MEMORANDUM

No 25/2003

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

Tao Zhang

ISSN: 0801-1117

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.oekonomi.uio.no/
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no/
e-mail: frisch@frisch.uio.no

List of the last 10 Memoranda:

No 24	Karine Nyborg and Kjetil Telle The Role of Warnings in Regulation: Keeping Control with Less Punishment. 31 pp.
No 23	Maria Kalvarskaia Savings behaviour when households havean access to occupational pensions. 50 pp.
No 22	Rolf Golombek and Michael Hoel Climate Policy under Technology Spillovers. 37 pp.
No 21	Jo Thori Lind Fractionalization and the size of government. 50 pp.
No 20	Michael Hoel, Tor Iversen, Tore Nilssen and Jon Vislie Genetic testing and repulsion from chance. 20pp.
No 19	Jon Vislie Domestic Environmental Policy under Asymmetric Information: The role of foreign ownership, outside options and market power 34 pp.
No 18	Karl Ove Moene and Michael Wallerstein Income Inequality and Welfare spending: A disaggregated Analysis. 53 pp.
No 17	Siwan Anderson, Jean-Marie Baland and Karl Ove Moene Sustainability and organizational design in informal groups, with some evidence from Kenyan Roscas. 49 pp.
No 16	Halvor Mehlum, Karl Ove Moene and Ragnat Torvik Parasites 19 pp.
No 15	Michael Wallerstein and Karl Ove Moene Does the Logic of Collective Action Explain the Logic of Corporatism?. 42 pp.

A complete list of this memo-series is available in a PDF® format at:
http://www.oekonomi.uio.no/memo/

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

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 $[d\text{-}1, d]$ (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 $[d\text{-}1, d]$, 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 $[d\text{-}1, d]$ 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 $[d\text{-}1, d]$ 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[1]. 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

[1] 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.[2] 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.

[2] 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 $\lambda = 0.10,\ \ \alpha = 0.90$	-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

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

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[3]. 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

[3] 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[4]. 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

[4] 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[5]. 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

[5] 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.

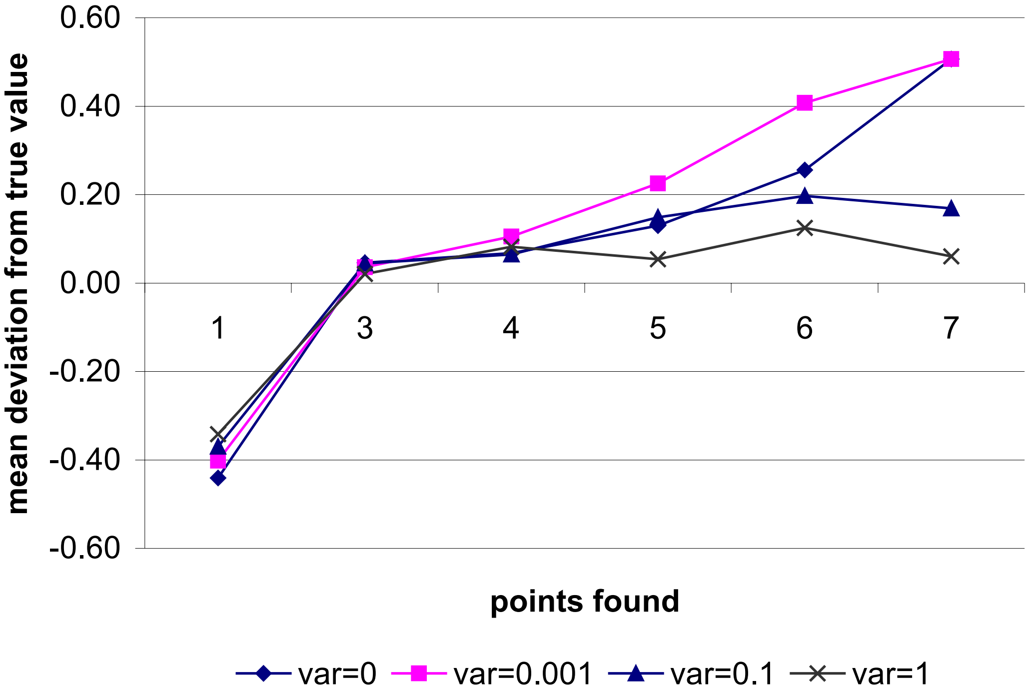
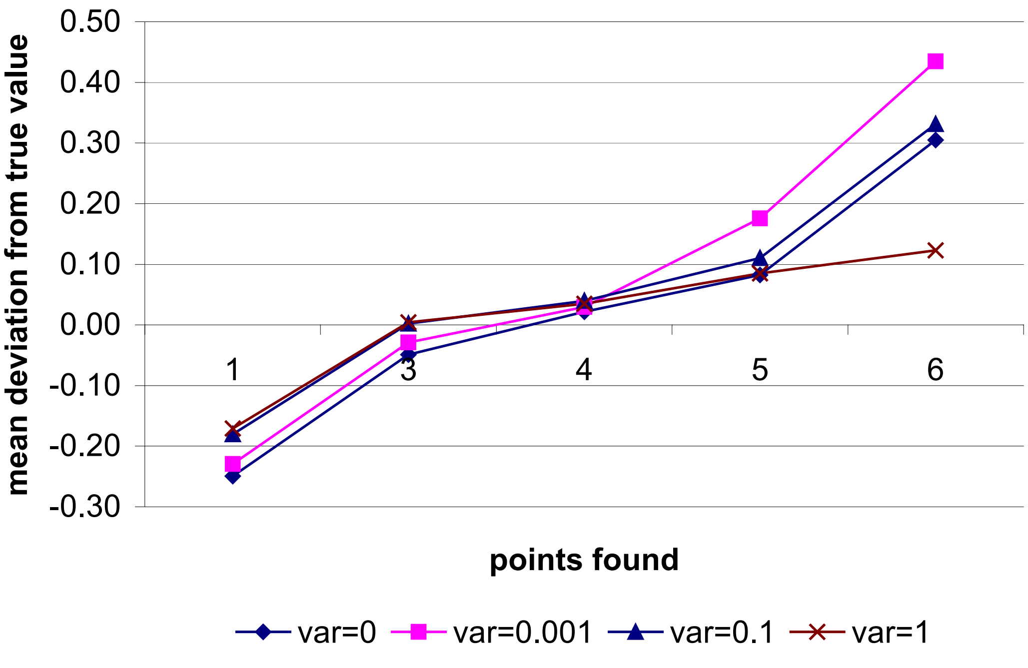


Figure 2: Mean deviations of estimated $\hat{\beta}$ from true value 1 by maximum number of mass points found by maximum loglikelihood. Weibull hazard, Gamma mixture in DGP. var represents calendar variation in DGP. Obs=5,000.



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[6].

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

[6] 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)[7], 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

[7] 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[8]. 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.

[8] 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.

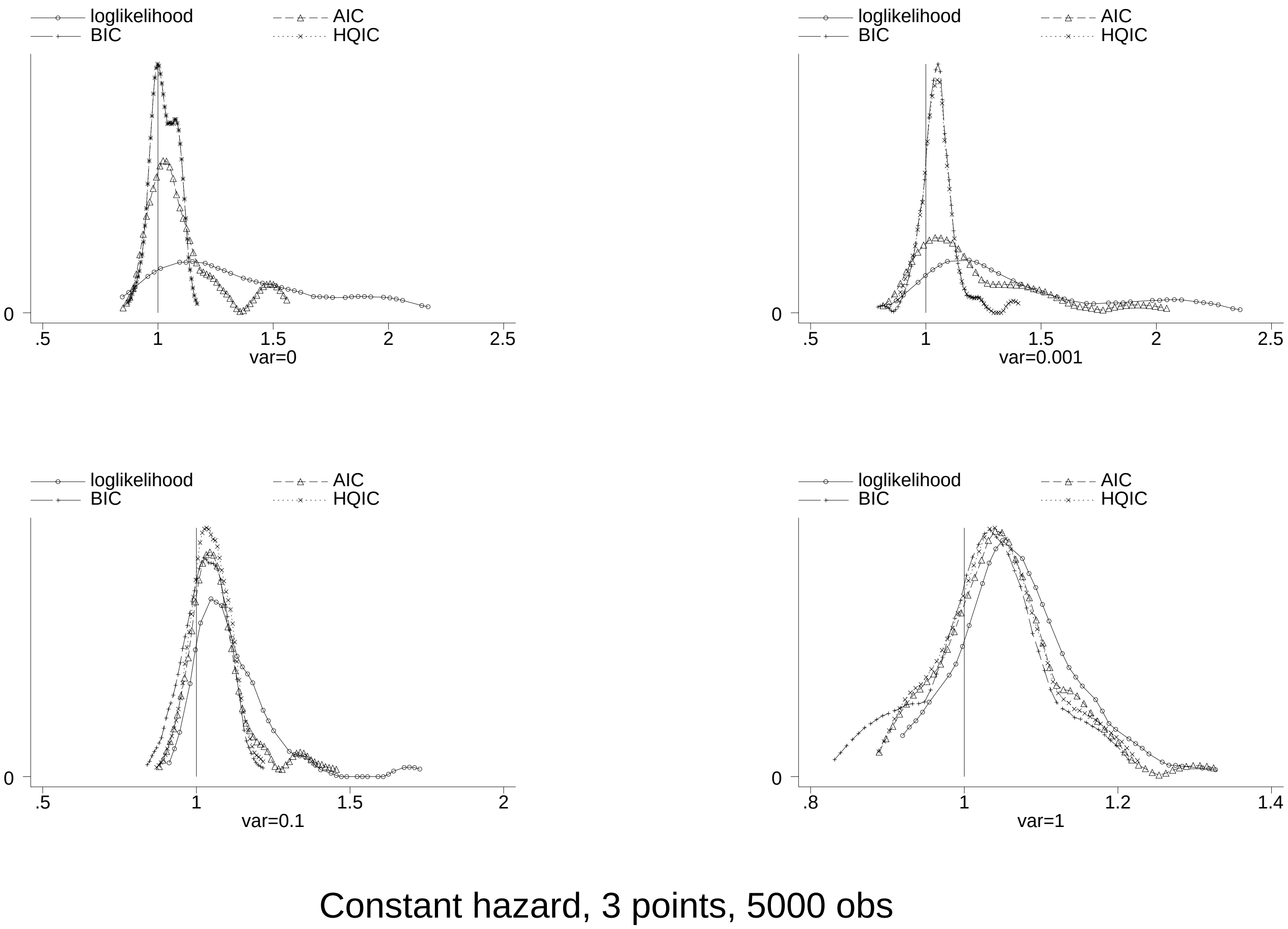
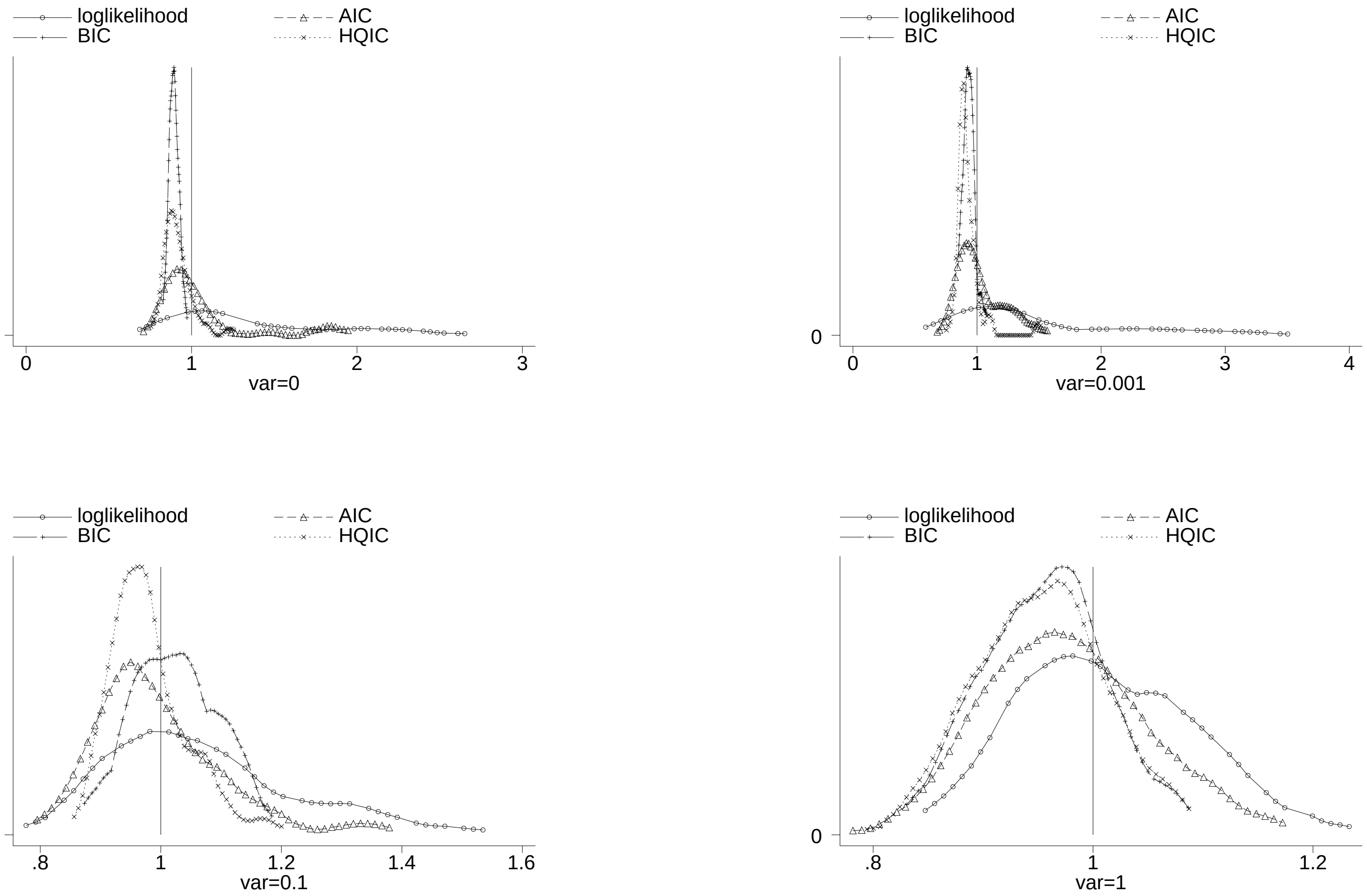


Figure 4: Kernel densities of estimated β. Weibull hazard, Gamma mixture, 5,000 individuals.



Weibull hazard, Gamma, 5000 obs

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

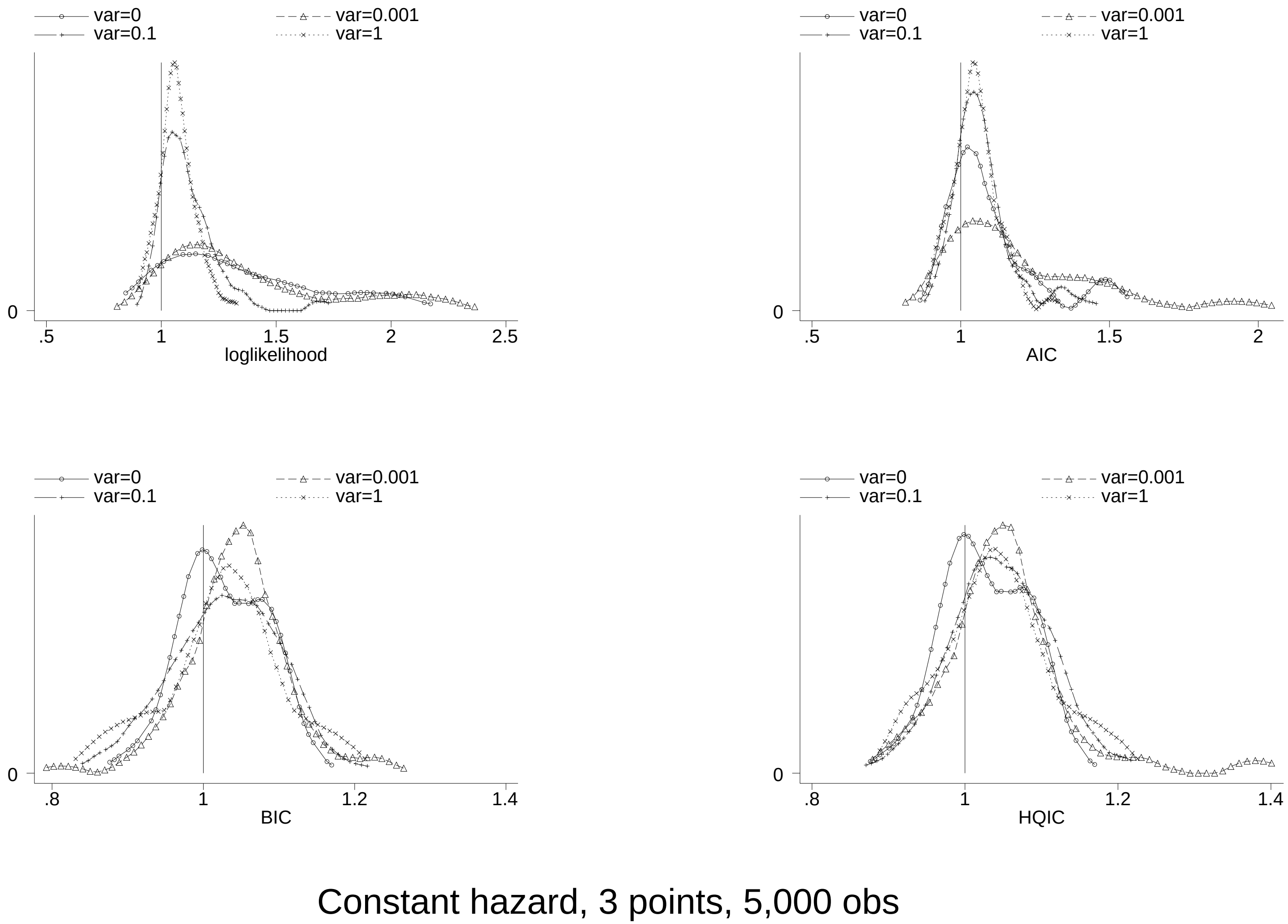
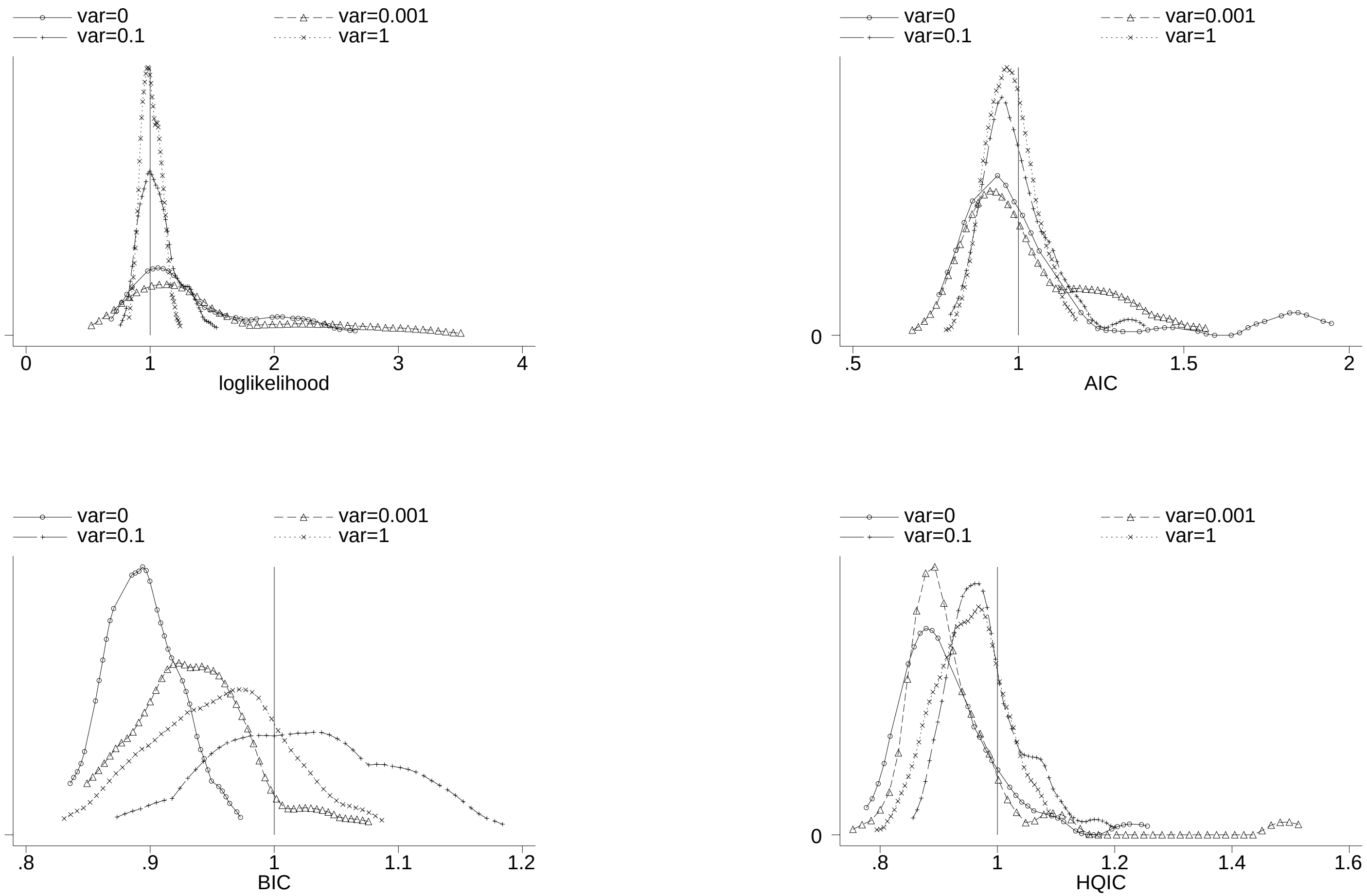
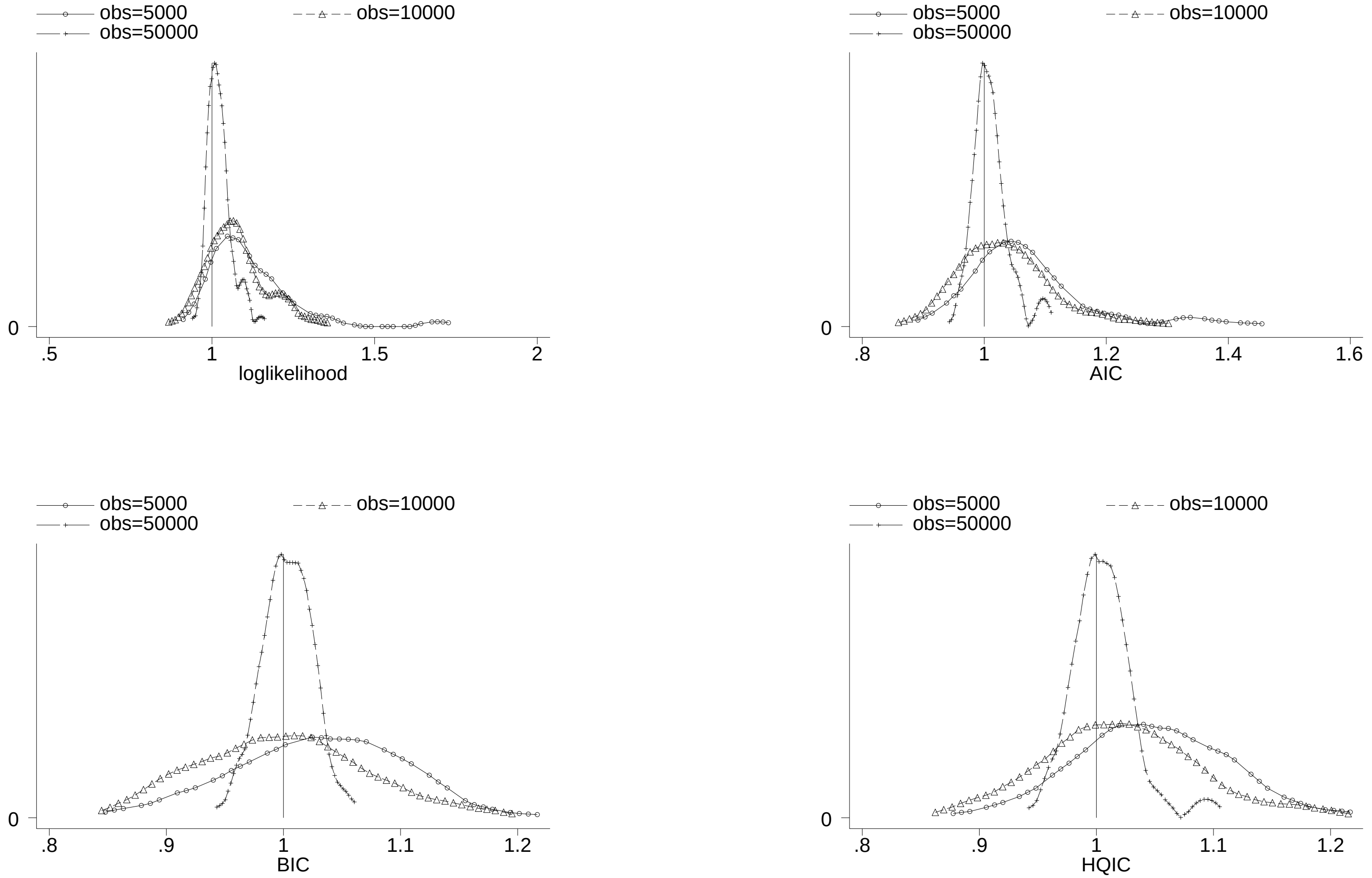


Figure 6: Kernel densities of estimated β by calendar variations. Weibull hazard, Gamma mixture, 5,000 individuals.



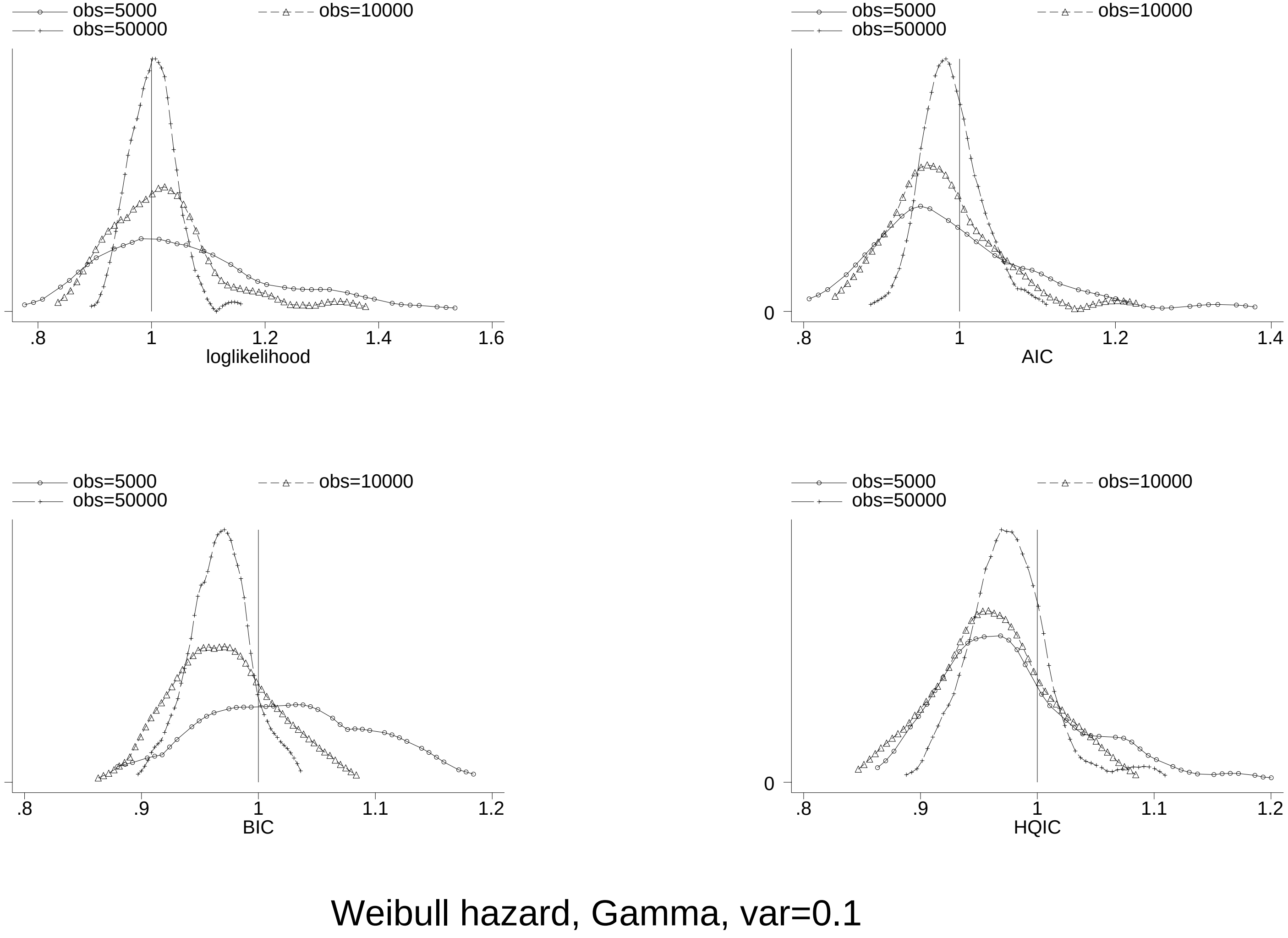
Weibull hazard, Gamma, 5,000 obs

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.



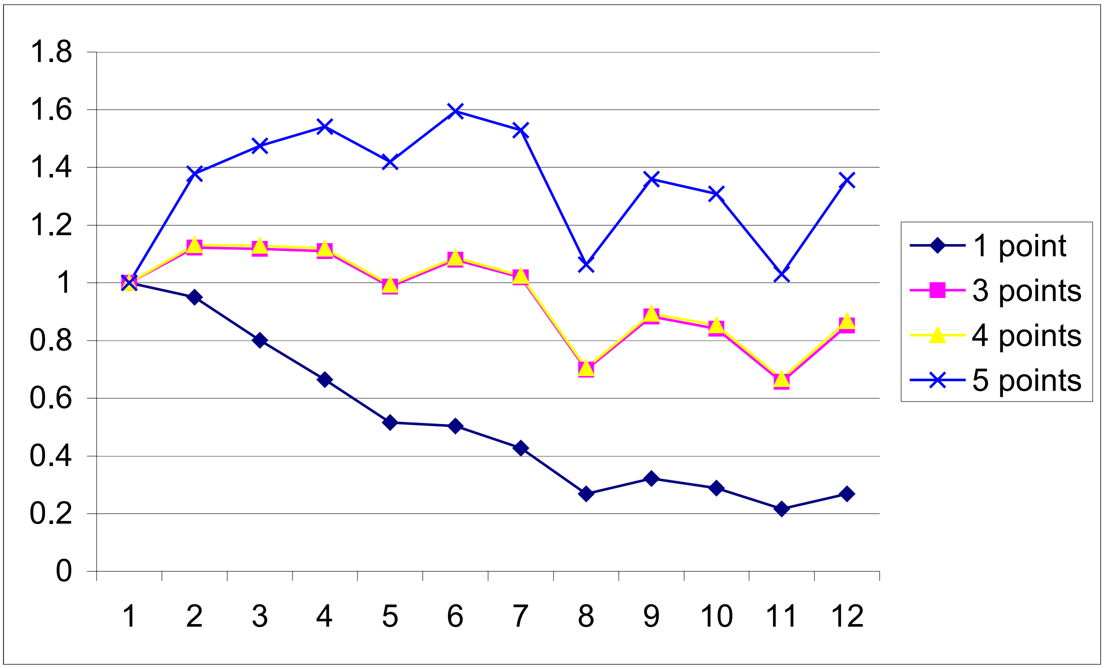
Figures 5 and 6 look into the effect of calendar variation on the estimated $\hat{\beta}$. One could come to the conclusion from figures that BIC and HQIC are less sensitive towards calendar variations than loglikelihood. At least for constant hazard model with 3 points distributed unobserved heterogeneity, 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

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**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

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**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[9]. 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.

[9] 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

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

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

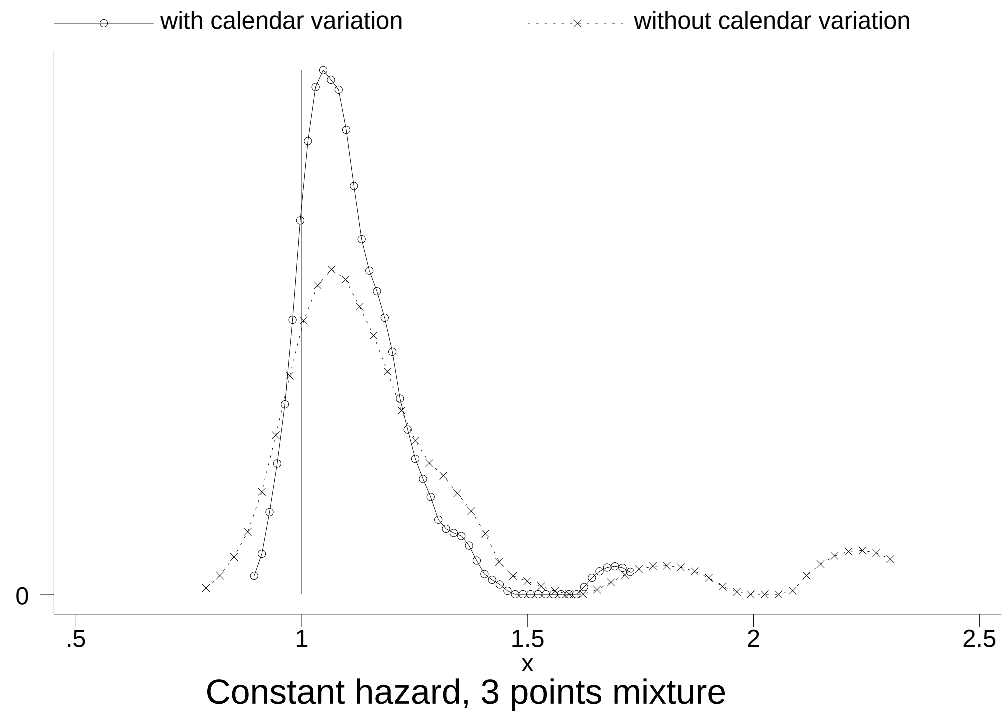
		Loglikelihood		**AIC**		**BIC**		**HQIC**	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**
5000	0	-	-	-	-	-	-	-	-
	0.001	0.4398	1.5500	0.0151	0.0116	0.0231	0.0078	0.0156	0.0117
	0.1	0.1245	0.2989	0.0144	0.0063	0.0129	0.0039	0.0135	0.0049
	1	0.0297	0.0306	0.0193	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.

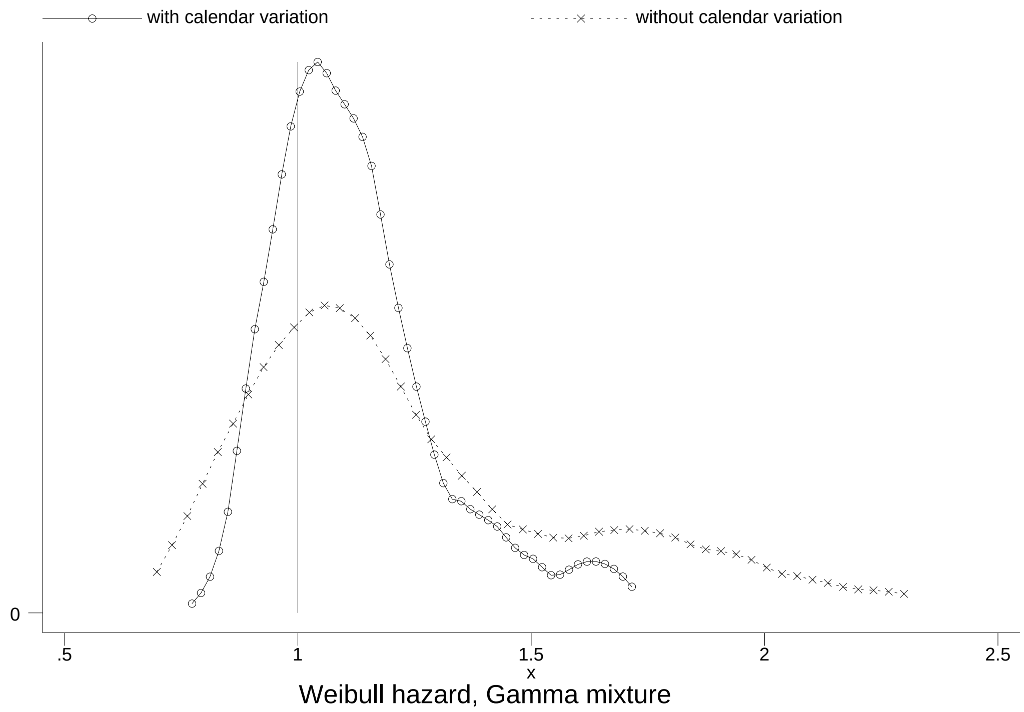
These estimated monthly hazard rates characterise the business and seasonal cycle conditions, as Gaure and Røed (2003) point out. As suggested earlier, the information on labour market conditions during the elapsed time of active spells would contribute to the identification of unobserved personal characteristics. 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[10]. 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,

[10] 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:

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

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-invaryin 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 $\mathbf{N}(0,1)$. 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 A13-1 and A13-2, 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 A14-1 and A14-2, 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

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**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

		Loglikelihood		AIC		BIC		HQIC	
# obs	**var(month)**	**mean**	**std.**	**mean**	**std.**	**mean**	**std.**	**mean**	**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

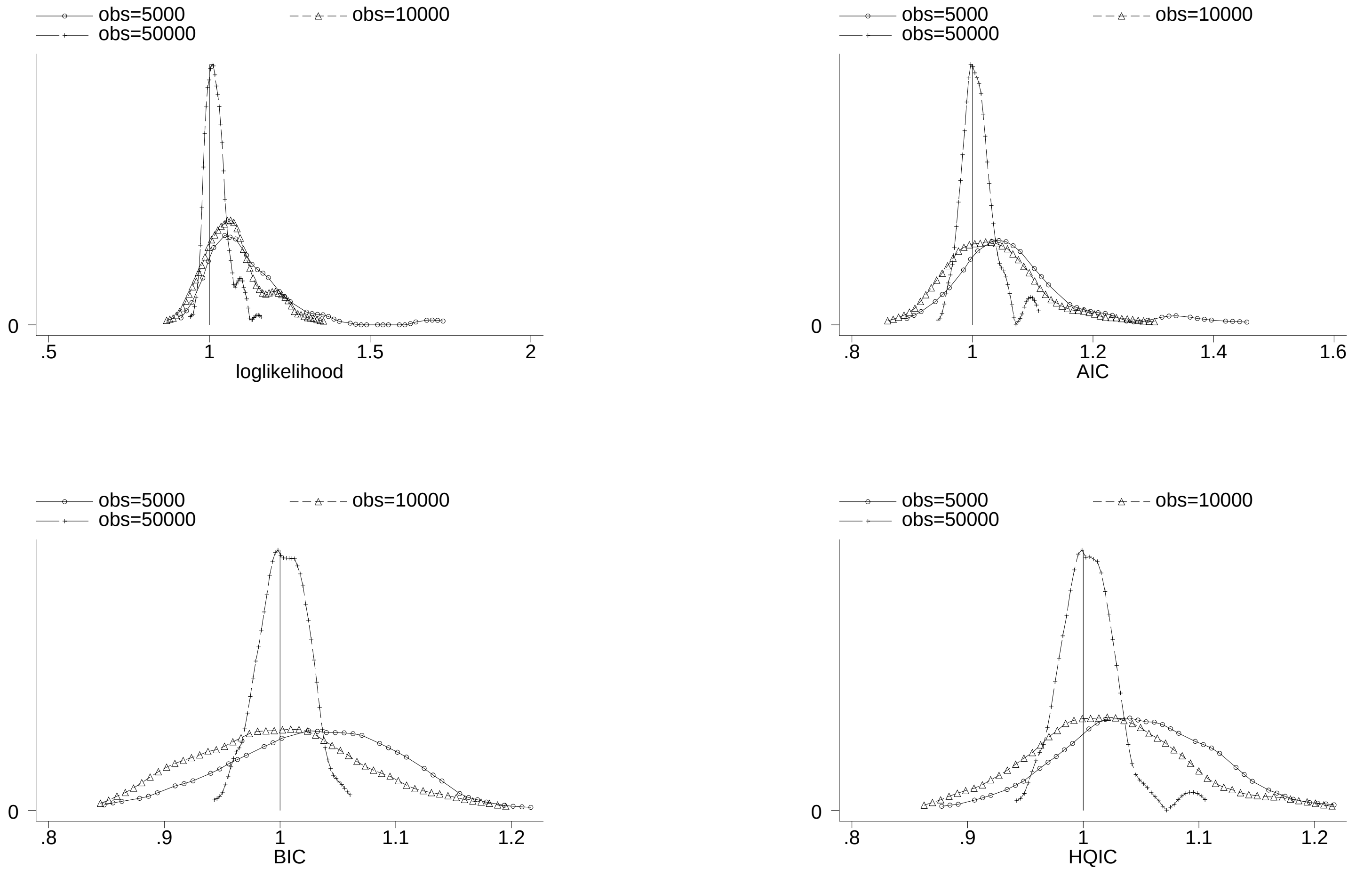
		Loglikelihood		AIC		BIC		HQIC	
# obs	var(month)	mean	std.	mean	std.	mean	std.	mean	std.
5000	0	0.0446	0.0663	0.0237	0.0117	0.0197	0.0021	0.0215	0.0039
	0.001	0.2027	1.5527	0.0274	0.0198	0.0187	0.0031	0.0222	0.0048
	0.1	0.0606	0.0967	0.0274	0.0158	0.0237	0.0039	0.0241	0.0048
	1	0.0387	0.0513	5.0320	45.0592	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.



Constant hazard, 3 points, var=0.1

Figure A2: Kernel densities of estimated β by sample sizes. Weibull hazard, Gamma mixture, var(month)=0.1.

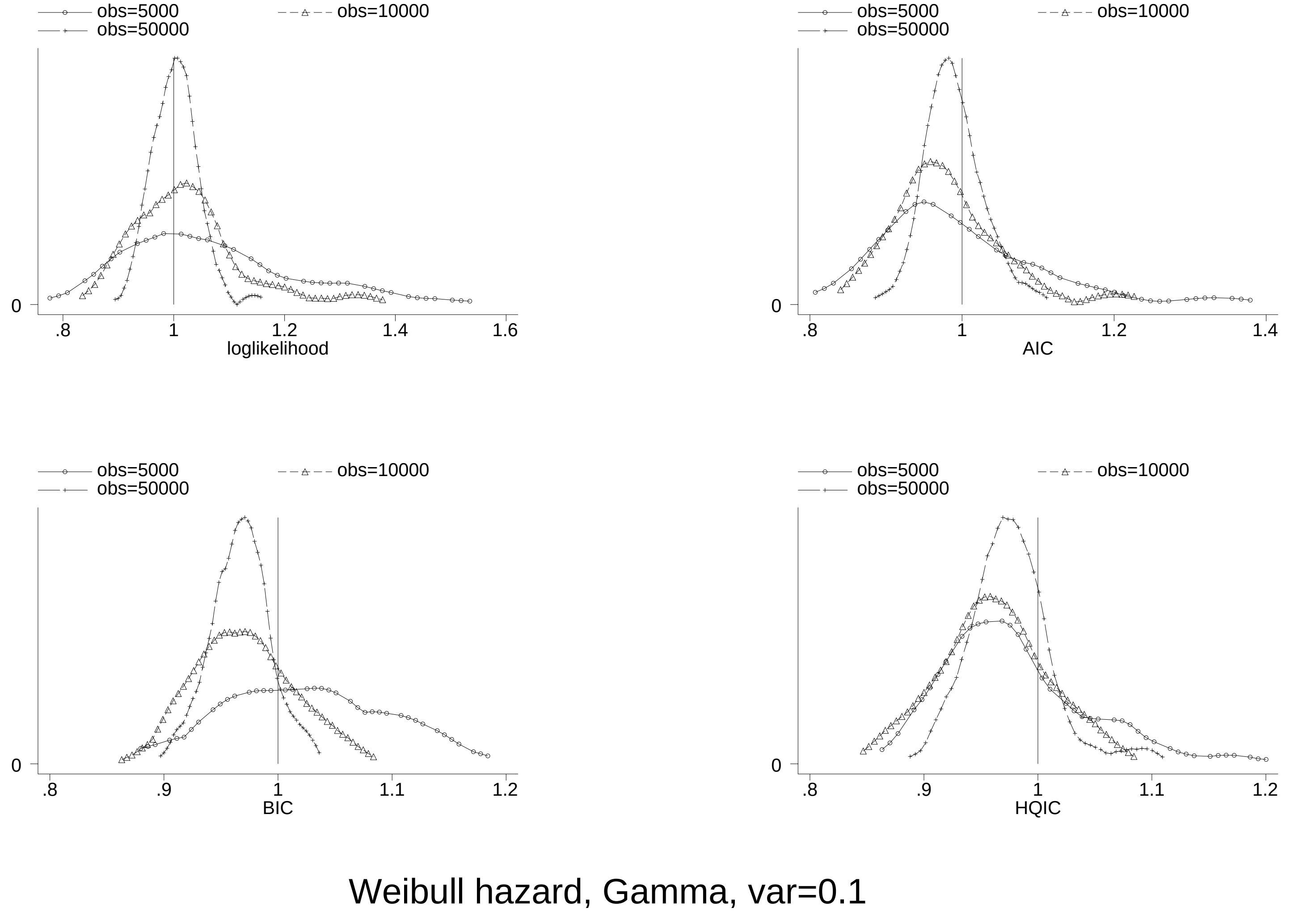
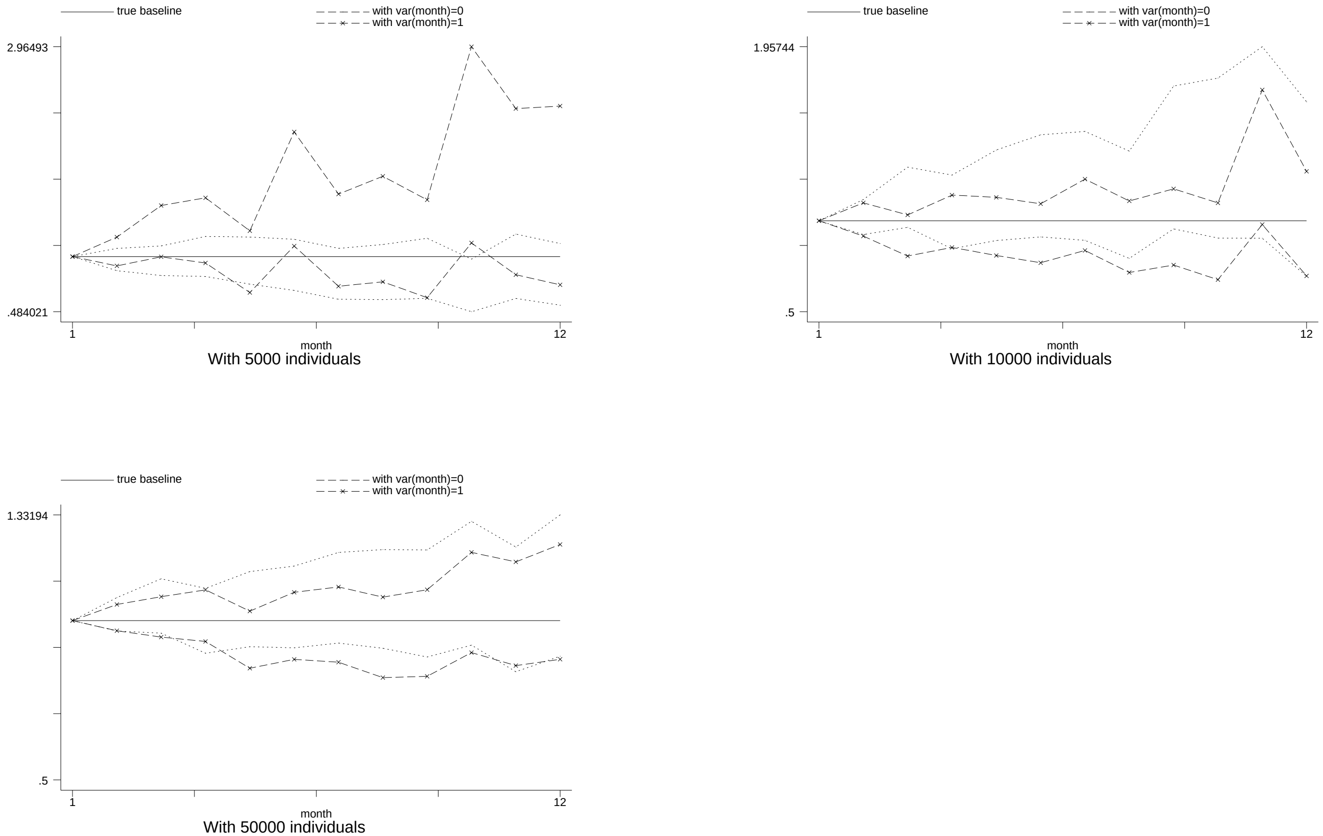
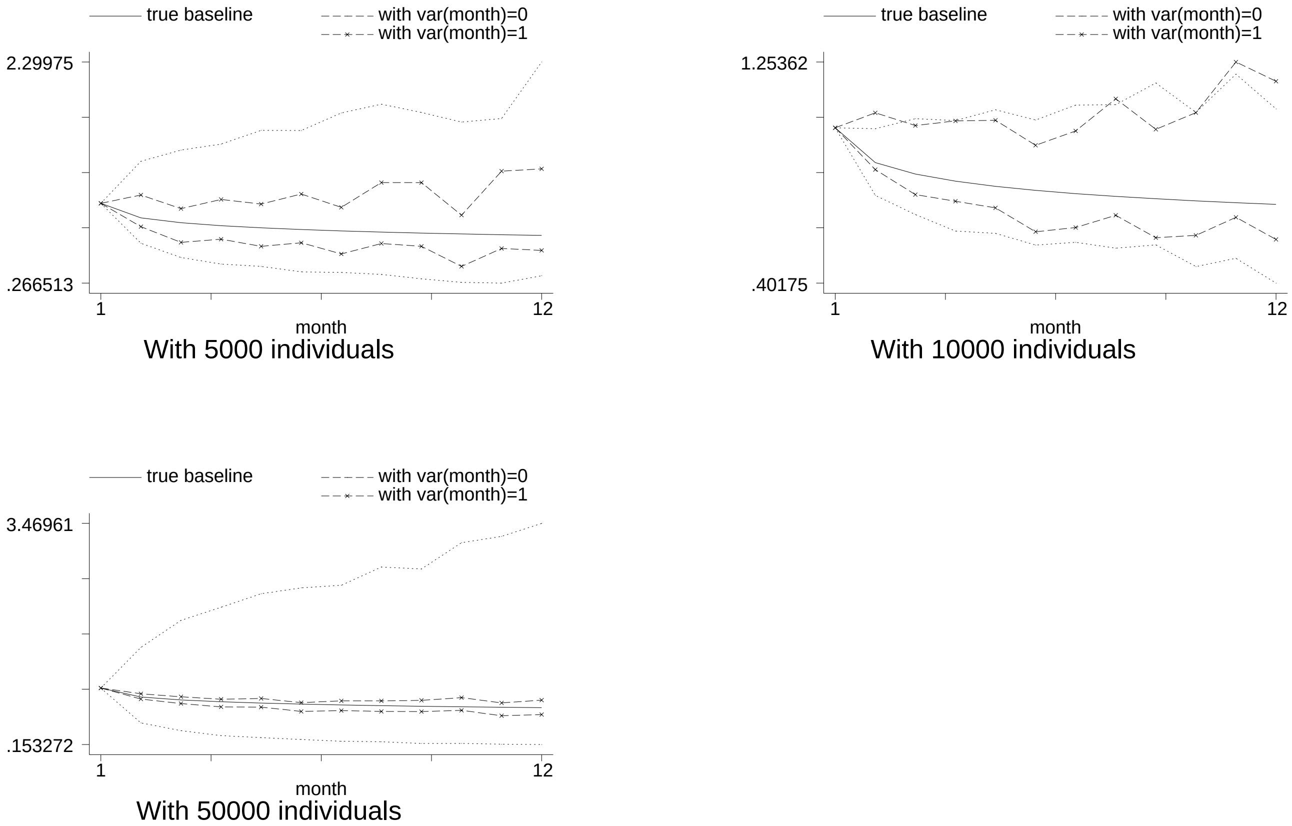


Figure A3: 95% Confidence intervals for baseline hazard rate estimates, across calendar variations. Constant hazard, 3 points mixture.



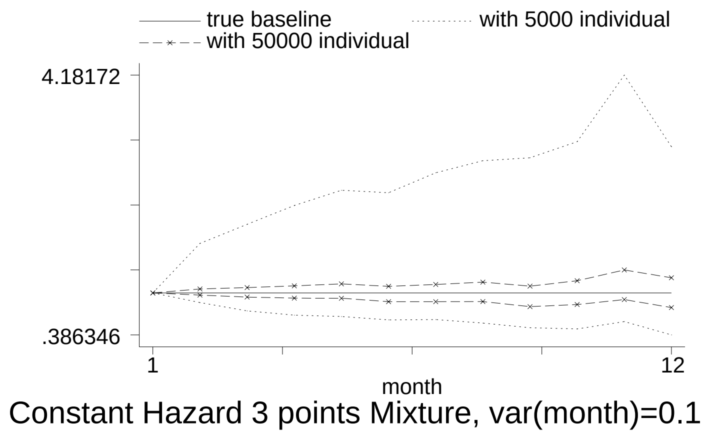
Note: confidence intervals are calculated based on the estimated standard errors (in exponential form) for the duration baseline estimators from the estimations that produce Table 4-1. Therefore they do not have the interpretation as confidence intervals for transition probabilities.

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.



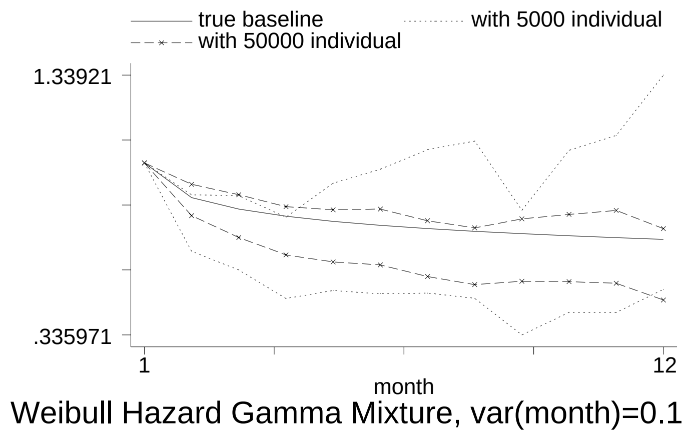
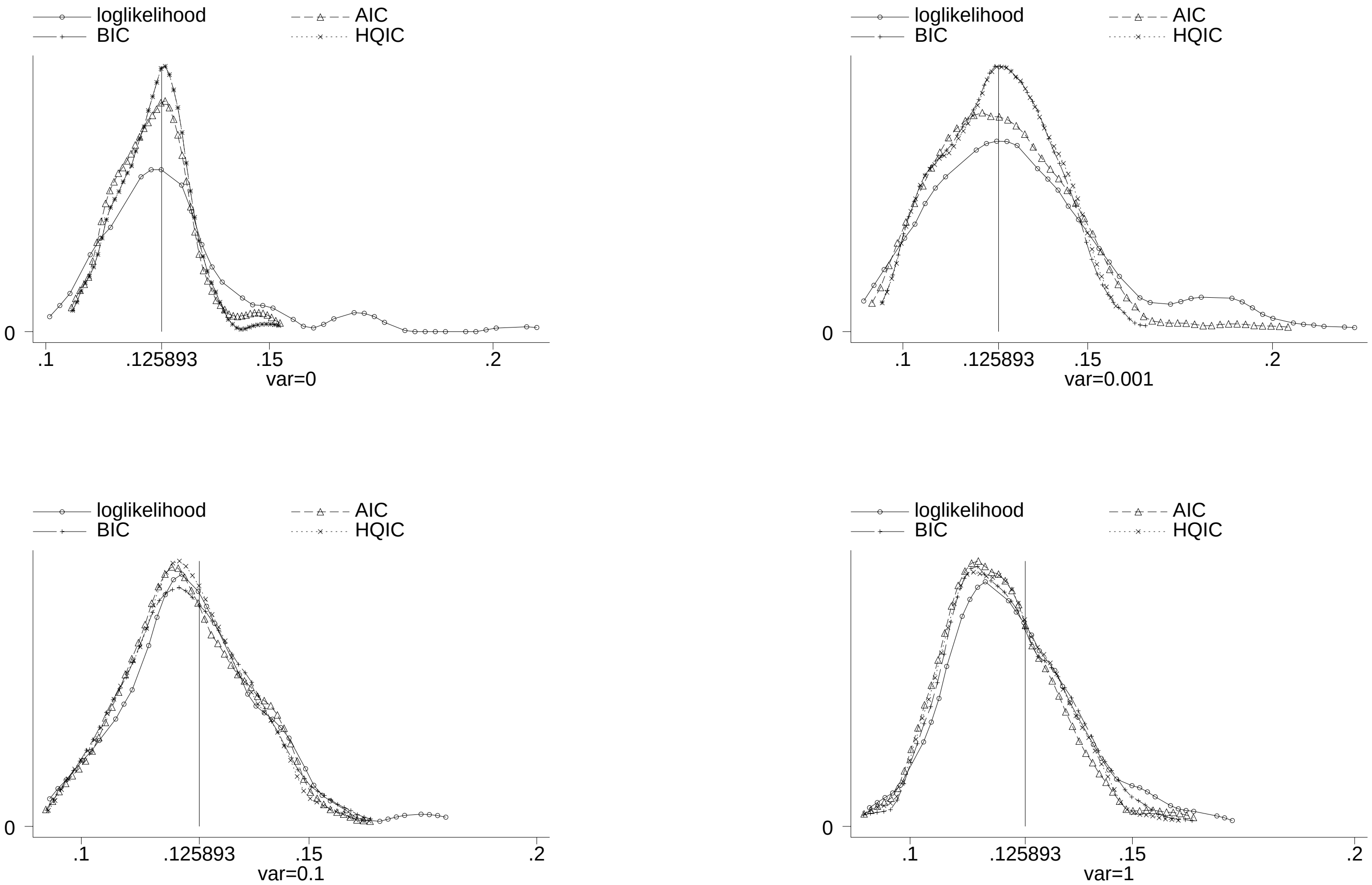
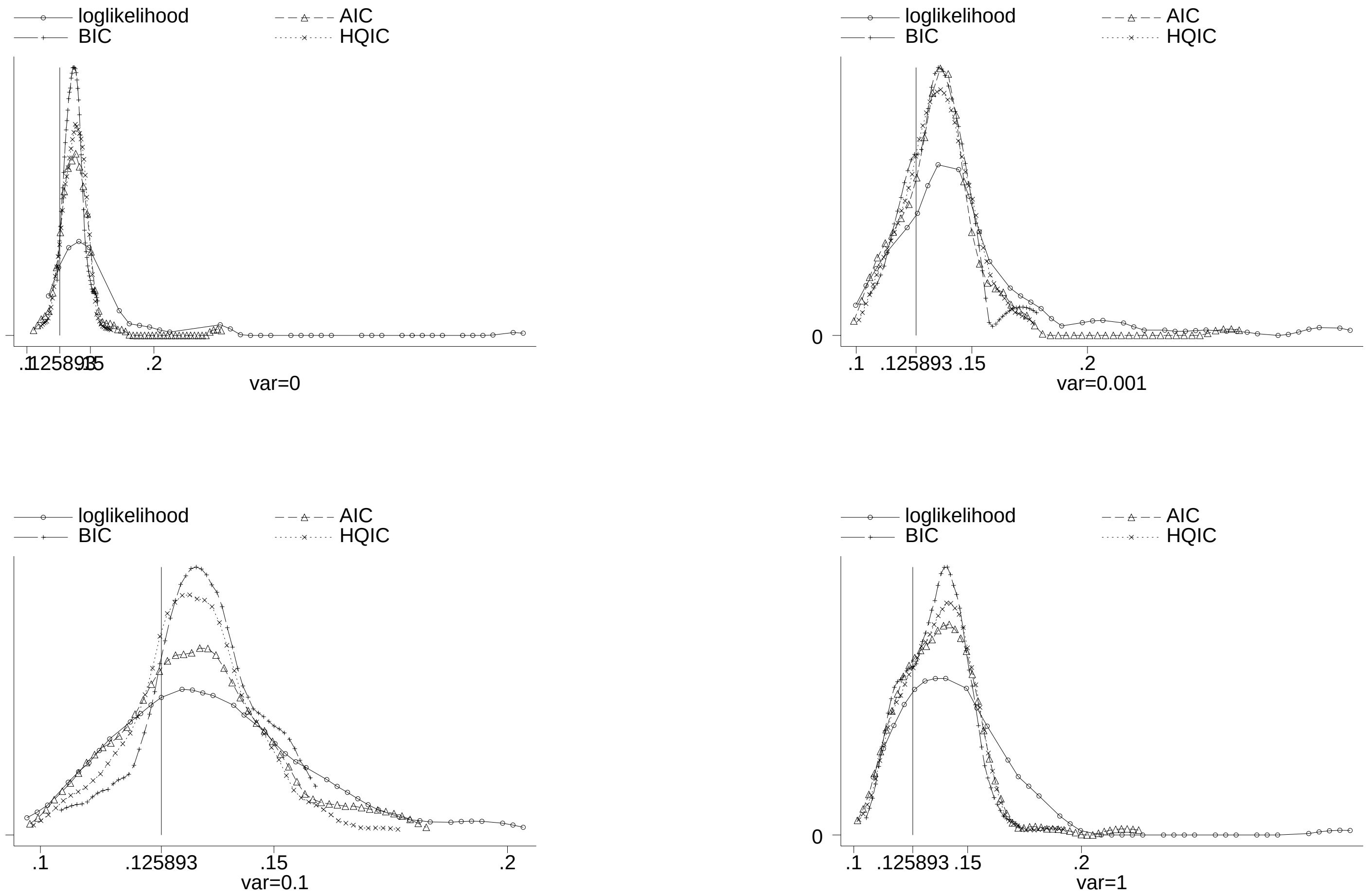


Figure A6: Kernel Densities of estimated expectation of $\hat{\mu}$. Constant hazard, 3 points mixture, 5,000 individuals.



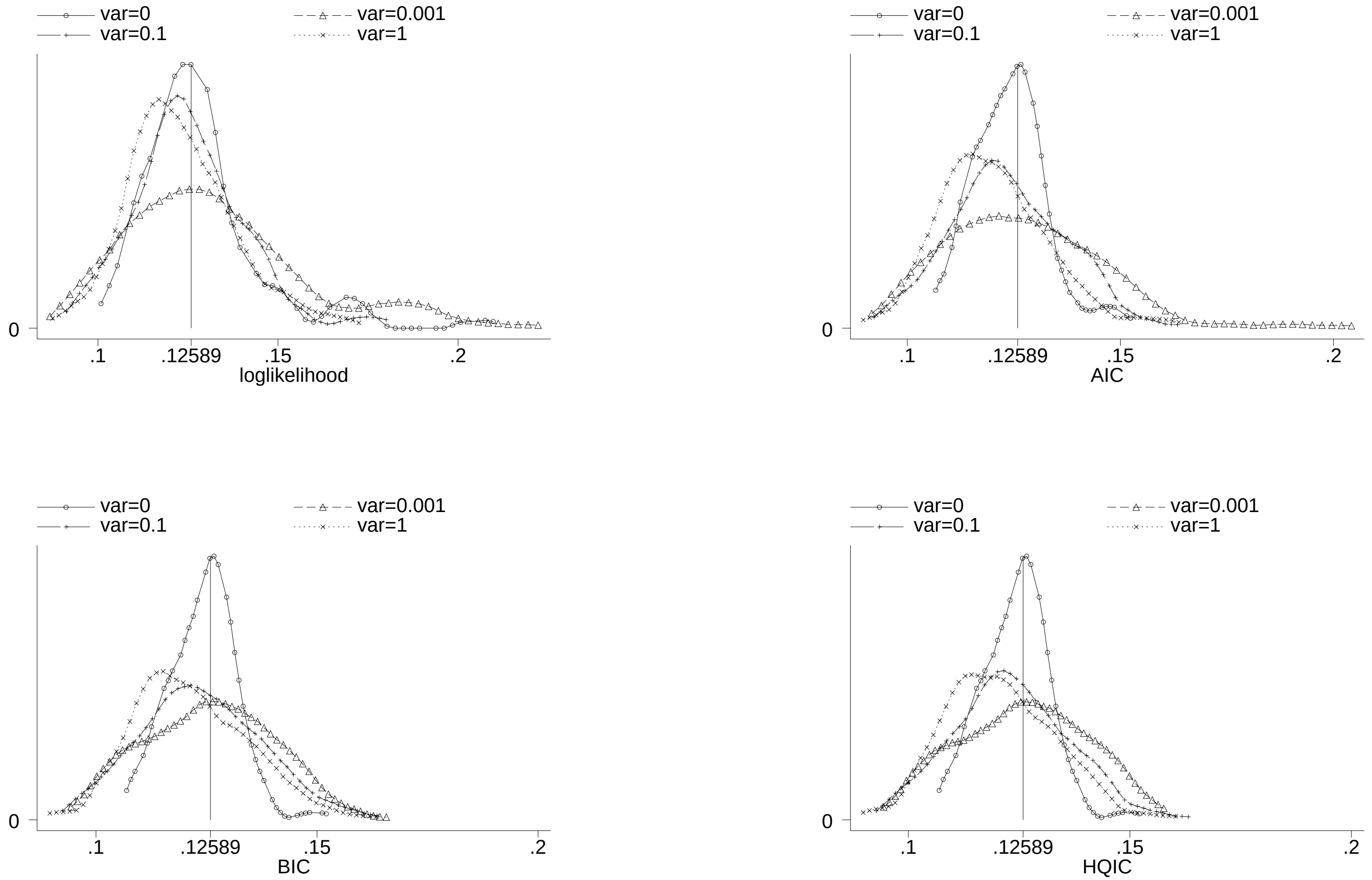
Constant hazard, 3 points, 5000 obs

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.

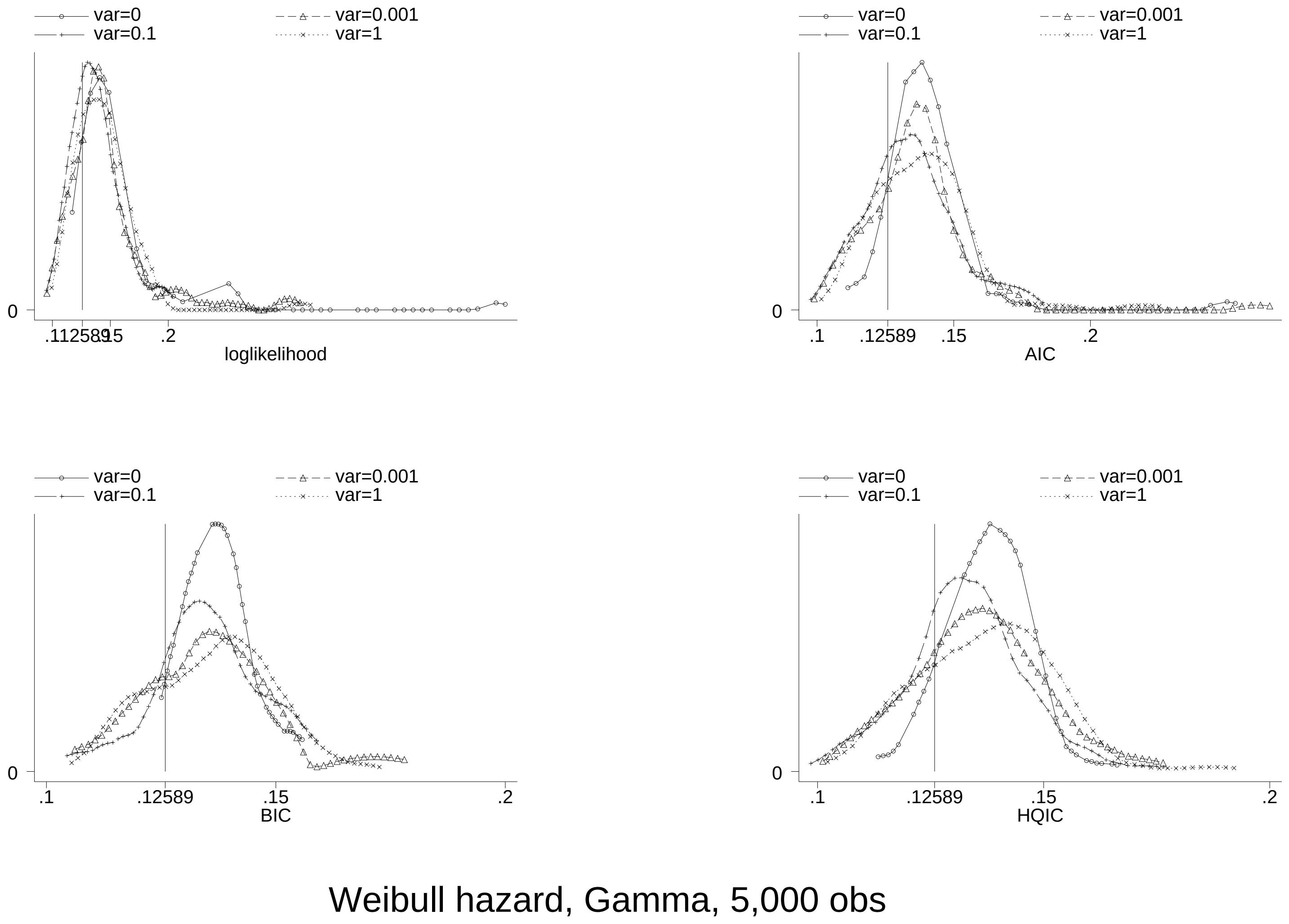
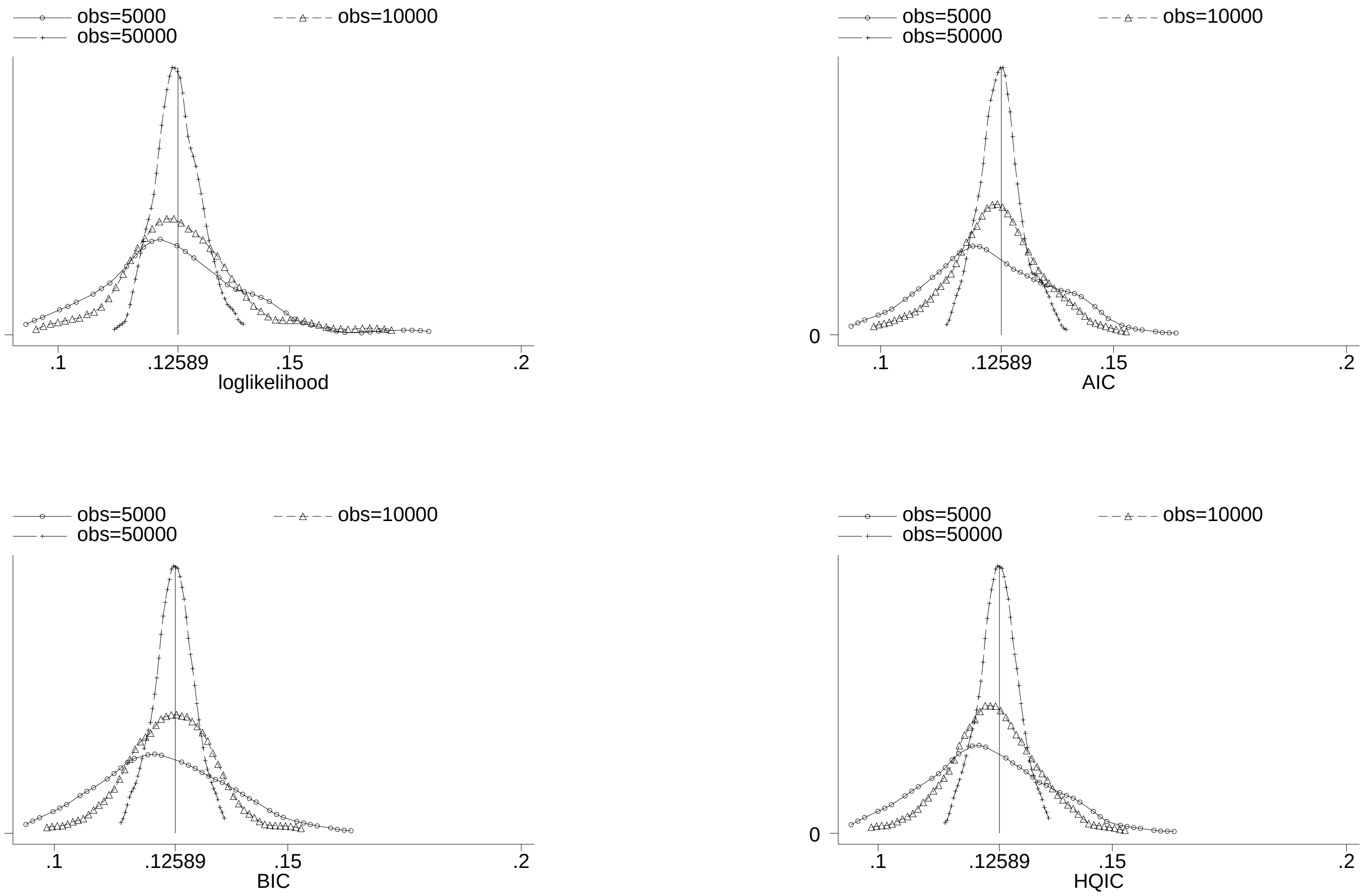
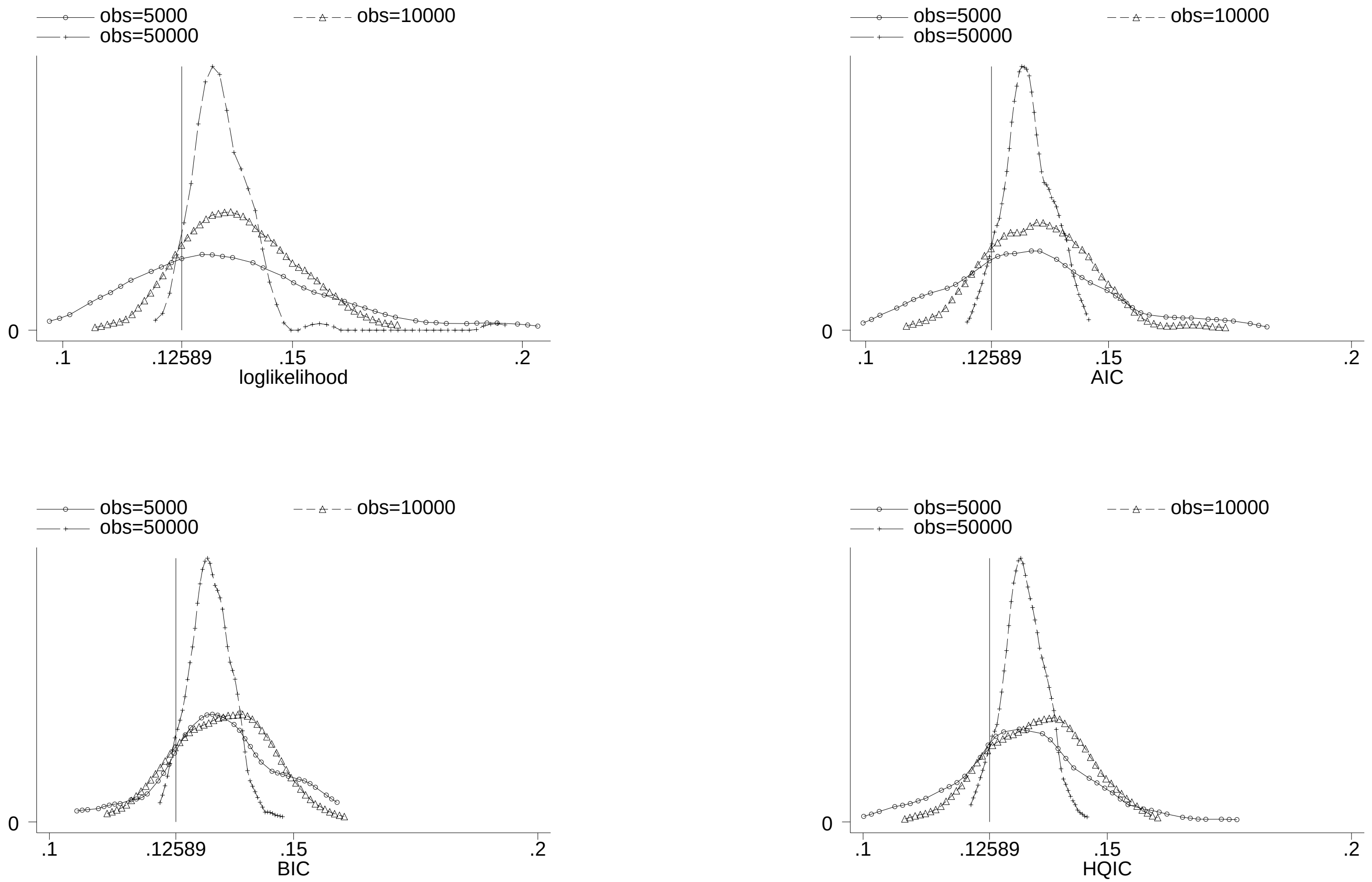


Figure A10: Kernel Densities of estimated expectation of $\hat{\mu}$ by sample sizes. Constant hazard, 3 points mixture, var(month)=0.1.



Constant hazard, 3 points, var=0.1

Figure A11: Kernel Densities of estimated expectation of $\hat{\mu}$ by sample sizes. Weibull hazard, Gamma mixture, var(month)=0.1.



Weibull hazard, Gamma, var=0.1

Figure A12-1: Kernel Densities of estimated $\hat{\beta}_1$. 10,000 individuals.

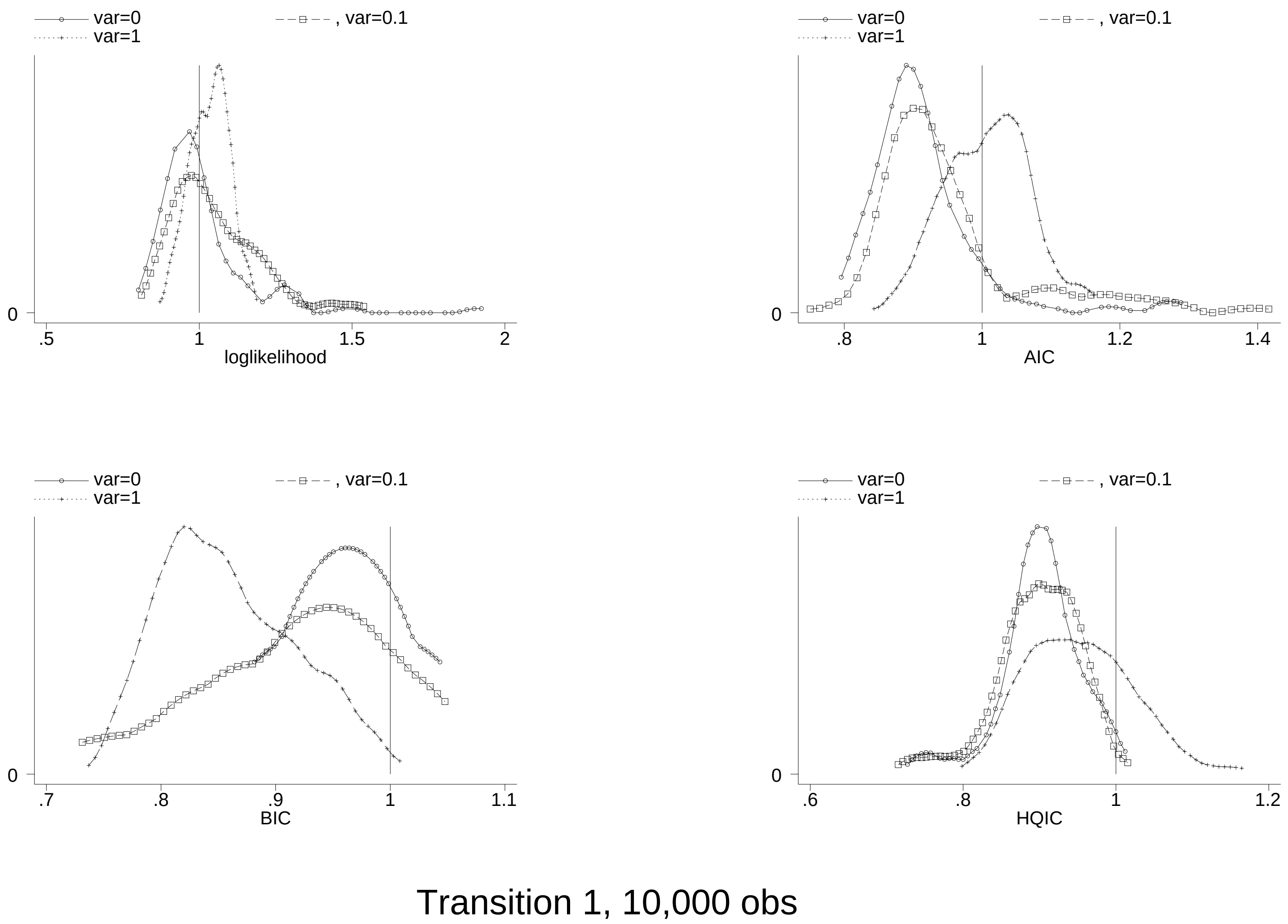
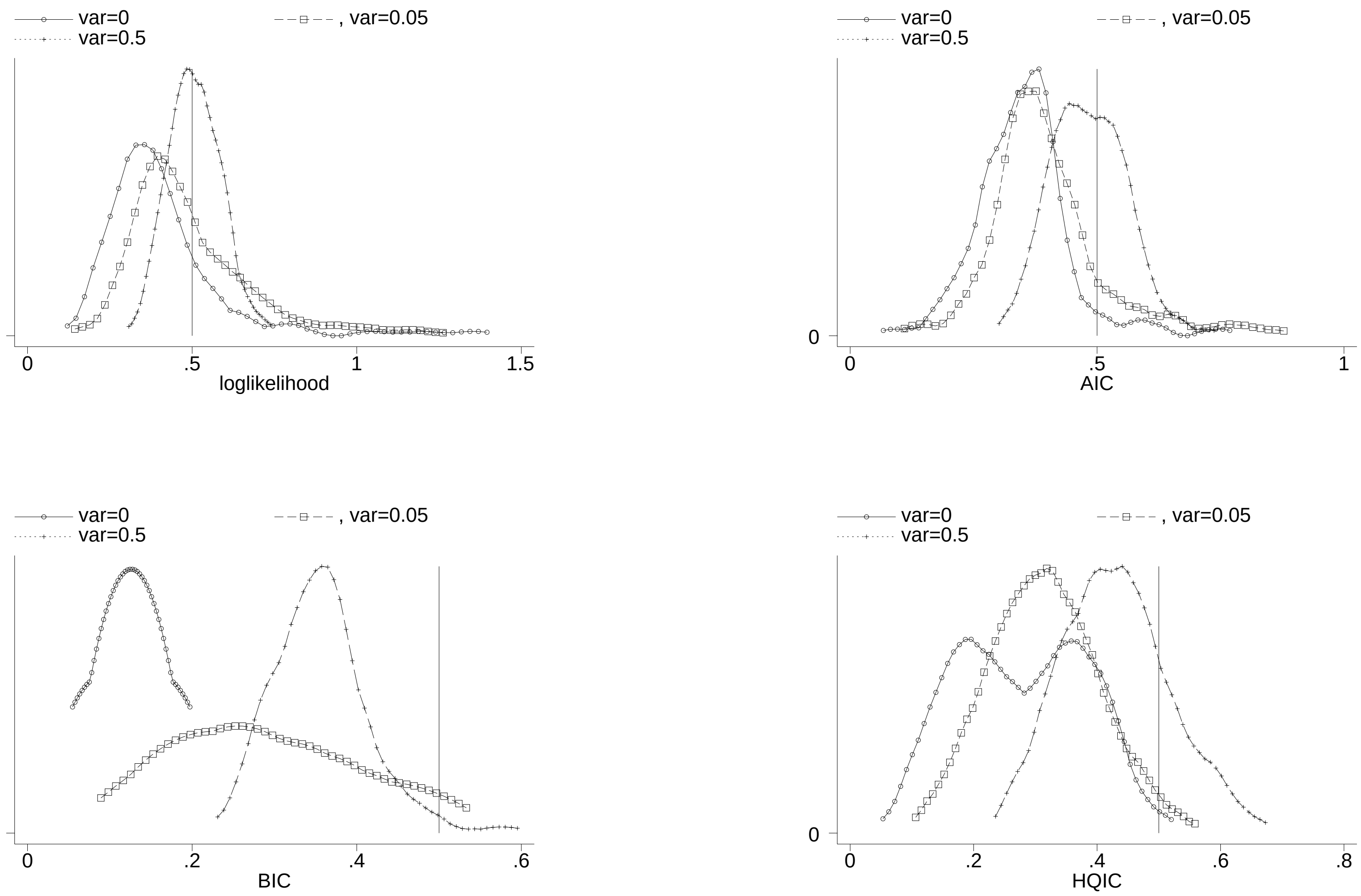


Figure A12-2: Kernel Densities of estimated $\hat{\beta}_2$. 10,000 individuals.



Transition 2, 10,000 obs

Figure A13-1: Kernel Densities of estimated $E(\hat{\mu}_1)$. 10,000 individuals.

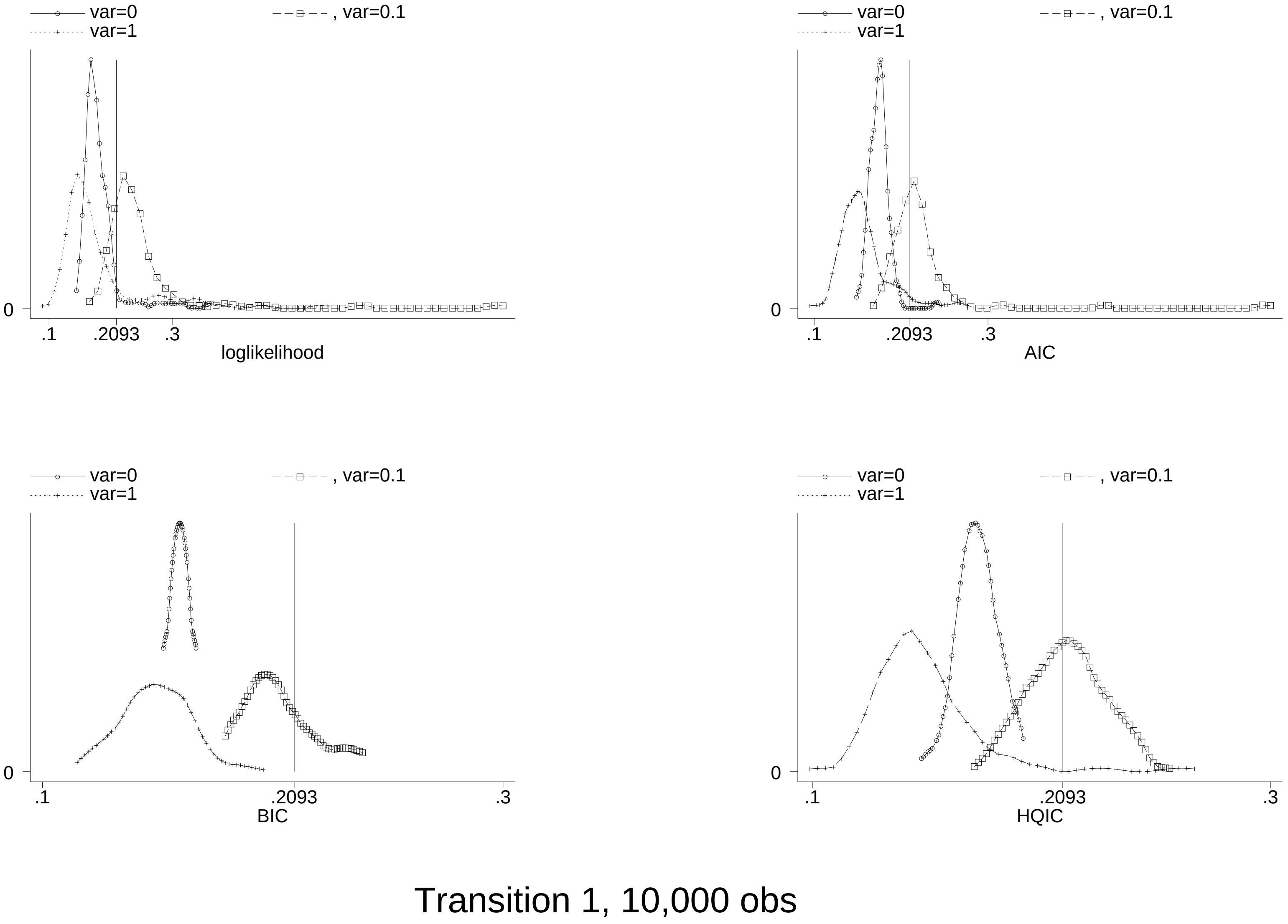
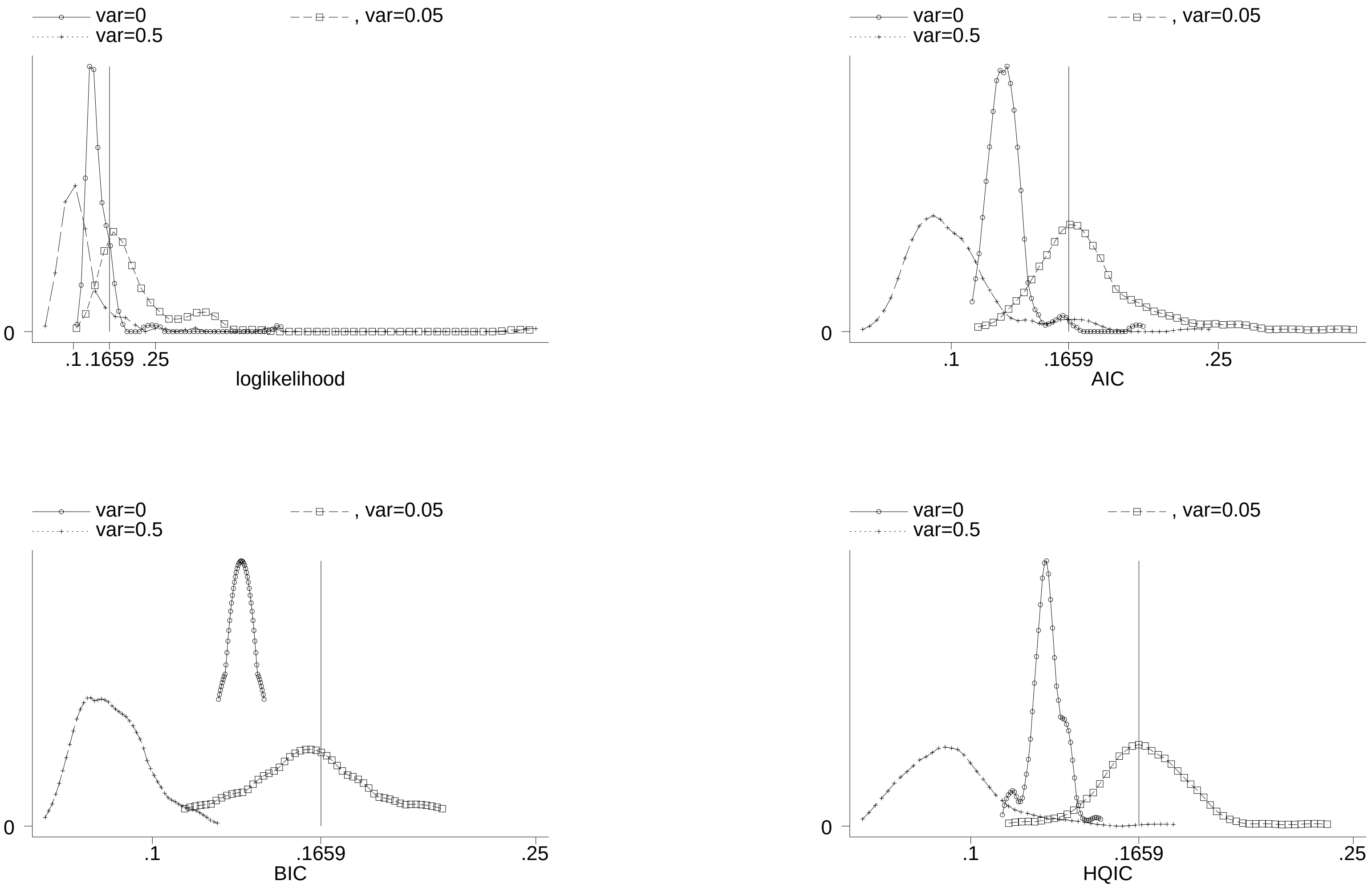


Figure A13-2: Kernel Densities of estimated $E(\hat{\mu}_2)$. 10,000 individuals.



Transition 2, 10,000 obs