The (mis)specification of discrete time duration models with unobserved heterogeneity: a Monte Carlo study

Cheti Nicoletti
Institute for Social and Economic Research
University of Essex

Concetta Rondinelli
Institute for Quantitative Methods
Bocconi University

ISER Working Paper
2006-53

Acknowledgement:
We would like to thank John Ermisch and Stephen Jenkins for useful suggestions and comments. The research was supported by the Economic and Social Research Council through their grant to the Research Centre on Micro-social Change in ISER.

Readers wishing to cite this document are asked to use the following form of words:

> Nicoletti, Cheti and Rondinelli, Concetta (November 2006) 'The (mis)specification of discrete time duration models with unobserved heterogeneity: a Monte Carlo study', ISER Working Paper 2006-53. Colchester: University of Essex.

The on-line version of this working paper can be found at http://www.iser.essex.ac.uk/pubs/workpaps/

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and

- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Institute for Social and Economic Research, University of Essex, Wivenhoe Park,
Colchester. Essex CO4 3SQ UK
Telephone: +44 (0) 1206 872957 Fax: +44 (0) 1206 873151 E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

The most popular statistical models among empirical researchers are usually the ones which can be easily estimated by using commonly available software packages. Sequential binary models with or without normal random effects are an example of such models, because they can be adopted to estimate discrete time duration models in presence of unobserved heterogeneity.

But an easy-to-implement estimation may incur a cost. In this paper we use Monte Carlo methods to analyze the consequences of omission or misspecification of unobserved heterogeneity in single spell discrete time duration models.

1 Introduction

One of the main issues concerning the estimation of hazard regressions is unobserved heterogeneity.[1] Ignoring unobserved individual characteristics may bias estimates of the effect of observed explanatory variables and of duration dependence in the hazard function.

The standard way of accounting for unobserved heterogeneity is to consider a random component, which represents a scalar function of time-invariant unobserved variables. The hazard function is then defined conditioning on observed explanatory variables and on the unobserved random component. It is then possible to estimate model parameters by maximizing the likelihood from which the unobserved random effect is integrated out. The resulting model is a mixture of hazard functions with respect to the unobserved random component. The estimation of these mixture models requires either to assume a specific parametric distribution for the random component, or to use a non-parametric maximum likelihood estimation.

In principle the non-parametric maximum likelihood estimation is the best solution to minimize the potential bias caused by improper parametric distributional assumptions.[2] Nevertheless, the computation of the non-parametric estimator is not usually possible using commands built into common software packages. For this reason many non-specialists adopt easier estimation methods by either imposing specific parametric distributions for the unobserved heterogeneity or by ignoring altogether the unobserved heterogeneity.

In continuous time duration models, the unobserved heterogeneity distribution is often chosen to be gamma for analytical convenience (see Lancaster, 1979) and theoretical reasons (see van den Berg, 2001; and Abbring and van den Berg, 2006). In discrete time duration models, instead, the assumption of a normal distribution can be computationally convenient. Under this assumption, discrete time duration models can be easily estimated as binary models with normal random effects using widely available statistical programs.[3]

[1] See Lancaster (1979, 1990), Heckman and Singer (1984) and van den Berg (2001).

[2] See, for continuous time, Heckman and Singer (1984) and, for discrete time, Baker and Melino (2000) and Zhang (2003).

[3] The fact that discrete time duration models can easily estimated by using widely available statistical programs was first noticed by Yamaguchi (1991) and Jenkins (1995). Stata provides for example the commands xtcloglog, xtlogit and xtprobit (cloglog, logit and probit) to estimate binary models with normal

In this paper we evaluate the consequences of ignoring the unobserved heterogeneity or misspecifying its parametric distribution when estimating single spell discrete time duration models. Similar studies have been already carried out by Baker and Melino (2000), Zhang (2003), Gaure et al. (2005).[4] They find that estimates are biased if unobserved heterogeneity is ignored or if its distribution is estimated non-parametrically using a discrete distribution but with an incorrect number of support points. Zhang (2003) and Gaure et al. (2005) find that the estimation bias is smaller if time-varying instead than time-invariant explanatory variables are used.

One important issue - overlooked by Baker and Melino (2000), Zhang (2003) and Gaure et al. (2005) - is that the residual variance in sequential binary models changes if unobserved heterogeneity is ignored or if a non-parametric distribution with too few or to many support points is used. Since the coefficients in binary models are usually normalized by dividing them by the residual standard deviation, models with high (low) residual variances produce coefficients which are attenuated (amplified). The attenuation (amplification) biases, Baker and Melino (2000), Zhang (2003) and Gaure et al. (2005) find, could be then due, at least in part, to their neglect of this issue. Mroz and Zaytas (2005) show that this is the case for the choice of the number of support points of the unobserved heterogeneity discrete distribution. Increasing (decreasing) the number of support points causes a reduction (rise) of the residual variance and an amplification (attenuation) of the coefficients. In this paper we show that the attenuation bias caused by the omission of the unobserved heterogeneity or the misspecification of its distribution can be also a consequence, at least in part, of the coefficient normalization issue.

We undertake a Monte Carlo study to evaluate the effects of two misspecification problems in sequential logit models with unobserved heterogeneity:

1. omission of the unobserved heterogeneity when using time-varying and/or time-invariant explanatory variables,

random effects (without normal random effects) and error terms with extreme value, logistic and normal distributions. For more details on discrete time duration models we refer to Holford (1976), Prentice and Gloeckler (1978), Allison (1982), Narendranathan and Stewart (1993) and Sueyoshi (1995).

[4] For continuous time duration models we refer instead to Heckman and Singer (1984), Lancaster (1985), Trussell and Richards (1985), Ridder (1987) and Dolton and van der Klaauw (1990).

2. imposing a normal distribution for the unobserved heterogeneity when instead the true distribution is a gamma or a discrete distribution.

The consequences of the first type of misspecification were already studied by Baker and Melino (2000) but they only considered time-invariant explanatory variables and, as already noted, without taking account of the normalization issue in binary models. The consequences of incorrectly imposing a normal random effect in discrete time duration model have not been studied before.

Finally, we consider the effect of misspecifying both the distribution of the unobserved heterogeneity and of the residual error.

The paper is organized as follows. Section 2 considers the effects of neglecting unobserved heterogeneity while Section 3 considers the effects of its misspecification. In both sections we first discusses the theoretical consequences of omitting or misspecifying the unobserved heterogeneity, and then we assess those possible consequences through a Monte Carlo simulation exercise. In Section 4 we summarize the main findings.

2 Ignoring unobserved heterogeneity

2.1 Consequences of ignoring unobserved heterogeneity

Ignoring unobserved heterogeneity in duration models can cause a bias in the estimation of the duration dependence. More precisely, the omission of the unobserved heterogeneity causes an overestimation of the negative duration dependence (see for example Lancaster, 1990; and van den Berg, 2001). This is because people who have a high unobserved random component are more likely to experience the event of interest early, so that the sample of individuals that survive is a selected sample with relatively small random effects.[5] This selection process is known as *weeding out* or *sorting effect*.

Omitting unobserved heterogeneity may also bias the coefficients of the explanatory variables in the hazard model. For example neglecting unobserved heterogeneity in mixed pro-

[5]Notice that, without loss of generality, we are assuming in this section that the unobserved random component be positively related to the hazard function.

portional (continuous time) hazard models causes an underestimation of the proportionate response of the hazard function with respect to the explanatory variables.[6]

The bias is again due to a weeding out effect. Let us assume that the unobserved heterogeneity is given by a time-invariant scalar random effect, θ, independent of the explanatory variables; while the observed heterogeneity is given by a scalar function $\mu = m(X; \beta)$, where X is a vector of individual time-invariant explanatory variables and β is the vector of the corresponding coefficients. Without loss of generality, we assume in this section that the hazard function conditional on the observed explanatory variables and the unobserved heterogeneity be positively related to both θ and μ. A hazard model ignoring the unobserved heterogeneity is a hazard function conditional on the observed characteristics, X, but unconditional on the unobserved heterogeneity, θ, which we call the observed hazard function. The difference in the observed hazard function between survived people with high and low values of μ reflects also a gap in their values of θ. Survivors with a large μ have on average a smaller θ than survived people with a small μ, so that the difference between the observed hazard functions is on average lower than the difference we would observe if the survivors had the same value for θ. If we fail to recognize that the lower difference between the observed hazards is due to a difference in the unobserved heterogeneity, we would erroneously estimate an attenuated effect of the explanatory variables on the hazard.

More rigorously, the weeding out effect on the covariate coefficients can be described as the consequence of a lack of independence between the random effect for a generic individual i, θ_i, and her (his or its) observed heterogeneity, $m(X_i; \beta)$, given a duration $T_i \geq \tau$, where τ is a scalar strictly higher than zero, say the failure of the condition $(\theta_i \perp\!\!\!\perp m(X_i; \beta) \mid T_i \geq \tau)$. Notice, instead, that hazard models assume that $(\theta_i \perp\!\!\!\perp X_i)$ which implies that $(\theta_i \perp\!\!\!\perp m(X_i; \beta) \mid T_i \geq 0)$. We assume here that $(T_i \mid X_i, \theta_i)$ be identically and independently distributed i.i.d. across individuals.

There are some continuous time duration models for which the attenuation bias due to omitted unobserved heterogeneity reduces to a rescaling by a factor (a bias proportionally identical) for all explanatory variables coefficients or to a bias only for the intercept. Lancaster (1985) proves analytically that the omission of unobserved heterogeneity in mixed

[6]See van den Berg (2001) for a formal proof.

proportional hazard models with baseline distribution given by a Weibull causes a rescaling by a constant factor for all coefficients. Ridder (1987) proves analytically that the omission in mixed proportional hazard models with known baseline hazard and with no right censuring causes a bias only for the intercept. Moreover, Ridder (1987) suggests that replacing the baseline with a non-parametric flexible specification should produce an almost unbiased estimation of the covariates coefficients.

Ridder's suggestion is supported by his Monte Carlo study and by some empirical studies: see Dolton and van der Klaauw (1995), Meyer (1990), and Trussell and Richards (1985). By contrast, the conjecture is not confirmed by the Monte Carlo experiment in Baker and Melino (2000), who consider discrete time duration models with single spell. But this contradictory result may be due to the fact that discrete time duration model coefficients are identified only up to a scale normalization and models with different specifications use different normalizations, which Baker and Melino (2000) do not consider.

It is possible to prove analytically that the omission of the unobserved heterogeneity causes only a rescaling by a factor of the covariate coefficients when considering sequential probit models with normal random effects θ_{it} that are i.i.d. across individuals and time t, and independent of the explanatory variables, X_{it}, and with known duration dependence function. This is because $(\theta_{it} \perp\!\!\!\perp m(X_{it};\beta) \mid T_i \geq \tau)$ for any $\tau \geq 0$. (see Appendix A for more details).

Similar analytical results do not exist instead for more general discrete time duration models. In this paper we consider the consequences of omitting unobserved heterogeneity in more general single spell discrete time duration models in the following cases:

a. the unobserved random effects is time-invariant and follows a normal, a gamma or a discrete distribution with two points of support,

b. the error distribution is logistic instead of normal,

c. the duration dependence is ignored or it is approximated by a flexible function,

d. the covariates are i.i.d. across individuals and time or i.i.d. across individuals but not time.

Cases described in (a) to (c) were already considered by Baker and Melino (2000). They find that ignoring unobserved heterogeneity component causes an attenuation bias for the covariate coefficients. In this paper we replicate their Monte Carlo study to evaluate again the consequences of ignoring unobserved heterogeneity but taking into account the issue of the normalization of the coefficients. Mroz and Zayats (2005) reconsider instead the Monte Carlo study of Baker and Melino (2000) to compare the effects of alternative non-parametric specifications of the unobserved heterogeneity distribution when taking account of the normalization issue. Baker and Melino (2000) find that non-parametric maximum likelihood estimation that penalizes specifications with many mass points for the unobserved heterogeneity distribution produces more reliable coefficients; but Mroz and Zayats (2005) give evidence that this result is a consequence of the normalization issue.

Case (d) is an extension necessary to understand how the estimation bias can depend on the types of covariates used. If the covariates, say X_{it} for individual i and duration (time) t, are i.i.d. across individuals and time, then estimation bias should reduce. This is because in this case the independence between the unobserved component and the observed covariates holds even when conditioning on survival until a time strictly greater than zero, that is $(\theta_i \perp\!\!\!\perp X_{it} \mid T_i \geq \tau)$, where $\tau > 0$.

If, instead, covariates are i.i.d. across individuals but time-invariant or correlated across time, then we would expect an attenuation bias. Nevertheless, the bias could consist in a proportional reduction, in absolute value, in all covariate coefficients, i.e. a rescaling by a constant factor.

In next section we describe the Monte Carlo experiment carried out in order to study the potential consequences of omitting time-invariant unobserved random effects for the cases described by (a) to (d).

2.2 Description of the Monte Carlo Simulation: Data Generating Processes

We consider the same data generating processes (DGPs from now on) used in the Monte Carlo study of Baker and Melino (2000) except that we use both time-varying and time-invariant explanatory variables while they use only a time-invariant one.

We assume that duration is measured in discrete time. This is quite often the case when observations are grouped into intervals or when the event, whose occurrence defines the end of a duration, can occur only in discrete time. We then record an event taking place in the interval $(t-1, t]$ as occurred in t.

We assume that the probability of experiencing an event in t (or in the time interval $(t-1, t]$) conditional on survival to $(t-1)$ for a generic individual i is given by :

$$Pr(d_{it} = 1 | d_{it-1} = 0) \quad = \quad Pr(z^*_{it} < 0 | z^*_{it-1} \geq 0) \tag{1}$$

where d_{it} is a dummy variable indicating the event occurrence at t for individual i, and z^*_{it} is a latent continuous variable which is lower than zero if $d_{it} = 1$ and higher or equal to zero otherwise. We assume that z^*_{it} obeys the following linear model:

$$z^*_{it} = X_{it}\beta - f(t) + \theta_i + \epsilon_{it} \tag{2}$$

where X_{it} is a vector of explanatory variables, β is the corresponding vector of parameters, $f(t)$ is a deterministic function of elapsed duration, θ_i is an individual random effect representing unobserved heterogeneity, ϵ_{it} is a residual error term distributed as a logistic with zero mean and variance $\pi^2/3$ and both θ_i and ϵ_{it} are independent of the explanatory variables[7]. We can then write the hazard probability conditional on the observed explanatory variables, X_{it}, and on the unobserved heterogeneity, θ_i, as

$$Pr(d_{it} = 1 | d_{it-1} = 0, X_{it}, \theta_i) \quad = \quad \frac{1}{1 + exp(z_{it})} \tag{3}$$

where

$$z_{it} \quad = \quad X_{it}\beta - f(t) + \theta_i. \tag{4}$$

By choosing different specifications for the observed explanatory variables, X_{it}, the duration dependence function, $f(t)$, and the unobserved heterogeneity, θ_i, we produce a set of different DGPs.

[7]The definition of the above discrete time hazard model and the notation used are consistent with Baker and Melino (2000) except for the negative sign in front of the duration dependence function $f(t)$. This is because it is counterintuitive to have a negative (positive) duration dependence when $f(t)$ is increasing (decreasing) in time.

We organize the simulations in two main sets. In the first set, exercise **A**, we focus our attention on the effect of omitting unobserved heterogeneity when using different types of explanatory variables. In particular we consider three DGPs using three different typologies of observed explanatory variables: **A1** time-varying variables, **A2** time-invariant variables and **A3** variables given by the sum of a time-invariant variable and a time-varying one, say mixture variables. For each choice of the covariates we consider two different types of duration dependence function, one increasing and one deceasing, and three types of distribution for the unobserved heterogeneity, a discrete (with two points of support), a gamma and a normal distribution. This provides us with 18 different DGPs.

In the second set of simulations, exercise **B**, we consider both time-invariant and time-varying covariates and focus attention on the effect of omitting the unobserved heterogeneity when considering or not considering duration dependence in the simulated and estimated models. Again we consider three different types of distribution of the unobserved heterogeneity, whereas we consider only one specification for the duration function and for the vector of covariates which includes both time-invariant and mixture variables. This second simulation exercise produces six different types of DGPs.

For each of the DGPs in simulation exercise **A** we consider two sample sizes: 500 and 1000 individuals. For simulation exercise **B**, we consider instead three sample sizes: 500, 1000 and 5000 individuals. The higher sample size of 5000 is motivated by the fact that in exercise **B** there are some small sample biases which decrease very slowly with the sample size.

As in Baker and Melino (2000) we draw 100 samples for each DGP, we follow the individuals for 40 periods and consider all durations greater than 40 as censored.

In the following, we discuss in more detail how we specify the explanatory variables, the duration dependence function and the unobserved heterogeneity distribution for different types of DGP.

Observed explanatory variables.

As in Baker and Melino (2000) we fix the variance of the observed heterogeneity in the hazard model, $Var(X_{it}\beta)$, to be equal to 0.25 for all our simulations.

In exercise **A** we specify the observed heterogeneity in the hazard model as:

$$X_{it}\beta \quad = \quad X_{1,\,it}\beta_1 + X_{2,\,it}\beta_2. \tag{5}$$

where $X_{1,it}$ and $X_{2,it}$ are normal random variables, and β_1 and β_2 are fixed parameters which we set to be equal to 1 and 0.5.

We consider three different simulations for the variables, $X_{1,it}$ and $X_{2,it}$:

A1 two independent time-varying variables identically and independently distributed (i.i.d.) across individuals and time with zero means and variances 0.125 and 0.5;

A2 two independent time-invariant variables i.i.d. across individuals with zero means and variances 0.125 and 0.5;

A3 two independent variables defined as the sum of a time-invariant variable and a time-varying one with equal variances, say mixture variables; more precisely, $X_{1,it}$ ($X_{2,it}$) is the sum of a time-varying variable defined as **A1** but with variance 0.0625 (0.25) and a time-invariant variable defined as in **A2** but with variance 0.0625 (0.25).

Simulation **A1** represents an extreme case which is interesting from a theoretical point of view but it is less interesting from an empirical one. In empirical examples explanatory variables are usually correlated across time so that the assumption of explanatory variables i.i.d. across individuals and time does not seem to be very plausible. Simulation **A2** represents the opposite extreme case where all the explanatory variables are supposed to be time-invariant. This is the case considered by Baker and Melino (2000). Finally, simulation **A3** represents an intermediate case where the explanatory variables are given by the sum of a time-invariant component and a time-varying one. Earnings and income can be examples of such types of variables. Earnings and income (or their logarithm transformations) are usually assumed by economists to be the sum of a permanent component and a transitory one (see for example Moffitt and Gottschalk, 2002).

In simulation exercise **B** we specify instead the observed heterogeneity in the hazard model as:

$$X_{it}\,\beta \quad = \quad X_{1,\,i}\,\beta_1 + X_{2,\,i}\,\beta_2 + X_{1,\,it}\,\beta_3 + X_{2,\,it}\,\beta_4 \tag{6}$$

where $X_{1,i}$ and $X_{2,i}$ are time-invariant variables, $X_{3,it}$ and $X_{4,it}$ are mixture variables and $\beta' = [1, 0.5, 1, 0.5]$. To be more specific $X_{1,i}$ and $X_{2,i}$ are time-invariant variables defined as in **A2** but with variances 0.0625 and 0.25, $X_{3,it}$ and $X_{4,it}$ are mixture variables defined as in **A3** but with variances 0.0625 and 0.25, and all explanatory variables are independent.

Duration Dependence.

In exercise **A** we consider, as in Baker and Melino (2000), the following deterministic time function

$$f(t) = 1 - \exp\left(\frac{1-t}{5}\right) \tag{7}$$

for a negative duration dependence and

$$f(t) = \exp\left(\frac{1-t}{5}\right) - 1 \tag{8}$$

for a positive duration dependence.

In simulation exercise **B** we consider instead $f(t) = 0$ for no duration dependence and again $f(t) = \exp\left(\frac{1-t}{5}\right) - 1$ for a positive duration dependence.

Unobserved Heterogeneity.

In both exercises **A** and **B** we consider three distributions for the unobserved heterogeneity θ_i: discrete, gamma and normal distribution. To be consistent with Baker and Melino(2000) we set $E(\theta_i) = 1.8$ and $Var(\theta_i) = 1$ and for the discrete distribution we consider two support points with equal probability, that is:

$$\theta_i = \begin{cases} 0.8 & \text{with probability } 0.5 \\ 2.8 & \text{with probability } 0.5\,. \end{cases} \tag{9}$$

2.3 Description of the Monte Carlo Simulation: estimation models

Using the data simulated in exercise **A** we estimate a sequential logit model as specified in (3) but ignoring the unobserved heterogeneity and approximating the duration dependence function with either a cubic polynomial in t or using a 'non-parametric' step function. As in Baker and Melino (2000) we consider a step function given by

$$\phi(t) = \sum_{\tau=1}^{40} \phi_\tau D_{t\tau} \tag{10}$$

where

$$D_{t\tau} = \begin{cases} 1 & \text{if } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

and ϕ_τ, $\tau = 1, ..., 40$, are the corresponding coefficients. However, because few individuals survive after 15 periods, we allow the coefficients to vary for each period only until $\tau = 14$ and then we impose constant coefficients within the following time intervals: $\tau = 15 - 19$, $20 - 24$, $25 - 29$, $30 - 40$.

Using the data simulated in exercise (**B**) we estimate again a sequential logit model ignoring the unobserved heterogeneity and approximating the duration dependence function with either a zero function (no duration dependence) or the above 'non-parametric' step function.

2.4 Results

In this section we present the results of the Monte Carlo simulation exercises **A** and **B**.

The results of the exercise **A** are reported in Table 1, which is divided in three panels giving the results for time-varying covariates (top panel **A1**), time-invariant covariates (middle panel **A2**) and mixture covariates (bottom panel **A3**). The results reported are the average and the standard deviation over 100 replications for the covariate coefficients, β_1 (the true value of which is 1) and β_2 (which true value is 0.5), and their ratio β_1/β_2. By row we specify the type of DGP used to generate the simulated data. More precisely, we consider six different types of DGPs: sequential logit model with negative or positive duration dependence and with unobserved heterogeneity following a discrete, a gamma or a normal distribution (labeled discrete, gamma and normal). By column we specify instead the sample size (500 or 1000 observations/individuals) and the type of estimation model used: sequential logit model omitting random effects and with duration dependence approximated by a step function (labeled Step DD) or by a cubic polynomial (labeled polynomial DD).

If the omission of the unobserved heterogeneity causes only a rescaling by a factor of the coefficients, then the coefficients would be biased towards zero (attenuation bias) but their ratio would still be correctly estimated. This seems supported by the results in Table 1 when using any type of covariates. Moreover, when using time-varying covariates which are i.i.d. across individuals and time (top panel **A1**), the attenuation problem for the coefficients does not seem to be significant. When using instead covariates which are i.i.d. across individuals and time-invariant, the attenuation problem is more severe. Finally, when using mixture covariates the attenuation bias magnitude seems to be intermediate between the two previous extreme cases.

Using different distributions for the simulated unobserved heterogeneity components and different specifications for the simulated duration dependence (negative or positive) produce some very small and insignificant differences in the coefficients.

Approximating the duration dependence function by using a step or a cubic polynomial function does not produce any significant difference in the results.

Finally, increasing the sample size from 500 to 1000 observations leads to a slight improvement in the results, meaning that the attenuation bias for β_1 and β_2 decreases a little and the average ratio between coefficients becomes even closer to the true value of two.

To summarize, ignoring unobserved heterogeneity in sequential logit models seems to cause an attenuation of the covariate coefficients due to a different normalization. This attenuation bias cancels almost completely when using covariates which are i.i.d. across individuals and time, while it does not when the covariates are highly autocorrelated.

As emphasized in Section 2.1, ignoring unobserved heterogeneity may cause a bias for the covariate coefficients as well as for the duration dependence function estimation. In Figures 1 and 2 we compare the true simulated duration dependence function (line labeled true duration) with the estimated duration dependence functions averaged over 100 samples of size 1000, simulated using the different DGPs in Monte Carlo exercise **A1**. We consider a negative true duration dependence function given by (7) in figure 1 and a positive one given by (8) in figure 2. In both Figures the three average estimated dependence functions (lines labeled discrete, gamma and normal) are obtained by considering a cubic polynomial in the duration and by using data simulated from three different DGPs assuming a discrete, a gamma or a normal distribution for the unobserved random effects.

It seems that ignoring the unobserved heterogeneity causes an overestimation of the negative duration dependence and a spurious negative dependence even when the true duration dependence is positive.

In Tables 2 and 3 we report the results for exercise **B** in which we consider a hazard model with two time-invariant and two time-varying covariates. The estimation models considered are sequential logit models ignoring unobserved heterogeneity. In Table 2 the duration dependence is estimated using a step function, whereas in Table 3 the duration dependence is ignored. In both Tables we report the average and the standard deviation over 100 replications for the covariate coefficients, $\beta' = [\beta_1, \beta_2, \beta_3, \beta_4]$ which have true values $[1, 0.5, 1, 0.5]$, and the ratios between any pair of coefficients. The estimation results for different simulated DGPs are reported by column. We consider six different DGPs, sequential logit models with three possible distributions for the unobserved heterogeneity (discrete, gamma or normal) and with either negative or no duration dependence. Both Tables are divided in 3 panels corresponding to three different sample sizes (500, 1000 and 5000).[8]

In Table 2 the covariate coefficients seem to be significantly underestimated. Moreover, the underestimation of the coefficients seems to be slightly larger for the pair of time-invariant variables than for the pair of time-varying ones. In other words, it seems that the rescaling factor is slightly dissimilar for different types of variables (time-varying and invariant variables). Indeed, ratios between coefficients seems to be correctly estimated when considering two variables of the same type (see β_1/β_2 and β_3/β_4) and to be slightly biased when considering the ratio between two different types of variables (see β_3/β_2, β_3/β_1, β_1/β_4 and β_4/β_3). Nevertheless, since the standard deviations for coefficient ratios are quite high, the differences in the rescaling factor are not significant. This result is confirmed even when using a larger sample size of 5000 observations. In conclusion, we find again that omitting the unobserved heterogeneity cause an attenuation of the covariate coefficients due to a rescaling factor which differs slightly and not significantly by typology of variable.

[8] As already said, the higher sample size of 5000 is motivated by the fact that in exercise **B** there are some small sample biases which decrease very slowly with the sample size.

Looking at the results in Table 3 where the estimation models ignore both the unobserved heterogeneity and the duration dependence, the underestimation of covariate coefficients reduces and the rescaling factor is more similar for variables of different types. The ratios between coefficients are not biased especially when considering a sample size of 5000 observations. This result is not unexpected because the unobserved heterogeneity component and the time duration component are negatively and positively related to the duration, so that the biases for ignoring those components should go in opposite directions and should offset each other, at least in part.

In conclusion, the two main findings of this section are that ignoring the unobserved heterogeneity in sequential logit models causes an overestimation of the negative duration dependence and an attenuation of the covariate coefficients but this attenuation is due to a rescaling by a factor of every coefficient by the same amount. Since coefficients in binary models are only identified up to a scale normalization, inference will not be affected by the unobserved heterogeneity omission except for the duration dependence.

3 Misspecifying the unobserved heterogeneity distribution

3.1 Consequences of misspecifying unobserved heterogeneity

Heckman and Singer (1984) argue that an incorrect assumption about the distribution of the unobserved heterogeneity in hazard models can have severe consequences. In particular, they find that the parameters estimates for a model with Weibull baseline hazard are very sensitive to changes in the distribution assumed for the unobserved heterogeneity. Similar results were found also by Trussell and Richards (1985), Hougaard et al. (1994), Baker and Melino (2000), Zhang (2003) and Gaure et al. (2005). However, Ridder and Verbakel (1983) criticize the findings of Heckman and Singer (1984) and highlight the fact that a non-flexible specification of the baseline hazard may explain their (Heckman and Singer) findings.

We reconsider the heterogeneity misspecification problem in the specific case of discrete time duration models with single spell specified as sequential binary models. We are particularly interested in evaluating the effect of imposing a normal distribution for the unobserved heterogeneity component when the true distribution is a gamma or a discrete distribution with two support points.

By contrast, Baker and Melino (2000) studied the effect of using different non parametric specification of the unobserved heterogeneity distribution. They find that if too many support points for the estimated heterogeneity distribution are used, the unobserved heterogeneity dispersion is overestimated and the covariate coefficients are biased away from zero (amplification bias). As explained by Mroz and Zayats (2005) this amplification bias may simply be due to a rescaling by a factor of the variables. Mroz and Zayats (2005) find indeed that covariates effects seem to be better estimated if the number of support points is large when taking into account the normalization problem.

In addition, we consider the potential consequences of misspecifying the distribution of the residual error as well as of the unobserved heterogeneity in the sequential binary models.

3.2 Description of the Monte Carlo simulation: DGPs and estimation models

As in Section 2, we carry out a Monte Carlo experiment by simulating 100 samples from a set of different DGPs (data generator processes).

The DGPs used to generate the data are sequential logit models with unobserved heterogeneity following three alternative types of distribution (discrete, gamma or normal), with a negative time duration dependence and two explanatory variables given by two mixture variables. For more details on the DGPs we refer to Monte Carlo exercise **A3** described in Section 2.2.

Our estimation models are instead given by sequential binary models with normal random effects and duration dependence approximated by a cubic polynomial in the duration. We consider three models: (1) sequential logit, (2) sequential probit and (3) sequential complementary log-log models. We estimate those sequential binary models with random effects by using Stata which considers an adaptive Gauss-Hermite quadrature to approximate the integral of the maximum likelihood function with respect to the random effects (see for more details StataCorp, 2005)[9].

The simulation exercise is carried out as the previous ones by drawing 100 samples for each DGP and three different sample sizes: 500, 1000 and 5000 individuals. We consider durations longer than 40 periods as censored.

3.3 Results

In Tables 4, 5 and 6 we report the results corresponding to the use of the three different estimation models: sequential logit, probit and or complementary log-log with normal random effects and cubic polynomial in the duration. The simulated data used in all three Tables are generated from the same DGP: a sequential logit model with negative duration dependence and unobserved heterogeneity following three alternative distributions (discrete, gamma or normal).

In each Table we report the average and the standard deviation over 100 replications for the two covariate (mixture variable) coefficients, β_1 (which true value is 1) and β_2 (which true value is 0.5), their ratio β_1/β_2, the fraction of residual variance explained by individual random effects (ρ), the average number of iterations and the number of cases out of 100 of successful convergence of the maximum likelihood algorithm.[10] Each Table is divided in three panels reporting results produced using three different sample sizes: 500, 1000 and 5000 observations.

Looking at the results in Table 4, where both estimation and simulated models are sequential logit models, the covariate coefficients do not seem to be underestimated. They seem to be well estimated even when the unobserved heterogeneity distribution is misspecified. This is an encouraging result for practitioners who would like to use easy-to-implement estimation methods to take account of unobserved heterogeneity.

In Table 5, where the estimation model is given by a sequential probit model while the true DGPs are given by sequential logit models, the two covariate coefficients are underestimated

[9] An alternative estimation methods is given by the simulated maximum likelihood, see for more details Gourieroux and Monfort (1996) and Train (2003).

[10] We report averages and standard deviations only for the cases where convergence was reached.

but the ratio between them is still unbiased. Again we do not find relevant differences when considering DGPs with different distributions for the random effects.

Since the logistic distribution is similar to the normal one but with heavier tails, the difference in the estimated coefficients when using probit instead of logit models is probably due to the different normalization implied by the different residual variances. Because the residual variance in logit models is normalized to $\pi^2/3$ while in probit models is normalized to one, the rescaling factor should be given by $\pi/\sqrt{3}$ (see Greene, 2003) . By multiplying the coefficients in Table 5 by this factor, we find that the estimated are very close to the ones reported in Table 4.

Finally, in Table 6, we change the estimation model to a sequential complementary log-log model. The two covariate coefficients seem to be slightly underestimated while the ratio between them is unbiased. The coefficients seem slightly lower than the ones shown in Table 4 and it seems that coefficients bias be due again to a rescaling. Again, the results do not seem to be affected by the distribution assumed for the unobserved heterogeneity in the DGPs.

Increasing the sample size has the same effect for all three types of models (logit, probit and complementary log-log): the attenuation bias does not change significantly, the standard deviations decrease, and the number of unsuccessful convergence cases reduces to zero.

The fraction of the residual variance explained by the individual unobserved heterogeneity, ρ, seems very slightly and insignificantly underestimated when using sequential logit models. It is still slightly and insignificantly underestimated when using sequential complementary log-log models, and it is more significantly underestimated when considering a sequential probit. Notice that a higher underestimation of the ρ coefficient seems to be associated with a higher attenuation bias for the coefficients. This result seems to confirm Baker and Melino's (2000) conclusion that an underestimation (overestimation) of the dispersion of the unobserved heterogeneity leads to an attenuation (amplification) of the covariate coefficients. However, we find that this attenuation (amplification) bias is simply due to a normalization problem that Baker and Melino (2000) did not notice.

To evaluate the effect of misspecifying the unobserved heterogeneity distribution on the duration dependence estimation, we plot the baseline hazard functions estimated using sequential logit, probit and complementary log-log with normal random effects: see Figures 3, 4 and 5 . In all three Figures we consider data simulated from a DGP given by a sequential logit model with the usual three types of distribution for the unobserved heterogeneity.

When both estimation and simulation models are given by a sequential logit model (Figure 3), the true baseline hazard (simulated) has a profile similar to the three estimated baseline hazards averaged over 100 samples (labeled discrete, gamma and normal) corresponding to three different DGPs, i.e. sequential logit models with random effects following a discrete, a gamma and a normal distribution.

When we change the estimation model to a sequential probit model with normal random effects (Figure 4), the estimated baseline hazards (labeled discrete, gamma and normal) have instead a different profile with respect to the true baseline hazard (labeled simulated) especially for long durations.

Finally, when using a sequential complementary log-log model with normal random effect for the estimation of the duration model, we find that the profile of the estimated baseline hazards (labeled discrete, gamma and normal) follow the true one (simulated), but the negative dependence is overestimated for short durations.

In summary, it seems that misspecification of the unobserved heterogeneity distribution does not seriously affect the estimation results. Changes in the error distribution (logistic, normal and extreme value) bias the duration dependence estimation but cause only a rescaling of the coefficients estimates by a constant factor. Because coefficients in binary models are identified only up to a scale normalization, the rescaling is not a genuine problem.

4 Conclusions

This paper assesses the effects of ignoring unobserved heterogeneity or misspecifying its distribution in single spell discrete time duration models. In particular, we focuse on assessing the consequences of adopting two models that can be easily estimated using standard software: sequential binary models with or without individual normal random effects.

The main findings from our Monte Carlo study can be summarized as follows. First, neglecting the unobserved heterogeneity seems to cause a bias in the duration dependence estimation. It does not seem to cause a bias in the covariate coefficients but rather a rescaling by a constant factor. Second, the rescaling factor is close to one when considering covariates i.i.d. across individuals and time, while it is significantly smaller than one for covariates that are i.i.d. across individuals and correlated across time. Third, misspecifying the random effects distribution biases neither the duration dependence nor the covariate coefficients estimation. Fourth, misspecifying the error distribution, assuming a normal or an extreme value distribution instead than a logistic one, seems to cause a bias in the duration dependence estimation while it seems to cause only a equiproportional rescaling of the covariate coefficients.

These findings are very encouraging for practitioners who estimate discrete time duration models with or without normal random effects using command built into standard statistical software packages.

References

Abbring J.H., van den Berg G.J. (2006) The unobserved heterogeneity distribution in duration analysis, Tinbergen Institute Discussion Papers 06-059/3, Tinbergen Institute.

Allison P. (1982) Discrete-time methods for the analysis of event histories, in: *Sociological Methodology*, Leinhardt S., ed., Jossey-Bass Publishers, San Francisco, 61–97.

Arumlampalam W. (1999) A note on estimated coefficients in random effects probit models, *Oxford Bulletin of Economics and Statistics*, 61, 4, 597–602.

Baker M., Melino A. (2000) Duration dependence and nonparametric heterogeneity: a Monte Carlo study, *Journal of Econometrics*, 96, 357–393.

Dolton P., van der Klaauw W. (1995) Leaving teaching in the UK: A duration analyis, *The Economic Journal*, 105, 429, 431–444.

Gaure S., Rø ed K., Zhang T. (2005) Time and causality: A Monte Carlo assessment of the timing of events approach, Memorandum 19/2005, Oslo University, Department of Economics.

Gourieroux C., Monfort A. (1996) *Simulation-based Econometric Methods*, Oxford University Press, New York, USA.

Greene W.H. (2003) *Econometric Analysis*, Prentice Hall, Fifth Edition, New York, USA.

Heckman J.J., Singer B. (1984) Econometric duration analysis, *Journal of Econometrics*, 24, 1-2, 63–132.

Holford T.R. (1976) Life tables with concomitant information, *Biometrics*, 32, 587–597.

Hougaard P., Myglegaard P., Borch-Johnsen K. (1994) Heterogeneity models of disease susceptibility with an application to diabetic nephropathy, *Biometrics*, 50, 1178–1188.

Jenkins S.P. (1995) Easy estimation methods for discrete-time duration models, *Oxford Bulletin of Economics and Statistics*, 57, 129–138.

Lancaster T. (1979) Econometric methods for the duration of unemployment, *Econometrica*, 47, 4, 939–956.

Lancaster T. (1985) Generalized residuals and heterogeneous duration models with applications to the Weibull model, *Journal of Econometrics*, 28, 113–126.

Lancaster T. (1990) *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Maddala G.S. (1987) Limited dependent variable models using panel data, *Journal of Human Resources*, 22, 307–337.

Meyer B.D. (1990) Unemployment insurance and unemployment spells, *Econometrica*, 58, 757–782.

Moffitt R.A., Gottschalk P. (2002) Trends in the transitory variance of earnings in the United States, *The Economic Journal*, 112, C68–C73.

Mroz T.A. , Zayats Y.V. (2005) Arbitrarily normalized coefficients, information sets and false reports of "biases" in binary outcome models, Working Paper 05-01, Department of Economics and the Carolina Population Center, University of North Carolina at Chapel Hill.

Narendranathan W., Stewart M.B. (1993) How does the benefit effect vary as unemployment spells lengthen?, *Journal of Applied Econometrics*, 8, 361–381.

Prentice R.L. , Gloeckler L.A. (1978) Regression analysis of grouped survival data with application to breast cancer data, *Biometrics*, 34, 57–67.

Ridder G. (1987) The sensitivity of duration models to misspecified unobserved heterogeneity and duration dependence, Working Paper, Groningen Univeristy, Groningen.

Ridder G. , Verbakel W. (1983) On the estimation of the proportional hazard model in the presence of unobserved heterogeneity, Working paper, University of Amsterdam, Faculty of Actuarial Science and Econometrics.

StataCorp (2005) *Longitudinal/Panel Data Reference Manual, Release 9*, Stata Corporation, College Station, TX.

Sueyoshi G.T. (1995) A class of binary response models for grouped duration data, *Journal of Applied Econometrics*, 10, 4, 411–431.

Train K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Trussel J., Richards T. (1985) Unemployment insurance and unemployment spells, *Sociological Methodology*, 15, 242–276.

Van den Berg G.J. (2001) Duration models: specification, identification and multiple durations, in: *Handbook of Econometrics, Volume V*, Heckman J.J., Leamer E., eds., North-Holland, Amsterdam.

Yamaguchi K. (1991) *Event History Analysis*, Sage Publications, Newbury Park, CA.

Zhang T. (2003) A Monte Carlo study on non-parametric estimation of duration models with unobserved heterogeneity, Memorandum 25/2003, Oslo University, Department of Economics.

Table 1: Means and standard deviations of the coefficients estimates over 100 samples. Monte Carlo exercise A.

	500 Observations						*1000 Observations*					
	Step DD			*Polynomial DD*			*Step DD*			*Polynomial DD*		
DGP	β_1	β_2	β_1/β_2	β_1	β_2	β_1/β_2	β_1	β_2	β_1/β_2	β_1	β_2	β_1/β_2
True Value	1	0.5	2	1	0.5	2	1	0.5	2	1	0.5	2
Time-varying covariates, ***A1***												
Positive duration dependence												
Discrete UH	0.896	0.446	2.061	0.892	0.445	2.058	0.934	0.473	1.999	0.932	0.472	1.998
(sd)	0.151	0.073	0.479	0.148	0.072	0.471	0.092	0.052	0.302	0.091	0.051	0.300
Gamma UH	0.967	0.464	2.144	0.963	0.462	2.146	0.951	0.478	2.015	0.948	0.477	2.015
(sd)	0.147	0.074	0.505	0.146	0.074	0.505	0.086	0.052	0.300	0.085	0.051	0.299
Normal UH	0.922	0.459	2.081	0.919	0.457	2.082	0.928	0.474	1.978	0.926	0.473	1.979
(sd)	0.158	0.087	0.526	0.158	0.086	0.526	0.099	0.050	0.296	0.099	0.049	0.295
Negative duration dependence												
Discrete UH	0.903	0.444	2.086	0.900	0.442	2.090	0.911	0.458	2.014	0.909	0.457	2.015
(sd)	0.142	0.070	0.472	0.142	0.069	0.469	0.090	0.053	0.308	0.090	0.053	0.307
Gamma UH	0.927	0.454	2.106	0.922	0.450	2.109	0.911	0.456	2.023	0.908	0.455	2.021
(sd)	0.152	0.074	0.531	0.148	0.072	0.525	0.107	0.057	0.306	0.107	0.056	0.306
Normal UH	0.895	0.446	2.047	0.891	0.443	2.051	0.905	0.464	1.977	0.903	0.463	1.977
(sd)	0.167	0.073	0.459	0.165	0.072	0.457	0.095	0.053	0.319	0.095	0.052	0.321
Time-invariant covariates, ***A2***												
Positive duration dependence												
Discrete UH	0.662	0.326	2.168	0.664	0.326	2.168	0.678	0.315	2.250	0.680	0.316	2.248
(sd)	0.147	0.077	0.790	0.147	0.077	0.787	0.112	0.060	0.675	0.112	0.060	0.672
Gamma UH	0.672	0.334	2.121	0.673	0.334	2.122	0.659	0.328	2.082	0.660	0.329	2.081
(sd)	0.144	0.075	0.688	0.145	0.075	0.687	0.104	0.061	0.533	0.104	0.061	0.530
Normal UH	0.730	0.350	2.277	0.731	0.350	2.274	0.684	0.341	2.051	0.685	0.341	2.051
(sd)	0.141	0.084	1.167	0.140	0.084	1.148	0.102	0.050	0.432	0.101	0.050	0.432
Negative duration dependence												
Discrete UH	0.740	0.355	2.161	0.739	0.354	2.167	0.726	0.351	2.113	0.724	0.350	2.113
(sd)	0.141	0.070	0.551	0.142	0.070	0.559	0.102	0.051	0.435	0.102	0.051	0.437
Gamma UH	0.625	0.316	2.075	0.622	0.314	2.075	0.613	0.303	2.105	0.611	0.303	2.105
(sd)	0.132	0.067	0.654	0.133	0.066	0.660	0.099	0.058	0.588	0.099	0.058	0.588
Normal UH	0.709	0.342	2.194	0.707	0.341	2.192	0.660	0.340	1.985	0.659	0.339	1.986
(sd)	0.136	0.082	0.681	0.136	0.081	0.677	0.102	0.052	0.427	0.102	0.052	0.428
Mixture covariates, ***A3***												
Positive duration dependence												
Discrete UH	0.789	0.390	2.114	0.788	0.390	2.117	0.809	0.399	2.068	0.809	0.399	2.069
(sd)	0.154	0.083	0.636	0.153	0.083	0.638	0.113	0.058	0.418	0.114	0.058	0.419
Gamma UH	0.810	0.408	2.077	0.807	0.407	2.077	0.794	0.404	2.005	0.793	0.403	2.007
(sd)	0.159	0.082	0.637	0.159	0.083	0.635	0.115	0.056	0.414	0.114	0.056	0.414
Normal UH	0.811	0.402	2.087	0.808	0.401	2.086	0.810	0.402	2.056	0.810	0.401	2.057
(sd)	0.148	0.073	0.546	0.148	0.073	0.548	0.105	0.058	0.386	0.106	0.058	0.388
Negative duration dependence												
Discrete UH	0.828	0.410	2.082	0.824	0.408	2.084	0.835	0.416	2.044	0.833	0.414	2.048
(sd)	0.157	0.079	0.534	0.157	0.078	0.537	0.101	0.054	0.391	0.102	0.054	0.398
Gamma UH	0.791	0.390	2.110	0.787	0.387	2.113	0.761	0.375	2.074	0.759	0.374	2.073
(sd)	0.141	0.074	0.598	0.141	0.074	0.593	0.109	0.055	0.419	0.108	0.055	0.422
Normal UH	0.796	0.393	2.083	0.792	0.391	2.086	0.786	0.393	2.025	0.785	0.392	2.027
(sd)	0.153	0.067	0.553	0.153	0.067	0.557	0.102	0.053	0.313	0.102	0.053	0.313

Note: Characteristics of the DGPs (data generator processes) and of the estimation models are given by row and by column. UH = unobserved heterogeneity. DD = duration dependence. Step = step function. Poynomial= cubic polynomial function.

Table 2: Means and standard deviations of the coefficients estimates over 100 samples. Monte Carlo exercise B. Estimation model with step function duration dependence.

	β_1	β_2	β_3	β_4	β_1/β_2	β_3/β_4	β_3/β_2	β_3/β_1	β_1/β_4	β_4/β_2
True Value	1	0.5	1	0.5	2	2	2	1	2	1
DGP										
					500 Observations:					
Positive duration dependence:										
Discrete UH	0.479	0.251	0.575	0.273	2.091	2.327	2.512	1.377	1.940	1.194
	(0.155)	(0.070)	(0.131)	(0.077)	(0.994)	(0.995)	(1.024)	(0.699)	(1.035)	(0.564)
Gamma UH	0.507	0.236	0.584	0.295	2.470	2.128	2.886	1.260	1.867	1.456
	(0.148)	(0.077)	(0.150)	(0.074)	(1.428)	(0.821)	(1.628)	(0.560)	(0.846)	(0.966)
Normal UH	0.489	0.239	0.574	0.283	2.664	2.225	3.237	1.295	1.863	1.531
	(0.145)	(0.080)	(0.170)	(0.071)	(3.397)	(1.116)	(5.616)	(0.601)	(0.858)	(2.068)
No duration dependence:										
Discrete UH	0.510	0.255	0.589	0.283	2.144	2.197	2.517	1.258	1.897	1.197
	(0.139)	(0.064)	(0.146)	(0.061)	(0.857)	(0.772)	(1.074)	(0.504)	(0.707)	(0.489)
Gamma UH	0.453	0.239	0.548	0.275	2.115	2.153	2.596	1.428	1.812	1.270
	(0.144)	(0.073)	(0.149)	(0.066)	(1.030)	(0.928)	(1.360)	(1.256)	(1.012)	(0.546)
Normal UH	0.471	0.235	0.553	0.278	2.276	2.119	2.640	1.348	1.792	1.319
	(0.148)	(0.072)	(0.144)	(0.067)	(1.242)	(0.808)	(1.738)	(0.734)	(0.758)	(0.647)
					1000 Observations:					
Positive duration dependence:										
Discrete UH	0.469	0.233	0.557	0.276	2.226	2.109	2.680	1.259	1.770	1.306
	(0.105)	(0.061)	(0.115)	(0.054)	(1.093)	(0.674)	(1.637)	(0.423)	(0.545)	(0.628)
Gamma UH	0.469	0.238	0.554	0.282	2.131	2.034	2.503	1.231	1.720	1.277
	(0.096)	(0.055)	(0.093)	(0.052)	(0.943)	(0.508)	(0.976)	(0.331)	(0.483)	(0.484)
Normal UH	0.498	0.239	0.566	0.284	2.216	2.053	2.518	1.197	1.806	1.272
	(0.101)	(0.054)	(0.112)	(0.050)	(0.799)	(0.542)	(0.856)	(0.389)	(0.498)	(0.462)
No duration dependence:										
Discrete UH	0.496	0.248	0.576	0.284	2.130	2.097	2.480	1.209	1.800	1.219
	(0.100)	(0.056)	(0.108)	(0.054)	(0.881)	(0.535)	(1.008)	(0.340)	(0.468)	(0.445)
Gamma UH	0.444	0.224	0.547	0.271	2.122	2.106	2.588	1.316	1.699	1.288
	(0.104)	(0.051)	(0.100)	(0.053)	(0.864)	(0.616)	(0.859)	(0.462)	(0.516)	(0.475)
Normal UH	0.487	0.233	0.552	0.286	2.259	1.995	2.563	1.194	1.754	1.327
	(0.106)	(0.057)	(0.100)	(0.050)	(0.934)	(0.518)	(0.982)	(0.351)	(0.514)	(0.514)
					5000 Observations:					
Positive duration dependence:										
Discrete UH	0.472	0.231	0.561	0.278	2.070	2.033	2.463	1.198	1.708	1.219
	(0.045)	(0.027)	(0.045)	(0.023)	(0.289)	(0.266)	(0.354)	(0.157)	(0.210)	(0.154)
Gamma UH	0.466	0.236	0.562	0.285	1.989	1.987	2.405	1.221	1.643	1.217
	(0.047)	(0.024)	(0.045)	(0.022)	(0.289)	(0.236)	(0.334)	(0.168)	(0.212)	(0.152)
Normal UH	0.481	0.243	0.568	0.283	2.004	2.019	2.362	1.192	1.709	1.177
	(0.049)	(0.023)	(0.047)	(0.024)	(0.303)	(0.233)	(0.300)	(0.144)	(0.222)	(0.144)
No duration dependence:										
Discrete UH	0.502	0.245	0.570	0.285	2.067	2.013	2.349	1.143	1.773	1.172
	(0.040)	(0.024)	(0.044)	(0.023)	(0.258)	(0.226)	(0.306)	(0.129)	(0.207)	(0.127)
Gamma UH	0.445	0.223	0.549	-0.279	2.014	1.979	2.490	1.248	1.603	1.264
	(0.046)	(0.025)	(0.043)	(0.023)	(0.303)	(0.218)	(0.348)	(0.161)	(0.213)	(0.162)
Normal UH	0.467	0.235	0.566	-0.279	2.006	2.039	2.431	1.224	1.682	1.198
	(0.047)	(0.023)	(0.046)	(0.020)	(0.285)	(0.225)	(0.304)	(0.150)	(0.207)	(0.140)

Note: Characteristics of the DGPs (data generator processes) are given by row. UH = unobserved heterogeneity.

Table 3: Means and standard deviations of the coefficients estimates over 100 samples. Monte Carlo exercise B. Estimation model ignoring duration dependence.

	β_1	β_2	β_3	β_4	β_1/β_2	β_3/β_4	β_3/β_2	β_3/β_1	β_1/β_4	β_4/β_2
True Value	1	0.5	1	0.5	2	2	2	1	2	1
DGP										
					500 Observations:					
Positive duration dependence:										
Discrete UH	0.627	0.330	0.636	0.297	2.099	2.412	2.140	1.189	2.393	1.001
	(0.216)	(0.094)	(0.155)	(0.088)	(1.054)	(1.184)	(0.985)	(0.657)	1.539)	(0.495)
Gamma UH	0.624	0.310	0.632	0.327	2.231	2.187	2.327	1.124	2.101	1.182
	(0.186)	(0.097)	(0.187)	(0.094)	(0.996)	(1.247)	(1.261)	(0.545)	(0.974)	(0.576)
Normal UH	0.638	0.311	0.635	0.312	2.676	2.266	2.722	1.108	2.234	1.314
	(0.198)	(0.109)	(0.196)	(0.086)	(3.082)	(1.219)	(3.943)	(0.525)	(1.109)	(1.688)
No duration dependence:										
Discrete UH	0.657	0.329	0.654	0.311	2.133	2.234	2.165	1.091	2.239	1.019
	(0.184)	(0.082)	(0.169)	(0.070)	(0.840)	(0.839)	(0.954)	(0.448)	(0.914)	(0.425)
Gamma UH	0.548	0.277	0.611	0.294	2.371	2.490	2.454	1.315	2.214	1.311
	(0.194)	(0.106)	(0.192)	(0.091)	(2.619)	(2.113)	(3.618)	(0.883)	(1.680)	(2.051)
Normal UH	0.599	0.301	0.611	0.307	2.275	2.159	2.302	1.193	2.092	1.146
	(0.195)	(0.095)	(0.165)	(0.081)	(1.262)	(0.918)	(1.532)	(0.709)	(0.941)	(0.594)
					1000 Observations:					
Positive duration dependence:										
Discrete UH	0.614	0.305	0.608	0.303	2.264	2.102	2.308	1.054	2.119	1.118
	(0.141)	(0.083)	(0.133)	(0.061)	(1.220)	(0.732)	(1.786)	(0.372)	(0.715)	(0.620)
Gamma UH	0.620	0.299	0.641	0.314	2.232	2.101	2.345	1.103	2.032	1.151
	(0.149)	(0.078)	(0.133)	(0.058)	(0.860)	(0.544)	(1.153)	(0.394)	(0.582)	(0.530)
Normal UH	0.652	0.309	0.624	0.316	2.240	2.047	2.150	1.013	2.135	1.094
	(0.138)	(0.072)	(0.132)	(0.058)	(0.806)	(0.589)	(0.740)	(0.339)	(0.609)	(0.400)
No duration dependence:										
Discrete UH	0.638	0.318	0.635	0.316	2.133	2.085	2.138	1.041	2.089	1.058
	(0.131)	(0.072)	(0.122)	(0.063)	(0.878)	(0.560)	(0.917)	(0.308)	(0.573)	(0.403)
Gamma UH	0.590	0.278	0.579	0.296	2.279	2.050	2.212	1.059	2.081	1.138
	(0.151)	(0.067)	(0.122)	(0.063)	(0.895)	(0.636)	(0.766)	(0.409)	(0.700)	(0.402)
Normal UH	0.622	0.295	0.606	0.318	2.277	1.970	2.215	1.030	2.013	1.167
	(0.140)	(0.074)	(0.115)	(0.057)	(0.934)	(0.555)	(0.833)	(0.317)	(0.606)	(0.445)
					5000 Observations:					
Positive duration dependence:										
Discrete UH	0.615	0.302	0.616	0.306	2.064	2.035	2.072	1.012	2.024	1.025
	(0.060)	(0.036)	(0.054)	(0.027)	(0.292)	(0.288)	(0.314)	(0.142)	(0.247)	(0.133)
Gamma UH	0.615	0.301	0.627	0.310	2.064	2.037	2.108	1.033	1.997	1.043
	(0.068)	(0.033)	(0.049)	(0.027)	(0.326)	(0.250)	(0.298)	(0.146)	(0.298)	(0.149)
Normal UH	0.626	0.317	0.631	0.312	1.997	2.036	2.009	1.018	2.021	0.994
	(0.065)	(0.031)	(0.058)	(0.028)	(0.310)	(0.248)	(0.273)	(0.131)	(0.272)	(0.130)
No duration dependence										
Discrete UH	0.645	0.316	0.632	0.316	2.060	2.012	2.021	0.987	2.053	1.009
	(0.053)	(0.031)	(0.052)	(0.026)	(0.255)	(0.234)	(0.265)	(0.119)	(0.244)	(0.114)
Gamma UH	0.552	0.287	0.603	0.299	1.956	2.034	2.134	1.107	1.862	1.056
	(0.068)	(0.034)	(0.057)	(0.028)	(0.370)	(0.281)	(0.358)	(0.160)	(0.293)	(0.153)
Normal UH	0.597	0.301	0.629	0.307	2.005	2.059	2.112	1.064	1.953	1.032
	(0.061)	(0.030)	(0.055)	(0.024)	(0.291)	(0.248)	(0.278)	(0.135)	(0.253)	(0.128)

Note: Characteristics of the DGPs (data generator processes) are given by row. UH = unobserved heterogeneity.

Table 4: Means and standard deviations of coefficient estimates over 100 samples. Estimation model: sequential logit. DGP: sequential logit.

	β_1	β_2	β_1/β_2	$\rho = \frac{\sigma_\theta^2}{\sigma_\epsilon^2+\sigma_\theta^2}$	Iterations	Convergence
True Value	1	0.5	2	$\frac{1}{\frac{\pi^2}{3}+1} = 0.233$		
DGP						
			500 Observations:			
Negative duration dependence:						
Discrete Unobserved Het.	0.927	0.455	2.130	0.174	6.908	98
	(0.170)	(0.093)	(0.611)	(0.143)	(2.981)	
Gamma Unobserved Het.	0.913	0.452	2.110	0.118	5.404	99
	(0.183)	(0.087)	(0.672)	(0.072)	(2.263)	
Normal Unobserved Het.	0.923	0.460	2.083	0.148	6.271	96
	(0.158)	(0.086)	(0.550)	(0.116)	(2.759)	
			1000 Observations:			
Negative duration dependence:						
Discrete Unobserved Het.	0.937	0.462	2.069	0.156	6.424	99
	(0.137)	(0.071)	(0.405)	(0.093)	(2.607)	
Gamma Unobserved Het.	0.942	0.474	2.012	0.151	5.889	99
	(0.123)	(0.060)	(0.317)	(0.090)	(2.788)	
Normal Unobserved Het.	0.944	0.470	2.048	0.170	6.316	98
	(0.121)	(0.066)	(0.393)	(0.118)	(2.775)	
			5000 Observations:			
Negative duration dependence:						
Discrete Unobserved Het.	0.956	0.477	2.013	0.202	6.850	100
	(0.079)	(0.040)	(0.178)	(0.099)	(2.851)	
Gamma Unobserved Het.	0.916	0.458	2.003	0.110	4.190	100
	(0.064)	(0.027)	(0.163)	(0.059)	(1.522)	
Normal Unobserved Het.	0.966	0.482	2.009	0.188	5.850	100
	(0.076)	(0.034)	(0.161)	(0.102)	(2.455)	

Note: Iterations = average number of iterations for the convergence of the likelihood maximization algorithm. Convergence = number of cases over 100 replications of successful convergence. ρ = fraction of residual variance explained by individual random effects.

Table 5: Means and standard deviations of coefficient estimates over 100 samples. Estimation model: sequential probit. DGP: sequential logit.

	β_1	β_2	β_1/β_2	$\rho = \frac{\sigma_\theta^2}{\sigma_\epsilon^2+\sigma_\theta^2}$	Iterations	Convergence
True Value	1	0.5	2	$\frac{1}{1+1} = 0.5$		
DGP						
500 Observations:						
Negative duration dependence:						
Discrete Unobserved Het.	0.519	0.255	2.128	0.306	8.690	100
	(0.099)	(0.058)	(0.587)	(0.140)	(1.495)	
Gamma Unobserved Het.	0.543	0.270	2.069	0.307	8.535	99
	(0.102)	(0.049)	(0.522)	(0.126)	(1.358)	
Normal Unobserved Het.	0.528	0.264	2.089	0.304	8.455	99
	(0.102)	(0.058)	(0.566)	(0.157)	(2.370)	
1000 Observations:						
Negative duration dependence:						
Discrete Unobserved Het.	0.525	0.257	2.078	0.292	8.848	99
	(0.082)	(0.039)	(0.407)	(0.104)	(1.480)	
Gamma Unobserved Het.	0.534	0.270	2.013	0.305	8.410	100
	(0.087)	(0.039)	(0.396)	(0.112)	(1.326)	
Normal Unobserved Het.	0.537	0.268	2.049	0.324	8.760	100
	(0.079)	(0.041)	(0.407)	(0.124)	(1.457)	
5000 Observations:						
Negative duration dependence:						
Discrete Unobserved Het.	0.524	0.261	2.015	0.311	9.010	100
	(0.036)	(0.017)	(0.174)	(0.044)	(1.259)	
Gamma Unobserved Het.	0.542	0.271	2.004	0.305	8.740	100
	(0.037)	(0.015)	(0.159)	(0.042)	(0.960)	
Normal Unobserved Het.	0.538	0.268	2.010	0.315	8.850	100
	(0.038)	(0.016)	(0.161)	(0.049)	(1.175)	

Note: Iterations = average number of iterations for the convergence of the likelihood maximization algorithm. Convergence = number of cases over 100 replications of successful convergence. ρ = fraction of residual variance explained by individual random effects.

Table 6: Means and standard deviations of coefficient estimates over 100 samples. Estimation model: sequential complementary log-log. DGP: sequential logit.

	β_1	β_2	β_1/β_2	$\rho = \frac{\sigma^2_\theta}{\sigma^2_\epsilon+\sigma^2_\theta}$	Iterations	Convergence
True Value	1	0.5	2	$\frac{1}{\frac{\pi^2}{6}+1} = 0.378$		
DGP						
			500 Observations:			
Negative duration dependence:						
Discrete Unobserved Het.	0.855	0.421	2.123	0.243	6.714	98
	(0.147)	(0.086)	(0.606)	(0.127)	(2.428)	
Gamma Unobserved Het.	0.861	0.442	2.010	0.222	5.890	100
	(0.156)	(0.072)	(0.545)	(0.112)	(2.238)	
Normal Unobserved Het.	0.859	0.430	2.076	0.224	6.358	95
	(0.152)	(0.081)	(0.556)	(0.130)	(2.475)	
			1000 Observations:			
Negative duration dependence:						
Discrete Unobserved Het.	0.871	0.429	2.069	0.233	6.602	98
	(0.123)	(0.062)	(0.410)	(0.094)	(2.560)	
Gamma Unobserved Het.	0.886	0.432	2.084	0.228	5.730	100
	(0.107)	(0.051)	(0.381)	(0.084)	(2.348)	
Normal Unobserved Het.	0.879	0.437	2.053	0.250	6.240	100
	(0.109)	(0.062)	(0.386)	(0.131)	(2.590)	
			5000 Observations:			
Negative duration dependence:						
Discrete Unobserved Het.	0.882	0.440	2.013	0.285	6.710	100
	(0.058)	(0.029)	(0.178)	(0.089)	(2.388)	
Gamma Unobserved Het.	0.878	0.438	2.011	0.224	5.310	100
	(0.054)	(0.028)	(0.160)	(0.075)	(2.246)	
Normal Unobserved Het.	0.901	0.450	2.008	0.287	6.290	100
	(0.060)	(0.027)	(0.161)	(0.097)	(2.262)	

Note: Iterations = average number of iterations for the convergence of the likelihood maximization algorithm. Convergence = number of cases over 100 replications of successful convergence. ρ = fraction of residual variance explained by individual random effects.

Figure 1: Estimated and true negative duration dependence functions. Monte Carlo exercise **A1**. Unobserved heterogeneity ignored.

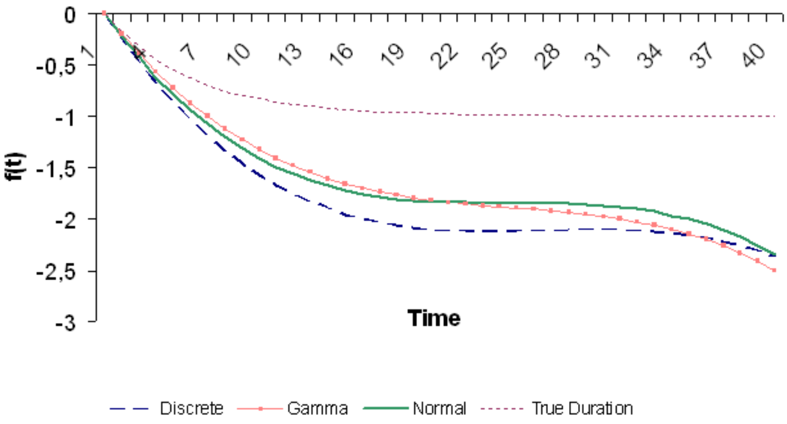


Figure 2: Estimated and true positive duration dependence functions. Monte Carlo exercise **A1**. Unobserved heterogeneity ignored.

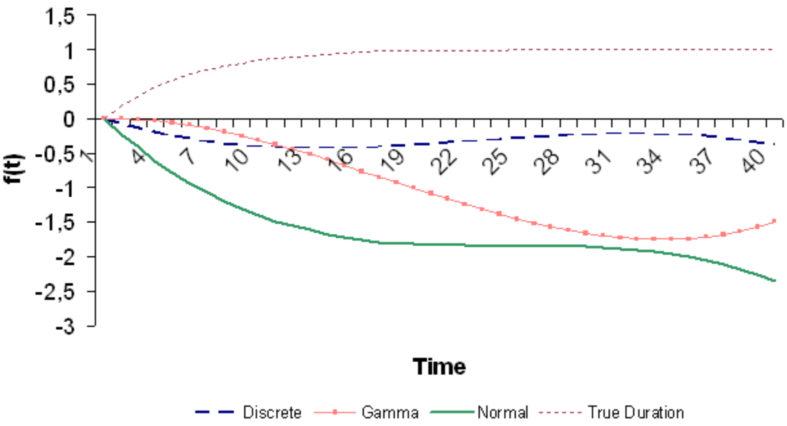


Figure 3: Estimated and true baseline hazards. Estimation model: sequential logit with normal random effects. DGP: sequential logit with unobserved heterogeneity.

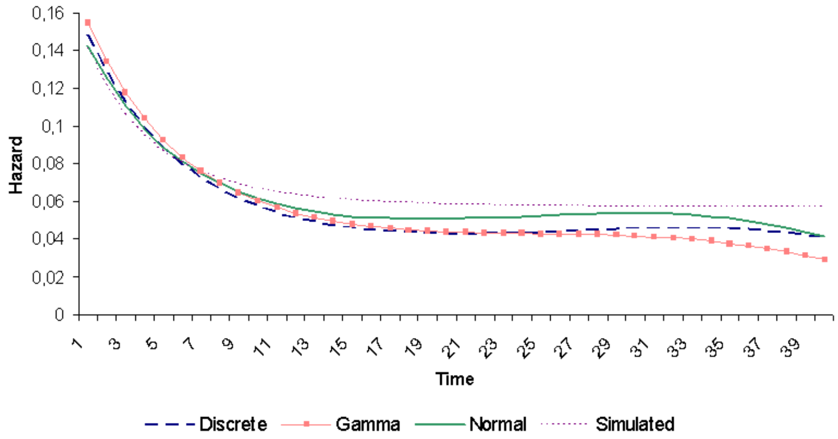


Figure 4: Estimated and true baseline hazards. Estimation model: sequential probit with normal random effects. DGP: sequential logit with unobserved heterogeneity.

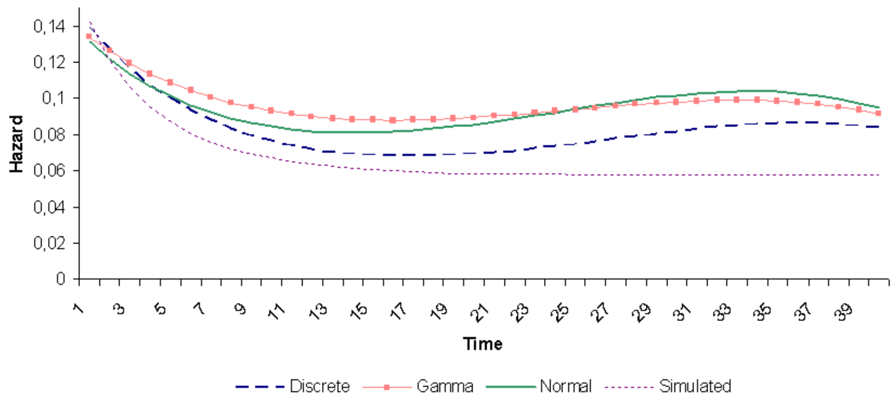
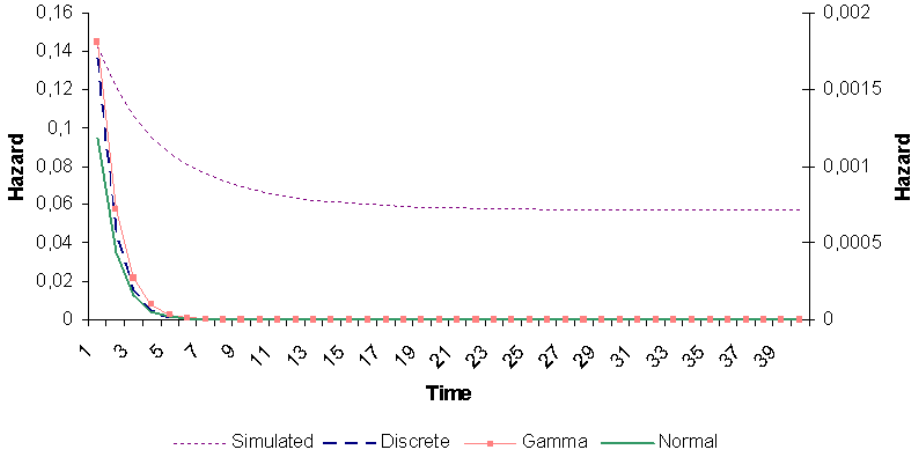


Figure 5: Estimated and true baseline hazards. Estimation model: sequential complementary log-log with normal random effects. DGP: sequential logit with unobserved heterogeneity.



A Appendix

Let us consider a sequential probit model with unobserved heterogeneity given by a normal random variable identically and independently distributed (i.i.d.) across individuals and time. Moreover, let d_{it} be a dummy variable indicating the risk event occurrence for the generic individual i ($i = 1, ..., n$) at time, duration, t ($t = 1, ..., T$). Then the hazard probability for a generic $i - th$ individual at time t is given by

$$Pr(d_{it} = 1 | d_{it-1} = 0) \quad = \quad \Phi(X_{it}\beta + f(t) + \theta_{it}), \tag{11}$$

where X_{it} is a vector of explanatory variables, β is the vector of the corresponding coefficients, $f(t)$ is a known deterministic function of the duration, θ_{it} is an unobserved random variable i.i.d. across individuals and time as normal with mean zero and variance σ^2_θ and independent of the explanatory variables, and Φ is the Gaussian cumulative distribution. If θ_{it} were observed and $f(t)$ were known, then the maximum likelihood estimation of the sequential probit conditioning to X_{it} and θ_{it} would produce consistent estimates for the β coefficients.

If we omit the unobserved heterogeneity, the hazard probability would be instead:

$$Pr(d_{it} = 1 | d_{it-1} = 0) \quad = \quad \Phi(X_{it}\tilde{\beta} + \tilde{f}(t)), \tag{12}$$

where $\tilde{\beta} = \frac{\beta}{\sigma}$, $\tilde{f}(\cdot)$ is given by $\frac{1}{\sigma} f(\cdot)$, and $\sigma = \sqrt{\sigma^2_\theta + 1}$. Assuming that $\tilde{f}(\cdot)$ is known, we can still use the maximum likelihood method to estimate consistently the explanatory variable coefficients (see Maddala, 1987). Notice that the weeding out effect does not operate here because the unobserved heterogeneity is given by a random variables i.i.d. across individuals and time. People who survive at time t because of a low unobserved random component may have a high unobserved random component in $(t + 1)$, so that θ_{it} and X_{it} are independent conditioning to $t = 0$ but also conditioning to $t > 0$.

Finally, notice that the new coefficients in (12) are rescaled and therefore not directly comparable with the coefficients in the hazard model (11) (see Arulampalam, 1999). Since the rescaling factor is given by $\frac{1}{\sigma}$ and $\sigma > 1$, ignoring the unobserved heterogeneity causes an attenuation bias for the covariate coefficients.