Elena Giarda

Persistency of financial distress amongst Italian households: evidence from dynamic probit models

Quaderni di Dipartimento

Serie Ricerche 2010, n. 3
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Persistency of financial distress amongst Italian households: Evidence from dynamic probit models

Elena Giarda*

This version: December 2010

Abstract

This paper analyses financial distress among Italian households using the longitudinal component of the Bank of Italy Survey on Household Income and Wealth (SHIW) for the period 1998-2006. It aims to test whether the probability of experiencing financial difficulties is persistent over time.

First we review the methodologies for estimating dynamic nonlinear panel data models, drawing attention to the problems to be dealt with to obtain consistent estimators. Specific attention is given to the initial condition problem introduced by the presence of the lagged dependent variable in the set of explanatory variables.

Second we provide an in-depth discussion of the alternative approaches proposed in the literature - subjective/qualitative versus quantitative indicators - to identify households in financial distress. We define a quantitative measure of financial distress based on the distribution of net wealth.

Finally we apply dynamic probit models to test for true state dependence in financial distress. The estimation uses four different methods: the pooled probit; the random effects probit with exogenous initial conditions; the Heckman model; and the more recent Wooldridge model. The results of all the estimators confirm the null hypothesis of true state dependence and show that, in line with the literature, less sophisticated models, namely pooled and exogenous models, tend to over-estimate this persistence.

Keywords: household financial distress; dynamic probit models; SHIW.

JEL classification: D14; C23; C35.

---

* Address for correspondence: Elena Giarda, Prometeia Associazione per le Previsioni Econometriche, Via G. Marconi 43, 40122 Bologna, Italy. E-mail: elena.giarda@prometeia.com

This paper is part of the doctoral thesis I discussed in March 2010 at the University of Bologna. I want to thank my advisor, Ignazio Drudi, for valuable suggestions and encouragement, and the commission members, Omar Paccagnella, Alessandro Viviani and Roberto Zelli for helpful comments.

The views expressed are mine and do not necessarily reflect those of Prometeia.

1 Introduction

Many households are likely to experience periods of temporary financial stress over the years, and will overcome them with varying degrees of difficulty. Periods of financial stress become more relevant when the financial difficulties persist over time. This paper focuses on households in financial distress and estimates the relevance of persistence over time of these situations.

In order to select households in financial distress, we need to define the character of their financial situations, and a threshold above which distress is considered to apply. There are two approaches in the literature. The first exploits subjective indicators of financial and economic stress derived from a number of household surveys. These indicators, where available, usually assess indebtedness problems and perceived financial hardship. The second approach exploits quantitative information, such as amounts of debt and wealth collected via survey questionnaires, to build measures of financial distress. Neither method is problem free: the former may be affected by low level of reliability in relation to responses, and corresponding misclassification problems; the latter is at odds with finding and treating quantitative measures to define an appropriate and reasonable criterion on which to split households into those with and those without financial difficulties.

This paper is in line with the second approach and follows the suggestions of Brown and Taylor (2008), which define financial pressure as the difference between total assets – financial and real – and liabilities: households with negative net wealth holdings are classified as being in financial distress. However this setting is not entirely satisfactory because positive amounts of total net wealth may hide a risky situation in the sub-balance of financial assets and liabilities. Therefore in our framework and definition of households in distress we include households with positive (even if small) net wealth holdings.

The methodology used in this paper is the estimation of dynamic nonlinear panel data models, where the coefficient of interest is the coefficient of the lagged dependent variable. Estimation of a dynamic model is aimed at distinguishing between *true state dependence* – the impact of the lagged dependent variable on the dependent variable, and *spurious state dependence* - caused by the presence of time-invariant unobserved heterogeneity. This requires resolution of the so-called *initial conditions problem,* which arises from the fact that the observed start of the stochastic process – period t_0, the first available observation – does not coincide with the true start of the process. It follows that the dependent variable at period t_0 cannot generally be considered to be an exogenous variable that gives rise to the process. We use the Heckman (1981b) as the standard parametric estimator for the probit model. We describe the econometric background to the estimation of panel data dynamic probit models, focusing first on Heckman's seminal work and then on developments proposed in the literature to remove, or make tractable, the computational difficulties of maximising the likelihood function implied by the Heckman method.

The empirical application uses the longitudinal component of the Bank of Italy Survey of Household Income and Wealth (SHIW) for the period 1998-2006, and estimates a range of dynamic probit models to test for the presence of true state dependence in relation to experiencing financial distress.

The remainder of the paper is organised as follows. Section 2 reviews the econometric approaches dealing with the estimation of dynamic, nonlinear, panel data models. It starts with the solution proposed in the seminal work by Heckman (1981a; 1981b) which solves the problem caused by the non-exogeneity of the initial conditions by maximising the likelihood function which allows for cross-correlation between the main and the initial period equations. Its computational cost has led researchers searching for simplified solutions to its implementation, among which Orme (2001) and Arulampalam and Stewart (2009). An alternative method is that proposed by Wooldridge (2005), who suggests an alternative conditional maximum likelihood estimator enabling the estimation of a random effects probit model that includes the explanatory variables as well as the lagged dependent variable, and the initial values and group means of the explanatory variables. The section also reviews the empirical literature, which covers a wide range of research areas, dominated by labour market studies; we focus on a representative selection that emphasises the most recent developments in the Heckman model estimation.

Section 3 reviews the small number of studies dealing with the determinants of household financial distress, and discusses the choice of the dependent variable used and the choice of explanatory variables.

Section 4 presents the criterion established in this paper for defining households in financial distress. This represents a contribution to the literature and is based on net wealth holdings. Households are considered to be under financial stress if they have negative net wealth holdings and if the combination of their real wealth and net financial balance (the difference between financial assets and liabilities) is below a certain threshold. This threshold is defined on the basis of net wealth distribution. This section presents the descriptive statistics for the dependent and exogenous variables commonly used in the literature on household wealth and debt.

Section 5 presents the results of the Heckman model estimations of the longitudinal component of the SHIW over the period 1998-2006. By taking account of both true state dependence and unobserved heterogeneity, the estimation methodology in this paper advances the existing empirical literature that uses Italian data to analyse the financial conditions of Italian households. The empirical analysis tests for true state dependence and identifies the factors that explain the probability of experiencing financial distress. Among these we include household-level variables, such as income, age, occupational status and other indicators of the “ability to pay”, and aggregate-level variables, such as the unemployment rate and house prices. This section also proposes some alternative estimation strategies: the pooled probit, the exogenous initial conditions random effects probit, and the Wooldridge models. Comparisons among these methodologies confirm the results of similar studies, with the exception of the coefficient of the lagged dependent variable obtained using the Wooldridge method, which shows a wider gap with the equivalent Heckman’s coefficient than is reported in the literature. Section 6 presents the conclusions.

All methods accept the null hypothesis that the lagged dependent variable is significantly different from zero, meaning that amongst Italian households the probability of experiencing financial distress is persistent over time and that movement along the net wealth distribution is sluggish. In line with the results of other studies, the state dependence coefficient obtained by assuming exogenous initial conditions is higher than the corresponding coefficient obtained by assuming the existence of some kind correlation between the initial and other values of the dependent variable, such as in the Heckman and Wooldridge methods.

2 The dynamic probit model and estimation methods

To model financial distress we use dynamic panel probit specifications on both unbalanced and balanced samples which include previous states of financial distress. The inclusion of a lagged dependent variable among the covariates allows us to test for the presence of state dependence in the experience of financial distress. One of the main issues in estimating dynamic panel data models consists of solving the initial conditions problem, which arises because the start of the observation period does not coincide with the start of the stochastic process that generates the observations of households in financial distress. To proceed to the estimation we need also to take account of unobserved heterogeneity which causes spurious state dependence.

Our dynamic probit model can be written as:

$$y_{it} = \mathbf{1}[y_{it}^{*} = x_{it}'\beta + \gamma y_{it-1} + \varepsilon_{it} > 0] \qquad i = 1,..., N \qquad t = 1,...,T \tag{1}$$

where y_{it} is the dichotomous dependent variable expressing distress/no distress, **1**(.) is the indicator variable, y_{it}^{*} the latent variable, x_{it} the explanatory variables, y_{it-1} the previous state of the endogenous variable and ε_{it} is the error term. The error term is decomposed as follows:

$$\varepsilon_{it} = \alpha_i + u_{it}$$

where α_i is unobservable individual heterogeneity and $u_{it} \approx N(0,1)$ is the idiosyncratic term. As in any panel data model, assumptions are required about α_i. In a fixed effects specification individual effects α_i are allowed to be correlated with the explanatory variables. This setting does not require specification of a functional distribution of α_i, as they are treated as parameters to be estimated together with the vector θ. However, this approach suffers from the so-called "incidental parameter problem" which, with a fixed T, causes inconsistency in the estimators of θ (Wooldridge, 2005). Honoré (1993) and Honoré and Kyriazidou (2000) suggest semi-parametric models that do not require specification of the distribution of individual effects. However, Wooldridge (2005) remarks that this requires strongly exogenous explanatory variables to resolve the identification problem.[1] For this reason the literature generally assumes a random effects specification of the model. The standard random effects specification assumes $\alpha_i \approx iidN(0,\sigma_\alpha^2)$ and zero correlation between individual effects and the exogenous variables, that is, $corr(\alpha_i, x_{it}) = 0$. Finally zero serial correlation is assumed in the idiosyncratic term u_{it} and, according to the mainstream literature, we assume "equicorrelation" of the composite error term ε_{it}:

$$\rho = corr(\varepsilon_{it}, \varepsilon_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} \qquad t, s = 1,..., T;\ t \neq s$$

In a probit model the conditional distribution of y_{it} is given by (Akay, 2009: 8):

$$f_{it}(y_{it} \mid x_{it}, y_{it-1}, \alpha_i; \Theta) = \Phi\{D_{it}(x_{it}'\beta + \gamma y_{it-1} + \sigma_\alpha \alpha_i)\} \qquad (2)$$

where $D_{it} = (2y_{it} - 1)$ and Φ is the standard normal distribution function. The joint density of y_{it} given ($y_{i0} = y_0$, $x_i = x$, $\alpha_i = \alpha$) is $\prod_{t=1}^{T} f_t(y_t \mid x_t, y_{t-1}, \alpha_i, \theta)$. In a random effects specification – where α is uncorrelated with x – the individual effects α_i follow the probability distribution $g(\alpha_i)$. In this case, the contribution of each individual i to the likelihood is (Verbeek, 2000: 341):

[1] Moreover for neither of their estimators can average partial effects be computed, making it unfeasible to quantify the impact of a change in the explanatory variables on the dependent variable.

$$f(y_{i0},...,y_{iT} \mid x_{i0},...,x_{iT},\Theta) = \int_{-\infty}^{+\infty} f(y_{i1},...,y_{iT} \mid x_{i1},...,x_{iT},\alpha_i,\Theta) f_0(y_{i0} \mid x_{i0},\alpha_i,\Theta) g(\alpha_i) d\alpha_i$$

$$= \int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T} f(y_{it} \mid y_{i0},x_{it},\alpha_i,\Theta) \right] f_0(y_{i0} \mid x_{i0},\alpha_i,\Theta) g(\alpha_i) d\alpha_i \qquad (3)$$

and the corresponding log-likelihood function is:

$$\ln L = \sum_{i=1}^{N} \ln \left[\int_{-\infty}^{+\infty} \left[\prod_{t=1}^{T} f(y_{it} \mid y_{i0},x_{it},\alpha_i,\Theta) \right] f_0(y_{i0} \mid x_{i0},\alpha_i,\Theta) g(\alpha_i) d\alpha_i \right] \qquad (4)$$

We are now at core of a dynamic, nonlinear, panel data model. Inclusion of the previous state to allow for state dependence requires some assumptions about the generation of the initial observations y_{i0}. The estimators proposed in the literature for estimating the lagged-variable coefficient γ differ in terms of how the initial condition problem is dealt with.

The simplest case treats the initial observations as exogenous, that is the distribution of y_{i0} does not depend on α_i. The likelihood function (4), therefore, can be conditioned only on the value y_{i0}, ignoring the term $f_0(y_{i0} \mid x_{i0},\Theta)$. The likelihood function thus consists of two independent terms, one relative to the initial period, the other to subsequent periods. It follows that the joint probability at $t = 1,...,T$ is maximised independent of the probability at time $t = 0$. For more realistic cases of endogenous initial conditions, methods have been proposed to integrate out unobserved heterogeneity from the likelihood function.

2.1 The Heckman model

Heckman (1981b) was the first to take explicit account of the initial conditions problem, assuming endogenous variables with a probability distribution conditional on the exogenous variables and unobserved heterogeneity. Heckman's is a simultaneous two stage approach. The first stage approximates the initial conditions by estimating a reduced form equation in which the explanatory variables are a set of instrumental variables.

Recall eq. (1), our dynamic random effects probit specification (in Heckman's terminology the "structural model"):

$$y_{it} = \mathbf{1}[y_{it}^* = x_{it}'\beta + \gamma y_{it-1} + \varepsilon_{it} > 0] \qquad i = 1,..., N \quad t = 1,...,T \qquad (5)$$

Let the first period equation (the "reduced form equation") be:

$$y_{i0} = \mathbf{1}[y_{i0}^* = z_i'\pi + \varepsilon_{i0} = z_i'\pi + \vartheta\alpha_i + u_{i0} > 0] \qquad (6)$$

where z_i is a vector of the exogenous variables, such as x_{i0},and the additional variables can be regarded as instruments (Akay, 2009; Arulampalam and Stewart, 2009). ε_{i0} is correlated with α_i, but uncorrelated with u_{i0}. u_{i0} is independent of α_i and the distributions are respectively $N(0,1)$ and $N(0, \sigma_\alpha^2)$. A test of $\vartheta = 0$ provides a test for exogeneity of the initial condition.

The conditional distribution of the structural model is the following:

$$f_{it}(y_{it} \mid y_{it-1}, x_{it}, \alpha_i; \Theta) = \Phi\{D_{it}(x_{it}\beta + \gamma y_{it-1} + \sigma_\alpha \alpha_i)\} \tag{7}$$

with $D_{it} = (2y_{it}-1)$ and $\Theta = [\beta, \gamma, \sigma_\alpha]$.

Similarly, the first period conditional distribution can be written as:

$$f_{i0}(y_{i0} \mid z_{i0}, \alpha_i; \pi, \vartheta) = \Phi\{D_{i0}(z_{i0}\pi + \vartheta\sigma_\alpha \alpha_i)\} \tag{8}$$

where $D_{i0} = (2y_{i0}-1)$.

Simultaneous estimation of the parameters of the structural and reduced models (7) and (8) can be achieved by substituting them into the log-likelihood function (4) and without imposing any restrictions (Heckman, 1981b; Hsiao, 2003). In the equi-correlated probit specification, the likelihood function for the individual i is thus:

$$L_i = \int_{-\infty}^{+\infty} \left\{ \Phi[(z_i'\pi + \vartheta\alpha_i)(2y_{i0} - 1)] \prod_{t=1}^{T} \Phi[(x_{it}'\beta + \gamma y_{it-1} + \alpha_i)(2y_{it} - 1)] \right\} g(\alpha_i) d\alpha_i \tag{9}$$

where $g(\alpha)$ is the probability density of unobserved heterogeneity and Φ is the standard normal cumulative function.

The main problem in the Heckman model is the computational burden of maximising the likelihood function, which requires simultaneous estimation of two composite functions.[2] Arulampalam and Stewart (2009) propose a shortcut implementation of Heckman's estimator of the dynamic probit and other nonlinear panel data models using standard software. It involves the creation of a set of T+1 dummy variables, such that $d_{it}^{(\tau)} = 1$ if the observation belongs to the initial period ($t = \tau$), $d_{it}^{(\tau)} = 0$ otherwise ($t \neq \tau$). Under the assumption of equi-correlation, the conditional probability deriving from equations (5) and (6) is:

---

[2] The integral in (8) can be computed by Gauss-Hermite quadrature (Butler and Moffit, 1982), based on approximation of the Gaussian integral $\int_{-\infty}^{+\infty} e^{-v^2} h(v) dv \cong \sum_{m=1}^{M} w_m h(v_m)$, where v_1, v_2,..., v_m are the roots of the Hermite polynomial $H(v)$, M is the number of evaluation points in the approximation process and w_m is the corresponding weight of v_m. For more detail on the formulation of the likelihood function in the Heckman and Wooldridge probit models see the appendix in Akay (2009). An application of the probit model in Stata is developed in Stewart (2006; 2007).

$$\Pr[y_{it} = 1 \mid y_{it-1}, x_{it}, z_i, \alpha_i] =$$

$$= \Phi[\gamma(1-d_{it}^{(0)})y_{it-1} + (1-d_{it}^{(0)})x_{it}'\beta + d_{it}^{(0)}z_i'\pi + (1-d_{it}^{(0)} + \vartheta d_{it}^{(0)})\alpha_i] \qquad (10)$$

Equation (9) is equivalent to a standard random effects specification, where $(1-d_{it}^{(0)} + \vartheta d_{it}^{(0)})\alpha_i$ is unobserved heterogeneity with a heteroskedastic factor loading. The authors suggest estimating this model using the routine "gllamm" in Stata, that allows for this form of heteroskedasticity.[3]

2.2 The Orme model

Orme (2001) suggests a two-step procedure to address the initial condition problem that is locally valid when the correlation between y_{i0} and y_{it} (ρ) tends to zero. Orme uses an approximation to substitute α_i with another unobservable component that is uncorrelated with the initial observation. By assuming bivariate normality of the composite error term ε_{i0} and unobserved heterogeneity α_i, that is $(\varepsilon_{i0}, \alpha_i) \approx BVN(0,0,\sigma_\varepsilon, \sigma_\alpha, \rho)$, individual effects can be defined as:

$$\alpha_i = \rho\frac{\sigma_\alpha}{\sigma_\varepsilon}\varepsilon_{i0} + \sigma_\alpha\sqrt{(1-\rho^2)}w_i$$

with $w_i \approx N(0,1)$ and orthogonal to ε_{i0} by construction and distributed as $N(0,1)$. The structural model thus becomes:

$$y_{it}^* = \gamma y_{it-1} + x_i'\beta + \left[\rho\frac{\sigma_\alpha}{\sigma_\varepsilon}\varepsilon_{i0} + \sigma_\alpha\sqrt{(1-\rho^2)}w_i\right] + u_{it} \qquad (11)$$

which encompasses two time-invariant components of unobserved heterogeneity, ε_{i0} and w_i. Orme suggests estimating the first period equation (6) to compute its generalised residual:

$$e_i \equiv E(\varepsilon_{i0} \mid y_{i0}) = (2y_{i0}-1)\sigma_\varepsilon\varphi(\pi^{*\prime}z_i / \sigma_\varepsilon)/\Phi[(2y_{i0}-1)\pi^{*\prime}z_i / \sigma_\varepsilon] \qquad (12)$$

and use it as an explanatory variable in (11), that is $\varepsilon_{i0} \equiv e_i$. Under the given distributional assumptions, Orme shows that this method approximates Heckman's solution as ρ approximates zero and can perform well also when ρ is not small.

[3] The routine *gllamm* is usually applied to multilevel models, but can be applied to panel data models as well. Longitudinal data are two-dimensional, with a cross-section and a temporal dimension. In a random effects specification (Pudney, 2008: 23), a longitudinal dataset "is a special case of the multilevel structure, with time observations (level 1) clustered within individuals (level 2)."

2.3 The Wooldridge model

Wooldridge (2005) proposes a conditional maximum likelihood estimator as an alternative to the Heckman model, suggesting that the distribution of unobserved heterogeneity should be modelled conditional on the initial value and any exogenous explanatory variables, in order to integrate out individual effects α_i . Attention is directed away from joint density $f(y_{i0},...,y_{iT} \mid x_i)$ in Heckman's approach and towards conditional density $f(y_{i1},...,y_{iT} \mid y_{i0},x_i)$. The contribution of each individual *i* to the likelihood is thus:

$$f(y_1,...,y_T \mid y_0) = \int_{-\infty}^{+\infty} f(y_1,...,y_T \mid y_0, x, \alpha) h(\alpha \mid y_0) d\alpha \tag{13}$$

where $h(\alpha \mid y_0)$ is the density of α conditional on initial observation y_0. While Heckman requires approximation of the joint density of y_0 and α, Wooldridge requires only an approximation of conditional density $h(\alpha \mid y_0)$. Moreover, as noted in Arulampalam and Stewart (2009: 666), while Wooldridge requires normality for the conditional distribution $\alpha_i \mid y_{i0}$, Heckman requires bivariate normality for the joint distribution (u_{i0}, α_i).

Wooldridge specifies a correlated random effects model in line with the Mundlak-Chamberlain approach (Mundlak, 1978; Chamberlain, 1984), which relaxes zero-correlation of the random effects model by assuming the following specification of unobserved individual effects:

$$\alpha_i = a_0 + \bar{x}_i' \xi + a_i \tag{14}$$

where $a_i \approx iidN(0, \sigma_a^2)$ and is independent of x_{it} and u_{it} for each *i* and *t*, and where $\bar{x}_i$ are the means over time (group-means) of the explanatory variables[4]. Wooldridge suggests a different specification for the individual effects α_i , which includes the initial values of the endogenous variable y_{i0}, in addition to the group means of the explanatory variables $\bar{x}_i$. Unobserved heterogeneity is thus modelled as:

$$\alpha_i = \xi_0 + \xi_1 y_{i0} + \bar{x}_i \xi + a_i \tag{15}$$

The dynamic correlated random effects probit model can be written as:

$$y_{it} = \mathbf{1}(x_{it}' \beta + \gamma y_{it-1} + \xi_0 + \xi_1 y_{i0} + \bar{x}_i' \xi + a_i + u_{it} > 0) \tag{16}$$

[4] According to this formulation the random effects model can be renamed the “correlated random effects model”.

where a_i is the "new" unobserved heterogeneity and u_{it} is the idiosyncratic term. The model assumes the distribution of α given y_{i0} and x_i be $\alpha_i \mid y_{i0}, x_i \approx N(\xi_0 + \xi_1 y_{i0} + \bar{x}_i \xi, \sigma_a^2)$ and the explanatory variables $z_{it} \equiv (x_{it}, y_{it-1}, y_{i0}, \bar{x}_i)$.

It follows that the likelihood function of individual i is specified as:

$$L_i = \int_{-\infty}^{+\infty} \left\{ \prod_{t=1}^{T} \Phi\left[(x_{it}'\beta + \gamma y_{it-1} + \xi_1 y_{i0} + \bar{x}_i \xi + a_i)(2y_{it} - 1)\right] \right\} g^*(a_i) da_i \quad (17)$$

where $g^*(a_i)$ is the normal density of the "new" unobserved heterogeneity a_i in equation (15). The likelihood function (17) is equivalent to the likelihood function of a static random effects probit model where the explanatory variables are $z_{it} \equiv (x_{it}, y_{it-1}, y_{i0}, \bar{x}_i)$ and the maximum likelihood estimator can be obtained via standard random effects probit estimation.

2.4 Related literature

The application of the methodology proposed by Heckman (1981a; 1981b) is infrequent, due to its computational complexity. The empirical literature evolved towards computationally less demanding solutions, e.g. Orme (2001), or towards simulated ML methods such as Hyslop (1999). Orme (2001) suggests a first computational simplification of the Heckman estimator, defined as a "two-step pseudo-ML estimator", which has been widely utilised in subsequent applications. Examples are the papers by Arulampalam et al. (2000) on unemployment dynamics, Henley (2004) on self-employment dynamics and Requena-Silvente (2005) on small and medium enterprises in the UK. Other examples in the area of welfare and social benefits are the papers by Chen and Enstrom-Host (2005) and Andrèn (2007) for Sweden, by Lee and Oguzoglu (2007) for Australia and by Cappellari and Jenkins (2009) for the UK. Propper (2000) analyses demand for private healthcare in the UK. May and Tudela (2005) estimate a dynamic probit model that accounts for the correlation between individual effects in the initial condition equation and in the structural equation, and for serial correlation in the error term (more details of this application are discussed in Section 3).

Following the circulation of a working paper by Wooldridge (2002b) and its publication (Wooldridge, 2005), the implementation of dynamic nonlinear models has become widely applicable. The author proposes a conditional ML estimator, "finding the distribution conditional on the initial value and the observed history of strictly exogenous explanatory variables", rather than attempting "to obtain the joint distribution of all outcomes of the endogenous variables" (Wooldridge, 2005: 39). The estimator is implementable using standard software for random effects probit models. Contoyannis et al. (2004) apply the Wooldridge method to a dynamic ordered probit on health status self-assessment and find strong, true state dependence. They estimate the model also on an unbalanced panel dataset and, after accounting for attrition through the inverse probability weighted estimator (Wooldridge, 2002a), show that attrition does not cause bias in the estimates.

Alessie et al. (2004) and Clark and Etilè (2006) move away from the univariate context to tackle bivariate models. The former, of specific interest in our research area, applies the Heckman model to the interaction between mutual funds and stocks. The main results are a positive correlation between ownership of one type of asset in one period, and ownership of the other in the subsequent period explained by correlated unobserved heterogeneity, and negative state dependence of lagged ownership of stocks on ownership of mutual funds. Clark and Etilé apply the Wooldridge method to examine interactions between spouses in terms of cigarette smoking (a univariate application relating to the smoking behaviour of single mothers can be found in Dorsett, 1999). Arulampalam and Bhalotra (2006) implement the Heckman methodology to test state dependence in infant mortality in India via a logit model. Benito and Young (2003) and more recently Loudermilk (2007), analyse firms' dividends, the former via the Heckman model on a Tobit specification, the latter via Wooldridge's model on a probit specification. At the macroeconomic level, Chauvin and Kraay (2007) apply the Wooldridge method to the probability that a low-income country will receive debt relief if it has been a recipient of it in the past.

Stewart (2007) started a line of research on comparisons among methods. Much of the evidence in the literature indicates that the Heckman, Orme, and Wooldridge methods produce comparable results. Stewart (2007) tests for true state dependence in unemployment and the role played by spells of low-wage employment, by presenting and comparing the estimates from the Heckman and the Wooldridge methods to assess the robustness of results.[5] Both methods produce similar results. Similarly, Sousounis (2008) finds equivalent results when applying the Heckman, Wooldridge and Orme methods to study state dependence in participation in work-related training programmes. In 2008, Arulampalam and Stewart circulated a working paper that was published in 2009, in which they provided a simplified implementation of the Heckman method, using established routines in statistical software such as Stata and Limdep.[6] They study the unemployment dynamics of male workers in the UK and compare the results for a range of estimators: exogenous initial conditions, Heckman, Orme, and Wooldridge. Akay (2009) studies the dynamics of the female labour market in Sweden by implementing a probit model on the probability of participation and a Tobit model on the hours worked following the Heckman, Orme and Wooldridge methodologies. Arulampalam and Stewart (2009) and Akay (2009)

[5] Stewart (2006) implements the Stata program "redprob" to estimate a dynamic probit model using the Heckman approach.

[6] E.g., using Stata the Heckman model can be estimated with the "gllamm" procedure (Rabe-Hesketh and Skrondal, 2005; Grilli and Rampichini, 2005), although its implementation is not straightforward.

conducted Monte Carlo experiments to asses the performance of the various methodologies on finite samples. The results of the simulations show that when one or both longitudinal dimensions (T and N) are relatively large, $T \geq 6$ and $N \geq 800$, the bias is relatively small for all three estimators, whereas for smaller sample sizes, the bias increases although none of the estimators dominates (Arulampalam and Stewart, 2009). According to Akay (2009), the Wooldridge method performs well for panels longer than five periods, and less well for shorter panels where the Heckman method is preferred. For lengths of 10-15 periods the three estimators produce equivalent results, with the bias diminishing with increasing lengths.

The shortcut suggested by Arulampalam and Stewart (2009) is applied by Narazani (2009) to a bivariate probit model on the interrelationship between the employment and capital adjustment decisions of Italian firms, using the routine “gllamm” in Stata.

An interesting alternative approach is that developed by Pudney (2008) to model the dynamics of individuals’ subjective assessments of their financial wellbeing, in a short panel. The originality of Pudney’s approach is in shifting the emphasis from the observed lagged dependent variable y_{it-1} to its latent counterpart y^*_{it-1}. He argues that state dependence models (SD) were developed primarily to explain labour market dynamics, where $y_{it} = 0$ and $y_{it} = 1$ indicate employment and unemployment at time t respectively. In this context the nature of the data is intrinsically discrete and the latent variable y^*_{it} represents an artificial construct; Pudney therefore sees no reason why the lagged latent variable y^*_{it-1} should appear among the covariates of the model. He argues that the concepts of wellbeing and living conditions are not inherently discrete and that the “true” behaviour is represented by the latent variable y^*_{it}. Consequently, in these cases, y^*_{it-1}, rather than y_{it-1}, should incorporate the feedback effect on the variable at time t. Pudney’s model can be defined therefore as *latent autoregressive* (LAR) and its estimation is carried out via simulated maximum likelihood maximisation in GAUSS. Compared to the SD models, the LAR model shows quite different dynamic properties which translate into higher state dependence.

3 Empirical literature on households’ financial conditions

The strand of the literature focusing on the analysis of subjective measures of financial distress relates mainly to questions about debt burdens, and exploits information contained in the British Household Panel Survey (BHPS). In what is perhaps the first work on households’ financial difficulties, Boheim and Taylor (2000) assess the incidence of housing finance problems by building a dichotomic variable that takes the value 1 if the interviewee answers “yes” to at least one of the following questions: “Did you have problems paying for your housing over the last 12 months?”, “Over the last 12 months were you ever 2 months or more behind with your rent/mortgage payments?", and "Did you have to borrow to pay the rent/mortgage?". The variable is zero otherwise. Boheim and Taylor estimate a dynamic probit model on an unbalanced panel dataset assuming exogeneity of initial conditions. The explanatory variables relate to the socio-economic characteristics of households and household-heads (income, equity value, mortgage value), and the aggregate variables (regional unemployment rate and interest rates). They include a variable for "financial surprise" following a suggestion made by Boheim and Ermish (2001) in a different context.

May and Tudela (2005), based on a different time span, exploit the answers to the first of the questions reported above to estimate a dynamic probit model, following the Orme (2001) methodology. Among the explanatory variables, in addition to the past value of the dependent variable, they include three dummies for the loan-to-value ratio, two dummies for the cost of servicing mortgage debt and its relative incidence on income, dummies for whether the household has any savings, whether the household head has moved into unemployment or has any health-related problems, and a set of regional dummies. Due to the nature of the dependent variable (housing-related payment problems over the previous 12 months) they lag all variables by one period to identify individual characteristics before the household experienced difficulties. Macroeconomic conditions are accounted for by introducing house prices growth rates at the regional level, the regional unemployment rate and effective mortgage interest rates, only this last is statistically significant. Amongst the instruments of the initial conditions equation, the authors include a dummy for house purchase before 1989, dummies for negative equity value and socio-economic characteristics such as sex, job qualification, ethnicity, number of dependents (according to May and Tudela (2005: 26) "these variables were at some stage included in the main regression but were dropped because they were not significant"). After controlling for unobserved heterogeneity and autocorrelated errors,[7] there is evidence of persistence in mortgage payment problems: 34 per cent of total variance is explained by unobserved heterogeneity compared to Boheim and Taylor's (2000) finding of 19 percent. A general result of these models is that, amongst British households, the probability of experiencing financial problems is persistent over time.

With a static ordered probit model based on the 1995 and 2000 waves of the BHPS, del Rio and Young (2008) estimate that the determinants of unsecured debt (consisting of overdrafts, credit card debt and personal loans) are the unsecured debt-income ratio, the mortgage income gearing, financial wealth, health, ethnicity and marital status. The probability of reporting a high debt burden increases for high debt-income households who have also experienced an adverse financial surprise.

[7] However, a likelihood ratio test indicated that autocorrelation is not statistically significant.

Pudney (2008) introduces a dynamic autoregressive latent model to estimate an ordered probit where the dependent variable is based on the responses to the question: “How well would you say you are managing financially these days?” Of interest is the higher persistence found using the latent model compared to the state dependence model, implying a longer duration of the adjustment process.

For Italy, only Boeri and Brandolini (2005) have tackled the issue of perceived financial distress and discontent. They research the factors underlying dissatisfaction in Italian households, by looking at the “horizontal distribution” of income among socio-demographic groups. Following an indication obtained from the European Community Household Panel (ECHP), according to which the perceived ability of Italian households to make ends meet deteriorated between 1996 and 2001, they exploit information from the European Commission Business and Consumer (BSC) survey, Eurobarometer and the SHIW to study trends in income growth and poverty measures in Italy. Eurobarometer and the BSC provide similar indications to the ECHP, with households reporting a more acute deterioration in their financial situation between 2000 and 2002. The question is whether this evidence is a result of worse overall economic conditions, and especially less equal income distribution; however, it is not possible to verify this because there is counterintuitive evidence of decreasing inequality from 1993 to 2002. The authors note that this approach focuses on the “vertical” distribution between rich and poor, in which inequality indices discriminate among households only in terms of income levels. They therefore analyse the “horizontal” allocation of income across socio-demographic groups. This reveals important changes in the income distribution among groups defined by occupational status of the household head. These changes also have an impact on group-specific poverty ratios and Gini indices.

Work on financial hardship using quantitative indicators is very limited, and is mostly descriptive rather than econometric analysis. Cox et al. (2002) and May et al. (2004) suggest a series of indicators to identify households suffering financial distress, which include flow and stock variables. They propose monthly income, savings (income minus expenditure) and the income gearing ratio as flow variables, and unsecured debt and mortgage commitment to income ratios as stock variables. The ratio between total assets (real and financial) and liabilities are other useful indicators. Barwell et al. (2006) address the issue more explicitly by suggesting analysis of the net worth distribution, emphasizing that to examine households’ financial conditions in more depth requires more than analysis of liabilities and also requires account to be taken of the levels and composition of assets.

To our knowledge the only econometric analysis using quantitative indicators to identify households in financial distress is by Brown and Taylor (2008) and uses a single cross section probit model for three countries, Great Britain (GB), Germany, and the United States

(US). It defines households in distress as those with negative wealth holdings. The model uses the standard microeconomic variables (age, education, income, households characteristics, etc.). The probability of negative net worth decreases monotonically with age for all three countries, more so in GB and the US than Germany, and decreases with income in all countries. Only in GB does the probability of negative wealth decrease with education. Following the suggestions in Cox et al. (2002) to look at some additional quantitative indicators of financial pressure, Brown and Taylor estimate a series of ordinary least squares (OLS) regressions in which the dependent variable in turn is the debt/income ratio, the savings/income ratio, and the cost of servicing debt to income. In GB and the US, unlike Germany, the unsecured debt to income ratio is higher for the younger age classes than for the older ones. In the two former countries the difference between the effects of being in the bottom income quartile and being in the top income quartile is not statistically significant. In the opinion of the authors, these two results are of concern as younger families might be unable to respond to adverse economic shocks.

This review of the literature seems from a methodological point of view to indicate that: (a) the recent innovations by Orme, Wooldridge and Arulampalam and Stewart have made the estimation of nonlinear dynamic panel data models more feasible; (b) the various estimation methods produce similar results when tested with Monte Carlo simulations and applied to real data, with values of the state dependence coefficient lower than the exogenous case. In terms of the definition of the dependent variable measuring financial distress, most of the literature is based on qualitative indicators related to perceived difficulties, one exception being the study by Brown and Taylor (2008).

4 The data and variables

To test for true state dependence of financial distress in Italian households, we use the Bank of Italy SHIW for the period 1998-2006, a total of six waves.

The survey collects detailed data on demographics, household consumption, income and balance sheet items. The first survey was in the mid-1960s and over time sample size and design, sampling methodology and questionnaire structure have evolved: consistent information over time is available from 1989. The survey is biannual, with the exception of a three-year gap between 1995 and 1998, and the number of households interviewed in each wave is around 8,000, providing a representative sample of the Italian resident population. Sampling is in two stages: municipalities in the first stage and households in the second stage. Municipalities are divided into 15 strata defined by 17 regions and 3 classes of population (more than 40,000, 20,000 to 40,000, and less than 20,000 inhabitants). Households are randomly selected from registry office archives. The net response rate in

1998 was 43.9 percent (56.9 percent in 1995) and in 2006 was 42.0 percent, considerably lower than in 1995, but at least an increase on the 2004 low of 36.4 percent.[8]

The survey also has a panel component, which we exploit in our empirical application.[9] We built two different longitudinal sub-samples. Sample (A) is an unbalanced panel with common entrance in 1998, and exits after at least three consecutive periods, with length $T \geq 3$ and a total number of observations of 8,619, of which 1,911 are observations relative to 637 households remaining in the sample for at least 3 waves, 1,328 observations of 332 households remaining in the sample for at least 4 periods, and 5,380 observations of 1,076 households in the sample for 5 periods. Sample (B) is a balanced panel, whose length is $T = 5$, with entries in 1998 and exits in 2006, covering 5,380 observations and 1,076 households.

4.1 Choice of the dependent variable

The model to be estimated assumes a relationship between the probability of a household experiencing financial distress and a set of variables for economic, demographic, social and macroeconomic factors. The model postulates the inclusion of time-invariant individual effects and past values of the dependent variable. The main issue is the choice or construction of the variable to define a household in financial difficulty and definition of the set of variables that may affect or determine the state of distress. We chose a quantitative indicator. We have highlighted that much of the literature deals with models where the variable of interest is qualitative and, in most cases, is derived from responses to questions about perceived hardship (debt burden, ability to make ends meet, etc.). In terms of the Italian SHIW dataset, only since 2002 to the time of writing this draft paper in winter 2009-2010, does the questionnaire ask about self-reported financial hardship, and only in three waves (2002, 2004 and 2006). This information is based on the question: “Does your household income allow your family to make ends meet?” (variable CONDGEN[10]). The use of this variable allows us to extend the estimation of the analogous models employed in other countries to the Italian case. However, the panel length is very short (only three periods) which makes estimation of a dynamic model unrealistic. For purely investigatory purposes, a dynamic random effects probit model was estimated on the balanced panel with T = 3 and

[8] See Brandolini and Cannari (1994) and Faiella (2008) for a detailed description of sampling method, attrition and other measurement issues.

[9] The SHIW longitudinal component required two data corrections in order to achieve information that is consistent over time. The first relates to years of birth, and sex within households, where there was some incoherence due to changes in household composition (e.g. the household split. or the head of household left). Variations of this kind apply to 5 percent of households in the entire sample: we decided to split these households according to change in head of household. The second relates to discontinuity in presence within the sample. This was very rare: 0.06 percent, but, we decided, anyway to drop them from the sample.

[10] Responses are: 1. finding it very difficult, 2. finding it difficult, 3. finding it quite difficult, 4. fairly easily, 5. easily, 6. very easily.

the answers recoded as y = 1 when CONDGEN equals 1, 2 or 3, and y = 0 when CONDGEN equals 4, 5 or 6. The explanatory variables are used for the estimation of the dynamic model in Section 5 (lagged dependent variable, age dummies, income quartiles, education levels, household composition, ownership of risky portfolio, homeownership, being indebted, regional house prices and unemployment rate). The likelihood ratio test confirms the null hypothesis of the absence of unobserved individual effects, indicating equivalence between the random effects and the pooled model, possibly to the low dimension of *T*, which reduces from 3 to 2 periods for the presence of the lagged dependent variable. The limited availability of an appropriate qualitative dependent variable is one of the reasons why this choice suggests the need for a quantitative indicator.

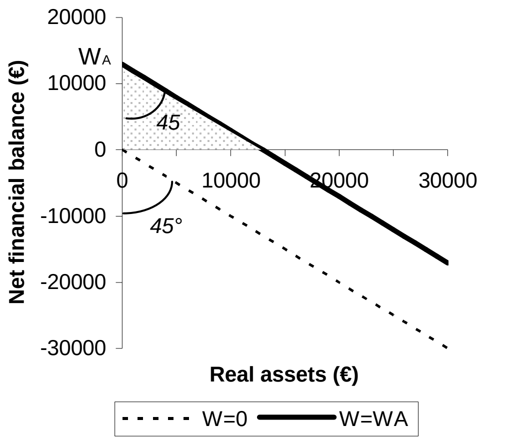
Another reason is related to some remarks in the household debt-related literature about whether outstanding debt, particularly in countries such as the US and the UK, is considered excessive, where "excessive" means carrying the risk of default or financial hardship in the event that the household is exposed to unexpected adverse shocks. The literature highlights that the riskiness associated to debt holdings increases with the income gearing ratio, but is softened by the coexistence of relatively liquid real or financial assets in the household portfolio. It follows that quantitative indicators of financial distress derive from a combination of both factors, assets and liabilities. It should be stressed that, in terms of the income gearing ratio, unlike in the case of surveys of other countries such as the UK, the SHIW does not provide very reliable information because of the very high number of missing values. In terms of the coexistence of debt and real and financial assets, household net worth can be defined as the difference between assets and liabilities. If a household taking out a loan has some financial assets which are either of no or lower value in absolute terms than the outstanding debt, then their net financial balance will be negative; if the real wealth value is smaller (in absolute terms) than the negative net financial balance, then the household owns "negative net worth"; "null net worth" corresponds to a situation where real assets values equal the negative net financial balance, in absolute terms. Null net worth can be determined by any value of debt and real assets, since it simply requires two factors to be cancelled out. Fig. 1 depicts the possible combinations of real wealth holdings (positive x-axis) and financial balance (y-axis). The negative y-axis describes a situation of negative net balance and null real wealth; the 45° line splitting the lower-right quadrant describes a situation on null wealth (W = 0); the area below the W=0 line encompasses all cases of negative net worth.

Two issues arise when we try to define households in "financial distress". The first is whether holding negative net wealth is a sufficient condition to identify a situation of financial stress. The second is whether positive or null net wealth values can be associated with financial vulnerability. In terms of Fig. 1 we can identify financially vulnerable households by:

(a) a positive answer to the first question, thus including all individuals on the negative y-axis and in the area underneath the 45° line;

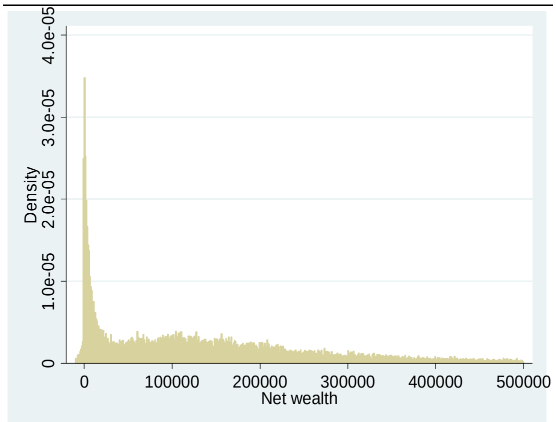
(b) a positive answer to the second question. This implies also including individuals with a "small" amount of positive net worth, given by the algebraic sum of positive net worth and negative financial balance, or by the sum of small values of real wealth and small values of positive financial balance.

Fig. 1 Households in financial distress (area below the line $W = W_A$)



For option (a), it is clear that negative net wealth holdings define a situation of financial fragility. However, it requires some thought about whether a "small" entity of negative worth can determine a critical situation, without examining the associated real wealth value. It could be assumed that small amounts of negative worth are not critical if associated with high values of real wealth. Equally, individuals with moderate amounts of real wealth, but above the absolute value of the net financial balance and therefore with positive net worth, could be assumed to be in economic distress. These observations suggest a choice amongst a range of solutions. Here we consider two. The first one, defined as option (a1), identifies financial distress with the area below a parallel line to W = 0 and shifted upwards by a certain amount, so that it intersects the positive y-axis at the level W_A as depicted in Fig. 1 (line $W = W_A$). The second, defined as option (a2), is similar to the first option, but with an increase, in absolute terms, of the slope of the new line in the second quadrant in order to reduce the risk of including among those in difficulty, households with very high real net wealth holdings.

Fig. 2 Net wealth distribution



For option (b), individuals in financial distress are those with combinations of real wealth and net financial balance lying in the triangle defined by the line $W = W_A$ and the x and y axes (dotted area in Fig. 1).

We choose to follow the criteria defined by options (a1) and (b) to define households in financial distress. The choice to include households with positive net wealth can also be justified by the fact that, according to the 2006

SHIW data, only 3 percent of Italian households have negative net wealth, a largely lower percentage than is observed in countries such as the US, the UK and France (Sierminska et al., 2008).

The choice of the threshold W_A was made with reference to the net wealth distribution from pooling the five waves of the SHIW. The distribution is depicted in Fig. 2: it is highly concentrated, with the majority of households owning low or null wealth. After careful consideration we decided to define the threshold as the level of wealth corresponding to the second decile of the distribution: this value is 13,000 euro (at 2006 prices). Households in financial distress, therefore, are defined as those whose net worth is equal to or below 13,000 euro. Fig. 3 is a scatter plot of real wealth and net financial balance using the SHIW data, corresponding to the real-data representation in Fig. 1. The distribution of net wealth of financially vulnerable households is depicted in Fig. 4.

Fig. 3 **Households in financial distress**

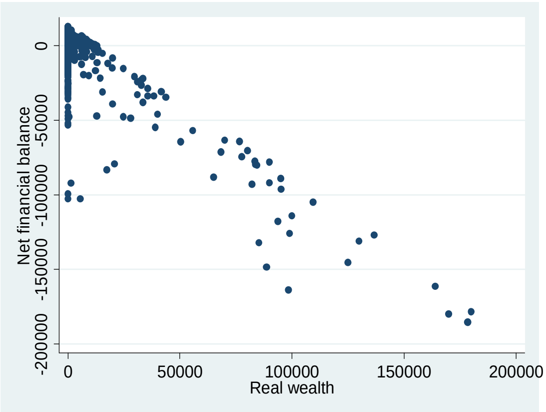
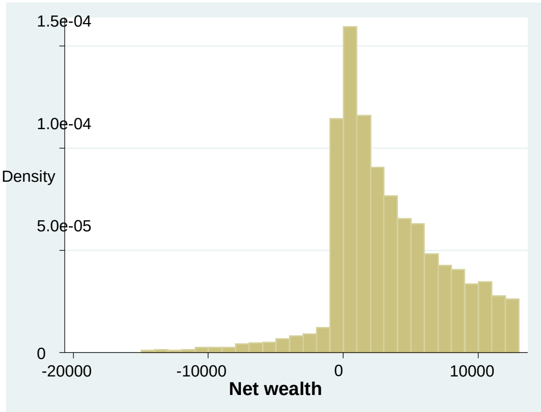


Fig. 4 **Net wealth distribution for values below threshold W_A**



Tab. 1 reports the percentages for observations in financial distress, in relation to two possible panel data structures of the data. Both cases show very similar percentages of observations with net worth lower than the threshold, 15.3 percent in the unbalanced case and 14.9 percent in the balanced one.

Tab. 1 **Distribution of the response variable** (percentages)

	(A) Panel T>=3 (1998)			**(B) Balanced panel T=5**		
	0	1	Tot.	0	1	Tot.
1998	82.8	17.2	100	83.6	16.4	100
2000	84.0	16.0	100	84.0	16.0	100
2002	84.9	15.1	100	84.7	15.3	100
2004	86.9	13.1	100	86.7	13.3	100
2006	86.7	13.3	100	86.7	13.3	100
Tot.	84.7	15.3	100	85.1	14.9	100

4.2 *Choice of explanatory variables*

The choice of the model covariates is based on the reduced form models for the determinants of household debt and financial assets (see, amongst others, Duca and Rosenthal,1993; Cox and Jappelli, 1993; Crook, 2001; Magri, 2007; Crook and Hochguertel, 2007) and the literature reviewed in Section 3 on household financial distress (e.g. May and Tudela, 2005; Brown and Taylor, 2008; del Rio and Young, 2008). The set of explanatory variables includes past values of the dependent variable, age, income, education levels, gender, household composition, ownership of risky assets, homeownership and indebtedness. Aggregate trends for the economy are captured by macro-area unemployment rates and the regional house price index. In what follows we justify our variables selection and provide descriptive statistics for them, and describe how they relate to the endogenous variable.[11]

In the literature, one of the most relevant factors is the so-called "ability to pay", indicated mainly by income: the probability of experiencing financial fragility will be an inverse function of the income level. A low income level, if persistent over time, generates null or limited savings, and likely induces indebtedness to sustain household consumption. Low levels of savings are nearly always a sufficient condition for low wealth levels. It is reasonable to associate low income with small, null or negative net worth and, therefore, with a high probability of financial distress. Tab. 2 reports average values in terms of real income, debt, and real and net wealth, for each wave of the sample, for households experiencing financial fragility. We observe wide discrepancies in the behaviour of these variables. For instance, the income differential between the two groups of households (in distress/not in distress) is 1 to 2, whilst liabilities on average are in the ratio 1 to 3, real assets 1 to 130 and net wealth 1 to 100. Also, for households in distress, that is, with low levels of net worth, real wealth on average is equivalent to outstanding debt and, therefore, the positive value of net wealth is determined by small amounts of financial assets. For the other group of households, positive values of net worth are determined mainly by high and increasing over time property values. Finally, we observe large differentials in the dynamics of the variables. Whilst average income growth rates are similar for both groups of households, stock variables behave differently: liabilities, real assets and net worth are essentially stable over time for households in distress, whilst they grow considerably - and at a higher pace than income - for households not experiencing financial difficulties. Net wealth growth is driven by growth in real wealth, which is financed only partially by loans.

[11] For illustrative purposes descriptive statistics on the links between the dependent and the exogenous variables refer only to the unbalanced panel dataset.

Tab. 2	**Dependent variable, income, liabilities, assets and wealth** (euro 2006)				
		Income	Liabilities	Real assets	Net wealth
No distress	1998	33045	4789	205661	237023
	2000	34504	4800	224829	256797
	2002	35290	4807	246959	277696
	2004	35879	6191	292278	321527
	2006	36594	6481	309711	344979
Distress	1998	16477	1683	1994	3288
	2000	17376	1471	2121	3341
	2002	17035	2462	2784	3194
	2004	18795	2263	2286	3176
	2006	18301	1812	2397	3045

Tab. 3	**Dependent variable and dummies** (percentages)								
	Risky portfolio			Homeowner			Indebted		
	No	Yes	Tot.	No	Yes	Tot.	No	Yes	Tot.
No distress	82.3	17.7	100.0	12.6	87.4	100.0	78.5	21.5	100.0
Distress	93.8	6.2	100.0	97.6	2.4	100.0	82.4	17.6	100.0
Tot.	84.0	16.0	100.0	25.5	74.5	100.0	79.1	20.9	100.0

In addition to income as an indicator of the household's "ability to pay", we introduce three dummies, risky portfolio ownership, homeownership, and being indebted. These variables integrate the descriptive power of income in selecting households, which, for their general economic conditions and their portfolio composition, are less likely to incur financial distress. Tab. 3 relates these variables to the dependent variable. For the first two variables, we expect a negative value: the more diversified the portfolio the lower will be the probability of incurring financial distress in the event of an adverse shock; this is also true if the household head is the homeowner. Being indebted, on the other hand, contributes to increasing exposure to potential financial fragility, despite the average low levels of indebtedness in Italian households and the fact that amounts of debt are very similar for both sub-groups of households. However, the data show that the percentage of indebted households is higher amongst those without financial problems than amongst the other group. Therefore, it is difficult to formulate an ex-ante hypothesis on the sign of the dummy for "being indebted".

Households owning risky portfolios, homeowners and the indebted on average have higher incomes than the others (Tab. 4).

Tab. 4 **Average income and household characteristics**
(euro 2006)

	1998	2000	2002	2004	2006
No risky portfolio	26858	28476	30117	31034	31706
Risky portfolio	39119	38505	41470	44006	47174
Non-homeowner	20786	22285	24438	22395	22854
Homeowner	31507	32523	34763	37383	37253
Non-indebted	26438	27854	30441	30963	31749
Indebted	34801	38520	38654	43930	43205

In the household debt literature, age plays an important role in explaining the extent of debt and the probability of being indebted. It enters usually in a nonlinear fashion, in accordance with life cycle and consumption smoothing theories, which predict a concave age-debt profile, peaking around middle age. Descriptive statistics for the relationship between age and financial distress reveal a relatively higher percentages of younger (household head aged under 40) and older (household heads over 60) than middle aged households in distress.

The literature also suggests the inclusion of aggregate variables such as national or regional unemployment rates, house prices, interest rates and the income gearing ratio.[12] The rationale for including aggregate explanatory variables is that, being annual, they can capture time effects, and being disaggregated at the territorial level, they capture regional or wider area effects. Following May and Tudela (2005) we include the unemployment rate by geographical area, and regional house prices.[13] For the former it is possible to have an ex-ante opinion on its sign, with lower probabilities of financial distress for households living in areas of lower unemployment. The latter variable, real estate value, constitutes the net worth component, which, more than any other component, explains the differentials amongst individuals. As well as increasing over time, property values have been the driver of net wealth growth. Expectations about its sign diverge: on the one hand, an increase in house prices can have a dampening effect on non-homeowners and make it more difficult to access the property market, on the other hand, it will positively affect house owners. Overall its sign is not predetermined ex-ante.

[12] In the literature (see e.g. May and Tudela, 2005), two additional variables are considered, the income gearing ratio and the loan to value ratio. However their use in our context is problematic for two main reasons. May and Tudela focus on indebted households and therefore both variables, if not missing in the survey, are available for each observation. Our study sample instead covers the whole survey sample and includes households without debt and/or without real wealth: this implies a large number of missing values. In addition, there are many missing values even in the case of debt holdings and home ownership. For both these reasons we excluded these variables from the analysis.

[13] Source for house prices: Muzzicato et al. (2008).

Explanatory variables of the initial conditions equation. The initial conditions equation in the Heckman model is estimated with the explanatory variables set including the initial values x_{i0} of the structural equation and three additional dummy instrumental variables for if the household lives in the South of Italy, if the head of household is self-employed, and if the household resides in a municipality with less than 20000 inhabitants. Percentage distributions of these variables are depicted in Tab. 5.

Tab. 5 **Initial value variables (1998)**
(percentages)

	Living in the South			Self-employed			Municipality<20000 inhab.		
	No	Sì	Tot.	No	Sì	Tot.	No	Sì	Tot.
No distress	68.0	32.0	100.0	85.7	14.3	100.0	69.8	30.2	100.0
Distress	50.6	49.4	100.0	95.2	4.8	100.0	72.7	27.3	100.0
Tot.	65.0	35.0	100.0	87.3	12.7	100.0	70.3	29.7	100.0

5 Model estimates

Model estimates were run on both longitudinal samples described at the beginning of Section 4, the unbalanced panel (A) with $T \geq 3$ and the balanced panel (B) with $T = 5$.

Tab. 6 reports the estimates of the initial conditions equation and the structural equation for the Heckman model,[14] where the main parameter of interest is the coefficient of the lagged dependent variable, γ. After controlling for unobserved heterogeneity, we find evidence of true state dependence, that is, the probability of experiencing financial distress at time (t) positively depends upon the probability of having experienced financial fragility at time (t-1). The previous state parameter is equal to 0.563 and is statistically significant at the 95 percent level.[15]

The results confirm the presence of unobserved individual effects, with a value of the LR test on ρ of 29.39 (p-value=0.000). According to Arulampalam (1999), the fraction of total explained variance due to unobserved individual characteristics can be derived from ρ as follows:

$$\sigma_{\alpha}^{2} = \rho/(1-\rho)$$

In our case about 32 percent of the total variance is explained by unobserved household-level characteristics. Boheim and Taylor (2000) and May and Tudela (2005) find evidence of

[14] The model is estimated in Stata using two routines: “redprob” (Stewart, 2006, 2007) and “gllamm”. I want to thank Prof. Wiji Arulampalam for useful suggestions on the use of “gllamm”.

[15] Estimates run with *gllamm* and *redprob* produce equivalent results. For instance, the previous state coefficients coincide at the second decimal point, differing by only 0.001. The joint significance of the initial values is not rejected, with a Chi-sqared of 34.5 and p-value of 0.011.

unobserved heterogeneity, with values of respectively 34 and 19 percent. The size of this parameter shows the importance of individual components in the analysis of household financial problems, and the adequateness of the panel data.

Tab. 6 Dynamic models estimation with the Heckman method
(Panel A: T>=3 and common entry in 1998)

	Initial condition equation		Structural equation	
	Coef.	t-stat	Coef.	t-stat
constant	2.943	5.42	-0.917	-2.29
		Lagged response variable		
distress (t-1)			0.563	5.42
		Socio-economic explanatory variables		
young	0.426	2.43	0.092	0.72
old	-0.011	-0.06	0.107	0.96
1st income quartile	0.375	2.38	0.453	4.52
3rd income quartile	-0.466	-2.44	-0.322	-2.77
4th income quartile	-0.756	-3.19	-0.588	-4.20
education: primary	0.695	3.45	0.623	4.72
education: lower secondary	0.132	0.78	0.380	3.30
eduation: university	-0.787	-2.19	-0.447	-2.06
female	0.243	1.58	0.104	1.03
no. components	-0.002	-0.04	0.019	0.47
risky portfolio	-0.295	-1.31	-0.438	-3.30
homeowner	-3.377	-10.05	-3.258	-14.80
indebted	0.621	3.63	0.303	2.90
		Aggregate explanatory variables		
unemployment	0.110	1.47	0.03	3.440
house prices	-4.240	-1.32	0.12	0.390
		Other instruments for initial conditions		
south-isles			-0.719	-0.81
self-employed			-1.085	-4.57
small area			0.217	1.49
rho			0.244	3.30
theta			1.233	2.62
Log-likelihood			-1279.3	
LR test: rho=0			chi2(1) =	29.39
			p-value =	0.000
No. of observations	8619			

The age structure is not very strong: for the younger age group (household heads under 40) and the older age group (household heads over 60) the dummies are not significant.[16] However their coefficients are positive, which may suggest greater distress than in the intermediate age group 40-60 years old.

[16] In an alternative specification (not shown) which excludes education level, the oldest age group dummy is significant.

A crucial variable in our model is income, the main indicator of the ability of the household to pay. It enters the equation in the form of income quartiles, with the second one as the reference category; the results are as expected. Lower income households display a higher probability of financial distress, whilst higher income households display lower probability to get into difficulties (the first quartile coefficient is positive, while the third and fourth quartile coefficients are negative and increasing in absolute terms). In other of the models proposed in the literature financial fragility is associated with the income gearing ratio and, therefore, implicitly expresses an inverse relationship with income (May and Tudela, 2005). Our result is coherent with Boheim and Taylor's (2000) model where income has a negative sign. In line with these results, the probability of financial distress displays a negative relationship with education levels.

The dummy for risky portfolio is significant with a negative coefficient. As risky portfolios are owned by higher income households, this variable can reinforce the role of income in defining the "ability" of the household to pay, and indicates lower exposure to financial fragility. The female dummy and the dummy for household composition show a positive but not significant coefficient. The dummies for homeownership and being indebted have the expected signs: respectively negative and positive.

Turning to the aggregate variables, unemployment rate by geographical location is statistically significant and positive, denoting a higher probability of financial distress among households in areas of high unemployment. The regional house price index is not significant. Both results are in line with May and Tudela (2005).

Finally, the t-test on coefficient ϑ (the parameter that defines the presence of individual effects η_i correlated with α_i in the initial conditions equation and defined as $\eta_i = \vartheta\,\alpha_i + \varepsilon_{i0}$) rejects the null of non-exogeneity of initial conditions, with a Chi-squared of 109.9.

5.1 Comparisons with alternative estimation methods

In order to compare estimation methods we focus on the lagged dependent variable, the main variable of interest in dynamic models. We compare the Heckman model with the pooled, exogenous random effects and Wooldridge models. The results are reported in Tab. 7.

The coefficient of the previous state γ is larger in the case of the random effects model with exogenous initial conditions than in the Heckman model: 0.790 compared to 0.563. This result is in line with the literature review in Section 2.4 and shows that the hypothesis of exogenous initial conditions tends to overestimate state persistence. The coefficients of the other variables are of the same magnitude, sign and significance as in the Heckman model.

Tab. 7 **Dynamic models estimation with other methods**
(Panel A: T>=3 and common entry in 1998)

	Pooled model		Exogenous RE model		Wooldridge model	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
constant	-0.905	-2.55	-0.912	-2.40	-2.16	-2.000
			Lagged response variable			
distress (t-1)	0.881	12.67	0.790	8.77	0.325	2.59
			Socio-economic explanatory variables			
young	0.073	0.70	0.094	0.81	0.906	3.05
old	0.116	1.25	0.116	1.13	0.241	0.92
1st income quartile	0.383	4.54	0.426	4.5	0.232	1.69
3rd income quartile	-0.258	-2.58	-0.284	-2.59	-0.219	-1.40
4th income quartile	-0.475	-4.06	-0.533	-4.04	-0.474	-2.32
education: primary	0.480	4.76	0.544	4.54	-1.192	-1.76
education: lower secondary	0.289	3.23	0.328	3.15	-0.413	-1.06
eduation: university	-0.374	-2.11	-0.399	-2	-0.082	-0.12
female	0.057	0.72	0.070	0.77	0.064	0.56
no. components	0.016	0.48	0.021	0.57	-0.100	-0.90
risky portfolio	-0.353	-3.15	-0.406	-3.23	-0.502	-2.88
homeowner	-2.681	-26.02	-2.940	-15.61	-3.560	-13.34
indebted	0.269	2.96	0.285	2.88	0.222	1.65
			Aggregate explanatory variables			
unemployment	0.023	3.35	0.024	3.16	0.041	0.95
house prices	0.076	0.28	0.081	0.28	0.028	0.06
			Initial value			
distress (t=0)					0.781	5.55
			Group-means			
young					-0.992	-2.91
old					-0.110	-0.36
1°income quart.					0.456	1.97
3°income quart.					-0.131	-0.48
4°income quart.					-0.181	-0.60
education: primary					1.791	2.56
education: lower secondary					0.852	2.07
eduation: university					-0.218	-0.30
no. components					0.146	1.18
risky portfolio					0.146	0.50
homeowner					0.360	1.57
indebted					0.236	0.96
unemployment					-0.019	-0.44
house prices					0.876	0.81
sigma_alpha			0.388	3.20	0.683	6.07
rho			0.131	1.84	0.318	4.45
Log-likelihood	-907.1		-905.3		-869.2	
LR test: rho=0			chi2(1) =	3.63	chi2(1) =	19.65
			p-value =	0.028	p-value =	0.000
No. of observations: 8619						

The results of the estimations of the Wooldridge model are more problematic.[17] In contrast to the applied literature on comparative evaluations of the Heckman and Wooldridge methods (e.g. Stewart, 2007; Sousounis, 2008; Arulampalam and Stewart, 2009; Akay, 2009), there is

[17] In the Wooldridge model neither year dummies nor time-invariant variables (such as sex) can be included in the set of explanatory variables. As the model includes group means, the time-invariant variables are equivalent to their group means, which introduces collinearity problems. In our model, we include the variable sex in the control variable, but not in the group means.

a remarkable difference in the value of the state coefficient. The coefficient γ in the Wooldridge model is equal to 0.325 (Tab. 7) compared to 0.563 for the Heckman model. Although we estimated a range of specifications using both methods, differences persisted. This area is worthy of further examination, particularly the characteristics of the exogenous variables. It might be that, over time, some of the control variables present low variation and, therefore, a degree of correlation with the group means. The individual group means show low statistical significance, despite being jointly significant (Chi-squared=23.37). When we test for joint significance in the group means and initial values we are testing the validity of the structure of the unobserved heterogeneity, in line with Wooldridge: the Chi-squared rejects the null of non-significance with a statistics value of 49.25. Finally, the fraction of variance explained by individual effects is 46.6 percent, higher than the 32.2 percent obtained with the Heckman model.

The final comparison is with the pooled model. However, the pooled and the random effects models involve different normalisations of the error term. Normalisation for error variance in the pooled model is $\sigma_\varepsilon^2 = 1$, and in the random effects model it is $\sigma_u^2 = 1$. To allow comparison of the coefficients, those in the random effects model need be multiplied by $\sigma_u / \sigma_\varepsilon = \sqrt{1-\rho}$, where $\rho = \sigma_\alpha^2 / (\sigma_\alpha^2 + 1)$ is the constant cross-period error correlation (Arulampalam, 1999). Scaled coefficients of lagged financial distress are 0.736 in the exogenous random effects model, 0.490 in the Heckman model and 0.268 in the Wooldridge model. The pooled model produces a coefficient of 0.881.

5.2 Robustness analysis

As a robustness check, we also estimated the model on the balanced panel data structure with T = 5, and a total number of observations of 5,380, for a total of 1,076 households.[18] Again we estimate four models (Tab. 8): the pooled, exogenous initial conditions, Heckman and Wooldridge models.

The previous results also hold with this alternative data structure, although the coefficients of the lagged dependent variable are slightly larger in all the models: 0.970, 0.852, 0.402 and 0.636 in the pooled, with exogenous initial conditions, Wooldridge and Heckman models respectively. It should be noted, however, that there is a slight reduction in the gap between the previous state variables in the Heckman and Wooldridge models, although they are still relevant. In the Wooldridge model the variance explained by individual effects is 0.439, and in the Heckman model it is 0.326. In the random effects models the null $\rho = 0$ is rejected at

[18] We are aware of two problems: the first is attrition (Wooldridge, 2002a), the second the bias induced by extracting a balanced panel dataset from an unbalanced one (Verbeek, 2000: 343). Neither of these issues is dealt with here; they are left for future developments.

the standard significance levels. Finally the null of joint non-significance of the instruments in the Heckman model is also rejected (Chi-squared=56.28).

Tab. 4.8 Dynamic models estimation on the balanced panel
(Panel B: T=5)

	Pooled model		Exogenous RE model		Wooldridge model		Heckman model	
	Coef.	z	Coef.	z	Coef.	z	Coef.	z
constant	-1.14	-2.66	-1.154	-2.49	-2.095	-1.15	-1.185	-2.42
				Lagged response variable				
distress (t-1)	0.970	11.02	0.852	6.87	0.402	2.65	0.636	4.70
				Socio-economic explanatory variables				
young	0.251	1.86	0.314	1.97	1.137	3.04	0.337	1.97
old	0.119	1.03	0.110	0.84	0.212	0.69	0.117	0.83
1st income quartile	0.317	3.04	0.364	3.08	0.255	1.58	0.362	2.92
3rd income quartile	-0.187	-1.48	-0.207	-1.49	-0.011	-0.06	-0.239	-1.64
4th income quartile	-0.407	-2.76	-0.451	-2.74	-0.118	-0.48	-0.497	-2.87
education: primary	0.450	3.58	0.511	3.38	-1.760	-2.10	0.587	3.54
education: lower secondary	0.233	2.07	0.260	1.97	-0.772	-1.70	0.309	2.13
eduation: university	-0.299	-1.43	-0.327	-1.36	1.259	0.99	-0.369	-1.41
female	0.082	0.81	0.103	0.87	0.133	0.91	0.146	1.11
no. components	0.014	0.34	0.018	0.37	-0.245	-1.81	0.018	0.34
risky portfolio	-0.441	-3.10	-0.498	-3.10	-0.483	-2.31	-0.533	-3.19
homeowner	-2.589	-20.69	-2.873	-11.74	-3.354	-10.51	-3.167	-11.65
indebted	0.300	2.71	0.312	2.59	0.285	1.80	0.341	2.69
				Aggregate explanatory variables				
unemployment	0.033	3.64	0.036	3.40	0.054	1.13	0.042	3.63
house prices	0.178	0.57	0.198	0.59	0.089	0.16	0.232	0.67
				Initial value				
difficoltà (t=0)					0.807	4.64		
				Joint significance of group-means				
chi2(14)					25.71			
Prob > chi2					0.028			
				Joint significance of initial conditions				
chi2(18)							56.28	
Prob > chi2							0.000	
sigma_alpha			0.407	2.59	0.662	4.91	0.571	
rho			0.142	1.51	0.305	3.53	0.246	2.64
theta							1.312	2.01
Log-likelihood	-585.0		-573.7		-544.1		-757.7	
LR test: rho=0			chi2(1) =	2.45	chi2(1) =	12.4	chi2(1) =	19.1
			p-value =	0.059	p-value =	0.000	p-value =	0.000
No. of obs.: 5380								

As a final robustness check we ran the estimates on an unbalanced panel, with households present in the sample for at least three consecutive years, but relaxing the constraint of common entrance in 1998, for a total of 13,209 observations.[19] Estimated coefficients of the lagged dependent variable are in line with previous results: 0.868 for the exogenous conditions model, 0.664 for the Heckman and 0.372 for the Wooldridge model.

[19] The Heckman model is estimated using the gllamm procedure, which is more flexible than redprob for dealing with a longitudinal dataset with different entry times.

6 Conclusions and further developments

Many households are likely to experience periods of temporary financial distress over the years, and will overcome them with varying degrees of difficulty. Periods of financial stress become more relevant when the financial difficulties persist over time. This paper focused on households in financial distress and estimated the relevance of persistence over time of these situations.

We built a quantitative measure of financial distress based on combinations of assets and liabilities. It is the sign and dimension of net wealth rather than just debt levels that identify households experiencing financial fragility (as suggested in Barwell et al., 2006; Brown and Taylor, 2008). We built on the literature on households considered to be under financial stress to show that this occurs when they have negative net wealth holdings and the combination of real wealth and net financial balance (the difference between financial assets and liabilities) is below a certain threshold. This threshold is defined on the basis of net wealth distribution and we defined the threshold as the level of wealth corresponding to the second decile of the distribution. Households in financial distress, therefore, are defined as those whose net worth is equal to or below 13,000 euro. In the setting proposed by Brown and Taylor (2008), financial pressure is defined by the difference between total assets and liabilities: households with negative net wealth holdings are classified as being in financial distress. This definition clearly excludes households with small amounts of net wealth. In our framework we include all households with positive (even if small) net wealth holdings. This decision is motivated by the fact that the Italian dataset is characterised by very few observations with negative net wealth and that households with limited wealth holdings can experience financial difficulties.

The methodology used in this paper is estimation of dynamic nonlinear panel data models, where the coefficient of interest is the coefficient of the lagged dependent variable. Estimating a dynamic model is aimed at distinguishing between *true state dependence* – the impact of the lagged dependent variable on the dependent variable, and *spurious state dependence* - caused by the presence of time-invariant unobserved heterogeneity. This requires resolution of the so-called *initial conditions problem,* which arises from the fact that the observed start and the true start of the stochastic process do not coincide. We use the Heckman (1981b) as the standard parametric estimator for the probit model. We describe the econometric background to the estimation of panel data dynamic probit models, focusing first on Heckman's seminal work and then on developments proposed in the literature to overcome, or make tractable, the computational difficulties of maximising the likelihood function implied by the Heckman method (Orme, 2001; Wooldridge, 2005).

The empirical application uses the longitudinal component of the Bank of Italy SHIW for the period 1998-2006, and estimates a range of dynamic probit models (e.g. Stewart, 2007; Sousounis, 2008; Akay, 2009; Arulampalam and Stewart, 2009) to test for the presence of true state dependence in relation to experiencing financial distress: (1) the Heckman model; (2) the Wooldridge model; (3) a random effects probit model that assumes exogeneity of initial conditions; (4) a pooled model on repeated cross-sections that includes the lagged dependent variable, but ignores the presence of unobserved heterogeneity and assumes exogeneity of initial conditions.

From the Heckman model estimation we obtain a statistically significant coefficient of the lagged dependent variable of 0.563, implying the existence of true state dependence: the probability that households experiencing financial distress at time t is positively related to the probability of having experienced distress at time t-1. We do not reject the null of non-significance of unobserved heterogeneity, with the fraction of variance explained by individual unobserved effects of value 32.2 percent. This is in line with May and Tudela's (2005) result of 34 percent. We also reject the hypotheses of non-exogeneity of initial conditions and joint non-significance of the initial values. We can say, therefore, that there is true state dependence in experiencing financial fragility among Italian households. This result also identifies low levels of mobility along the net wealth distribution, particularly when we remember that the data are biannual. Hence, if a household's net wealth is below the threshold in 2004, it is probable that the same household will be below the threshold in 2006.

In terms of the other explanatory variables, "ability to pay" confirms our expectations: higher income, higher education and owning a risky portfolio, lower the probability of experiencing financial distress. Age is not very relevant, with most dummies not significant. There is evidence also that households with a female head have a higher probability of incurring financial fragility. For aggregate variables, higher unemployment positively affects the probability of distress, whereas house prices are not significant.

When we compare methods, the results of the pooled and random effects probit models are in line with the findings in the literature, with coefficients of the previous state showing higher values than the Heckman model, indicating that taking no account of unobserved heterogeneity or of exogenous initial conditions leads to overestimation of the coefficient of interest. However, we observe an unexplained difference between the Wooldridge and Heckman model estimates, which deserves further investigation as it is undocumented in the literature.

Robustness checks, consisting of estimation of the same set of models on two different panel data structures, confirm our results, and especially the relative positions of the four methodologies, with the less sophisticated methods providing over-estimations of state dependence.

In addition to examining the differences between the coefficients of the lagged dependent variable in the Wooldridge and Heckman models, two other aspects are worthy of further study. First, we should look at computations of the average partial effects (Wooldridge, 2005) in order to quantify state dependence. Second, we should estimate the Orme model and compare estimates with the results of other methods.

Bibliography

Akay, A. (2009), "The Wooldridge method for the initial values problem is simple: what about performance?", IZA Discussion Paper no. 3943.

Alessie, R., Hochguertel, S., van Soest, A. (2004), "Ownership of stocks and mutual funds: a panel data analysis," *Review of Economics and Statistics*, vol. 86, pp. 783-796.

Andrén, T. (2007), "The Persistence of Welfare Participation", School of Business, Economics and Law, Goteborg University, Working paper no. 266.

Arulampalam, W. (1999), "A note on estimated effects in random effect probit models", *Oxford Bulletin of Economics and Statistics*, vol. 61, pp. 597-602.

Arulampalam, W. and Bhalotra, S. (2006), "Sibling death clustering in India: state dependence versus unobserved heterogeneity", *Journal of the Royal Statistical Society, Series A*, vol. 169, part 4, pp. 829-848.

Arulampalam, W., Booth, A.L., Taylor, M.P. (2000), "Unemployment persistence", *Oxford Economic Papers*, vol. 52, pp. 24-50.

Arulampalam, W. and Stewart, M.B. (2008), "Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators", Warwick Research Papers no. 884.

Arulampalam, W. and Stewart, M.B. (2009), "Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators", *Oxford Bulletin of Economics and Statistics*, vol. 71, no. 5, pp. 659-681.

Barwell, R., May, O., Pezzini, S. (2006), "The distribution of assets, income and liabilities across UK households: results from the 2005 NMG Research survey", *Bank of England Quarterly Bulletin*, Spring.

Benito, A. and Young, G. (2003), "Hard times or great expectations? Dividend omissions and dividend cuts by UK firms", *Oxford Bulletin of Economics and Statistics*, vol. 65, pp. 531-555.

Boeri, T. and Brandolini, A. (2005), "The age of discontent: Italian households at the beginning of the decade", IZA Discussion Paper no. 1530.

Boheim, R. and Ermisch, J. (2001), "Partnership Dissolution in the UK – the Role of Economic Circumstances", *Oxford Bulletin of Economics and Statistics*, vol. 63, no. 2, pp. 197-208.

Boheim, R. and Taylor, M. P. (2000), "My home was my castle: evictions and repossessions in Britain", *Journal of Housing Economics*, vol. 9, pp. 287-319.

Boushey, H. and Weller, C.E. (2008), "Has growing inequality contributed to rising household economic distress?", *Review of Political Economy*, vol. 20, no. 1, pp. 1-22.

Brown, S. and Taylor, K. (2008), "Household debt and financial assets: evidence from Germany, Great Britain and the USA", *Journal of the Royal Statistical Society Series A*, vol. 171, part 3, pp. 615–643.

Cappellari, L. and Jenkins, S.P. (2009), "The dynamics of social assistance receipt: Measurement and modelling issues, with an application to Britain", *ISER Essex*, Working Paper no. 2008-34.

Chamberlain, G. (1984), "Panel data", in Griliches, Z., Intrilligator, M. (eds.), *Handbook of Econometrics*, North-Holland: Amsterdam.

Chauvin, N. and Kraay, A. (2007), "Who gets debt relief?", *Journal of the European Economic Association*, vol. 5, pp. 333-342.

Chen, J. and Entrom Ost, C. (2005), "Housing allowance and the recipient's homeownership: Evidence from a panel data study in Sweden", *Housing Studies*, vol. 20, pp. 605-625.

Clark A.E. and Etilé, F. (2006), "Don't give up on me baby: Spousal correlation in smoking behaviour", *Journal of Health Economics*, vol. 25, pp. 958-978.

Contoyannis, P., Jones, A.M., Rice, N. (2004), "The dynamics of health in the British Household Panel Survey", *Journal of Applied Econometrics*, vol. 19, pp. 473-503.

Cox, D. and Jappelli, T. (1993), "The effect of borrowing constraints on consumer liabilities", *Journal of Credit and Banking*, vol. 25, no. 2, pp. 197-213.

Cox, P., Whitley, J., Brierley, P. (2002), "Financial pressure in the UK household sector: evidence from the British Household Panel Survey", *Bank of England Quarterly Bulletin*, vol. 42, no. 4, pp. 410-419.

Crook, J. (2001), "The demand for household debt in the USA: evidence from the 1995 Survey of Consumer Finance", *Applied Financial Economics*, vol. 11, no. 1, pp. 83-91.

Crook, J. and Hochguertel, S. (2007), "US and European household debt and credit constraints, Tinbergen Institute, Discussion paper n. TI 2007-087/3.

del Rio, A. and Young, G. (2008), "The impact of unsecured debt on financial pressure among British households", *Applied Financial Economics*, vol. 18, no. 15, pp. 1209-1220.

Dorsett, R. (1999), “An econometric analysis of smoking prevalence among lone mothers”, *Journal of Health Economics*, vol. 18, pp. 429-441.

Duca, J. V., Rosenthal, S. S. (1993), “Borrowing Constraints, Household Debt, and Racial Discrimination in Loan Markets”, *Journal of Financial Intermediation*, vol. 3:1, pp. 77-103.

Grilli, L. and Rampichini, C. (2005), “A review of random effects modelling using *gllamm* in Stata”, downloadable from: http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-software/gllamm.shtml.

Hall, R. E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence", *Journal of Political Economy*, December, vol. 86, pp. 971-87.

Heckman, J. (1981a), “Heterogeneity and state dependence”, in S. Rosen (ed.), *Studies in Labor Markets*, Chicago Press: Chicago, IL.

Heckman, J. (1981b), “The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process”, in Manski, C.F., McFadden, D. (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press: Cambridge, MA; pp. 179-195.

Henley, A. (2004), “Self-employment status: The role of state dependence and initial circumstances”, *Small Business Economics*, vol. 22, pp. 67-82.

Honoré, B. E. (1993), "Orthogonality conditions for tobit models with fixed effects and lagged dependent variables", *Journal of Econometrics*, vol. 59, pp. 35-61.

Honoré, B. E., and Kyriazidou, E. (2000), "Panel data discrete choice models with lagged dependent variables", *Econometrica*, vol. 68, pp. 839-874

Hsiao, C. (2003), *Analysis of panel data*, Second edition, Cambridge University Press.

Hyslop, D.E. (1999), “State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women”, *Econometrica*, vol. 67, pp. 1255-1294.

Lee, W. and Oguzoglu, U. (2007), “Income support and stigma effects for young Australians”, *Australian Economic Review*, vol. 40, pp. 369-384.

Lillard L. A. and Willis, R.J. (1978), “Dynamic aspects of earnings mobility”, *Econometrica*, vol. 46, pp. 985-1012.

Loudermilk, MS. (2007), “Estimation of fractional dependent variables in dynamic panel data models with an application to firm dividend policy”, *Journal of Business and Economic Statistics*, vol. 25, pp. 462-472.

Magri, S. (2007), “Italian households’ debt: the participation to the debt market and the size of the loan”, *Empirical Economics*, vol. 33, pp. 401-426.

May, O. and Tudela, M. (2005), “When is mortgage indebtedness a financial burden to British households? A dynamic probit approach”, Bank of England, Working paper no. 277.

May, O., Tudela, M., Young, G. (2004), “British household indebtedness and financial stress: a household level picture”, *Bank of England Quarterly Bulletin*, no. 44: 4, pp. 414-428.

Mundlak, Y. (1978), “On the pooling of time series and cross section data”, *Econometrica*, vol. 46, pp. 69-85.

Muzzicato, S., Sabbatini, R., Zollino, F. (2008), "Prices of residential property in Italy: constructing a new indicator", Questioni di Economia e Finanza no. 17, Bank of Italy.

Narazani E. (2009), “Labour and capital adjustment dynamics using bivariate random probit”, Working paper ChilD no. 07/2009.

Orme, C.D. (1997), “The Initial Conditions Problem and Two-Step estimation in Discrete Panel Data Models”, mimeo, Department of Economic Studies, University of Manchester, UK. Non più reperibile.

Orme, C.D. (2001), “Two-step inference in dynamic non-linear panel data models”, School of Economic Studies, University of Manchester, mimeo, downloadable from: http://personalpages.manchester.ac.uk/staff/chris.orme/

Propper, C. (2000), “The demand for private health care in the UK”, *Journal of Health Economics*, vol. 19, pp. 855-876.

Pudney, S. (2008), “The dynamics of perception: Modelling subjective wellbeing in a short panel”, *Journal of the Royal Statistical Society, Series A*, vol. 171, pp. 21-40.

Rabe-Hesketh, S. and Skrondal, A. (2005), *Multilevel and longitudinal modeling using Stata*, Stata Press, StataCorp LP, College Station, Texas.

Requena-Silvente, F. (2005), “The decision to enter and exit foreign markets: Evidence from U.K. SMEs”, *Small Business Economics*, vol. 25, pp. 237-53.

Sousounis, P. (2008), “State dependence in work-related training participation among British employees: A comparison of different random effects probit estimators”, MPRA Paper no. 14261.

Stewart, M.B. (2006), “-redprob- A Stata program for the Heckman estimator of the random effects dynamic probit model”, downloadable from: http://www2.warwick.ac.uk/fac/soc/economics/staff/academic/stewart/stata/redprobnote.pdf

Stewart, M.B. (2007), “Inter-related dynamics of unemployment and low wage employment”, *Journal of Applied Econometrics*, vol. 22, pp. 511-531.

Wooldridge, J. M. (2002a), “Inverse probability weighted M-estimators for sample stratification, attrition and stratification”, *Portuguese Economic Journal*, vol. 1, pp. 117-139.

Wooldridge, J. M. (2002b), “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity”, The Institute for Fiscal Studies, Dept. of Economics, UCL, cemmap working paper no. 18/02.

Wooldridge, J. M. (2005), “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity”, *Journal of Applied Econometrics*, vol. 20, pp. 39–54.