The persistence of poverty: true state dependence or unobserved heterogeneity? Some evidence from the Italian Survey on Household Income and Wealth

Anna Giraldo, Enrico Rettore and Ugo Trivellato
Dipartimento di Scienze Statistiche, Università di Padova

Via Cesare Battisti, 241 - 35121 Padova - Italy
Ph. ++39/0498274168 - Fax ++39/0498274170 - E-mail: enrico.rettore@stat.unipd.it

Summary
Evidence from several countries is that any household experiencing poverty today is much more likely to experience it again, which may be due to both unobserved heterogeneity (UH) and true state dependence (TSD). We point out that in this context there are two sources of UH: (1) the household ability to obtain income at a specific time period and (2) the way in which this ability evolves from that time period onwards. We test for TSD using a panel from Italy. After testing for the ignorability of the massive attrition plaguing the panel and accepting it, we do not find any sign of TSD.

Keywords: Attrition ignorability, Discrete response panel data models, Poverty dynamics.

JEL-code: I32, C23, C25

1. Introduction

Evidence from several countries is that any household experiencing a poverty spell today is much more likely to experience it again in the future (for comparative cross-country analyses, see Duncan *et al.*, 1993, Oxley, Dang and Antolín, 2000, Mejer and Linden, 2000, and OECD, 2001). Let y_{it}=1 if the *i*-th household disposable income falls below the poverty line at time *t*, and y_{it}=0 otherwise. As an example referring to Italy, using data from the panel component of the Italian Survey on Household Income and Wealth (SHIW), late 1980s/beginning of the 1990s, Trivellato (1998) obtained the following figures:

$\Pr(y_{it}=1 | y_{it-1}=0) \cong .05$

$\Pr(y_{it}=1 | y_{it-1}=1) \cong .50.$

Patently, they document a particularly high degree of persistence of poverty, as measured on income.

There are two logically distinct (albeit possibly concomitant) processes which might generate such a persistence of poverty. It might be that households are heterogeneous with respect to characteristics which are (i) relevant for the chance of falling into poverty and (ii) persistent over time. If this were the case, then a household who is likely to experience poverty at time *t* because of (possibly unobserved) adverse characteristics will also be likely to experience poverty in any other period because of the very same adverse characteristics. We refer to this process as steered by *unobserved heterogeneity* (UH).

On the other hand, it might be that the fact of experiencing poverty in a specific time period causes further poverty in subsequent periods. Since Heckman (1978) such a process is said to exhibit *true state dependence* (TSD).

Distinguishing between the two processes is crucial, since the policy implications are largely different. If the persistence of poverty is (at least partly) due to TSD, then it makes sense forcing households out of poverty at time *t* in order to reduce their chance of experiencing poverty in the future. On the other hand, if the persistence of poverty is due only to UH, any policy aimed at breaking the 'vicious circle' *via* monetary transfers to the poor is pointless: forcing households out of poverty today does not affect their adverse characteristics, hence does not reduce their chance of experiencing poverty spells in subsequent periods[1].

It is worth noting that much of the empirical literature just descriptively juxtaposes the two potential sources of poverty persistence, without trying to ascertain whether, after accounting for UH, there is TSD and without assessing their respective importance. For instance, Oxley, Dang and Antolín (2000, p. 6) summarise the key results of their study across six OECD countries in the following terms: "*(ii) The probability of exiting poverty falls with previous experiences in poverty. At the same time, there is a high probability of falling back into poverty. Thus, for the longer-term poor, low probability of exit and high probability of re-entry tend to reinforce each other. … (iv) The characteristics of*

[1] The argument is relevant for programmes of income support viewed (also) as an active policy, inspired by an efficiency rationale – helping the poor to exit from poverty in order to steadily stay out of it. Of course, it does not call into question the reasonableness of such programmes simply as a passive welfare policy, resting on equity grounds. On the other side, it clearly leaves room for other anti-poverty active programmes, aimed at contrasting the adverse personal and/or household characteristics and the permanent shocks responsible for the chance of persistently experiencing poverty.

households experiencing shorter spells in poverty tend to be different from those of the longer- term poor. A large share of the longer-term poor would appear to be women, lone parents and elderly single individuals. A significant share of the longer-term poor are in paid work."

In this paper we test for TSD while allowing for the presence of UH. We use a panel sample from SHIW, a survey carried out on a two-year basis, over the period 1989-1995. Since Heckman (1978, 1981a), it is well known that panel data allow one to tackle the issue. By studying the pattern of the sequence $\{y_{i1}, y_{i2}, \ldots, y_{iT}\}$ we can identify whether TSD is at work. Recent papers focusing on the issue of UH and TSD in poverty dynamics, and on the related issues of endogeneity of initial conditions and of panel attrition, include Stevens (1999), Devicienti (2000) and Cappellari and Jenkins (2001)[2].

As for the substantive issue of interest, we make clear that to properly test for TSD in poverty/non poverty sequences one needs to account for two sources of UH: (1) the household ability to obtain income at a specific, initial time period and (2) the way in which this ability evolves from that time period onwards. A consequence of such a double source of UH is that simple models for TSD in the presence of UH (*e.g.*, fixed-effect models) might badly miss the point, as shown in section 3. In section 4 we develop a richer model, allowing for a more complex dynamics.

Since the SHIW panel is plagued by massive attrition, preliminarily we develop a test on whether such sample selection is ignorable to the purpose of testing for TSD (section 5). The main results on the model of interest are presented in section 6, and are clearly in favour of a parsimonious specification of the two sources of heterogeneity, with no evidence of TSD.

Final results are in Section 7, and can be summarised in two statements. Firstly, while it is apparent that the panel sample is biased by attrition, with households less likely to experience poverty surviving longer in the sample, we also find clear cut evidence that attrition is ignorable to the specific purpose of testing for TSD. Second, after accounting for the two sources of heterogeneity we do not find any sign of TSD.

2. Testing for TSD in the presence of UH: the textbook model

The textbook model to test for TSD in the presence of UH (see for instance Hsiao, 1986) is the following:

$$y_{it}^{*} = \alpha_i + \rho \; y_{it-1} + \varepsilon_{it}\,, \qquad (1)$$

where:

- y_{it}^{*} is unobservable: instead, it is the binary variable y_{it}, to be observable which is equal to 1 if $y_{it}^{*} < 0$ and zero otherwise[3];
- the model allows for UH through α_i, an unobserved characteristic that makes individuals heterogeneous in a time invariant way: the lower α_i the higher the chance for the *i*-th individual to experience y_{it}=1 in each time period;

[2] Related models have been applied to studies of income mobility. See Cappellari (1999) and Stewart and Swaffield (1999), among others.

[3] With respect to the conventional notation we reverse the inequality to ease the comparison with subsequent models.

- $\{\varepsilon_{it}\}$ is a sequence of serially independent zero mean identically distributed random variables.

The value of ρ determines whether the sequence $\{y_{it}\}$ features TSD. If ρ<0, then experiencing y_{it-1}=1 *causes* an increase in the chance to experience y_{it}=1:

$$\Pr(y_{it}=1|\, y_{it-1}=1, \alpha_i) = \Pr(\varepsilon_{it} < -\alpha_i - \rho) > \Pr(\varepsilon_{it} < -\alpha_i) = \Pr(y_{it}=1|\, y_{it-1}=0, \alpha_i).$$

With reference to this set up, an adequate representation for UH is crucial to properly testing for TSD. A direct check on whether $\Pr(y_{it}=1|\, y_{it-1}=1)$ is larger than $\Pr(y_{it}=1|\, y_{it-1}=0)$ does not provide the required test, since in the presence of UH ($\mathrm{var}\{\alpha_i\}>0$) we are bound to observe $\Pr(y_{it}=1|\, y_{it-1}=1) > \Pr(y_{it}=1|\, y_{it-1}=0)$ even if ρ=0.

Alternative strategies for testing for TSD in the presence of UH (see Arellano and Honoré, 2001 for an up-to-date review) include (i) conditioning on a sufficient statistic for α_i and (ii) imposing some structure on the distribution of α_i.

As for the first strategy, it has been pioneered by Chamberlain (1985). It works in those instances in which, with reference to model (1), a sufficient statistic SS_i, say, exists for the parameter α_i. Exploiting such sufficient statistic, $\Pr(y_{i1}, \ldots, y_{iT}|\, SS_i; \rho, \alpha_i)$ – the probability to observe a specific sequence on the *i*-th unit conditional on SS_i – turns out to be independent of α_i, thus allowing to infer on ρ.

As for the second strategy, by assuming that UH is distributed in a specific way, one can obtain a likelihood function for ρ by integrating out the unobserved α. There is an additional problem here with the initial condition y_{i1}, because it is very often the case that the analyst does not know whether y_{i1} has been generated by the same model as the subsequent observations (see Heckman, 1981b).

3. How does income evolve over time? A flexible specification for UH

In this section we show why the textbook model (1) does not provide an adequate representation of the features of a poverty/non poverty sequence.

Let I_{it} be the *current disposable income* for the *i*-th household at time *t*. Let us represent I_{it} as:

$$I_{it} = I^p_{it-1} + S_{it}, \qquad I^p_{it-1} \perp S_{it}. \tag{2}$$

Here I^p_{it-1} represents the *expected income* for time *t* on the basis of the information available up to time *t*-1. In the absence of any surprise, current income at time *t* would equal I^p_{it-1}. S_{it} represents unexpected (as seen from time *t*-1) departures of current income from I^p_{it-1}. Being a prediction error, S_{it} is orthogonal to the predicted value I^p_{it-1}.

Moreover, let us represent S_{it} as:

$$S_{it} = u_{it} + v_{it}\,, \tag{3}$$

where u_{it} is the permanent component of the shock, which lastingly affects income *from time t onwards*, and v_{it} is the transitory component of the shock, which affects income *only at time t*.

As a consequence, the sequence of expected incomes follows a random walk:

$$I_{it}^{p} = I_{it-1}^{p} + u_{it}\,, \qquad (4)$$

while the sequence of first differences in current income follows a MA(1,1) process:

$$\Delta I_{it} = u_{it} + v_{it} - v_{it-1}\,, \qquad (5)$$

and, consequently, the sequence of current incomes an IMA (1,1) process:

$$I_{it} = I_{it-1} + u_{it} + v_{it} - v_{it-1}\,.$$

As compared to model (1), there are *two sources* of across households heterogeneity here. Households differ with respect to their expected income at time t=1 and they differ also with respect to the way in which the sequence of permanent shocks u_{it} shapes the pattern of expected income from period t=1 onwards.

In this set up, TSD adds a further source of serial dependence:

$$I_{it} = I_{i0}^{p} + u_{i2} + \ldots + u_{it} + \rho\, y_{it-1} + v_{it}\,, \quad t=1,T, \qquad (6)$$

with y_{it}=1 if I_{it} is below the poverty line and y_{it}=0 otherwise.

The qualitative difference made by TSD (ρ<0) is the following. If ρ=0, then:

$$y_{it} \perp v_{is}, \qquad \forall\; s \neq t, \qquad (7)$$

i.e., the transitory shock affects only contemporary income. On the contrary, if ρ<0 then y_{it} is *not* independent of lagged values of the transitory shock v_{it}.

In the following we model the sequence $\{y_{it}\}$ according to (2)-(3), that is to say, maintaining the hypothesis:

$$H_0\!: \rho=0, \qquad (8)$$

and develop simple tests for such hypothesis.

To the purpose of testing the model, we will also largely rely on the following auxiliary assumptions:

- the sequences $\{u_{it}\}$ and $\{v_{it}\}$ are homoskedastic, (9)
- the sequence $\{I_{it}\}$ is jointly Normal. (10)

Assumption (9) is just a convenient assumption to start with, which will be relaxed in the sequel (see sections 4 and 6). Besides, assumption (10) is less restrictive than it looks like *prima facie*[4].

It is worth adding a word of caution on how a rejection of the null hypothesis should be interpreted. In principle, rejecting H_0, *i.e.*, obtaining evidence that y_{it} is *not* independent of lagged values of v_{it}, needs not to be due to TSD, in that serially correlated transitory shocks would also induce a departure from (7). Note, however, that on accepting (7) we would unambiguously conclude against TSD.

---

[4] Let $g(.)$ be any strictly increasing monotonic mapping. Since the inequality $I_{it} < c$ holds if and only if the inequality $g(I_{it}) < g(c)$ holds, y_{it} turns out to be invariant with respect to the choice of $g(.)$. Otherwise stated, the information on income necessary to develop our analysis is defined up to a strictly increasing monotonic mapping. Within the class of strictly increasing monotonic mappings, we are free to choose the particular one fitting our needs better. Then, (10) amounts to assuming that there exists a strictly increasing monotonic mapping $g(.)$ such that $g(I_{it})$ can be represented as in (2)-(3) with $\{I_{it}\}$ jointly Normal.

Also note that equation (6) allows us to assess the consequences of mistakenly testing for TSD in poverty/non poverty sequences within model (1), *i.e.*, omitting the across-household heterogeneity due to the sequence of permanent shocks. Let ρ=0 in (6). To exemplify, consider the couple of observations (y_{i1}, y_{i2}). Despite the absence of TSD, conditional on I_{i0}^{p} they are *not* independent, since they are both affected by the permanent shock u_{i1}. Formally:

Pr(y_{i2} = 1| y_{i1}=1, I_{i0}^{p}) > Pr(y_{i2} = 1| | y_{i1}=0, I_{i0}^{p}).

Since model (1) does not account for this (positive) dependence of y_{i2} on $y_{i1,}$, within it such a dependence is picked up by the TSD parameter. Once again, it looks like TSD but in fact it is only omitted heterogeneity.

4. Testing for TSD in poverty/non poverty sequences

Let c_t be the poverty line according to which we define our sequence {y_{it}}:

y_{it} = I(I_{it}<c_t), t=1,T,

where I(.) is the indicator function equal to 1 if the event within brackets takes place and equal to 0 otherwise.

Under H_0 (*i.e.*, no TSD) and the auxiliary assumptions (9)-(10), the degree of dependence between y_{it} and y_{is} is determined by the Pearson correlation coefficient between I_{it} and I_{is} which according to (6) is given by:

$$\mathrm{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2+(s-1)\sigma_u^2}{\sqrt{(\sigma_I^2+(s-1)\sigma_u^2+1)(\sigma_I^2+(t-1)\sigma_u^2+1)}}, \qquad t>s, \tag{11}$$

$$\sigma_I^2 = \mathrm{var}\{I_1^p\}, \quad \sigma_u^2 = \mathrm{var}\{u_{it}\}^5.$$

Let us have a T-wave panel available. The pattern of dependence in the sequence {y_{it}} is determined by a correlation matrix made up of T(T-1)/2 correlation coefficients defined as in (11), each of them depending on the couple (σ_I^2, σ_u^2), *i.e.*, the amount of UH in the population.

Note that as soon as T>2, model (6) imposes restrictions on the correlation matrix. For instance, with T=4 the pattern of dependence is determined by *six* correlation coefficients, while there are only *two* free parameters in the model.

4.1. A goodness-of-fit test for TSD

Within the model developed under the hypothesis (8) (and the additional assumptions (9)-(10)), a straightforward test for TSD exploits the goodness-of-fit statistic, which contrasts

[5] As usual in binary response models, the variance of the transitory shock is normalised to 1 meaning that we estimate the ratio of the remaining variances to the transitory shock one.

the *free* estimates of the correlation coefficients, $\hat{\rho}_{ts}$, to their estimates implied by the maintained model, $\rho_{ts}\left(\hat{\sigma}_I^2, \hat{\sigma}_u^2\right)$. As for the free estimate of the correlation matrix, it suffices to exploit the joint normality of the sequence of the current disposable income implied by (10). For instance, the probability to observe a sequence of T poverty episodes is the following:

$$\Pr(y_{it}=1,\ t=1,T) = \Pr(I_{it}<c,\ t=1,T) = \Phi(\mu_1, \mu_2, \ldots, \mu_T, R), \tag{12}$$

where $\Phi(.)$ is the T-variate standard normal cdf, $\mu_t = (c - \mathrm{E}\{I_{it}\}/\sqrt{\mathrm{var}\{I_{it}\}})$ and R is the vector obtained by stacking the $T(T-1)/2$ correlation coefficients. Maximising the resulting log-likelihood function provides the required estimates, $\hat{R}$, and an estimate of its covariance matrix, $\hat{\Sigma}_R$ [6].

As for the estimate of the correlation coefficients under the null hypothesis, let $R(\sigma_I^2, \sigma_u^2)$ be the vector obtained by stacking the $T(T-1)/2$ correlation coefficients as functions of (σ_I^2, σ_u^2), according to (11). By minimising:

$$S(\sigma_I^2, \sigma_u^2) = (\hat{R} - R(\sigma_I^2, \sigma_u^2))'\hat{\Sigma}_R^{-1}(\hat{R} - R(\sigma_I^2, \sigma_u^2)) \tag{13}$$

we obtain the minimum chi-squared estimates for (σ_I^2, σ_u^2) (see Amemiya, 1985).

Evaluating function (13) at the minimum provides the required test. If (8) holds (along with (9)), $S(\hat{\sigma}_I^2, \hat{\sigma}_u^2)$ is distributed as a χ^2 with $T(T-1)/2$ -2 degrees of freedom. Large values of the statistic point to a misspecification of the model.

4.2. Refined tests for TSD

The goodness-of-fit statistic $S(\hat{\sigma}_I^2, \hat{\sigma}_u^2)$ is not enough to properly test for TSD, since it is a general purpose one. On the one hand, if it does not reject the model, it might simply miss to detect TSD because of lack of power against *that* specific alternative. On the other hand, if it does reject the model, it needs not be due to TSD. Model rejection might be due to a violation of assumption (9) on the variance of the shocks. This is why we need refined tools.

The second test we consider is specifically targeted to detect departures from (7). As pointed out in section 3, if ρ is not zero in equation (6) then current income depends on lagged values of the transitory shock. Thus, instead of testing for the dependence of y_{it} on y_{it-1} as in (6), we test for its dependence on the *lagged transitory shock* v_{it-1} in the following equation:

[6] Note that to obtain the free estimate of R we do not need to evaluate the T-dimensional integral in (12) (and his analogues corresponding to the other feasible sequences of poverty/non poverty episodes). This is because ρ_{ts} is a feature of the bivariate distribution of (y_{it}, y_{is}) and to estimate it only requires the evaluation of two dimensional integrals. Once R (and $\mu_1, \mu_2, \ldots, \mu_T$) have been estimated, their covariance matrix is evaluated exploiting a straightforward reparameterisation from the T-dimensional 2*2*….*2 contingency table summarising the sample evidence on the poverty sequences to the postulated model.

$I_{it} = I_{i1}^{p} + u_{i2} + \ldots + u_{it} + \lambda\, v_{it-1} + v_{it}, \quad t=1,T.$ (14)

On accepting the hypothesis $\lambda = 0$, we would conclude in favour of the absence of TSD. On the other hand, on rejecting that hypothesis, the question of whether we reject it because of TSD or because of serially correlated transitory shocks would emerge.

The test is again a goodness-of-fit one and develops in strict analogy to what we presented in section 4.1. Moving from (14), the Pearson correlation coefficient between I_{it} and I_{is}, $s<t$, which determines the degree of dependence between y_{it} and y_{is}, becomes:

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2 + (s-1)\sigma_u^2}{\sqrt{(\sigma_I^2 + (s-1)\sigma_u^2 + \lambda^2 + 1)(\sigma_I^2 + (t-1)\sigma_u^2 + \lambda^2 + 1)}} \qquad \text{if } s < t-1 \quad (15.1)$$

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2 + (t-2)\sigma_u^2 + \lambda}{\sqrt{(\sigma_I^2 + (s-1)\sigma_u^2 + \lambda^2 + 1)(\sigma_I^2 + (t-1)\sigma_u^2 + \lambda^2 + 1)}} \qquad \text{if } s = t-1. \quad (15.2)$$

Exploiting (15) and the same free estimates of R as defined above, we obtain a minimum chi-squared estimate of model (14) by minimising:

$$S(\sigma_I^2, \sigma_u^2, \lambda) = (\hat{R} - R(\sigma_I^2, \sigma_u^2, \lambda))'\hat{\Sigma}_R^{-1}(\hat{R} - R(\sigma_I^2, \sigma_u^2, \lambda)), \quad (16)$$

where $R(\sigma_I^2, \sigma_u^2, \lambda)$ are the correlation coefficients defined in (15). $S(\sigma_I^2, \sigma_u^2, \lambda)$ evaluated at the minimum provides the required test. Under the null hypothesis, it is distributed as a χ^2 with $T(T\text{-}1)/2\text{-}3$ degrees of freedom.

Note that for model (14) to impose restrictions on the pattern of the correlation matrix it is crucial to have at least a 4-wave panel available.

The third test we consider is still targeted to detect departures from (7), as the previous one, but it rests on a feature of the correlation matrix whose occurrence does not rely on the homoskedasticity assumption (9). In this sense, this test is more robust than the two previous ones. To exemplify, consider the case T=4. By allowing for heteroskedastic shocks, the correlation coefficient between I_{it} and I_{is}, $s<t$, in the absence of TSD (λ=0 in (14)) is:

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2}{\sqrt{(\sigma_I^2 + 1)(\sigma_I^2 + \sum_{j=2}^{t}\sigma_j^2 + 1)}} \qquad \text{if } t > s = 1 \quad (17.1)$$

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2 + \sum_{j=2}^{s}\sigma_j^2}{\sqrt{(\sigma_I^2 + \sum_{j=2}^{s}\sigma_j^2 + 1)(\sigma_I^2 + \sum_{j=2}^{t}\sigma_j^2 + 1)}} \qquad \text{if } t > s > 1, \quad (17.2)$$

where $\sigma_j^2 = \text{var}\{u_{ij}\}$[7].

It is straightforward to check that the correlation coefficients (17) are such that the following equality holds:

[7] The variance of v_{it} is normalised to 1 (see footnote 5).

$$\frac{\rho_{13}}{\rho_{14}} = \frac{\rho_{23}}{\rho_{24}}. \tag{18}$$

The counterpart of (17) when allowing for the dependence of I_{it} on the lagged transitory shock as in (14) is the following:

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2 + \sum_{j=2}^{t-1} \sigma_j^2 + \lambda}{\sqrt{(\sigma_I^2 + \sum_{j=2}^{t-1} \sigma_j^2 + \lambda^2 + 1)(\sigma_I^2 + \sum_{j=2}^{t} \sigma_j^2 + \lambda^2 + 1)}} \qquad \text{if } t-1 = s > 1 \tag{19.1}$$

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2 + \lambda}{\sqrt{(\sigma_I^2 + \lambda^2 + 1)(\sigma_I^2 + \sigma_2^2 + \lambda^2 + 1)}} \qquad \text{if } t-1 = s = 1 \tag{19.2}$$

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2 + \sum_{j=2}^{s} \sigma_j^2}{\sqrt{(\sigma_I^2 + \sum_{j=2}^{s} \sigma_j^2 + \lambda^2 + 1)(\sigma_I^2 + \sum_{j=2}^{t} \sigma_j^2 + \lambda^2 + 1)}} \qquad \text{if } t-1 > s > 1 \tag{19.3}$$

$$\text{corr}\{I_{it}, I_{is}\} = \frac{\sigma_I^2}{\sqrt{(\sigma_I^2 + \lambda^2 + 1)(\sigma_I^2 + \sum_{j=2}^{t} \sigma_j^2 + \lambda^2 + 1)}} \qquad \text{if } t-1 > s = 1 \tag{19.4}$$

where $\sigma_j^2 = \text{var}\{u_{ij}\}$.

Now, (19) is such that condition (18) *does not* hold. Having the usual free estimate of R available, testing the hypothesis (18) is readily done. Again note that this test requires at least T=4, the same as the previous one.

5. Is attrition in the SHIW panel ignorable to the purpose of testing for TSD?

SHIW is conducted by the Bank of Italy. Starting from 1987 it has been carried out on a two-yearly basis; since 1989 it features a panel component, according to a split panel design[8]. In principle, the four-wave panel over the years 1989, 1991, 1993 and 1995 should allow us to implement our tests for TSD, since it matches the condition required by the more demanding among the tests we developed in the previous section: at least T=4. Unfortunately, our four-wave panel sample suffers from severe attrition (see Table 1). As a consequence of the attrition process (i) its size is rather small (N=827), and what is more troublesome, (ii) it results to be severely biased.

Table 1 about here

To document the bias due to attrition, we consider the six mutually exclusive panel samples originated by the interaction between the SHIW design and the attrition process.

[8] For an accurate description of SHIW, see Brandolini (1999). For an analysis of the attrition process plaguing its panel component over the period 1989-1995, see Giraldo, Rettore and Trivellato (2001).

There are 8,274 households entering the survey in 1989, among which 2,187, 1,050, 827 were still in the panel in 1991, 1993 and 1995 respectively. As a result, three mutually exclusive panels originated out of the 1989 sample: a two-wave panel made up of the households exiting the survey after the 1991 interview; a three-wave panel made up of the households exiting the survey after the 1993 interview; finally, a four-wave panel made up of the households still in the sample at the 1995 interview. By the same token, we get one two-wave panel and one three-wave panel from the pool of households entering the survey in 1991 and one further two-wave panel from the pool of households entering the survey in 1993.

The poverty headcount ratios in the *six panels* over the four years are reported in Table 2. Patently, the number of waves a household stays in the panel is correlated to its probability to experience a poverty spell. In particular, households belonging to the four-wave panel are less likely to experience a poverty spell than households belonging to shorter panels.

Table 2 about here

Note, however, that the apparent bias due to the attrition process is not necessarily a problem to our test for TSD. In fact, to discriminate between alternative explanations of the persistence of poverty it is the *degree of association* between the components of the sequence $\{y_{it}\}$ to matter, not the size of the headcount ratios. If attrition did not bias the estimation of the correlation coefficients, which in our model measure the degree of association between subsequent outcomes of y_{it}, then attrition would be *ignorable* to the purpose of testing for TSD.

To test for the ignorability of the attrition process, we estimate the correlation coefficients exploiting the six panels separately (see Table 3). For instance, $\rho_{89,91}$ is estimated on the two-wave panel 1989-'91, on the three-wave panel 1989-'91-'93 and on the four-wave panel. The three estimates are independent, since the panels we are dealing with are mutually exclusive.

Table 3 about here

The tests for TSD developed in section 4 depend on the data only through the estimate of *R*. Consequently, if we could conclude that the estimation of *R* is not affected by the number of waves sample households survive in the panel, then our tests for TSD would not be affected by attrition[9].

The test for the ignorability of attrition amounts to checking the equality of the correlation coefficients across the six panels. The test statistic is distributed as a χ^2_9 under the null hypothesis. In our case it yields the value 6.8 with an associated *p*-value equal to .66, which neatly points to the ignorability of the attrition process as far as the estimation of *R* is concerned.

On the other hand, the same test applied to the μ's parameters in (12) *strongly rejects* the hypothesis of ignorability: the observed value is equal to 38.7 with an

[9] Of course, within this framework we cannot test whether the six panels are biased with respect to the cross-section sample, since no correlation coefficient can be estimated out of it!

associated *p*-value smaller that .01. Summing up the two results, we conclude that attrition in SHIW affects the *level of poverty*, but it does not affect its *dynamics*.

Having settled the potential bias problem raised by attrition, it is worth noting that by pooling the six panels we end up with 14,975 households-years as compared to the 3,308 households-years in the four-wave panel, gaining a lot in term of standard errors of the estimated correlation coefficients. Table 4 presents the standard errors of the estimated R as they result from using the four-wave panel only and from pooling the six panels, respectively.

Table 4 about here
----

6. Testing TSD: results

In the application, we set the 1989 poverty line for a two-component household at the sample mean of the per capita disposable income – the relative standard followed in Italy to produce official statistics on poverty (see Inquiry Commission on Poverty, 1997)[10]. The 1989 poverty line is then updated to 1991, 1993 and 1995 by using a consumer price index.

To account for differently sized households, disposable income is made comparable by means of the equivalence scale currently used in Italy, which only considers the number of households members[11] (elasticity≅.7) (see Table 5).

Table 5 about here

Moving from the estimate of R obtained by pooling the six SHIW panels, as explained in the previous section, we fitted the simplest model (no TSD, homoskedastic shocks) as it results from (6) – after setting ρ to zero - (9) and (10). By minimising (13) we got $\hat{\sigma}_I^2$ = 5.7359 (.6711) and $\hat{\sigma}_u^2$ = 1.2023 (.3828) (here and in the sequel, standard errors in parentheses). The associated goodness-of-fit statistic (a χ_4^2 under the null hypothesis) is 26.47, *p*-value ≅.0, which *strongly rejects* the model.

To settle the issue we take three complementary routes:

(a) checking whether the rejection is due to the omission of v_{it-1} in (14);
(b) testing for TSD in a manner robust to omitted heteroskedasticity; and, as a final check,
(c) searching for a parsimonious model, able to survive the goodness-of-fit test.

By allowing for the dependence of y_{it} on $v_{it\text{-}1}$ as in (14) and by minimising (16), we got $\hat{\sigma}_I^2$=5.7359 (.6711), $\hat{\sigma}_u^2$=1.2023 (.3828) and $\hat{\lambda}$=.0780 (.1523). The associated goodness-of-fit statistic (a χ_3^2 under the null hypothesis) still rejects the model (26.36, *p*-

[10] Note, however, that current statistics on poverty in Italy rely on a definition of the state of poverty in relation to mean consumption expenditure and are based on Istat (the national statistical institute) data from the annual Household Budget Survey.

[11] The scale is often referred to as the 'Carbonaro scale', from the name of the author: see Inquiry Commission on Poverty (1997), p. 37.

value ≅.0). Given the size of $\hat{\lambda}$ and of the associated t statistic, it is apparent that the rejection of the model is not due to omitted TSD.

This is confirmed by the test robust to omitted heteroskedasticity (a χ_1^2 under H_0), designed to check (18). It yields 1.5991, *p*-value=.2060, which is favourable to the null hypothesis of no TSD.

Finally, to investigate why the model does not fit the data we estimate equation (6) under the null hypothesis ρ=0 but allowing for some heteroschedasticity. Specifically, we allow both the transitory and the permanent shock to feature a time-specific variance in 1993. This is because in 1993 the Italian economy went through a deep recession – indeed, the deepest one since World War II (see Miniaci and Weber, 1999), which might have increased the across-household heterogeneity of the shocks. The resulting goodness-of-fit statistic (a χ_2^2 under the null hypothesis) is as large as 3.16 with an associated *p*-value equal to .20[12] which is much better than before and favourable to the null hypothesis.

It is worth noting that by estimating the textbook model that includes just *one source of heterogeneity* and year-specific dummies to account for macro shocks, namely:

$$I_{it} = I_{i1}^{p} + u_t + \rho y_{it-1} + v_{it}\,{}^{13}, \tag{20}$$

that is to say, a model disregarding the heterogeneity component given by the permanent shocks u_{it}, we would get $\hat{\rho}$ = -.7861 with an associated *t* statistic equal to 3.8, which would lead us to mistakenly conclude for the presence of a sizeable TSD!

7. Concluding remarks

We summarize our results in few statements. As regard attrition in the SHIW panel, we got clear-cut evidence that it affects the usual poverty incidence index – the headcount ratio: the longer the household survives in the panel the lower its probability to experience a poverty spell. On the other hand, attrition does not affect the dynamics of poverty: the length of the panel does not make any significant difference for the degree of dependence between the states in different time periods.

As for the issue of interest, we pointed out that to properly test for the presence of TSD in income based poverty/non poverty sequences it is essential to recognize that there are *two* sources of across households unobserved heterogeneity: (1) their ability to obtain income at a specified time period and (2) the way in which this ability evolves over time. By accounting for such heterogeneity, we do not find *any sign of TSD*. On the other hand, a standard textbook model designed to account for time-invariant unobserved heterogeneity

[12] Estimated parameters are $\hat{\sigma}_I^2$=4.8313 (1.0864), $\hat{\text{var}}\{u_{it}\}$=.2165 (.4165) t≠1993, $\hat{\text{var}}\{u_{i93}\}$=2.0246 (.8778), $\hat{\text{var}}\{v_{i93}\}$=1.1924 (.4340). Apparently, it is the permanent shock in 1993 which features a peculiar variance, much larger than those relative to the other periods. In fact, the model obtained by maintaining the homoschedasticity of the transitory shock fits the data even better (goodness-of-fit statistic=3.2853, *p*-value=.3497).

[13] The initial condition problem is dealt with by specifying a reduced form equation for the first observation, whose disturbance term is allowed to be correlated to the disturbance term in (20) (see Amemiya, 1986, pp. 169-172).

would have mistakenly led us to conclude that poverty sequences are affected by substantial TSD.

References

Amemiya, T. (1985), *Advanced econometrics*, Harvard University Press, Cambridge (MA).

Arellano, M., and B. Honoré (2001): 'Panel data models: some recent developments', in J.J. Heckman and E.E. Leamer (eds.), *Handbook of Econometrics, Vol. V*, North-Holland, Amsterdam (forthcoming).

Banca d'Italia (1997), 'I bilanci delle famiglie italiane nell'anno 1995', a cura di G. D'Alessio, *Supplemento al Bollettino Statistico* (nuova serie), Banca d'Italia, Roma, n. 14, Marzo.

Brandolini, A. (1999), 'The distribution of personal income in post–war Italy: source description, data quality, and the time pattern of income inequality', *Giornale degli Economisti e Annali di Economia*, 58, 183-239.

Cappellari, L. (1999), 'Low-pay transitions and attrition bias in Italy: an analysis using simulation based estimation', Warwick Economics Research Papers No. 532, Coventry, University of Warwick.

Cappellari, L., and S.P. Jenkins (2001), 'Who stays poor? Who becomes poor? Evidence from the British Household Panel Survey', Institut for Social and Economic Research, University of Essex (mimeo).

Chamberlain, G. (1982), 'Panel data', in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics, Vol. II*, North-Holland, Amsterdam.

Chamberlain, G. (1985), 'Heterogeneity, omitted variable bias, and duration dependence', in J. Heckman and B. Singer (eds.), *Longitudinal analysis of labor market data*, Cambridge University Press, Cambridge.

Devicienti, F. (2000), 'Estimating poverty persistence in Britain', LABORatorio R.R. – Centre for Employment Studies, Turin (mimeo).

Duncan, G.J., B. Gustafsson, R. Hauser, G. Schmauss, H. Messinger, R. Muffels, B. Nolan, and J.C. Ray (1993), 'Poverty dynamics in eight countries', *Journal of Population Economics*, 6, 215-234.

Giraldo, A., E. Rettore and U. Trivellato (2001), 'Attrition bias in the Bank of Italy's Survey of Household Income and Wealth', paper presented at the International Conference on Quality in Official Statistics, Stockholm, 14-15 May, 2001.

Heckman, J.J. (1978), 'Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence', *Annales de l'INSEE*, 30-31.

Heckman, J.J. (1981a), 'Statistical models for discrete panel data', in C. Manski and D. McFadden (eds.), *Structural analysis of discrete data with econometric applications*, The MIT Press, Cambridge (MA).

Heckman, J.J. (1981b), 'The incidental parameters problem and the problem of initial conditions in estimating a discrete time – discrete data stochastic process', in C. Manski and D. McFadden (eds.), *Structural analysis of discrete data with econometric applications*, The MIT Press, Cambridge (MA).

Hsiao, C. (1986), *Analysis of panel data*, Cambridge University Press, Cambridge.

Inquiry Commission on Poverty (1997), *Poverty in Italy 1980-1995*, Presidency of the Council of Ministers, Department of Information and Publishing, Rome.

Mejer, L., and G. Linden (2000), ‘Persistent income poverty and social exclusion in the European Union’, *Statistics in Focus*, Population and Social Conditions, Theme 3, 13/2000.

Miniaci R. and G. Weber (1999), ‘The Italian Recession of 1993: Aggregate Implications of Microeconomic Evidence, *Review of Economics and Statistics.*

Oxley, H., T.T. Dang, and P. Antolín (2000), ‘Poverty dynamics in six OECD countries’, *OECD Economic Studies*, No. 30, 2000/I, 7-52.

Stevens, A.H. (1999), ‘Climbing out of poverty, falling back in. Measuring the persistence of poverty over multiple spells’, *Journal of Human Resources*, 34, 557-588.

Stewart, M.B., and J.K. Swaffield (1999), ‘Low pay dinamics and transition probabilities’, *Economica*, 66, No. 261, 23-42.

Trivellato, U. (1998), ‘Il monitoraggio della povertà e della sua dinamica: questioni di misura e evidenze empiriche’, *Statistica*, 58, 549-574.

Table 1: *Sample size of SHIW by year of the first wave and by wave, 1989-1995 (source: Bank of Italy, 1997)*

Year of the first wave	**Survey occasion**			
	1989	**1991**	**1993**	**1995**
1989	8,274	2,187	1,050	827
1991		6,001	2,420	1,752
1993			4,619	1,066
1995				4,490
Overall sample size	8,274	8,188	8,089	8,135

Table 2: *Poverty headcount ratios from SHIW, two- , three- and four-wave panels over the years 1989, 1991, 1993 and 1995 (sample size in parenthesis)*

Panel	**1989**	**1991**	**1993**	**1995**
’89-’91-’93-’95 (827)	8.3	6.7	10.2	8.7
’89-’91-’93 (223)	13.9	12.6	13.5	-
’91-’93-’95 (1752)	-	8.3	15.5	13.5
’89-’91 (1137)	12.3	10.4	-	-
’91-’93 (668)	-	12.0	16.3	-
’93-’95 (1066)	-	-	17.0	14.7

Table 3: *Correlation coefficients as estimated on the SHIW two- , three- and 4-wave panels (sample size in parenthesis)*

	’89-’91 (1137)	**’91-’93** (668)	**’93-’95** (1066)	**’89-’91-’93** (223)	**’91-’93-’95** (1752)	**’89-’91-’93-’95** (827)
$\rho_{89,91}$	.8311			.8494		.7430
$\rho_{89,93}$				.7589		.6390
$\rho_{89,95}$						.5366
$\rho_{91,93}$		.7098		.6361	.7163	.7044
$\rho_{91,95}$					.7085	.6567
$\rho_{93,95}$			.8557		.8483	.8106

Table 4: S*tandard errors of the correlation coefficients as estimated on the SHIW four-wave panel and on the pooled SHIW panels*

	$\rho_{89,91}$	$\rho_{89,93}$	$\rho_{89,95}$	$\rho_{91,93}$	$\rho_{91,95}$	$\rho_{93,95}$
Four-wave panel	.05472	.06210	.07437	.05786	.06529	.04067
Pooled estimates	.02340	.04315	.05935	.02428	.02935	.01518

Table 5: *The equivalence scale*

No. Household's components	1	2	3	4	5	6	7 +
Equivalence coefficient	.599	1	1.335	1.632	1.905	2.150	2.401