BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

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

by Cheti Nicoletti and Concetta Rondinelli

March 2009

Number 705

THE (MIS)SPECIFICATION OF DISCRETE DURATION MODELS WITH UNOBSERVED HETEROGENEITY: A MONTE CARLO STUDY

by Cheti Nicoletti[†] and Concetta Rondinelli[‡]

Abstract

Empirical researchers usually prefer statistical models that can be easily estimated using standard software packages. One such model is the sequential binary model with or without normal random effects; such models can be adopted to estimate discrete duration models with unobserved heterogeneity. But ease of estimation may come at a cost. In this paper we conduct a Monte Carlo simulation to evaluate the consequences of omitting or misspecifying the unobserved heterogeneity distribution in single-spell discrete duration models.

JEL Classification: C23, C25.
Keywords: discrete duration models, unobserved heterogeneity, Monte Carlo simulations.

Contents

[†] Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK.
[‡] Bank of Italy, Economic Outlook and Monetary Policy Department, Rome, Italy.

1 Introduction[1]

This paper provides a guide for applied econometricians in using conventional software to estimate single-spell discrete duration models. To this aim we carry out a set of Monte Carlo exercises to evaluate the consequences, in terms of bias, of using sequential binary models with or without normal random effects.

As first noticed by Yamaguchi (1991) and Jenkins (1995), single- spell discrete duration models can be easily estimated by considering parametric models for repeated binary measures such as probit, logit or complementary log-log. Moreover, they can be extended to take account of unobserved heterogeneity by introducing a random component, which represents a scalar function of time-invariant unobserved variables.[2] It is then possible to estimate model parameters by maximizing the likelihood function integrated over the unobserved random effect. The resulting model is a mixture of hazard functions with respect to the unobserved random component. The estimation of these mixture models requires either assuming a specific parametric distribution for the random component, or using a non-parametric maximum likelihood estimation.

In principle non-parametric maximum likelihood estimation is the best solution to minimize the potential bias caused by improper parametric distributional assumptions.[3] Nevertheless, the computation of the non-parametric estimator is not usually feasible using commands built into common software packages. For this reason many non-specialists adopt easier estimation methods either by imposing specific parametric distributions for the unobserved heterogeneity or by ignoring the unobserved heterogeneity altogether.

[1]We would like to thank John Ermisch, Stephen Jenkins, Alfonso Rosolia, Ignazio Visco and seminar participants at the European Economic Association Annual Meeting 2007 (Budapest) and at the North American Summer Meeting of the Econometric Society 2008 (Pittsburgh) for their helpful comments and suggestions. The research was supported by the Economic and Social Research Council through their grant to the Research Centre on Micro-social Change in ISER. The usual disclaimer applies: the opinions expressed in this paper are those of the authors and do not involve the Bank of Italy.

[2]This way to control for unobserved heterogeneity was first introduced for continuous duration models; see Lancaster (1979, 1990), Heckman and Singer (1984) and van den Berg (2001).

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

In discrete duration models the assumption of a normal distribution can be computationally convenient.[4] Under this assumption, discrete duration models can be easily estimated as binary models with normal random effects using widely available statistical softwares.[5]

In this paper we use a Monte Carlo study to evaluate the consequences of ignoring the unobserved heterogeneity or misspecifying its parametric distribution when estimating single-spell discrete duration models. While Baker and Melino (2000), Zhang (2003), Gaure et al. (2007) and Mroz and Zayats (2008) consider the consequences of choosing different numbers of support points when imposing a non-parametric distribution for the unobserved heterogeneity,[6] we evaluate the consequences of imposing a normal random effect. On the other hand, similarly to Baker and Melino (2000), Zhang (2003) and Gaure et al. (2007), we consider the effect of neglecting unobserved heterogeneity.

One important issue - overlooked by Baker and Melino (2000), Zhang (2003) and Gaure et al. (2007) - 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 too 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 that are attenuated (amplified). As suggested by Mroz and Zayats (2008), the attenuation (amplification) biases that Baker and Melino (2000), Zhang (2003) and Gaure et al. (2007) find could be due at least in part to this neglected issue.

In this paper we show that the coefficient bias, caused by the omission of the unobserved heterogeneity or by imposing an incorrect normality assumption, can be a consequence of the

[4]In continuous 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, 2007).

[5]For example, Stata provides the commands `xtcloglog`, `xtlogit` and `xtprobit` (`cloglog`, `logit` and `probit`) to estimate binary models with normal random effects (without normal random effects) and error terms with extreme value, logistic and normal distributions. For more details on discrete duration models we refer to Holford (1976), Prentice and Gloeckler (1978), Allison (1982), Narendranathan and Stewart (1993) and Sueyoshi (1995).

[6]Studies of the consequences of misspecification of the unobserved heterogeneity in continuous duration models have been conducted by Heckman and Singer (1984), Lancaster (1985), Trussell and Richards (1985), Ridder (1987), Huh and Sickles (1994) and Dolton and van der Klaauw (1995).

coefficient normalization issue. Given this difficulty in comparing covariate coefficients across different models, we also consider another way to evaluate the consequences of adopting simplified duration models. We check whether their predicted effects of changes in covariates on expected duration and survival probabilities are close to the true effects.

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 discuss the theoretical consequences and then assess these possible consequences through a Monte Carlo simulation exercise. In Section 4 we summarize the main findings.

2 Ignoring unobserved heterogeneity

2.1 Consequences

Ignoring unobserved heterogeneity in duration models can cause a bias in estimating the duration dependence. More precisely, omitting the unobserved heterogeneity causes an overestimation of the negative duration dependence (see for example Lancaster, 1990 and van den Berg, 2001). People who have a high unobserved random component are more likely to complete their duration early, so that the sample of individuals that survive is a selected sample with relatively small random effects.[7] This selection process is known as "weeding out" or "sorting effect".

Omitting unobserved heterogeneity may also bias the coefficients estimation of the explanatory variables in the hazard model. For example, neglecting unobserved heterogeneity in mixed proportional (continuous time) hazard models causes an underestimation of the proportionate response of the hazard function with respect to the explanatory variables.[8]

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 cor-

[7] Notice that, without loss of generality, we assume that the unobserved random component is positively related to the hazard function.

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

responding coefficients. Without loss of generality, we assume in this section that the hazard function conditional on the observed explanatory variables and the unobserved heterogeneity is 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 survivors 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 do survivors with a small μ, so that the difference between the observed hazard functions is on average lower than that 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 individual i, θ_i, and 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 $(T_i \mid X_i, \theta_i)$ is identically and independently distributed (i.i.d.) across individuals.

There are some continuous 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 omitting unobserved heterogeneity in mixed proportional hazard models with baseline distribution given by a Weibull distribution causes a rescaling by a constant factor for all coefficients. Ridder (1987) analytically shows 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 other 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), when they consider the omission of unobserved heterogeneity in discrete duration models with single spells. This contradictory result may be due to the fact that in discrete duration models the 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 easy 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 $\tau \geq 0$. The proof is an application of the analytical results in Arulampalam (1999).

Similar analytical results do not exist for more general discrete duration models. In this paper we carry out a Monte Carlo simulation exercise to study the consequences of omitting unobserved heterogeneity in more general cases. In particular, we simulate sequential binary models where

a. the unobserved random effect 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. and the covariates are i.i.d. across individuals and time or i.i.d. across individuals but not time.

Cases (a) and (b) were considered by Baker and Melino (2000) who found that ignoring unobserved heterogeneity component causes an attenuation bias for the covariate coefficients. We replicate their Monte Carlo study to re-evaluate the consequences of ignoring unobserved heterogeneity, but take into account the coefficient normalization.

Mroz and Zayats (2008) reconsider 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) conclude that non-parametric maximum likelihood estimation that penalizes specifications with many mass points produces better results; however, Mroz and Zayats (2008) present opposite results.

Case (c) is a useful extension to understand whether the estimation bias depends on the type of covariates used. If the covariates X_{it} are i.i.d. across individuals and time, then the estimation bias should be reduced because the independence between the unobserved component and the observed covariates tends to hold even when conditioning to $T_i > 0$, that is $(\theta_i \perp\!\!\!\perp X_{it} \mid T_i \geq \tau)$, where $\tau > 0$.

If, on the contrary, covariates are i.i.d. across individuals but time-invariant or correlated across time, then we expect an attenuation bias. However, this bias could consist of a rescaling by a constant factor for all covariate coefficients.

2.2 Monte Carlo simulation: Data Generating Processes

We consider the same data generating processes (DGPs) used in the Monte Carlo study of Baker and Melino (2000) and generalize them to consider both time-varying and time-invariant explanatory variables.

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 (terminating event) can occur only in discrete time. We record an event taking place in the interval $(t-1, t]$ as occurred in t.

We assume that the probability that an individual i experiences a terminating event in t conditional on survival to $(t-1)$ 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 in t for individual i, and z^*_{it} is a continuous latent 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.[9] Then we can 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 differing specifications for the observed explanatory variables, X_{it}, the duration dependence function, $f(t)$, and the unobserved heterogeneity, θ_i, we produce different DGPs.

We organize the simulations in two main sets. In the first set, exercise **A**, we focus on the effect of omitting unobserved heterogeneity when using different types of explanatory variables. In particular the three DGPs use three 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 of these DGPs we consider two types of duration dependence function, one increasing and one decreasing, and three distributions for the unobserved heterogeneity, a discrete (with two support points), 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 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 types of distribution for the unobserved heterogeneity, whereas only one specification is given to the duration function and the vector of covariates that includes both time-invariant and mixture variables. This second simulation exercise produces six types of DGPs.

[9] The definition of the above discrete hazard model and the notation used are consistent with Baker and Melino (2000).

For each of the DGPs in simulation exercises **A** and **B** we consider three sample sizes: 500, 1000 and 5000 individuals.

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

In the following, we discuss in more detail how the explanatory variables, the duration dependence function and the unobserved heterogeneity distribution are specified 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** the observed heterogeneity of the hazard model is specified as follows:

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

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 and two independent variables defined as the sum of a time-invariant variable and a time-varying one, 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 that is interesting from a theoretical viewpoint but less common 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: 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 the observed heterogeneity of the hazard model as:

$$X_{it}\,\beta \;=\; X_{1,i}\,\beta_1 + X_{2,i}\,\beta_2 + X_{3,it}\,\beta_3 + X_{4,it}\,\beta_4 \tag{6}$$

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 positive duration dependence and

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

for a negative 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 negative duration dependence.

Unobserved Heterogeneity.

In both exercises **A** and **B** three distributions for the unobserved heterogeneity θ_i are assumed: discrete, gamma and normal. 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 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 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 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 step function.

2.4 Results

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

The results of exercise **A** are reported in Table 1, which is divided into three panels providing the estimated coefficients for time-varying covariates (top panel **A1**), time-invariant covariates (middle panel **A2**) and mixture covariates (bottom panel **A3**). We report 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 types of DGPs: sequential logit model with negative or positive duration dependence and with unobserved heterogeneity following a discrete with two mass points, a gamma or a normal distribution (labeled "Discrete UH", "Gamma UH" and "Normal UH"). By column we specify instead the sample size (500 or 1000 observations or 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 an attenuation bias because of a rescaling by a constant factor of the coefficients, then the ratio between coefficients would be correctly estimated. This seems supported by the results in Table 1 when using any type of covariates. Moreover, when using time-varying covariates that are i.i.d. across individuals and time (top panel **A1**), the attenuation problem for the coefficients does not seem significant. When, instead, the covariates are i.i.d. across individuals and time-invariant, the attenuation problem is more severe (middle panel **A2**). Finally, the attenuation bias magnitude seems to be intermediate between the two previous extreme cases for mixture covariates (bottom panel **A3**).

Using different distributions for the simulated unobserved heterogeneity components and different specifications for the simulated duration dependence (negative or positive) produces some very small and insignificant differences in the coefficients. Similarly, the way we estimate the duration dependence (by considering either a step or a cubic polynomial function) does not affect 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 its true value. We find again a slight improvement in the results when the number of observation are increased to 5000 (results are not reported but are available upon request to the authors).

To summarize, ignoring unobserved heterogeneity in sequential logit models seems to cause an attenuation of the covariate coefficients due to a rescaling by a constant factor. This attenuation bias is alomost completely canceled when using covariates that are i.i.d. across individuals and time, while it is very significant when the covariates are highly autocorrelated.

As emphasized in Section 2.1, ignoring unobserved heterogeneity may cause an estimation bias for the covariate coefficients as well as for the duration dependence function. To check whether the duration dependence is well estimated we compare the true (simulated) and estimated duration dependence functions under the different DGPs simulated in Monte Carlo exercise **A1** (see Figures 1). We consider a negative and a positive true dependence function, equations (8) and (7), in the Figure 1 panels (a) and (b) respectively. The estimated dependence functions are computed by using the estimated intercept and coefficients (both averaged across the 100 replications considered in our Monte Carlo exercise) of the cubic polynomial used to approximate the duration dependence function. We draw three estimated duration dependence functions, one for each type of unobserved heterogeneity distribution simulated (labeled as before "Discrete UH", "Gamma UH" and "Normal UH").

Ignoring the unobserved heterogeneity causes an overestimation of the negative duration dependence and a spurious negative dependence when the true one is positive. Furthermore, it seems that the duration dependence function is better estimated at low durations.

Sometimes empirical researchers are interested in the effect of covariates on survival probabilities and expected duration. For this reason we also report the true (simulated) and estimated effects of changes in the covariate X_1 on the survival function and expected duration in Table 5.[10] For each row the results are obtained under some of the DGPs simulated in Monte Carlo exercise **A1**. The simulated DGPs are sequential logit models with two possible choices for the duration dependence (positive and negative) and three types of unobserved heterogeneity distribution (labeled as before "Discrete UH", "Gamma UH" and "Normal UH"). The true and estimated survival functions (expected duration) are computed fixing the variable X_2 at its mean, zero, and the variable X_1 at three values, its mean (zero) and its mean plus or minus half its standard deviation. The expected duration is given by $E(TI_{T\leq 40} \mid X_1, X_2 = 0)$ where I denotes the indicator function taking value 1 for durations shorter or equal to 40 and 0 otherwise; the survival function is given by

[10]Since both the estimated survival functions and the expected durations are based on hazard functions that ignore the unobserved heterogeneity, we consider their true (simulated) counterparts after integrating out the random effect through simulation.

$\Pr(T > t \mid X_1, X_2 = 0)$ and is computed at $t = 5$ and 10.

Both the survival function and the effects of changes in X_1 are well estimated at durations of 5 and 10, while their estimation slightly deteriorates with longer durations.

The true and estimated expected duration values are very close; as a consequence, the effects of changes in X_1 on the expected duration is well estimated. These results are quite encouraging if compared with similar effects computed by Mroz and Zayats (2008) (see their Table 2) when using non-parametric maximum likelihood estimation to take account of unobserved heterogeneity as in Baker and Melino (2000). Our results of the estimated effects of X_1 on the expected duration are much better compared with the ones that are computed using non-parametric maximum likelihood estimation and adopt the Hannan-Quinn Information Criterion suggested by Baker and Melino (2000).

Furthermore, the differences between the true and estimated survival function (expected duration) as well as the true and estimated effects of changes in X_1 on the survival function (expected duration) do not change across DGPs considered in the full Monte Carlo exercise **A**.[11]

When we simulate a hazard model with two time-invariant and two time-varying variables (see equation [6] and simulation exercise **B**) and estimate it ignoring the unobserved heterogeneity, the covariate coefficients seem again to be significantly underestimated. Moreover, the underestimation of the coefficients tends to be slightly larger for the pair of time-invariant variables than for the pair of time-varying ones.[12] In other words, it seems that the rescaling factor is slightly dissimilar for different types of variables (time-varying and invariant variables). Indeed, the ratios between coefficients seem to be correctly estimated when considering two variables of the same type and to be slightly biased when considering the ratio between two different types of variables. 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 causes an attenuation of the covariate coefficients due to a rescaling factor that differs slightly and not significantly by

[11]The results of the entire Monte Carlo exercise **A** are available from the authors upon request.

[12]The detailed results of exercise **B** are not reported but are available from the authors upon request.

typology of variable.

More generally the rescaling factor of a covariate could depend on its association with the duration and with other variables.[13] Nevertheless, even if each type of covariate had a different rescaling factor, we could still infer the statistical significance of each variable and compare significance across variables. This is because tests of significance, such as the Wald chi-square, are based on the ratio between the estimated covariate coefficient and its standard error, so that the rescaling factor cancels out.

When the estimation models ignore both the unobserved heterogeneity and the duration dependence, the underestimation of covariate coefficients is reduced and the rescaling factor becomes more similar for variables of different types. The ratios between coefficients are not biased, especially when considering a sample size of 5000 observations.

In conclusion, the two main findings of this section are that ignoring the unobserved heterogeneity in sequential logit models causes a rescaling of the covariates coefficients and an underestimation of the duration dependence, but the effect of covariates on the survival function and expected duration does not seem to be badly estimated. Since coefficients in binary models are only identified up to a scale normalization, applied researchers should not be concerned about the rescaling problem.

3 Misspecifying the unobserved heterogeneity distribution

3.1 Consequences

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 parameter estimates for a model with Weibull baseline hazard are very sensitive to changes in the distribution assumed for the unobserved heterogeneity. Similar

[13]Note that the use of variables that change almost monotonically with the duration or of covariates which are collinear can cause identification problems.

results are found also by Trussell and Richards (1985), Hougaard et al. (1994), Baker and Melino (2000), Zhang (2003) and Gaure et al. (2007). 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, 1984) findings.

In this paper we consider the heterogeneity misspecification problem in single-spell discrete duration models specified as sequential binary models.

Since sequential binary models with a normal distribution for the unobserved heterogeneity term can be easily estimated using conventional softwares, we wonder if practitioners should worry about the possible consequences of an incorrect normality assumption. For this reason, we carry out a Monte Carlo exercise where we evaluate the effect of imposing a normal distribution for the unobserved heterogeneity component when its true distribution is a gamma or a discrete distribution with two support points.

In addition, we consider the potential consequences of misspecifying the distribution of the residual error as well as of the unobserved heterogeneity in sequential binary models. This can be useful to guide empirical researchers in the choice of the sequential binary models (probit, logit or complementary log-log) to estimate discrete hazard models.

Before presenting the Monte Carlo exercise, we emphasize that identification of unobserved heterogeneity and duration dependence in duration models with single spells can be problematic.

In case of continuous duration and single spells, Elbers and Ridder (1982) prove that it is possible to non-parametrically identify mixed proportional hazard models with covariates. The identification is possible because this model is multiplicative in the duration and in the covariates, whereas the observed hazard function (i.e. the hazard function integrated over the unobserved heterogeneity component) is not. This implies that the interactions between duration and covariates in the observed hazard allow identifying the unobserved heterogeneity and the duration dependence in the mixed proportional hazard models (see van den Berg, 2001). On the contrary, a mixed hazard model that allows for interaction between covariates and duration would not be identified, except when using time varying covariates (see Brinch, 2007).

Similarly, single-spell discrete duration models with unobserved heterogeneity cannot be identified if we allow covariate coefficients to change with duration and consider only time invariant covariates. This result is emphasized by Mroz and Zayats (2008) and Mroz (2008).[14]

The duration models with normal random effect considered in the next section are theoretically identified because they include time varying covariates without interaction with duration.

3.2 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 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 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 these sequential binary models with random effects by using Stata, which approximates the integral of the likelihood function with respect to the random effects by using an adaptive Gauss-Hermite quadrature (see StataCorp, 2005).[15]

The simulation exercise is carried out as the previous ones were 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 being censored.

[14]Mroz (2008) considers this identification problem for sequential binary models with unobserved heterogeneity used to estimate count models with only one observed count for each individual. These models are analytically equivalent to the single-spell discrete duration models with unobserved heterogeneity that are considered in this paper.

[15]An alternative estimation method is the simulated maximum likelihood. See Gourieroux and Monfort (1996) and Train (2003).

3.3 Results

This section provides the results of the effects of misspecifying the unobserved heterogeneity or the residual error distribution.

Tables 2, 3 and 4 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.[16] Each Table considers a different estimation model (sequential logit, probit or complementary log-log) and is divided into three panels reporting results produced using three different sample sizes: 500, 1000 and 5000 observations. 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 with two mass points, gamma and normal that are labeled "Discrete UH", "Gamma UH" and "Normal UH").

Looking at the results in Table 2, 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 is erroneously assumed to follow a normal distribution instead of a gamma or a discrete distribution. 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 3, where the estimation model is given by a sequential probit model while the true DGPs are sequential logit models, the two covariate coefficients are underestimated but their ratio is still unbiased. Again we do not find relevant differences when considering DGPs with different distributions for the random effects.

Finally, in Table 4 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 2 and the bias is reduced again to a rescaling. The results are not affected by the distribution

[16]We report averages and standard deviations only for cases where convergence was achieved.

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 is reduced 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 or complementary log-log models, and it is more significantly underestimated when considering a sequential probit. Note that a higher underestimation of the ρ coefficient seems to be associated with a higher attenuation bias for the coefficients. This 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 conclude that this attenuation (amplification) bias is a consequence of the covariate coefficient normalization.

To evaluate the effect on the duration dependence estimation of misspecifying the unobserved heterogeneity distribution, we plot true and estimated baseline hazard functions (see Figure 2). The estimated baseline hazards are predicted fixing the covariates and random effects at their mean values and using the estimated coefficients (average across 100 replications) of a sequential logit (panel a), probit (panel b), and complementary log-log (panel c) model with normal random effects. The true baseline hazard is computed using the simulation model, which is a sequential logit model with covariates and random effects fixed at their mean values and coefficients fixed at their simulated values (see simulation exercise **A3** for details). For each panel there are three estimated baseline hazards corresponding to the three different types of simulated unobserved heterogeneity (labeled "Discrete UH", "Gamma UH" and "Normal UH"). Since the true baseline hazard is computed fixing the random effect at its mean, the true baseline does not depend on the simulated distribution for the unobserved heterogeneity term.

In panel (a) Figure 2, where both estimated and simulated models are given by a sequential logit, the true baseline hazard has a profile similar to the three estimated baseline hazards.

When we change the estimation model to a sequential probit as in Figure 2 panel (b), the estimated baseline hazards have a slightly different profile with respect to the true baseline hazard but they are similar for short durations.

Finally, when using a sequential complementary log-log model to estimate the duration model (see Figure 2, panel c), we find that the profile of the estimated baseline hazards follows the true one.

In Table 6, we study the difference between the true and estimated survival function (expected duration), as well as the true and estimated effects of changes in the covariate X_1 on the survival function (expected duration). The expected duration is given by $E(TI_{T\leq 40} \mid X_1, X_2 = 0, \theta = 1.8)$, while the survival function is given by $\Pr(T > t \mid X_1, X_2 = 0, \theta = 1.8)$ and is computed at $t = 5$ and 10.

Both the survival function and the expected duration value are computed fixing the variable X_2 and the unobserved heterogeneity at their means (zero and 1.8 respectively) and the variable X_1 at three different values, its mean and its mean plus or minus half its standard deviation. We consider again the DGPs in simulation exercise **A3**. The estimation models are either sequential logit models, panel (a); or sequential probit models, panel (b); or sequential complementary log-log models, panel (c); simulated unobserved heterogeneity distributions (labeled "Discrete UH", "Gamma UH" and "Normal UH").

The survival function at low durations seems to be well estimated. Furthermore, differences between true and estimated effects of changes in the variable X_1 on the survival function are never higher than 7%, at durations of 5, and of 10% at durations of 10.

The true and estimated expected duration values are close and the effects of changes in covariates on the expected duration have a correct sign and are quite well estimated. These covariate effects seem to be better estimated than when computed using non-parametric maximum likelihood estimation and adopting the Hannan-Quinn Information Criterion suggested by Baker and Melino (2000) (see Mroz and Zayats, 2008, Table 2).

In summary, the 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. Furthermore, the estimated effects of changes in X_1 on the survival function and expected duration have a correct sign and are relatively close to the true ones.

4 Conclusions

In this paper we assess the consequences of estimating single-spell discrete duration models by adopting two types of models that can be easily estimated using standard software: sequential binary models with or without individual normal random effects.

When using a sequential binary model neglecting the unobserved heterogeneity we find that the duration dependence is underestimated, but the covariate coefficients are consistently estimated up to a scale factor. Applied researchers should not be concerned about the rescaling factor because all binary models have coefficients that are identified only up to a scale. We find that the rescaling factor could change slightly across types of variables, but even in this case we can still make a correct inference on the statistical significance of each covariate and compare the significance across variables. Neglecting the unobserved heterogeneity does not cause any relevant bias in estimating the survival and expected duration functions and in evaluating the effect of changes in the covariates on these two functions.

An incorrect normality assumption for the unobserved heterogeneity distribution biases neither the duration dependence nor the covariate coefficients estimation. On the other hand, misspecifying the error distribution, assuming a normal or an extreme value distribution instead than a logistic one, seems to cause a slight bias in the duration dependence estimation but only a proportional rescaling of the covariate coefficients. Again, there are no major biases in estimating the survival (at least at low durations) and expected duration functions and in predicting the effect of covariate changes on these two functions.

These findings are very encouraging for the practitioner who would like to adopt sequential binary models (with or without normal random effects) because they are easy to estimate using conventional statistical software. These models allow empirical researchers to correctly answer the main research questions addressed in survival analysis, but not the ones on the duration dependence.

The strategy used in this paper to study the consequences of omitting or misspecifying the unobserved heterogeneity can be easily extended to more general data generator processes. We leave these possible extensions for future research.

References

Abbring J.H., van den Berg G.J. (2007) The unobserved heterogeneity distribution in duration analysis, *Biometrika*, 94, 1, 87–99.

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.

Arulampalam 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.

Brinch C.N. (2007) Nonparametric identification of the mixed hazards model with time-varying covariates, *Econometric Theory*, 23, 349–354.

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

Elbers C., Ridder G. (1982) True and spurious duration dependence: The identifiability of the proportional hazard model, *Review of Economic Studies*, 49, 3, 403–09.

Gaure S., Rø ed K., Zhang T. (2007) Time and causality: A monte carlo assessment of the timing-of-events approach, *Journal of Econometrics*, 141, 2, 1159–1195.

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

Huh K. , Sickles R.C. (1994) Estimation of the duration model by nonparametric maximum likelihood, maximum penalized likelihood, and probability simulators, *The Review of Economics and Statistics*, 76, 4, 683–94.

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.

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. (2008) A simple, flexible estimator for count and other ordered discrete data, Working paper, John E. Walker Department of Economics, Clemson University.

Mroz T.A., Zayats Y.V. (2008) Arbitrarily normalized coefficients, information sets, and false reports of "biases" in binary outcome models, *The Review of Economics and Statistics*, 90, 3, 406–413.

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***												
Negative 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
Positive 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***												
Negative 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
Positive 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***												
Negative 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
Positive 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 coefficient estimates over 100 samples. Estimation model: sequential logit. 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	$\frac{1}{\frac{\pi^2}{3}+1} = 0.233$		
DGP						
			500 Observations:			
Negative duration dependence:						
Discrete UH	0.927	0.455	2.130	0.174	6.908	98
	(0.170)	(0.093)	(0.611)	(0.143)	(2.981)	
Gamma UH	0.913	0.452	2.110	0.118	5.404	99
	(0.183)	(0.087)	(0.672)	(0.072)	(2.263)	
Normal UH	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 UH	0.937	0.462	2.069	0.156	6.424	99
	(0.137)	(0.071)	(0.405)	(0.093)	(2.607)	
Gamma UH	0.942	0.474	2.012	0.151	5.889	99
	(0.123)	(0.060)	(0.317)	(0.090)	(2.788)	
Normal UH	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 UH	0.956	0.477	2.013	0.202	6.850	100
	(0.079)	(0.040)	(0.178)	(0.099)	(2.851)	
Gamma UH	0.916	0.458	2.003	0.110	4.190	100
	(0.064)	(0.027)	(0.163)	(0.059)	(1.522)	
Normal UH	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 3: 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 DGP	1	0.5	2	$\frac{1}{1+1} = 0.5$		
500 Observations:						
Negative duration dependence:						
Discrete UH	0.519	0.255	2.128	0.306	8.690	100
	(0.099)	(0.058)	(0.587)	(0.140)	(1.495)	
Gamma UH	0.543	0.270	2.069	0.307	8.535	99
	(0.102)	(0.049)	(0.522)	(0.126)	(1.358)	
Normal UH	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 UH	0.525	0.257	2.078	0.292	8.848	99
	(0.082)	(0.039)	(0.407)	(0.104)	(1.480)	
Gamma UH	0.534	0.270	2.013	0.305	8.410	100
	(0.087)	(0.039)	(0.396)	(0.112)	(1.326)	
Normal UH	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 UH	0.524	0.261	2.015	0.311	9.010	100
	(0.036)	(0.017)	(0.174)	(0.044)	(1.259)	
Gamma UH	0.542	0.271	2.004	0.305	8.740	100
	(0.037)	(0.015)	(0.159)	(0.042)	(0.960)	
Normal UH	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 4: 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	$\rho = \frac{\sigma_\theta^2}{\sigma_\epsilon^2+\sigma_\theta^2}$	Iterations	Convergence
True Value	1	0.5	2	$\frac{1}{\frac{\pi^2}{6}+1} = 0.378$		
DGP						
			500 Observations:			
Negative duration dependence:						
Discrete UH	0.855	0.421	2.123	0.243	6.714	98
	(0.147)	(0.086)	(0.606)	(0.127)	(2.428)	
Gamma UH	0.861	0.442	2.010	0.222	5.890	100
	(0.156)	(0.072)	(0.545)	(0.112)	(2.238)	
Normal UH	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 UH	0.871	0.429	2.069	0.233	6.602	98
	(0.123)	(0.062)	(0.410)	(0.094)	(2.560)	
Gamma UH	0.886	0.432	2.084	0.228	5.730	100
	(0.107)	(0.051)	(0.381)	(0.084)	(2.348)	
Normal UH	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 UH	0.882	0.440	2.013	0.285	6.710	100
	(0.058)	(0.029)	(0.178)	(0.089)	(2.388)	
Gamma UH	0.878	0.438	2.011	0.224	5.310	100
	(0.054)	(0.028)	(0.160)	(0.075)	(2.246)	
Normal UH	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.

Table 5: True and estimated expected duration and survival function. True and estimated effects of X_1 on expected duration and survival function.

DGP	$Pr(T>t\|X_1,X_2=0)$												Impact of X_1 on $Pr(T>t\|X_1,X_2=0)$							
	at $X_1=-0.5\sigma$				at $X_1=0$				at $X_1=0.5\sigma$				From $X_1=0$ to $X_1=0.5\sigma$				From $X_1=-0.5\sigma$ to $X_1=0$			
	t=5		t=10		t=5		t=10		t=5		t=10		t=5		t=10		t=5		t=10	
	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR
Negative Dur.																				
Discrete	0.47	0.48	0.34	0.35	0.52	0.52	0.39	0.39	0.57	0.56	0.45	0.43	0.05	0.04	0.06	0.04	0.05	0.04	0.06	0.04
Gamma	0.46	0.47	0.32	0.33	0.51	0.51	0.37	0.37	0.57	0.56	0.43	0.42	0.05	0.05	0.06	0.05	0.05	0.05	0.06	0.04
Normal	0.47	0.51	0.33	0.38	0.53	0.55	0.39	0.43	0.58	0.60	0.45	0.47	0.05	0.04	0.06	0.04	0.05	0.04	0.06	0.04
Positive Dur.																				
Discrete	0.34	0.34	0.14	0.15	0.39	0.38	0.18	0.18	0.45	0.42	0.23	0.22	0.05	0.04	0.05	0.03	0.05	0.04	0.04	0.03
Gamma	0.32	0.33	0.12	0.13	0.37	0.37	0.16	0.16	0.42	0.42	0.21	0.19	0.05	0.04	0.05	0.03	0.05	0.04	0.04	0.03
Normal	0.33	0.37	0.13	0.16	0.38	0.41	0.17	0.19	0.44	0.45	0.21	0.23	0.05	0.04	0.05	0.03	0.05	0.04	0.04	0.03

DGP	$E(TI_{T\leq 40}\|X_1,X_2=0)$						Effect of X_1 on $E(TI_{T\leq 40}\|X_1,X_2=0)$			
	at $X_1=-0.5\sigma$		at $X_1=0$		at $X_1=0.5\sigma$		From $X_1=0$ to $X_1=0.5\sigma$		From $X_1=-0.5\sigma$ to $X_1=0$	
	ES	TR	ES	TR	ES	TR	ES	TR	ES	TR
Negative Dur.										
Discrete	6.71	6.61	6.89	6.70	6.92	6.78	0.03	0.08	0.18	0.09
Gamma	6.73	6.56	6.99	6.96	7.10	7.33	0.11	0.37	0.26	0.40
Normal	7.05	7.13	7.31	7.37	7.41	7.53	0.10	0.16	0.26	0.23
Positive Dur.										
Discrete	5.41	5.41	6.22	6.06	7.12	6.74	0.90	0.69	0.81	0.65
Gamma	5.04	5.35	5.65	5.87	6.23	6.41	0.59	0.55	0.60	0.52
Normal	5.24	5.81	5.93	6.36	6.65	6.93	0.72	0.57	0.70	0.55

Note: ES= estimated, TR= true. $I_{T\leq 40}$ denotes the indicator function. σ is the standard deviation of X_1 whose mean is zero. Sample size: 5000. Estimated cubic polynomial duration dependence.

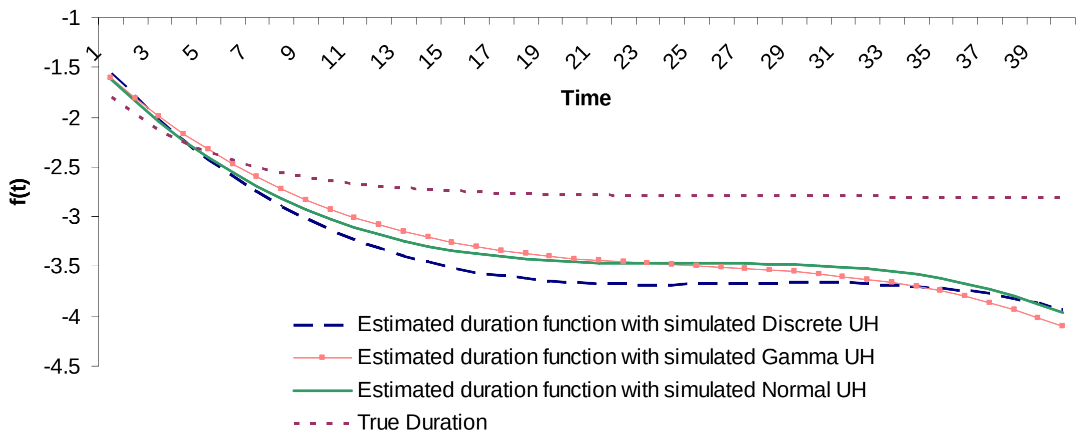
Table 6: True and estimated expected duration and survival function. True and estimated effects of X_1 on expected duration and survival function.

	$Pr(T > t \mid X_1, X_2 = 0, \theta = 1.8)$						Effect of X_1 on $Pr(T > t \mid X_1, X_2 = 0, \theta = 1.8)$			
	at $X_1 = -0.5\sigma$		at $X_1 = 0$		at $X_1 = 0.5\sigma$		From $X_1 = 0$ to $X_1 = 0.5\sigma$		From $X_1 = -0.5\sigma$ to $X_1 = 0$	
	t=5	t=10	t=5	t=10	t=5	t=10	t=5	t=10	t=5	t=10
True	0.50	0.32	0.56	0.38	0.61	0.44	0.05	0.06	0.06	0.06
(a) Sequential logit										
Discrete UH	0.49	0.32	0.54	0.38	0.59	0.44	0.05	0.06	0.05	0.06
Gamma UH	0.47	0.31	0.52	0.36	0.57	0.42	0.05	0.06	0.05	0.06
Normal UH	0.49	0.32	0.55	0.38	0.60	0.44	0.05	0.06	0.05	0.06
(b) Sequential probit										
Discrete UH	0.48	0.28	0.53	0.34	0.59	0.41	0.05	0.06	0.06	0.06
Gamma UH	0.47	0.25	0.53	0.31	0.58	0.38	0.06	0.06	0.06	0.06
Normal UH	0.49	0.28	0.54	0.34	0.60	0.40	0.06	0.07	0.06	0.06
(c) Sequential complementary log-log										
Discrete UH	0.50	0.33	0.55	0.39	0.60	0.45	0.05	0.06	0.05	0.06
Gamma UH	0.50	0.33	0.56	0.39	0.61	0.45	0.05	0.06	0.05	0.06
Normal UH	0.50	0.33	0.56	0.39	0.61	0.44	0.05	0.06	0.05	0.06

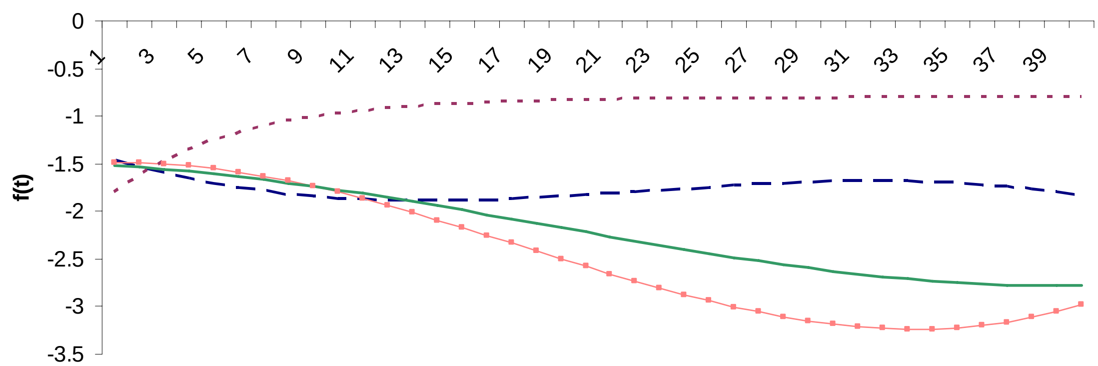
	$E(TI_{T \leq 40} \mid X_1, X_2 = 0, \theta = 1.8)$			Effect of X_1 on $E(TI_{T \leq 40} \mid X_1, X_2 = 0, \theta = 1.8)$	
	at $X_1 = -0.5\sigma$	at $X_1 = 0$	at $X_1 = 0.5\sigma$	From $X_1 = 0$ to $X_1 = 0.5\sigma$	From $X_1 = -0.5\sigma$ to $X_1 = 0$
True	8.35	9.15	9.76	0.61	0.80
(a) Sequential logit					
Discrete UH	8.04	8.61	8.99	0.38	0.57
Gamma UH	7.54	8.04	8.36	0.33	0.49
Normal UH	8.21	8.89	9.37	0.48	0.67
(b) Sequential probit					
Discrete UH	8.09	9.12	10.04	0.92	1.03
Gamma UH	7.56	8.64	9.71	1.06	1.08
Normal UH	8.00	9.13	10.22	1.09	1.13
(c) Sequential complementary log-log					
Discrete UH	8.03	8.55	8.88	0.33	0.51
Gamma UH	7.97	8.46	8.77	0.31	0.49
Normal UH	8.29	8.94	9.40	0.46	0.65

Note: UH= Unobserved Heterogeneity. $I_{T \leq 40}$ denotes the indicator function. σ is the standard deviation of X_1 whose mean is zero. Sample size: 5000. (a), (b), (c) define three different estimated models and data simulated with Discrete UH, Gamma UH and Normal UH.

Figure 1: Estimated and true negative/positive duration dependence functions. Monte Carlo exercise A1. Unobserved heterogeneity ignored. Sample size: 1000.

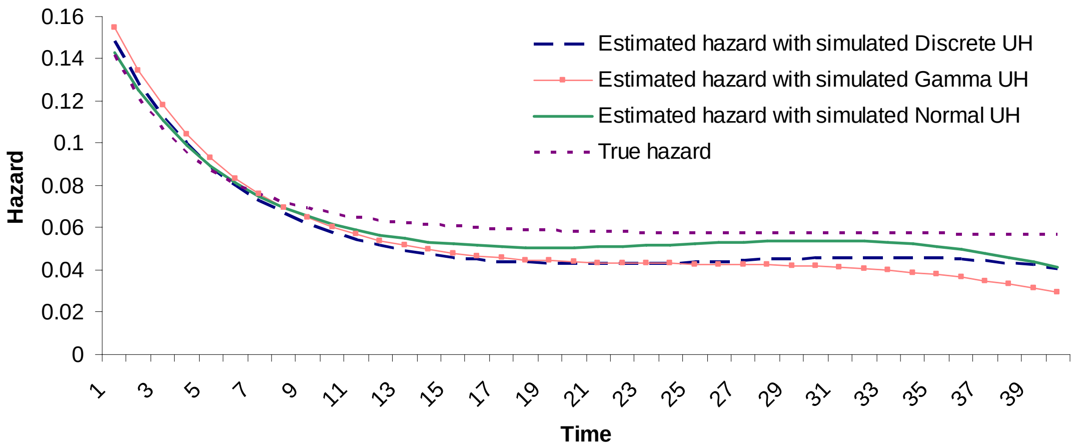


(a) negative duration dependence

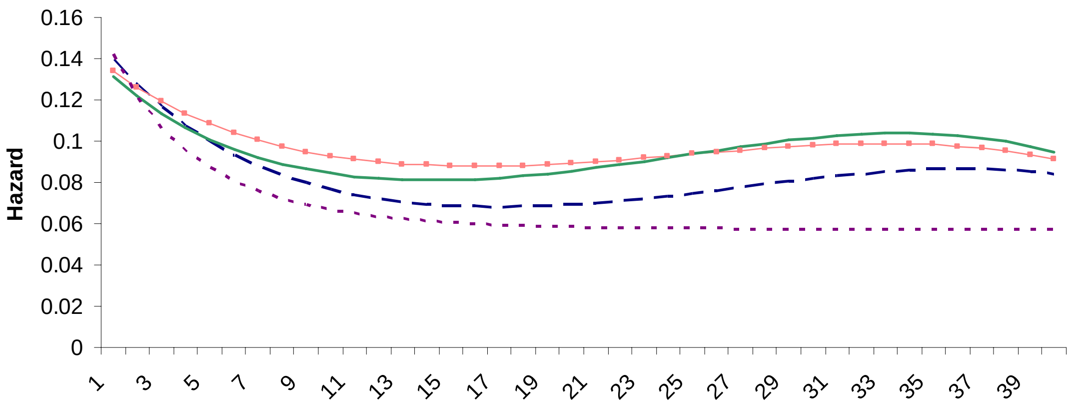


(b) positive duration dependence

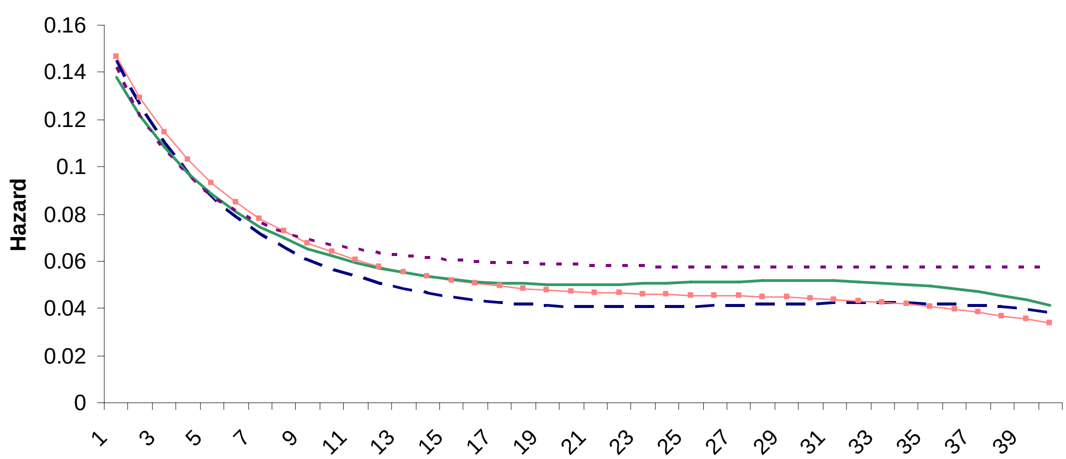
Figure 2: Estimated and true baseline hazards. DGP: sequential logit with unobserved heterogeneity. Estimation model: sequential logit/probit/cloglog with normal random effect. Sample size: 5000.



(a) Sequential logit



(b) Sequential probit



(c) Sequential complementary log-log

RECENTLY PUBLISHED "TEMI" (*)

N. 680 – *Family succession and firm performance: Evidence from Italian family firms*, by Marco Cucculelli and Giacinto Micucci (June 2008).

N. 681 – *Short-term interest rate futures as monetary policy forecasts*, by Giuseppe Ferrero and Andrea Nobili (June 2008).

N. 682 – *Vertical specialisation in Europe: Evidence from the import content of exports*, by Emanuele Breda, Rita Cappariello and Roberta Zizza (August 2008).

N. 683 – *A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with imperfect compliance*, by Andrea Mercatanti (August 2008).

N. 684 – *Balancing work and family in Italy: New mothers' employment decisions after childbirth,* by Piero Casadio, Martina Lo Conte and Andrea Neri (August 2008).

N. 685 – *Temporal aggregation of univariate and multivariate time series models: A survey*, by Andrea Silvestrini and David Veredas (August 2008).

N. 686 – *Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG*, by Luca Arciero, Claudia Biancotti, Leandro D'Aurizio and Claudio Impenna (August 2008).

N. 687 – *The labor market impact of immigration in Western Germany in the 1990's*, by Francesco D'Amuri, Gianmarco I. P. Ottaviano and Giovanni Peri (August 2008).

N. 688 – *Agglomeration and growth: the effects of commuting costs*, by Antonio Accetturo (September 2008).

N. 689 – *A beta based framework for (lower) bond risk premia*, by Stefano Nobili and Gerardo Palazzo (September 2008).

N. 690 – *Nonlinearities in the dynamics of the euro area demand for M1*, by Alessandro Calza and Andrea Zaghini (September 2008).

N. 691 – *Educational choices and the selection process before and after compulsory schooling*, by Sauro Mocetti (September 2008).

N. 692 – *Investors' risk attitude and risky behavior: a Bayesian approach with imperfect information*, by Stefano Iezzi (September 2008).

N. 693 – *Competing influence*, by Enrico Sette (September 2008).

N. 694 – *La relazione tra gettito tributario e quadro macroeconomico in Italia*, by Alberto Locarno and Alessandra Staderini (December 2008).

N. 695 – *Immigrant earnings in the Italian labour market*, by Antonio Accetturo and Luigi Infante (December 2008).

N. 696 – *Reservation wages: explaining some puzzling regional patterns*, by Paolo Sestito and Eliana Viviano (December 2008).

N. 697 – *Technological change and the demand for currency: an analysis with household data*, by Francesco Lippi and Alessandro Secchi (December 2008).

N. 698 – *Immigration and crime: an empirical analysis*, by Milo Bianchi, Paolo Buonanno and Paolo Pinotti (December 2008).

N. 699 – *Bond risk premia, macroeconomic fundamentals and the exchange rate*, by Marcello Pericoli and Marco Taboga (January 2009).

N. 700 – *What determines debt intolerance? The role of political and monetary institutions*, by Raffaela Giordano and Pietro Tommasino (January 2009).

N. 701 – *On analysing the world distribution of income*, by Anthony B. Atkinson and Andrea Brandolini (January 2009).

N. 702 – *Dropping the books and working off the books*, by Rita Cappariello and Roberta Zizza (January 2009).

N. 703 – *Measuring wealth mobility*, by Andrea Neri (January 2009).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2006

F. BUSETTI, *Tests of seasonal integration and cointegration in multivariate unobserved component models,* Journal of Applied Econometrics, Vol. 21, 4, pp. 419-438, **TD No. 476 (June 2003).**

C. BIANCOTTI, *A polarization of inequality? The distribution of national Gini coefficients 1970-1996*, Journal of Economic Inequality, Vol. 4, 1, pp. 1-32, **TD No. 487 (March 2004).**

L. CANNARI and S. CHIRI, *La bilancia dei pagamenti di parte corrente Nord-Sud (1998-2000),* in L. Cannari, F. Panetta (a cura di), Il sistema finanziario e il Mezzogiorno: squilibri strutturali e divari finanziari, Bari, Cacucci, **TD No. 490 (March 2004).**

M. BOFONDI and G. GOBBI, *Information barriers to entry into credit markets,* Review of Finance, Vol. 10, 1, pp. 39-67, **TD No. 509 (July 2004).**

W. FUCHS and LIPPI F., *Monetary union with voluntary participation,* Review of Economic Studies, Vol. 73, pp. 437-457 **TD No. 512 (July 2004).**

E. GAIOTTI and A. SECCHI, *Is there a cost channel of monetary transmission? An investigation into the pricing behaviour of 2000 firms*, Journal of Money, Credit and Banking, Vol. 38, 8, pp. 2013-2038 **TD No. 525 (December 2004).**

A. BRANDOLINI, P. CIPOLLONE and E. VIVIANO, *Does the ILO definition capture all unemployment?,* Journal of the European Economic Association, Vol. 4, 1, pp. 153-179, **TD No. 529 (December 2004).**

A. BRANDOLINI, L. CANNARI, G. D'ALESSIO and I. FAIELLA, *Household wealth distribution in Italy in the 1990s,* in E. N. Wolff (ed.) International Perspectives on Household Wealth, Cheltenham, Edward Elgar, **TD No. 530 (December 2004).**

P. DEL GIOVANE and R. SABBATINI, *Perceived and measured inflation after the launch of the Euro: Explaining the gap in Italy,* Giornale degli economisti e annali di economia, Vol. 65, 2 , pp. 155-192, **TD No. 532 (December 2004).**

M. CARUSO, *Monetary policy impulses, local output and the transmission mechanism*, Giornale degli economisti e annali di economia, Vol. 65, 1, pp. 1-30, **TD No. 537 (December 2004).**

L. GUISO and M. PAIELLA, *The role of risk aversion in predicting individual behavior, In P. A. Chiappori e C. Gollier (eds.) Competitive Failures in Insurance Markets: Theory and Policy Implications, Monaco, CESifo*, **TD No. 546 (February 2005).**

G. M. TOMAT, *Prices product differentiation and quality measurement: A comparison between hedonic and matched model methods*, Research in Economics, Vol. 60, 1, pp. 54-68, **TD No. 547 (February 2005).**

L. GUISO, M. PAIELLA and I. VISCO, *Do capital gains affect consumption? Estimates of wealth effects from Italian household's behavior*, in L. Klein (ed), Long Run Growth and Short Run Stabilization: Essays in Memory of Albert Ando (1929-2002), Cheltenham, Elgar, **TD No. 555 (June 2005).**

F. BUSETTI, S. FABIANI and A. HARVEY, *Convergence of prices and rates of inflation*, Oxford Bulletin of Economics and Statistics, Vol. 68, 1, pp. 863-878, **TD No. 575 (February 2006).**

M. CARUSO, *Stock market fluctuations and money demand in Italy, 1913 - 2003,* Economic Notes, Vol. 35, 1, pp. 1-47, **TD No. 576 (February 2006).**

R. BRONZINI and G. DE BLASIO, *Evaluating the impact of investment incentives: The case of Italy's Law 488/92.* Journal of Urban Economics, Vol. 60, 2, pp. 327-349, **TD No. 582 (March 2006).**

R. BRONZINI and G. DE BLASIO, *Una valutazione degli incentivi pubblici agli investimenti,* Rivista Italiana degli Economisti , Vol. 11, 3, pp. 331-362, **TD No. 582 (March 2006).**

A. DI CESARE, *Do market-based indicators anticipate rating agencies? Evidence for international banks,* Economic Notes, Vol. 35, pp. 121-150, **TD No. 593 (May 2006).**

R. GOLINELLI and S. MOMIGLIANO, *Real-time determinants of fiscal policies in the euro area*, Journal of Policy Modeling, Vol. 28, 9, pp. 943-964, **TD No. 609 (December 2006).**

2007

S. SIVIERO and D. TERLIZZESE, *Macroeconomic forecasting: Debunking a few old wives' tales*, Journal of Business Cycle Measurement and Analysis , v. 3, 3, pp. 287-316, **TD No. 395 (February 2001).**

S. MAGRI, *Italian households' debt: The participation to the debt market and the size of the loan*, Empirical Economics, v. 33, 3, pp. 401-426, **TD No. 454 (October 2002).**

L. CASOLARO. and G. GOBBI, *Information technology and productivity changes in the banking industry*, Economic Notes, Vol. 36, 1, pp. 43-76, **TD No. 489 (March 2004).**

G. FERRERO, *Monetary policy, learning and the speed of convergence*, Journal of Economic Dynamics and Control, v. 31, 9, pp. 3006-3041, **TD No. 499 (June 2004).**

M. PAIELLA, *Does wealth affect consumption? Evidence for Italy*, Journal of Macroeconomics, Vol. 29, 1, pp. 189-205, **TD No. 510 (July 2004).**

F. LIPPI. and S. NERI, *Information variables for monetary policy in a small structural model of the euro area*, Journal of Monetary Economics, Vol. 54, 4, pp. 1256-1270, **TD No. 511 (July 2004).**

A. ANZUINI and A. LEVY, *Monetary policy shocks in the new EU members: A VAR approach*, Applied Economics, Vol. 39, 9, pp. 1147-1161, **TD No. 514 (July 2004).**

D. JR. MARCHETTI and F. Nucci, *Pricing behavior and the response of hours to productivity shocks*, Journal of Money Credit and Banking, v. 39, 7, pp. 1587-1611, **TD No. 524 (December 2004).**

R. BRONZINI, *FDI Inflows, agglomeration and host country firms' size: Evidence from Italy*, Regional Studies, Vol. 41, 7, pp. 963-978, **TD No. 526 (December 2004).**

L. MONTEFORTE, *Aggregation bias in macro models: Does it matter for the euro area?*, Economic Modelling, 24, pp. 236-261, **TD No. 534 (December 2004).**

A. NOBILI, *Assessing the predictive power of financial spreads in the euro area: does parameters instability matter?*, Empirical Economics, Vol. 31, 1, pp. 177-195, **TD No. 544 (February 2005).**

A. DALMAZZO and G. DE BLASIO, *Production and consumption externalities of human capital: An empirical study for Italy*, Journal of Population Economics, Vol. 20, 2, pp. 359-382, **TD No. 554 (June 2005).**

M. BUGAMELLI and R. TEDESCHI, *Le strategie di prezzo delle imprese esportatrici italiane*, Politica Economica, v. 23, 3, pp. 321-350, **TD No. 563 (November 2005).**

L. GAMBACORTA and S. IANNOTTI, *Are there asymmetries in the response of bank interest rates to monetary shocks?*, Applied Economics, v. 39, 19, pp. 2503-2517, **TD No. 566 (November 2005).**

P. ANGELINI and F. LIPPI, *Did prices really soar after the euro cash changeover? Evidence from ATM withdrawals*, International Journal of Central Banking, Vol. 3, 4, pp. 1-22, **TD No. 581 (March 2006).**

A. LOCARNO, *Imperfect knowledge, adaptive learning and the bias against activist monetary policies*, International Journal of Central Banking, v. 3, 3, pp. 47-85, **TD No. 590 (May 2006).**

F. LOTTI and J. MARCUCCI, *Revisiting the empirical evidence on firms' money demand*, Journal of Economics and Business, Vol. 59, 1, pp. 51-73, **TD No. 595 (May 2006).**

P. CIPOLLONE and A. ROSOLIA, *Social interactions in high school: Lessons from an earthquake*, American Economic Review, Vol. 97, 3, pp. 948-965, **TD No. 596 (September 2006).**

L. DEDOLA and S. NERI, *What does a technology shock do? A VAR analysis with model-based sign restrictions*, Journal of Monetary Economics, Vol. 54, 2, pp. 512-549, **TD No. 607 (December 2006).**

F. VERGARA CAFFARELLI, *Merge and compete: strategic incentives for vertical integration*, Rivista di politica economica, v. 97, 9-10, serie 3, pp. 203-243, **TD No. 608 (December 2006).**

A. BRANDOLINI, *Measurement of income distribution in supranational entities: The case of the European Union*, in S. P. Jenkins e J. Micklewright (eds.), Inequality and Poverty Re-examined, Oxford, Oxford University Press, **TD No. 623 (April 2007).**

M. PAIELLA, *The foregone gains of incomplete portfolios*, Review of Financial Studies, Vol. 20, 5, pp. 1623-1646, **TD No. 625 (April 2007).**

K. BEHRENS, A. R. LAMORGESE, G.I.P. OTTAVIANO and T. TABUCHI, *Changes in transport and non transport costs: local vs. global impacts in a spatial network*, Regional Science and Urban Economics, Vol. 37, 6, pp. 625-648, **TD No. 628 (April 2007).**

M. BUGAMELLI, *Prezzi delle esportazioni, qualità dei prodotti e caratteristiche di impresa: analisi su un campione di imprese italiane*, v. 34, 3, pp. 71-103, Economia e Politica Industriale, **TD No. 634 (June 2007).**

G. ASCARI and T. ROPELE, *Optimal monetary policy under low trend inflation*, Journal of Monetary Economics, v. 54, 8, pp. 2568-2583, **TD No. 647 (November 2007).**

R. GIORDANO, S. MOMIGLIANO, S. NERI and R. PEROTTI, *The Effects of Fiscal Policy in Italy: Evidence from a VAR Model*, European Journal of Political Economy, Vol. 23, 3, pp. 707-733, **TD No. 656 (January 2008).**

B. ROFFIA and A. ZAGHINI, *Excess money growth and inflation dynamics*, International Finance, v. 10, 3, pp. 241-280, **TD No. 657 (January 2008).**

G. BARBIERI, P. CIPOLLONE and P. SESTITO, *Labour market for teachers: demographic characteristics and allocative mechanisms*, Giornale degli economisti e annali di economia, v. 66, 3, pp. 335-373, **TD No. 672 (June 2008).**

E. BREDA, R. CAPPARIELLO and R. ZIZZA, *Vertical specialisation in Europe: evidence from the import content of exports*, Rivista di politica economica, numero monografico,**TD No. 682 (August 2008).**

2008

P. ANGELINI, *Liquidity and announcement effects in the euro area*, Giornale degli Economisti e Annali di Economia, v. 67, 1, pp. 1-20, **TD No. 451 (October 2002).**

P. ANGELINI, P. DEL GIOVANE, S. SIVIERO and D. TERLIZZESE, *Monetary policy in a monetary union: What role for regional information?*, International Journal of Central Banking, v. 4, 3, pp. 1-28, **TD No. 457 (December 2002).**

F. SCHIVARDI and R. TORRINI, *Identifying the effects of firing restrictions through size-contingent Differences in regulation*, Labour Economics, v. 15, 3, pp. 482-511, **TD No. 504 (June 2004).**

L. GUISO and M. PAIELLA,, *Risk aversion, wealth and background risk*, Journal of the European Economic Association, v. 6, 6, pp. 1109-1150, **TD No. 483 (September 2003).**

C. BIANCOTTI, G. D'ALESSIO and A. NERI, *Measurement errors in the Bank of Italy's survey of household income and wealth*, Review of Income and Wealth, v. 54, 3, pp. 466-493, **TD No. 520 (October 2004).**

S. MOMIGLIANO, J. HENRY and P. HERNÁNDEZ DE COS, *The impact of government budget on prices: Evidence from macroeconometric models*, Journal of Policy Modelling, v. 30, 1, pp. 123-143 **TD No. 523 (October 2004).**

L. GAMBACORTA, *How do banks set interest rates?*, European Economic Review, v. 52, 5, pp. 792-819, **TD No. 542 (February 2005).**

P. ANGELINI and A. GENERALE, *On the evolution of firm size distributions*, American Economic Review, v. 98, 1, pp. 426-438, **TD No. 549 (June 2005).**

R. FELICI and M. PAGNINI, *Distance, bank heterogeneity and entry in local banking markets*, The Journal of Industrial Economics, v. 56, 3, pp. 500-534, **No. 557 (June 2005).**

S. DI ADDARIO and E. PATACCHINI, *Wages and the city. Evidence from Italy*, Labour Economics, v.15, 5, pp. 1040-1061, **TD No. 570 (January 2006).**

M. PERICOLI and M. TABOGA, *Canonical term-structure models with observable factors and the dynamics of bond risk premia*, Journal of Money, Credit and Banking, v. 40, 7, pp. 1471-88, **TD No. 580 (February 2006).**

E. VIVIANO, *Entry regulations and labour market outcomes. Evidence from the Italian retail trade sector*, Labour Economics, v. 15, 6, pp. 1200-1222, **TD No. 594 (May 2006).**

S. FEDERICO and G. A. MINERVA, *Outward FDI and local employment growth in Italy*, Review of World Economics, Journal of Money, Credit and Banking, v. 144, 2, pp. 295-324, **TD No. 613 (February 2007).**

F. BUSETTI and A. HARVEY, *Testing for trend*, Econometric Theory, v. 24, 1, pp. 72-87, **TD No. 614 (February 2007).**

V. CESTARI, P. DEL GIOVANE and C. ROSSI-ARNAUD, *Memory for prices and the Euro cash changeover: an analysis for cinema prices in Italy*, In P. Del Giovane e R. Sabbatini (eds.), The Euro Inflation and Consumers' Perceptions. Lessons from Italy, Berlin-Heidelberg, Springer, **TD No. 619 (February 2007).**

H. BRONWYN, F. LOTTI and J. MAIRESSE, *Employment, innovation and productivity: evidence from Italian manufacturing microdata*, Industrial and Corporate Change, v. 17, 4, pp. 813-839, **TD No. 622 (April 2007).**

J. SOUSA and A. ZAGHINI, *Monetary policy shocks in the Euro Area and global liquidity spillovers*, International Journal of Finance and Economics, v.13, 3, pp. 205-218, **TD No. 629 (June 2007).**

M. DEL GATTO, GIANMARCO I. P. OTTAVIANO and M. PAGNINI, *Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms,* Journal of Regional Science, v. 48, 1, pp. 97-129, **TD No. 635 (June 2007).**

P. DEL GIOVANE, S. FABIANI and R. SABBATINI, *What's behind "inflation perceptions"? A survey-based analysis of Italian consumers,* in P. Del Giovane e R. Sabbatini (eds.), *The Euro Inflation and Consumers' Perceptions. Lessons from Italy,* Berlin-Heidelberg, Springer, **TD No. 655 (January 2008).**

B. BORTOLOTTI, and P. PINOTTI, *Delayed privatization,* Public Choice, v. 136, 3-4, pp. 331-351, **TD No. 663 (April 2008).**

R. BONCI and F. COLUMBA, *Monetary policy effects: New evidence from the Italian flow of funds,* Applied Economics , v. 40, 21, pp. 2803-2818, **TD No. 678 (June 2008).**

M. CUCCULELLI, and G. MICUCCI, *Family Succession and firm performance: evidence from Italian family firms,* Journal of Corporate Finance, v. 14, 1, pp. 17-31, **TD No. 680 (June 2008).**

2009

S. MAGRI, *The financing of small innovative firms: The Italian case,* Economics of Innovation and New Technology, v. 18, 2, pp. 181-204, **TD No. 640 (September 2007).**

L. ARCIERO, C. BIANCOTTI, L. D'AURIZIO and C. IMPENNA, *Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG,* Journal of Artificial Societies and Social Simulation, v. 12, 1, **TD No. 686 (August 2008).**

A. CALZA and A. ZAGHINI, *Nonlinearities in the dynamics of the euro area demand for M1,* Macroeconomic Dynamics, v. 13, 1, pp. 1-19, **TD No. 690 (September 2008).**

FORTHCOMING

L. MONTEFORTE and S. SIVIERO, *The Economic Consequences of Euro Area Modelling Shortcuts,* Applied Economics, **TD No. 458 (December 2002).**

M. BUGAMELLI and A. ROSOLIA, *Produttività e concorrenza estera*, Rivista di politica economica, **TD No. 578 (February 2006).**

R. BRONZINI and P. PISELLI, *Determinants of long-run regional productivity with geographical spillovers: the role of R&D, human capital and public infrastructure,* Regional Science and Urban Economics, **TD No. 597 (September 2006).**

U. ALBERTAZZI and L. GAMBACORTA, *Bank profitability and the business cycle,* Journal of Financial Stability, **TD No. 601 (September 2006).**

A. CIARLONE, P. PISELLI and G. TREBESCHI, *Emerging Markets' Spreads and Global Financial Conditions*, Journal of International Financial Markets, Institutions & Money, **TD No. 637 (June 2007).**

L. FORNI, L. MONTEFORTE and L. SESSA, *The general equilibrium effects of fiscal policy: Estimates for the euro area*, Journal of Public Economics, **TD No. 652 (November 2007).**

R. GOLINELLI and S. MOMIGLIANO, *The cyclical reaction of fiscal policies in the Euro Area. A critical survey of empirical research*, Fiscal Studies, **TD No. 654 (January 2008).**

A. ACCETTURO, *Agglomeration and growth: the effects of commuting costs,* Regional Science, **TD No. 688 (September 2008).**