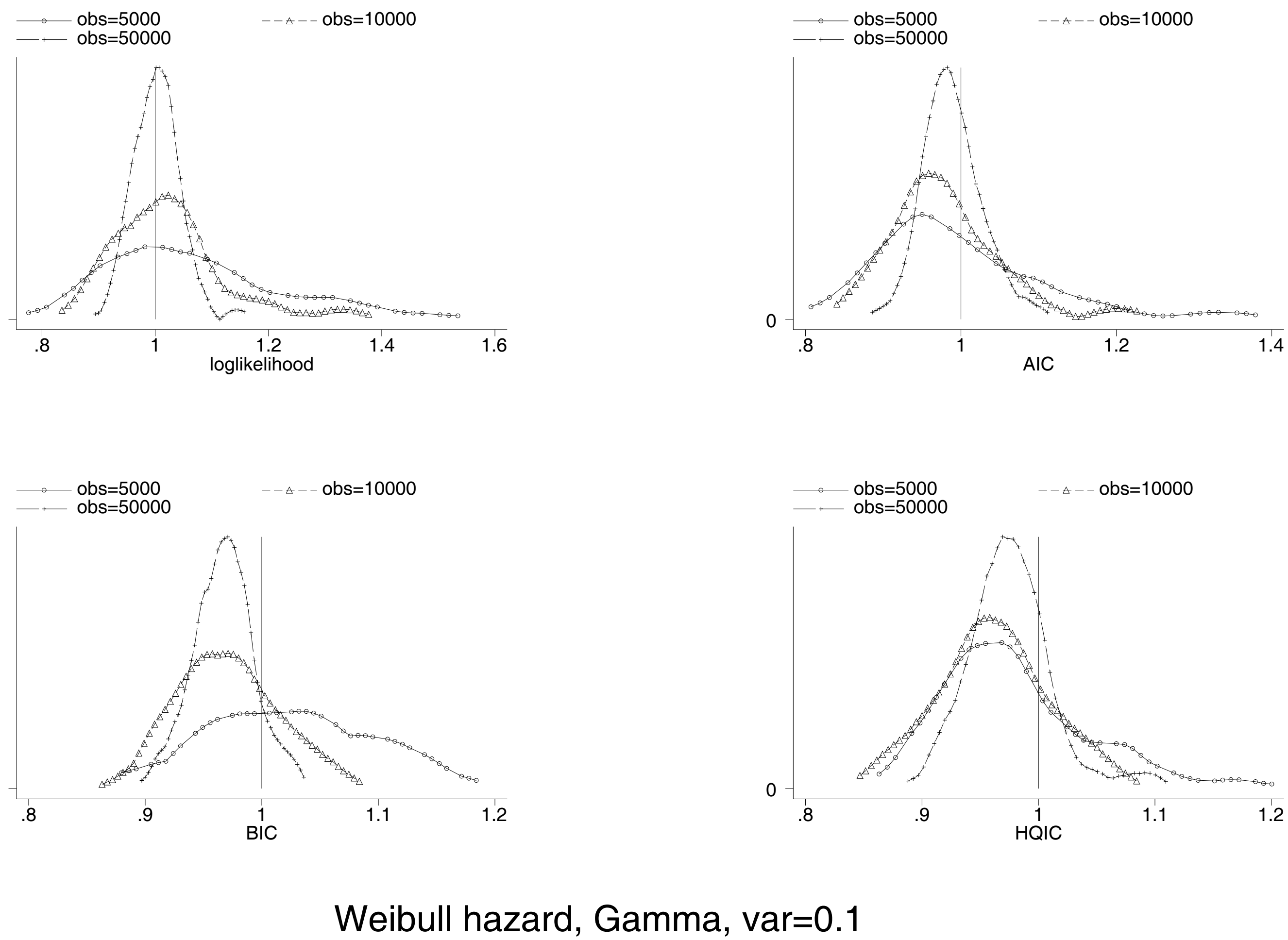
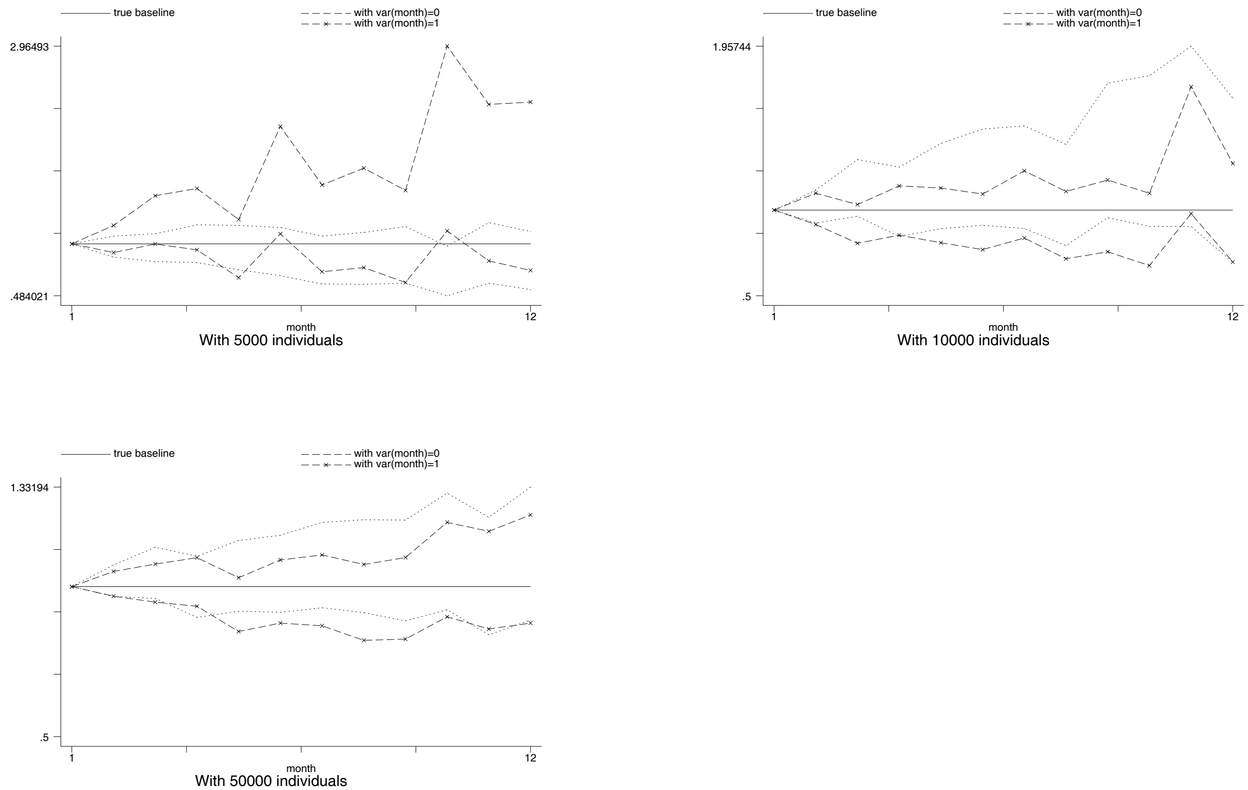
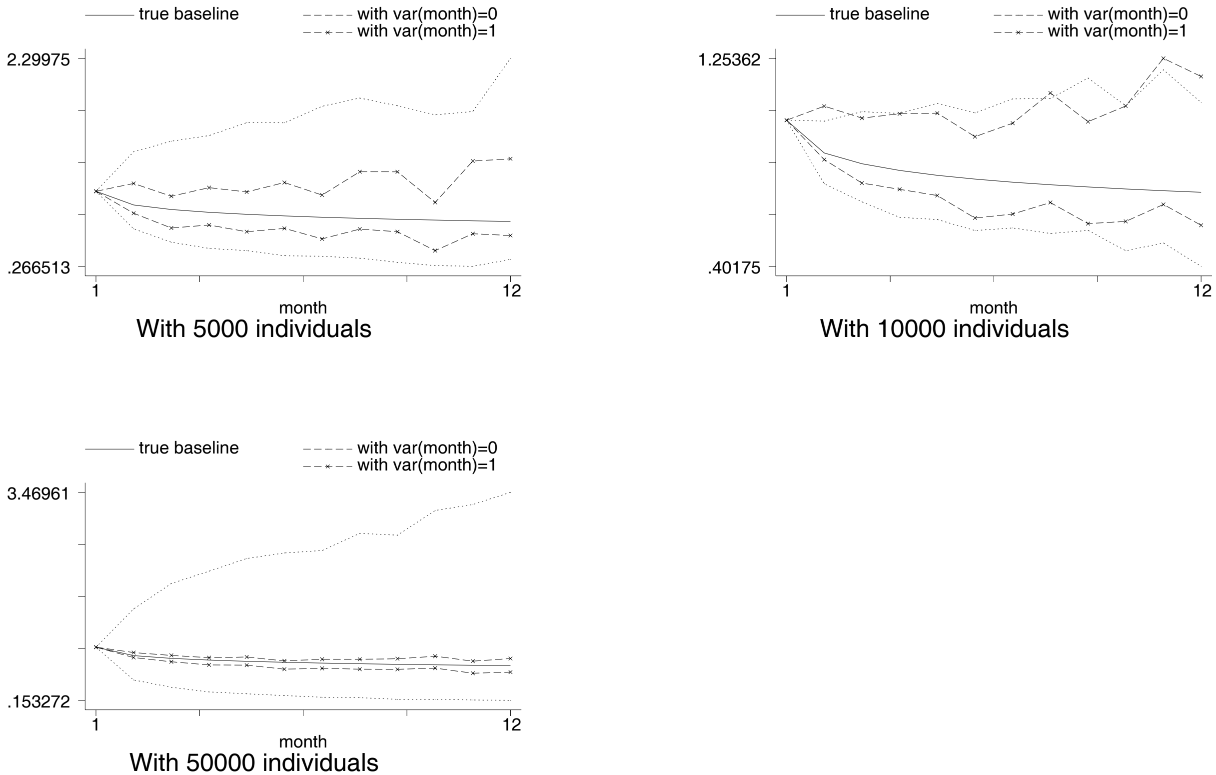


Figure A3: 95% Confidence intervals for baseline hazard rate estimates, across calendar variations. Constant hazard, 3 points mixture.



Note: confidence intervals are calculated based on the estimated standard errors (in exponential form) for the duration baseline estimators from the estimations that produce Table 4-1. Therefore they do not have the interpretation as confidence intervals for transition probabilities.

Figure A4: 95% Confidence intervals for baseline hazard rate estimates, across calendar variations. Weibull hazard, Gamma mixture.



Note: confidence intervals are calculated based on the estimated standard errors (in exponential form) for the duration baseline estimators from the estimations that produce Table 4-2. Therefore they do not have the interpretation as confidence intervals for transition probabilities.

Figure A5: 95% Confidence intervals for baseline hazard rate estimates, across sample sizes, var(month)=0.1.

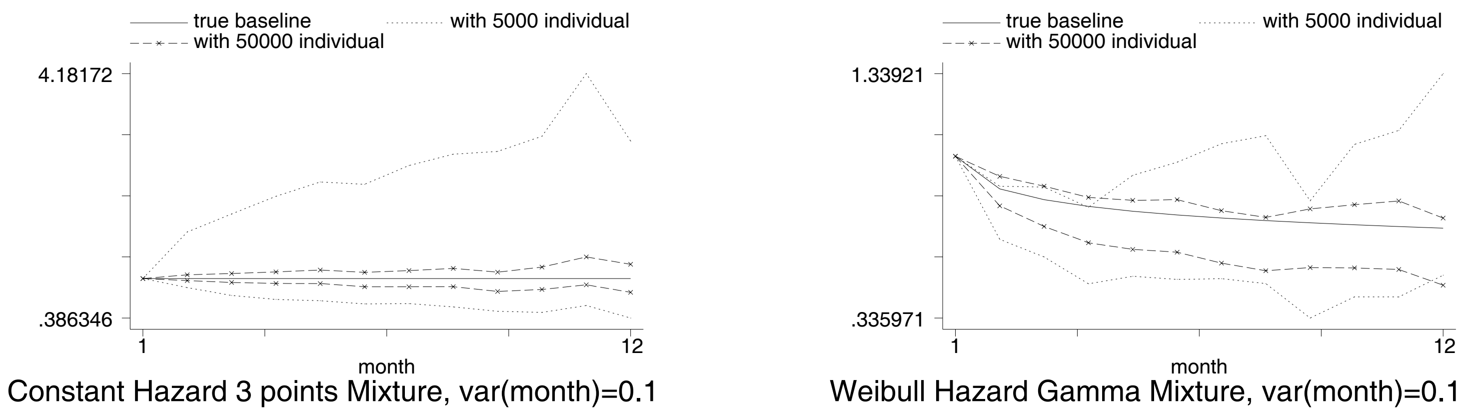
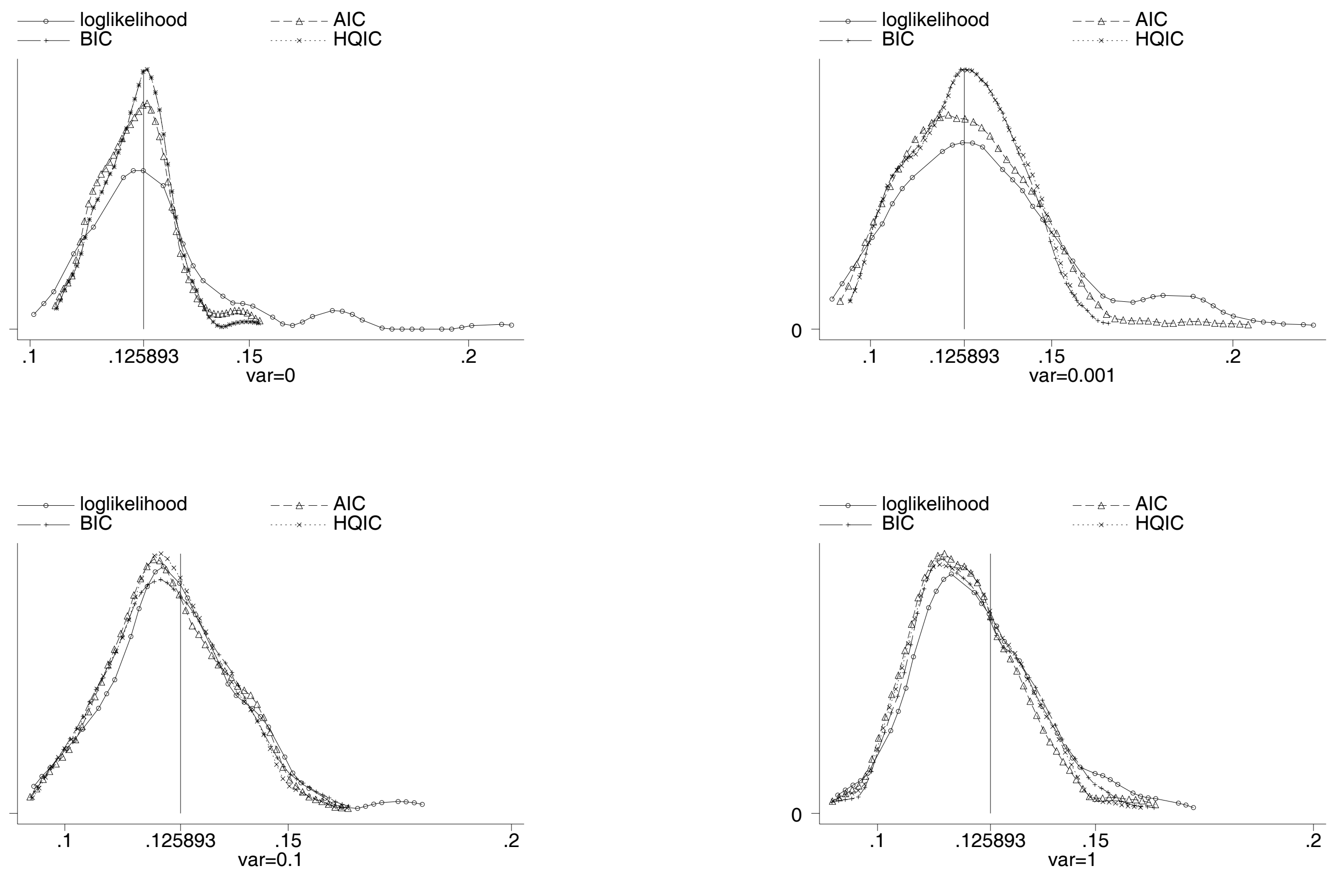
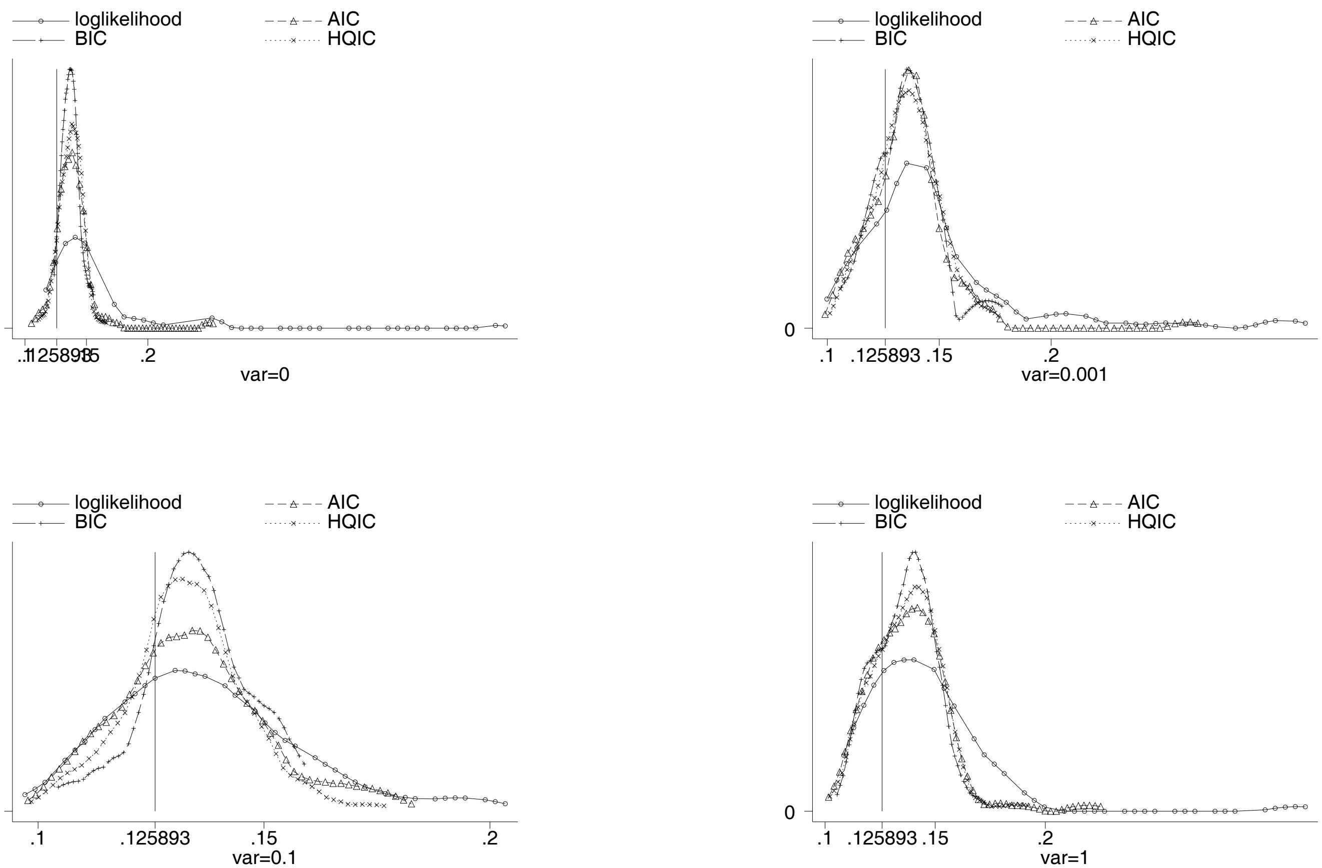


Figure A6: Kernel Densities of estimated expectation of $\hat{\mu}$. Constant hazard, 3 points mixture, 5,000 individuals.



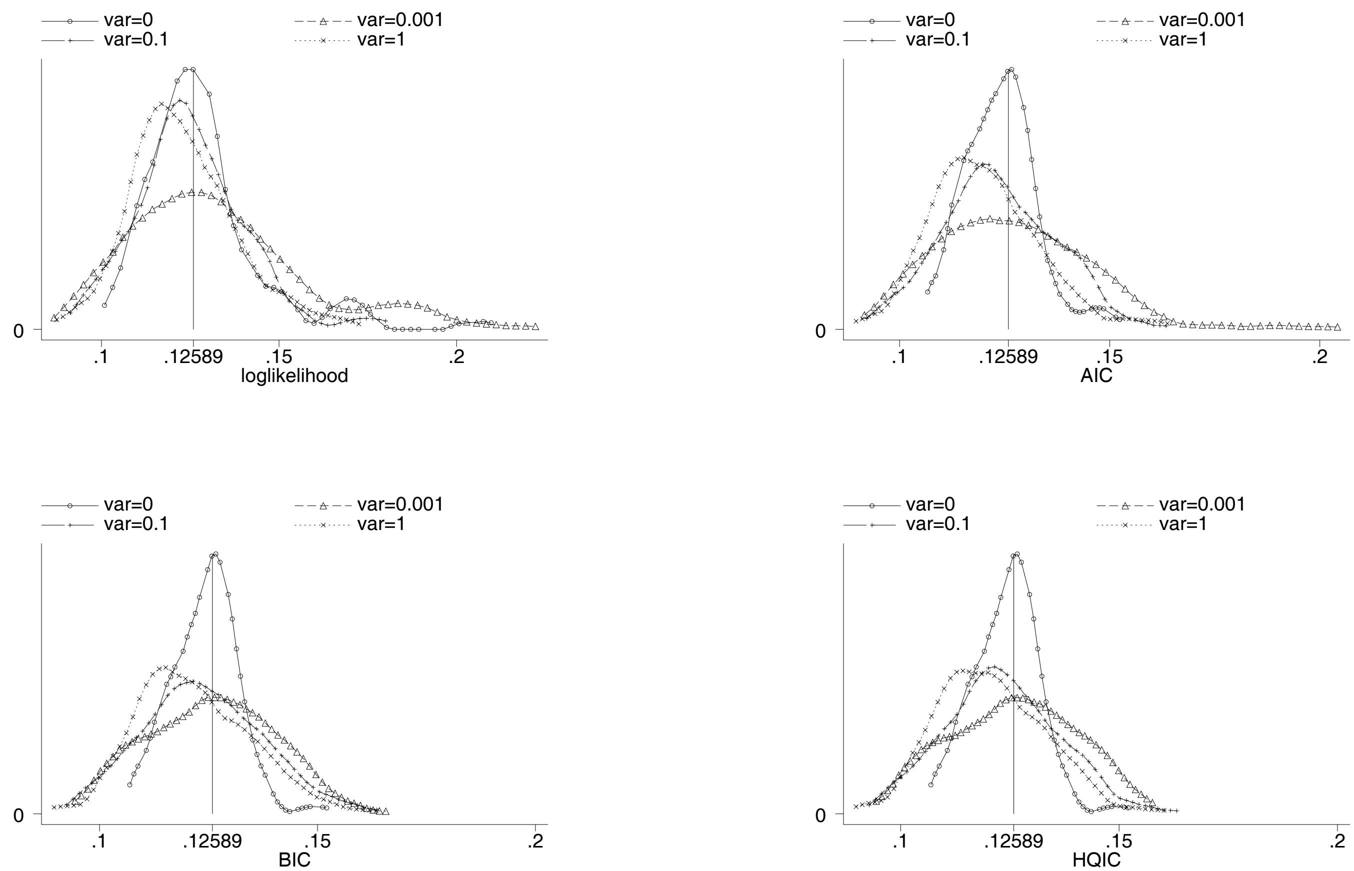
Constant hazard, 3 points, 5000 obs

Figure A7: Kernel Densities of estimated expectation of $\hat{\mu}$. Weibull hazard, Gamma mixture, 5,000 individuals.



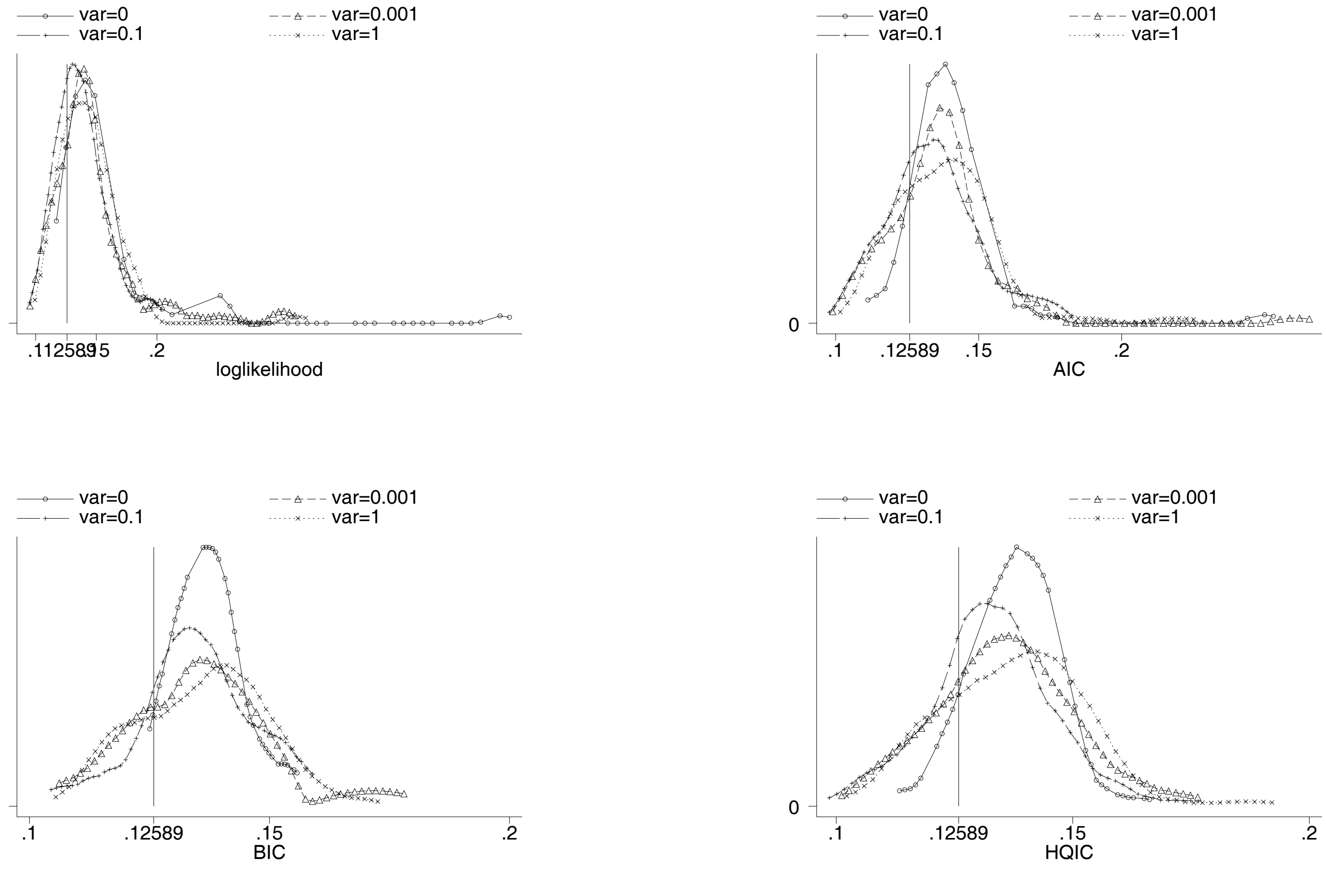
Weibull hazard, Gamma, 5000 obs

Figure A8: Kernel Densities of estimated expectation of $\hat{\mu}$ by calendar variations. Constant hazard, 3 points mixture, 5,000 individuals.



Constant hazard, 3 points, 5,000 obs

Figure A9: Kernel Densities of estimated expectation of $\hat{\mu}$ by calendar variations. Weibull hazard, Gamma mixture, 5,000 individuals.



Weibull hazard, Gamma, 5,000 obs

Figure A10: Kernel Densities of estimated expectation of $\hat{\mu}$ by sample sizes. Constant hazard, 3 points mixture, var(month)=0.1.

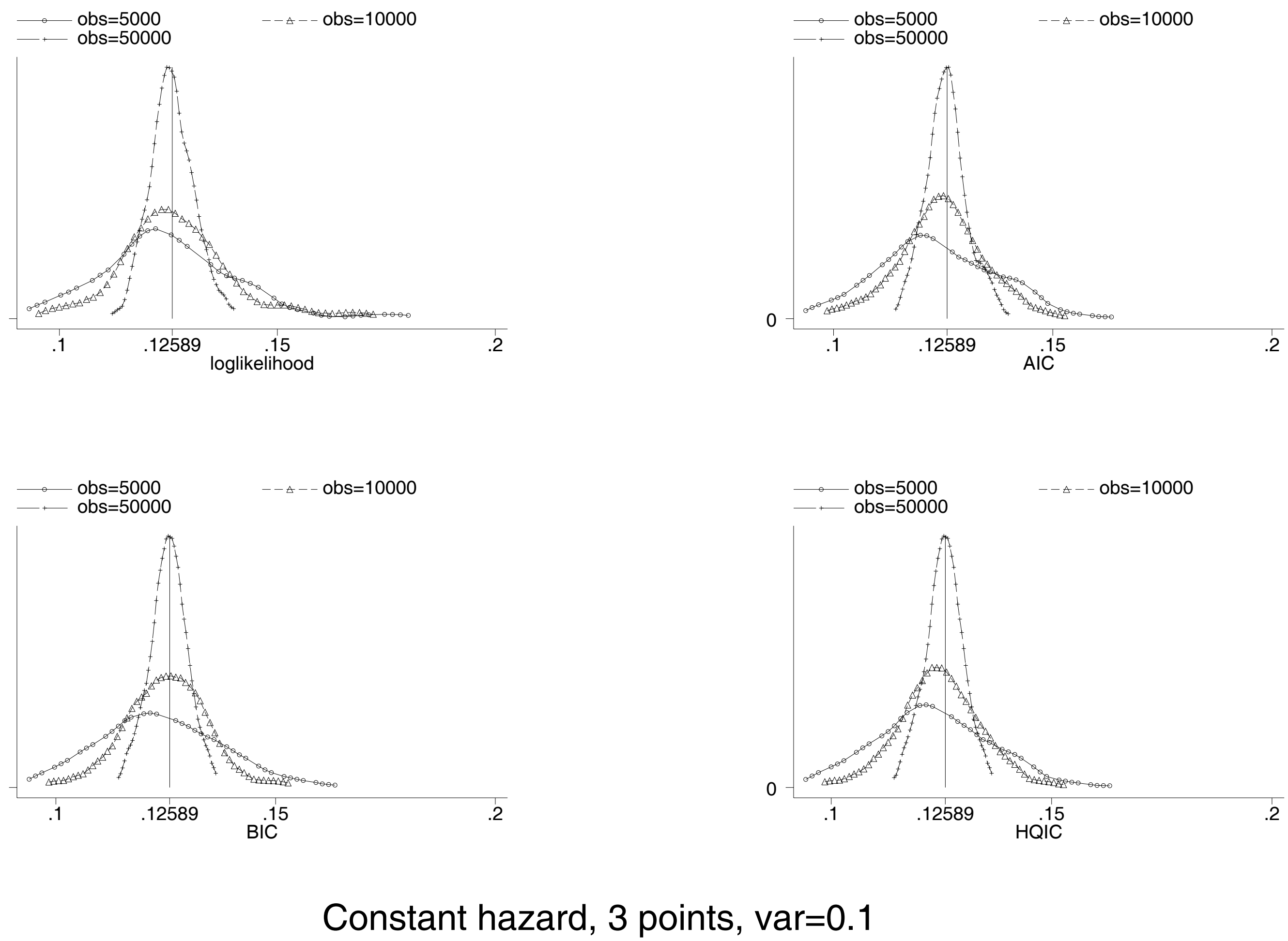


Figure A11: Kernel Densities of estimated expectation of $\hat{\mu}$ by sample sizes. Weibull hazard, Gamma mixture, var(month)=0.1.

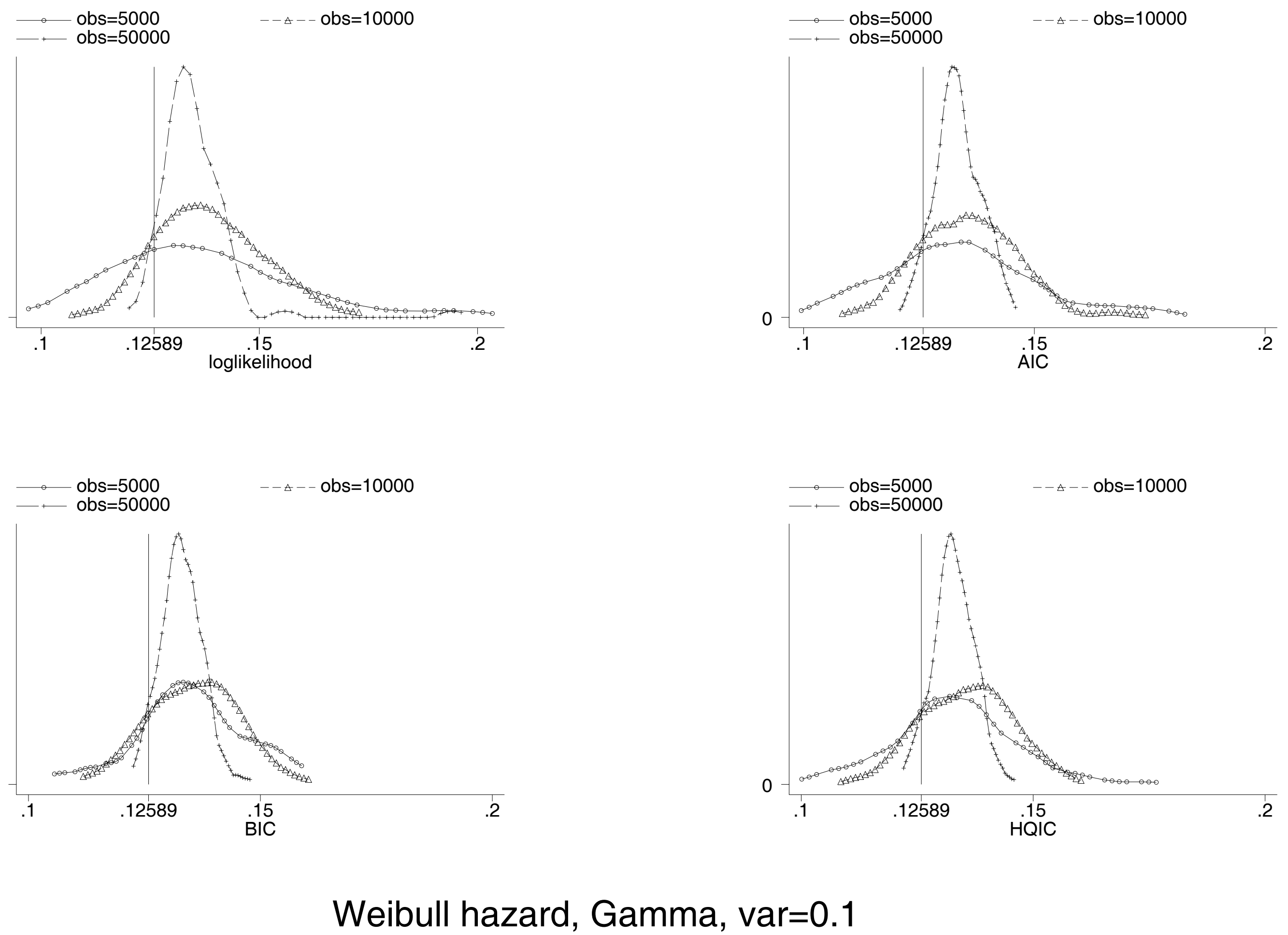


Figure A12-1: Kernel Densities of estimated $\hat{\beta}_1$. 10,000 individuals.

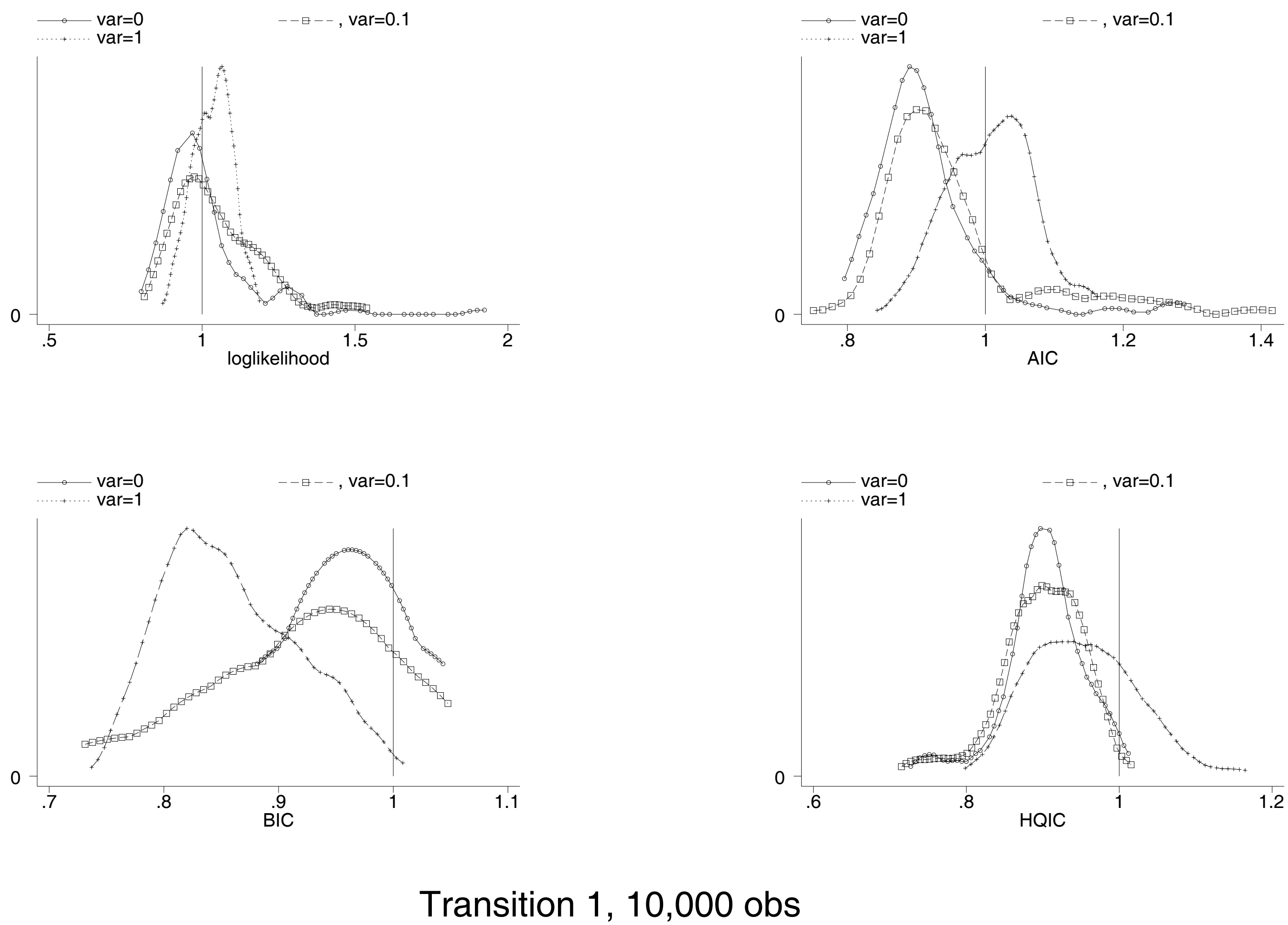
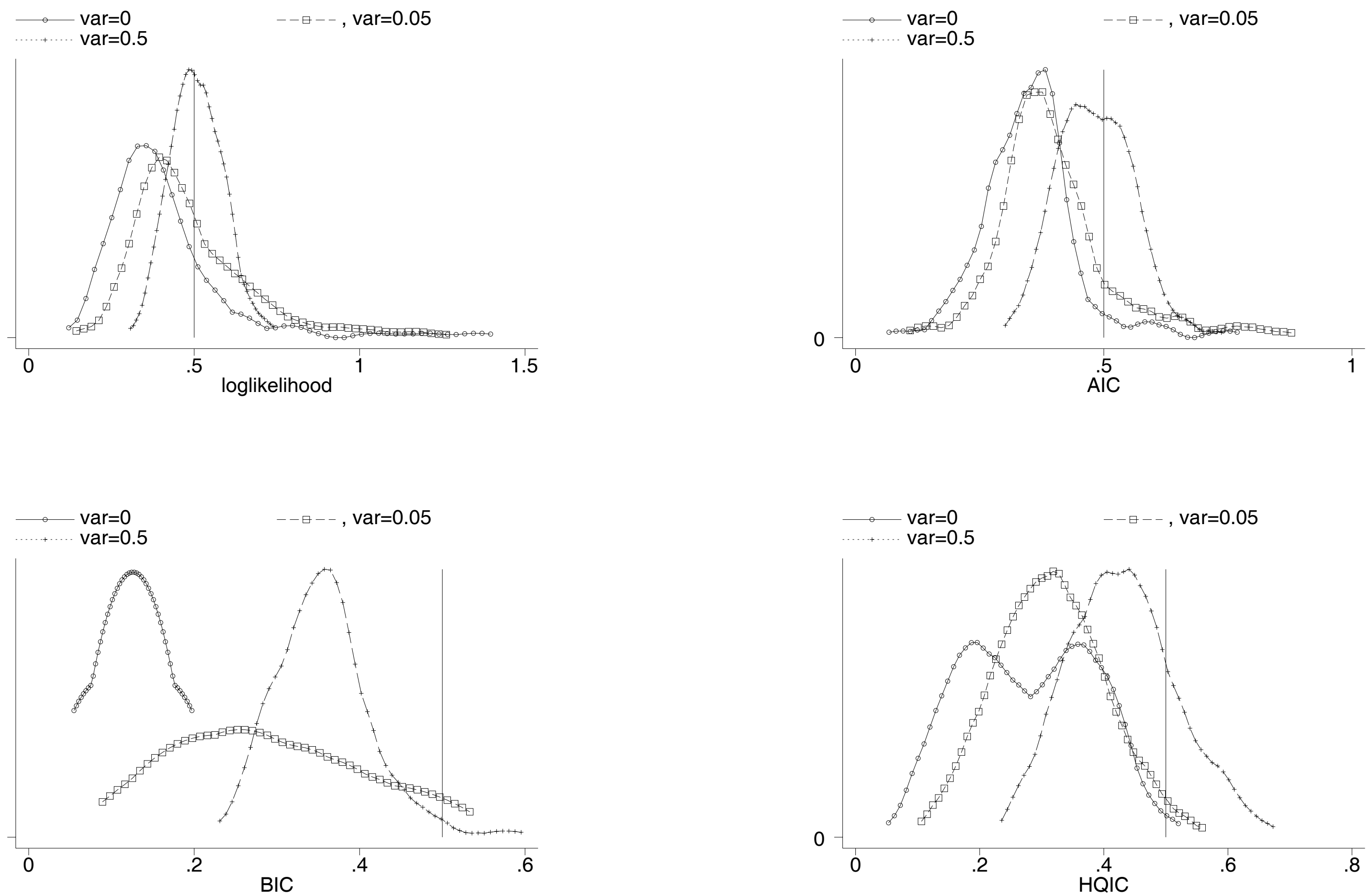
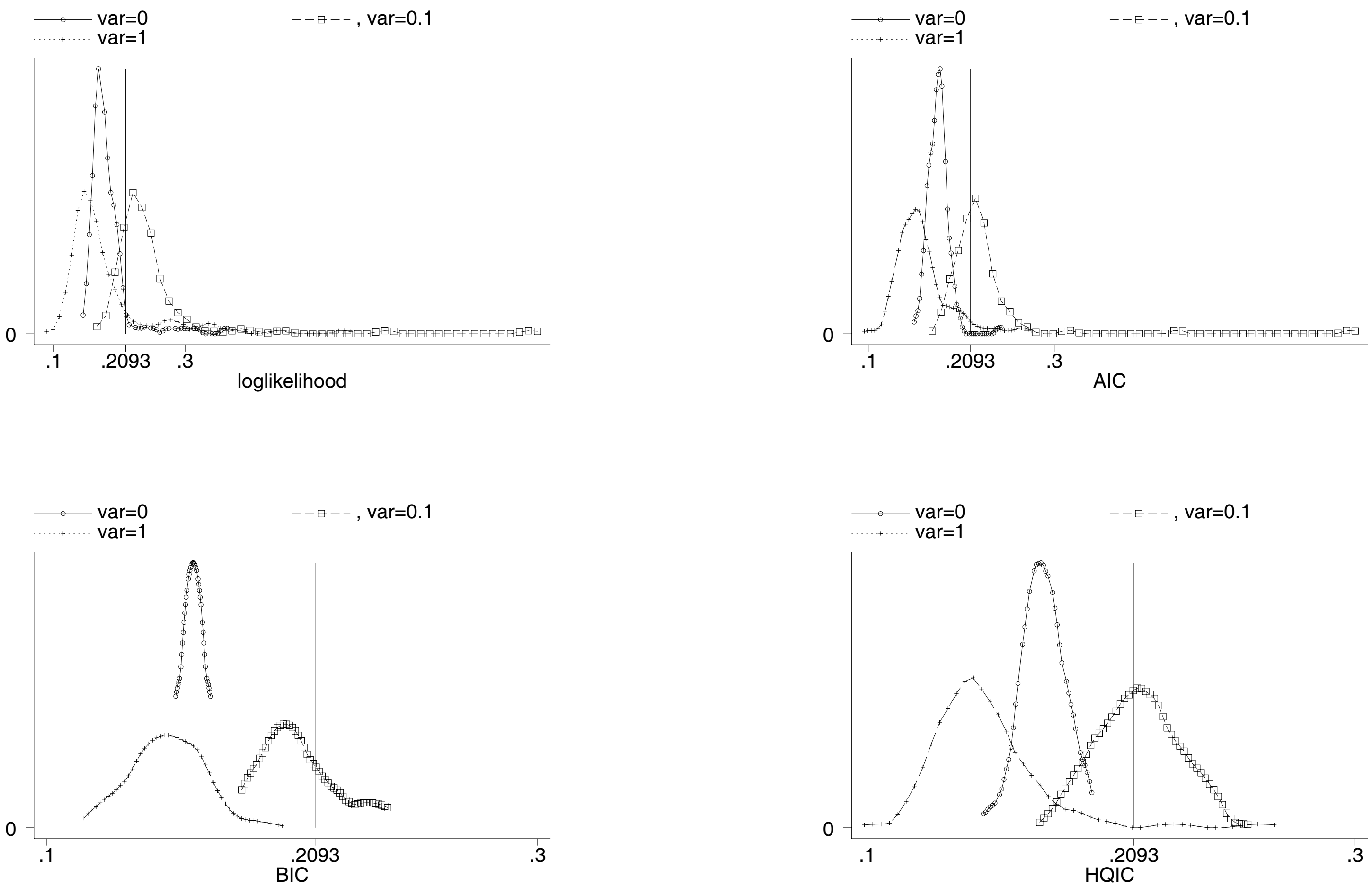


Figure A12-2: Kernel Densities of estimated $\hat{\beta}_2$. 10,000 individuals.



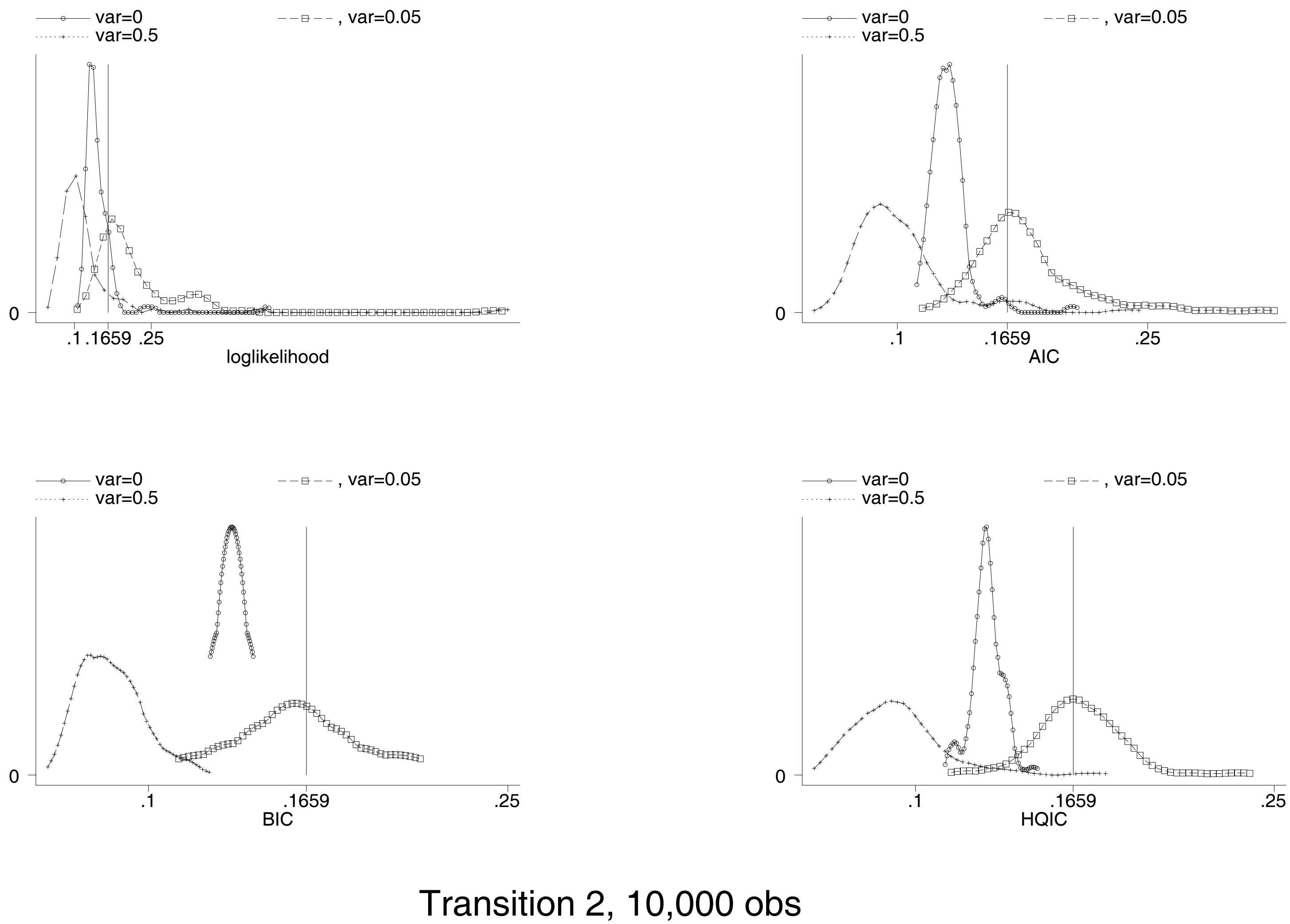
Transition 2, 10,000 obs

Figure A13-1: Kernel Densities of estimated $E(\hat{\mu}_1)$. 10,000 individuals.



Transition 1, 10,000 obs

Figure A13-2: Kernel Densities of estimated $E(\hat{\mu}_2)$. 10,000 individuals.



Frisch Centre Publications

All publications are available in Pdf-format at : www.frisch.uio.no

Rapporter (Reports)

1/1999	**Arbeidsledighet, arbeidsmarkedspolitikk og jobbsøking i Norge**	Knut Røed, Hege Torp, Tom Erik Aabø
2/1999	**Egenskaper ved tildelingsformer for nasjonale klimagasskvoter**	Rolf Golombek, Michael Hoel, Snorre Kverndokk, Ove Wolfgang
3/1999	**Regionale virkninger av økte elektrisitetspriser til kraftkrevende industri**	Nils-Henrik M. von der Fehr, Trond Hjørungdal
4/1999	**Bedriftsnedleggelse og klimakvoter i norsk industri**	Rolf Golombek, Arvid Raknerud
5/1999	**Utdanning og livsinntekt i Norge**	Oddbjørn Raaum, Tom Erik Aabø, Thomas Karterud
1/2000	**Hvem er de ledige? En økonometrisk analyse av arbeidsledighetens sammensetning i Norge på 1990-tallet**	Morten Nordberg
2/2000	**Effektivitet i pleie- og omsorgssektoren**	Dag F. Edvardsen, Finn R. Førsund, Eline Aas
3/2000	**Norge i liberalisert europeisk energimarked**	Finn Roar Aune, Rolf Golombek, Knut Einar Rosendahl, Sverre A.C. Kittelsen
4/2000	**Hvem vil og hvem får delta? Analyser av rekruttering og utvelgelse av deltakere til arbeidsmarkedstiltak i Norge på 1990-tallet**	Knut Røed, Hege Torp, Irene Tuveng, Tao Zhang
5/2000	**Deregulering av det vest-europeiske gassmarkedet - korttidseffekter**	Rolf Golombek, Sverre A.C. Kittelsen, Ove Wolfgang
6/2000	**Oversikt over litteratur om svart arbeid og skatteunndragelser**	Erling Eide
7/2000	**Arbeidstilbud i vedvarende gode tider**	Christian Brinch
8/2000	**Miljøreguleringer av norsk treforedlingsindustri**	Rolf Golombek, Arent Greve, Ken Harris
1/2001	**Analyse av inntektsfordeling og inntektsulikhet basert på registerdata. En kartlegging av**	Remy Åserud

	muligheter og begrensninger	
2/2001	**Miljøvirkninger av norsk eksport av gass og gasskraft**	Finn Roar Aune, Rolf Golombek, Sverre A.C. Kittelsen, Knut Einar Rosendahl
3/2001	**De statlige høgskolene som produsenter: Ressursbruk og resultater 1994-1999**	Dag Fjeld Edvardsen, Finn R. Førsund
4/2001	**Ragnar Frisch's bibliography**	Kåre Edvardsen
5/2001	**Tidlig arbeidsledighet og marginalisering**	Christer Af Geijerstad
1/2002	**En studie av fattigdom basert på registerdata**	Taryn Ann Galloway
2/2002	**Utstøting fra arbeidsmarkedet og tiltaksapparatets rolle**	Morten Nordberg, Knut Røed
3/2002	**Svart arbeid fra 1980 til 2001**	Tone Ognedal, Harald Goldstein, Wiljar G. Hansen, Steinar Strøm
4/2002	**Yrkesdeltakelse, pensjoneringsatferd og økonomiske insentiver**	Erik Hernæs, Knut Røed, Steinar Strøm
5/2002	**Miljø og nytte-kostnadsanalyse. Noen prinsipelle vurderinger**	Karine Nyborg
6/2002	**Transboundary environmental problems and endogenous technological change.** ***A survey with particular emphasize on the climate problem***	Maria Chikalova
7/2002	**Langsiktige samfunnsøkonomiske konsekvenser av kapasitetsproblember i dekningen av kraftetterspørselen**	Helge Dønnum, Rolf Golombek, Pål Føyn Jespersen, Snorre Kverndokk
1/2003	**En enkel dynamisk modell for skatteunndragelse med normer og rasjonell adferd**	Erling Eide
2/2003	**Marginalkostnader i jernbanenettet**	Øystein Børnes Daljord
3/2003	**Økonomiske prinsipper for fastsettelse av priser for jernbaneinfrastruktur**	Dag Morten Dalen, Nils-Henrik M. von der Fehr
4/2003	**Utnyttelse av vannkraftmagasiner**	Finn R. Førsund, Rolf Golombek, Michael Hoel, Sverre A.C. Kittelsen
1/2004	**Causality and Selection in Labour Market Transitions. Dissertation for the Dr.Polit Degree**	Tao Zhang

Arbeidsnotater (Working papers)

1/1999	**Kan markedskreftene temmes i lønnsdannelsen?**	Colin Forthun
2/1999	**Inntektseffekter av utdanning i Norge – en litteraturoversikt**	Oddbjørn Raaum
1/2000	**Empirical Specification of the Model in "Early Retirement and Economic Incentives"**	Erik Hernæs, Steinar Strøm
2/2000	**Forholdene på arbeidsmarkedet, økonomiske incentiver og risikoen for å bli yrkeshemmet**	Christian L. Wold Eide
3/2000	**Koordinering av inntektsoppgjørene i Norge og Sverige 1961-1999**	Bergljot Bjørnson Barkbu
4/2000	**Insentivvirkninger av skatte- og pensjonsregler**	Fredrik Haugen
5/2000	**Dynamisk arbeidstilbud**	Merethe Nordling
1/2001	**LIBEMOD – LIBEralisation MODel for the European Energy Markets: A Technical Description**	Finn Roar Aune, Rolf Golombek, Sverre A.C. Kittelsen, Ove Wolfgang
1/2002	**Forklaringer på forskjeller i effektivitet**	Finn R. Førsund, Dag Fjeld Edvardsen
2/2002	**Implisitte skattesatser i pensjonssystemet**	Guro Engstrøm Nilsen
1/2003	**AFP and OP data construction techniques**	Fedor Iskhakov, Maria Kalvarskaia

Memoranda (Discussion papers)

The series is published by Department of Economics, University of Oslo, in co-operation with the Frisch Centre. This list includes memoranda related to Frisch Centre projects.
The complete list of memoranda can be found at www.sv.uio.no/sosoek/memo/.

3/1999	**The Economics of Screening Programs**	Steinar Strøm
7/1999	**What hides behind the rate of unemployment? Micro evidence from Norway**	Knut Røed, Tao Zhang
9/1999	**Monte Carlo Simulations of DEA Efficiency Measures and Hypothesis Tests**	Sverre A.C. Kittelsen
11/1999	**Efficiency and Productivity of Norwegian Colleges**	Finn R. Førsund, Kjell Ove Kalhagen

13/1999	**Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies**	Tor Jakob Klette, Jarle Møen, Zvi Griliches
14/1999	**Unemployment Duration in a Non-Stationary Macroeconomic Environment**	Knut Røed, Tao Zhang
16/1999	**The effect of schooling on earnings: The role of family background studies by a large sample of Norwegian twins**	Oddbjørn Raaum, Tom Erik Aabø
17/1999	**Early Retirement and Economic Incentives**	Erik Hernæs, Marte Sollie, Steinar Strøm
18/1999	**Fewer in Number but Harder to Employ: Incidence and Duration of Unemployment in an Economic Upswing**	Erik Hernæs
19/1999	**Progressiv Taxes and the Labour Market**	Knut Røed, Steinar Strøm
22/1999	**Inequality, Social Insurance and Redistribution**	Karl Ove Moene, Michael Wallerstein
24/1999	**Do Voluntary Agreements Lead to Cost Efficiency**	Rolf Golombek, Espen R. Moen
25/1999	**Rent Grabbing and Russia's Economic Collapse**	Sheetal K. Chand and Karl Ove Moene
28/1999	**The role of foreign ownership in domestic environmental regulation under asymmetric infortmation**	Jon Vislie
29/1999	**Labor unions versus individualized bargaining with heterogeneous labor**	Jon Strand
32/1999	**Efficiency in the Provision of Municipal Nursing – and Home-Care Services: The Norwegian Experience**	Espen Erlandsen, Finn R. Førsund
33/1999	**Effects of Progressive Taxes under Decentralized Bargaining and Heterogeneous Labor**	Jon Strand
34/1999	**Reflections on Abatement Modelling**	Ove Wolfgang
35/1999	**Crime Induced Poverty Traps**	Halvor Mehlum, Karl Ove Moene, Ragnar Torvik
36/1999	**Statistical Discrimination and the Returns to Human Capital and Credentials**	Christian Brinch
38/1999	**Relative Unemployment Rates and Skill-Biased Technological Change**	Knut Røed
2/2000	**Married Men and Early Retirement Under the AFP Scheme**	Ole J. Røgeberg

4/2000	**Family Labor Supply when the Husband is Eligible for Early Retirement: Some Empirical Evidences**	Jia Zhiyang
5/2000	**Earnings Assimilation of Immigrants in Norway - A Reappraisal**	Pål Longva, Oddbjørn Raaum
9/2000	**Influencing bureaucratic Decisions**	Nils-Henrik von der Fehr, Lone Semmingsen
13/2000	**Family Labour Supply when the Husband is Eligible for Early Retirement**	Erik Hernæs, Steinar Strøm
15/2000	**Labour Market Transitions and Economic Incentives**	Knut Røed, Tao Zhang
16/2000	**Transboundary environmental problems with a mobile population: is there a need for a central policy**	Michael Hoel, Perry Shapiro
19/2000	**Have the Relative Employment Prospects for the Low-Skilled Deteriorated After All?**	Knut Røed, Morten Nordberg
23/2000	**A Note on the Weibull Distribution and Time Aggregation Bias**	Knut Røed, Tao Zhang
24/2000	**On The Origins of Data Envelopment Analysis**	Finn R. Førsund, Nikias Sarafoglou
27/2000	**Predator or Prey? Parasitic enterprises in economic development**	Halvor Mehlum, Karl Ove Moene, Ragnar Torvik
31/2000	**Genetic testing when there is a mix of public and private health insurance**	Michael Hoel, Tor Iversen
33/2000	**Competitive effort and employment determination with team production**	Jon Strand
34/2000	**CO2 mitigation costs and ancillary benefits in the Nordic countries, the UK and Ireland: A survey**	Snorre Kverndokk, Knut Einar Rosendahl
35/2000	**Tax distortions, household production and black market work**	Jon Strand
36/2000	**A turning point in the development of Norwegian economics - the establishment of the University Institute of Economics in 1932**	Olav Bjerkholt
40/2000	**Health Insurance: Treatment vs. Compensation**	Geir B. Asheim, Anne Wenche Emblem, Tore Nilssen
41/2000	**Private health care as a supplement to a public health system with waiting time for treatment**	Michael Hoel, Erik Magnus Sæther
11/2001	**Wage coordination and unemployment dynamics in Norway and Sweden**	Bergljot Bjørnson Barkbu, Ragnar Nymoen, Knut Røed

12/2001	**Temporary Layoffs and the Duration of Unemployment**	Knut Røed, Morten Nordberg
14/2001	**Liberalising the Energy Markets of Western Europe - A Computable Equilibrium Model Approach**	Finn Roar Aune, Rolf Golombek, Sverre A.C. Kittelsen, Knut Einar Rosendahl
23/2001	**Eco-Correlation in Acidification Scenarios**	Ove Wolfgang
24/2001	**Errors in Survey Based Quality Evaluation Variables in Efficiency Models of Primary Care Physicians**	Sverre A.C. Kittelsen, Guri Galtung Kjæserud, Odd Jarle Kvamme
26/2001	**Climate policies and induced technological change: Which to choose the carrot or the stick?**	Snorre Kverndokk, Knut Einar Rosendahl, Tom Rutherford
30/2001	**Cost-effective Abatement of Ground-level Ozone in Cities and for larger Regions: Implication of Non-monotonicity**	Ove Wolfgang
33/2001	**Labour Supply Effects of an Early Retirement Programme**	Christian Brinch, Erik Hernæs, Steinar Strøm
34/2001	**The Compensation Mechanism in the RAINS Model: The Norwegian Targets for Acidification**	Finn R. Førsund, Ove Wolfgang
35/2001	**International Benchmarking of Electricity Distribution Utilities**	Finn R. Førsund, Dag Fjeld Edvardsen
36/2001	**The neighbourhood is not what it used to be: Has there been equalisation of opportunity across families and communities in Norway?**	Oddbjørn Raaum, Kjell G. Salvanes, Erik O. Sørensen
3/2002	**Explaining Variations in Wage Curves: Theory and Evidence**	Erling Barth, Bernt Bratsberg, Robin A. Naylor, Oddbjørn Raaum
6/2002	**The Duration and Outcome of Unemployment Spells-The role of Economic Incentives**	Knut Røed, Tao Zhang
7/2002	**Characterization and Measurement of Duration Dependence in Hazard Rates Models**	Rolf Aaberge
9/2002	**Unemployment Duration, Incentives and Institutions - A Micro-Econometric Analysis Based on Scandinavian Data**	Knut Røed, Peter Jensen and Anna Thoursie
10/2002	**Plunder & Protections Inc.**	Halvor Mehlum, Kalle Moene, Ragnar Torvik
11/2002	**Battlefields and Marketsplaces**	Halvor Mehlum, Kalle Moene
12/2002	**Do Business Cycle Conditions at the Time of Labour**	Oddbjørn Raaum and Knut

	Market Entry Affect Future Unemployment?	Røed
14/2002	**Business cycles and the impact of labour market programmes**	Oddbjørn Raaum, Hege Torp and Tao Zhang
15/2002	**Do individual programme effects exceed the costs? Norwegian evidence on long run effects of labour market training**	Oddbjørn Raaum, Hege Torp and Tao Zhang
19/2002	**Local Unemployment and the Earnings Assimilation of Immigrants in Norway**	Erling Barth, Bernt Bratsberg and Oddbjørn Raaum
20/2002	**Local Unemployment and the Relative Wages of Immigrants: Evidence from the Current Population Surveys**	Erling Barth, Bernt Bratsberg and Oddbjørn Raaum
26/2002	**At Last! An Explicit Solution for the Ramsey Saddle Path**	Halvor Mehlum
28/2002	**Public-good valuation and intrafamily allocation**	Jon Strand
29/2002	**Institutions and the resource curse**	Halvor Mehlum, Kalle Moene, Ragnar Torvik
3/2003	**Optimal Provision of Public Goods with Rank Dependent Expected Utility**	Erling Eide
4/2003	**A Mixture Model of Household Retirement Choice**	Zhiyang Jia
7/2003	**Assimilation Effects on Poverty Among Immigrants in Norway**	Taryn Ann Galloway, Rolf Aaberg
9/2003	**How Tight is the Labour Market? A Micro-Based Macro Indicator**	Simen Gaure, Knut Røed
10/2003	**Retirement in Italy and Norway**	Ugo Colombino, Erik Hernæs, Zhyiang Jia, Steinar Strøm
13/2003	**The Effect of Programme Participation on the Transition Rate from Unemployment to Employment**	Knut Røed, Oddbjørn Raaum
16/2003	**Parasites**	Halvor Mehlum, Karl-Ove Moene, Ragnar Torvik
22/2003	**Climate Policy under Technology Spillovers**	Rolf Golombek, Michael Hoel
23/2003	**Savings behaviour when households have an access to occupational pension**	Maria Kalvarskaia
24/2003	**The Role of Warnings in Regulation: Keeping control with less punishment**	Karine Nyborg, Kjetil Telle
25/2003	**A Monte Carlo study on non-parametric estimation of duration models with unobserved heterogeneity**	Tao Zhang

26/2003	**Identifying treatment effects of active labour market programmes for Norwegian adults**	Tao Zhang
28/2003	**Quasi-dynamic forward-looking model for joint household retirement decision under AFP scheme**	Fedor Iskhakov
31/2003	**Green consumers and public policy: On socially contingent moral motivation**	Karine Nyborg, R.B. Howarth, Kjell Arne Brekke
33/2003	**The Effects of Taxes and Socioeconomic Variables on Market Work and Home Production in Norway in the Years 1970 to 2000**	Christoph Schwierz
34/2003	**Does Increased Extraction of Natural Gas Reduce Carbon Emissions?**	Finn Roar Aune, Rolf Golombek, Sverre A.C. Kittelsen
2/2004	**A dissolving paradox: Firms' compliance to environmental regulation**	Karine Nyborg, Kjetil Telle

The Frisch Centre

The Ragnar Frisch Centre for Economic Research is an independent research institution founded by the University of Oslo. The Frisch Centre conducts economic research in co-operation with the Department of Economics, University of Oslo. The research projects are mostly financed by the Research Council of Norway, government ministries and international organisations. Most projects are co-operative work involving the Frisch Centre and researchers in other domestic and foreign institutions.

Ragnar Frisch Centre for Economic Research
Gaustadalléen 21
N-0349 Oslo, Norway
T + 47 22 95 88 10
F + 47 22 95 88 25
frisch@frisch.uio.no
www.frisch.uio.no