CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Warwick Research Archives Portal Repository

# Low-pay transitions and attrition bias in Italy: An analysis using simulation based estimation

Lorenzo Cappellari[#]
Department of Economics
University of Warwick
Coventry, CV4 7AL, UK
e-mail: l.cappellari@warwick.ac.uk

October 1999

**Abstract**

This paper analyses the extent to which existing econometric models of low-pay transition probabilities in Italy are biased by the presence of endogenous panel attrition. Non-random exits from the sample of wage earners may result from both demand and supply side factors and this could lead to under- or overestimation (respectively) of the extent of low-wage persistence. The analysis is carried out by extending the bivariate probit model used in Cappellari [1999] (where starting state and transition probabilities are jointly modelled thus tackling the endogeneity of the conditioning starting wage state) with a third equation which controls for the non-randomness of panel attrition. The resulting trivariate probit model with endogenous switching, whose evaluation is not feasible within the routines customarily available in microeconometric packages, is implemented by applying simulation estimation techniques. Results show the ignorability of attrition in SHIW data, thus pointing towards the robustness of the results previously obtained without controlling for attrition.

Keywords: Low-pay transitions, Attrition bias, Simulation-based estimation
JEL code: J31, D31, C25

[#] I wish to thank Carlo Dell'Aringa, Claudio Lucifora and Mark Stewart for helpful suggestions. I am also grateful to seminar participants at the Catholic University of Milan for their comments. Data availability form the Bank of Italy through IGIER and financial support from the ESRC (award no. R00429724408) and the Catholic University of Milan are gratefully acknowledged. The usual disclaimers apply.

**Contents**

## 1. Introduction

Existing evidence on individual transitions at the bottom of the Italian wage distribution indicates that while factors traditionally thought of as “wage determinants”, such as education and labour market experience, have, if any, a limited impact in generating mobility across pre-defined low-wage thresholds, the experience of low-pay has, *per se*, a clear role in determining the future occurrence of low-pay episodes, i.e. a *low-pay stigma* effect appears to be in place (see Cappellari [1999]). Such evidence has been produced using panel data from the Survey on Household Income and Wealth (SHIW)[1] for the period 1993-1995 and estimating a bivariate probit model with endogenous switching which allows assessment of the so called *initial conditions problem* (Heckman [1981]), i.e. the potential endogeneity of the conditioning starting wage state: workers assignment above or below the low-pay threshold at the beginning of the transition could, in general, be correlated with unobserved determinants of transition probabilities and the joint estimation of state and transition probabilities accounts for this source of endogeneity.[2] Results show how, similarly to other studies (see Stewart and Swaffield [1999]), the endogeneity of initial conditions should not be ignored, the correlation between state and transition probabilities being statistically significant.

Results above are based on a sample for which a valid wage is observed at both ends of the transition investigated, while observations available only at the beginning or at the end of the transition are discarded from the analysis. Such a sample selection rule may lead to biased parameters’ estimates if the propensity to be observed in both of the SHIW waves considered is not randomly distributed across individuals, but is correlated with unobservables in the transition equation. In other words, the presence of panel attrition is a second potential source of endogeneity inherent to the modelling of wage transition probabilities: the aim of this paper is to investigate the extent to which the existing evidence on Italy is plagued by endogenous attrition. Such a task is pursued by augmenting the bivariate probit model with a third equation which accounts for the probability of belonging to the balanced sample. The resulting set-up is a

[1] The SHIW is produced by the Bank of Italy and is based on questionnaires which, apart from information on various aspects of the households’ economic behaviour, report detailed information on labour market variables for their members. See Cappellari [1999] for a description of the sample.

[2] A bivariate probit with endogenous switching and partial observability of the arrival wage distribution has been originally proposed by Stewart and Swaffield [1999] to model low-pay transitions on BHPS data. Cappellari [1999] extends this framework along two directions: first, the partial observability hypothesis is removed and, secondly, the binary probit specification is replaced by an ordered probit which accounts for the width of transitions.

trivariate probit with double endogenous switching. The use of trivariate normal integrals, which are not commonly packaged in statistical softwares, poses a computational difficulty for the implementation of maximum likelihood estimation. Such a problem is tackled here by means of simulation techniques, in particular by implementing the GHK simulator within STATA's maximum likelihood routines.

The sign of the correlation between attrition and low-wage transition probabilities is not clear a priori. On the one hand, attrition could be determined by demand side factors, with workers abandoning the sample of wage earners as a consequence of events such as layoffs. In such a case, individuals dropping out from the sample over time are likely to be characterised by a low degree of attachment to the labour market and their characteristics (both observed and unobserved by the researcher) would probably positively influence the propensity to persist in low-pay. The exclusion of these observations from the analysis would then lead to underestimation of both aggregate low-pay persistence and the effect of observable characteristics on transition probabilities. At the other extreme, exits from the data set may be the result of supply side decisions which, had the sample unit been observed in subsequent time periods, would have generated mobility out of low-pay, which is instead not observed due to the inability to track the missing observation. In this occurrence attrition would negatively covary with low-pay persistence, so that inferences based on the "balanced" sample would overestimate both aggregate persistence and the effect of observable characteristics on transition probabilities. The actual situation will probably result from the interplay of these two effects and in the analysis which follows attention will be focused on the net result.

The case of endogenous panel attrition is an example of what Verbeek and Nijman [1992] define as a *non-ignorable sample selection rule*: conducting inference on the selected sample is legitimate only if conditioning on the availability of observations does not alter the joint density of the variables under examination, and only in this case the selection rule may be deemed ignorable. As pointed out in this study, given a set of incomplete data, there are three strategies which could be pursued. Data may first of all be imputed, i.e. missing bits of information are replaced by their prediction based on the available sample. Alternatively, available observations could be weighted in some way, in order to reconstruct their relative importance to what it should have been in a random (i.e. non attrited) sample. Finally, a model based strategy can be pursued. In this case, the treatment of the missing data process is deferred to the analysis stage, where the probability of belonging to the available data set is modelled jointly with the economic relation of interest. In the case of non-ignorability of the sample selection rule, this last strategy is superior in that both imputation and weighting would need to be model based in order to be properly carried out.

Such a modelling approach to attrition characterises the few studies which address the problem in the context of panel data on earnings. Hausman and Wise [1979] are concerned with endogenous selection in wage equations estimated on a sample of participants in the Gary Income Maintenance Experiment, in particular with attrition of subsequent responses once the observational unit has already been part of the sample. They apply Heckman' s correction techniques to wage equations and find that the extent of bias is limited in statistical significance, while its sign implies that high wage workers tend to drop out from the experiment: this is consistent with the fact that high wage individuals benefit less from the experiment and are thus more likely to abandon it. They also find that attrition is more severe in simple analyses of variance rather than in structural models and suggest that this occurrence arises from the fact that in the latter case the conditioning set already includes the factors determining attrition (this point is also noted by Verbeek and Nijiman [1992]). Keane et al. [1988] are interested in analysing self-selection over the business cycle and thus to investigate the issue of wage cyclicality when macroeconomic shocks do not hit workers at random. The framework is again that of selectivity correction. They find that attrition significantly biases wages in a procyclical direction, suggesting that high wage workers exit the employed pool during downturns. The issue of attrition in the context of wage mobility modelling[3] is addressed in Bingley et al. [1995], who use a trivariate probit model to tackle selectivity of both initial conditions and attrition. Their results point towards a statistically significative impact of attrition on mobility, with attrition probabilities positively correlated with upward mobility.

The paper is organised as follows. Section 2 describes the features of the attrition process in the SHIW data, while in section 3 the trivariate switching probit is set out. Section 4 presents the results from the simulated maximum likelihood

[3] In the study of Stewart and Swaffield [1999], the impact of attrition on the bivariate probit estimates of low-pay persistence is investigated by amalgamating exits from the sample together with persistence in low-pay, not moving up the wage distribution being the common factor between these two

estimation of the model, while in section 4 some conclusions are drawn. Details on the implementation of the maximum likelihood estimator are given in the Appendix.

## 2. Attrition in the SHIW data

Before moving on to the modelling stage, this section describes the features of the attrition process in the Bank of Italy's data set. In this context, an important aspect of the sampling design is the distinction between panel and non-panel households, the first group corresponding to those households sampled in (at least) two consecutive waves. Assignment to this group is carried out in two steps. In the first step, which takes place at the date of the first wave's interview (1993 in our case), each household is asked whether or not it is willing to be re-interviewed in the subsequent wave. At the second step, which takes place previous to interviews for the subsequent wave, roughly 50% of those households available for re-interview are sampled to take part in the new wave. In 1993, 87% of interviewed households (7040 out of 8089 households) gave their availability for a new contact in 1995; of these, 47% where actually re-interviewed. A limited number of households (299) were also re-sampled among those answering NO or DON'T KNOW at the question on availability for future interviews. On the whole, of the 8089 households forming the 1993 wave of the SHIW, 3645 belong to the panel sub-group.

Moving from the household to the individual level and focusing on the group of full-time wage earners with valid wage observations aged between 18 and 65 in 1993, which is the sample implicitly deemed to be randomly selected in Cappellari [1999] when cross-sectional probit regressions for the probability of being low-paid were carried out, such a sampling design implies that of the 5708 valid observations, only 2734 belong to panel households, of which 2160 (see again Table 4.1) have a valid wage in both 1993 and 1995.

The sampling process just described suggests that some caution should be exerted when defining the control group for the attrition analysis: a considerable number of cases exit the sample at random, i.e. from a decision of the survey builders, and not for economic or demographic reasons. It would clearly make no sense to include these observations in an analysis of the probability of staying in the sample.

---

outcomes. They find that results are not dramatically different when compared to those obtained on the balanced sample.

Table 1: Transition probabilities into 1995 status for the sample of 1993 wage earners aged from 18 to 65 belonging to panel households (low pay defined as bottom quintile of hourly wage distribution)

| **1993 wage status** | Low-pay | High-pay |
|---|---|---|
| **1995 status** | | |
| low-pay | 0.388 | 0.051 |
| high-pay | 0.304 | 0.761 |
| missing wage; part-time | 0.049 | 0.037 |
| self employed | 0.027 | 0.01 |
| entrepreneur | 0.012 | 0.01 |
| unemployed | 0.092 | 0.019 |
| retired | 0.022 | 0.068 |
| other | 0.004 | 0 |
| housewife | 0.027 | 0.002 |
| not observed | 0.076 | 0.043 |
| Total obs | 490 | 2244 |

The analysis of this paper thus enlarges the estimation sample to include those observations which belong to a panel household in 1993 but don't have an observable wage in 1995: these are observations which could have potentially stayed in the sample of wage earners, but are not observed in the arrival wage distribution, either because they left the employed labour force or the household of origin. Potentially, also those belonging to households refusing to cooperate (and not actually re-sampled) could have been used to form the control group; however, the reasons behind the willingness to cooperate in the subsequent wave are not clear and it has been preferred not to include these cases (a total of 367 individuals) in the analysis.[4]

In order to get an illustration of the kind of movements out from the wage distribution which determine the attrition process, Table 1 gives the destinations in 1995 for the sample of wage earners belonging to a panel household and aged between 18 and 65 in 1993, i.e. the estimation sample for the present paper. Focusing on the comparison between low- and high-paid in 1993, it can be seen how the low-paid are characterised by higher transition rates especially in the group of the unemployed, and, to a minor extent, in the housewives and the "not observed" classes. On the other hand, the high paid have higher transition rates into retirement.

[4] Similarly, individuals belonging to panel households and with a valid wage only in 1995 (34 cases) are not included in the control group of the attrition equation.

An alternative illustration of the attrition process is given in Table 2, which reports results from probit regressions for the probability of having a valid wage in both 1993 and 1995 on a set of personal characteristics: the event under investigation is persistence in the sample of valid wage earners, while explanatory variables are measured at the beginning of the transition. The estimation sample differs from the one considered in Table 1 due to the presence of missing values in some of the explanatory variables; the same remark applies to the sample used in estimating the model of the next section.

Column 1 considers the effect of the wage determinants used in the reduced form low-pay probits of Cappellari [1999] (i.e. the selection equations of the bivariate switching probit), but without controlling for parental backgrounds. We can observe that the probability of persisting in sample displays an inverted u-shaped profile in labour market experience, indicating a higher sample attachment towards the central part of the working career, with maximum probability approximately 18 years after the beginning of the first job. Education has a positive impact on such a probability, while being female reduces it by 4 percentage points.

Table 2: Probit estimates (marginal effects) for the probability of having a valid wage in 1993 and 1995 (asymptotic t-ratios in parentheses).

| | **1** | | **2** | |
|---|---|---|---|---|
| experience/10 | 0.219 | (9.73) | 0.146 | (5.74) |
| (experience/10)^2 | -0.059 | -(11.30) | -0.044 | -(7.77) |
| education>=high school | 0.038 | (1.73) | 0.048 | (2.20) |
| female | -0.043 | -(2.46) | 0.017 | (0.64) |
| living in the north | -0.005 | -(0.30) | 0.008 | (0.49) |
| non-manual | 0.006 | (0.29) | 0.010 | (0.46) |
| firm size>=100 | 0.049 | (2.36) | 0.047 | (2.23) |
| public sector | 0.104 | (4.63) | 0.097 | (4.35) |
| agricolture | -0.038 | -(0.74) | -0.046 | -(0.89) |
| service sector | -0.010 | -(0.46) | -0.009 | -(0.41) |
| dependent children | | | 0.076 | (2.96) |
| dependent children*female | | | -0.096 | -(2.16) |
| married | | | 0.099 | (3.22) |
| married*female | | | -0.048 | -(1.22) |
| per capita equivalized household wealth (millions of lire) | | | -0.153 | -(2.99) |
| n. obs | 2716 | | 2716 | |
| pseudo r2 | 0.0792 | | 0.0971 | |

Among the other variables considered, while geographical location, occupation and sectoral affiliations (within the private sector) have no significant impact, affiliation to the public sector or employment in large firms positively influence retention probabilities on a scale between 5 and 10 percentage points. Column 2 augments the same specification with some reservation wage indicators; these are dummies for the presence of dependent children (i.e. aged less than 14) in the household and for being married, interacted with the gender dummy, and the per capita equivalised household wealth.[5] The first two variables are assumed to potentially influence workers' motivation to participate in the labour market in different directions depending upon gender: while for males the presence of family responsibilities is supposed to raise incentives to participate, for females negative signs could arise from the social structuring of child and household care. On the other hand, wealth is assumed to raise the reservation wage irrespective of gender. We can first of all observe that, with the exception of the gender dummy, all the other effects are robust to the inclusion of reservation wage indicators. The reasons for the loss in size and significance of the gender dummy become clear as we move to consider the new variables included in the regression. In particular, considering the indicator for the presence of dependent children and its interaction with the gender dummy, we can see how these effects come with the expected sign: a male with a dependent children within the household has a higher probability (7.6%) of staying in sample than an otherwise identical worker, while for females this probability is 2% lower than for otherwise identical males without dependent children within the household.[6] Taking the effect of marriage into account, it is rather strong (10%) and positive for males, while for females it is less intense and statistically significant. Finally, the household wealth indicator displays the expected negative sign.[7]

The probit analysis above has shed some light on the factors influencing exits from the sample of wage earners in 1995; in the next paragraph I will build on this in trying to model attrition bias within the model of low-pay transitions.

---

[5] Compulsory education usually lasted until 13 in the years examined. The equivalising factor for the wealth indicator is the square root of the number of household members.

[6] Recall that the sample is not selected with respect the position in the household, i.e. this evidence is based also on 357 sons and 226 daughters, plus 55 other relatives. Thus, the sample departs a bit from a stylised model of labour supply allocation between husband and wife. Here I am implicitly assuming that the factors determining the allocation of family responsibilities in the case of husbands and wives also influence the decisions of working sons and daughters.

[7] In a separate non reported analysis, the effect of wealth have been differentiated by sex, finding that for females such effect is still negative, but greater that the male one, although at the 17% level of significance.

## 3. A sequentially nested trivariate probit model for the analysis of attrition bias

This section describes the modelling approach adopted in order to assess to what extent the Italian evidence on low-pay transition probabilities is plagued by the presence of attrition bias. Is it correct to focus the analysis of wage transitions only on those individuals for whom a valid wage is available in each panel wave? Or, on the contrary, are these individuals systematically different from the ones dropping out from the survey, so that their selection for estimation is endogenous, thus biasing estimation results?

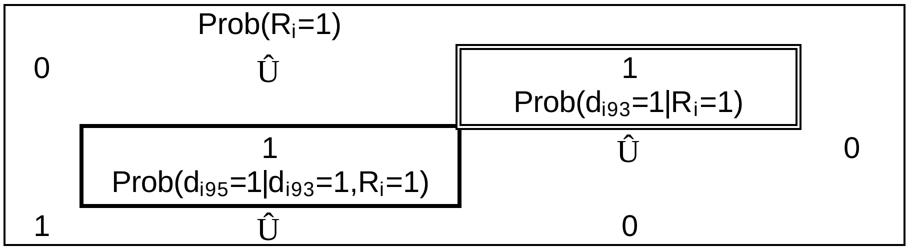
To provide an answer to these questions, the (structure of the) model used in Cappellari [1999] has been extended to allow for a third dichotomic event which interacts with the ones previously considered (i.e. low-pay/high-pay at both ends of the transition investigated) in determining the likelihood of the data. The resulting set-up is a trivariate conditional probit model which allows for sequential nesting of the equation of interest. The sequential nesting structure resembles the one used by Bingley et al. [1995], with two relevant differences: first, in this chapter attrition is considered only with respect to the arrival wage distribution (as in Hausman and Wise), while in Bingley et al. attrition in the starting wage distribution is also taken into account; secondly Bingley et al. consider discrete indicators of wage changes conditional on starting wage levels, while here wage levels are conditioned on starting wage levels. However, in the case of mobility from the tails of the distribution, the two specifications are observationally equivalent.

The modelling of attrition is carried out expanding the model for low-pay persistence proposed by Stewart and Swaffield [1999], i.e., differently from the model of Cappellari [1999], the 1995 wage outcome is assumed to be observable only for the 1993 low-paid. The model is expanded by acknowledging that it can be actually estimated only using observations for which a wage is observable in 1993 and 1995. Let $R_i$ be a dummy partitioning the sample of 1993 wage earners[8] depending upon their wage observability in 1995; let us also define $d_{it}$ as dummy variables for low-pay

[8] Recall from section 2 that these are wage earners belonging to panel households, this last state being assumed exogenous to low-pay transitions.

occurrence in year t. The sequential nesting structure of the model is represented in Figure 1.

Fig. 1: The structure of the trivariate nested probit




At the first nesting level, a probit is estimated for the probability of having a valid wage in both periods. At the second nesting level, only those observations for which $R_i$=1 are utilised to estimate a probit for the probability of low-pay in 1993. Finally, the sample of the 1993 low-paid observed in both waves is used to estimate a probit equation for low-pay in 1995; of course, the nesting sequence just described is simultaneously estimated. It is worth stressing that the multivariate normal density assumed allows for unrestricted correlation between the errors, thus allowing a proper assessment of potential endogeneity issues among the three events investigated.

More formally, let's assume that the propensity to stay in sample (retention propensity) is a latent variable R*; when R* overcomes an unobservable (possibly individual specific) threshold $\tau^*$, observations remain in the sample of wage earners in both waves. R* is assumed to be a function of observable characteristics, and we only observe a dummy indicator R signalling whether or not $R^* > \tau^*$:

$$\begin{aligned} R_i^* &= x'_{Ri}\,\delta_R + v_i \\ R_i &= I(R_i^* > \tau_i^*) \\ v_i &\sim N(0,1) \end{aligned} \qquad (1)$$

where $x'_{Ri}$ contains the whole set of explanatory variables used in the model and I(A) is a dummy equal to 1 when A is true and to 0 otherwise.

The second stage can be formalised according to the discussion in Cappellari [1999] and assuming partial observability of the 1993 low-pay outcome:

$$
\begin{aligned}
& g(w_{i93}) = x'_i\,\delta + u_i \qquad if \qquad R_i = 1 \\
& d_{i93} = I(g(w_{i93}) \le g(\lambda_{93})) \\
& u_i \sim N(0,1)
\end{aligned}
\tag{2}
$$

where g( ) is a monotonic transformation such that u is normally distributed, w is the relevant wage rate, variables in x are a subset of those in $x_R$ and $\lambda_{93}$ is the low-pay threshold in the 1993 wage distribution.

The headline equation of interest is a probit equation for the occurrence of low-pay in 1995 for which two sources of partial observability are assumed:

$$
\begin{aligned}
& h_1(w_{i95}) = z_i'\eta_1 + \varepsilon_{1i} \qquad if \qquad R_i = 1 \quad and \quad d_{i93} = 1 \\
& d_{i95} = I(h_1(w_{i95}) \le h_1(\lambda_{95})) \\
& \varepsilon_1 \sim N(0,1)
\end{aligned}
\tag{3}
$$

where variables in z are a subset of those in x. Assuming that error terms in the three equations are jointly distributed as a tri-variate normal

$$
\begin{pmatrix} v_i \\ u_i \\ \varepsilon_{i1} \end{pmatrix} \sim N_3 \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & & \\ \rho_1 & 1 & \\ \rho_2 & \rho_3 & 1 \end{pmatrix} \right]
\tag{4}
$$

and that observations are iid, the log-likelihood function of the model may be written as[9]:

$$
\begin{aligned}
\log L = \sum_i \{ & R_i d_{i93} d_{i95} \log \Phi_3 (x'_{Ri}\,\delta_R, x'_i\,\beta, z'_i\,\gamma_1; \rho_1, \rho_2, \rho_3) + \\
& R_i d_{i93} (1 - d_{i95}) \log \Phi_3 (x'_{Ri}\,\delta_R, x'_i\,\beta, -z'_i\,\gamma_1; \rho_1, -\rho_2, -\rho_3) + \\
& R_i (1 - d_{i93}) \log \Phi_2 (x'_{Ri}\,\delta_R, -x'_i\,\beta; -\rho_1) + \\
& (1 - R_i) \log \Phi(-x'_{Ri}\,\delta_R) \}.
\end{aligned}
\tag{5}
$$

[9] The way $\delta$ commutes into $\beta$ can be derived by (3):

$$
prob(d_{i93} = 1) = prob(w_{i93} \le l_{93}) = prob(g(w_{i93}) \le g(l_{93})) = \Phi(g(l_{93}) - x_i'd) = \Phi(x_i'b)
$$

where the new constant term in $\beta$ subsumes the difference between g($l$) and the old constant in $d$ and the coefficients associated with the individual characteristics in $b$ are the same as in $d$, but with opposite sign. Similar remarks apply to $\eta 1$ and $\gamma 1$.

The nested structure of the selection processes of the model implies a nested structure of identifying restrictions, i.e. $z \subset x \subset x_R$ . [10] For what concerns the first nesting level, here I assume that variables entering only $x_R$ are some of the reservation wage indicators included in Table 1, namely the dummy for the presence of dependent children in the household interacted with the gender dummy. Such variables have been chosen on the basis of a reduced form bivariate probit model in which $R_i$ and $d_{93}$ have been conditioned on a general specification of $x_{Ri}$[11]; results from the reduced form show how these two variables do not enter the low-pay equation significantly[12]; it is then assumed that their effect on wages in both time periods only works through participation in 1995. The choice of such variables is also in line with previous studies of attrition bias in panel wage analysis, namely with Keane et al. [1988], who use the number of kids as an instrument in their employment equation. As mentioned in the introduction, attrition may well result from demand side factors, and it could also be argued that such factors are more relevant at the lower end of the wage distribution, where monopsonistic behaviour is likely to characterise the labour market. However, it is difficult to imagine demand side factors, among the available information, which do not enter the wage equation directly. For what concerns identification of the transition equation with respect to the starting state equation, i.e. the second selection process where variables are needed which enter only x (and $x_R$) but not z, here I follow the identification strategy adopted in Cappellari [1999] and firstly proposed by Stewart and Swaffield [1999] and use a set of parental background indicators, the assumption being that they affect wage levels but, given this, they have no direct effect on wage changes. Moreover, given its nature of wage change equation, the transition equation doesn't include the square of labour market experience.[13]

As mentioned in the introduction, we can see from (5) that the log-likelihood function involves the c.d.f. of the trivariate normal distribution, whose evaluation has

---

[10] Due to the presence of non-linearities, the model would be statistically identified also without such restrictions.

[11] The specification includes the whole set of variables used in column 2 of Table 1 plus a set of parental background indicators.

[12] The p-value for these variables in the 1993 low-pay equation is .77 for the male dummy and .9 for the female one.

[13] Abstracting for a moment from the selectivity for attrition, this last restriction, which comes from the very structure of the model, gives exact identification in a bivariate probit for state and transition

been implemented via simulation estimation: the Appendix illustrates the practical implementation of such an estimator.

## 4. Results

Results from the simulated maximum likelihood analysis are given in Table 3, which reports SML estimated coefficients and asymptotic t-ratios for the two nesting equations and the low-pay transition equation; the analysis is restricted to the low-pay threshold defined in terms of the bottom quintile of the hourly wage distribution, while, as a benchmark, the first column of the Table reports results obtained with a nested bivariate model which only controls for the endogeneity of initial conditions.[14] The simulated likelihood function is computed using 75 random draws from the truncated normal distributions of interest.[15]

---

probabilities, enabling tests of the validity of parental backgrounds as overidentifying instruments. Such tests have been carried out in Cappellari [1999], and results support the validity of instruments.

[14] Results in column 1 are taken from Table 7 in Cappellari [1999], where bivariate and ordered probit specifications of the low-pay persistence equation were compared.

[15] In the appendix the performance of the SML estimator at different choices of the number of draws is checked, showing how estimates are robust to such a choice.

Table 3: Simulated maximum likelihood estimates (asymptotic t-ratios) of the sequentially nested trivariate probit. GHK simulator with 75 draws.

| | (1):bivariate probit without attrition | | (2): unrestricted trivariate probit | | (3): $\rho(R_i, d_{i93}) = \rho(R_i, d_{i95}) * \rho(d_{i93}, d_{i95})$ | | (4): restricted conditioning set | | (5): (3)&(4) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Low-pay 1995 | | | | | | | | | | |
| experience/10 | -0.072 | (0.680) | -0.079 | (0.531) | -0.057 | (0.408) | 0.120 | (1.456) | 0.122 | (1.546) |
| edu>=high sch. | -0.646 | (2.723) | -0.657 | (2.464) | -0.627 | (2.381) | | | | |
| female | 0.247 | (1.253) | 0.259 | (1.055) | 0.225 | (0.955) | 0.034 | (0.262) | 0.030 | (0.234) |
| non-manual | -0.139 | (0.567) | -0.133 | (0.474) | -0.110 | (0.395) | | | | |
| size>=100 | -0.405 | (1.383) | -0.432 | (1.318) | -0.399 | (1.233) | | | | |
| public sector | 0.068 | (0.205) | 0.033 | (0.088) | 0.066 | (0.183) | | | | |
| agriculture | 0.082 | (0.262) | 0.112 | (0.350) | 0.095 | (0.303) | | | | |
| service sector | 0.258 | (1.515) | 0.264 | (1.536) | 0.260 | (1.518) | | | | |
| living in the north | -0.382 | (2.300) | -0.392 | (2.101) | -0.370 | (1.980) | | | | |
| constant | 0.956 | (5.115) | 0.968 | (2.488) | 1.041 | (3.319) | 0.713 | (2.269) | 0.747 | (2.764) |
| Low-pay 1993 | | | | | | | | | | |
| experience/10 | -1.109 | (7.988) | -1.057 | (5.865) | -1.084 | (6.071) | -0.963 | (4.946) | -1.000 | -(6.387) |
| edu>=high sch. | -0.357 | (3.063) | -0.342 | (2.855) | -0.351 | (2.956) | | | | |
| female | 0.636 | (6.914) | 0.618 | (6.258) | 0.629 | (6.543) | 0.236 | (3.042) | 0.242 | (3.192) |
| non-manual | -0.540 | (4.494) | -0.533 | (4.416) | -0.538 | (4.472) | | | | |
| size>=100 | -0.668 | (5.787) | -0.648 | (5.307) | -0.658 | (5.499) | | | | |
| public sector | -1.010 | (8.056) | -0.977 | (6.908) | -0.995 | (7.204) | | | | |
| agriculture | 0.546 | (2.582) | 0.530 | (2.486) | 0.539 | (2.532) | | | | |
| service sector | 0.103 | (0.974) | 0.098 | (0.927) | 0.101 | (0.956) | | | | |
| living in the north | -0.159 | (1.848) | -0.158 | (1.835) | -0.157 | (1.820) | | | | |
| (exp./10)^2 | 0.202 | (6.338) | 0.186 | (4.136) | 0.194 | (4.334) | 0.175 | (3.518) | 0.185 | (4.729) |
| father blue coll. | -0.089 | (0.908) | -0.082 | (0.827) | -0.086 | (0.861) | 0.160 | (1.965) | 0.160 | (1.956) |
| father not empl. | 0.264 | (0.876) | 0.277 | (0.900) | 0.276 | (0.893) | 0.717 | (2.751) | 0.716 | (2.731) |
| father missing | 0.277 | (1.725) | 0.263 | (1.580) | 0.266 | (1.584) | 0.506 | (3.107) | 0.520 | (3.314) |
| mother blue coll. | 0.152 | (0.823) | 0.154 | (0.824) | 0.148 | (0.790) | 0.187 | (1.187) | 0.187 | (1.183) |
| mother not empl. | 0.239 | (1.587) | 0.233 | (1.542) | 0.237 | (1.567) | 0.124 | (0.981) | 0.127 | (1.001) |
| mother missing | 0.521 | (3.056) | 0.472 | (2.465) | 0.503 | (2.778) | 0.565 | (2.922) | 0.599 | (3.661) |
| father edu.>=hs | -0.386 | (2.082) | -0.382 | (2.058) | -0.388 | (2.099) | -0.631 | (3.801) | -0.643 | -(3.975) |
| mother edu>=hs | -0.284 | (1.191) | -0.281 | (1.182) | -0.284 | (1.194) | -0.383 | (1.889) | -0.388 | -(1.904) |
| constant | 0.488 | (2.325) | 0.393 | (1.428) | 0.441 | (1.625) | -0.543 | (2.502) | -0.505 | -(2.660) |

Table 3: continued

| Retention | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| experience/10 | | | 0.461 | (4.755) | 0.464 | (4.781) | 0.509 | (5.370) | 0.511 | (5.400) |
| edu>=high sch. | | | 0.160 | (1.944) | 0.160 | (1.936) | | | | |
| female | | | -0.038 | (0.510) | -0.037 | (0.498) | 0.028 | (0.394) | 0.028 | (0.387) |
| non-manual | | | 0.031 | (0.366) | 0.033 | (0.385) | | | | |
| size>=100 | | | 0.177 | (2.152) | 0.176 | (2.148) | | | | |
| public sector | | | 0.365 | (4.240) | 0.364 | (4.232) | | | | |
| agriculture | | | -0.139 | (0.772) | -0.144 | (0.802) | | | | |
| service sector | | | -0.030 | (0.381) | -0.030 | (0.377) | | | | |
| living in the north | | | 0.003 | (0.044) | 0.002 | (0.040) | | | | |
| (exp./10)^2 | | | -0.151 | (7.047) | -0.151 | (7.092) | -0.161 | (7.640) | -0.161 | -(7.734) |
| dep. children | | | 0.354 | (3.884) | 0.353 | (3.863) | 0.343 | (3.834) | 0.340 | (3.819) |
| dep.chil.*female | | | -0.433 | (3.240) | -0.438 | (3.276) | -0.425 | (3.285) | -0.428 | -(3.308) |
| father blue coll. | | | 0.114 | (1.643) | 0.113 | (1.636) | 0.036 | (0.540) | 0.035 | (0.529) |
| father not empl. | | | 0.306 | (1.146) | 0.307 | (1.151) | 0.242 | (0.912) | 0.244 | (0.922) |
| father missing | | | -0.107 | (0.852) | -0.116 | (0.933) | -0.210 | (1.677) | -0.218 | -(1.778) |
| mother blue coll. | | | -0.057 | (0.446) | -0.061 | (0.480) | -0.081 | (0.648) | -0.084 | -(0.677) |
| mother not empl. | | | -0.108 | (1.145) | -0.108 | (1.153) | -0.089 | (0.973) | -0.090 | -(0.980) |
| mother missing | | | -0.513 | (4.233) | -0.512 | (4.221) | -0.561 | (4.688) | -0.563 | -(4.708) |
| father edu.>=hs | | | -0.002 | (0.015) | -0.008 | (0.070) | 0.070 | (0.630) | 0.065 | (0.594) |
| mother edu>=hs | | | -0.072 | (0.540) | -0.071 | (0.532) | -0.005 | (0.040) | -0.005 | -(0.037) |
| constant | | | 0.531 | (3.454) | 0.533 | (3.463) | 0.764 | (5.524) | 0.768 | (5.582) |
| | | | | | | | | | | |
| r(Ri,di93) | | | 0.193 | (0.591) | | | 0.325 | (0.777) | | |
| r(Ri,di95) | | | -0.062 | (0.110) | -0.187 | (0.377) | -0.357 | (0.796) | -0.412 | -(1.088) |
| r(di93,di95) | -0.451 | (1.769) | -0.420 | (1.008) | -0.485 | (1.337) | -0.516 | (2.271) | -0.523 | -(2.449) |
| n.obs | | 2148 | | 2716 | | 2716 | | 2716 | | 2716 |
| logLik | | -800.65 | | -2053.78 | | -2053.87 | | -2263.69 | | -2263.7 |

None of the correlation coefficients is significant at conventional levels and the more precisely estimated is the correlation between low-pay level and low-pay persistence, which also preserves the sign and size it had in the analysis of Cappellari [1999]. The correlation between retention and low-pay in the starting year is positive, a result which also arises in Bingley et al. [1995]. The correlation between retention and low-pay persistence is instead negative, meaning that those staying in sample have a lower propensity to remain in low-pay. However, neither is significantly different from zero.

Recalling the discussion from the introduction, such a result seems to indicate that observations abandoning the sample correspond to "weaker" labour market participants, and thus that their exclusion from the analysis could lead us to overestimate the whole phenomenon of persistence. As stressed above, however, the extent of the bias is irrelevant from the viewpoint of statistical significance.

As a next step in the analysis, restrictions are tested on the general model: after all, the previous finding of statistical insignificance of attrition bias could arise from the fact the structure imposed on the data is too complex to be precisely estimated, so that before concluding in favour of the irrelevance of attrition it is worth checking whether or not the finding is also supported by restricted specifications of the model in column 2.

A first restricted version of the model is proposed in column 3, where the null hypothesis that the correlation between retention and initial low-pay is the product of the other two correlation coefficients is tested: $\rho(R_i,d_{i93})=\rho(R_i,d_{i95})*\rho(d_{i93},d_{i95})$. The hypothesis means that the correlation between initial low-pay and retention only works through the combination of the correlation between retention and final low-pay ($\rho(R_i,d_{i95})$) and the correlation between initial and final low-pay, i.e. the individual effect in earnings ($\rho(d_{i93},d_{i95})$); apart from this combined effect, there's no direct correlation between $R_i$ and $d_{i93}$. This hypothesis is adopted from the outset of the analysis by Hausman and Wise [1979]. The structure of the attrition process in their data is pretty similar to ours in that observations availability in the starting year is assumed to be exogenous, and only non-random attrition in the second year is tested (i.e. no observations available only in the second year are utilised in estimation), which makes the hypothesis worth testing. Also, the signs (but not the sizes) of the correlation coefficients in column 2 are in accordance with this hypothesis. Results are given in column 3. By first considering the maximised simulated likelihood and comparing it to the one of the unrestricted model in column 2 via a Likelihood Ratio test, we obtain a $\chi 2$ statistic of 0.18, which strongly supports the non-rejection of the null. The impact on the estimated parameters is negligible as far as the effect of explanatory variables is concerned. Taking the two remaining correlation coefficients into account, we can see that they both gain in size and precision, with the correlation between initial low-pay and low-pay persistence approaching conventional levels of statistical significance. This remark does not apply to the correlation coefficient between low-pay persistence and sample retention (the main object of the analysis in this paper), whose sign still indicates that retention is negatively correlated with low-pay persistence.

Results up to this point indicate that the extent of attrition bias is pretty weak, a finding which also arises in Hausman and Wise [1979]. As shown by those authors, this finding emerges in a "structural" model of earnings, i.e. where the conditioning set contains a set of explanatory variables deemed to "cause" earnings, and which are likely also to determine the attrition process. This is actually true in their study: in a simple variance decomposition analysis of earnings they obtain a significant effect of attrition on earnings. A natural question which arises is then whether or not the finding of non significant attrition bias in our model of low-pay persistence is also due to the features of the conditioning set. To provide an answer to such a question, column 4 of Table 3 further simplifies the model of low-pay persistence, excluding both demand and supply side factors appearing in the transition equation from the model; consistently, such variables are also excluded from the selection equations.[16] By comparing the maximised simulated likelihood function with the one from column 2 with a Likelihood Ratio test, we obtain a $\chi 2$ statistic of 419.82, which is well above the critical values of the $\chi 2$ distribution with 21 degrees of freedom (7 variables are excluded from each equation) at usual confidence levels, thus clearly rejecting the restriction imposed. We can notice how the exclusion of the set of explanatory variables brings labour market experience to the verge of statistical significance in the low-pay transition equation, although with a reverse sign which arises from the fact that we are not controlling for other factors, for example education. On the other hand, the coefficient on the gender dummy looses both size and significance. Focusing on

[16] The two explanatory variables left in the transition equation are the gender dummy and the linear term in labour market experience. The reason for leaving these variables in the model is to maintain a quadratic profile in experience for the low-pay selection equation, and a comparison category for the "instruments" of the retention selection equation.

the estimated error covariance matrix, we can observe a general rise in size and precision for each correlation coefficient, in particular for $\rho(d_{i93},d_{i95})$, which is now significantly different from zero at conventional levels. Gains in precision also characterise the estimate of $\rho(R_i,d_{i95})$, but not enough to conclude in favour of the relevance of attrition bias. Thus, some effect of attrition seems to be present in this less-strucured specification, but both the fact that the restricted model is not supported by the data and the unsatisfactory precision characterising the attrition bias parameter even in this case clearly suggest that attrition can be deemed ignorable in this case.

A final test for the relevance of attrition bias is reported in column 5, which combines the restrictions of columns 3 and 4, i.e. $\rho(R_i,d_{i93})=\rho(R_i,d_{i95})*\rho(d_{i93},d_{i95})$ with the exclusion of structural explanatory variables from the conditioning set. Again, these restrictions are clearly rejected at conventional levels (the unrestricted model is the one in column 2). Taking the attrition bias parameter into account, we can observe a further gain in size and precision, which is, however, not enough to conclude in favour of the relevance of attrition bias.[17] Thus even if the combination of a restricted covariance matrix and a restricted conditioning set was supported by the data, its effect on the attrition parameter would not lead us to reject the analysis of Chapter 4 for suffering from attrition bias.

## 5. Summary and conclusions

This paper has investigated the extent to which the existing econometric models of low-pay transitions in Italy are affected by the presence of attrition bias, i.e. by the non-randomness of the propensity with which workers with a valid wage at the beginning of the transition observed leave the sample of wage earners during such a transition.

Focusing on this problem of sample selection involved some computational difficulties: controlling for attrition bias required expanding the bivariate probit framework of Cappellari [1999] to include a third limited dependent variable equation and hence the resulting likelihood function included trivariate normal integrals which are not packaged within statistical software. The problem has been tackled by implementing a simulated maximum likelihood estimator, in particular adopting the so-called GHK simulator, using STATA's maximum likelihood routine; details on the construction of the simulated estimator are given in the Appendix.

Results obtained point towards the ignorability of attrition bias: various versions of a sequentially nested trivariate probit model in which the first stage controls for non-random attrition have been estimated on the SHIW data and in no case does the parameter measuring the extent of attrition reach statistical significance at conventional levels. The maximum level of precision for this parameter has been reached by restricting both the error covariance matrix and the conditioning set, with this last restriction not supported by the data. The sign of the parameter indicates that persisting in sample and persisting in low-pay negatively covary; had this parameter been significantly different from zero, this would have meant that the exclusion of attrited observations lead us to underestimate the true extent of low-pay persistence.

The finding of irrelevant attrition mirrors previous results from Hausman and Wise [1979] in the context of structural models of earnings. Opposite conclusions have been obtained in models of wage mobility by Bingley et al. [1995].

A final word of caution has to be issued in order to correctly interpret the results of this Chapter. The sampling design of the SHIW panel is peculiar in that about half of workers observed in the 1993 wave leave the sample at random due to a decision of the survey builders, and such observations have not been used in the estimation of the trivariate probit described above. As a consequence, the control group, i.e. attrited observations, in the analysis is relatively small (the balanced sample is enlarged by 26%), so that it may be that it doesn't provide enough variability to capture the effect of attrition. Results presented have thus to be viewed as contingent on the peculiar sample structure, and confirmation of such findings should be pursued in the future on panel data sets with more conventional sample design.

[17] The parameter is now significant at the 28% confidence level.

## Appendix. Simulated maximum likelihood estimation of the trivariate probit model

The sequentially nested trivariate probit model utilised in this paper requires evaluation of trivariate normal integrals, thus posing computational problems due to the fact that the function evaluating such integrals is not available among those usually packaged within commonly used econometric software. Moreover, multiple integrals are hardly tractable by usual linear numerical approximations such as those based on the Newton Raphson method, and produce unreliable results in terms of the goodness of approximation (Hajivassiliou and Ruud [1994]).

As an alternative to numerical approximations, simulation based inference has been developed in recent years (see Stern [1997] for a survey and Gourieroux and Monfort [1996] for an extensive presentation of simulation estimation techniques and its applications in various context; see also Börsch-Supan et al. [1992], Börsch-Supan and Hajivassiliou [1993] and Hajivassiliou and Ruud [1994] for applications of the GHK (Geweke-Hajivassiliou-Keane)-smooth recursive conditioning simulators in the context of ML estimation of limited dependent variable models). The basic idea of simulated maximum likelihood estimation (SML) is to replace the intractable bit of the likelihood function by its simulated counterpart. This appendix illustrates how this is practically done in the case of the trivariate probit model, and shows the implementation of the SML-GHK estimator using STATA's maximum likelihood routine. The illustration of the method is carried out in terms of a (complete) trivariate probit, i.e., when full observability of the three variables is assumed.

The model of interest is a (seemingly unrelated) trivariate probit:

$$
\begin{aligned}
& Y_{ij}^{*} = X_{ij}\beta_j + u_{ij} \\
& Y_{ij} = I(Y_{ij}^{*} > 0) \\
& j = 1,2,3 \\
& i = 1,..,N \\
& \begin{pmatrix} u_{i1} \\ u_{i2} \\ u_{i3} \end{pmatrix} \sim \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & & \\ \rho_{12} & 1 & \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix} \right]
\end{aligned}
\tag{A.1}
$$

where I(A) is a dummy indicating whether or not A is true. Assuming observations are i.i.d., the log-likelihood function of the model is:

$$\log L(\beta_j, \rho_{jl}; X_j, Y_j) = \sum_i \log\{\Phi_3[K_{i1}X_{i1}\beta_1, K_{i2}X_{i2}\beta_2, K_{i3}X_{i3}\beta_3; \\ K_{i1}K_{i2}\rho_{12}, K_{i1}K_{i3}\rho_{13}, K_{i2}K_{i3}\rho_{23}]\} \tag{A.2}$$

$$j,l = 1,2,3$$
$$K_{ij} = 2Y_{ij} - 1$$

which involves the trivariate standard normal c.d.f. $\Phi_3$ and is hardly tractable with traditional numerical approximation.

The main intuition behind the GHK smooth recursive conditioning simulator is to exploit the definition of conditional distribution functions[18]:

$$\Pr(u_1 \le X_1\beta_1, u_2 \le X_2\beta_2, u_3 \le X_3\beta_3) = \\ \Pr(u_3 \le X_3\beta_3 | u_2 \le X_2\beta_2, u_1 \le X_1\beta_1) * \Pr(u_2 \le X_2\beta_2 | u_1 \le X_1\beta_1) * \\ \Pr(u_1 \le X_1\beta_1) \tag{A.3}$$

and to replace the joint multivariate normal with the product of sequentially conditioned univariate normal distribution functions. The expression in (A.3) involves conditioning upon unobservables: if some approximation for these conditional distributions can be found, then the likelihood function only requires evaluation of univariate integrals which is feasible within ordinary statistical packages.

Consider the Cholesky decomposition of the errors' covariance matrix:

$$E(uu')=\Sigma=Cee'C' \tag{A.4}$$

where C is the lower triangular Cholesky factor of $\Sigma$ and $e \sim N_3(0,I_3)$, from which it follows that:

[18] I drop i indices for notational convenience.

$$u_1 = c_{11}e_1$$
$$u_2 = c_{21}e_1 + c_{22}e_2 \qquad (A.5)$$
$$u_3 = c_{31}e_1 + c_{32}e_2 + c_{33}e_3$$

where $c_{ji}$ is the C element in position ji.

Thus, we can re-write (A.3) as:

$$\begin{aligned}&\Pr(u_1 \leq X_1\beta_1, u_2 \leq X_2\beta_2, u_3 \leq X_3\beta_3) = \\ &\quad \Pr(e_3 \leq (X_3\beta_3 - c_{32}e_2^* - c_{31}e_1^*)/c_{33})\Pr(e_2 \leq (X_2\beta_2 - c_{21}e_1^*)/c_{22}) \\ &\quad \Pr(e_1 \leq X_1\beta_1/c_{11})\end{aligned} \qquad (A.6)$$

where $e^*_1$ and $e^*_2$ come from standard normal distributions with upper truncation points at $X_1\beta_1/c_{11}$ and $(X_2\beta_2 - c_{21}e^*_1)/c_{22}$ respectively, i.e. they satisfy the conditioning events in (A.3). It is worth stressing that we are now working with uncorrelated errors (the vector e) and that correlation between the elements of the original vector of errors u has been transferred to the truncation points of the sequential conditioning via the Cholesky decomposition of $\Sigma$.

Evaluation of the probability in (A.6) involves unobservable terms $e^*_2$ and $e^*_1$. Let's introduce R random draws of $e^*_1$ and $e^*_2$, i.e. random draws of $e_1$ and $e_2$ from upper truncated standard normals, with truncation points given above. The GHK simulator of (A.6) is the arithmetic mean of the R probabilities we obtain for each of these draws:

$$\begin{aligned}\tilde{P}_{GHK} &= \frac{1}{R}\sum_{r=1}^{R} \{\Pr(e_3 \leq (X_3\beta_3 - c_{32}\tilde{e}^r{}_2 - c_{31}\tilde{e}^r{}_1)/c_{33}) \\ &\quad \Pr(e_2 \leq (X_2\beta_2 - c_{21}\tilde{e}^r{}_1)/c_{22})\Pr(e_1 \leq X_1\beta_1/c_{11})\} = \\ &\quad \frac{1}{R}\sum_{r=1}^{R} \{\Phi(X_3\beta_3 - c_{32}\tilde{e}^r{}_2 - c_{31}\tilde{e}^r{}_1)/c_{33}) \\ &\quad \Phi((X_2\beta_2 - c_{21}\tilde{e}^r{}_1)/c_{22})\Phi(X_1\beta_1/c_{11})\}\end{aligned} \qquad (A.7)$$

where $\tilde{e}^q{}_j$ is the q-th draw for $e^*_j$. The SML estimator is then obtained by replacing the cumulative trivariate normal distributions in (A.2) by their simulated counterparts from (A.7). Note that the resulting maximand will be conditional on the set of draws: for computational stability it is then important that such draws do not change with the parameter values during optimization steps (Hajivassiliou [1997]).

The last thing which is to be explained is how to generate random variables from upper truncated normal distributions. Such variables can be obtained by exploiting random number generators on the unit interval available in statistical packages and the inversion formula given, among others, in Stern [1997]. First of all, let's consider the relationship between draws from the uniform distribution on the unit interval (say v) and the corresponding random draws (say z) from the standard normal distribution; such a relationship is given by:

$$z = \Phi^{-1}(v)\,. \tag{A.8}$$

Draws for (say) upper truncated standard normals can be similarly obtained by recalling that in this case $F(z)=\Phi(z)/\Phi(b)$ where F(.) indicates the cumulative density function of the truncated variable and b is the upper truncation point; replacing F(z) by the uniform on the unit interval and solving the expression for z we get:

$$z=\Phi^{-1}(v\Phi(b)). \tag{A5.9}$$

Börsch-Supan and Hajivassiliou [1993] highlight the key features of the GHK simulator in the context of multivariate normal LDV models:

- simulated probabilities are unbiased;
- such probabilities are bounded in the (0,1) interval;
- the simulator is a continuous and differentiable function of the model's parameters.

They also show that GHK is more efficient, in terms of variance of probabilities' estimates, than other simulators such as the acceptance-rejection or the Stern simulator. Note that unbiasedness of simulated probabilities doesn't translate into unbiasedness of the logs of such probabilities, which is what is needed to compute the log-likelihood function. However, such bias becomes negligible as the number of draws is raised with the sample size (Hajivassiliou [1997]).

Table A.1: Comparison of bivariate probit ML and SML (R=75) estimates (asymptotic standard errors)

| | ML | | SML | |
|---|---|---|---|---|
| Y1 | | | | |
| x11 | -1.2057 | (0.3394) | -1.1987 | (0.3395) |
| x12 | 0.2212 | (0.0900) | 0.2200 | (0.0901) |
| x13 | 0.1244 | (0.2338) | 0.1360 | (0.2340) |
| x14 | 0.1025 | (0.2869) | 0.0898 | (0.2866) |
| Y2 | | | | |
| x21 | -0.3997 | (0.3480) | -0.3945 | (0.3479) |
| x22 | 0.0455 | (0.0932) | 0.0438 | (0.0932) |
| x23 | 0.3155 | (0.2191) | 0.3231 | (0.2192) |
| x24 | -0.5707 | (0.3047) | -0.5718 | (0.3049) |
| rho | 0.7450 | (0.0901) | 0.7530 | (0.0870) |
| n.obs | 200 | | 200 | |
| logLik | -146.97 | | -146.603 | |

Table A.2: Comparison of SML (R=100) trivariate probit estimates (asymptotic standard errors) between LIMDEP and STATA

| | LIMDEP | | STATA | |
|---|---|---|---|---|
| Y1 | | | | |
| x11 | -0.9230 | (0.2954) | -0.9227 | (0.2616) |
| x12 | 0.7476 | (0.1245) | 0.7474 | (0.1224) |
| x13 | -1.6734 | (0.2555) | -1.6537 | (0.2392) |
| Y2 | | | | |
| x21 | 0.1721 | (0.2627) | 0.1661 | (0.2607) |
| x22 | -1.2202 | (0.3306) | -1.1926 | (0.3172) |
| x23 | 0.2218 | (0.0861) | 0.2150 | (0.0835) |
| Y3 | | | | |
| x31 | 0.7563 | (0.2536) | 0.7607 | (0.2307) |
| x32 | -0.3801 | (0.1051) | -0.3834 | (0.0966) |
| x33 | 0.5648 | (0.2048) | 0.5554 | (0.1915) |
| rho12 | -0.1127 | (0.1810) | -0.0950 | (0.1758) |
| rho13 | 0.0742 | (0.1453) | 0.0878 | (0.1438) |
| rho23 | -0.5409 | (0.1081) | -0.5445 | (0.1181) |
| n.obs | 200 | | 200 | |
| logLik | -272.069 | | -272.145 | |

Table A.3: Behaviour of the STATA's SML estimator by different choices of R (asymptotic standard errors)

| | R=75 | | R=100 | | R=150 | |
|---|---|---|---|---|---|---|
| Y1 | | | | | | |
| x11 | -0.9207 | (0.2610) | -0.9227 | (0.2616) | -0.9241 | (0.2620) |
| x12 | 0.7461 | (0.1221) | 0.7474 | (0.1224) | 0.7477 | (0.1226) |
| x13 | -1.6178 | (0.2304) | -1.6537 | (0.2392) | -1.6610 | (0.2415) |
| Y2 | | | | | | |
| x21 | 0.1647 | (0.2618) | 0.1661 | (0.2607) | 0.1505 | (0.2595) |
| x22 | -1.1752 | (0.3179) | -1.1926 | (0.3172) | -1.1647 | (0.3129) |
| x23 | 0.2095 | (0.0837) | 0.2150 | (0.0835) | 0.2088 | (0.0826) |
| Y3 | | | | | | |
| x31 | 0.7642 | (0.2311) | 0.7607 | (0.2307) | 0.7571 | (0.2305) |
| x32 | -0.3864 | (0.0966) | -0.3834 | (0.0966) | -0.3825 | (0.0965) |
| x33 | 0.5436 | (0.1919) | 0.5554 | (0.1915) | 0.5593 | (0.1915) |
| rho12 | -0.0920 | (0.1710) | -0.0950 | (0.1758) | -0.1243 | (0.1717) |
| rho13 | 0.0765 | (0.1306) | 0.0878 | (0.1438) | 0.0676 | (0.1435) |
| rho23 | -0.5155 | (0.1206) | -0.5445 | (0.1181) | -0.5504 | (0.1185) |
| n.obs | 200 | | 200 | | 200 | |
| logLik | -272.8304 | | -272.145 | | -272.2269 | |

The remaining part of this appendix reports some robustness checks performed on the SML estimator built in STATA.

The robustness checks of SML estimator built using STATA's maximum likelihood routines are quite interesting. Table A.1 compares a simulated bivariate probit estimate against its exact (or, more correctly, numerically approximated) counterpart, which is packaged within STATA. The sample utilised consists of 200 observations randomly extracted from the SHIW data set. Either size and significance of estimated parameters are very close between the two estimators; this is also true for the maximum of the log-likelihood function.

Table A.2 compares estimates between the SML trivariate probit built in STATA and the one available in LIMDEP 7.0, which also uses the GHK simulator. It is worth stressing that this packaged estimator wouldn't have been sufficient for the analyses of this chapter, given that the model utilised here allows for sequential nesting via partial observability of two of the variables in the model, while the LIMDEP estimator is a pure multivariate probit, i.e. with full observability of each variable. Again size of coefficients, their standard errors and the value of the maximised likelihood functions are very similar across models. Finally Table A.3 checks the sensitivity of the SML estimator with respect to the number of random draws used to approximate the trivariate integral: as can be seen, there are only minor differences when moving from one choice to another, suggesting that R=75 is sufficient. On the whole, evidence from the three Tables is supportive of the estimator built in STATA.

**Bibliographical references**


Bingley, P. - Bjørn, J. - Westergård Nielsen, N. [1995]: "Determinants of Wage Mobility in Denmark 1980-1990", University of Aarhus, CLS working paper 10-95.

Börsch Supan, A. - Hajivassiliou, V. - Kotlikoff, L. - Morris, J. [1992]: "Health, Children and Elderly Living Arrangements. A Multiperiod - Multinomial Probit Model with Unobserved Heterogeneity and Autocorrelated Errors", in Topics in the Economics of Ageing, D. Wise (ed.), The University of Chicago Press, pp. 79-104.

Börsch Supan, A. - Hajivassiliou, V. [1993]: "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models", Journal of Econometrics, 58 (3), pp.347-368.

Cappellari, L. [1999]: "Low-Wage Mobility in the Italian Labour Market", Warwick Economics Research Paper, no. 531.

Gourieroux, C. - Monfort, A [1996]: "Simulation-Based Econometric Methods", CORE Lectures, Oxford University Press.

Hajivassiliou, V. - Ruud, P. [1994]: "Classical Estimation Methods for LDV Models Using Simulation", in Handbook of Econometrics, vol. IV, Engle and McFadden (eds.), pp. 2383-2441.

Hajivassiliou, V. [1997]: "Some Practical Issue in Maximum Simulated Likelihood", STICERD discussion paper no. EM/97/340.

Hausman, J. - Wise, D. [1979]: "Attrition Bias in Experimental Panel Data: The Gary Income Maintenance Experiment", Econometrica, 47, pp.. 455-473.

Heckman, J.J. [1981a]: "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time - Discrete Data Stochastic Process", in Structural Analysis of Discrete Data with Economic Applications, Manski C.F. and McFadden D. (eds.), Mit Press.

Keane, M - Moffit, R. - Runkle, D. [1988]: "Real Wages over the Business Cycle: Estimating the Impact of Heterogeneity with Micro-Data", Journal of political Economy, 96, pp.1232-1266.

Stern, S. [1997]: "Simulation-Based Estimation", Journal of Economic Literature, vol XXXV, pp 2006-2039.

Stewart, M.B. - Swaffield, J.K. [1999]: "Low Pay Dynamics and Transition Probabilities", Economica, 66, pp. 23-42.

Verbeek, M. - Nijman, T. [1992]: “Incomplete Panels and Selection Bias”, in The Econometrics of Panel Data, Mátyás and Sevestre (eds.), Kluwer Academic Publishers, pp. 262-302.